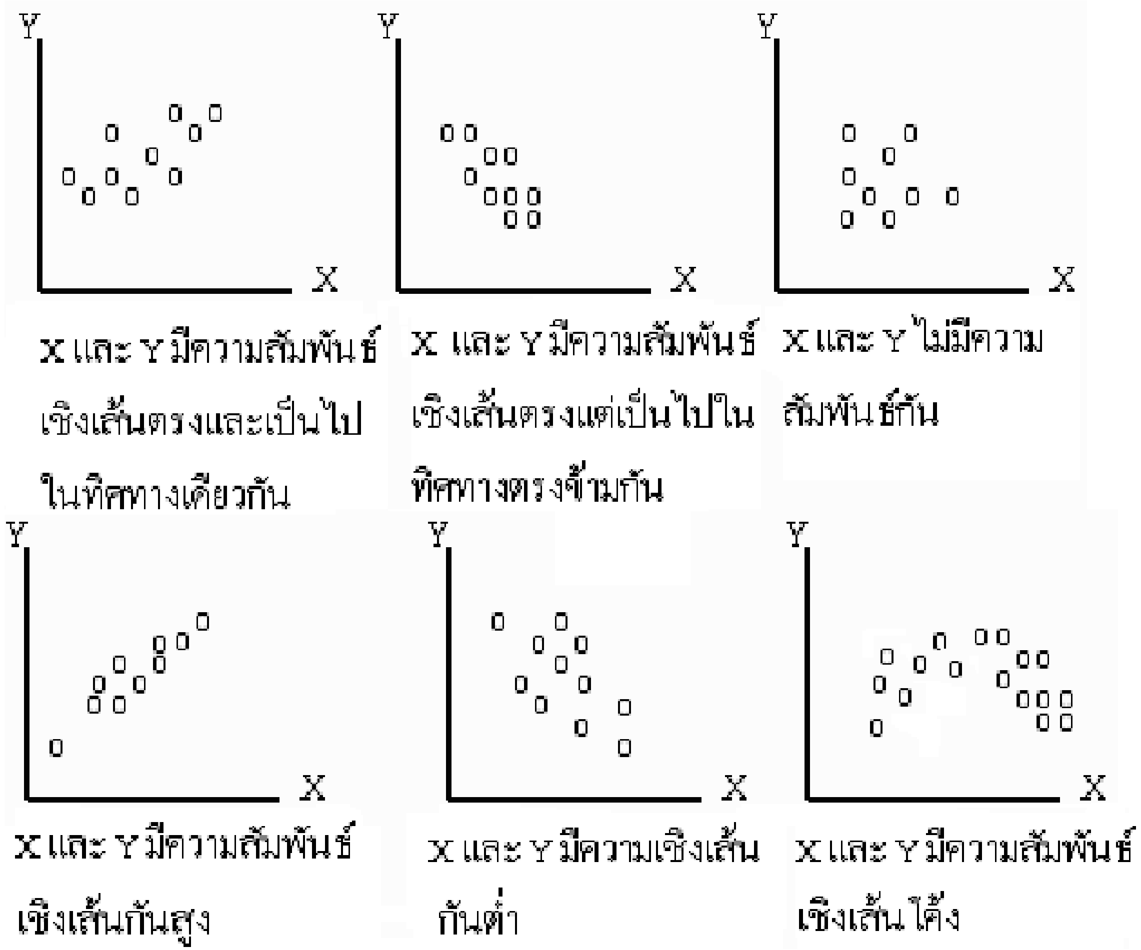
chapter 10

Simple Linear Regression

and Correlation Analysis

1. Plot scatter plot เพื่อดูความสัมพันธ์ของตัวแปร X และตัวแปร Y



รูปที่ 10.1 ตัวอย่างของแผนภาพการกระจายแบบต่างๆ

2. หาตัวแบบความสัมพันธ์ โดยวิธี OLS

$$Y = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; i = 1, 2, n \qquad (10.1)$$

โดย Y_i คือ ตัวแปรตาม (Dependent variable)

X_i คือ ตัวแปรอิสระ (Independent variable)

β_0 คือ ระยะตัดแกน Y หรือ Y - intercept

β_1 คือ ความชันของเส้นถดถอย ซึ่งเป็นค่าที่แสดงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนไป 1 หน่วย

3. หาตัวประมาณแบบจุด

3.1 ค่า พารามิเตอร์ $\beta_1 = b_1$ และ $\beta_0 = b_0$

$$b_0 = \overline{Y} - b_1 \overline{X}$$

$$b_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2} \qquad (10.10)$$

ตัวอย่าง 1 ฝ่ายวิจัยของบริษัทแห่งหนึ่ง ต้องการหาความสัมพันธ์ระหว่างค่าโฆษณากับยอดขาย เพื่อพยากรณ์ยอดขายเดือนหน้า ถ้ากำหนดค่าโฆษณาในเดือนหน้าเป็น 450,000 บาท จึงใช้ข้อมูลค่าโฆษณาและยอดขายรายเดือนของปีที่ผ่านมาดังนี้

เดือนที่	โฆษณา (แสนบาท): X	ยอดขาย (ล้านบาท): Y
1	0.8	22
2	1.0	28
3	1.6	22
4	2.0	26
5	2.2	34
6	2.6	18
7	3.0	30
8	3.0	38
9	4.0	30
10	4.0	40
11	4.0	50
12	4.6	60

กำหนดให้

$$\sum X = 32.8$$

$$\sum Y = 398$$

$$\sum X^2 = 106.96$$

$$\sum Y^2 = 14,852$$

$$\sum XY = 1214.4$$

วิธีทำ จาก ตัวอย่างที่ 1

$$\sum X = 32.8$$
$$\sum Y = 398$$
$$\sum X^2 = 106.96$$
$$\sum Y^2 = 14,852$$
$$\sum XY = 1214.4$$

1. $$b_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}$$

2. $$b_0 = \overline{Y} - b_1 \overline{X}$$

- สมการพยากรณ์คือ
- ถ้างบโฆษณาเป็น 450,000 คือ 4.5 แสนบาท
- อธิบายความหมาย b_0
- อธิบายความหมาย b_1

3.2 ค่าความแปรปรวน σ^2 => ทำการประมาณโดยค่าความแปรปรวนของตัวอย่าง

$$S^2 = \frac{\sum Y_i^2 - b_0 \sum Y_i - b_1 \sum X_i Y_i}{n - 2}$$

ดังนั้นจะได้ความคลาดเคลื่อนมาตรฐานในการประมาณค่าคือ $s = \sqrt{s^2}$

4. หาตัวประมาณแบบช่วง

4.1 ค่า พารามิเตอร์ β_1 และ β_0

ช่วงความเชื่อมั่นของ $(1-\alpha)100\%$ ของ β_0 คือ

$$b_0 \pm t_{\alpha/2,(n-2)} S_{b_0}$$ เมื่อ $$S_{b_0} = \sqrt{\frac{s^2 \sum X_i^2}{n(\sum X_i^2 - n\bar{X}^2)}}$$

ช่วงความเชื่อมั่นของ $(1-\alpha)100\%$ ของ β_1 คือ

$$b_1 \pm t_{\alpha/2,(n-2)} S_{b_1}$$ เมื่อ $$S_{b_1} = \sqrt{\frac{s^2}{(\sum X_i^2 - n\bar{X}^2)}}$$

5. ทดสอบสมมติฐานของพารามิเตอร์ β_1 และ β_0 ว่าควรนำไปใช้ประมาณสมการถดถอยหรือไม่ สามารถทำได้ 2 วิธี คือ ใช้ตัวสถิติ t หรือใช้ ANOVA

$H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

$$t = \frac{b_0 - \beta_0}{S_{b_0}} = \frac{b_0}{S_{b_0}} \quad ; df = n-2$$

ปฏิเสธ H_0 $t < -t_{\alpha, n-2}$ หรือ $t > t_{\alpha, n-2}$

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$$t = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{b_1}{S_{b_1}} \quad ; df = n-2$$

ปฏิเสธ H_0 $t < -t_{\alpha, n-2}$ หรือ $t > t_{\alpha, n-2}$

ตัวอย่างที่ 2 จากตัวอย่าง 1 จงทำการทดสอบว่าค่าโฆษณากับยอดขาย มีความสัมพันธ์กันหรือไม่ที่ระดับนัยสำคัญ 0.05 และหาช่วงเชื่อมั่น 95% ของ β_1

$$\sum X = 32.8$$
$$\sum Y = 398$$
$$\sum X^2 = 106.96$$
$$\sum Y^2 = 14,852$$
$$\sum XY = 1214.4$$

$$b_1 \pm t_{\alpha/2,(n-2)} S_{b_1}$$

$$S_{b_1} = \sqrt{\frac{s^2}{(\sum X_i^2 - n\overline{X}^2)}}$$

หาความคลาดเคลื่อนมาตรฐานการประมาณค่า $s = \sqrt{s^2}$

$$S^2 = \frac{\sum Y_i^2 - b_0 \sum Y_i - b_1 \sum X_i Y_i}{n-2}$$

ตัวอย่างที่ 2 (ต่อ) จากตัวอย่าง 1 หาช่วงเชื่อมั่น 95% ของ β_1

$$b_1 \pm t_{\alpha/2,(n-2)} S_{b_1}$$

$$S_{b_1} = \sqrt{\frac{s^2}{(\sum X_i^2 - n\overline{X}^2)}}$$

$$\sum X = 32.8$$
$$\sum Y = 398$$
$$\sum X^2 = 106.96$$
$$\sum Y^2 = 14{,}852$$
$$\sum XY = 1214.4$$

6. ช่วงเชื่อมั่นสำหรับค่าเฉลี่ยและค่าพยากรณ์

นำมาใช้ในการประมาณค่าแบบช่วงได้ 2 แบบคือ การประมาณค่าเฉลี่ย(mean value)ของ Y ทั้งหมด $(\mu_{Y.X})$ เมื่อกำหนดค่า X มาให้ เช่นกำหนด $X=X_0$ และการประมาณค่า Y แต่ละค่า (Individual value) เมื่อกำหนดค่า $X=X_0$

ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของ Y เมื่อกำหนดค่า $X = X_0$

ช่วงความเชื่อมั่น (1-α)100% ของ $\mu_{Y.X}$ เมื่อกำหนดค่า $X = X_0$ คือ

$$\hat{Y} \pm t_{\alpha/2,\,n-2}\, S_{\hat{Y}_0} \qquad (10.26)$$

เมื่อ $S^2_{\hat{Y}_0} = S^2\left[\frac{1}{n} + \frac{(X_0-\overline{X})^2}{\sum X_i^2 - n\overline{X}^2}\right]$ ดังนั้น $S_{\hat{Y}_0} = \sqrt{S^2\left[\frac{1}{n} + \frac{(X_0-\overline{X})^2}{\sum X_i^2 - n\overline{X}^2}\right]}$

ช่วงความเชื่อมั่นสำหรับค่าพยากรณ์ของ Y เมื่อกำหนดค่า $X = X_0$

ช่วงความเชื่อมั่น (1-α)100% ของ Y เมื่อกำหนดค่า $X = X_0$ คือ

$$\hat{Y} \pm t_{\alpha/2,\,n-2}\, S_{Y_0} \qquad (10.27)$$

เมื่อ $S_{Y_0}^2 = S^2 + S_{\hat{Y}_0}^2$ ฉะนั้น $S_{Y_0} = \sqrt{S^2\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X_i^2 - n\bar{X}^2}\right]}$

ตัวอย่างที่ 3 จากตัวอย่าง 1 ถ้าผู้จัดการต้องการ**ประมาณช่วงเชื่อมั่นของยอดขายเฉลี่ย** เมื่อกำหนดให้ค่าโฆษณาเป็น 250,000 บาท ที่ระดับความเชื่อมั่น 95%

$$S^2 = \frac{\sum Y_i^2 - b_0 \sum Y_i - b_1 \sum X_i Y_i}{n-2}$$

$$S_{\hat{Y}_0} = \sqrt{S^2\left[\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum X_i^2 - n\overline{X}^2}\right]}$$

$$\hat{Y} \pm t_{\alpha/2,\, n-2}\, S_{\hat{Y}_0}$$

ตัวอย่างที่ 4 จากตัวอย่าง 1 ถ้าผู้จัดการต้องการ**ประมาณช่วงเชื่อมั่นค่าพยากรณ์ยอดขาย** เมื่อกำหนดให้ค่าโฆษณาเป็น 250,000 บาท ที่ระดับความเชื่อมั่น 95%

$$\hat{Y} \pm t_{\alpha/2,n-2} S_{Y_0}$$

$$S_{Y_0} = \sqrt{S^2\left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum X_i^2 - n\overline{X}^2}\right]} \longrightarrow S_{Y_0}^2 = S^2 + S_{\hat{Y}_0}^2$$

สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination) (R^2)

$$R^2 = \frac{(n\sum X_iY_i - (\sum X_i)(\sum Y_i))^2}{(n\sum X_i^2 - (\sum X_i)^2)(n\sum Y_i^2 - (\sum Y_i)^2)} \qquad (0 \le R^2 \le 1)$$

- ถ้า R^2 มีค่าเข้าใกล้ 1 หมายความว่า X สามารถอธิบายการเปลี่ยนแปลงของ Y ได้ดี เนื่องจาก X และ Y มีความสัมพันธ์กันมาก หรือสมการถดถอยสามารถอธิบายค่าของตัวแปรตามได้ดี
- ถ้า R^2 มีค่าเข้าใกล้ 0 หมายความว่า X สามารถอธิบายการเปลี่ยนแปลงของ Y ได้น้อย เนื่องจาก X และ Y มีความสัมพันธ์กันน้อย หรือสมการถดถอยอธิบายค่าของตัวแปรตามได้ไม่ดี

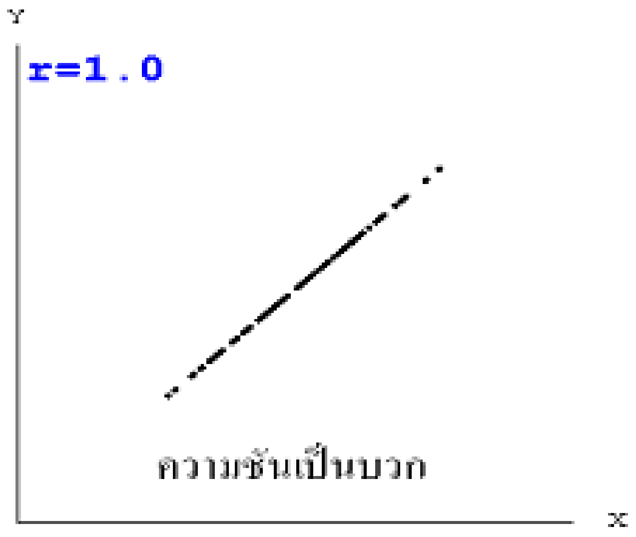
การวิเคราะห์สหสัมพันธ์เชิงเส้นอย่างง่าย

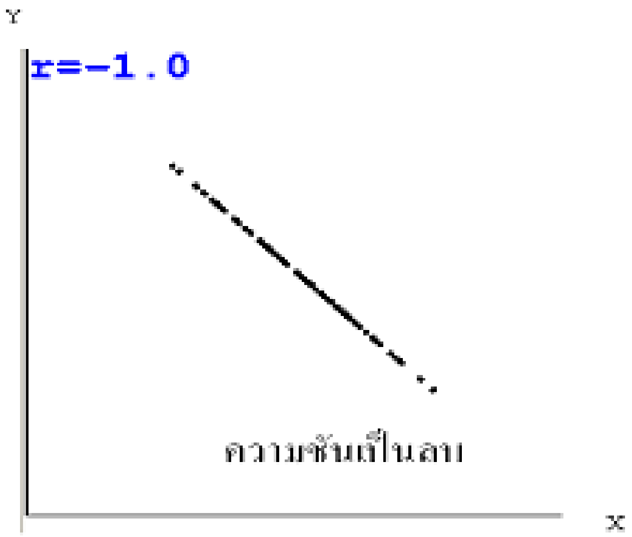
สัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) เป็นค่าที่ใช้วัดความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปรใดๆ ว่ามีความสัมพันธ์กันมากน้อยแค่ไหน สัมประสิทธิ์สหสัมพันธ์ของประชากร (Population correlation coefficient) เขียนแทนด้วย ρ ในกรณีที่ใช้ข้อมูลตัวอย่างจะประมาณค่าของ ρ ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง (Sample correlation coefficient) เขียนแทนด้วย r ซึ่งมีสูตรการคำนวณดังต่อไปนี้

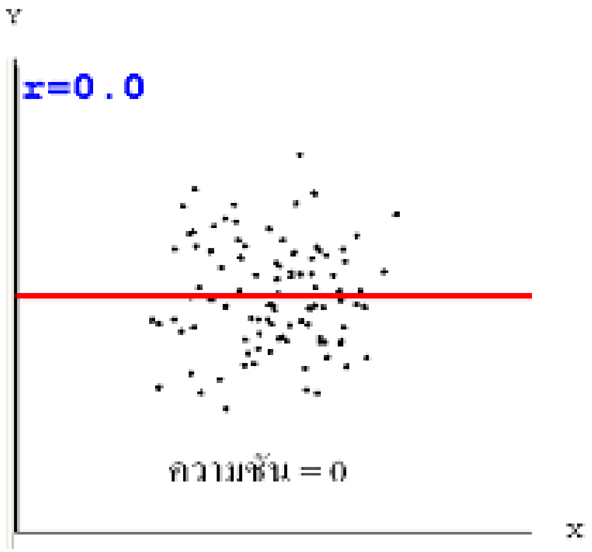
$$\boxed{r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}} \qquad (10.29)$$

หรือ

$$\boxed{r = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{(n(\sum X_i^2) - (\sum X_i)^2)(n(\sum Y_i^2) - (\sum Y_i)^2)}}} \qquad (10.30)$$







โดยค่า r จะมีค่าตั้งแต่ -1 ถึง 1 ($-1 \leq r \leq 1$) และมีความหมายดังต่อไปนี้

- ถ้า r มีค่าเข้าเป็นบวก หมายความว่า X และ Y มีความสัมพันธ์ในทิศทางเดียวกัน
- ถ้า r มีค่าเป็นลบ หมายความว่า X และ Y มีความสัมพันธ์ในทิศทางตรงข้ามกัน
- ถ้า r มีค่าเข้าใกล้ 1 หมายความว่า X และ Y มีความสัมพันธ์ในทิศทางเดียวกันและมีความสัมพันธ์กันมาก
- ถ้า r มีค่าเข้าใกล้ -1 หมายความว่า X และ Y มีความสัมพันธ์ในทิศทางตรงข้ามกันและมีความสัมพันธ์กันมาก
- ถ้า r มีค่าเข้าใกล้ 0 หมายความว่า X และ Y *มีความสัมพันธ์กันน้อย*

การทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์ ρ

เป็นการทดสอบว่าตัวแปรอิสระ (X) และตัวแปรตาม (Y) มีความสัมพันธ์กันหรือไม่

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

ตัวสถิติทดสอบ คือ $t = \dfrac{(r - \rho)}{S_r}$; $df = n - 2$

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

ปฏิเสธ H_0 $t < -t_{\alpha, n-2}$ หรือ $t > t_{\alpha, n-2}$

ตัวอย่าง 5 จากตัวอย่าง 1 จงทดสอบเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์เพื่อตรวจสอบความสัมพันธ์ระหว่างค่าโฆษณาและยอดขาย

$$r = \frac{n\sum X_iY_i - (\sum X_i)(\sum Y_i)}{\sqrt{(n(\sum X_i^2) - (\sum X_i)^2)(n(\sum Y_i^2) - (\sum Y_i)^2)}}$$

$\sum X = 32.8$

$\sum Y = 398$

$\sum X^2 = 106.96$

$\sum Y^2 = 14,852$

$\sum XY = 1214.4$

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

ตัวสถิติทดสอบ คือ $t = \dfrac{(r - \rho)}{S_r}$

การทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์ในสมการถดถอยโดยใช้การวิเคราะห์ความแปรปรวน

การทดสอบสมมติฐานเกี่ยวกับค่า β_1 ซึ่งเป็นการทดสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปร X และ Y นอกจากจะใช้ตัวสถิติทดสอบ t แล้ว ยังสามารถใช้หลักการของการวิเคราะห์ความแปรปรวนมาทดสอบได้อีกด้วย โดยการพิจารณาแยกความผันแปรหรือความแปรปรวนทั้งหมดของตัวแปรสุ่ม Y ออกเป็นส่วนๆ ตามแหล่งที่มาของความผันแปร แสดงดังต่อไปนี้

$$SST = SSR + SSE \qquad (10.22)$$

เมื่อ SST คือค่าความผันแปรทั้งหมดของ Y (Total Sum of Squares)

SSR คือผลบวกกำลังสองของการถดถอย (Regression Sum of Squares) หรือความผันแปรที่สามารถอธิบายค่าได้ด้วยเส้นถดถอย (Explained Variation)

SSE คือผลบวกกำลังสองของความคลาดเคลื่อน (Error Sum of Squares)

$$SST = \sum (Y_i - \bar{Y})^2 = \sum Y^2 - n\bar{Y}^2$$

$$= \sum Y^2 - \frac{(\sum Y)^2}{n} = SYY \qquad (10.23)$$

$$SSR = \sum (\hat{Y}_i - \bar{Y}_i)^2 = b_1 \left(\sum XY - n\bar{X}\,\bar{Y} \right) = b_1 SXY \qquad (10.24)$$

$$= b_1^2 \left(\sum X^2 - n\bar{X}^2 \right) \qquad (10.24ก)$$

$$SSE = \sum (Y_i - \hat{Y}_i)^2 = SST - SSR \qquad (10.25)$$

สมมติฐานการทดสอบ คือ

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

ตาราง ANOVA

SOV	*df*	*SS*	*MS*	F_0
Regression	*1*	*SSR*	$MSR = \frac{SSR}{1}$	$\frac{MSR}{MSE}$
Error	$n-2$	*SSE*	$MSE = \frac{SSE}{(n-2)}$	
Total	$n-1$	*SST*		

จะปฏิเสธ $H_0 : \beta_1 = 0$ ถ้า

F_0 ในตาราง ANOVA > $F_{\alpha,(1,n-2)}$

จากตัวอย่าง 10.1

ผลการวิเคราะห์

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.980026
R Square	0.960451
Adjusted R Square	0.955508
Standard Error	2.27572
Observations	10

Multiple R =0.980026 คือค่าสหสัมพันธ์ (r) แต่ทิศทาง + หรือ -

พิจารณาจากค่า สัมประสิทธิ์ของ X หรือค่า β_1 จากตารางด้านล่าง

ค่า R Square หรือ $= r^2 = 0.960451$ ค่า Standard Error $= \sigma$

$\sqrt{MSE} = \sqrt{5.178904} = 2.27572$

ค่า Observations คือจำนวนค่าสังเกตุ มีค่าเท่ากับ 10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F* หรือ ค่า *Sig.*
Regression	1	1006.169	1006.169	194.2822	6.8E-07
Residual	8	41.43123	5.178904		
Total	9	1047.6			

ค่า significance F หรือ ค่า Sig. หรือค่า p-value ไว้พิจารณาว่าจะปฏิเสธ หรือยอมรับสมมติฐาน H_0 โดยจะปฏิเสธสมมติฐานเมื่อ ค่า sig. หรือค่า p-value น้อยกว่าหรือเท่ากับ ระดับนัยสำคัญ α

ดังในตัวอย่างนี้กำหนดระดับนัยสำคัญ 0.05 (α=0.05) ดังนั้นใน ANOVA จึง ปฏิเสธสมมติฐาน H_0

	Coefficients (ค่าสัมประสิทธิ์)	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	154.725	2.373452	65.18987	3.41E-12	149.2519	160.1982
X	-6.58143	0.472176	-13.9385	6.8E-07	-7.67027	-5.49259

$\beta_0 = 154.725$ ส่วนเบี่ยงเบนมาตรฐาน $S_{\beta_0} = 2.373$

สามารถหาช่วงความเชื่อมั่นคือของ β_0 มีค่าเท่ากับ (149.2519, 160.1982)

$H_0 : \beta_0 = 0$ $H_1 : \beta_0 \neq 0$ พบว่า p-value=3.41×10^{-12} ดังนั้นจึงปฏิเสธสมมติฐาน H_0

$\beta_1 = -6.58143$ ส่วนเบี่ยงเบนมาตรฐาน $S_{\beta_1} = 0.472176$

สามารถหาช่วงความเชื่อมั่นคือของ β_1 มีค่าเท่ากับ (-7.67027, -5.49259)

$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$ พบว่า p-value =6.8×10^{-7} ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ตัวแปร X มีความสัมพันธ์กับตัวแปร Y และมีความสัมพันธ์ในทิศทางตรงกันข้าม

ตัวอย่าง 6 (อ่าน output) จากการศึกษาคุณภาพยางรถยนต์รุ่นหนึ่งเมื่อรถทำการเบรคกระทันหัน โดยเก็บข้อมูลความเร็วก่อนเหยียบเบรคกับระยะที่รถไถล (หน่วย: เซนติเมตร) ได้ข้อมูลดังตาราง

ครั้งที่	X: ความเร็ว (km/hr.)	Y: ระยะรถไถล (cm)
1	76	49.3774
2	89	49.3798
3	100.5	49.3799
4	101	49.3835
5	103	55.3876
6	104	55.4386
7	112	55.4413
8	122.5	56.5496
9	150	62.0461

ก. จงหาค่า b_0 และ b_1

ข. สมการถดถอยคือ

ค. จงอธิบายความหมายของค่า b_0 และ b_1 ที่คำนวณได้

ง. จงประมาณช่วงเชื่อมั่น 95 % ของ β_1

จ. จงหาค่าประมาณระยะไถลถ้ารถวิ่งด้วยความเร็ว 130 km/hr

ฉ. จงหาช่วงเชื่อมั่น 95% ของระยะไถลถ้ารถวิ่งด้วยความเร็ว 130 km/hr

กำหนดให้ $\sum X = 958$, $\sum X^2 = 105,473.5$, $\sum Y = 482.38$, $\sum Y^2 = 26,016.14$, $\sum XY = 52,012.10$

Regression Statistics	
Multiple R	0.885
R Square	0.784
Adjusted R Square	0.753
Standard Error	2.232
Observations	9.000

ANOVA					
	df	SS	MS	F	Significance F
Regression	1.000	126.370	126.370	25.374	0.002
Residual	7.000	34.862	4.980		
Total	8.000	161.232			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Intercept	33.371	4.084	8.172	0.000	23.715	43.028	19.080	47.662
X	0.190	0.038	5.037	0.002	0.101	0.279	0.058	0.322

ก. จงหาค่า b_0 และ b_1

ข. สมการถดถอยคือ

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Intercept	33.371	4.084	8.172	0.000	23.715	43.028	19.080	47.662
X	0.190	0.038	5.037	0.002	0.101	0.279	0.058	0.322

ค. จงอธิบายความหมายของค่า b_0 และ b_1 ที่คำนวณได้

ง. จงประมาณช่วงเชื่อมั่น 95 % ของ β_1

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Intercept	33.371	4.084	8.172	0.000	23.715	43.028	19.080	47.662
X	0.190	0.038	5.037	0.002	0.101	0.279	0.058	0.322

จ. จงหาค่าประมาณระยะไถลถ้ารถวิ่งด้วยความเร็ว 130 km/hr

ฉ. จงหาช่วงเชื่อมั่น 95% ของระยะไถลถ้ารถวิ่งด้วยความเร็ว 130 km/hr

$$S_{Y_0} = \sqrt{S^2\left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum X_i^2 - n\overline{X}^2}\right]}$$

Regression Statistics	
Multiple R	0.885
R Square	0.784
Adjusted R Square	0.753
Standard Error	2.232
Observations	9.000

ANOVA					
	df	SS	MS	F	Significance F
Regression	1.000	126.370	126.370	25.374	0.002
Residual	7.000	34.862	4.980		
Total	8.000	161.232			

ช. ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากับ

ฉ. สัมประสิทธิ์การตัดสินใจมีค่าเท่ากับ

ณ. ความคลาดเคลื่อนในการประมาณค่าสมการถดถอยมีค่าเท่ากับ

ผู้จัดการธนาคารแห่งหนึ่งต้องการสร้างสมการถดถอยเชิงเส้นอย่างง่ายเพื่อใช้ในการพยากรณ์เงินออมจากรายได้ของลูกค้า จึงสุ่มลูกค้าของธนาคารมาจำนวน 5 ราย มีรายละเอียดดังนี้

ลูกค้า (คนที่)	1	2	3	4	5
รายได้ (หน่วย: พันบาท)	26.9	28.7	30.2	31.8	33.5
เงินออม (หน่วย: พันบาท)	5.5	5.9	6.0	6.5	6.3

กำหนดให้ X แทนรายได้ของลูกค้า และ Y แทนเงินออมของลูกค้า จงตอบคำถามต่อไปนี้

ลูกค้า	เงินออม (Y)	รายได้ (X)	Y^2	X^2	XY
1	5.5	26.9	30.25	723.61	147.95
2	5.9	28.7	34.81	823.69	169.33
3	6	30.2	36	912.04	181.2
4	6.5	31.8	42.25	1011.24	206.7
5	6.3	33.5	39.69	1122.25	211.05
รวม	**30.2**	**151.1**	**183**	**4592.83**	**916.23**

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.9038713
R Square	0.816983327
Adjusted R Square	0.75597777
Standard Error	0.190040233
Observations	5

ANOVA					
	df	SS	MS	F	Significance F
Regression	1	0.48365413	0.4836541	13.391949	0.035257382
Residual	3	0.10834587	0.0361153		
Total	4	0.592			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.964141718	1.117012163	1.7583888	0.1769227	-1.590689512	5.518972948
รายได้ (X)	0.134872875	0.036855535	3.6595012	0.0352574	0.017582114	0.252163636