บทที่ 5

การถดถอยและสหสัมพันธ์

(Regression and Correlation)

1. สรุปสูตรในการวิเคราะห์การถดถอย (Regression analysis)

ตัวแบบเชิงเส้นตรง (Straight line model)

$Y_i = \beta_0 + \beta_1 x_i + E_i$ หรือ $\mu_{Y|x_i} = \beta_0 + \beta_1 x_i$

$i = 1, \ldots, n$

โดยที่ Y = ตัวแปรตาม (dependent variable)

x = ตัวแปรอิสระ (independent variable)

β_0 = ค่าของ $\mu_{Y|x}$ เมื่อ $x = 0$

β_1 = สัมประสิทธิ์การถดถอย (regression coefficient)

E_i = random error

ข้อมูล : (x_i, y_i), $i = 1, \ldots, n$

คำนวณ $S_{xx} = \Sigma x_i^2 - (\Sigma x_i)^2/n$

$S_{yy} = \Sigma y_i^2 - (\Sigma y_i)^2/n$

$S_{xy} = \Sigma x_i y_i - \Sigma x_i \Sigma y_i/n$

$\bar{x} = \Sigma x_i/n$

$\bar{y} = \Sigma y_i/n$

1) ค่าของ b_0 และ b_1 ซึ่งเป็นค่าของ B_0 และ B_1 (least square estimators และเป็น unbiased estimators ของ β_0 และ β_1 ตามลำดับ)

$b_1 = S_{xy}/S_{xx}$

$b_0 = \bar{y} - b_1\bar{x}$

2) Sample regression equation หรือ prediction equation

$\hat{y} = b_0 + b_1 x$

3) $SSE = \sum_{i=1}^{n} e_i^2 = S_{yy} - b_1 S_{xy} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$

4) $\hat{\sigma}^2 = s^2 = MSE = \frac{SSE}{n-2}$

5) $(1-\alpha)100\%$ C.I. ของ β_0 คือ $b_0 \pm t_{n-2,\,\alpha/2}\, s \frac{\sqrt{\Sigma x_i^2}}{\sqrt{nS_{xx}}}$

6) $(1-\alpha)$ 100% C.I. ของ β_1 คือ $b_1 \pm t_{n-2,\,\alpha/2} \frac{s}{\sqrt{S_{xx}}}$

7) $(1-\alpha)$ 100% C.I. ของ mean response μ_{Y}/x_0) คือ

$$\hat{y} \pm t_{n-2,\,\alpha/2}\, s \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}\ \text{โดยที่}\ \hat{y} = b_0 + b_1 x_0$$

8) $(1-\alpha)$ 100% C.I. ของ individual y_0 $(y/x = x_0)$ คือ

$$\hat{y} \pm t_{n-2,\,\alpha/2}\, s \sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}\ ,\ \text{โดยที่}\ \hat{y} = b_0 + b_1 x_0$$

9) ทดสอบ $H_0 : \beta_0 = \beta_0'$ ใช้ $t_c = \frac{b_0 - \beta_0'}{s\sqrt{\Sigma x_i^2 / n S_{xx}}}$ ซึ่งเป็นค่าของตัวสถิติ T ซึ่งถ้า H_0 จริงมันจะมีการกระจายเป็น t-distribution ที่มี df = n -2

10) ทดสอบ $H_0 : \beta_1 = \beta_1'$ ใช้ $t_c = \frac{b_1 - \beta_1'}{s/\sqrt{S_{xx}}}$ ซึ่งเป็นค่าของตัวสถิติ T ซึ่งถ้า H_0 จริง มันจะมีการกระจายเป็น t-distribution ที่มี df. $= n-2$

11) ทดสอบ $H_0 : \beta_1 = 0,\ H_1 : \beta_1 \neq 0$ อาจใช้ t – test หรือ F – test

t – test : ใช้ $t_c = \frac{b_1}{s/\sqrt{S_{xx}}}$ ซึ่งถ้า H_0 จริงมันจะเป็นค่าของตัวสถิติ T ซึ่งมีการกระจาย

เป็น t-distribution ที่มี df = n-2

F-test : ใช้ $f_c = \frac{MSR}{MSE} = \frac{SSR}{s^2}$ ซึ่งถ้า Ho จริงมันจะเป็นค่าของตัวสถิติ F ซึ่งมีการกระจายเป็น F-distribution ที่มี df. = (1, n −2), โดยที่

$$MSR = SSR = b_1 S_{xy}$$

$$MSE = \frac{SSE}{n-2} = \frac{S_{yy} - b_1 S_{xy}}{n-2}$$

2. การทดสอบการเท่ากันของ slopes จากสมการ regression 2 สมการ ($H_0 : \beta_1^{(1)} = \beta_1^{(2)}$) และช่วงความเชื่อมั่นของ $\beta_1^{(1)} - \beta_1^{(2)}$

โครงสร้างของข้อมูล

x_1 ของวิธีการที่ 1	x_{11} x_{12} . . . x_{1i} . . . x_{1n_1}
y_1	y_{11} y_{12} ... y_{1i} ... y_{1n_1}

x_2 ของวิธีการที่ 2	x_{21} x_{22} . . . x_{2i} ... x_{2n_2}
y_2	y_{21} y_{22} ... y_{2i} ... y_{2n_2}

Model I : $y_{1i} = \beta_0^{(1)} + \beta_1^{(1)} x_{1i} + E_{1i}$, $i = 1, \ldots . n_1$

Model 2: $y_{2i} = \beta_0^{(2)} + \beta_1^{(2)} x_{2i} + E_{2i}$, $i = 1, \ldots . n_2$

กำหนดให้

$$\widehat{\beta_1^{(1)}} = b_1^{(1)}$$

$$\widehat{\beta_1^{(2)}} = b_1^{(2)}$$

SSE (I) และ SSE (2) แทน error sum of squares จาก model 1 และ 2 ตามลำดับ

$$S_{xx}^{(1)} = \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 = \sum_{i=1}^{n_1} x_{1i}^2 - \left(\sum_{i=1}^{n_1} x_{1i}\right)^2 / n_1$$

$$S_{xx}^{(2)} = \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2 = \sum_{i=1}^{n_2} x_{2i}^2 - \left(\sum_{i=1}^{n_2} x_{2i}\right)^2/n_2$$

pooled estimate ของ σ^2 คือ $\hat{\sigma}^2 = s^2_{pooled} = \dfrac{SSE(1) + SSE(2)}{n_1 + n_2 - 4}$

2.1 สรุปขั้นการทดสอบ

1. $H_0 : \beta_1^{(1)} = \beta_1^{(2)}$
2. $H_1 : \beta_1^{(1)} \neq \beta_1^{(2)}$
3. กำหนด α
4. CR : $|T| > t_{n_1+n_2-4,\ \alpha/2}$
5. $t_c = \dfrac{b_1^{(1)} - b_1^{(2)}}{s_{pooled}\sqrt{\dfrac{1}{S_{xx}^{(1)}} + \dfrac{1}{S_{xx}^{(2)}}}}$
6. สรุปผล

2.2 ช่วงความเชื่อมั่นของ $\beta_1^{(1)} - \beta_1^{(2)}$

(1 -a) 100% Confidence interval ของ $\beta_1^{(1)} - \beta_1^{(2)}$ คือ

$$(b_1^{(1)} - b_1^{(2)}) \pm t_{n_1+n_2-4,\ \alpha/2}\, S_{pooled}\sqrt{\frac{1}{S_{xx}^{(1)}} + \frac{1}{S_{xx}^{(2)}}}$$

3. การทดสอบ Lack of fit

ตารางแสดงข้อมูลที่สามารถจะนำไปใช้ในการทดสอบ Lack of fit

ค่าต่าง ๆ ของ x	ค่า y's
x_1	$y_{11}, \ldots, y_{1n_1}$
.	
.	
.	
x_k	$y_{k1}, \ldots, y_{kn_k}$

สรุปขั้นของการทดสอบ

1. H_0 : ไม่มี lack of fit หรือตัวแบบเป็นสมการเส้นตรง
2. H_1 : มี lack of fit หรือตัวแบบไม่เป็นสมการเส้นตรง
3. กำหนด α
4. CR : $F > f_{(k-2,\ n-k),\ \alpha}$
5. $f_c = \dfrac{MS(lof)}{MS(p.e.)} = \dfrac{[SSE - SS(p.e.)]/(k-2)}{SS(p.e.)/(n-k)}$

โดยที่ $SS(p.e.) = \sum_{i=1}^{k}\left[\sum_{j=1}^{n_i} y_{ij}^2 - \dfrac{\left(\sum_{j=1}^{n_i} y_{ij}\right)^2}{n_i}\right]$

6. สรุปผล

4. การทดสอบสมมติฐานและการหาช่วงความเชื่อมั่นของ ρ

4.1 สรุปขั้นในการทดสอบสมมติฐาน $\rho = 0$

1. $H_0 : \rho = 0$
2. $H_1 : \rho \neq 0$
3. กำหนด a
4. CR : $|T| > t_{n-2,\ \alpha/2}$
5. $t_c = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ โดยที่ $r = \dfrac{S_{xy}}{\sqrt{S_{xx}\,S_{yy}}}$

$$= \frac{\Sigma xy - \Sigma x\,\Sigma y/n}{\sqrt{[\Sigma x_i^2 - (\Sigma x)^2]\,[\Sigma y_i^2 - (\Sigma y)^2]}}$$

6. สรุปผล

4.2 สรุปขั้นในการทดสอบสมมติฐาน $\rho = \rho_0$ ($\rho_0 \neq 0$)

1. $H_0 : \rho = \rho_0$
2. $H_1 : \rho \neq \rho_0$
3. กำหนด α

4. CR : $|Z| > z_{\alpha/2}$

5. $z_c = \frac{\sqrt{n-3}}{2}\ln\left[\frac{(1+r)}{(1-r)}\frac{(1-\rho_0)}{(1+\rho_0)}\right]$

$= \frac{z_r - z_{\rho_0}}{\sqrt{\frac{1}{n-3}}}$, โดยที่ $z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$ และ $z_{\rho_0} = \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)$

6. สรุปผล

4.3 ช่วงความเชื่อมั่นของ ρ

$(1-\alpha)$ 100% C.I. ของ ρ คือ

$\left(\tanh\left(z_r - z_{\alpha/2}\frac{1}{\sqrt{n-3}}\right), \tanh\left(z_r + z_{\alpha/2}\frac{1}{\sqrt{n-3}}\right)\right)$,

โดยที่ $z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$

5. การทดสอบการเท่ากันของ Correlation coefficient 2 ตัว

สรุปขั้นการทดสอบ

1. $H_0 : \rho_1 = \rho_2$
2. $H_1 : \rho_1 \neq \rho_2$
7. กำหนด a
4. CR : $|Z| > z_{\alpha/2}$

5. $z_c = \frac{z_1 - z_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$, โดยที่ $z_1 = \frac{1}{2}\ln\left(\frac{1+r_1}{1-r_1}\right)$,

$z_2 = \frac{1}{2}\ln\left(\frac{1+r_2}{1-r_2}\right)$,

$\sigma_1^2 = \frac{1}{n_1-3}$ และ $\sigma_2^2 = \frac{1}{n_2-3}$

6. สรุปผล

6. Serial correlation

$x_1, \ldots, x_n$ แทนค่าสังเกตของตัวแปร x ในเวลา n จุดที่ต่อเนื่องกัน

ก) สร้าง Scatter diagram ของจุด (n – 1) จุดคือ $(x_1, x_2), \ldots (x_{i-1}, x_i), \ldots (x_{n-1}, x_n)$

ข) คำนวณค่า lag 1 correlation หรือ first-order antocorrelation

$$r_1 = \frac{\sum_{i=2}^{n} (x_{i-1} - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

ถ้าคำนวณโดยใช้เครื่องคิดเลข อาจใช้สูตร

$$r_1 = A/B$$

โดยที่

$$A = \sum_{i=2}^{n} x_{i-1} x_i - \dot{x} \left[\sum_{i=1}^{n-1} x_i + \sum_{i=2}^{n} x_i \right] + (n-1)\,\bar{x}^2$$

และ $$B = \sum_{i=1}^{n} x_i^2 \quad \left(\sum_{i=1}^{n} x_i \right)^2 / n$$

ค) สรุปขั้นของการทดสอบสมมติฐานเกี่ยวกับ ρ_1 (เมื่อตัวอย่างมีขนาดใหญ่)

1) $H_0 : \rho_1 = 0$

2) H_1 : 1) $\rho_1 \neq 0$ หรือ 2) $\rho_1 > 0$ หรือ 3) $\rho_1 < 1$

3) กำหนด α

4) CR : 1) $Z > z_{\alpha/2}$ และ $Z < z_{\alpha/2}$

3) $Z > z_\alpha$

3) $Z < -z_\alpha$

5) $z_c = \sqrt{n}\, r_1$

6) สรุป ก) ถ้า z_c ตกอยู่ใน CR. เราปฏิเสธ H_0 ที่ระดับนัยยะสำคัญ = α นั่นคือข้อมูลชุดนี้มีความสัมพันธ์กันในตัวข้อมูลเอง

ข) ถ้า z_c ไม่ตกอยู่ใน CR. เราไม่ปฏิเสธ H_0 ที่ระดับนัยยะสำคัญ = α นั่นคือข้อมูลชุดนี้ไม่มีความสัมพันธ์กันในตัวข้อมูลเอง หรือข้อมูลมีคุณสมบัติของการสุ่ม (randomness)

เฉลยแบบฝึกหัดบทที่ 5

5.1 พนักงานป่าไม้ต้องการทราบว่าเขาจะใช้เส้นผ่าศูนย์กลางในการทำนายความสูงของต้นไม้ได้หรือไม่ เขาได้ทำการวัดเส้นผ่าศูนย์กลางของต้นไม้ตรงส่วนสูง 4.5 ฟุต จากพื้นดิน และความสูงของต้นเมเปิลที่ให้น้ำตาล 12 ต้น ผลปรากฏดังนี้

x : เส้นผ่าศูนย์กลาง (นิ้ว)	.9	1.2	2.3	3.1	3.3	3.9	4.3	6.2	9.6	12.6	16.1	25.8
y : ความสูง (ฟุต)	18	26	32	36	44.5	35.6	40.5	57.5	67.3	84	67	87.5

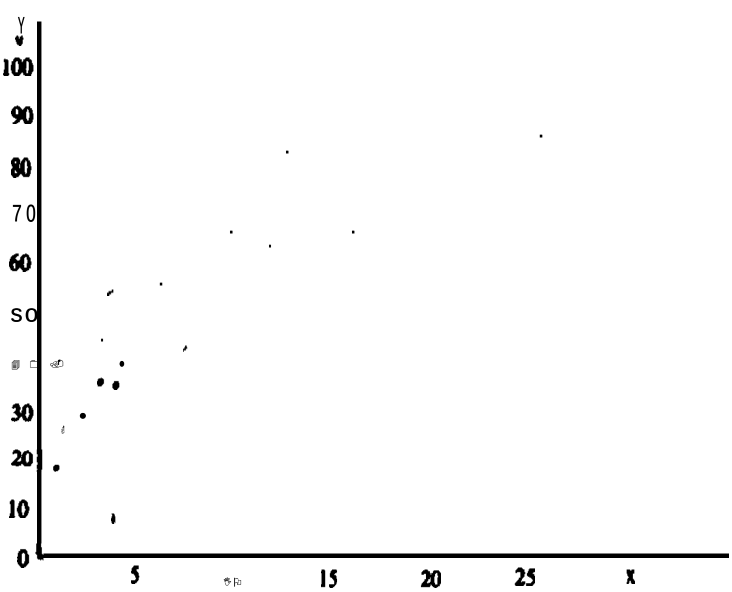
tl) จงสร้าง Scatter diagram เพื่อดูว่าควรใช้ Straight line made1 หรือไม่

ข) ถ้ากำหนดให้ $x' = \log_e x$ และ $y' = \log_e y$ จงสร้าง Scatter diagram ของ (x', y')

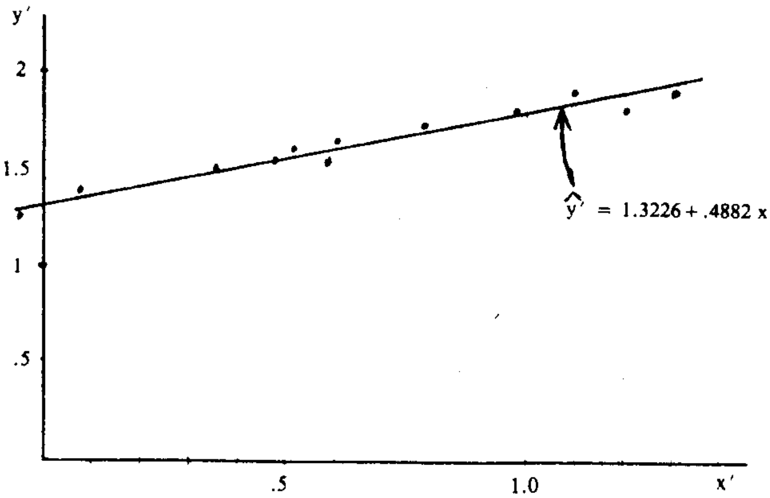
X'	-.05	.08	.36	.49	.52	.59	.63	.79	.98	1.10	1.21	1.41
Y'	1.26	1.41	1.51	1.56	1.65	1.55	1.61	1.76	1.83	1.92	1.83	1.94

ค) rIn model $y_i' = \beta_0 + \beta_1 x_i' + E_i$ จงหา Sample regression equation

ก) Plot (x_i, y_i), i = 1, ... 12



ข) Plot (x_i: y_i), i = 1, ... 12



ค) Model : $y_i' = \beta_0 + \beta_1 x_i' + E_i$

$\Sigma x_i' = 8.11$ $\Sigma y_i' = 19.83$

$\Sigma x_i'^2 = 7.6407$ $\Sigma y_i'^2 = 33.2519$

$\bar{x}' = .6758$ $\bar{Y}' = 1.6525$

$\Sigma x_i' y_i' = 14.4561$

$S_{x'x'} = 2.1597$

$S_{y'y'} = .4828$

$S_{x'y'} = 1.0543$

$$b_1 = \hat{\beta}_1 = \frac{S_{x'y'}}{S_{x'x''}} = .4882$$

$$b_0 = \hat{\beta}_0 = \bar{y}' - \hat{\beta}_1 \bar{x}' = 1.6525 - (.4882)(.6758) = 1.3226$$

Sample regression equation

$$\hat{y}' = b_0 + b_1 x'$$

$$\hat{Y}' = 1.3226 + .4882 x'$$

5.2 ในการเตรียมการที่จะคัดเลือกนักศึกษาปีที่ 1 เข้าเรียนวิชาเอกคณิตศาสตร์ โดยจะใช้วิธีการสอบวิชาคณิตศาสตร์พื้นฐาน (คะแนนเต็ม 10 คะแนน) เพื่อจะศึกษาดูว่าการใช้คะแนนวิชานี้จะทำนายผลสำเร็จในการเรียนวิชาเอกได้หรือไม่ วิทยาลัยแห่งหนึ่งจึงได้เอานักศึกษาที่จบแล้ว และได้คะแนนวิชาพื้นฐานต่าง ๆ กันมา 8 คน แล้วจดคะแนนตอนจบปริญญาตรี (grade point average) ของแต่ละคนดังนี้

x : คะแนนคณิตศาสตร์พื้นฐาน	4	6	8	6.5	3.5	8	7	5
y : คะแนนจบปริญญาตรี	2.1	3.1	3.4	2.9	2.1	3.3	2.8	2.3

กำหนดค่าที่คำนวณจากข้อมูลดังนี้

$\Sigma x_i = 46$ $\qquad \Sigma y_i = 22$

$\Sigma x_i^2 = 308.5$ $\qquad \Sigma y_i^2 = 62.42$

$\bar{x} = 6$ $\qquad \bar{y} = 2.15$

$\Sigma x_i y_i = 137.9$

$S_{XX} = \Sigma (x_i - \bar{x})^2 = 20.5$

$S_{yy} = \Sigma (y_i - \bar{y})^2 = 1.92$

$S_{xy} = \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 5.9$

สมมุติว่าความสัมพันธ์ของ x และ Y อยู่ในรูป $\mu_{Y|x} = \beta_0 + \beta_1 x$

ก) จงหา Least squared estimate ของ β_0 และ β_1

ข) จงเขียน Prediction equation

fl) จงหา Point estimate ของ $\mu_{Y|7.5}$ และอธิบายความหมาย และถ้านาย ก. ได้คะแนนวิชาพื้นฐาน 5.5 คะแนน จงประมาณว่านาย ก. จะเรียนจบด้วยคะแนนเฉลี่ยเท่าใด

ง) จงทดสอบ $H_0 : \beta_1 = 0$ โดยใช้ F-test กำหนด a = .01

จ) จงหา $100r^2\%$ และอธิบายความหมาย

ฉ) ถ้าความสัมพันธ์ของ X และ Y อยู่ในรูปนี้จริง ค่าประมาณของ σ^2 เท่ากับเท่าใด

ก) $\hat{\beta}_1 = b_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{5.9}{20.5} = .2878$

$\hat{\beta}_0 = bo = \bar{y} - b_1 \bar{x} = 2.75 - (.2878)(6) = 1.0232$

ข) $\hat{y} = 1.0232 + .2878\ x$

ค) $\hat{\mu}_{Y/7.5} = 1.0232 + .2878\ (7.5) = 3.1817$

หมายความว่านักศึกษาคนใดก็ตามที่ทำคะแนนคณิตศาสตร์พื้นฐานได้ 7.5 คะแนน คาดว่าจะทำคะแนนตอนจบปริญญาตรีได้โดยเฉลี่ย 3.1817

$\hat{y}_{5.5} = 1.0232 + (.2878)(5.5) = 2.6061$

หมายความว่า ถ้านาย ก. ได้คะแนนคณิตศาสตร์พื้นฐาน 5.5 คะแนน คาดว่าเขาจะเรียนจบด้วยคะแนน 2.6061

ง) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = .01$

CR : $F > f_{(1,\ 6),\ .01} = 17.75$

$n = 8$, $SST = S_{yy} = 1.92$

$SSR = b_1 S_{xy} = (.2878)(5.9) = 1.6980$

$SSE = 1.92 - 1.6980 = .222$

ANOVA

S.V.	df.	ss	MS	f
Regression	1	1.698	1.698	f_c = 45.8919
Error	6	.222	.037	
Total	7	1.920		

$$f_c = \frac{MSR}{MSE} = \frac{1.698}{.037} = 45.8919$$

$\because$ f_c ตกอยู่ใน CR. เราปฏิเสธ H_0 ที่ $\alpha = .01$ นั่นคือคะแนนคณิตศาสตร์พื้นฐานมีความสัมพันธ์กับคะแนนตอนจบปริญญาตรีในลักษณะเหตุและผล

จ) $100\, r^2\% = \frac{SSR}{SST} \times 100 = \frac{1.698}{1.92} \times 100 = 88.44\%$

88.44% ของความแปรปรวนของ Y เนื่องมาจากความสัมพันธ์เชิงเส้นตรงกับ X

ฉ) $\hat{\sigma}^2 = MSE = .037$

5.3 ในปี 2519 บริษัทผู้ผลิตผงซักฟอกยี่ห้อต่าง ๆ มีรายจ่ายในการโฆษณา (x) และยอดขาย (y) ดังนี้

ยี่ห้อ	x (ล้านบาท)	y (ล้านบาท)
บรีส	7	9
ริน	9	12
ควิก	5	7
นิส	6	8
แฟบ	3	4

สมมุติรูปความสัมพันธ์คือ $\mu_{Y/x} = \beta_0 + \beta_1 x$

ก) จงเขียน Prediction equation

ข) จงอธิบายค่า $\hat{\beta}_0 = b_0$ และ $\hat{\beta}_1 = b_1$ ที่หาได้

ค) ถ้าบริษัทผลิตควิกมีรายจ่ายในการโฆษณา 4 ล้านบาท จงประมาณว่าจะได้ยอดขายกี่ล้านบาท

ง) จงทดสอบที่ $\alpha = .05$ $H_0 : \beta_1 = 0$ โดยใช้ F-test

จ) ถ้าหา 95% C.I. ของ $\mu_{Y/6}$ ได้ (7.631, 8.369) จงอธิบายความหมาย

$n = 5$

$\Sigma x_i = 30$ $\Sigma y_i = 40$

$\Sigma x_i^2 = 200$ $\Sigma y_i^2 = 354$

$\bar{x} = 6$ $\bar{y} = 8$

$\Sigma x_i y_i = 266$

$S_{xx} = 20$

$S_{yy} = 34$

$S_{xy} = 26$

ก) $b_1 = \frac{26}{20} = 1.3$

$b_0 = 8-(1.3)(6) = .2$

Prediction equation : $\hat{y} = .2 + 1.3x$

ข) b_0 = ยอดขายเมื่อไม่จ่ายค่าโฆษณาเลย

b_1 = ยอดขาย (ล้านบาท) ที่จะเพิ่มขึ้นเมื่อรายจ่ายในการโฆษณาเพิ่มขึ้น 1 ล้านบาท

ค) $\hat{y}_4 = .2 + 1.3(4) = 5.4$ ล้านบาท

ง) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = .05$

CR : $F > f_{(1,3),.05} = 10.13$

n = S

$SST = S_{yy} = 34$

$SSR = b_1 S_{xy} = 1.3(26) = 33.8$

$SSE = 34-33.8 = .2$

ANOVA

S.V.	df.	ss	MS	f
Regression	1	33.8	33.8	f_c = 506.1466
Error	3	.2	.0667	
Total	4	34		

$$f_c = \frac{MSR}{MSE} = \frac{33.8}{.0667} = 506.1466$$

$\because$ f_c ตกอยู่ใน CR.เราปฏิเสธ H_0 ที่ $\alpha = .05$ นั่นคือยอดขายมีความสัมพันธ์กับรายจ่ายในการโฆษณาในลักษณะผลและเหตุ

จ) ถ้าหา 95% C.I. ของ $\mu_{Y/6}$ ได้ (7.631, 8.369) นั่นคืออาจกล่าวด้วยความเชื่อมั่น 95% ว่าบริษัทที่ลงทุนโฆษณา 6 ล้านบาท คาดว่ายอดขายเฉลี่ยจะอยู่ระหว่าง 7.631 ล้านบาท ถึง 8.369 ล้านบาท

5.4 จากข้อมูลที่กำหนดให้

x	1	1	1	2	3	3	4	5	5
Y	9	7	8	10	15	12	19	24	21

ก) จงสร้าง Scatter diagram

ข) จงเขียน Prediction equation ของ model : $\mu_{Y/x} = \beta_0 + \beta_1 x$

ค) จงทดสอบ Lack of fit ที่ a = .05

$n = 9$

$\Sigma x_i = 25$ $\Sigma y_i = 125$

$\Sigma x_i^2 = 91$ $\Sigma y_i^2 = 2041$

$\bar{x} = 2.1778$ $\bar{y} = 13.8889$

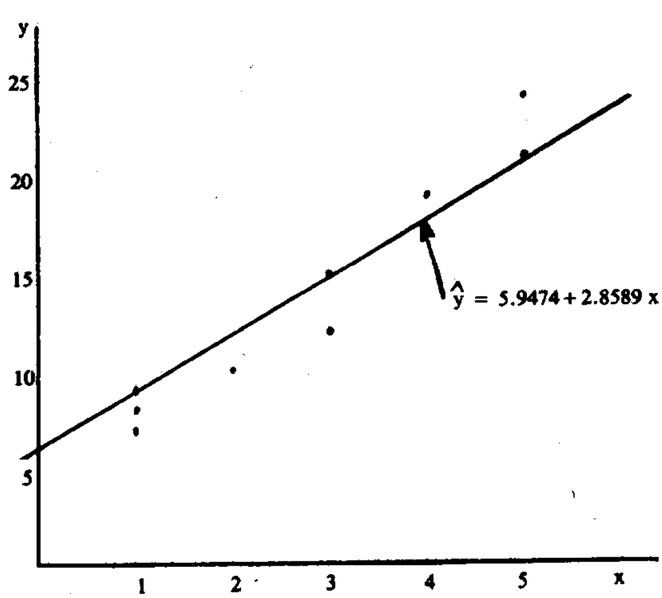
$\Sigma x_i y_i = 426$

$S_{xx} = 21.5556$

$S_{yy} = 304.8889$

$S_{xy} = 78.1718$

ก) Plot (x_i, y_i), $i = 1, \ldots, 9$



ข) $b_1 = \frac{S_{xx}}{S_{xy}} = 2.8589$

$b_0 = 13.8889 - (2.8589)(2.7778) = 5.9474$

Prediction equation : $\hat{y} = 5.9474 + 2.8589\, x$

ค)

x	y_i's	$\bar{y}_i$	n_i	df	SS_i
1	9, 7, 8	8	3	2	2
2	10	10	1	0	0
3	15, 12	13.5	2	1	4.5
4	19	19	1	0	0
5	24, 21	22.5	2	1	4.5
			$n = 9$	4	11 = SS (p.e.)

$SST = S_{yy} = 304.8889$

$SSR = b_1 S_{xy} = (2.8589)(78.7778) = 225.2179$

$SSE = 304.8889 - 225.2179 = 79.6710$

$SS\,(p.e.) = 11$

$\therefore SS\,(lof) = 76.6710 - 11 = 68.6710$

ANOVA

S.V.	df	ss	MS	f
Regression	1	225.2179	225.2179	$f_1 = 19.7879$
Error	7	79.6710	11.3816	
Lack of fit	3	68.6710	22.8903	$f_2 = 8.3238$
Pure error	4	11	2.15	
Total	8	304.8889		

H_0 : There is no lack of fit

H_1 : There is a lack of fit

a = .05

CR : F > $f_{(3,\ 4),\ .05}$ = 6.59

f_c = f_2 = 8.3238

∵ f_c ตกอยู่ใน CR. เราปฏิเสธ H_0 ที่ α = .05 นั่นคือไม่ยอมรับว่า X และ Y มีความสัมพันธ์เชิงเส้นตรง

หมายเหตุ ถ้าทดสอบที่ a = .01, CR : F > $f_{(3,\ 4),\ .01}$ = 16.69

เราไม่สามารถปฏิเสธ H_0 ที่ a = .01

5.5 ปริมาณสารเคมี y (กรัม) ซึ่งละลายในน้ำ 100 กรัม ณ อุณหภูมิต่าง ๆ กันเป็นดังนี้

x (°c)	y (กรัม)
0	8, 6, 8
15	12, 10, 14
30	25, 21, 24
45	31, 33, 28
60	44, 39, 42
75	48, 51, 44

n) **จงสร้าง** Scatter diagram

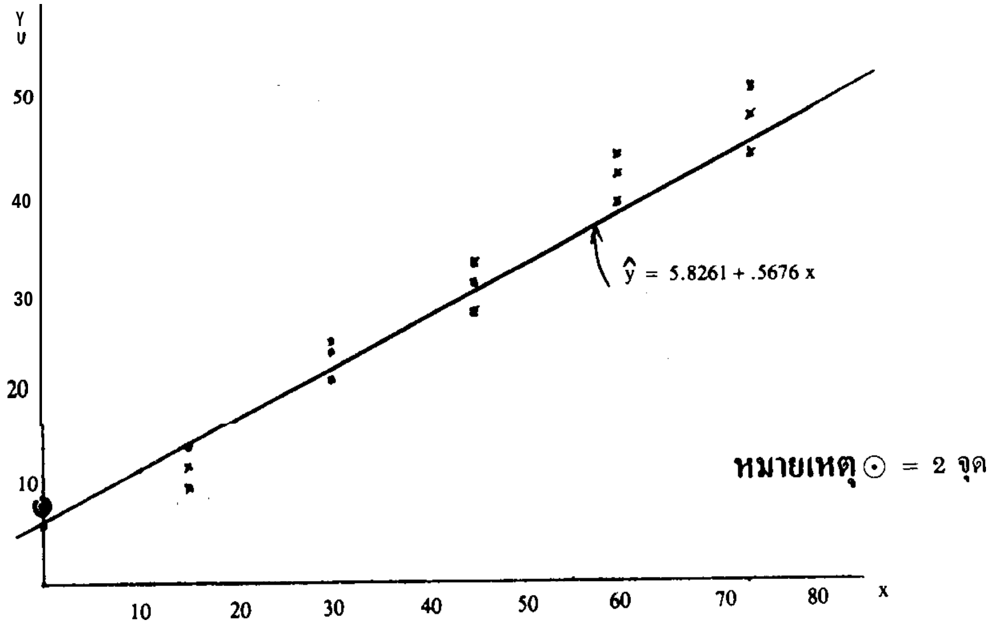
ข) จงเขียน Prediction equation **ของ** model : $\mu_{Y/x} = \beta_0 + \beta_1 x$

ค) จงประมาณปริมาณของสารเคมีที่จะละลายในน้ำ 100 **กรัม ที่อุณหภูมิ** 50°C

ง) จงทดสอบที่ a = .01, $H_0 : \beta_0 = 6$, $H_1 : \beta_0 \neq 6$

จ) จงทดสอบว่า Straight line model ที่ใช้อยู่นี้พอเพียงหรือไม่ นั่นคือทดสอบ H_0 ไม่มี lack of fit

fl) Plot (x_i, y_i), i = 1, ..., 18



$n = 18$

$\Sigma\ x_i = 675$ $\qquad$ $\Sigma\ y_i = 488$

$\Sigma\ x_i^2 = 37125$ $\qquad$ $\Sigma\ y_i^2 = 17142$

$\bar{x} = 37.5$ $\qquad$ $\bar{y} = 27.1111$

$\Sigma\ x_i\ y_i = 25005$

$S_{xx} = 11812.5$

$S_{yy} = 3911.7778$

$S_{xy} = 6705$

ข) $b_1 = \dfrac{S_{xy}}{S_{xx}} = .5676$

$b_o = \bar{y} - b_1\bar{x} = 27.1111 - (.5676)(37.5) = 5.8261$

. $\hat{y} = 5.8261 + .5676\ x$

ค) $x = 50°C$

$\hat{y} = 5.8261 + .5676\ (50) = 34.2061$

นั่นคือคาดว่าปริมาณสารเคมีจำนวน 34.2061 กรัม จะละลายในน้ำ 100 กรัม ที่อุณหภูมิ 50°C

ง) SST = S_{yy} = 3911.7778, df = 17

SSR = $b_1 s_{xy}$ = .5676 (6705) = 3805.758, df = 1

SSE = 3911.7778 – 3805.758 = 106.0198 , df = 16

$$\therefore \text{MSE} = \frac{\text{SSE}}{16} = \frac{106.0198}{16} = 6.6262 = s^2$$

s = 2.5741

Ho : β_0 = 6

H_1 : $\beta_0 \neq 6$

a = .01

CR : $|T| > t_{16,.05}$ = 1.746

$$tc = \frac{bo-6}{s\sqrt{\Sigma x_i^2/n S_{xx}}} = \frac{5.8261-6}{2.5741\sqrt{\frac{37125}{18(11812.5)}}} = \frac{-.1739}{1.0756017}$$

$$= -.1617$$

∵ t_c ตกอยู่นอก CR. เราไม่สามารถปฏิเสธ H_0 ที่ a = .Ol นั่นคือ สารเคมีจำนวน 6 กรัมสามารถละลายในน้ำ 100 กรัมที่อุณหภูมิ 0°C

จ) H_0 : There is no lack of lit

H_1 : There is a lack of fit

a = .Ol

CR : F > $f_{(4,12),.01}$ = 5.41

x	Y'S	Σy	Σy²	SSᵢ	nᵢ	df
0	8, 6, 8	22	164	2.6667	3	2
15	12, 10, 14	36	440	8	3	2
30	25, 21, 24	70	1642	8.6667	3	2
45	31, 33, 28	92	2834	12.6667	3	2
60	44, 39, 42	125	5221	12.6667	3	2
75	48, 51, 44	143	6841	24.6667	3	2
	รวม	488	17142	69.3333	18	12

SS (p.e.) = 69.3333

ss (lof) = SSE − SS (p.e.) = 106.0198 -69.3333 = 36.6865

ANOVA

S.V.	df	SS	MS	f
Regression	1	3805.758	3805.758	f_1 = 574.35
Error	16	106.0198	6.6262	
Lack of fit	4	36.6865	9.1716	f_2 = 1.5874
Pure error	12	69.3333	5.7778	
Total	17	3911.7778		

$f_c = f_2 = 1.5874$ ซึ่งไม่ตกอยู่ใน CR. เราไม่สามารถปฏิเสธ H_0 ที่ $\alpha = .01$ นั่นคือ X และ Y มีความสัมพันธ์กันในเชิงเส้นตรง หรือ Straight line model ใช้ได้ดีกับข้อมูลนี้

หมายเหตุ ถ้าทำการทดสอบที่ $\alpha = .05$, CR : F > $f_{(4,\ 12),\ .05} = 3.26$ เราก็ยังคงไม่สามารถปฏิเสธ H_0 ได้ที่ $\alpha = .05$

5.6 กำหนดข้อมูลให้ต่อไปนี้

x	2.1	2.6	2.7	2.7	2.7	3.0	3.0	3.0	3.1	3.1	3.2	3.3	3.6	3.8
y	2.6	2.8	2.9	3.6	3.6	3.1	3.4	3.1	3.9	3.1	4.1	3.6	4.0	3.6

$\Sigma x_i = 41.9$ $\qquad \Sigma y_i = 41.4$

$\Sigma x_i^2 = 127.79$ $\qquad \Sigma y_i^2 = 163.26$

$\Sigma x_i y_i = 143.59$

$n\bar{y}^2 = 160.4829$

$S_{xx} = \Sigma (x_i - \bar{x})^2 = 2.3893$

$S_{yy} = \Sigma (y_i - \bar{y})^2 = 2.7771$

$S_{xy} = \Sigma (x_i - \bar{x})(y_i - \bar{y}) = 1.7286$

Simple regression model : $\mu_{Y/x} = \beta_0 + \beta_1 x$ หรือ $Y = \beta_0 + \beta_1 x + E$

กำหนด $\hat{\beta}_0 = b_0 = 1.2203$ และ

$\hat{\beta}_1 = b_1 = 0.1235$

ก) จงเขียน Prediction equation หรือ Fitted regression equation

ข) จงสร้างตาราง ANOVA เพื่อทดสอบที่ $a = .05$

1. Ho : $\beta_1 = 0$

2. H_0 : ไม่มี lack of fit, เราต้องปรับปรุง model ใหม่หรือไม่

ค) คำนวณ $100\ r^2\%$

ก) $\hat{y} = b_0 + b_1 x = 1.2203 + .7235x$

ข) SST $= S_{yy} = 2.7771$

SSR $= b_1\ S_{xy} = .7235\ (1.7286) = 1.2506$

SSE $= 2.7771 - 1.2506 = 1.5265$

x	y's	SS_i	df
2.1	2.6	0	0
2.6	2.8	0	0
2.1	2.9, 3.6, 3.6	.3267	2
3.0	3.1, 3.4, 3.1	.06	2
3.1	3.9, 3.1	.32	1
3.2	4.1	0	0
3.3	3.6	0	0
3.6	4.0	0	0
3.8	3.6	0	0
	รวม	.7067	5

$\Sigma\ y = 10.1$, $\Sigma\ y^2 = 34.33$

$\Sigma\ y = 9.6$, $\Sigma\ y^2 = 30.78$

$\Sigma\ y = 7$, $\Sigma\ y^2 = 24.82$

ANOVA

S.V.	df	ss	MS	f
Regression	1	1.2506	1.2506	f_1 = 9.8318
Error	12	1.5265	.1272	
Lack of fit	7	.8198	.1171	f_2 = .8287
Pure error	5	.7067	.1413	
Total	13	2.7771		

1. Ho : $\beta_1 = 0$

   $H_1 : \beta_1 \neq 0$

   a = .05

   CR : $F > f_{(1,\ 12),\ .05} = 4.75$

   $f_c = f_1 = 9.8318$ ซึ่งตกอยู่ใน CR. เราปฏิเสธ H_0 ที่ a = .05 นั่นคือ X และ Y มีความสัมพันธ์กันในรูปเหตุและผล

2. H_0 : There is no lack of fit

   H_1 : There is a lack of fit

   a = .05

   CR : $F > F_{(7,\ 5),\ .05} = 4.88$

   $f_c = f_2 = .8287$ ซึ่งไม่ตกอยู่ใน CR. เราไม่สามารถปฏิเสธ H_0 ที่ $\alpha = .05$ นั่นคือ X และ Y มีความสัมพันธ์กันเชิงเส้นตรง เราไม่ต้องปรับปรุง model

ค. $100\ r^2\% = \frac{SSR}{SST} \times 100 = \frac{1.2506}{2.7771} \times 100 = 36.01\%$

36% ของความแปรปรวนของ Y เนื่องมาจากความสัมพันธ์เชิงเส้นตรงกับ X

5.7 สุ่มคนไข้ที่เป็นโรคชนิดหนึ่งมา 12 คน หลังจากที่ให้การรักษาด้วยยาชนิดหนึ่งแล้ว 3 วัน ได้วัดอุณหภูมิ (X) และจำนวนการเต้นของชีพจรต่อ 1 นาที (Y) ของคนไข้แต่ละคน ผลปรากฏดังนี้

x	99.4	100.1	98.6	97.9	101.4	102.6	99.0	98.6	101.1	98.0	100.5	99.3
y	78	83	70	76	75	89	80	77	86	79	84	87

$\Sigma x_i = 1196.5$ $\qquad \Sigma y_i = 964.0$

$\Sigma x_i^2 = 119324.37$ $\qquad \Sigma y_i^2 = 77786$

$\Sigma x_i y_i = 96170.2$

ก) คำนวณค่า r (Sample correlation coefficient)

ข) ทดสอบที่ $\alpha = .05$, $H_0: \rho = 0$, $H_1: \rho \neq 0$

ค) ทดสอบที่ $\alpha = .05$, $H_0: \rho = .7$, $H_1: \rho \neq .7$

ง) จงหา 95% C.I. ของ ρ

$n = 12$

$S_{xx} = 23.3492$

$S_{yy} = 344.6667$

$S_{xy} = 51.3667$

ก) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \dfrac{51.3667}{\sqrt{23.3492\ (344.6667)}}$

$= .5726$

ข) $H_0: \rho = 0$

$H_1: \rho \neq 1$

$\alpha = .05$

$CR: |T| > t_{10,\ .025} = 2.228$

$t_c = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

$$\frac{.5726\sqrt{10}}{\sqrt{1-(.5726)^2}}$$

$= 2.2086$

$\because$ t_c ไม่ตกอยู่ใน CR. เราไม่สามารถปฏิเสธ H_0 ได้ที่ $\alpha = .05$

ค) $H_0 : \rho = .7$

$H_1 : \rho \neq .7$

$\alpha = .05$

CR : $|Z| > z_{.025} = 1.96$

$$\left.\begin{aligned} z_r &= z_{.573} = .6520 \\ z_{\rho_0} &= z_{.7} = .8673 \end{aligned}\right\} \text{จากตาราง A10}$$

$$z_c = \frac{z_r - z_{\rho_0}}{\sqrt{\frac{1}{n-3}}}$$

$= (.652 - .8673)\sqrt{9} = .6459$

$\because$ z_c ไม่ตกอยู่ใน CR. เราไม่สามารถปฏิเสธ H_0 ที่ $\alpha = .05$

ง) 95% C.I. ของ z_ρ คือ $z_r \pm z_{\alpha/2}\sqrt{\frac{1}{n-3}}$

$= .652 \pm 1.96\frac{1}{\sqrt{9}}$

$= .652 \pm .6533$

$= (-.0013, 1.3053)$

$\therefore$ 95% C.I. ของ ρ คือ (tanh (- .0013), tanh (1.3053))

$= (-.0013, .8630)$

5.8 สุ่มนักเรียนมา 8 คน แต่ละคนวัด I.Q. (X) และวัดความเร็วในการอ่าน (Y) ผลปรากฏดังนี้

X	102	95	115	112	86	102	121	108
Y	14.7	12.0	17.1	15.8	10.1	11.6	19.8	17.5

n) คำนวณค่า r

ข) ทดสอบ Ho : $\rho = 0$, $H_1 : \rho \neq 0$ ที่ $\alpha = .05$

ค) จงหา 95% C.I. ของ ρ

$n = 8$

$\Sigma x_i = 841$ $\Sigma y_i = 1186$

$\Sigma x_i^2 = 89303$ $\Sigma y_i^2 = 1837$

$\Sigma x_i y_i = 12713.1$

$S_{xx} = 892.875$

$S_{yy} = 78.755$

$S_{xy} = 245.275$

ก) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = .925$

ข) Ho : $\rho = 0$

HI : $\rho \neq 0$

$a = .05$

CR : $|T| > t_{6 \; .025} = 2.447$

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{.925\sqrt{6}}{\sqrt{1-(.925)^2}}$$

≈ 5.9631

$\because$ t_c ตกอยู่ใน CR. เราปฏิเสธ H_0 ที่ $\alpha = .05$

n) 95% C.I. ของ z_ρ คือ $z_r \pm z_{.025}\sqrt{\dfrac{1}{n-3}}$

จากตาราง A10 : $z_r = z_{.925} \approx 1.6226$

$\therefore$ 95% C.I. ของ $z_r = 1.6226 \pm 1.96\sqrt{\dfrac{1}{5}}$

$= 1.6226 \pm .8765$

$= (0.7461, \; 2.4991)$

$\therefore$ 95% C.I. ของ ρ คือ (tanh (.7461), tanh (2.4991))

$= (.6328, \; .9866)$

5.9 จากคะแนนวิชาคณิตศาสตร์ และคะแนนภาษาอังกฤษของนักศึกษา 6 คนที่สุ่มมาเป็นตัวอย่าง

x : คะแนนคณิตศาสตร์	70	92	80	74	65	83
y : คะแนนภาษาอังกฤษ	74	84	63	87	78	90

ก) จงคำนวณค่า r

ข) ทดสอบ $H_0 : \rho = 0,\ H_1 : \rho \neq 0$ ที่ $\alpha = .10$

ค) จากข้อ ข) ถ้าเราไม่สามารถปฏิเสธ H_0 เราจะสรุปผลการทดสอบอย่างไร

$n = 6$

$\Sigma x_i = 464$ $\quad\quad$ $\Sigma y_i = 476$

$\Sigma x_i^2 = 36354$ $\quad\quad$ $\Sigma y_i^2 = 38254$

$\Sigma x_i y_i = 36926$

$S_{xx} = 471.3333$

$S_{yy} = 491.3333$

$S_{xy} = 115.3333$

ก) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = .2397$

ข) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

$\alpha = .10$

CR : $|T| > t_{4,.05} = 2.132$

$t_c = \dfrac{(.2397)\sqrt{4}}{\sqrt{1-(.2397)^2}} = .4938$

$\because$ t_c ไม่ตกอยู่ใน CR. เราไม่ปฏิเสธ H_0 ที่ $\alpha = .10$

ค) จากผลการทดสอบในข้อ ข) สรุปได้ว่า X และ Y ไม่มีความสัมพันธ์กันเชิงเส้นตรง แต่อาจมีความสัมพันธ์กันในรูปอื่น หรือไม่มีความสัมพันธ์กันเลย

5.10 ต่อไปนี้เป็นความสูงและน้ำหนักของมิสอเมริกา 17 คน

x : ความสูง (นิ้ว)	65	67	66	65.5	65	66.5	66	67	66
y : น้ำหนัก (ปอนด์)	114	120	116	118	115	124	124	115	116

x : ความสูง	69	67	65.5	68	67	68	69	68
Y : น้ำหนัก	135	125	110	121	118	120	125	119

ก) จงสร้าง Scatter dirgram

ข) คำนวณค่า r และอธิบายความหมาย

ค) ทดสอบที่ $\alpha = .10$, $H_0 : \rho = 0$

$n = 17$

$\Sigma x_i = 1135.5$ $\qquad$ $\Sigma y_i = 2035$

$\Sigma x_i^2 = 75870.75$ $\qquad$ $\Sigma y_i^2 = 244135.0$

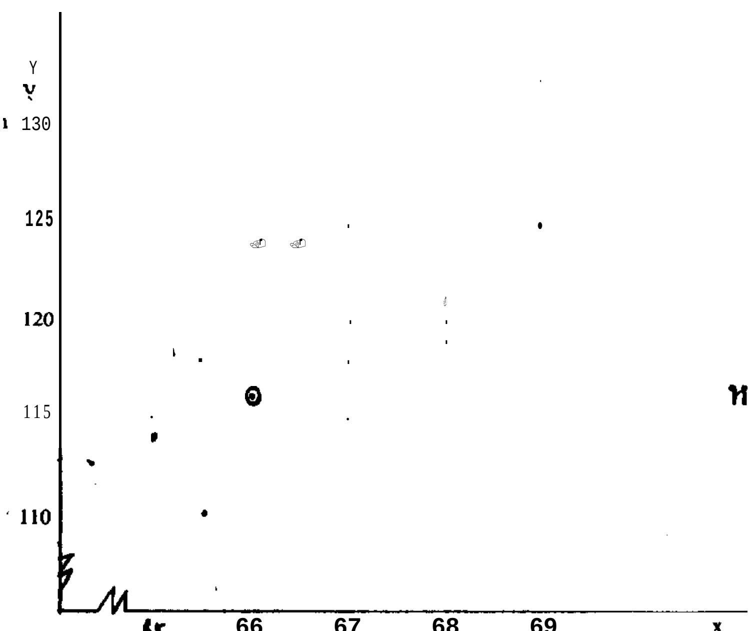
$\Sigma x_i y_i = 13600.7$

$S_{xx} = 26.0294$

$S_{yy} = 533.5294$

$S_{xy} = 80.9706$

ก. Plot (x_i, y_i), i = 1,......17



หมายเหตุ ⊙ = 2 จุด

ข) r = .6871 (r เป็นค่าบวก)

หมายความว่าถ้า variable ตัวหนึ่งมีค่าเพิ่มขึ้นหรือลดลง ค่าของ variable อีกตัวหนึ่งจะเพิ่มขึ้นหรือลดลงตามไปด้วย

ค) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

$a = .10$

CR: $|T| > t_{15,.05} = 1.753$

$$t_c = \frac{.6871\sqrt{15}}{\sqrt{1-(.6871)^2}} = 3.6626$$

∴ tc ตกอยู่ใน CR. เราปฏิเสธ Ho ที่ $\alpha = .10$ นั่นคือ X และ Y มีความสัมพันธ์กันในเชิงเส้นตรง

5.11 จากข้อมูลที่กำหนดให้ 2 ชุด จงทดสอบ $H_0 : \rho_1 = \rho_2$ ที่ $\alpha = .05$

x_{1i}	.20	.25	.25	.30	.40	.50	.50
y_{1i}	30	26	40	35	54	56	65

x_{2i}	.20	.25	.30	.40	.40	.50
y_{2i}	23	24	42	49	55	70

$n_1 = 7$

$S_{xx}^{(1)} = .0921$

$S_{yy}^{(1)} = 1301.4286$

$S_{xy}^{(1)} = 10.1857$

$r_1 = .930,\ a : = \frac{1}{n_1-3} = \frac{1}{4}$

จากตาราง A10 : $z_1 = z_{r_1} = z_{.93}$

$= 1.6584$

$n_2 = 6$

$S_{xx}^{(2)} = .0621$

$S_{yy}^{(2)} = 1666.8333$

$S_{xy}^{(2)} = 9.9417$

$r_2 = ,977,\ \sigma_2^2 = \frac{1}{n_2-3} = \frac{1}{3}$

$z_2 = z_{r_2} = z_{.977} = 2.2269$

$Ho : \rho_1 = \rho_2$

$H_1 : \rho_1 \neq \rho_2$

a = .05

CR : $|Z| > z_{.025} = 1.96$

$$z_c = \frac{z_1 - z_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

$$= \frac{1.6584-2.2269}{\sqrt{\frac{1}{4}+\frac{1}{3}}}$$

$$= -.7444$$

$\because$ z_c ไม่ตกอยู่ใน CR. เราไม่ปฏิเสธ H_0 ที่ α = .05 นั่นคือสรุปว่า $\rho_1 = \rho_2$

5.12 จากการวัดความถี่ของเสียงที่สี่แยกหนึ่งในเวลาที่ต่อเนื่องกันได้ข้อมูลดังนี้ 65, 64, 63, 61 60, 58, 63, 64, 62, 64, 63, 63, 62, 60, 62, 64, 66, 68, 68, 69 จงทำการสำรวจเพื่อตรวจดูว่าอาจมี lag 1 independence หรือไม่โดย

ก) สร้าง Scatter diagram

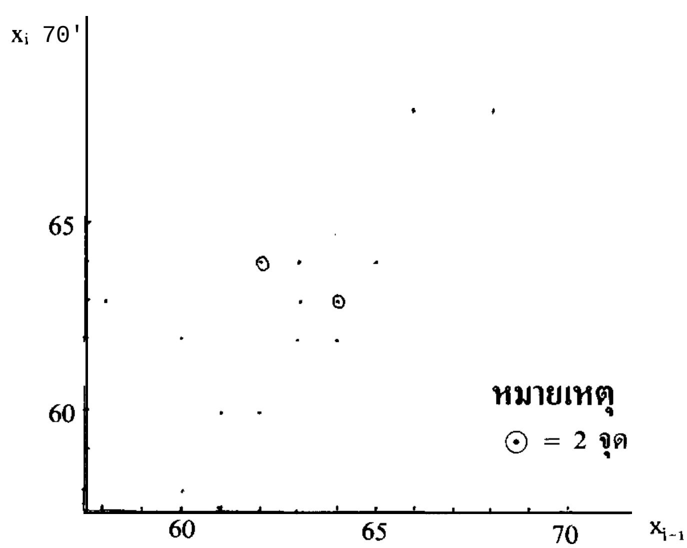
ข) คำนวณค่า r_1 (first serial correlation coefficient)

ค) ทดสอบ $H_0 : \rho_1 = 0,\ H : \rho_1 \neq 0$ ที่ $\alpha = .05$

กำหนด $\sum_{i=1}^{n} x_i^2 = 80667$, $\sum_{i=2}^{n} x_{i-1}\, x_i = 76140$ และ $\sum_{i=1}^{n} x_i = 1269$

วิธีทำ

ก) จุด 19 จุดที่ใช้สร้าง Scatter diagram คือ (65, 64), (64, 63), (63, 61), (61, 60), (60, 58), (58, 63), (63, 64), (64, 62), (62, 64), (64, 63), (63, 63), (63, 62), (62, 60), (60, 62), (62, 64), (64, 66), (66, 68), (68, 68), (68, 69)



ข) $n = 20$, $\sum_{i=1}^{n} x_i^2 = 80667$, $\sum_{i=1}^{n} x_i = 1269$, $\bar{x} = 63.45$

$\sum_{i=2}^{n} x_i = 1204$, $\sum_{i=1}^{n-1} x_i = 1200$, $\sum_{i=2}^{n} x_{i-1} x_i = 76140$

$$\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2/n = 80667 - (1269)^2/20$$

$$= 80667 - 80518.05$$

$$= 148.95$$

$$, \quad \sum_{i=2}^{n} x_{i-1} x_i - \bar{x}\left[\sum_{i=1}^{n-1} x_i + \sum_{i=2}^{n} x_i\right] + (n-1)\bar{x}^2$$

$$= 76140 - (63.45)\left[1200 + 1204\right] + (20-1)(63.45)^2$$

$$= 98.3475$$

$$\therefore \quad r_1 = 98.3475/148.95 = .6603$$

ค) 1) $H_0 : \rho_1 = 0$

2) $H_1 : \rho_1 \neq 0$

3) $\alpha = .05$

4) $CR : Z > z_{.025} = 1.96$ และ $Z < -1.96$

5) $z_c = \sqrt{n}\, r_1 = \sqrt{20}\,(.6603) = 2.9530$

6) $\because z_c > 1.96$ เราปฏิเสธ H_0 ที่ $\alpha = .05$ นั่นคือข้อมูลชุดนี้มีความสัมพันธ์กันในตัวข้อมูลเอง

5.13 จากข้อมูลที่วัดต่อเนื่องกันของปริมาณซัลเฟอร์ของโลหะชนิดหนึ่งที่อยู่ในเครื่องทำความร้อนในบ้าน คือ 2.0, 2.8, 2.0, 1.9, 1.1, 1.4, 1.2, 2.4, 2.5, 2.3, 3.1, 5.2, 2.8, 2.7, 2.6 2.4, 3.3, 3.9, 3.0, 3.8, 2.9

ก) จงสร้าง Scatter diagram

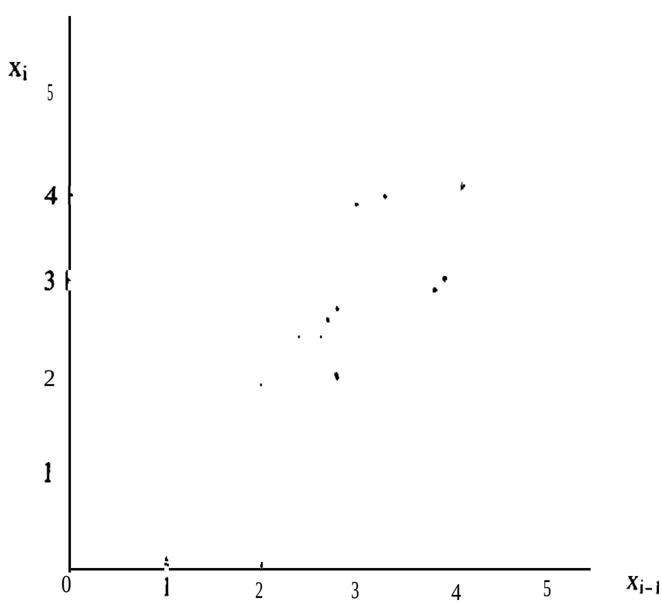
ข) จงคำนวณค่า r_1

ค) จงทดสอบ $H_0 : \rho_1 = 0, H_1 : \rho_1 \neq 0$ ที่ $\alpha = .05$

กำหนด $\sum_{i=1}^{n} x_i^2 = 163.61$, $\sum_{i=2}^{n} x_{i-1} x_i = 148.48$, และ $\sum_{i=1}^{n} x_i = 55.3$

วิธีทำ

ก) จุด 20 จุดสำหรับสร้าง Scatter diagram คือ (2.0, 2.8), (2.8, 2.0), (2.0, 1.9) (1.9, 1.1), (1.1, 1.4), (1.4, 1.2), (1.2, 2.4), (2.4, 2.5), (2.5, 2.3), (2.3, 3.1), (3.1, 5.2), (5.2, 2.8), (2.8, 2.7), (2.7, 2.6), (2.6, 2.4), (2.4, 3.3), (3.3, 3.9), (3.9, 3.0), (3.0, 3.8), (3.8, 2.9)



ข) $n = 21, \sum_{i=1}^{n} x_i^2 = 163.61, \sum_{i=1}^{n} x_i = 55.3, \bar{x} = 2.6333$

$$\sum_{i=2}^{n} x_i = 53.3, \sum_{i=1}^{n-1} x_i = 52.4, \sum_{i=2}^{n} x_{i-1} x_i = 148.48$$

$$\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2/n = 163.61 - (55.3)^2/21$$

$$= 163.61 - 145.6233$$

$$= 11.9861$$

$$\sum_{i=2}^{n} x_{i-1} x_i - \bar{x}\left[\sum_{i=1}^{n-1} x_i + \sum_{i=2}^{n} x_i\right] + (n-1)\bar{x}^2$$

$$= 148.48 - (2.6333)[52.4 + 53.3] + (21-1)(2.6333)^2$$

$$= 8.8262$$

$\therefore \quad r_1 = 8.8262/17.9867 = .4907$

ค) 1) $H_0 : \rho_1 = 0$

2) $H_1 : \rho_1 \neq 0$

3) $\alpha = .05$

4) CR: $Z > 1.96$ และ $Z < -1.96$

5) $z_c = \sqrt{n}\, r_1 = \sqrt{21}\,(.4907) = 2.2487$

6) $\because z_c > 1.96$ เราปฏิเสธ H_0 ที่ $\alpha = .05$ นั่นคือข้อมูลชุดนี้มีความสัมพันธ์กันในตัวข้อมูลเอง