11. ความถดถอยเชิงเดี่ยวและสหสัมพันธ์

- การประมาณค่าโดยเส้นถดถอย
- การวิเคราะห์ความถดถอย
- การวิเคราะห์สหสัมพันธ์

11.1 จงบอกลักษณะของสมการข้างล่างว่า อันใดเป็นแบบเชิงเดี่ยว, เชิงซ้อน, เส้นตรง และเส้นโค้ง

ก) $Y = 2 + 3Y$

ข) $Y = 1000 - 5000X$

ค) $Y = 10 + 2X_1 + 3X_2 + 4X_3$

ง) $Y = 2 + 3X + X^2$

ข้อ (ก) $Y = 2 + 3X$ เป็นสมการเส้นตรงแบบเชิงเดี่ยว

ข้อ (ข) $Y = 1000 - 5000X$ เป็นสมการเส้นตรงแบบเชิงเดี่ยว

ข้อ (ค) $Y = 10 + 2X_1 + 3X_2 + 4X_3$ เป็นสมการเชิงซ้อนของระนาบ (hyperplane)

ข้อ (ง) $Y = 2 + 3X + X^2$ เป็นสมการเชิงเดี่ยวแบบเส้นโค้ง

11.2 จงบอกจุดประสงค์ของการวิเคราะห์ความถดถอยและการวิเคราะห์สหสัมพันธ์

การวิเคราะห์ความถดถอยเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรคู่หนึ่ง ซึ่งอาจอยู่ในรูปเส้นตรงหรือเส้นโค้ง โดยใช้ตัวหนึ่งเป็นหลัก เรียกว่า ตัวแปรอิสระ ซึ่งเราสามารถกำหนดค่าต่าง ๆ ได้ล่วงหน้า ส่วนตัวแปรตามหรือตัวแปรไม่เป็นอิสระ จะเป็นผลที่ตามของตัวแปรอิสระ และไม่สามารถกำหนดค่าได้ล่วงหน้า เพราะเป็นตัวแปรเชิงสุ่ม หากมีตัวแปรอิสระหลายตัว เรียกว่า ความถดถอยเชิงซ้อน อย่างไรก็ตาม เรามุ่งหวังที่จะหาสมการที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม เพื่อจะได้ใช้สมการนั้นพยากรณ์ค่าตัวแปรตามต่อไป ส่วนการศึกษาสหพันธ์เป็นการวัดขนาดของความเกี่ยวพันกันของตัวแปรคู่หนึ่ง (degree of association) ว่ามีมากน้อยแค่ไหน

11.3 จงอธิบายความหมาย "ความถดถอย Y บน X" ตัวใดเป็นตัวแปรอิสระ และตัวใดเป็นตัวแปรพึ่งพา

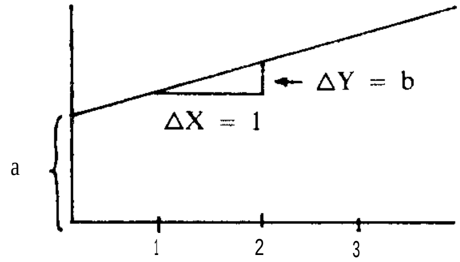
ความถดถอย Y บน X หรือ Regression Y on X หมายถึง การศึกษาความถดถอยโดยมีตัวแปร X เป็นตัวแปรอิสระ (คือ สามารถกำหนดค่าล่วงหน้าได้) ส่วน Y เป็นตัวแปรพึ่งพาหรือตัวแปรตาม ซึ่งค่าต่าง ๆ เป็นแบบเชิงสุ่ม และขึ้นอยู่กับอิทธิพลของ X ด้วย

11.4 จากสมการ $Y = a + bX$ จงอธิบายความหมายของค่า a และ b โดยรูปกราฟ

a จุดตัดที่แกน Y หรือ Y - intercept

b คือความลาดชันของเส้นถดถอย หรือ $\Delta Y/\Delta X$

ในเรื่องความถดถอย b หมายถึง อัตราการเปลี่ยนแปลงของ Y ต่อ 1 หน่วยของ X ที่เปลี่ยนแปลง



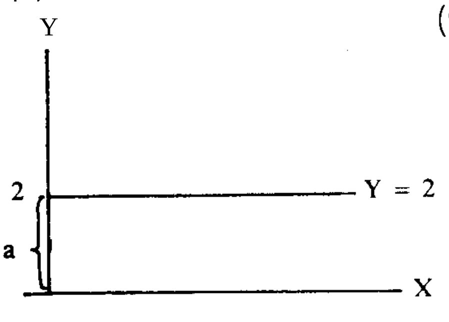
11.5 จงเขียนกราฟสมการ $Y = a + bX$ เมื่อ a และ b มีค่าต่าง ๆ ดังนี้

(ก) $a = 2, b = 0$ (ง) $a = -2, b = 2$

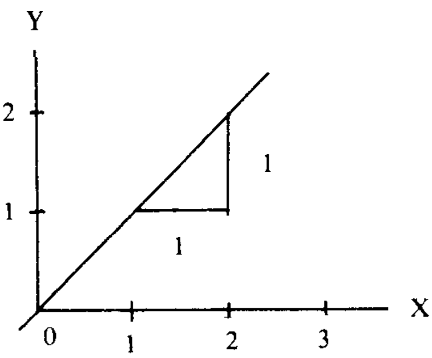
(ข) $a = 0, b = 1$ (จ) $a = 4, b = -2$

(ค) $a = 2, b = 1$

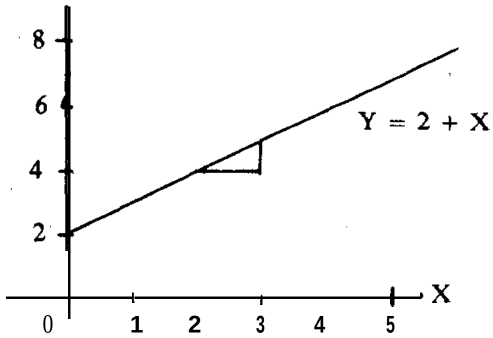
(ก)



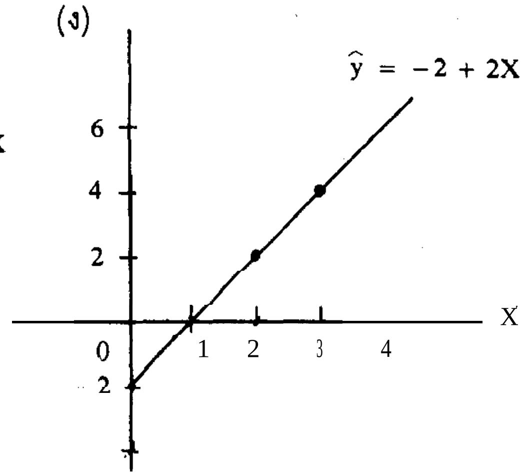
(ข)



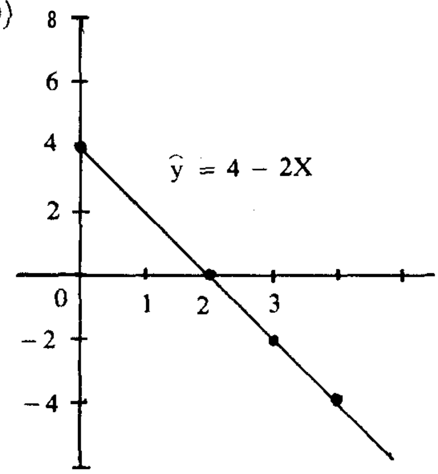
(ค)



(ง)



(จ)



11.6 จงใช้ข้อมูลที่กำหนดให้

ก) พล็อตใน scatter diagram

ข) สร้างสมการประมาณค่าที่ใช้แทนข้อมูลได้ดีที่สุด

ค) พยากรณ์ค่า Y เมื่อ X = 4, 9 และ 12

X	7	10	8	5	11	3	7	11	12	6
Y	2.0	3.0	2.4	1.8	3.2	1.5	2.1	3.8	4.0	2.2

$\Sigma X = 80$, $n = 10$, $\bar{X} = 8$, $\Sigma X^2 = 718$

$\Sigma Y = 26$, $\bar{Y} = 2.6$, $\Sigma Y^2 = 74.18$

$\Sigma XY = 229.6$

$$b = \frac{\Sigma XY - (\Sigma X)(\Sigma Y)/n}{\Sigma X^2 - (\Sigma X)^2/n} = \frac{229.6 - (80)(26)/10}{718 - (80)^2/10}$$

$$= \frac{21.6}{78} = .2769$$

$$a = \bar{Y} - b\bar{X} = 2.6 - (.2769)(8) = .3848$$

ข) $Y = .3848 + .2769X$

ค) เมื่อ $x = 4$

$\hat{Y} = .3848 + .2769(4) = 1.49$

เมื่อ $x = 9$

$\hat{Y} = .3848 + .2769(9) = 2.8769$

เมื่อ $x = 12$

$\hat{Y} = .3848 + .2769(12) = 3.71$

11.7 **จงใช้ข้อมูลที่กำหนดให้ข้างล่าง**

ก) **สร้าง** scatter diagram

ข) **สร้างสมการถดถอย**

ค) **พยากรณ์ค่า** Y **เมื่อ** X = 12, 14 **และ** 18

X	20	11	15	10	17	19
Y	5	15	14	17	8	9

(ข) $n = 6,\ \Sigma X = 92,\quad \bar{X} = 15.33,\ \Sigma X^2 = 1496$

$\Sigma Y = 68,\quad \bar{Y} = 11.33,\ \Sigma Y^2 = 880$

$\Sigma XY = 952$

$b = \frac{\Sigma XY - (\Sigma X)(\Sigma Y)/n}{\Sigma X^2 - (\Sigma X)^2/n} = \frac{952 - (92)(68)/6}{1496 - (92)^2/6} = \frac{-90.67}{85.53} = -1.06$

$a = \bar{Y} - b\bar{X} = 11.33 - (-1.06)(15.33) = 27.58$

สมการถดถอยคือ $Y = 27.58 - 1.06\ X$

(ค) เมื่อ $x = 12$

$\hat{Y} = 27.58 - 1.06\ (12) = 14.86$

เมื่อ $x = 14$

$\hat{Y} = 27.58 - 1.06\ (14) = 12.74$

เมื่อ $x = 18$

$\hat{Y} = 27.58 - 1.06\ (18) = 8.5$

11.8

X	56	48	42	58	40	39	50
Y	9.5	7.5	7.0	9.5	6.2	6.6	8.7

ก) หาเส้นตรงที่ปรับเข้ากับข้อมูลได้ดีที่สุด

ข) หาความคลาดเคลื่อนมาตรฐานของค่าประมาณ

(standard error of estimate หรือ $S_{Y/X}$

ค) หา 95% ช่วงเชื่อมั่นของค่าพยากรณ์ เมื่อ X = 44

$n = 7,\ \Sigma X = 333,\quad \bar{X} = 47.57,\quad \Sigma X^2 = 16{,}189$

$\Sigma Y = 55,\quad \bar{Y} = 7.857,\quad \Sigma Y^2 = 443.44$

$\Sigma XY = 2{,}677.4$

$\Sigma(X - \bar{X})^2 = \Sigma X^2 - (\Sigma X)^2/n = 347.72$

$\Sigma(Y - \bar{Y})^2 = \Sigma Y^2 - (\Sigma Y)^2/n = 11.30$

$\Sigma(X - \bar{X})(Y - \bar{Y}) = \Sigma XY - (\Sigma X)(\Sigma Y)/n = 60.97$

$b = 60.97/347.72 = 0.175$

$a = (\bar{Y} - b\bar{X}) = 7.857 - (0.175)(47.57) = -0.47$

สมการถดถอยคือ $Y = -0.47 + 0.175X$

ข) $S_{Y/X} = \sqrt{\dfrac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n-2}}$

$= \sqrt{\dfrac{443.44 + 0.47(55) - 0.175(2677.4)}{5}}$

$= \sqrt{\dfrac{.745}{5}} = \sqrt{.149} = .386$

ค) 95% ช่วงเชื่อมั่นของค่าพยากรณ์ เมื่อ $X = 44$ คือ

$\hat{Y} \pm t_{\alpha/2,\gamma} S_{\hat{Y}}$

เมื่อ $X = 44$, $\hat{Y} = -0.47 + .175(44) = 7.23$

$t_{\alpha/2,\gamma} = t_{.025,5} = \pm 2.571$

$S_{\hat{Y}} = S_{Y/X} \sqrt{1/n + \dfrac{(X - \bar{X})^2}{\Sigma(X - \bar{X})^2}}$

$= .386 \sqrt{\dfrac{1}{7} + \dfrac{(44 - 47.57)^2}{341.72}}$

$= .1635$

ดังนั้น 95% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ คือ

$7.23 \pm (2.571)(.1635)$

$= 7.23 \pm .42$

$= 6.81,\ 7.65$ หรือ $6.81 < \mu_{Y/X} < 7.65$

11.9 อุดมเป็นผู้จัดการร้านขายเครื่องไฟฟ้า เขาต้องการทราบว่า เงินที่เขาใช้โฆษณาโดยการซื้อเวลาจากโทรทัศน์และวิทยุ มีความสัมพันธ์กับจำนวนขายอย่างไร เขาได้เก็บข้อมูล คือ เวลาโฆษณาเป็นนาที (X) และจำนวนสินค้าหลักที่ขายได้ต่อสัปดาห์ (Y) ใน 7 สัปดาห์ มีดังนี้

x (นาที)	25	18	32	21	35	28	30
Y	16	11	20	15	26	32	20

$n = 7, \quad \Sigma X = 189, \quad \Sigma X^2 = 5323, \quad \bar{X} = 27$

$\Sigma Y = 140, \quad \Sigma Y^2 = 3102, \quad \bar{Y} = 20$

$\Sigma XY = 3959$

$\Sigma(X - \bar{X})^2 = \Sigma X^2 - (\Sigma X)^2/n = 220$

$\Sigma(X - \bar{X})(Y - \bar{Y}) = \Sigma XY - (\Sigma X)(\Sigma Y)/n = 179$

$\Sigma(Y - \bar{Y})^2 = \Sigma Y^2 - (\Sigma Y)^2/n = 302$

$b = \Sigma(X - \bar{X})(Y - \bar{Y})/\Sigma(X - \bar{X})^2 = 179/220 = .8136$

$a = \bar{Y} - b\bar{X} = 20 - (.8136)(27) = 1.97$

ก) จงสร้างสมการถดถอย

สมการถดถอย คือ

$$\hat{Y} = -1.97 + 0.8136X$$

ข) จงหา standard error of estimate

$$SY/X = \sqrt{\frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n - 2}}$$

$$= \sqrt{\frac{3102 + (1.97)(140) - (0.8136)(3959)}{7 - 2}}$$

$$= \sqrt{31.35} = 5.599 = 5.6$$

ค) จงหาช่วงเชื่อมั่นของค่าพยากรณ์ Y **เมื่อ** X $= 27, \alpha = .10$

90% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ คือ

$$\hat{Y} \pm t_{.05,5} S_{\hat{Y}}$$

เมื่อ $X = 27$, $\hat{Y} = -1.97 + 0.8136\ (27) = 19.997 = 20$

$t_{.05,5} = 2.015$

$$S_{\hat{Y}} = S_{Y/X} \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{\Sigma(X - \bar{X})^2}}$$

$$= 5.6 \sqrt{\frac{1}{7} + \frac{(27 - 27)^2}{220}} = (5.6)\ (.378)$$

$$= 2.1168$$

ดังนั้น **90%** ช่วงเชื่อมั่นของ $\mu_{Y/X}$ คือ

$20 \pm (2.015)\ (2.1168)$

$= 20 \pm 4.265$

$= 15.735 < \mu_{Y/X} < 24.265$

11.10 **ที่ปรึกษาฝ่ายผลิตของโรงงานหนึ่งให้คำแนะนำว่า โรงงานควรพิจารณาแจกงานซึ่งต้องใช้แรงงานสูงให้เหมาะสมกับอายุของพนักงาน เขาได้สุ่มพนักงานมา** 10 **คน และได้บันทึกระยะเวลาที่สามารถแบกหามของหนัก ดังนี้**

อายุ	42	27	36	25	22	39	57	19	33	30
นาทีของงานหนัก	2	7	5	9	10	4	4	8	6	5

ก) จงพล็อตข้อมูล

ข) จงสร้างสมการ แสดงความสัมพันธ์ของอายุและความสามารถของร่างกาย

ให้ X = อายุ, Y = นาทีแบกหามของหนัก

$n = 10,\ \Sigma X = 330,\ \bar{X} = 33,\ \Sigma X^2 = 12{,}018$

$\Sigma Y = 60,\ \bar{Y} = 6,\ \Sigma Y^2 = 416$

$\Sigma XY = 1782$

$\Sigma(X - \bar{X})^2 = \Sigma X^2 - (\Sigma X)^2/n = 1128$

$\Sigma(X - \bar{X})(Y \quad \hat{Y}) = \Sigma XY - (\Sigma X)(\Sigma Y)/n = -198$

$\Sigma(Y - \bar{Y})^2 = \Sigma Y^2 - (\Sigma Y)^2/n = 56$

$b = \Sigma(X - \bar{X})(Y - \bar{Y})/\Sigma(X - \bar{X})^2 = -198/1128 = -0.175$

$a = \bar{Y} - b\bar{X} = 6 - (-.175)(33) = 11.775$

สมการแสดงความสัมพันธ์ระหว่างอายุและเวลาแบกหามของหนักคือ

$Y = 11.775 - 0.175x$

ค) **เราจะคาดว่าพนักงานที่มีอายุ 30 ปี คนหนึ่ง จะสามารถแบกหามของหนักได้นานเท่าใด**

เมื่อ $x = 30,\ \hat{Y} = 11.775 - 0.175\ (30)$

$= 6.525$

นั่นคือ พนักงานที่มีอายุ 30 ปี จะแบกหามของหนักได้โดยเฉลี่ย 6.525 นาที

11.11 จงแสดงว่า เส้น $Y = a + bX$ และ $Y = \bar{Y} + b(X - \bar{X})$ คือเส้นเดียวกัน

$Y = a+bX$

$= (\bar{Y} - b\bar{X}) + bX$ (แทนค่า a)

$= \bar{Y} - b(X - \bar{X})$

11.12 บริษัทประกันชีวิตแห่งหนึ่งต้องการทราบความสัมพันธ์ของประสบการณ์ กับจำนวนขาย จึงสุ่มพนักงานมา 9 คน ให้ X คือจำนวนปีที่ทำงาน และ Y คือ จำนวนขายต่อปี มีหน่วยเป็น 100,000 บาท ได้ข้อมูลดังนี้

จำนวนขาย = Y	1	2	3	4	5	6	7	8	9
จำนวนปี = x	2	1	3	3	4	5	6	5	7

ก) จงประมาณจำนวนขายต่อปีของพนักงานผู้หนึ่ง ซึ่งทำงานมา 10 ปี

ข) มีหลักฐานพอเพียงที่จะแสดงว่า β ไม่เป็น 0 ที่ a = .05 ไหม?

ค) จงหา 95% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ เมื่อ x = 4.5 และ x = 10

ง) จงสร้าง 95% ช่วงเชื่อมั่นของ Y_a เมื่อ x = 10

$n = 9,\ \Sigma X = 36,\ \bar{X} = 4,\ \Sigma X^2 = 174,\ \Sigma(X - \bar{X})^2 = 30$

$\Sigma Y = 45,\ \bar{Y} = 5,\ \Sigma Y^2 = 285,\ \Sigma(Y - \bar{Y})^2 = 60$

$\Sigma XY = 220,\ \Sigma(X - \bar{X})(Y \quad \bar{Y}) = 40$

$b = \Sigma(X - \bar{X})(Y - \bar{Y})/\Sigma(X - \bar{X})^2 = 40/30 = 1.3333$

$a = \bar{Y} - b\bar{X} = 5 - (1.3333)(4) = -0.3333$

สมการประมาณค่า คือ $Y = -0.3333 + 1.3333\ x$

ก) ถ้าพนักงานผู้หนึ่งทำงานมา 10 ปี จะประมาณจำนวนขายต่อปีได้ดังนี้

$\hat{Y} = -0.3333 + 1.3333\,(10) = 13.0$ แสนบาท

นั่นคือ พนักงานที่ทำงานมา 10 ปี จะมีจำนวนขายปีละ 1,300,000 บาท

ข) $Ho: \beta = 0,\ Ha: \beta \neq 0,\ a = .05$

วิธีที่ 1 ใช้ t-test

ตัวสถิติที่ใช้ทดสอบคือ $T = b/S_b$

$S_b = S_{Y/X}/\sqrt{\Sigma(X - \bar{X})^2},$

$$S_{Y/X} = \sqrt{\frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n-2}} = \sqrt{\frac{285 - (-.3333)(45) - (1.3333)(220)}{9-7}}$$

$$= \sqrt{6.6667/7} = \sqrt{.9524} = .9759$$

$$S_b = .9759/\sqrt{30} = .178$$

ดังนั้น $T = \frac{b}{S_b} = \frac{1.3333}{.178} = 7.49$

และ T จะมีการแจกแจงแบบ $t_{.025,7} = \pm 2.365$

T = 7.49 > 2.365 จึงปฏิเสธ Ho และสรุปว่า $\beta \neq 0$ นั่นคืออายุการทำงาน และจำนวนขาย มีความสัมพันธ์แบบเชิงเส้น

วิธีที่ 2. ใช้ F-test

$SST = \Sigma(Y - \bar{Y})^2 = 60$

$SSR = \hat{\beta}\ \Sigma(X - \bar{X})(Y - \bar{Y}) = 53.3333$

$SSE = SST - SSR = 6.6661$

SOV	df	SS	MS	F
Regression	1	53.3333	53.3333	55.998
Error	7	6.6667	.9524	
รวม	8	60		

F จะมีการแจกแจงแบบ $f_{1,7,.05} = 5.59$

แต่ F คำนวณได้ 55.998 > 5.59 จึงปฏิเสธ Ho และมีผลสรุปเช่นเดียวกับ t-test

พึงสังเกตุว่า 1. $T = 1.49 = \sqrt{F} = \sqrt{55.998}$

2. $\sqrt{MSE} = \sqrt{.9524} = .9759 = S_{Y/X}$

ค. 95% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ เมื่อ x = 4.5 และ x = 10 คือ

$$\hat{Y} \pm t_{.025,7}\, S_{\hat{Y}}$$

เมื่อ $X = 4.5$, $\hat{Y} = -0.3333 + 1.3333\ (4.5) = 5.67$

และ $S_{\hat{Y}} = S_{Y/X}\sqrt{\dfrac{1}{n} + \dfrac{(X - \bar{X})^2}{\Sigma(X - \bar{X})^2}}$, $t_{.025,7} = 2.365$

$$= .9759\sqrt{\frac{1}{9} + \frac{(4.5 - 4.0)^2}{30}} = .337$$

ดังนั้น 95% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ เมื่อ x = 4.5 คือ

$$5.67 \pm (2.365)\ (.337) = 5.67 \pm .797$$

$$= 4.873;\ 6.467$$

นั่นคือ 95% ช่วงเชื่อมั่นของผู้ทำงานมา 4.5 ปี จะมีจำนวนขายเฉลี่ยต่อปี ระหว่าง 487,300 ถึง 646,700 บาท

และเมื่อ $x = 10$; $\hat{Y} = -0.3333 + 1.3333\ (10) = 13.0$

$$S_{\hat{Y}} = .9759\sqrt{\frac{1}{9} + \frac{(10 - 4)^2}{30}} = 1.117$$

ดังนั้น 95% ของ $\mu_{Y/X}$ เมื่อ x = 10 คือ

$$13.0 \pm (2.365)\ (1.117) = 13.0 \pm 2.64 = 10.36;\ 15.64$$

J) จงสร้าง 95% ช่วงเชื่อมั่นของ Y_a เมื่อ X = 10

$$\hat{Y}_a = -0.3333 + 1.3333\ (10) = 13.0$$

$$S_{\hat{Y}_a} = S_{Y/X}\sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\Sigma(X - \bar{X})^2}}$$

$$= .9759\sqrt{1 + \frac{1}{7} + \frac{(10-4)^2}{30}} = 1.48$$

ดังนั้น ช่วงเชื่อมั่น 95% ของ Y_a คือ

$$13 \pm (2.365)\ (1.48) = 13 \pm 3.5 = 9.5;\ 16.5$$

11.13 ในการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนผลงานของวิศวกร 10 คน และระยะเวลาที่ใช้วิชาชีพ ได้ข้อมูลดังนี้

Y = คะแนนผลงาน	1	2	3	4	5	5	6	7	8	9
X = อายุการใช้วิชาชีพ	1	3	2	5	5	4	7	6	9	8

$$n = 10,\ \Sigma X = 50,\ \bar{X} = 5,\ \Sigma X^2 = 310,\ \Sigma(X - \bar{X})^2 = 60$$

$$\Sigma Y = 50,\ \hat{Y} = 5,\ \Sigma Y^2 = 310,\ \Sigma(Y - \bar{Y})^2 = 60$$

$$\Sigma XY = 306,\ \Sigma(X - \bar{X})(Y - \bar{Y}) = 56$$

$$b = \Sigma(X - \bar{X})(Y - \bar{Y})/\Sigma(X - \bar{X})^2 = 56/60 = .93333$$

$$a = \bar{Y} - b\bar{X} = 5 - .93333\ (5) = .33333$$

ก) จงหาเส้นกำลังสองน้อยที่สุด เพื่อ 'fit" ข้อมูล

สมการกำลังสองน้อยที่สุด คือ

$$Y = a + bX = .33333 + .93333\ X$$

ข) จงประมาณคะแนนผลงานของผู้ใช้วิชาชีพมา 10 ปี

$$\hat{Y} = .33333 + 9.3333\ (10) = 93$$

ค) จงทดสอบความสัมพันธ์เชิงเส้นระหว่างคะแนนผลงาน และอายุการใช้วิชาชีพ โดยใช้ $\alpha = 0.01$

ก. ใช้ t-test

$$T = b/S_b \, , \; S_b = \frac{S_{Y/X}}{\sqrt{\Sigma(X-\bar{X})^2}}$$

$$S_{Y/X} = \sqrt{\frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n-2}} = \sqrt{\frac{310 - (.33333)(50) - .93333(306)}{8}}$$

$$= \sqrt{7.73361/8} = \sqrt{.9667} = .9832$$

$$S_b = .9832/\sqrt{60} = .12693$$

$$T = b/S_b = .93333/.12693 = 7.35$$

$T \sim t_{.005,8} = \pm 3.355$ แต่ $T = 7.35 > 3.355$ จึงปฏิเสธ Ho และสรุปว่า คะแนนผลงานและอายุการทำงานมีความสัมพันธ์แบบเชิงเส้น

ข. ใช้ F-test

$$SST = \Sigma(Y - \bar{Y})^2 = 60,$$

$$SSR = \hat{\beta}\, \Sigma(X - \bar{X})(Y - \bar{Y}) = 52.266661$$

$$SSE = 7.733333$$

SOV	df	SS	MS	F-ratio
regression	1	52.266661	52.266667	54.07
error	8	7.733333	0.966667	
รวม	9	60	$f_{1,8,.01}$ =	11.26

$F = 54.07 > 11.26$ จึงปฏิเสธ Ho และยอมรับ $Ha = \beta \neq 0$ และได้ข้อสรุปเหมือน t-test และพึงสังเกตว่า $\sqrt{F} = \sqrt{54.07} = 7.35 = T$ และ $\sqrt{MSE} = \sqrt{.966667} = .9832 = S_{Y/X}$

ง) จงหา 99% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ เมื่อ $X = 5$ และ $X = 1$

เมื่อ $X = 5, \hat{Y} = .33333 + 9.3333(5)) = 46.33335$

เมื่อ $X = 1$, $\hat{Y} = .33333 + 9.3333\,(1) = 9.00000$

$$S_{\hat{Y}} = S_{Y/X}\sqrt{\frac{1}{n} + \frac{(X-\bar{X})^2}{\Sigma(X-\bar{X})^2}}$$

เมื่อ $X = 5$, $= .9832\sqrt{\frac{1}{10} + \frac{(5-5)^2}{60}} = (.9332)\,(.316) = .31$

เมื่อ $X = 1$, $= .9832\sqrt{\frac{1}{10} + \frac{(1-5)^2}{60}} = (.9832)\,(.60553) = .60$

$t_{\alpha/2,\nu} = t_{.005,8} = \pm 3.355$

ดังนั้น 99% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ เมื่อ $X = 5$ คือ

$46.33335 \pm (3.355)\,(.31) \simeq 46.33335 \pm 1.04 = 45.29;\ 47.37$

และ 99% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ เมื่อ $X = 5$ คือ

$9 \pm (3.355)\,(.6) = 9 \pm 2.013 = 6.987;\ 11.013$

จ) จงหา 99% ช่วงเชื่อมั่นของ Y_a เมื่อ $X = 10$

$\hat{Y}_a = .33333 + 9.3333\,(10) = 93.6663$

$$S_{\hat{Y}_a} = S_{Y/X}\sqrt{1 + \frac{1}{n} + \frac{(X-\bar{X})^2}{\Sigma(X-\bar{X})^2}} = .9832\sqrt{1 + \frac{1}{10} + \frac{(10-5)^2}{60}}$$

$= (.9832)\,(1.23)$

$\simeq 1.21$

ดังนั้น ช่วงเชื่อมั่นคือ

$93.6663 \pm (3.355)\,(1.21) = 93.6663 \pm 4.05955$

$= 89.60675;\ 97.72585$

11.14 สถิติค่าใช้จ่ายในการซ่อมแซมเครื่องจักรชนิดเดียวกัน แต่มีอายุการใช้งานต่างกัน จำนวน 6 เครื่อง โดยให้ X คือ อายุการใช้งาน และ Y คือ ค่าบำรุงรักษา มีหน่วยเป็น 1,000 บาท มีดังนี้

X = อายุการใช้งาน	2	1	3	2	1	3
Y = ค่าบำรุงรักษา	70	40	100	80	30	100

$n = 6,\ \Sigma X = 12,\ \bar{X} = 2,\ \Sigma X^2 = 28,\ \Sigma(X - \bar{X})^2 = 4$

$\Sigma Y = 420,\ \bar{Y} = 70,\ \Sigma Y^2 = 33800,\ \Sigma(Y - \bar{Y})^2 = 4400$

$\Sigma XY = 970,\ \Sigma(X - \bar{X})(Y - \bar{Y}) = 130$

$b = \Sigma(X - \bar{X})(Y - \bar{Y})/\Sigma(X - \bar{X})^2 = 130/4 = 32.50$

$a = \bar{Y} - b\bar{X} = 70 \quad 32.50\ (2) = 5$

ก) จงหาสมการถดถอย Y **บน** X

สมการถดถอยคือ $Y = a + bX = 5 + 32.5\ X$

ข) จะต้องใช้ค่าบำรุงรักษาเท่าใดสำหรับเครื่องจักรที่มีอายุ 4 ปี

$\hat{Y} = 5 + 32.5\ (4) = 135$ (พันบาท)

$= 135{,}000$ บาท

ค) จงทดสอบความสัมพันธ์เชิงเส้นระหว่างอายุการใช้งานและค่าบำรุงรักษา. $\alpha = .01$

$$S_{y/x} = \sqrt{\frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{n - 2}}$$

$$= \sqrt{\frac{33800 - (5)(420) - (32.5)(970)}{4}}$$

$$= \sqrt{\frac{175}{4}} = \sqrt{43.75} = 6.614$$

$$S_b = \sqrt{\frac{S^2_{Y/X}}{\Sigma(X - \bar{X})^2}} = \sqrt{\frac{43.75}{4}} = \sqrt{10.9375} = 3.307$$

Ho : $\beta = 0$, Ha : $\beta \neq 0$

$T = b/S_b = 32.5/3.307 = 9.8276$

$t_{.005,4} = 4{,}604$, $T = 9.8276 > 4.604$ จึงปฏิเสธ Ho และยอมรับ Ha : $\beta \neq 0$ นั่นคือ อายุการใช้งาน (X) และค่าบำรุงรักษา (Y) มีความสัมพันธ์แบบเชิงเส้นตรง

ง) **จงหา** 99% **ช่วงเชื่อมั่นของ** $\mu_{Y/X}$ **เมื่อ** X = 2

$$\hat{Y} = 5 + 32.5\ (2) = 70$$

$$S_{\hat{Y}} = 6.614\sqrt{\frac{1}{6} + \frac{(2-2)^2}{4}} = (6.614)\ (.4082) = 2.7$$

ช่วงเชื่อมั่น 99% **คือ**

$$70 \pm (4.604)\ (2.7) = 70 \pm 12.43 = 57.57;\ 82.43$$

จ) จงหา 99% ของ Y_a เมื่อ x = 2

$$\hat{Y}_a = 5 + 32.5\ (2) = 70$$

$$S_{\hat{Y}_a} = 6.614\sqrt{1 + \frac{1}{6} + \frac{(2-2)^2}{4}} = 7.1439$$

ช่วงเชื่อมั่น 99% ของ Y_a คือ

$$70 \pm (4.604)\ (7.1439)$$

$$= 70 \pm 32.89$$

$$= 37.11;\ 102.89$$

นั่นคือ เครื่องจักรที่มีอายุการใช้งาน 2 ปี จะมีค่าบำรุงรักษา 37,110 ถึง 102,890 บาท ด้วยความเชื่อมั่น 99%

11.15 สถิติการป่วยต่อปีของพนักงาน 5 คน มีดังนี้

X = จำนวนปีที่ทำงาน	15	9	13	11	12
Y = จำนวนวันลาต่อปี	10	16	14	15	15

$n = 5,\ \Sigma X = 60,\ \overline{X} = 12,\ \Sigma X^2 = 740,\ \Sigma(X - \overline{X})^2 = 20$

$\Sigma Y = 70,\ \overline{Y} = 14,\ \Sigma Y^2 = 1002,\ \Sigma(Y - \overline{Y})^2 = 22$

$\Sigma XY = 821,\ \Sigma(X - \bar{X})(Y - \bar{Y}) = \quad 19$

$b = -19/20 = -0.95$

$a = 14 - (-.95)(12) = 25.4$

ก) จงหาสมการถดถอย Y บน X

สมการถดถอยคือ

$y = a + bX = 25.4 - 0.95\ X$

ข) จงหาจำนวนวันลาป่วยต่อปีของพนักงานที่ทำงานมา 14 ปี

$\hat{Y} = 25.4 - 0.95\ (14) = 12.1$ วัน

ค) มีเหตุผลเพียงพอที่จะเชื่อว่า X และ Y มีความสัมพันธ์เชิงเส้นไหม? $\alpha = .05$

$$S^2_{Y/X} = \frac{1}{(n-2)}(\Sigma Y^2 - a\Sigma Y - b\Sigma XY)$$

$$= \frac{1}{3}\{1002 - (25.4)(70) + (0.95)(821)\}$$

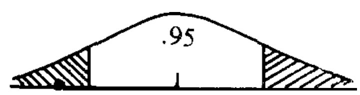
$$= \frac{1}{3}(3.95) = 1.31667$$

$$S_{Y/X} = \sqrt{1.31667} = 1.14746$$

$$S_b = \sqrt{S^2_{Y/X}/\Sigma(X - \bar{X})^2} = \sqrt{1.31667/20} = .2566$$

$Ho : \beta = 0,\ Ha : \beta \neq 0,\ a = .05$

$T = b/S_b = -0.95/.2566 = -3.702$



$t_{.025,3} = \pm 3.182$

$T = -3.702 < -3.182$ จึงปฏิเสธ Ho, ยอมรับ $Ha : \beta \neq 0$ และสรุปว่า จำนวนวันลา และอายุการทำงานมีความสัมพันธ์แบบเชิงเส้น

ง) จงหา 95% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ **เมื่อ** $X = 12$ **และ** $X = 15$

เมื่อ $X = 12,\ \hat{Y} = 25.4 - 0.95\ (12) = 14$ วัน

เมื่อ $X = 15,\ \hat{Y} = 25.4 - 0.95\ (15) = 11.15$ วัน

เมื่อ $X = 12$ จะมี $S_{\hat{Y}} = 1.14746\sqrt{\frac{1}{5} + \frac{(2-12)^2}{20}} = .513$

เมื่อ $X = 15$, จะมี $S_{\hat{Y}} = 1.14746\sqrt{\frac{1}{5} + \frac{(15-12)^2}{20}} = .925$

ดังนั้น 95% ช่วงเชื่อมั่นของ $\mu_{Y/X}$ คือ

เมื่อ $X = 12 \quad = \quad 14 \pm (3.182)\ (.513) = 14 \pm 1.63$

$= \quad 9.37;\ 15.63$

เมื่อ $X = 15 \quad = \quad 11.15 \pm (3.182)\ (.925) = 11.15 \pm 2.94$

$= \quad 8.21,\ 14.09$

11.16 ให้ X คือ คะแนนทดสอบความถนัด Y คือผลผลิตต่อชั่วโมง ข้อมูลสรุปจากพนักงาน 10 คน มีดังนี้

$n = 10,\ \Sigma X = 550,\ \Sigma Y = 680,\ \Sigma XY = 45{,}900$

$\Sigma X^2 = 38{,}500,\ \ \Sigma Y^2 = 56{,}000$

ก) จงหาสมการถดถอย Y on x

$\Sigma(X - \overline{X})^2 = 8250,\ \Sigma(Y - \overline{Y})^2 = 9760,\ \Sigma(X - \overline{X})(Y - \overline{Y}) = 8500$

$b = 8500/8250 = 1.03$

$a = \frac{680}{10} - 1.03\ \frac{(550)}{10} = 11.35$

สมการถดถอยคือ

$y = 11.35 + 1.03\ x$

ข) มีหลักฐานเพียงพอที่จะสรุปว่า สัมประสิทธิ์ความถดถอย β มีค่าแตกต่างจาก 0 ด้วย $\alpha = .02$ ไหม?

$$S^2_{Y/X} = \frac{1}{n-2}\left\{ \Sigma Y^2 - a\Sigma Y - b\Sigma XY \right\}$$

$$= \frac{1}{8}\left\{ 56{,}000 - (11.35)\ (680) - (1.03)\ (45{,}900) \right\}$$

$$= 1005/8 = 125.625$$

$S_{Y/X} = \sqrt{125.625} = 11.208$

$S_b = \sqrt{S^2_{Y/X}/\Sigma(X-\bar{X})^2} = \sqrt{125.625/8250} = 0.1234$

Ho : $\beta = 0$, Ha : $\beta \neq 0$, $a = .02$, $t_{.01,8} = \pm 2.896$

$T = b/S_b = 1.03/.1234 = 8.35$

$T = 8.35 > 2.896$ จึงปฏิเสธ Ho, ยอมรับ Ha : $\beta \neq 0$ และสรุปว่า คะแนนทดสอบความถนัด และจำนวนผลผลิตต่อชั่วโมงมีความสัมพันธ์แบบเชิงเส้น

ค) **จงหา** 08% **ช่วงเชื่อมั่นของ** $\mu_{Y/X}$ **เมื่อ** x = 40, x = 55, x = 70

เมื่อ $X = 40$, $\hat{Y} = 11.35 + 1.03\ (40) = 52.55$

เมื่อ $X = 55$, $\hat{Y} = 11.35 + 1.03\ (55) = 69.1$

เมื่อ $x = 70$, $\hat{Y} = 11.35 + 1.03\ (70) = 83.45$

และ $S_{\hat{Y}} = S_{Y/X}\sqrt{\frac{1}{n} + \frac{(X-\bar{X})^2}{\Sigma(X-X)^2}}$

$= 11.208\sqrt{\frac{1}{10} + \frac{(X-55)^2}{8250}}$

เมื่อ $x = 40$, $S_{\hat{Y}} = 3.998$

เมื่อ $x = 55$, $S_{\hat{Y}} = 3.54$

เมื่อ $X = 70$, $S_{\hat{Y}} = 3.998$

ดังนั้น 98% ช่วงเชื่อมันของ $\mu_{Y/X}$ คือ

$\hat{Y} \pm t_{.01,8} S_{\hat{Y}}$, $t_{.01,8}$ 3 $\pm$ 2.896

เมื่อ $x = 40$; $= 52.55 \pm (2.896)\ (3.998)$

$= 52.55 \pm 11.58 = 40.97,\ 64.13$

เมื่อ $x = 55$; $= 69.1 \pm (2.896)\ (3.54)$

$= 69.1 \pm 10.25 = 88.85,\ 79.35$

เมื่อ $x = 70$; $= 83.45 \pm (2.896)\ (3.998)$

$= 83.45 \pm 11.58 = 73.87,\ 95.03$

ง) **จงหา 98% ช่วงเชื่อมั่นของ Y_a**

$\hat{Y}_a$ จะมีค่า = $\hat{Y}$ สำหรับค่าประมาณแบบจุด

แต่จะมีความคลาดเคลื่อนมาตรฐานต่างจากเดิม เพราะ

$$S_{\hat{Y}_a} = S_{Y/X}\sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\Sigma(X - X)^2}}$$

$$= 11.208\sqrt{1 + \frac{1}{10}\ \frac{(X - 55)^2}{8250}}$$

เมื่อ $X = 40,\ S_{\hat{Y}_a} = 11.208\sqrt{1 + \frac{1}{10} + \frac{(40 - 55)^2}{8250}} = 11.9$

เมื่อ $X = 55,\ S_{\hat{Y}_a} = 11.208\sqrt{1 + \frac{1}{10} + \frac{(55 - 55)^2}{8250}} = 11.76$

เมื่อ $x = 70,\ S_{\hat{Y}_a} = 11.208\sqrt{1 + \frac{1}{10} + \frac{(70 - 55)^2}{8250}} = 11.9$

ดังนั้น 98% ช่วงเชื่อมั่นของ Y_a คือ

เมื่อ $x = 40;\ = 52.55 \pm (2.896)\ (11.9)$

$= 52.55 \pm 34.46$

$= 18.09,\ 87.01$

เมื่อ $X = 55\ = 69.1 \pm (2.896)\ (11.76) = 69.1 \pm 34.06$

$= 35.04,\ 103.16$

เมื่อ $x = 70\ = 83.45 \pm (2.896)\ (11.9)$

$= 83.45 \pm 34.46 = 48.99,\ 117.91$

จ) **ความแตกต่างของคำตอบในข้อ (ค) และ (ง) คือ** ค่าประมาณในข้อ (ค) เป็นค่าประมาณของค่าเฉลี่ย $\mu_{Y/X}$ ส่วน $\hat{Y}_a$ เป็นค่าประมาณของ Y อื่น ๆ ที่ไม่ใช่ค่าเฉลี่ย ดังนั้น $v(\hat{Y})$ จึงเล็กกว่า $V(\hat{Y}_a)$ เนื่องจากเป็นความแปรปรวนของค่าเฉลี่ย จึงทำให้ค่าประมาณในข้อ (ค) มีช่วงแคบกว่าค่าประมาณในข้อ (ง) แม้จะใช้ระดับความเชื่อมั่นเท่ากัน และพยากรณ์ที่จุดเดียวกันของ X

11.17 ข้อมูลข้างล่างคือรายได้ต่อครัวเรือนและค่าใช้จ่ายสำหรับอาหารของครอบครัวที่สุ่มมา 8 ครอบครัว ดังนี้

รายได้ (1000 บาท)	8	12	9	24	13	37	19	16
เปอร์เซนต์รายจ่ายค่าอาหาร	36	25	33	15	28	19	20	22

ก) จงสร้างสมการประมาณค่าที่ดีที่สุดที่ใช้อธิบายความสัมพันธ์ของข้อมูล

ข) จงหา $S_{Y/X}$

ค) จงหาช่วงเชื่อมั่น 90% ของเปอร์เซนต์รายจ่าย สำหรับอาหารของครอบครัวที่มีรายได้ 25,000 บาท

ให้ x = รายได้ และ Y = เปอร์เซนต์รายจ่ายค่าอาหาร เพราะในข้อ (n) ต้องพยากรณ์เปอร์เซนต์ค่าอาหารโดยกำหนดรายได้

$n = 8,\ \Sigma X = 138,\ \bar{X} = 17.25,\ \Sigma X^2 = 3020,\ \Sigma(X - \bar{X})^2 = 639.50$

$\Sigma Y = 198,\ \bar{Y} = 24.75,\ \Sigma Y^2 = 5264, \Sigma(Y - \bar{Y})^2 = 363.5$

$\Sigma XY = 3044,\ \Sigma(X - \bar{X})(Y - \bar{Y}) = -371.5$

$b = -371.5/639.5 = -0.5809$

$a = 24.75 + (.5809)(17.25) = 34.77$

ก) ดังนั้น สมการประมาณค่าที่ดีที่สุดคือ

$$\hat{Y} = 34.77 - 0.5809\ X$$

ข) $S_{Y/X} = \frac{1}{\sqrt{(n-2)}}\sqrt{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}$

$$\frac{1}{\sqrt{6}}\sqrt{5264 - (34.77)(198) + (.5809)(3044)}$$

$$= \sqrt{147.7996/6} = \sqrt{24.6333} = 4.963$$

ค) ช่วงเชื่อมั่น 90% ของ $\mu_{Y/X}$ เมื่อ x = 25 คือ

$$\hat{Y} \pm t_{.05,6}\, S_{\hat{Y}}$$

$$\hat{Y} = 34.17 - 0.5809\,(25) = 20.2475$$

$$S_{\hat{Y}} = 4.963\sqrt{\frac{1}{8} + \frac{(25 - 17.25)^2}{639.50}} = 2.32$$

$$t_{.05,6} = 1.943$$

ดังนั้น ช่วงเชื่อมั่น $\mu_{Y/X}$ คือ

$$20.2475 \pm (1.943)\,(2.32)$$

$$20.2475 \pm 4.51$$

$$= 15.7375, \quad 24.1575$$

นั่นคือ กล่าวด้วยความเชื่อมั่น 90% ได้ว่า ครอบครัวที่รายได้เดือนละ 25,000 บาท จะใช้จ่ายสำหรับอาหารโดยเฉลี่ยระหว่าง 15.7375% ถึง 24.7575%

11.18 ข้อมูลข้างล่าง คือจำนวนขายเป็นพันบาทต่อเดือน และค่าประมาณของจำนวนขายเป็นเวลา 10 เดือน ของบริษัทขายเครื่องกีฬา จงหาสัมประสิทธิ์การตัดสินใจจากตัวอย่าง และสัมประสิทธิ์สหสัมพันธ์จากตัวอย่าง

Y	55	64	54	63	68	70	76	66	75	74
$\hat{Y}$	55.5	59.5	60.5	63.5	67.5	65.5	73.5	70.5	72.5	76.5
$Y - \hat{Y}$	-0.5	4.5	-6.5	-0.5	0.5	4.5	2.5	-4.5	2.5	-2.5

$$\Sigma(Y - \hat{Y})^2 = (-0.5^2 + 4.5^2 + \ldots + -2.5^2) = 122.5 = SSE$$

$$\Sigma Y = 665,\; n = 10,\; \Sigma Y^2 = 44763$$

$\Sigma(Y - \bar{Y})^2 = \Sigma Y^2 - (\Sigma Y^2)/n = 540.5 = SST$

$SSR = \Sigma(\hat{Y} - \bar{Y})^2 = SST - SSE = 418$

ดังนั้น r^2 = สัมประสิทธิ์การตัดสินใจจากตัวอย่าง

= SSR/SST

= $418/540.5 = 0.7733$

และ r = สัมประสิทธิ์สหสัมพันธ์จากตัวอย่าง = $\sqrt{r^2}$

= $\sqrt{.7733} = .87991$

11.19 ผู้จัดการบริษัทขายน้ำกรองสังเกตว่ามีการเปลี่ยนแปลงจำนวนการบริโภคน้ำ ซึ่งแต่เดิมพบว่า เส้นแสดงความสัมพันธ์ระหว่างจำนวนลูกค้า (บ้าน) กับจำนวนบริโภค มีความลาดชัน 13 หน่วย จงใช้ $\alpha = .10$ ทดสอบว่า ความลาดชันเพิ่มสูงขึ้นหรือไม่ จากข้อมูลต่อไปนี้

จำนวนบ้าน I(1000 หลัง)	8.1	7.8	8.4	7.6	8.0	8.1
จำนวนการบริโภคน้ำกรอง (1,000 แกลลอน)	94	83	97	85	89	92

ให้ จำนวนบ้าน = X และ Y = จำนวนการบริโภค

$n = 6,\ \Sigma X = 48,\ \bar{X} = 8,\ \Sigma X^2 = 384.38,\ \Sigma(X - \bar{X})^2 = 0.38$

$\Sigma Y = 540,\ \bar{Y} = 90,\ \Sigma Y^2 = 48744,\ \Sigma(Y - \bar{Y})^2 = 144$

$\Sigma XY = 4326.8,\ \Sigma(X - \bar{X})(Y - i') = 6.8$

$b = 6.8/.38 = 17.894736$

$SSR = \hat{\beta}\ \Sigma(X \quad \bar{X})\ (Y - \bar{Y}) = 121.68$

SST = Σ(Y − i')' = 144

$SSE = SST - SSE = 22.32$

$MSE = SSE/(n-2) = 22.32/4 = 5.5789$

$S_{Y/X} = \sqrt{MSE} = 2.36$

$$S_b = \sqrt{\frac{MSE}{\Sigma(X - \bar{X})^2}} = 3.8316367$$

1) Ho : $\beta = 13$

2) Ha : $\beta > 13$

3) a = .10

4) จะปฏิเสธ Ho เมื่อ $T > t_{4,.10} = 1.533$

5) ตัวสถิติที่ใช้ทดสอบคือ

$$T = \frac{b - \beta_0}{S_b} = \frac{17.894736 - 13}{3.8316367} = 1.277$$

6) $T = 1.277 < 1.533$ จึงยังปฏิเสธ Ho ไม่ได้ นั่นคือ ยังต้องยอมรับ Ho ว่าความลาดชัน เป็น 13 หน่วยเท่าเดิม

11.20 ในการสำรวจตลาดใน 10 ท้องที่ ของบริษัทจำหน่ายรถจักรยานยนต์ พบว่า จำนวนเงินที่ใช้โฆษณา และจำนวนขายมีความสัมพันธ์แบบเชิงเส้น โดยมีความลาดชันเป็น 1.5 หน่วย และความคลาดเคลื่อนมาตรฐานของ b = .35 หน่วย ข้อมูลนี้จะขัดแย้งที่ระดับนัยสำคัญ 10% กับคำกล่าวอ้างของฝ่ายตลาด ที่ว่าจะขายมอเตอร์ไซด์ ได้เพิ่มขึ้น 60,000 คัน สำหรับรายจ่ายค่าโฆษณาที่เพิ่มขึ้น 25 ล้านบาทหรือไม่?

X มีหน่วยเป็น 1000 คัน, Y มีหน่วยเป็นล้านบาท

$b = 1.5,\ S_b = .35,\ t_{.05,8} = \pm 1.86$

ดังนั้น ช่วงเชื่อมั่น 90% ของ β คือ

$$
\begin{aligned}
& b \pm t_{.05,(n-2)} (S_b) \\
= \;& 1.5 \pm (1.86)(.35) \\
= \;& 1.5 \pm .651 \\
= \;& .849,\ 2.151
\end{aligned}
$$

นั่นคือ ถ้าเพิ่มค่าโฆษณา 1 ล้านบาท จำนวนขายจะเพิ่มขึ้นระหว่าง 849 ถึง 2,151 คัน ด้วยความเชื่อมั่น 90%

ดังนั้น ถ้าเพิ่มค่าโฆษณา 25 ล้านบาท จำนวนขายจะเปลี่ยนแปลงระหว่าง

$.849(25) = 21.225 = 21{,}225$ คัน และ

$2.151(25) = 53.775 = 53{,}775$ คัน

จะเห็นว่าช่วงเชื่อมั่นไม่รวม $\beta = 60$ (60,000 คัน) ดังนั้น ข้อมูลที่ได้จึงขัดแย้งกับคำกล่าวของฝ่ายตลาด เพราะอัตราการเพิ่มต่ำกว่า 60,000 คัน

นอกจากการวิเคราะห์ด้วยช่วงเชื่อมั่นแล้ว ยังใช้วิธีทดสอบสมมติฐานก็ได้ ดังนี้

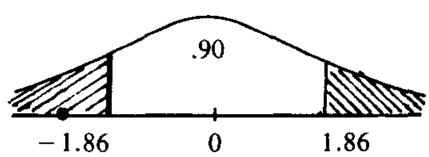
ถ้าเพิ่มค่าโฆษณา 25 ล้านบาท ทำให้จำนวนขายเพิ่มขึ้น 60 (พันคัน)

นั่นคือ ค่าโฆษณา 1 ล้านบาท ทำให้จำนวนขายเพิ่มขึ้น 60/25 = 2.4 (พันคัน)

นั่นคือ β' หรือ $\beta_0 = 2.4$

Ho : $\beta = 2.4$, Ha : $\beta \neq 2.4$, $\alpha = .10$

$$T = \frac{b - \beta'}{S_b} = \frac{1.5 - 2.4}{.35} = -2.57$$



T = −2.57 อยู่ในเขตวิกฤต จึงปฏิเสธ Ho และยอมรับ Ha : $\beta \neq 2.4$

นั่นคือ อัตราเปลี่ยนแปลงต่ำกว่าที่คาดไว้ เพราะค่าสถิติ T ตกอยู่ในเขตวิกฤตด้าน $-\infty$ ($t_{.05,8} = \pm 1.86$)

11.21 ข้อใดคือความแตกต่างระหว่างสัมประสิทธิการตัดสินใจ และสัมประสิทธิสหสัมพันธ์

ก) บอกได้ว่าความลาดชันของเส้นถดถอยเป็นแบบเชิงบวกหรือเชิงลบ

r^2 ไม่บอกทิศทางความลาดชัน เพราะมีเครื่องหมายเป็น + เสมอ แต่ r บอกทิศทางของความลาดชัน เพราะ $-1 \leq r \leq 1$

ข) วัดความแรงของความเกี่ยวพันระหว่าง 2 ตัวแปร

r วัดความแรงของความเกี่ยวพันกันแบบเชิงเส้น ถ้า r มีค่าใกล้เคียง ±1 แสดงว่า มีความเกี่ยวพันสูง แต่ถ้า r มีค่าใกล้ 0 แสดงว่า มีความเกี่ยวพันกันน้อย ในขณะที่ r^2 วัดความผันแปรของ Y ที่อธิบายได้โดยตัวแปร X หรือโดยสมการเชิงเส้น ถ้า $r^2 \rightarrow 1$ แสดงว่า X และ Y มีความสัมพันธ์เชิงเส้นสูง เพราะความผันแปรส่วนใหญ่ สามารถอธิบายได้โดย X แต่ถ้า $r^2 \rightarrow 0$ แสดงว่า X และ Y มีความสัมพันธ์เชิงเส้นค่อนข้างน้อย เส้นตรง (X) จึงอธิบายความผันแปรของ Y ได้น้อยมาก

ค) จะมีค่าไม่เกิน 1

ทั้ง r และ r^2 จะมีค่าไม่เกิน 1

ง) วัดเปอร์เซนต์ของความผันแปรที่อธิบายโดยความถดถอย

r^2 วัดเปอร์เซนต์ความผันแปรของ Y ที่อธิบายโดยเส้นถดถอย หรือโดยตัวแปร X แต่ r เป็นเพียงสัมประสิทธิ์ที่แสดง degree ของความสัมพันธ์แบบเชิงเส้นของ X และ Y

11.22. โรงงานผลิตเครื่องปรับอากาศทราบว่าจำนวนขายในฤดูร้อนจะไม่แน่นอน และจะมีความสัมพันธ์กับอุณหภูมิ จากการศึกษาข้อมูลเดิม เมื่อ 6 ปีก่อน มีดังนี้

อุณหภูมิ (°ซ)	29”	33”	35”	34”	36”	31”
จำนวนขาย (เครื่อง)	66	74	72	76	78	72

จงสร้างสมการที่อธิบายความสัมพันธ์ของข้อมูลนี้ได้ดีที่สุด

ให้ X = อุณหภูมิ และ Y = จำนวนขาย และสมมุติว่า X และ Y มีความสัมพันธ์แบบเชิงเส้น จึงมีสมการแสดงความสัมพันธ์ $Y = a + bX$ ซึ่งจะสร้างจากข้อมูล ดังนี้

$n = 6,\ \Sigma = 198,\ \overline{X} = 33,\ \Sigma X^2 = 6568,\ \Sigma(X - \overline{X})^2 = 34$

$\Sigma Y = 438,\ \overline{Y} = 73,\ \Sigma Y^2 = 32060,\ \Sigma(Y - \overline{Y})^2 = 86$

$\Sigma XY = 14500,\ \Sigma(X - \overline{X})(Y - \overline{Y}) = 46$

$b = 46/34 = 1.35294$

$a = 73 - (1.35294)(33) = 28.35$

ดังนั้น สมการที่อธิบายความสัมพันธ์ของจำนวนขายและอุณหภูมิ คือ

$Y = 28.35 + 1.35294\ X$

สมการนี้ จะเป็นสมการที่ดีที่สุด หรือจะใช้ได้หรือไม่ต้องทดสอบว่า β มีค่าต่างจาก 0 หรือไม่ ดังนี้

$SST = \Sigma(Y - \overline{Y})^2 = 86$

$SSR = b\,\Sigma(X - \overline{X})(Y - \overline{Y}) = 62.2353 = MSR$ และมี 1 df

$SSE = SST - SSR = 23.7647$

$MSE = SSE/(n - 2) = 23.7647/4 = 5.9412$ และมี 4 df

$Ho : \beta = 0,\ Ha : \beta \neq 0,\ f_{1,4,.05} = 7.71$

$F = MSR/MSE = 62.2353/5.9412 = 10.48$

ค่าสถิติ $F = 10.48 > 7.71$ จึงปฏิเสธ Ho และยอมรับ $Ha : \beta \neq 0$

นั่นคือ X และ Y มีความสัมพันธ์แบบเชิงเส้นดังนี้

$Y = 28.35 + 1.35294\ X$

นั่นคือ ถ้าอุณหภูมิเพิ่มสูงขึ้น 1° จะทำให้จำนวนขายเครื่องปรับอากาศเพิ่มขึ้น 1.35 เครื่อง

11.23 ในการวิเคราะห์ความถดถอย เราสามารถกำหนดค่าของ X ได้ล่วงหน้า แต่ค่าของ Y เป็นตัวแปรเชิงสุ่ม อยากทราบว่าค่าของ X และ Y จะเป็นแบบใด เมื่อเราวิเคราะห์สหสัมพันธ์

ในการวิเคราะห์สหสัมพันธ์ เราไม่สนใจว่าตัวแปรหนึ่งมีอิทธิพลเหนือตัวแปรอีกตัวหนึ่ง แต่มุ่งสนใจว่า ตัวแปรคู่นั้นมีความสัมพันธ์กันมากหรือน้อยเพียงใด ดังนั้น ตัวแปรทั้งคู่ จึงเป็นตัวแปรเชิงสุ่ม ซึ่งมีการแจกแจงร่วมกันแบบโค้งปกติ ซึ่งเรียกว่า การแจกแจงแบบ bivariate normal

11.24 ถ้าเรามีตัวอย่างอยู่ 2 ชุด และมีค่า $r = 0.6$ และ $r = 0.8$ เราจะกล่าวถึงความสัมพันธ์ระหว่าง X และ Y ว่าอย่างไร เมื่อเรานำข้อมูล 2 ชุดนี้มาเปรียบเทียบกัน

เราไม่นิยมใช้ค่า r เปรียบเทียบกัน เพราะไม่ชัดเจน เพราะเราทราบเพียงว่า $r = .8$- สูงกว่า $r = .6$ แต่ไม่ทราบว่าดีกว่าแค่ไหน ควรหาค่า r^2 ดังนี้ เมื่อ $r = .6, r^2 = .36$ และเมื่อ $r = .8, r^2 = .64$ นั่นคือ ในชุดแรก ตัวแปรอิสระ X อธิบายความผันแปรของ Y ได้เพียง 36% แต่ในชุดหลัง ตัวแปรอิสระสามารถอธิบายความผันแปรของ Y ได้ 64%

11.25 บริษัทโฆษณาต้องการทราบว่า จำนวนครั้งที่โฆษณาทางโทรทัศน์ต่อสัปดาห์ และจำนวนขายสินค้าชนิดหนึ่งต่อสัปดาห์ มีสหสัมพันธ์กันหรือไม่ จึงได้ศึกษาจากเมืองตัวอย่าง 9 เมือง ดังนี้

เมือง	A	B	C	D	E	F	G	H	I
จำนวนโฆษณา (X)	1	2	3	4	5	6	7	8	9
จำนวนขาย (Y) (100 หน่วย)	2	1	3	3	4	5	6	5	7

ก) จงหาสัมประสิทธิ์สหสัมพันธ์

ข) จงทดสอบ Ho : $\rho = 0$, a = .05

$n = 9$, $\Sigma X = 45$, $\bar{X} = 5$, $\Sigma X^2 = 285$, $\Sigma(X - \bar{X})^2 = 60$

$\Sigma Y = 36$, $\bar{X} = 4$, $\Sigma Y^2 = 174$, $\Sigma(Y - \bar{Y})^2 = 30$

$\Sigma XY = 220$, $\Sigma(X - \bar{X})(Y - \bar{Y}) = 40$

$SSR = \{\Sigma(X - \bar{X})(Y - \bar{Y})\}^2/\Sigma(X - \bar{X})^2 = 26.6667$

$SST = \Sigma(Y - \bar{Y})^2 = 30$

$SSE = SST - SSR = 3.3333$

$r^2 = SSR/SST = .88889$

ก) $r = \sqrt{r^2} = \sqrt{.88889} = .9428$

ข) Ho : $\rho = 0$, Ha: $\rho \neq 0$, a = .05

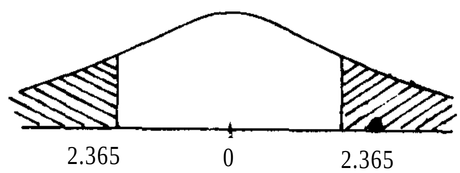
$t_{.025,7} = \pm 2.365$

ตัวสถิติที่ใช้ทดสอบคือ

$$T = \frac{r\sqrt{(n-2)}}{\sqrt{1-r^2}} = \sqrt{\frac{r^2(n-2)}{1-r^2}}$$

$$= \sqrt{\frac{.88889\,(7)}{1 - .88889}} = \sqrt{\frac{6.22223}{.11111}}$$

$$= \sqrt{56} = 7.48$$



T = 7.48 > 2.365 จึงปฏิเสธ Ho และสรุปว่า $\rho \neq 0$ คือ X และ Y มีสหสัมพันธ์แบบเชิงเส้น

11.26 งานทดลองเพื่อศึกษาความสัมพันธ์ของปริมาณน้ำฝน และผลผลิตข้าวมีดังนี้

ปริมาณน้ำฝน (นิ้ว) = X	1	2	3	4	5	5	6	7	8	9
ผลผลิต (บุชเชล) = Y	1	3	2	5	5	4	7	6	9	8

ก) จงหาสัมประสิทธิ์สหสัมพันธ์

ข) ทดสอบ Ho : $\rho = 0$, a = .025

$n = 10,\ \Sigma X = 50,\ \overline{X} = 5,\ \Sigma X^2 = 310,\ \Sigma(X - \overline{X})^2 = 60$

$\Sigma Y = 50,\ \overline{Y} = 5,\ \Sigma Y^2 = 310,\ \Sigma(Y - \overline{Y})^2 = 60$

$\Sigma XY = 306,\ \Sigma(X \quad \overline{X})(Y - \overline{Y}) = 56$

$SSR = \{\Sigma(X - \overline{X})(Y - \overline{Y})\}^2/\Sigma(X - \overline{X})^2 = (56)^2/60 = 52.266667$

$SST = \Sigma(Y - \overline{Y})^2 = 60$

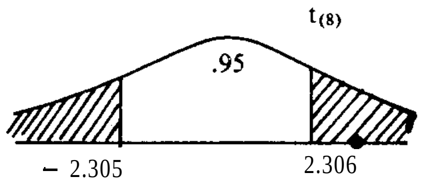
$r^2 = SSR/SST = 52.266667/60 = .871111$

ก) $r = \sqrt{r^2} = \sqrt{.871111} = .93333$

ข) Ho : $\rho = 0$, Ha : $\rho > 0$, a = .05, $t_{.025,8} = 2.306$

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \sqrt{\frac{r^2(n-2)}{1-r^2}} = \sqrt{\frac{(.87111)(8)}{.1288889}}$$

$$= \sqrt{54.0689} = 7.35$$



T = 7.35 > 2.306 จึงปฏิเสธ Ho และยอมรับ Ha : $\rho \neq 0$

นั่นคือ ปริมาณน้ำฝนและผลผลิตมีสหสัมพันธ์แบบเชิงเส้น