VI การวิเคราะห์เกี่ยวกับการถดถอย (Regression Analysis)

ความหมายและสมการของการถดถอย

ในชีวิตประจำวันเราอาจจะต้องการข้อเท็จจริง เกี่ยวกับสิ่งต่าง ๆ ควบกันไป ต้องการดูความสัมพันธ์ระหว่างสิ่งต่าง ๆ เช่นจำนวนผลผลิตที่ได้เกี่ยวข้องกับ จำนวนน้ำและปุ๋ย หรือราคาสินค้า เกี่ยวข้องกับ เวลา ค่าวัตถุดิบ ค่าแรงงาน เป็นต้น

เราอาจจะแสดงความสัมพันธ์ระหว่างตัวแปรที่สนใจในรูปฟังก์ชัน เช่น Y เป็นฟังก์ชันของ X เราใช้คำว่า "การถดถอย" (Regression) แสดงความสัมพันธ์ระหว่าง Y กับ X ค่าของ Y ขึ้นอยู่กับค่าของ X เราเรียก X ว่าตัวแปรอิสระ (Independent Variable) และเรียก Y ว่าตัวแปรพึ่งพิง (Dependent Variable)

จุดมุ่งหมายที่สำคัญของการถดถอย ก็คือ เพื่อศึกษาว่าค่าของ Y ขึ้นอยู่กับค่าของ X หรือไม่ หรือเพื่อที่จะทำนายค่าของ Y โดยอาศัยค่าของ X หรือศึกษารูปร่างลักษณะของโค้งการถดถอย

ตัวแบบถดถอยเชิงเส้นแบบเชิงเดี่ยว (Simple Linear Regression Model)

ตัวแบบของการถดถอยเชิงเส้น จะอยู่ในรูป

$$Y = \beta_1 + \beta_2 X + \varepsilon$$

X เป็นตัวแปรอิสระหรือตัวแปรอธิบายได้ X, Y สามารถวัดหรือสังเกตได้

ε เป็นตัวคลาดเคลื่อน ซึ่งไม่สามารถวัดและสังเกตได้

β_1, β_2 เป็นพารามีเตอร์ของการถดถอยที่ไม่ทราบค่า

ข้อสมมติเบื้องต้นเกี่ยวกับตัวแบบถดถอยเชิงเส้น คือ

(1) ε เป็นตัวแปรเชิงสุ่มที่มีการกระจายแบบปกติ $N(0, \sigma^2)$

(2) ตัวแปรอิสระ X เป็นตัวแปรแบบกำหนดได้ หรือแบบคงที่ ซึ่งมีค่าจำกัดในตัวอย่างซ้ำ ๆ และสำหรับตัวอย่างใด จะได้ว่า $\Sigma(X - \bar{X})^2/n$ เป็นจำนวนเลขนับได้ที่ต่างจากศูนย์

(3) เมื่อกำหนดค่าของ X ให้ ค่าของ Y จะมีการแจกแจงปกติที่มีค่าเฉลี่ย (เขียนแทนด้วย μ หรือ $E(Y|X)$) เท่ากับ $\beta_1 + \beta_2 X$ และความแปรปรวน σ^2 เรากล่าวได้ว่า ตัวแปร X และ ε เป็นมูลเหตุที่ทำให้เกิดค่าของตัวแปร Y

จะเห็นได้ว่า สำหรับทุก ๆ ค่าของ X ที่กำหนดมาให้ จะได้ค่าเฉลี่ยของ Y จาก

$$E(Y|X) = \beta_1 + \beta_2 X$$

สมการนี้ได้ชื่อว่า เส้นถดถอยประชากร (Population Regression Line) ซึ่งเขียนใหม่ได้เป็นสมการเส้นถดถอย

$$Y = \beta_1 + \beta_2 X$$

β_1 : เป็นจุดตัดแกนของเส้น ซึ่งจะวัดค่าเฉลี่ยของ Y เมื่อ X = 0

β_2 : เป็นความชันของเส้นถดถอย ซึ่งแสดงถึง อัตราการเปลี่ยนแปลงของค่า Y ต่อหนึ่งหน่วยค่า X ที่เปลี่ยนไป เราเรียก β_2 ว่า สัมประสิทธิ์การถดถอย (Regression Coefficient)

ทั่วไปเราใช้ตัวอย่างสุ่มประมาณค่าของ β_1 และ β_2 ดังนั้นเส้นตรงที่ได้เราเรียกเส้นถดถอยตัวอย่าง ซึ่งทำหน้าที่เป็นค่าประมาณของเส้นถดถอยประชากร เมื่อ $\hat{\beta}_1, \hat{\beta}_2$ เป็นตัวประมาณค่าของ β_1, β_2 แล้วเส้นถดถอยตัวอย่าง จะเป็น

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

หรือ $$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X + e$$

เมื่อ e เป็นความคลาดเคลื่อน ซึ่งเป็นผลต่างของ Y กับ $\hat{Y}$ โดยทั่วไป e ต่างจาก ε และ e เป็นค่าประมาณของ ε

การประมาณค่าของพารามีเตอร์ถดถอย

การประมาณทำได้หลายวิธี ในที่นี้จะกล่าวเฉพาะการประมาณค่าโดยวิธีกำลังสองต่ำสุด ซึ่งจะได้ค่าประมาณ $\hat{\beta}_1$ และ $\hat{\beta}_2$ ดังนี้

$$\hat{\beta}_2 = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$$

หรือ $$\hat{\beta}_2 = \frac{\Sigma xy}{\Sigma x^2}$$ เมื่อ $x = X - \bar{X}$ และ $y = Y - \bar{Y}$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

การประมาณความแปรปรวนของ $\hat{\beta}_1$ และ $\hat{\beta}_2$

ตัวประมาณที่ไม่เอียงเฉของ σ^2 คือ S^2

$$S^2 = \frac{\Sigma Y^2 - \hat{\beta}_1 \Sigma Y - \hat{\beta}_2 \Sigma XY}{n - 2}$$

หรือ $$S^2 = \frac{n-1}{n-2}(S_Y^2 - \hat{\beta}_2^2 S_X^2)$$

เรียกรากที่ 2 ของ S^2 ว่า ความคลาดเคลื่อนมาตรฐานของค่าประมาณ (Standard error of estimate)

ตัวประมาณความแปรปรวนของ $\hat{\beta}_1$ และ $\hat{\beta}_2$ ตามลำดับ คือ $S^2_{\hat{\beta}_1}$ และ $S^2_{\hat{\beta}_2}$

$$S^2_{\hat{\beta}_1} = S^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X-\bar{X})^2}\right] = S^2 \cdot \frac{\Sigma X^2}{n\Sigma X^2 - (\Sigma X)^2}$$

$$S^2_{\hat{\beta}_2} = \frac{S^2}{\Sigma(X-\bar{X})^2} = \frac{nS^2}{n\Sigma X^2 - (\Sigma X)^2}$$

เรียกรากที่ 2 ของ $S^2_{\hat{\beta}_1}$ และ $S^2_{\hat{\beta}_2}$ ว่า ความคลาดเคลื่อนมาตรฐานของ $\hat{\beta}_1$ และ $\hat{\beta}_2$ ตามลำดับ และจะใช้เป็นมาตราวัดความเที่ยงตรงของ $\hat{\beta}_1$ และ $\hat{\beta}_2$

ช่วงเชื่อมั่นของ β_1, β_2 และ σ^2

สิ่งที่แสดงถึงความเที่ยงตรง (Precision) ของ $\hat{\beta}_1, \hat{\beta}_2$ และ S^2 ก็คือการสร้างช่วงเชื่อมั่นของ β_1, β_2 และ σ^2 ซึ่งสามารถสร้างได้ดังนี้

ช่วงเชื่อมั่น $100(1-\alpha)\%$ ของ σ^2 คือ

$$\frac{(n-2)S^2}{\chi^2_{\alpha/2,(n-2)}} < \sigma^2 < \frac{(n-2)S^2}{\chi^2_{1-\alpha/2,(n-2)}}$$

ช่วงเชื่อมั่น $100(1-\alpha)\%$ ของ β_1 และ β_2 ก็คือ

$$\hat{\beta}_1 - t_{\alpha/2,(n-2)} \cdot S_{\hat{\beta}_1} < \beta_1 < \hat{\beta}_1 + t_{\alpha/2,(n-2)} \cdot S_{\hat{\beta}_1}$$

$$\hat{\beta}_2 - t_{\alpha/2,(n-2)} \cdot S_{\hat{\beta}_2} < \beta_2 < \hat{\beta}_2 + t_{\alpha/2,(n-2)} \cdot S_{\hat{\beta}_2}$$

ช่วงเชื่อมั่นของค่าเฉลี่ย Y เมื่อกำหนดค่าของ X ให้

เมื่อกำหนดค่าของ X เป็น X_0 เราจะได้ค่าประมาณของค่าเฉลี่ย Y ดังนี้

$$E(Y|X_0) = \hat{\beta}_1 + \hat{\beta}_2 X_0 = \hat{Y}$$

ความคลาดเคลื่อนมาตรฐานของ $E(Y|X_0)$ หรือ $\hat{Y}$ คือ $S_{\hat{Y}}$ เป็นรากที่ 2 ของ $S^2_{\hat{Y}}$ ซึ่งกำหนดไว้ดังนี้

$$S^2_{\hat{Y}} = S^2\left(\frac{1}{n} + \frac{(X_0-\bar{X})^2}{\Sigma(X-\bar{X})^2}\right)$$

ช่วงเชื่อมั่น $100(1-\alpha)\%$ ของ $E(Y|X)$ หรือ Y คือ

$$\hat{Y} - t_{\alpha/2,(n-2)} \cdot S_{\hat{Y}} < E(Y|X) < \hat{Y} + t_{\alpha/2,(n-2)} \cdot S_{\hat{Y}}$$

ตัวอย่าง กำหนดจำนวนผลผลิตที่ได้ X พันหน่วย และราคา Y ล้านบาท ของสินค้าชนิดหนึ่ง ดังตารางต่อไปนี้

ครั้งที่	1	2	3	4	5	6	7	8	9	10	11	12
X	65	50	55	65	70	55	70	65	55	70	50	55
Y	85	74	76	90	87	85	98	94	81	91	76	74

จงหา ก) เส้นประมาณการถดถอยของราคาสินค้าบนจำนวนผลผลิตที่ได้

ข) ค่าเฉลี่ยของราคาสินค้า เมื่อจำนวนผลผลิตที่ได้ เท่ากับ 60,000 หน่วย และจงหาช่วงเชื่อมั่น 95% ของค่าเฉลี่ยนี้ด้วย

ค) ช่วงเชื่อมั่น 98% ของ β_1 และ β_2

ง) ช่วงเชื่อมั่น 90% ของ σ^2

วิธีทำ

X	Y	XY	X^2	Y^2
65	85	5525	4225	7225
50	74	3700	2500	5476
55	76	4180	3025	5776
65	90	5850	4225	8100
70	87	6090	4900	7569
55	85	4675	3025	7225
70	98	6860	4900	9604
65	94	6110	4225	8836
55	81	4455	3025	6561
70	91	6370	4900	8281
50	76	3800	2500	5776
55	74	4070	3025	5476
725	1011	61685	44475	85905

$$\hat{\beta}_2 = \frac{12 \times 61685 - 725 \times 1011}{12 \times 44475 - (725)^2} = .897$$

$$\bar{X} = 725/12 = 60.417 \quad , \quad \bar{Y} = 1011/12 = 84.25$$

$$\hat{\beta}_1 = 84.25 - .897 \times 60.417 = 30.056$$

เส้นประมาณการถดถอย ก็คือ

$$\bar{Y} = 30.056 + .897X \qquad \text{.....(ก)}$$

$$S^2 = \frac{85905 - 30.056 \times 1011 - .897 \times 61685}{12 - 2} = 18.6939$$

$$E(Y|60) = 30.056 + .897 \times 60 = 83.876$$

เมื่อจำนวนผลผลิตเป็น 60,000 หน่วย ค่าเฉลี่ยของราคาสินค้า เท่ากับ 83.876 ล้านบาท

$$\Sigma(X - \bar{X})^2 = 44475 - (725)^2/12 = 672.917$$

$$t_{.025,10}\, S_{\hat{Y}} = 2.228 \times \sqrt{18.6939\left(\frac{1}{12} + \frac{(60 - 60.417)^2}{672.917}\right)} = 2.7851$$

เมื่อจำนวนผลผลิตเป็น 60,000 หน่วยช่วงเชื่อมั่น 95% ของค่าเฉลี่ยของราคาสินค้า ก็คือ

$$83.876 - 2.7851 < Y < 83.867 + 2.7851$$

หรือ $$81.09 < Y < 86.661 \qquad \text{.....(ข)}$$

$$t_{.01,10} \cdot S_{\hat{\beta}_1} = 2.764 \times \sqrt{18.6939\left(\frac{1}{12} + \frac{(60.417)^2}{672.917}\right)} = 28.0463$$

ดังนั้น ช่วงเชื่อมั่น 98% ของ β_1 ก็คือ

$$30.056 - 28.0463 > \beta_1 < 30.056 + 28.0463$$

หรือ $$2.0097 < \beta_1 < 58.1023$$

$$t_{.01,10}\, S_{\hat{\beta}_2} = 2.764 \times \sqrt{18.6939/672.917} = .4607$$

ดังนั้น ช่วงเชื่อมั่น 98% ของ β_2 ก็คือ

$$.897 - .4607 < \beta_2 < .897 + .4607$$

หรือ $$.4363 < \beta_2 < 1.3577 \qquad \text{.....(ค)}$$

$$\chi^2_{.05,10} = 18.307 \quad , \quad \chi^2_{.95,10} = 3.940$$

ช่วงเชื่อมั่น 90% ของ σ^2 กำหนดได้ดังนี้

$$\frac{(12-2)\times 18.6939}{18.307} < \sigma^2 < \frac{(12-2)\times 18.6939}{3.940}$$

หรือ $$10.21 < \sigma^2 < 47.45 \qquad \text{.....(ง)}$$

การทดสอบสมมติฐานเกี่ยวกับการถดถอยเชิงเส้น

สมมติฐานที่เราต้องการทดสอบในเรื่องที่เกี่ยวกับการถดถอย ที่สำคัญที่สุดก็คือ สมมติฐานที่เกี่ยวกับ β_2 นั่นคือ การทดสอบว่า ไม่มีความสัมพันธ์ระหว่างตัวแปรที่อธิบายได้ X กับตัวแปรตาม Y

ในตัวแบบ

$$Y = \beta_1 + \beta_2 X + \varepsilon$$

เราอาจจะทดสอบสมมติฐานเกี่ยวกับ β_1 และ β_2 ได้ดังนี้

1. ทดสอบสมมติฐาน

$$H_0 : \beta_1 = \beta_{10}$$

ตัวสถิติทดสอบที่ใช้คือ

$$t = (\hat{\beta}_1 - \beta_{10})/S_{\hat{\beta}_1}$$

เกณฑ์ที่ใช้ในการตัดสินว่าจะยอมรับหรือจะปฏิเสธ H_0 ขึ้นอยู่กับสมมติฐานรอง H_a และระดับนัยสำคัญ α โดยเปรียบเทียบกับค่าทีจากตารางที่มีองศาอิสระ $n - 2$

ในกรณีที่ $\beta_{10} = 0$ แสดงว่าการทดสอบนี้ต้องการดูว่า เส้นถดถอยประชากรจะผ่านจุดกำเนิดหรือไม่

2. ทดสอบสมมติฐาน

$$H_0 : \beta_2 = \beta_{20}$$

ตัวสถิติทดสอบที่ใช้คือ

$$t = (\hat{\beta}_2 - \beta_{20})/S_{\hat{\beta}_2}$$

เทียบค่าที่คำนวณได้กับค่าที ที่มีองศาอิสระ $n - 2$

ในกรณีที่ $\beta_{20} = 0$ หรือทดสอบ $H_0 : \beta_2 = 0$ นั่นคือทดสอบว่า ไม่มีความสัมพันธ์ระหว่าง X และ Y หรือต้องการทดสอบว่า Y ขึ้นอยู่กับ X หรือไม่ หรืออีกนัยหนึ่งก็คือต้องการทดสอบว่าจะมีการถดถอยของ Y บน X หรือไม่

ถ้าเราไม่มีความรู้เกี่ยวกับค่าของพารามีเตอร์ถดถอยมาก่อน แล้วสมมติฐานรอง H_a จะเป็น

$$H_a : \beta_2 \neq 0$$

เราสามารถใช้การวิเคราะห์ความแปรปรวนกับการถดถอยได้ นั่นคือ ในการทดสอบ

$$H_0 : \beta_2 = 0$$

เราทำตารางวิเคราะห์ความแปรปรวนได้ดังนี้

SOV	df	SS	MS	F-ratio
เนื่องจากการถดถอย (ตัวแปรอธิบายได้)	1	$SSR = \hat{\beta}_2 \Sigma xy = (\Sigma xy)^2/\Sigma x^2$	$MSR = SSR$	$\frac{MSR}{S^2}$
ความคลาดเคลื่อน	$n - 2$	$SSE = SST - SSR$	$S^2 = \frac{SSE}{n-2}$	
ยอดรวม	$n - 1$	$SST = \Sigma y^2$		

เราจะไม่ยอมรับ H_0 หรือสรุปว่า Y และ X มีความสัมพันธ์กัน ถ้า

$$F = \frac{MSR}{S^2} > f_{\alpha(1,n-2)}$$

ตัวอย่าง ในการศึกษาเพื่อดูความสัมพันธ์ ระหว่างเวลาที่ใช้ในการบรรจุ กับปริมาตรของที่บรรจุ พบว่า ถ้าของที่บรรจุมีปริมาตร X ลูกบาศก์ฟุต จะใช้เวลาในการบรรจุ Y นาที และคำนวณได้ค่าต่าง ๆ ดังนี้

n	ΣY	ΣX	ΣXY	ΣY^2	ΣX^2
9	133.8	8.37	135.492	2165.26	8.5079

จะใช้ข้อเท็จจริงจากตัวอย่างที่รวบรวมไว้ สรุปผลที่ระดับนัยสำคัญ .05 ว่า

(ก) เส้นถดถอยประชากรผ่านจุดกำเนิด

(ข) เวลาที่ใช้ในการบรรจุไม่ขึ้นอยู่กับปริมาตรของที่บรรจุ

ได้หรือไม่

วิธีทำ 1. (ก) $H_0 : \beta_1 = 0$

$H_a : \beta_1 \neq 0$

(ข) $H_0 : \beta_2 = 0$

$H_a : \beta_2 \neq 0$

2. $\alpha = 0.05, n - 2 = 7$

3. เราจะไม่ยอมรับ H_0 ถ้า

$t_{\text{คำนวณ}} > t_{.025,7} = 2.365$

4. $\hat{\beta}_2 = \dfrac{9 \times 135.492 - 133.8 \times 8.37}{9 \times 8.5079 - (8.37)^2} = 15.2777$

$\overline{Y} = 133.8/9 = 14.87 , \overline{X} = 8.37/9 = 0.93$

$\hat{\beta}_1 = 14.87 - 15.2777 \times 0.93 = .6617$

$S^2 = \dfrac{2165.26 - .6617 \times 133.8 - 15.2777 \times 135.492}{9 - 2} = .9598$

$n\Sigma X^2 - (\Sigma X)^2 = 6.5142$

$S_{\hat{\beta}_1} = \sqrt{.9598 \times 8.5079/6.5142} = 1.1196$

$S_{\hat{\beta}_2} = \sqrt{.9598 \times 9/6.5142} = 1.1515$

(ก) $t = .6617/1.1196 = .591$

(ข) $t = 15.2777/1.1515 = 13.2676$

5. (ก) ยอมรับ H_0 ถือว่า เส้นถดถอยประชากรผ่านจุดกำเนิด

(ข) ไม่ยอมรับ H_0 ถือว่าเวลาที่ใช้ในการบรรจุขึ้นอยู่กับปริมาตรของที่บรรจุ

หมายเหตุ ในกรณีของข้อ ข เราใช้ตารางวิเคราะห์ได้ดังนี้

$SSR = \hat{\beta}_2\Sigma xy = 15.2777 \times (135.492 - 133.8 \times 8.37/9) = 168.9408$

$SST = \Sigma y^2 = 2165.26 - (133.8)^2/ 9 = 176.1$

$SSE = 176.1 - 168.9408 = 7.1592$

ตารางวิเคราะห์จะเป็นดังนี้

SOV	df	SS	MS	F-ratio
เนื่องจากการถดถอย	1	168.9408	168.9408	165.1845
ความคลาดเคลื่อน	7	7.1592	1.02274	
ยอดรวม	9	176.1		

เปรียบเทียบค่า F ที่ได้กับ $f_{.05(1,7)} = 5.59$ นั่นคือ ไม่ยอมรับ H_0 เช่นเดียวกับการทดสอบโดยใช้สถิติทดสอบที

การทำนาย (Prediction)

ประโยชน์ที่สำคัญอีกประการหนึ่งของการถดถอย ก็คือ ใช้ในการทำนายหรือการพยากรณ์ บ่อยครั้งเราสนใจทำนายหรือพยากรณ์ค่า Y แต่ละค่า ในเมื่อกำหนดค่าของ X ให้ นั่นคือ ถ้าค่าของ $X = X_0$ เราจะพยากรณ์ค่า Y ของประชากรได้ ดังนี้

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_0$$

ถ้าเราทราบพารามีเตอร์ของประชากร จะได้ค่าจริงของ Y ที่จุด $X = X_0$ ดังนี้

$$Y = \beta_1 + \beta_2 X_0 + \varepsilon$$

ค่าจริงของ Y กับค่าทำนายของ Y (หรือ $\hat{Y}$) นั้น ส่วนมากจะต่างกัน นั่นคือ เกิดความคลาดเคลื่อน เราสามารถเขียนความคลาดเคลื่อนของการพยากรณ์ $(Y - \hat{Y})$ ได้ดังนี้

$$Y - \hat{Y} = (\beta_1 + \beta_2 X_0 + \varepsilon) - (\hat{\beta}_1 + \hat{\beta}_2 X_0)$$

ดังนั้น เราจะได้

$$S_f^2 = S^2 + S_{\hat{Y}}^2$$

ในเมื่อ S_f^2 = ตัวประมาณค่าของความแปรปรวนของความคลาดเคลื่อนการพยากรณ์

S^2 = ตัวประมาณค่าของความแปรปรวนของตัวคลาดเคลื่อนเชิงสุ่ม

$S_{\hat{Y}}^2$ = ตัวประมาณค่าของความแปรปรวนของค่าเฉลี่ย หรือความแปรปรวนของความคลาดเคลื่อนจากตัวอย่าง

ดังนั้น

$$S_f = S\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\Sigma(X - \bar{X})^2}}$$

การพยากรณ์ด้วยช่วงเชื่อมั่น อาจกระทำได้โดยอาศัยค่าที นั่นคือ

ช่วงเชื่อมั่น $100(1 - \alpha)\%$ ของแต่ละค่า Y กำหนดไว้ดังนี้

$$(\hat{\beta}_1 + \hat{\beta}_2 X_0) - t_{\alpha/2,(n-2)} S_f < Y < (\hat{\beta}_1 + \hat{\beta}_2 X_0) + t_{\alpha/2,(n-2)} S_f$$

หรือ

$$\hat{Y} - t_{\alpha/2,(n-2)} S_f < Y < \hat{Y} + t_{\alpha/2,(n-2)} S_f$$

ข้อสังเกตุ เมื่อกำหนดค่า $X = X_0$ มาให้ ค่าเฉลี่ยของ Y หรือแต่ละค่า Y ที่เรียกว่าค่าพยากรณ์

จะเป็นค่าเดียวกัน คือ

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

แต่ความคลาดเคลื่อนมาตรฐานที่ใช้ แตกต่างกัน

ตัวอย่าง ในการทดลองเกี่ยวกับผลิตผลของข้าวสาลีต่อจำนวนปุ๋ยที่ใช้ พบว่า ถ้าใช้ปุ๋ย X ปอนด์/เอเคอร์ จะให้ผลผลิตข้าวสาลี Y บุชเชลต่อเอเคอร์ ผลจากการทดลองใน 7 แปลงทดลอง ประมาณการถดถอยได้ดังนี้

$$\hat{Y} = 32.8 + .068X$$

จงทำนายผลผลิตของข้าวสาลีที่จะได้ ถ้าใช้ปุ๋ย 500 ปอนด์ต่อเอเคอร์

ถ้า $\overline{X} = 400$ ปอนด์ต่อเอเคอร์, $\Sigma(X - \overline{X})^2 = 280{,}000$ และ $S^2 = 12.144$ จงทำนายช่วงเชื่อมั่น 95% ของจำนวนผลผลิตข้าวสาลีที่ได้ เมื่อใช้ปุ๋ย 500 ปอนด์ต่อเอเคอร์

วิธีทำ

$$\hat{Y} = 32.8 + .068 \times 500 = 66.8$$

เมื่อใช้ปุ๋ย 500 ปอนด์ต่อเอเคอร์ คาดว่าจะได้ข้าวสาลี 66.8 บุชเชลต่อเอเคอร์

$$S_f^2 = 12.144\left(1 + \frac{1}{7} + \frac{(500 - 400)^2}{280{,}000}\right) = 14.312569$$

$$t_{.025,5}\, S_f = 2.571 \times 3.783 = 9.726$$

ดังนั้น ช่วงเชื่อมั่น 95% ของ Y เมื่อ X = 500 ก็คือ

$$66.8 - 9.726 < Y < 66.8 + 9.726$$

หรือ $$57.074 < Y < 76.526$$

สัมประสิทธิ์การตัดสินใจ (Coefficient Determination)

การแยกความผันแปรของ Y จาก

$$SST = SSR + SSE$$

ในเมื่อ $SST = \Sigma y^2 = \Sigma Y^2 - (\Sigma Y)^2/n$

และ $SSR = \hat{\beta}_2 \Sigma xy = \hat{\beta}_2 \mid \Sigma XY - (\Sigma X)(\Sigma Y)/n \mid$

นั้น จะได้มาตรการอันหนึ่งเกี่ยวกับ "การปรับที่ดีที่สุด" (Goodness of fit) เรียกว่า "สัมประสิทธิ์การตัดสินใจ" ใช้แทนด้วย R^2 และนิยามไว้ดังนี้

$$R^2 = \frac{SSR}{SST} = \frac{\hat{\beta}_2 \Sigma xy}{\Sigma y^2} = \hat{\beta}_2 \frac{S_X^2}{S_Y^2}$$

R^2 เป็นมาตรการที่ใช้บรรยายว่า "เส้นถดถอยตัวอย่างปรับเข้าได้ กับข้อมูลที่สังเกตุได้ดีเพียงไร" R^2 จะเป็นตัวประมาณค่าที่ไม่เอียงเฉของสัมประสิทธิ์การตัดสินใจของประชากร ρ^2

รากที่ 2 ของ R^2 คือ r เรียกว่าสัมประสิทธิ์สหสัมพันธ์ ซึ่งจะเป็นค่าประมาณที่ไม่เอียงเฉของสัมประสิทธิ์สหสัมพันธ์ ρ เราใช้ r เป็นเครื่องวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว นั่นคือ วัดขนาดของการเปลี่ยนแปลงของตัวแปร 2 ตัว

ดังนั้น

$$r = \sqrt{\frac{\hat{\beta}_2 \Sigma xy}{\Sigma y^2}} = \hat{\beta}_2 \frac{S_X}{S_Y}$$

หรือ

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{\{n\Sigma X^2 - (\Sigma X)^2\}\{n\Sigma Y^2 - (\Sigma Y)^2\}}}$$

ค่าของ r จะอยู่ระหว่าง −1 กับ 1 ค่า r ที่เป็นบวก แสดงว่า X กับ Y มีค่าเพิ่มขึ้นด้วยกัน หรือลดลงด้วยกัน ค่า r ที่เป็นลบ แสดงว่า X เพิ่มขึ้น Y จะลดลง

การทดสอบสมมติฐานเกี่ยวกับ ส.ป.ส.สหสัมพันธ์

การทดสอบ

$$H_0 : \rho = \rho_0$$

เราใช้ข้อสมมติที่ว่าค่า X และ Y มาจากค่าสังเกตุที่มีการแจกแจงแบบปกติร่วม 2 ตัว (Bivariate normal distribution) ดังนั้นค่าสถิติที่ใช้คือ

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

ในกรณีที่ $\rho_0 = 0$ นั่นคือ ทดสอบ

$$H_0 : \rho = 0$$

นอกจากจะใช้ตัวสถิติทดสอบ Z ดังที่กล่าวแล้ว เราอาจจะใช้ตัวสถิติทดสอบ

$$t = r/S_r = r\sqrt{(n-2)/(1-r^2)}$$

ได้ และเปรียบเทียบค่า t ที่คำนวณได้นี้กับค่า t_{n-2} จากตาราง

ตัวอย่าง จากตัวอย่างในเรื่องการศึกษาความสัมพันธ์ ของเวลาที่ใช้ในการบรรจุ Y กับปริมาตรของที่บรรจุ X เราทราบว่า จากการทดลอง 9 ครั้ง ได้ $\Sigma X = 8.37$, $\Sigma Y = 133.8$, $\Sigma XY = 135.492$, $\Sigma X^2 = 8.5079$, $\Sigma Y^2 = 2165.26$

ก) จงหาสัมประสิทธิ์ระหว่างเวลาที่ใช้ในการบรรจุ กับปริมาตรของที่บรรจุ

ข) จงใช้ระดับนัยสำคัญ .05 ทดสอบความมีนัยสำคัญของค่าที่ได้จากข้อ ก

วิธีทำ

ก) $$r = \frac{9 \times 135.492 - 8.37 \times 133.8}{\sqrt{|9 \times 8.5079 - (8.37)^2|\,|9 \times 2165.26 - (133.8)^2|}} = 0.9794$$

ข) 1. $H_0 : \rho = 0$

$H_a : \rho > 0$

2. $\alpha = .05$, $n - 2 = 7$

3. ปฏิเสธ H_0 ถ้า

$$t_{\text{คำนวณ}} > t_{.05,7} = 1.895$$

4. $$t = r\sqrt{\frac{n-2}{1-r^2}} = .9794\sqrt{\frac{7}{1-(.9794)^2}} = 12.832$$

5. ปฏิเสธ H_0 สรุปว่ามีนัยสำคัญ

การนำเสนอผลของการวิเคราะห์การถดถอย

รูปแบบการนำเสนอ อาจเป็น

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \qquad R^2 = \ldots\ldots\ldots\ldots$$
$$(S_{\hat{\beta}_1})(S_{\hat{\beta}_2})$$

หรือ

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i \qquad R^2 = \ldots\ldots\ldots\ldots$$
$$(S_{\hat{\beta}_1})(S_{\hat{\beta}_2})$$

นักสถิติบางคนเสนอเป็นแบบค่า t นั่นคือ จาก $S_{\hat{\beta}_1}$ เป็น $\hat{\beta}_1/S_{\hat{\beta}_1}$ และจาก $S_{\hat{\beta}_2}$ เป็น $\hat{\beta}_2/S_{\hat{\beta}_2}$

ตัวแบบถดถอยเชิงซ้อน (Multiple Regression Model)

ตัวแบบประกอบด้วยตัวแปรอิสระมากกว่า 1 ตัว การเปลี่ยนแปลงในตัวแปรหนึ่งสามารถอธิบายได้ด้วยการเปลี่ยนแปลงของตัวแปรอื่น ๆ หลายตัว

รูปแบบการถดถอยเชิงซ้อน จะเป็น

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + \varepsilon_i$$

หรือ

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i$$

ข้อสมมติเบื้องต้นจะเป็นแบบเดียวกับการถดถอยเชิงเดี่ยว แต่เพิ่มเข้ามา 2 ข้อ คือ ให้มีจำนวนองศาแห่งความเป็นอิสระเพียงพอในการประมาณค่า $n > k$ และไม่มีตัวแปรอิสระใดมีสหสัมพันธ์สมบูรณ์กับตัวแปรอิสระอื่นใด

ค่าเฉลี่ยและความแปรปรวน จะเป็น

$$E(Y_i) = \beta_1 + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}$$

$$V(Y_i) = E|Y_i - E(Y_i)|^2 = \sigma^2$$

คำถาม-คำตอบ

1. สมการถดถอยสร้างไว้เพื่อจุดประสงค์ใด

   ก) ใช้ทำนายตัวแปรตาม ข) ใช้ทำนายเหตุและผลของตัวแปร

   ค) ใช้ทำนายความสัมพันธ์ ง) ใช้ทำนายความผันแปร

2. ตัวแบบถดถอยเชิงเส้น ที่อยู่ในรูปสมการ

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$$

   ε_i ในตัวแบบนี้ หมายถึงอะไร

   ก) พารามีเตอร์ของการถดถอยที่ไม่ทราบค่า

   ข) ตัวแปรอิสระหรือตัวแปรอธิบายได้

   ค) ตัวแปรที่สามารถวัดหรือสังเกตได้

   ง) ตัวคลาดเคลื่อนซึ่งไม่สามารถวัดหรือสังเกตได้

3. ข้อใดเป็นเกณฑ์สมมติ (assumption) ของตัวแบบถดถอย

   ก) ε_i มีการแจกแจงปกติ $(0, \sigma^2)$ ข) X_i มีการแจกแจงปกติ $(0, \sigma^2)$

   ค) Y_i มีการแจกแจงปกติ $(0, \sigma^2)$ ง) X_i และ Y_i มีความสัมพันธ์กัน

4. ที่กล่าวว่า X_i เป็นตัวแปรแบบกำหนดได้หรือแบบคงที่ มีความหมายว่าอย่างไร

   ก) ค่าของ X_i ในตัวอย่างหนึ่งต้องเท่ากันหมด

   ข) ค่าของ X_i จะเป็นเช่นเดียวกันในทุกตัวอย่าง

   ค) ค่าของ X_i สามารถควบคุมหรือทำนายได้

   ง) ค่าของ X_i จะไม่เพิ่มขึ้นหรือลดลงโดยปราศจากขีดจำกัด

5. ความคลาดเคลื่อนของตัวอย่าง จะทำให้เส้นถดถอยตัวอย่างไม่เป็นเส้นเดียวเส้นถดถอยประชากร เราจะลดความคลาดเคลื่อนนี้ได้อย่างไร

   ก) เก็บตัวอย่างโดยวิธีสุ่ม ข) เพิ่มขนาดตัวอย่างให้มากขึ้น

   ค) ลดระดับนัยสำคัญ ง) เพิ่มขนาดตัวอย่างและลดระดับนัยสำคัญ

6—11 จากรายงานผลการซ่อมแซมเครื่องจักรชนิดหนึ่ง 10 เครื่อง พบว่า มีอายุใช้งาน (X) โดยเฉลี่ย 3 ปี และค่าบำรุงรักษา (Y) โดยเฉลี่ย 67 พันบาท ส.ป.ส.การถดถอยของค่าบำรุงรักษาบนอายุใช้งานเท่ากับ 20.5

6. จงหาสมการถดถอยของค่าบำรุงรักษาบนอายุใช้งาน

   ก) $\hat{Y} = 67 - 20.5X$ ข) $\hat{Y} = 67 + 20.5X$

   ค) $\hat{Y} = 5.5 - 20.5X$ ง) $\hat{Y} = 5.5 + 20.5X$

7. ถ้าเครื่องจักรมีอายุ 5 ปี จะใช้ค่าบำรุงรักษาประมาณเท่าใด

ก) 97 พันบาท ข) 108 พันบาท

ค) 130 พันบาท ง) 169.5 พันบาท

8.,9.,10.,11. การทดสอบเพื่อดูว่า ค่าบำรุงรักษา จะมีความสัมพันธ์เชิงเส้นตรงกับอายุการใช้งานหรือไม่ ที่ระดับนัยสำคัญ 0.05

สมมติฐานในการทดสอบ จะเป็นแบบใด

ก) $H_0 : \beta_1 = 0, H_a : \beta_1 \neq 0$ ข) $H_0 : \beta_1 = 0, H_a : \beta_1 > 0$

ค) $H_0 : \beta_2 = 0, H_a : \beta_2 \neq 0$ ง) $H_0 : \rho = 0, H_a : \rho > 0$

ตัวสถิติทดสอบ จะเป็นเท่าใด

ก) $5.5/S_{\hat{\beta}_1}$ ข) $20.5/S_{\hat{\beta}_2}$

ค) $|67 - 20.5|/S_{\hat{\beta}_2}$ ง) $(8)(20.5)/\sigma_Y^2$

ตัวสถิติทดสอบจะมีการแจกแจงแบบใด

ก) แบบที ข) แบบเอฟ

ค) แบบปกติ ง) แบบคายกำลังสอง

จะสรุปว่า ค่าบำรุงรักษา มีความสัมพันธ์เชิงเส้นกับอายุใช้งาน ถ้าค่าของตัวสถิติทดสอบมากกว่าค่าใด

ก) 1.860 ข) 1.960

ค) 2.306 ง) 15.507

12–14 ศึกษาถึงผลของอุณหภูมิที่มีต่อผลผลิต จากการทดลอง 12 ครั้ง ประมาณได้เส้นถดถอยของจำนวนผลผลิต (Y) บนอุณหภูมิที่ใช้ (X) เป็น

$$\hat{Y} = 9.27 + 1.44X$$

และคำนวณได้ว่า $S^2 = 2.36$, $S^2_{\hat{\beta}_1} = 3.12$, $S^2_{\hat{\beta}_2} = 2.97$, $S^2_{\hat{Y}} = 3.01$ และ $S^2_f = 5.37$

12. ช่วงความเชื่อมั่น 95% ของ ส.ป.ส.การถดถอยของจำนวนผลผลิตบนอุณหภูมิ จะเป็นเท่าใด

ก) $9.27 \pm 2.228\sqrt{2.36}$ ข) $9.27 \pm 2.228\sqrt{3.12}$

ค) $1.44 \pm 2.228\sqrt{2.97}$ ง) $1.44 \pm 2.228\sqrt{3.01}$

13. ต้องการทำนายจำนวนผลผลิตที่จะได้เมื่อคาดว่าอุณหภูมิจะเป็น 3 ช่วงความเชื่อมั่น 95% ของค่าที่ทำนายนี้จะเป็นเท่าใด

ก) $9.27 \pm (1.44)(3)$ ข) $13.59 \pm 1.96\sqrt{5.37}$

ค) $13.59 \pm 2.228\sqrt{3.01}$ ง) $13.59 \pm 2.228\sqrt{5.37}$

14. ช่วงความเชื่อมั่น 95% ของ σ^2 จะเป็นเท่าใด

ก) $\frac{23.6}{20.483}$, $\frac{23.6}{3.247}$

ข) $\frac{31.2}{20.483}$, $\frac{31.2}{3.247}$

ค) $\frac{31.96}{21.920}$, $\frac{31.96}{3.816}$

ง) $\frac{59.07}{21.920}$, $\frac{59.07}{3.816}$

คำตอบ : ก ง ก ค ข ง ข ค ข ก ค ค ง ก

แบบฝึกหัดที่ 6

1. กำหนด

$$\Sigma X = 766 \quad , \quad \Sigma Y = 1700 \quad , \quad \Sigma XY = 109380$$

$$\Sigma X^2 = 49068 \quad , \quad \Sigma Y^2 = 246100 \quad , \quad n = 12$$

ก) จงประมาณเส้นถดถอยของ Y บน X

ข) จงทำนายค่า Y เมื่อ X = 80 และจงหาช่วงเชื่อมั่น 95% ของค่าทำนายนี้

ค) จงทดสอบที่ระดับนัยสำคัญ .05 ว่า Y และ X ไม่มีความสัมพันธ์กัน

2. กำหนดข้อมูลดังต่อไปนี้

X	0.5	1.5	3.2	4.2	5.1	6.5
Y	1.3	3.4	6.7	8.0	10.0	13.2

ก) จงหาเส้นประมาณการถดถอย

ข) จงทดสอบที่ระดับนัยสำคัญ .01 ว่า เส้นถดถอยจะผ่านจุดกำเนิด

3. ในการศึกษาความสัมพันธ์ระหว่าง ค่าจ้างแรงงานกับระดับการผลิต พบว่า ถ้าค่าแรงงานเป็น X แสนบาท จะมีระดับการผลิต Y ล้านหน่วย

ก) เราจะตั้งสมมติฐานเพื่อทดสอบความสัมพันธ์อย่างไร

ข) จากการสุ่มตัวอย่างการผลิต 11 ครั้ง สรุปในตารางวิเคราะห์ความแปรปรวน ดังนี้

SOV	SS	df	MS	F-ratio
เนื่องจากการถดถอย	3.6			
ความคลาดเคลื่อน				
ผลรวม	7.202			

จงเขียนตารางให้สมบูรณ์ และสรุปผลที่ได้ ที่ระดับนัยสำคัญ .05

4. เป็นที่เชื่อว่า จำนวนชิ้นส่วนที่ชำรุดของผลผลิตมีความสัมพันธ์กับอัตราเร็วที่ผู้ควบคุมเครื่องจักรตั้งไว้ จากการสุ่มตัวอย่างผลผลิตที่ได้ในช่วงเวลาที่กำหนดให้ ปรากฏผลดังนี้

อัตราเร็ว (รอบ/นาที)	10	12	15	13	14	17	13	18
จำนวนชิ้นส่วนที่ชำรุด	3	4	8	5	6	7	6	9

ก) จงหาเส้นประมาณการถดถอยของจำนวนชิ้นส่วนที่ชำรุดบนอัตราเร็วของเครื่องจักรที่ตั้งไว้

ข) จงทดสอบความมีนัยสำคัญของ ส.ป.ส.การถดถอย ที่ระดับนัยสำคัญ .05

ค) จงประมาณค่าเฉลี่ยของจำนวนชิ้นส่วนที่ชำรุด ถ้าอัตราเร็วของเครื่องจักรเป็น 15 รอบ/นาที

5. โรงงานอุตสาหกรรมผลิตอาหารสำเร็จรูป ต้องการศึกษาความสัมพันธ์ ระหว่างจำนวนขายกับค่าใช้จ่ายในการโฆษณา จากการสุ่มตัวอย่างมา 10 เขต ได้ผลดังตารางต่อไปนี้

เขต	ก	ข	ค	ง	จ	ฉ	ช	ด	ต	ท
จำนวนขาย (หมื่นบาท)	6	10	5	10	7	9	5	8	12	4
ค่าโฆษณา (หมื่นบาท)	1.0	3.0	1.5	4.5	2.5	4.0	2.0	3.5	5.0	0.5

ผู้จัดการโรงงานเชื่อว่า จำนวนขายมีความสัมพันธ์อย่างใกล้ชิดกับค่าใช้จ่ายในการโฆษณา และต้องการใช้การวิเคราะห์การถดถอยในการยืนยันความเชื่อของเขา

ก) เส้นประมาณการถดถอยจะเป็นอย่างไร

ข) ข้อเท็จจริงที่ได้จากตัวอย่าง จะสรุปว่า จำนวนขายมีความสัมพันธ์กับค่าใช้จ่ายในการโฆษณาด้วยความเชื่อมั่น 95% ได้หรือไม่

ค) จงประมาณค่าคาดหมายของจำนวนขาย เมื่อค่าใช้จ่ายในการโฆษณา เท่ากับ 50,000 บาท และจงประมาณค่านี้ในช่วงเชื่อมั่น 95%

ง) จงทำนายจำนวนขาย ถ้าค่าใช้จ่ายในการโฆษณา เท่ากับ 40,000 บาท และจงประมาณค่าทำนายนี้ในช่วงเชื่อมั่น 95%

6. บริษัทจัดส่งสินค้าออกได้สุ่มตัวอย่างจำนวนสินค้าที่ส่งออกและราคาของมันมา 9 ปี พบว่า ถ้าปริมาณส่งออกเท่ากับ X พันตัน แล้วราคาของสินค้าจะเท่ากับ Y ล้านบาท ตามตารางต่อไปนี้

ปีที่	1	2	3	4	5	6	7	8	9
ปริมาณส่งออก	1.5	1.8	2.4	3.0	3.5	3.9	4.4	4.8	5.0
ราคา	4.8	5.7	7.0	8.3	10.9	12.4	13.1	13.6	15.3

ก) เส้นประมาณการถดถอยจะเป็นอย่างไร

ข) ผู้จัดการส่งออกอ้างว่า ส.ป.ส.การถดถอยจะมีค่าไม่เกิน 2.45 ด้วยความเชื่อมั่น 99% ท่านเห็นด้วยหรือไม่

ค) ถ้าไม่เห็นด้วย ช่วงเชื่อมั่น 95% ของ ส.ป.ส.การถดถอยควรจะมีค่าอย่างไร

7. โรงงานอุตสาหกรรมขนาดย่อมพบว่าจำนวนขายของสินค้าชนิดหนึ่งมีความสัมพันธ์กับดัชนีเศรษฐกิจ จากการรวบรวมจำนวนขายที่ผ่านมาและดัชนีเศรษฐกิจในช่วงเดียวกัน ปรากฏผลดังตารางต่อไปนี้

ช่วงที่	1	2	3	4	5	6	7	8	9	10
ปริมาณขาย (ล้านบาท)	2.1	1.9	2.3	1.5	1.2	2.7	3.6	1.4	0.9	2.0
ดัชนีเศรษฐกิจ	104	101	106	99	95	109	120	98	90	103

ก) ขนาดของความสัมพันธ์จะเป็นเท่าใด

ข) จงหาเส้นประมาณการถดถอยของปริมาณขายบนดัชนีเศรษฐกิจ

ค) ถ้าค่าของดัชนีในช่วงต่อไป คาดหมายว่าจะเท่ากับ 112 จงทำนายปริมาณขายในช่วงนั้น

8. ห้างสรรพสินค้าแห่งหนึ่งได้รวบรวมข้อมูล เกี่ยวกับ ค่าจ้างพนักงานขายและปริมาณขายในแต่ละแผนกขาย ซึ่งมีอยู่ 8 แผนก ผลที่ได้จากเดือนที่แล้ว ปรากฏผลดังตารางต่อไปนี้

แผนกขาย	ก	ข	ค	ง	จ	ฉ	ช	ท
ค่าจ้างพนักงาน (แสนบาท)	.51	.30	.41	.62	.25	.46	.43	.38
ปริมาณขาย (แสนบาท)	4.7	3.6	4.4	5.4	2.9	4.6	3.9	3.4

ก) จงหาเส้นประมาณการถดถอยของค่าจ้างบนปริมาณขาย

ข) ถ้าปริมาณขายเท่ากับ 420,000 บาท จงทำนายค่าจ้างพนักงาน

ค) จงหาช่วงเชื่อมั่น 95% ของค่าทำนายค่าจ้างพนักงาน ถ้าปริมาณขายเท่ากับ 6.4 แสนบาท

ง) จงหาขนาดของความสัมพันธ์

9. ผู้จัดการโรงงานต้องการศึกษาว่าอัตราเร็วของเครื่องจักรที่ตั้งไว้ จะมีผลต่อคุณภาพการผลิตหรือไม่ จากการสุ่มตัวอย่าง 12 ครั้ง ได้ผลดังตารางต่อไปนี้

อัตราเร็วของเครื่องจักร (รอบต่อนาที)	15	16	9	14	17	13	16	10	13	9	12	12
เปอร์เซนต์ผลิตภัณฑ์คุณภาพต่ำกว่ามาตรฐาน	5	6	1	3	8	3	5	2	4	2	5	4

ก) จงหาสัมประสิทธิ์สหสัมพันธ์ และสัมประสิทธิ์การตัดสินใจ

ข) ทดสอบความมีนัยสำคัญของ ส.ป.ส.สหสัมพันธ์ ที่ระดับนัยสำคัญ .05

ค) ถ้าการทดสอบสรุปว่ามีนัยสำคัญ จงประมาณเส้นการถดถอยของเปอร์เซนต์ผลิตภัณฑ์คุณภาพต่ำกว่ามาตรฐานบนอัตราเร็วของเครื่องจักรที่ตั้งไว้

ง) ถ้าให้เครื่องจักรทำงานในอัตราเร็ว 15 รอบต่อนาที จงทำนายเปอร์เซนต์ผลิตภัณฑ์คุณภาพต่ำกว่ามาตรฐาน ในช่วงเชื่อมั่น 95%

10. นักเศรษฐศาสตร์สุ่มตัวอย่างการบริโภคและรายได้ในช่วงเดียวกัน มา 13 ปีติดต่อกัน เพื่อดูความสัมพันธ์ของมัน ได้ผลดังตารางต่อไปนี้

ปีที่	รายได้ (พันบาท)	การบริโภค (พันบาท)
1	18.7	10.0
2	21.5	10.5
3	18.5	11.0
4	19.6	11.5
5	18.2	12.0
6	20.8	12.5
7	21.6	13.0
8	22.4	13.5
9	23.3	14.0
10	19.6	14.5
11	23.8	15.0
12	21.7	15.5
13	23.2	16.0

นักเศรษฐศาสตร์สรุปผลที่ได้ว่า รายได้และการบริโภคไม่มีความสัมพันธ์กัน ท่านเห็นด้วยหรือไม่ (กำหนดระดับนัยสำคัญ .05)

ถ้าไม่เห็นด้วย จงประมาณค่าเฉลี่ยของรายได้ถ้าการบริโภคในปีหนึ่ง เท่ากับ 12,000 บาท

11. ค่าของ r ในแต่ละหัวข้อต่อไปนี้ควรเป็นอย่างไร ให้เหตุผล

ก) รายได้ต่อคนกับจำนวนประชากร

ข) ความหนาแน่นของประชากรกับอัตราอาชญากรรม

ค) การเจริญเติบโตทางเศรษฐกิจกับการว่างงาน

ง) การใช้ปุ๋ยกับจำนวนผลผลิตที่เก็บได้

จ) จำนวนสัปดาห์ของการอบรมวิธีการขายกับจำนวนที่ขายได้ต่อปีของพนักงานขาย

ฉ) รายได้ต่อปีของครอบครัวกับค่าใช้จ่ายสำหรับอาหาร