บทที่ 4
การวิเคราะห์สหสัมพันธ์
(Correlation Analysis)

การวิเคราะห์ regression ในบทที่ 2-3 จะบอกให้ทราบว่า ตัวแปรมีความสัมพันธ์กันอย่างไร (How variables are related) โดยชี้ให้เห็นถึงการเปลี่ยนแปลงในตัวแปรตาม จะเป็นไปตามการเปลี่ยนแปลงในตัวแปรอิสระ แต่ในการวิเคราะห์สหสัมพันธ์นั้น มุ่งแสดงให้ทราบว่า ตัวแปรมีความสัมพันธ์กันในระดับใด มากน้อยแค่ไหน (The degree to which variables are related) ความสัมพันธ์ที่ว่านี้เป็นความสัมพันธ์ซึ่งกันและกันโดยไม่มีความหมายว่า ตัวแปรใดเป็นเหตุและตัวแปรใดเป็นผลของความสัมพันธ์นั้น

ในการวิเคราะห์สหสัมพันธ์มีข้อสมมติเกี่ยวกับการกระจายของ X และ Y เป็นการกระจายร่วมแบบแจกแจงปกติชนิด 2 ตัวแปร (Bivariate Normal Distribution) โดยที่ทั้ง X และ Y ต่างก็เป็นตัวแปรสุ่ม (random variables) และในสหสัมพันธ์ของตัวแปรที่มากกว่า 2 ตัวขึ้นไป การกระจายของตัวแปรจะเป็นแบบ Multivariate Normal Distribution.

<u>หลักในการพิจารณาสหสัมพันธ์เบื้องต้น</u>

ในเบื้องต้น ให้ตัวแปร 2 ตัว X และ Y มีความสัมพันธ์กันเป็นเส้นตรง (linear relationship) ซึ่งอาจจะแสดงในรูปสมการเส้นตรงของ Y on X ($Y = a + bX$) หรือ X on Y ($X = a + bY$) สหสัมพันธ์จะเป็นดังนี้คือ

1. ลักษณะของสหสัมพันธ์

ก. ถ้ามีสหสัมพันธ์กันอย่างสมบูรณ์ (Perfect Correlation) สมการนั้นจะฟิตกับข้อมูลทุกตัว จุดทุกจุดจะอยู่บนเส้นตรง (รูป b และ d)

ข. ถ้าสหสัมพันธ์เป็นบวก (Positive Correlation) เส้นแสดงความสัมพันธ์

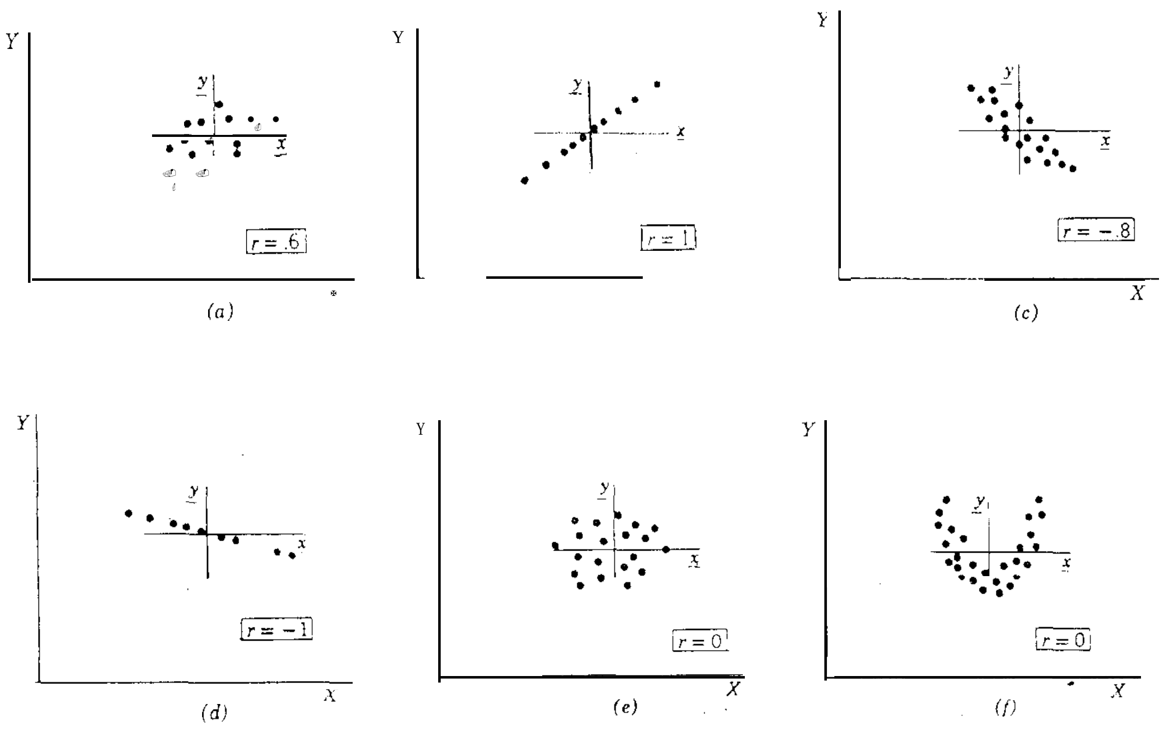
$\hat{Y} = \hat{a} + \hat{b}X$ จะเอียงขึ้น แสดงว่า X และ Y มีความสัมพันธ์ไปทางเดียวกัน (รูป a และ b)

ค. ถ้าสหสัมพันธ์เป็นลบ (Negative Correlation) เส้นแสดงความสัมพันธ์ $\hat{Y} = \hat{a} + \hat{b}X$ จะเอียงลง แสดงว่า X และ Y มีความสัมพันธ์ในทางตรงกันข้าม (รูป c และ d)

ง. ถ้าไม่มีสหสัมพันธ์ (No Correlation) ไม่สามารถจะหาเส้นแสดงความสัมพันธ์ได้ ลักษณะของข้อมูลจะรวมกันอยู่เป็นกลุ่มหรือกระจุกและอาจจะไม่อยู่ในแนวเส้นตรง (รูป e และ f)

2. ระดับสหสัมพันธ์

ระดับสหสัมพันธ์ระหว่างตัวแปร (X และ Y) จะมากน้อยแค่ไหนวัดด้วยสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ใช้สัญลักษณ์ว่า r ดังนั้นสัมประสิทธิ์สหสัมพันธ์ คือ การวัดระดับซึ่งตัวแปรมีความสัมพันธ์ซึ่งกันและกัน โดย r จะมีค่าตั้งแต่ -1ถึง 1 ถ้าตัวแปรไม่มีความสัมพันธ์กัน (No Correlation) r=0 แต่ถ้ามีความสัมพันธ์กันอย่างสมบูรณ์ r =-1 กับ 1



รูปที่ 4.1 ลักษณะและระดับของสัมประสิทธิ์สหสัมพันธ์ของข้อมูล

ในทางปฏิบัติ มักจะไม่ใคร่เห็นความสัมพันธ์ในลักษณะสมบูรณ์ ปกติ r จะเป็นจำนวนบวกหรือลบระหว่าง -1 กับ 1 ยิ่งใกล้ -1 และ 1 ความสัมพันธ์ยิ่งมีมาก ความสัมพันธ์ใกล้ -1 และ 1 จุดกระจายของข้อมูลจะอยู่ใกล้ ๆ เส้นตรง แต่โดยมากจุดกระจายของข้อมูลมักจะเบนออกจากเส้นตรง ก็จะทำให้ r ยิ่งใกล้ 0 ในรูป (a)

ข้อแตกต่างระหว่าง Correlation กับ Regression

Correlation และ Regression จะให้ความรู้ที่ต่างกันแต่มีการเชื่อมระหว่างกัน ข้อที่แตกต่างกันมีดังนี้

1. การวิเคราะห์ regression เลือกตัวแปรหนึ่งเป็นตัวแปรตามหรือตัวถูกพยากรณ์และตัวแปรอื่นเป็นตัวแปรอิสระซึ่งเป็นลักษณะสัมพันธ์กันอย่างเป็นฟังชั่น (funtional relationship) ระหว่างตัวแปร ส่วน Correlation ไม่ได้สัมพันธ์กันอย่างเป็นฟังชั่นหากแต่มีความสัมพันธ์กันอย่างธรรมดาเท่านั้น ใน Correlation บอกว่า มีความแปรปรวนร่วมระหว่างตัวแปร แต่มิได้บอกว่าความแปรปรวนใน Y เป็นสาเหตุมาจากความแปรปรวนใน X ดังนั้นในบางครั้ง Correlation อาจจะมาจากตัวแปรที่ไม่ได้เกี่ยวข้องกัน เช่นความสัมพันธ์ระหว่างจำนวนนักศึกษารามคำแหงกับจำนวนฝนตกในอเมริกา ซึ่งความสัมพันธ์เช่นนี้ เรียกว่า ความสัมพันธ์ที่เหลวไหล (Nonsense Correlation)

2. สัมประสิทธิ์สหสัมพันธ์เมื่อกำหนดขึ้นมาแล้วจะไม่เปลี่ยนแปลงแม้เส้นตรงจะเปลี่ยนไป เช่นเมื่อเราเลื่อนทุกจุดในรูป (b) ขึ้นหรือลงเป็นจำนวนเท่า ๆ กัน สัมประสิทธิ์สหสัมพันธ์ก็ยังคงเท่ากับ 1 ส่วนใน regression เราจะได้ Y-intercept ใหม่ ทำนองเดียวกัน ถ้าหมุนทุกจุดรอบ ๆ จุด ๆ หนึ่งแลจุดเหล่านี้จะตกอยู่บนเส้นตรงต่าง ๆ กันด้วย slope ที่ต่างกัน สัมประสิทธิเส้นถดถอย b (regression Coefficient) จะเปลี่ยนไปแต่สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) จะยังคงเดิม

3. สมการ regression สามารถจะนำไปใช้ในการพยากรณ์ต่อไปได้ซึ่งใน Correlation จะทำไม่ได้

อย่างไรก็ตามแม้จะมีข้อแตกต่างกันดังกล่าว แต่ก็มีการเชื่อมโยงระหว่าง Correlation กับ regression เราทราบว่า r เป็นบวกเส้นตรงต้องมี slope เป็นบวก และเมื่อ b เป็น slope ของเส้นตรง ค่า b ต้องเป็นบวกด้วย ตรงกันข้าม ถ้า r เป็นลบ b ต้องเป็นลบ และถ้า r = 0 , b ก็ต้องเป็น 0

<u>การคำนวณสัมประสิทธิ์สหสัมพันธ์</u>

สัมประสิทธิ์สหสัมพันธ์มีสูตรในการคำนวณดังนี้

1. สูตร Product Moment Method (โดยไม่ต้องคำนวณสมการ regression ก่อน)

ก. ใช้ข้อมูลโดยตรง

** $$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\Sigma(Y-\bar{Y})^2}} \qquad (1)$$

สูตร Product Moment Method คำนวณมาจาก

$$r = \frac{Cov.X,Y}{\sqrt{Var.X \cdot Var.Y}}$$

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

แล้วจะได้เป็นสูตร (1) จากสูตรนี้อาจจะเขียนได้อีกรูปหนึ่งเป็น

$$r = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\sqrt{(\Sigma X^2 - n\bar{X}^2)(\Sigma Y - n\bar{Y}^2)}} \qquad (2)$$

จากสูตร (1) จะได้ (2) ดังนี้

$$r = \frac{\Sigma(XY - X\bar{Y} - Y\bar{X} + \bar{X}\bar{Y})}{\sqrt{\Sigma(X^2 - 2X\bar{X} + \bar{X}^2)\Sigma(Y^2 - 2Y\bar{Y} + \bar{Y}^2)}}$$

$$= \frac{\Sigma XY - \Sigma X\bar{Y} - \bar{X}\Sigma Y + n\bar{X}\bar{Y}}{\sqrt{(\Sigma X^2 - 2\bar{X}\Sigma X + n\bar{X}^2)(\Sigma Y^2 - 2\bar{Y}\Sigma Y + n\bar{Y}^2)}}$$

$$= \frac{\Sigma XY - \Sigma X\frac{\Sigma Y}{n} - \frac{\Sigma X}{n}\Sigma Y + n\frac{\Sigma X}{n}\frac{\Sigma Y}{n}}{\sqrt{(\Sigma X^2 - 2\frac{\Sigma X}{n}\Sigma X + n\frac{\Sigma X\Sigma X}{n\,n})(\Sigma Y^2 - 2\frac{\Sigma Y}{n}\Sigma Y + n\frac{\Sigma Y\Sigma Y}{n\,n})}}$$

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

$$= \frac{\Sigma XY - n\frac{\Sigma X}{n}\frac{\Sigma Y}{n}}{\sqrt{(\Sigma X^2 - n\frac{\Sigma X^2}{n \cdot n})(\Sigma Y^2 - n\frac{\Sigma Y^2}{n \cdot n})}} \qquad \frac{\Sigma XY - n\bar{X}\bar{Y}}{\sqrt{(\Sigma X^2 - n\bar{X}^2)(\Sigma Y - n\bar{Y}^2)}}$$

ข. เขียนในรูปเบี่ยงเบนจาก mean (จากสูตร 1) จะเป็น

** $$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \qquad (3)$$

$x = X - \bar{X}$, $y = Y - \bar{Y}$

** หรือ $$r = \frac{S_{xy}}{S_x S_y} \qquad (4)$$

ซึ่งมาจาก $$r = \frac{\text{Cov.}X,Y}{\sqrt{\text{Var.}X \cdot \text{Var.}Y}} = \frac{\text{Cov.}X,Y}{\text{S.Dของ}X\ \text{S.Dของ}Y}$$

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

$$= \frac{S_{xy}}{S_x S_y}$$

การคำนวณโดยวิธีนี้ เครื่องหมายแสดงลักษณะของ r (+ หรือ -) เป็นไปตามผลที่คำนวณได้จากสูตร

สูตรจากวิธี Product Moment จะทำให้สัมประสิทธิ Y on X (r_{yx}) กับสัมประสิทธิสหสัมพันธ์ X on Y (r_{xy}) เท่ากันเสมอ เพราะ $r_{xy} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$ และ $r_{yx} = \frac{\Sigma yx}{\sqrt{\Sigma y^2 \; \Sigma x^2}}$ ซึ่งมีค่าเดียวกัน

2. สูตรมาจากความสัมพันธ์ระหว่าง Correlation และ Regression (ต้องคำนวณสมการ regression)

$$r = b \cdot \frac{S_x}{S_y} \qquad (5)$$

$$S_x = \sqrt{\frac{\Sigma x^2}{n-1}} \quad , \quad S_y = \sqrt{\frac{\Sigma y^2}{n-1}}$$

$$r = \pm\sqrt{R^2} \qquad (6)$$

สูตร R^2 ของ Simple Regression มีหลายสูตรดูจากบทที่ 2 โดยสูตรนี้เครื่องหมายของ r จะเป็นไปตาม b

สูตรที่กล่าวทั้งหมดนี้ใช้หาสัมประสิทธิสหสัมพันธ์จากตัวอย่าง (Sample Correlation Coefficient)

ความสัมพันธ์ระหว่าง Correlation และ Regression

เราอาจจะแสดงความสัมพันธ์ในเชิงคณิตศาสตร์ระหว่าง Correlation และ regression ได้ดังนี้

จากการวิเคราะห์ Regression' จะได้

$$b = \frac{\Sigma xy}{\Sigma x^2} \qquad (1)$$

แต่ r จากสูตรของวิธีที่ 2 ข้างต้น

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}} \qquad (2)$$

(1) $\div$ (2) จะได้

$$\frac{b}{r} = \frac{\Sigma xy/\Sigma x^2}{\Sigma xy/\sqrt{\Sigma x^2 \, \Sigma y^2}}$$

$$= \frac{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}{\Sigma x^2}$$

$$= \frac{\sqrt{\Sigma y^2}}{\sqrt{\Sigma x^2}}$$

หารทั้งเศษและส่วนภายใน Square root ด้วย n-1

$$\frac{r}{b} = \sqrt{\frac{\Sigma y^2/n-1}{\Sigma x^2/n-1}} = \frac{S_y}{S_x}$$

$$r = \frac{bS_x}{S_Y}$$

ความสัมพันธ์ระหว่าง r กับ R^2

จาก Regression

$$R^2 = \frac{b\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(Y - \bar{Y})^2}$$

แต่ $b = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2}$ (ดูบทที่ 2 หน้า 33)

แทนค่า b จะได้

$$R^2 = \frac{[\Sigma(X - \bar{X})(Y - \bar{Y})]^2}{\Sigma(X - \bar{X})^2 \, \Sigma(Y - \bar{Y})^2}$$

จาก Correlation

$$r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2} \sqrt{\Sigma(Y - \bar{Y})^2}}$$

ยกกำลังสองทั้งสองข้างจะได้

$$r^2 = \frac{[\Sigma(X - \bar{X})(Y - \bar{Y})]^2}{\Sigma(X - \bar{X})^2 \, \Sigma(Y - \bar{Y})^2}$$

$$r^2 = R^2$$

$$r = \pm\sqrt{R^2}$$

การทดสอบนัยสำคัญของ r

เมื่อคำนวณ r ได้แล้ว จะต้องทดสอบว่ามีนัยสำคัญทางสถิติหรือไม่ การทดสอบทำได้ 2 วิธีคือ

1. **ทดสอบด้วยค่า r ที่คำนวณได้**

โดยมีสมมติฐาน $r = 0$ ทดสอบค่า r ตามระดับความเชื่อมั่นที่กำหนดจากตาราง r , $d.f = n-2$ ถ้า r ที่คำนวณได้มากกว่าค่า r ในตาราง r จะปฏิเสธสมมติฐาน $r = 0$ กล่าวได้ว่า r มีนัยสำคัญ ถือว่ามีสหสัมพันธ์ระหว่างตัวแปรจริง ถ้าตรงข้าม ก็ยอมรับ $r = 0$ r ไม่มีนัยสำคัญนั่นคือไม่มีสหสัมพันธ์กันระหว่างตัวแปร

2. ทดสอบด้วยค่า t

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

โดยการแจกแจง t มี $d.f = n-2$ สมมติฐานที่ทำการทดสอบคือ $r = 0$ เปรียบเทียบค่า t ที่คำนวณจากสูตรข้างบนกับค่า t จากตาราง t ตามระดับความเชื่อมั่นที่กำหนด ถ้า $t > t$ จากตาราง จะปฏิเสธสมมติฐาน และสรุปได้เช่น-เดียวกับข้างต้น

ข้อสังเกตที่ได้จากการทดสอบ r :

(1) ค่า r ยิ่งสูง แนวโน้มที่จะมีนัยสำคัญจะมากขึ้น สำหรับขนาดของตัวอย่างนั้น

(2) การทดสอบสมมติฐาน $r = 0$ และ $b = 0$ ด้วยค่า t ซึ่งค่า t สำหรับทดสอบ r ($t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$) และ t สำหรับทดสอบ b ($t = \frac{\hat{b}}{S_{\hat{b}}}$) จะได้ค่าเท่ากัน (อาจจะต่างกันเพียงจุดทศนิยม) แสดงว่า เมื่อ $r = 0$ หมายความว่า b ก็จะเท่ากับ 0 ด้วย

<u>ตัวอย่างที่ 4.1</u>

การขายยาสีฟันขึ้นอยู่กับระดับการโฆษณาอย่างหนัก ดังข้อมูลค่าใช้จ่ายในการโฆษณาและการขายประจำปี ของยาสีฟันยี่ห้อที่รู้จักกันดี มีดังนี้คือ

	คอลเกต	ซื่อสัตย์	ใกล้ชิด	สินไทย
การโฆษณาประจำปี (X) (ล้านบาท)	2	4	3	1
การขายประจำปี (Y) (ล้านบาท)	5	7	6	2

จงหาสัมประสิทธิ์สหสัมพันธ์

วิธีทำ

ตารางที่ 4.1

Y	X	Y - Y = y	x-x= x	xy	y^2	x^2
5	2	0	-0.5	0	0	0.25
7	4	2	1.5	3.0	4	2.25
6	3	1	0.5	0.5	1	0.25
2	1	-3	-1.5	4.5	9	2.25
20	10			8.0	14	5.00

$\bar{Y} = 5$ $\bar{X} = 2.5$

สูตร $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$

$= \dfrac{8}{\sqrt{(5)(14)}} = \dfrac{8}{8.37}$

$= 0.96$

หรือคำนวณสมการ regression ก่อน

$$\hat{b} = \frac{\Sigma xy}{\Sigma x^2} = \frac{8}{5} = 1.6$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 5 - (1.6)\ (2.5) = 1$$

$$Y = 1 + 1.6\sim$$

$$r = b.\frac{S_x}{S_Y}$$

$$S_x = \sqrt{\frac{\Sigma x^2}{n-1}} = \sqrt{\frac{5}{4-1}} = 1.29$$

$$S_Y = \sqrt{\frac{\Sigma y^2}{n-1}} = \sqrt{\frac{14}{4-1}} = 2.16$$

$$r = (1.6)\frac{(1.29)}{2.16} = \frac{2.064}{2.16}$$

$$= 0.96$$

แสดงว่า การโฆษณาและการขายมีความสัมพันธ์กันถึง 96%

<u>ทดสอบ</u> r มีนัยสำคัญณ..ระดับความเชื่อมั่น 95% (r จากตาราง = 0.95)

<u>สหสัมพันธ์ของประชากร</u> (Population Correlation)

ค่า r ส่วนใหญ่จะหาจาก sample สหสัมพันธ์จากประชากร (population) ก็อาจจะหาได้เช่นกัน โดยสัมประสิทธิสหสัมพันธ์จากประชากร ซึ่งนิยามเหมือนกับของ r คือ

$$\rho = \frac{1}{N}\sum_{i=1}^{N} \underline{x}\ \underline{y}$$

ρ = สัมประสิทธิสหสัมพันธ์ของประชากร (อ่านว่า rho)

N = ขนาดของประชากร

$$x = \frac{X - \mu_X}{\sigma_X}$$ (μ คือ mean ของประชากร)

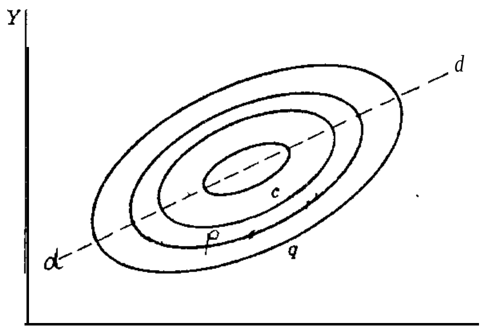
$$\underline{y} = \frac{Y - \mu_Y}{\sigma_Y}$$

ฉะนั้นอาจจะเขียน ρ ในรูป expectation ได้ดังนี้

$$\rho = E(\underline{x}\ \underline{y})$$

$$= E\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{a_Y}\right)$$

ρ มีค่าเหมือนกับ r คือมีค่าอยู่ระหว่าง -1 กับ 1 ค่า ρ อาจจะแสดงได้โดยเส้นความน่าจะเป็นเท่ากัน (isoprobability)

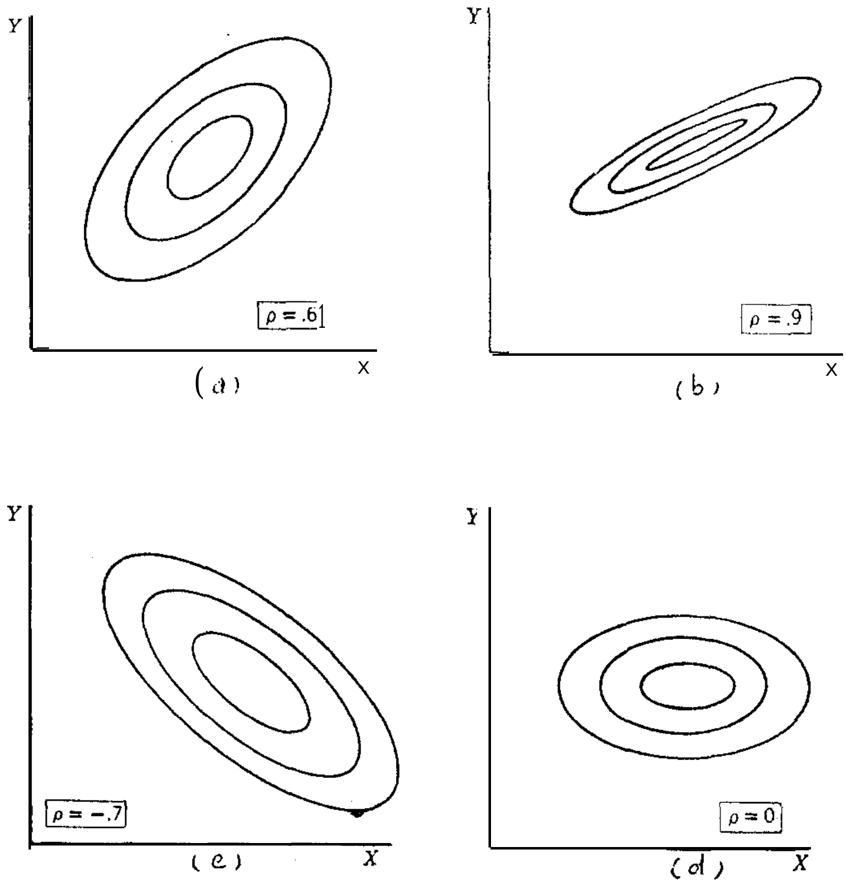


รูปที่ ๔.๒ รูปวงรี isoprobability แสดงถึง

Bivariate Normal Distribution

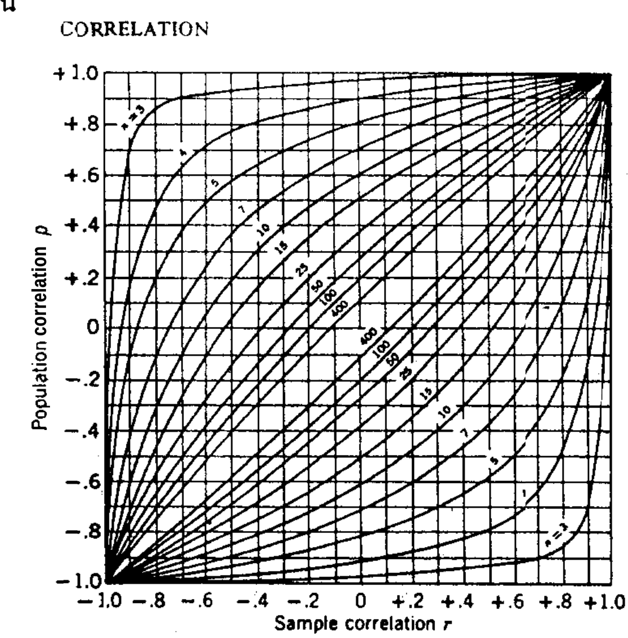
จากรูปที่ 4.2 รูปวงรี c จะมี probability มากกว่าวงรี p วงรี p มี prob. มากกว่า q ฉะนั้น c จึงมี prob. มากกว่าทั้ง p และ q เส้น dd เป็นเส้นตรงที่ตัดส่วนกว้างที่สุดของวงรี เส้นนี้เรียกว่า เส้นหลัก (marjor axis) สำหรับเป็นหลักในการตีความค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร ถ้าหากการกระจายของประชากร

รวมอยู่ใกล้เส้นหลัก ค่า ρ จะมีมาก (รูป 4.3b) ถ้าค่าการกระจายเท่า ๆ กันรวมอยู่รอบ ๆ เส้นหลักห่างออกมา ค่า ρ จะน้อย (รูป 4.3d) และค่า ρ จะเป็น + หรือ - ขึ้นอยู่กับเส้น d ว่าเอียงไปด้านใด



รูปที่ ๔.๓ สหสัมพันธ์ของประชากร

ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ) มีเพียงค่าเดียวจึงเป็นค่าที่คงที่ ส่วนสัมประสิทธิ์สหสัมพันธ์จากตัวอย่าง (r) เปลี่ยนแปลงได้สุดแต่ความแตกต่างของตัวอย่างที่นำมาศึกษาจะนำมาต่างกันอย่างไร อย่างไรก็ตาม สัมประสิทธิ์สหสัมพันธ์จากประชากรและจากตัวอย่างจะมีความสัมพันธ์กัน (ดังรูปที่ 4.4) ในช่วงความเชื่อมั่น 95% ถ้าขนาดของตัวอย่าง (Sample size : n) ใหญ่เส้นน่าจะเป็นซึ่งแสดงความสัมพันธ์ (มี 2 เส้น) จะแคบและจะกว้างขึ้นถ้าตัวอย่างเล็กลง จากรูปนี้ซึ่งเป็นรูปแสดง 95% ของช่วงความเชื่อมั่น (รูป 99% ก็มี) เราสามารถที่จะประมาณค่า ρ โดยอาศัยค่า r ได้ สมมติว่า r ที่คำนวณได้จากขนาดตัวอย่าง $n = 25$ มีค่าเท่ากับ 0.2 ค่า ρ ในช่วงความเชื่อมั่น 95% จะเป็นเท่าไร ก็ดูตามเส้นจาก $r = 0.2$ ขนานกับแกน ρ ตัดกับเส้นโค้งที่เกิดจาก $n = 25$ สองเส้นที่ใดก็อ่านค่าบนแกน ρ ซึ่งในที่นี้จะได้ค่า ρ ประมาณ -0.2 และ 0.45 นั่นคือค่า ρ จะอยู่ระหว่าง -0.2 ถึง 0.45 $(-0.2 < \rho < 0.45)$ ถ้า $r = 0.2$ เหมือนเดิม แต่ $n = 400$ ค่า ρ จะอยู่ระหว่าง 0.1 กับ 0.3 จะเห็นว่าช่วงจะแคบเข้า ทำให้ค่า r และ ρ ใกล้เคียงกัน



รูปที่ ๔.๔ ความเชื่อมั่น ๙๕% สำหรับสหสัมพันธ์ ρ ของประชากร และ r ของตัวอย่างขนาด n ต่าง ๆ กัน

สรุปได้ว่า การหาค่าสัมประสิทธิจากประชากรและจากตัวอย่างในระดับความเชื่อมั่น 95% ขึ้นไป ถ้าหากขนาดของตัวอย่าง (n) ใหญ่ ค่าสัมประสิทธิสหสัมพันธ์ยิ่งใกล้กันมากขึ้น เพราะช่วงค่าของ ρ จะแคบแสดงว่า จะให้ค่าไม่แตกต่างจาก r มากนัก ดังนั้น ในทาง Statistical Inference หรือทางปฏิบัติทั่วไปจึงหาสัมประสิทธิสหสัมพันธ์จากตัวอย่างมากกว่าจากประชากร

<u>สหสัมพันธ์เชิงซ้อน</u> (Multiple Correlation)

สหสัมพันธ์เชิงซ้อนวัดได้ด้วยสัมประสิทธิสหสัมพันธ์เชิงซ้อน (Coefficient of Multiple Correlation) ใช้สัญลักษณ์ว่า R หรือ $R_{1.2,3...k}$

สัมประสิทธิสหสัมพันธ์เชิงซ้อน ((R หรือ $R_{1.2,3...k}$) คือการวัดความสัมพันธ์ของตัวแปรหนึ่งกับกลุ่มตัวแปรที่เหลือทั้งหมด หรือระดับ (degree) ความสัมพันธ์ของตัวแปรตาม Y หรือ X_1 กับตัวแปรอิสระ $X_2, X_3, ...X_k$

จากแนวความคิดการวัดสหสัมพันธ์ตามวิธี Moment Product

** $$R = \frac{Cov.(y.x_2, x_3, ...)}{\sqrt{V(y)\ ,\ V(x_2, x_3 ...)}}$$

$$= \frac{S_{Y,x_2,x_3...}}{\sqrt{S_y}\sqrt{S_{x_2,x_3...}}}$$

หรือจาก regression

$$R = \sqrt{R^2}$$

(ดูสูตร R^2 ของ Multiple Regression มีหลายสูตร)

การพิสูจน์ว่า $R = \sqrt{R^2}$ **ทำนองเดียวกับ ในสหสัมพันธ์อย่างง่าย**

ก็จะผ่านไป จากสูตรนี้ เราก็อาจจะได้นิยามของสัมประสิทธิสหสัมพันธ์เชิงซ้อน R อีกอย่างได้ว่า R คือ สัมประสิทธิสหสัมพันธ์อย่างง่ายของค่าสังเกต Y กับ $\hat{Y}$ ที่คำนวณได้

การทดสอบ ;

นำค่า R ที่คำนวณได้ไปเปรียบเทียบกับตาราง R ตามระดับความเชื่อมั่นที่กำหนด โดยมีสมมติฐาน R = O หรือไม่มีสหสัมพันธ์ระหว่างกลุ่มตัวแปร ถ้าที่คำนวณสูงกว่า R ในตารางก็ปฏิเสธสมมติฐานว่าไม่เป็นความจริง (มีนัยสำคัญ)

การทดสอบอาจจะใช้ค่าสถิติ F เหมือนใน Multiple Regression

($F = \dfrac{R^2/k}{1 - R^2/n-k-1}$) และเปิดตาราง F เปรียบเทียบ

<u>สหสัมพันธ์บางส่วน</u> (Partial Correlation)

มีบ่อยครั้งที่เราต้องการทราบสหสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระตัวใดตัวหนึ่ง โดยเฉพาะโดยที่ตัวแปรอิสระอื่น ๆ ที่มีอิทธิพลต่อตัวแปรทั้งสองถูกกำหนดให้คงที่ นั่นคือ อิทธิพลของตัวแปรอื่น ๆ ทั้งหมดถูกกำจัดออกไป ความสัมพันธ์เช่นนี้เรียกว่า สหสัมพันธ์บางส่วน (Partial Correlation) อย่างไรก็ตาม สหสัมพันธ์บางส่วนจะเกิดขึ้นได้ ตัวแปรจะต้องมีการแจกแจงแบบปกติร่วมกันหลายตัวแปร (multivariate normal distribution)

เมื่อเราต้องการหาสหสัมพันธ์ระหว่าง 2 ตัวแปร เราอาจจะพบว่าตัวแปรทั้งสองยังมีความสัมพันธ์กับตัวแปรที่สาม เช่น ความสูง น้ำหนัก และอายุ ของเด็กที่ไปโรงเรียนเป็นตัวแปร 3 ตัว ให้ X_1 เป็นความสูง X_2 เป็นน้ำหนัก และ X_3 เป็นอายุ เราต้องการหาสหสัมพันธ์ระหว่างความสูงและน้ำหนัก แต่โดยที่ทั้งความสูงและน้ำหนักสัมพันธ์กับอายุ (X_3) เราจะได้สหสัมพันธ์ที่แท้จริงระหว่าง X_1 และ X_2 ถ้าอิทธิพลของ X_3 ถูกกำจัดไปจาก X_1 และ X_2 สำหรับเทคนิคในการกำจัด X_3 ซึ่งจะทำให้ X_1 และ X_2 มีอิสระจากอิทธิพล

ของ X_3 จะผ่านไปไม่กล่าวถึง

การวัดสหสัมพันธ์บางส่วน เรียกว่า สัมประสิทธิ์สหสัมพันธ์บางส่วน (Partial Correlation Coefficient) ดังนั้น สัมประสิทธิ์สหสัมพันธ์บางส่วน คือ การวัดสหสัมพันธ์ระหว่าง 2 ตัวแปร ภายหลังที่กำจัดอิทธิพลของตัวแปรอื่นออกไปหรือโดยกำหนดให้ตัวแปรอื่นอยู่คงที่ มีสูตรว่า

$$** \qquad r_{12.3} = \frac{r_{12} - (r_{13})(r_{23})}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$r_{12.3}$ = สัมประสิทธิ์สหสัมพันธ์ระหว่าง X_1 และ X_2 โดยกำหนดให้ X_3 คงที่

r_{12} , r_{13} และ r_{23} = สัมประสิทธิ์สหสัมพันธ์อย่างง่าย ระหวาง X_1 กับ X_2 , X_1 กับ X_3 และ X_2 กับ X_3 ตามลำดับ

ทำนองเดียวกัน ถ้าจะหา Partial Corretation Coefficient ระหว่าง X_1 กับ X_3 โดย X_2 อยู่คงที่ $(r_{13\ 2})$ หรือ X_2 กับ X_3 เมื่อ X_1 คงที่ $(r_{23\ 1})$

$$** \qquad r_{13.2} = \frac{r_{13} - (r_{12})(r_{32}) -}{\sqrt{(1 - r_{12}^2)(1 - r_{32}^2)}} \text{ , } r_{23.1} = \frac{r_{23} - (r_{21})(r_{31})}{(1-r_{21}^2)(1-r_{31}^2)}$$

<u>ตัวอย่างที่ 4.2</u>

จงคำนวณหา Partial Correlation Coefficient ของข้อมูลต่อไปนี้

X_1	3	5	6	9	7
X_2	4	7	8	12	9
X_3	5	8	12	15	20

<u>วิธีทำ</u>

ตารางที่ ๔.๒

X_1	X_2	X_3	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_2x_3	x_3x_1
3	4	5	-3	-4	-7	9	16	49	12	28	21
5	7	8	-1	-1	-4	1	1	16	1	4	4
6	8	12	0	0	0	0	0	0	0	0	0
9	12	15	3	4	3	9	16	9	12	12	9
7	9	20	1	1	8	1	1	64	1	a	8
30	40	60				20	34	138	26	52	42

$\bar{X}_1 = 6$ $\quad \bar{X}_2 = 8$ $\quad \bar{X}_3 = 12$

1. หาสหสัมพันธ์อย่างง่ายทีละคู่ (โดยสูตร r ที่กล่าวมาแล้วข้างต้น)

$$r_{12} = \frac{\Sigma x_1 x_2}{\sqrt{\Sigma x_1^2 \ \Sigma x_2^2}}$$

$$= \frac{26}{\sqrt{(20)(34)}} = \frac{26}{26.07} = 0.997$$

$$r_{23} = \frac{\Sigma x_2 x_3}{\sqrt{\Sigma x_2^2 \ \Sigma x_3^2}}$$

$$= \frac{52}{\sqrt{(34)(138)}} = \frac{52}{68.5} = 0.759$$

$$r_{31} = \frac{x_3 x_1}{\sqrt{x_3^2\ x_1^2}}$$

$$= \frac{42}{\sqrt{(138)(20)}} \approx \frac{42}{52.5} = 0.80$$

2. Partial Correlation จะหาได้จากสูตร

$$r_{12.3} = \frac{r_{12} - (r_{13})(r_{23})}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

$$= \frac{0.997 - (0.80)(0.759)}{(1 - 0.80^2)(1 - 0.759^2)} = \frac{0.390}{0.3906} = 0.998$$

$$r_{23.1} = \frac{r_{23} - (r_{21})(r_{31})}{\sqrt{(1 - r_{21}^2)(1 - r_{31}^2)}}$$

$$= \frac{0.759 - (0.997)(0.80)}{\sqrt{(1 - 0.997^2)(1 - 0.80')}} = \frac{-0.03874}{0.044} \approx -0.88$$

$$r_{31.2} = \frac{r_{31} - (r_{32})(r_{12})}{\sqrt{(1 - r_{32}^2)(1 - r_{12}^2)}}$$

$$= \frac{0.80 - (0.997)(0.759)}{\sqrt{(1 - 0.997^2)(1 - 0.759)^2}} = \frac{0.04295}{0.0477} = 0.90$$

การทดสอบนัยสำคัญของ Partial Corretation Coefficient:

$$H_0 : r_{12.3} = 0$$

ทดสอบด้วย t

$$t = r_{12\,3} \frac{n - k - 1}{1 - r^2_{12.3}}$$

<u>สหสัมพันธ์ลำดับที่</u> (Rank Correlation)

เมื่อตัวแปรที่นำมาหาสหสัมพันธ์ไม่สามารถจะวัดออกมาเป็นตัวเลขที่แน่นอนได้ เช่น รสนิยม ความชอบของบุคคล คุณภาพ หรือคุณลักษณะบางประการ การหาสหสัมพันธ์จะทำได้โดยนำคุณลักษณะของตัวแปรเหล่านี้มาจัดลำดับที่ การวัดสหสัมพันธ์ชนิดนี้จึงเรียกว่า สัมประสิทธิ์สหสัมพันธ์ลำดับที่ (Rank Correlation Coefficient) Spearman เป็นผู้คิดสูตรขึ้นมาคือ

$$r_S = 1 - \frac{6(\Sigma D^2)}{n(n^2 - 1)}$$

r_S = Spearman's Correlation

D = ค่าแตกต่างของลำดับที่แต่ละคู่ $(R_1 - R_2)$ ของตัวแปรที่ต้องการหาสหสัมพันธ์

n = จำนวนคู่ของข้อมูล

r_S จะมีค่าเช่นเดียวกับ r คือค่าอยู่ระหว่าง -1 ถึง 1 ถ้า r_S เป็น -1 หรือ +1 แสดงถึงความสัมพันธ์กันอย่างสมบูรณ์ (perfect correlation) และ 0 แสดงว่าไม่มีความสัมพันธ์กัน

<u>ตัวอย่างที่ 4.3</u>

บริษัทใหญ่แห่งหนึ่งได้ออกสอบถามกลุ่มลูกค้า 2 กลุ่มเพื่อที่จัดลำดับความชอบสินค้าชนิดหนึ่งซึ่งมียี่ห้อต่าง ๆ กัน และได้จัดลำดับ (rank) ความชอบดังนี้คือ

สินค้ายี่ห้อ	A	B	C	D	E	F	G	H	I	J
ลูกค้ากลุ่มที่ 1 (R_1)	1	2	3	4	5	6	7	8	9	10
ลูกค้ากลุ่มที่ 2 (R_2)	1	4	5	2	7	3	6	9	10	8

จงหาสัมประสิทธิสหสัมพันธ์ของความชอบของลูกค้า 2 กลุ่มนี้

วิธีทำ

ตารางที่ 4.3

สินค้ายี่ห้อ	R_1	R_2	$D = R_1 - R_2$	D^2
A	1	i	0	0
B	2	4	-2	4
C	3	5	-2	4
D	4	2	2	4
E	5	7	-2	4
F	6	3	3	9
G	7	6	1	1
H	8	9	-1	1
I	9	10	-1	1
J	10	8	2	4
				32

$$r_S = 1 - \frac{6(\Sigma D^2)}{n(n^2 - 1)}$$

$$= 1 - \frac{6(32)}{10(10^2-1)} = \frac{192}{990}$$

$$= 0.19$$

แสดงว่า ความชอบของลูกค้าทั้ง 2 กลุ่มนี้ มีความสัมพันธ์กันเพียง 19%

อย่างไรก็ตาม แม้ตัวแปรที่วัดออกเป็นตัวเลขได้ แต่ถ้าเราไม่ทราบลักษณะการกระจายของมันตามข้อสมมติเบื้องต้นของสหสัมพันธ์ เราก็อาจจะหาสหสัมพันธ์ได้โดย Rank Correlation Coefficient ซึ่งนับว่าเป็นข้อดีประการหนึ่งของ Rank Correlation และข้อดีอีกประการหนึ่งคือ วิธีการคำนวณ r_s ง่ายกว่าสัมประสิทธิ์สหสัมพันธ์ r และถ้าข้อมูลไม่ซ้ำ (no tie) กัน r_s กับ r จะให้ค่าเดียวกัน แต่ถ้าซ้ำกัน r_s กับ r จะต่างกันเล็กน้อย (ปกติถ้าต่างกันน้อยเราจะไม่สนใจ) เพราะการจัดลำดับจะต้องเป็นลำดับเดียวกันโดยการเฉลี่ยสำหรับค่าที่ซ้ำ

<u>ตัวอย่างที่ 4.4</u>

ความสัมพันธ์ระหว่างการลดเวลาการทำงาน (X) กับการลดประสิทธิภาพ (Y) ดังข้อมูลต่อไปนี้

การลดเวลาทำงาน (X) (ช.ม)	14	34	18	40	48	18	44	30	26	38
การลดประสิทธิภาพ (Y) (หนวยงาน)	55	60	44	78	75	50	62	65	55	70

จงหาสัมประสิทธิ์สหสัมพันธ์โดยวิธี Rank Correlation Coefficient

<u>วิธีทำ</u>

ตารางที่ 4.4

X	Y	ลำดับของ X	ลำดับของY	ลำดับX - ลำดับ Y = D	D^2
14	55	10	7.5	2.5	6.25
34	60	5	6	-1	1
18	44	8.5	10	-1.5	2.25
40	78	3	1	2	4
48	75	1	2	-1	1
18	50	8.5	9	-0.5	0.25
44	62	2	5	-3	9
30	65	6	4	2	4
26	55	7	7.5	-0.5	0.25
38	70	4	3	1	1
					29

$$r_S = 1 - \frac{\Sigma(D^2)}{n(n^2-1)}$$

$$= 1 - \frac{6(29)}{10(100-1)}$$

$$= 0.824$$

ถ้าหาโดยสัมประสิทธิสหสัมพันธ์ r (ใช้สูตร r) จะได้ $= 0.829$ ซึ่งจะเห็นว่าค่า r_s กับ r ใกล้เคียงกัน แต่ข้อบกพร่องของ r_s คือ การคำนวณ r_s จะขาดข้อมูล

ที่เป็นจริงบางอย่างไป เป็นการคำนวณโดยประมาณ จึงทำให้ไม่เหมาะที่จะใช้กับข้อมูลที่วัดเป็นตัวเลขที่มีหน่วยแน่นอน

การทดสอบ :

$$H_o : r_S = 0$$

เนื่องจากเราใช้ตัวอย่าง จึงทดสอบด้วย t

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

วิธีการทดสอบว่าจะยอมรับหรือปฏิเสธสมมติฐานเหมือนกับที่กล่าวมาแล้ว

<u>แบบฝึกหัด</u>

๑. สมมุติสุ่มตัวอย่าง ๕ ครอบครัวให้ข้อมูลดังต่อไปนี้.-

หน่วย : บาท

ครอบครัวที่	การออม S	รายได้ Y	ทรัพย์สิน W
๑	๖๐๐	๘,๐๐๐	๑๒,๐๐๐
๒	๑,๒๐๐	๑๑,๐๐๐	๖,๐๐๐
๓	๑,๐๐๐	๙,๐๐๐	๖,๐๐๐
๔	๗๐๐	๖,๐๐๐	๓,๐๐๐
๕	๓๐๐	๖,๐๐๐	๑๘,๐๐๐

(คำถาม ๑.๑ ถึง ๑.๕ เป็นเรื่อง regression)

๑.๑ ฟีต Simple regression : S on Y และ Multiple regression : S on Y on W

๑.๒ สัมประสิทธิของ Y ต่างกันหรือไม่? สัมประสิทธิใดจะอธิบายความสัมพันธ์ S ต่อ Y ได้ดีกว่า.

๑.๓ ครอบครัวที่มีทรัพย์สิน ๕,๐๐๐ บาท และรายได้ ๘,๐๐๐ บาท การออมจะเป็นเท่าไร?

๑.๔ ถ้าครอบครัวมีรายได้เพิ่มขึ้น ๒,๐๐๐ บาท โดยที่ทรัพย์สินคงที่ การออมของเขาจะเพิ่มเป็นเท่าไร?

๑.๕ ถ้าครอบครัวมีรายได้เพิ่ม ๑,๐๐๐ บาท ทรัพย์สินเพิ่ม ๓,๐๐๐ บาท การออมจะเพิ่มเป็นเท่าไร?

๑.๖ หา Simple Correlation Coefficient ของ S และ Y (r_{sy})

๑.๗ หา Partial Correlation Coefficient ของ S และ Y โดยให้ W คงที่ ($r_{sy.w}$)

๑.๘ หา Multiple Correlation Coefficient ของ S on Y and W (R)

9.d r_{sy} หรือ $r_{sy.w}$ ตัวใดที่จะ เป็น เครื่องวัดที่ดีกว่ากันว่า "S และ y สัมพันธกัน อย่างไร เมื่อให้สิ่งอื่น เท่ากัน"

๒. มีการสุ่มผู้สูงอายุถามคำถาม เกี่ยวกับกีฬาและ เหตุการณ์ของโลก ซึ่งผลคำถามได้ แปลออกมาใน รูปของ "คะแนนความรู้" ดังนี้

หน่วย : คะแนน

ชื่อ	ก	ข	ค	ง	จ	ฉ	ช	ซ	ณ	ญ
กีฬา	๔๗	๑๒	๖๒	๘๑	๙๐	๓๕	๖๒	๘๗	๕๙	๔๐
เหตุการณ์ของโลก	๔๙	๑๐	๗๖	๙๒	๘๖	๔๒	๖๑	๗๖	๘๖	๖๑

จงหาระดับความสัมพันธ์ลำดับที่

คำตอบข้อ ๑

๑.๑ S = ๗๖๐ + ๐.๑๑๔y , S = ๗๖๐ + ๐.๑๑๕y - ๐.๐๒๙๔ W

๑.๒ ๐.๑๑๕ ดีกว่า เพราะแสดงความสัมพันธ์ของ S ต่อ Y เมื่อ W คงที่

๑.๓ ๘๗๘

๑.๔ ๒๓๐

๑.๕ ๒๗

๑.๖ ๐.๘๗๔

๑.๗ ๐.๙๘๒

๑.๘ ๐.๙๙๒

๑.๙ $r_{sy.w}$