บทที่ 11
สหสัมพันธ์และสมการถดถอยเชิงเส้นตรง
(CORRELATION AND LINEAR REGRESSION EQUATION)

วัตถุประสงค์

เมื่อท่านศึกษาเนื้อหาบทที่ 11 โดยละเอียดแล้ว ควรมีความสามารถ ดังนี้

1. แปลความหมายค่าสัมประสิทธิ์สหพันธ์ระหว่างตัวแปร 2 ตัวได้
2. คำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์เชิงเส้นตรงได้
3. ทดสอบความมีนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์ได้
4. สร้างสมการถดถอยเชิงเส้นตรงได้
5. วิเคราะห์ข้อมูลทางการศึกษาโดยใช้วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ และการสร้างสมการถดถอยเชิงเส้นตรงได้

เนื้อหา

11.1 ความหมายของค่าสัมประสิทธิ์สหสัมพันธ์

11.2 วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์

11.3 การทดสอบความมีนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์ได้

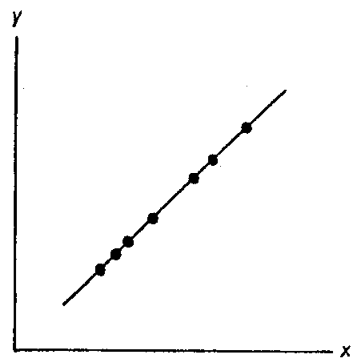
11.4 สมการถดถอยเชิงเส้นตรง (Linear Regression Equation)

เนื้อหาที่ 11.1 ความหมายของค่าสัมประสิทธิ์สหสัมพันธ์

เมื่อตัวแปรสุ่ม X และ Y มีความสัมพันธ์กัน เราอาจหาค่าแสดงความสัมพันธ์ระหว่าง X กับ Y ได้ ค่าที่แสดงความสัมพันธ์นี้เรียกว่า สัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) เขียนแทนด้วยสัญลักษณ์ r_{xy} r_{xy} จะมีค่าอยู่ระหว่าง + 1.00 ถึง –1.00

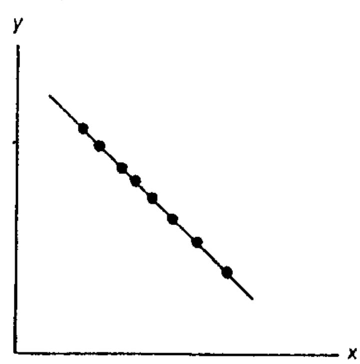
ถ้า r_{xy} มีค่าเท่ากับ + 1.00 แสดงว่า X และ Y มีความสัมพันธ์กันอย่างใกล้ชิด ซึ่งเรียกได้ว่าเป็นความสัมพันธ์ในทางตรง นั่นคือถ้า X มีค่ามาก Y ก็จะมีค่ามากตามไปด้วย และถ้า X มีค่าน้อย Y ก็จะมีค่าน้อยตามไปด้วยเช่นกัน เราสามารถเขียนกราฟแสดงความ

สัมพันธ์ระหว่าง X กับ Y ได้ กราฟนี้เรียกว่าสแคทเตอร์แกรม (scattergram) รูปที่ 11.1 จะเป็นกราฟแสดงความสัมพันธ์ เมื่อ $r_{xy} = +1.00$



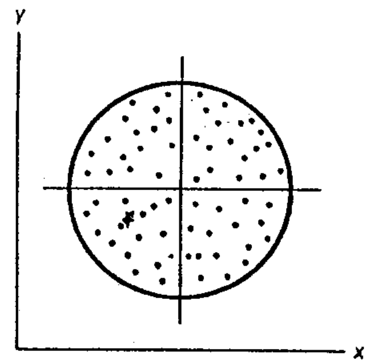
รูป 11.1 *กราฟแสดงความสัมพันธ์ระหว่าง* X *และ* Y *เมื่อ* $r_{xy} = +1.00$

ถ้า r_{xy} มีค่าเท่ากับ -1.00 แสดงว่าตัวแปร X และ Y มีความสัมพันธ์ในทางกลับกัน นั่นคือ ถ้า X มีค่ามาก Y จะมีค่าน้อย แต่ถ้า X มีค่าน้อย Y จะมีค่ามาก ความสัมพันธ์เชิงกลับระหว่าง X กับ Y แสดงให้เห็นได้อย่างชัดเจน จากกราฟแสดงความสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y ในรูป 11.2



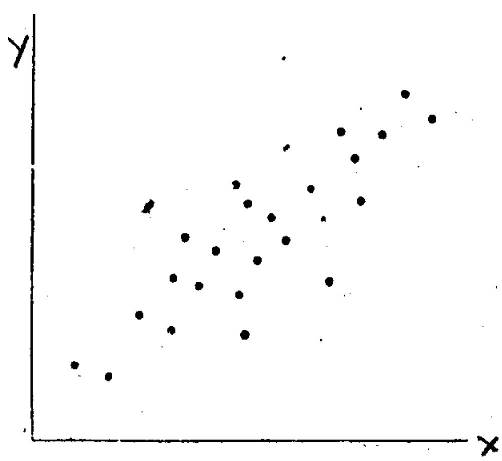
รูป 11.2 *กราฟแสดงความสัมพันธ์ระหว่าง* X *และ* Y *เมื่อ* $r_{xy} = -1.00$

ถ้า r_{xy} มีค่าเป็นศูนย์ ($r_{xy} = 0.00$) แสดงว่าตัวแปร X และตัวแปร Y **ไม่มีความสัมพันธ์**กันในเชิงเส้นตรง กราฟแสดงความสัมพันธ์ระหว่างตัวแปร X และ Y เมื่อ $r_{xy} = 0.00$ แสดงไว้ในรูป 11.3

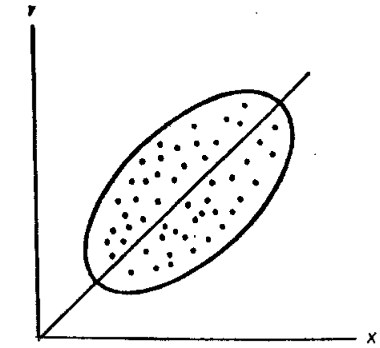


รูป 11.3 *กราฟแสดงความสัมพันธ์ระหว่าง* X *และ* Y *เมื่อ* $r_{xy} = 0.00$

ในกรณีที่ตัวแปร X และตัวแปร Y มีความสัมพันธ์กันสูงมากในทางบวก แต่ค่าน้อยกว่า + 1.00 ลักษณะของกราฟแสดงความสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y จะเป็นดังรูปที่ 11.4 และ 11.5 ตามลำดับ

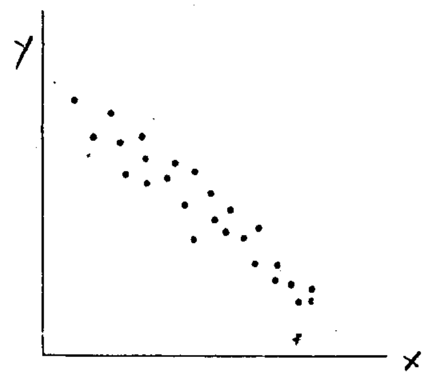


รูป 11.4 *กราฟแสดงความสัมพันธ์ระหว่าง* X *และ* Y *เมื่อ* X *และ* Y *มีความสัมพันธ์สูงมากในทางบวก* ($0.0 < r < +1.00$)

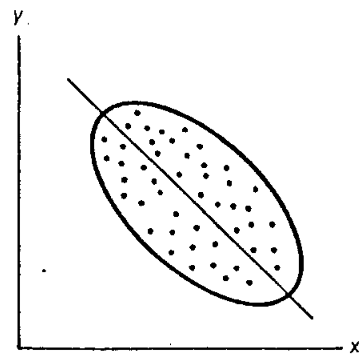


รูป 11.5 *กราฟแสดงความสัมพันธ์ระหว่าง X และ Y เมื่อ X และ Y มีความสัมพันธ์ในทางบวก* (0.0 < r < + 1.00)

ในกรณีที่ตัวแปร X และตัวแปร Y มีความสัมพันธ์กันสูงในทางลบ ลักษณะของกราฟแสดงความสัมพันธ์ระหว่างตัวแปร X และ Y จะเป็นดังรูป 11.6 และ 11.7



รูป 11.6 *กราฟแสดงความสัมพันธ์ระหว่าง X และ Y เมื่อ X และ Y มีความสัมพันธ์กันสูงมากในทางลบ* (–1.00 < r < 0.0)



รูป 11.7 *กราฟแสดงความสัมพันธ์ระหว่าง X และ Y เมื่อ X และ Y มีความสัมพันธ์ในทางลบ* (–1.00 < r < 0.0)

นอกจากค่าสหสัมพันธ์จะหมายถึงความสัมพันธ์ระหว่างตัวแปร 2 ตัวแล้ว ยังหมายถึงความสอดคล้องของอันดับของตัวแปร 2 ตัวด้วย ตัวอย่างเช่น สมมุติว่าทำการทดสอบแบบทดสอบ 4 ฉบับ คือ ก ข ค และ ฉบับ ง กับนักเรียน 5 คน นำผลของแบบทดสอบฉบับ ก มาเปรียบเทียบกับฉบับ ข ฉบับ ค และฉบับ ง ดังนี้

กรณีที่ 1			**กรณีที่ 2**			**กรณีที่ 3**		
นักเรียน	ก	ข	นักเรียน	ก	ค	นักเรียน	ก	ง
1	15	53	1	15	64	1	15	102
2	14	52	2	14	65	2	14	100
3	13	51	3	13	66	3	13	104
4	12	50	4	12	67	4	12	103
5	11	49	5	11	68	5	11	101

ถ้านำคะแนนของนักเรียนมาจัดอันดับโดยกำหนดให้อันดับที่ 1 คือผู้ที่สอบได้คะแนนสูงสุด อันดับที่ 5 คือผู้ที่สอบได้คะแนนต่ำสุด จะได้ผลดังนี้

กรณีที่ 1		**กรณีที่ 2**		**กรณีที่ 3**	
ก	ข	ก	ค	ก	ง
1	1	1	5	1	3
2	2	2	4	2	2
3	3	3	3	3	5
4	4	4	2	4	4
5	5	5	1	5	1
$r_{xy} = 1.00$		$r_{xy} = -1.00$		r_{xy} มีค่าต่ำและติดลบ	

จากตัวอย่างจะเห็นว่า ในกรณีที่ 1 นั้นอันดับที่ของนักเรียนในแบบทดสอบฉบับ ก กับฉบับ ข สอดคล้องต้องกัน นั่นคือ นักเรียนคนที่สอบได้ที่ 1 จากข้อสอบฉบับ ก ก็สอบได้ที่ 1 จากข้อสอบฉบับ ข ด้วย ในทำนองเดียวกันคนที่สอบได้คะแนนเป็นอันดับ 4 จากข้อสอบฉบับ ข ได้เป็นอันดับ 4 ด้วย เนื่องจากนักเรียนทั้งฉบับ ก และฉบับ ข ได้คะแนนเป็นอันดับที่เดียวกัน ค่า r_{xy} จึงเท่ากับ 1.00

เหตุการณ์จากกรณีที่ 2 เป็นเหตุการณ์ที่ตรงข้ามกับกรณีที่ 1 ทั้งนี้เพราะนักเรียนคนที่ทำข้อสอบฉบับ ก ได้คะแนนสูงเป็นอันดับ 1 กลับทำข้อสอบฉบับ ค ได้คะแนนเป็นอันดับสุดท้าย ส่วนนักเรียนคนที่สอบข้อสอบฉบับ ก ได้คะแนนเป็นอันดับสุดท้าย กลับทำข้อสอบฉบับ ค ได้เป็นอันดับหนึ่ง ค่า r_{xy} จึงเท่ากับ—1.00

เหตุการณ์ในกรณีที่ 3 ก็สามารถอธิบายได้ทำนองเดียวกันกับสองกรณีที่กล่าวมาแล้วข้างต้น

เนื้อหาที่ 11.2 วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์

วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์มีหลายวิธี แต่ในที่นี้จะอธิบายถึงวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ที่ง่าย ๆ และเป็นที่นิยมใช้กันทั่วไป 2 วิธี คือ

11.2.1 แบบเปียร์สัน (Pearson Product Moment Correlation)

11.2.2 แบบสเปียร์แมน (Spearman Rank Correlation Coefficient)

11.2.1 วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเปียร์สัน

ในกรณีที่ตัวแปร X และ Y เป็นตัวแปรสุ่มแบบต่อเนื่อง (นั่นคือต้องเป็นผลที่ได้จากการวัดในมาตรอันตรภาคหรือมาตราอัตราส่วน) เราสามารถหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และ Y โดยวิธีของเปียร์สัน (Pearson Product Moment Correlation) จากสูตร 11.1 ดังนี้

สูตร 11.1

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} \quad ....(11.1)$$

เมื่อ
r_{xy} = ค่าสัมประสิทธิ์สหสัมพันธ์
N = จำนวนกลุ่มตัวอย่าง
ΣX = ผลรวมของตัวแปร X
ΣY = ผลรวมของตัวแปร Y
ΣXY = ผลบวกของผลคูณระหว่าง X กับ Y
ΣX^2 = ผลรวมของ X^2
ΣY^2 = ผลรวมของ Y^2

ตัวอย่าง 11.1 ในการทดสอบย่อยวิชาคณิตศาสตร์และวิทยาศาสตร์ของนักเรียน 10 คน ปรากฏผลดังตารางข้างล่างนี้

คณิตศาสตร์ (X)	1	2	2	3	4	5	6	8	9	10
วิทยาศาสตร์ (Y)	2	3	2	5	5	7	5	6	7	8

ให้หาค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนคณิตศาสตร์และวิทยาศาสตร์

วิธีทำ

คำนวณหา $\sum X$, $\sum Y$, $\sum X^2$, $\sum Y^2$ และ $\sum XY$ ดังนี้

นักเรียนคนที่	X	Y	X^2	Y^2	XY
1	1	2	1	4	2
2	2	3	4	9	6
3	2	2	4	4	4
4	3	5	9	25	15
5	4	5	16	25	20
6	5	7	25	49	35
7	6	5	36	25	30
8	8	6	64	36	48
9	9	7	81	49	56
10	10	8	100	64	80
$\sum$	50	50	340	290	296

จากตารางข้างบน

$$\sum X = 50$$

$$\sum Y = 50$$
$$\sum X^2 = 340$$
$$\sum Y^2 = 290$$
$$\sum XY = 296$$
$$N = 10$$

แทนค่าในสูตร

$$r_{xy} = \frac{N\sum XY - \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{10\,(296) - (50)\,(50)}{\sqrt{[10\,(340) - (50)^2][10\,(290) - (50)^2]}}$$

$$r_{xy} = \frac{2960 - 2500}{\sqrt{(3400-2500)\,(2900-2500)}}$$

$$= \frac{460}{\sqrt{360000}}$$

$$= \frac{460}{600}$$

$$r_{xy} = .77$$

เมื่อ $r_{xy} = .77$ แสดงว่าคะแนนคณิตศาสตร์และคะแนนวิทยาศาสตร์มีความสัมพันธ์ค่อนข้างสูงในทางบวก

ตัวอย่าง 11.2 ให้หาค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลซึ่งเป็นน้ำหนักและส่วนสูงของสิ่งของบางอย่าง ดังตาราง

น้ำหนัก X	12	10	14	10	13	9	15
ส่วนสูง Y	18	15	20	18	20	13	19

วิธีทำ

คำนวณหาค่า $\sum X$, $\sum Y$, $\sum X^2$, $\sum Y^2$ และ $\sum XY$ ดังนี้

เลขที่	น้ำหนัก	ส่วนสูง	X^2	Y^2	XY
1	12	18	144	324	216
2	10	15	100	225	150
3	14	20	196	400	280
4	10	18	100	324	180
5	13	20	169	400	260
6	9	13	81	169	117
7	15	19	225	361	285
$\sum$	83	123	1015	2203	1488

จากตารางข้างบน $\sum X = 83$, $\sum Y = 123$,
$\sum X^2 = 105$, $\sum Y^2 = 2203$ $\sum XY = 1488$, $N = 7$

แทนค่าในสูตร

$$r_{xy} = \frac{7\,(1488)-(83)\,(123)}{\sqrt{[7\,(1015)-(83)^2]\,[7\,(2203)-(123)^2]}}$$

$$= \frac{10416-10209}{\sqrt{(7105-6889)\,(15414-15129)}}$$

$$= \frac{207}{\sqrt{(216)\,(285)}}$$

$$= \frac{207}{248.11}$$

$$r_{xy} = .83$$

แสดงว่าน้ำหนักและส่วนสูงของสิ่งของบางอย่างมีความสัมพันธ์กันสูงมากในทางบวกซึ่งความสัมพันธ์นี้มีลักษณะเกือบเป็นเส้นตรง

11.2.2 วิธีหาค่าสัมประสิทธิ์สัมพันธ์แบบสเปียร์แมน

ในกรณีที่ตัวแปร X และ Y เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง (นั่นคือเป็นผลที่ได้จากการวัดในมาตราจัดอันดับ) เราสามารถหาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปร X และ Y จากสูตรของ Spearman ดังนี้

$$r_{xy} = 1-\frac{6\sum d^2}{N^3-N} \qquad \text{.......(11.2)}$$

เมื่อ N = จำนวนกลุ่มตัวอย่าง

d = ความแตกต่างของอันดับของ X และ Y

ตัวอย่าง 11.3 ครู 2 คน ตรวจคะแนนวิชาเรียงความของนักเรียน 5 คน โดยกำหนดให้ผลงานที่ดีที่สุดให้เป็นอันดับ 1 และอันดับที่ 5 คือ ผลงานที่แย่ที่สุด ผลของการจัดอันดับของครูแสดงไว้ในตารางข้างล่าง จงหาค่าสัมประสิทธิ์สหสัมพันธ์ของการจัดอันดับของครูทั้งสองคนนี้

นักเรียน คนที่	อันดับของครู คนที่ 1 (X)	อันดับของครู คนที่ 2 (Y)	d	d^2
1	1	3	−2	4
2	2	1	1	1
3	3	4	−1	1
4	4	2	2	4
5	5	5	0	0

วิธีทำ

$$\sum d^2 = 10,\ N = 5$$

แทนค่าในสูตร

$$r_{xy} = 1-\frac{6\sum d^2}{N^3-N}$$

$$= 1-\frac{6(10)}{5^3-5}$$

$$= 1-\frac{60}{125-5}$$

$$r_{xy} = .50$$

แสดงว่าการตรวจให้คะแนนวิชาเรียงความของครู 2 คน มีความสัมพันธ์กันในทางบวก แต่ไม่สูงมากนัก

เนื้อหาที่ 11.3 การทดสอบความมีนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์

ในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ (r_{xy}) นั้นในบางครั้งเราต้องการทดสอบดูว่าค่า r_{xy} นั้นมีนัยสำคัญทางสถิติหรือไม่ นั่นคือตัวแปร X และ Y มีความสัมพันธ์กันอย่างเชื่อมั่นได้หรือไม่ เราสามารถทำการทดสอบความมีนัยสำคัญของ r_{xy} ได้จากสูตร 11.3 ดังนี้

$$t = r_{xy}\sqrt{\frac{N-2}{1-r^2_{xy}}},\ df = N-2 \qquad ....(11.3)$$

N = จำนวนกลุ่มตัวอย่าง

11.3.1 การเขียนสมมุติฐาน ในการทดสอบความมีนัยสำคัญของ r_{xy}

ในการทดสอบความมีนัยสำคัญของ r_{xy} นั้นเราสามารถตั้งสมมุติฐานเผื่อเลือก (H_1) ได้ 3 แบบ เช่นเดียวกับการทดสอบแบบ t และ ρ (โร) คือค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร ในขณะที่ r_{xy} คือค่าสัมประสิทธิ์สหสัมพันธ์ของกลุ่มตัวอย่าง สมมุติฐานกลางและสมมุติฐานเผื่อเลือก เขียนได้ดังนี้

1) $H_O : \rho = 0$
   $H_1 : \rho \neq 0$

หรือ 2) $H_O : \rho = 0$
   $H_1 : \rho > 0$

หรือ 3) $H_O : \rho = 0$
   $H_1 : \rho \angle 0$

การตั้งสมมุติฐานดังกล่าว หมายความว่า ถ้าเราปฏิเสธสมมุติฐานกลาง แสดงว่าตัวแปร X และ Y มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ นั่นคือตัวแปรทั้งสองมีความสัมพันธ์กันอย่างเชื่อถือได้ที่ระดับความเชื่อมั่น (∞) (100) %

11.3.2 กำหนดขอบเขตวิกฤต

ใช้หลักการเดียวกันกับการกำหนดขอบเขตวิกฤตของการทดสอบแบบที (t-test) ในบทที่ 10 โดยกำหนดขอบเขตวิกฤตไว้เป็นหลักในการพิจารณา ดังนี้

1) ในกรณีที่ได้มีการกำหนดสมมุติฐานเผื่อเลือกไว้เป็นแบบมีทิศทาง ขอบเขตวิกฤตหรือขอบเขตที่จะใช้เป็นหลักในการพิจารณาว่าจะปฏิเสธสมมุติฐานกลาง ($H_o : \rho = 0$) หรือไม่ ดังนี้

ก. ถ้าสมมุติฐานเผื่อเลือกตั้งไว้ว่า $H_1 : \rho > 0$ เราจะปฏิเสธสมมุติฐานกลางถ้าค่า t ที่คำนวณได้จากสูตร 11.3 มีค่า **มากกว่า** $t_{(1-\infty, df)}$

ข. ถ้าสมมุติฐานเผื่อเลือกตั้งไว้ว่า $H : \rho < 0$ เราจะปฏิเสธสมมุติฐานกลางถ้า ค่า t ที่คำนวณได้จากสูตร 11.3 มีค่าน้อยกว่า $-t_{(1-\infty, df)}$

2) ในกรณีที่สมมุติฐานเผื่อเลือกกำหนดไว้อย่างไม่มีทิศทาง นั่นคือ $H_1 : \mu \neq \mu_o$ ขอบเขตวิกฤตกำหนดไว้ว่า เราจะปฏิเสธสมมุติฐานกลาง ($H_o : \rho = 0$) ถ้า t ที่คำนวณได้จากสูตร 11.3 มีค่า **มากกว่า** $t_{(1-\infty/2,\ df)}$ หรือ **น้อยกว่า** $-t_{(1-\infty/2,\ df)}$

ตัวอย่าง 11.4 จากตัวอย่างที่ 1.3 อยากทราบว่าการจัดอันดับของครู 2 คน มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติหรือไม่ที่ระดับความเชื่อมั่น 95% ($\infty = .05$)

วิธีทำ

$\infty = .05$, $N = 5$, $df = N-2 = 3$, $r_{xy} = .50$

$$H_O : \rho = 0$$
$$H_1 : \rho \neq 0$$

$$t = r_{xy}\sqrt{\frac{N-2}{1-r^2_{xy}}}$$
$$= .5\sqrt{\frac{5-2}{1-(.5)^2}}$$
$$= .5\sqrt{\frac{3}{.75}}$$
$$= 1.0$$

จากตาราง $t_{(1-\alpha/2;\,df)} = t_{.975,\,3}$ ซึ่งมีค่าเท่ากับ 3.182 เนื่องจากค่า t ที่คำนวณได้มีค่าน้อยกว่า $t_{.975,\,3}$ เราจึงไม่ปฏิเสธสมมุติฐานกลาง แสดงว่าการจัดอันดับของครู 2 คน มีความสัมพันธ์กันอย่างไม่มีนัยสำคัญทางสถิติ

ในกรณีที่ผู้ทำวิจัยไม่ต้องการทดสอบความมีนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์โดยใช้สูตร 11.3 ก็อาจตรวจสอบความมีนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์โดยเทียบค่าจากตารางสำเร็จที่ 11.1 ข้างล่างนี้ หรือตรวจเทียบค่าจากตาราง 5 ในภาคผนวกก็ได้เช่นกัน การตรวจสอบค่าความมีนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์จากตาราง 11.1 มีวิธีการดังนี้

ตาราง 11.1 ค่าสัมประสิทธิ์สหสัมพันธ์ที่ระดับนัยสำคัญทางสถิติที่ระดับ .05 และ .01

Degrees of freedom $(N-2)$	.05	.01	Degrees of freedom $(N-2)$	.05	.01
1	.997	1.000	24	.388	.496
2	950	.990	25	.381	.487
3	.878	.959	26	.374	.478
4	.811	.917	27	.367	.470
5	.754	.874	28	.361	.463
6	.707	.834	29	.355	.456
7	.666	.798	30	.349	.449
8	.632	.765	35	.325	.418
9	.602	.735	40	.304	.393
10	.576	.708	45	.288	.372
11	.553	.684	50	.273	.354
12	.532	.661	60	.250	.325
13	.514	.641	70	.232	.302
14	.497	.623	80	.217	.283
15	.482	.606	90	.205	.267
16	.468	.590	100	.195	.254
17	.456	.575	125	.174	.228
18	.444	.561	150	.159	.208
19	.433	.549	200	.138	.181
20	.423	.537	300	.113	.148
21	.413	.526	400	.098	.128
22	.404	.515	500	.088	.115
23	.396	.505	1000	.062	.081

(ที่มา : Garrett, H.E. Statistics in Psychology and Education, Vakils, Feffer and Simons Private Ltd., 1967, p.201.)

จากตัวอย่างที่ 11.3 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการจัดอันดับคะแนนวิชาเรียงความของครู 2 คนมีค่าเท่ากับ 0.50 อยากทราบว่าคะแนนการจัดอันดับของครู 2 คนมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติหรือไม่

จากตาราง 11.1 เมื่อ N = 5 ค่าสัมประสิทธิ์สหสัมพันธ์จะมีนัยสำคัญที่ระดับ 0.05 ได้ต้องมีค่า r ตั้งแต่ 0.878 ขึ้นไปและจะมีนัยสำคัญที่ระดับ 0.01 ได้ ต้องมีค่า r ตั้งแต่ 0.959 ขึ้นไป (N = 5, ดังนั้น df = 5-2 = 3) แต่จากตัวอย่างที่ 11.3 ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.5 (r = 0.5) ซึ่งน้อยกว่า 0.878 แสดงว่าความสัมพันธ์นี้ **ไม่มี**นัยสำคัญทางสถิติ ซึ่งก็ได้ผลเช่นเดียวกับการทดสอบ โดยใช้สูตร 11.3 เช่นเดียวกัน

ในกรณีที่ผู้ทำวิจัยต้องการตรวจสอบ ความมีนัยสำคัญทางสถิติของตัวแปร 2 ตัว ที่ระดับ .02 และ .10 ก็อาจตรวจสอบได้ โดยดูจากตาราง 5 ในภาคผนวก

เนื้อหาที่ 11.4 สมการถดถอยเชิงเส้นตรง

สมการถดถอยเชิงเส้นตรงเป็นสมการแสดงความสัมพันธ์ของตัวแปร X และ Y อีกแบบหนึ่ง ถ้า X และ Y มีความสัมพันธ์กันแบบเส้นตรง เราอาจเขียนแสดงความสัมพันธ์ระหว่าง X กับ Y ในรูปของสมการดังนี้

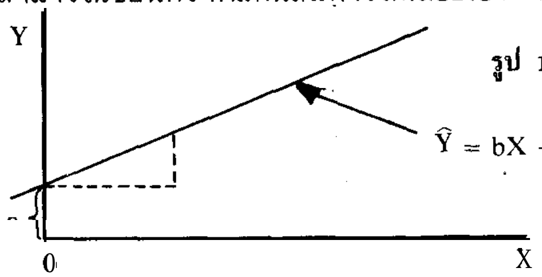
$$\hat{Y} = bx + a$$

เมื่อ b คือ slope ของสมการ

a คือ ความยาวจากจุด (0, 0) ถึงเส้นที่ตัดกับแกน Y

$\hat{Y}$ (อ่านว่า วายแฮ็ท) คือค่าของตัวแปร Y ที่ทำนายจากค่าของตัวแปร X

จากสมการข้างบนเมื่อเราทราบค่าของตัวแปร X ก็สามารถจะทำนายค่าของตัวแปร Y ได้ทันที สมการถดถอยเชิงเส้นตรงนี้เป็นสมการที่สร้างขึ้นมาจากค่าของตัวแปร X และ Y ที่ได้จากกลุ่มตัวอย่างจำนวนมาก แล้วใช้วิธีการทางสถิติขั้นสูงที่เรียกว่า Method of least squares ปรับให้เป็นสมการมาตรฐานเพื่อใช้พยากรณ์ค่าของตัวแปร Y ในอนาคตเราสามารถเขียนกราฟแทนสมการถดถอยเชิงเส้นตรงได้ดังรูป 11.8



รูป 11.8 *กราฟแสดงความสัมพันธ์ของตัวแปร X และ Y*

ในการสร้างสมการถดถอยเชิงเส้นตรงขึ้นมาเพื่อให้พยากรณ์ค่าของ Y ในอนาคตนั้น จะต้องคำนวณหาค่าของ b และ a เสียก่อน โดยหาได้จากสูตร

$$b = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} \qquad ....(11.4)$$

และ $a = \overline{Y} - b\overline{x}$ ....(11.5)

ตัวอย่าง 11.5 จงสร้างสมการถดถอยเชิงเส้นตรงจากข้อมูลข้างล่างนี้

$\overline{X}$	=	60.83	$\sum X$	=	730
$\overline{Y}$	=	84.75	$\sum Y$	=	1017
N	=	12	$\sum X^2$	=	44932
			$\sum XY$	=	62352

วิธีทำ

คำนวณหาค่า b และ a จากสูตร

$$b = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2}$$

$$b = \frac{12\,(62352) - (730)\,(1017)}{12\,(44932) - (730)^2}$$

$$= \frac{748224 - 742410}{539184 - 532900}$$

$$= \frac{5814}{6284}$$

$$b = .93$$

$$a = \overline{Y} - b\overline{x}$$

$$= 84.75 - (.93)\,(60.83)$$

$$= 28.18$$

ดังนั้นสมการถดถอย คือ

$$\hat{Y} = .93\,X + 28.18$$

ตัวอย่าง .11.6 ผลการสอบคัดเลือกกับคะแนนการเรียนหมวดวิชาคณิตศาสตร์ของนักเรียน 10 คน แสดงไว้ในตารางข้างล่าง จงหา

1) สมการถดถอย

ถ้านักเรียนคนหนึ่งสอบคัดเลือกได้คะแนน 60 ให้พยากรณ์ว่าเขาจะได้คะแนนคณิตศาสตร์เท่าไร

นักเรียนคนที่	1	2	3	4	5	6	7	8	9	10
คะแนนสอบคัดเลือก (X)	55	61	65	63	50	71	70	67	53	70
คะแนนคณิตศาสตร์ (Y)	62	70	68	67	62	80	75	72	53	75

วิธีทำ สร้างตารางเพื่อหาค่า $\sum X$, $\sum Y$, $\sum XY$ และ $\sum X^2$

นักเรียน	คะแนนสอบคัดเลือก X	คะแนนคณิตศาสตร์ Y	XY	X^2
1	55	62	3410	3025
2	61	70	4270	3721
3	65	68	4420	4225
4	63	67	4221	3969
5	50	62	3100	2500
6	71	80	5680	5041
7	70	75	5250	4900
8	67	72	4824	4489
9	53	53	2809	2809
10	70	75	5250	4900
ผลรวม ($\sum$)	625	684	43234	39579

จากสูตร

$$b = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2}$$

$$= \frac{10\,(43234) - (625)\,(684)}{10\,(39579) - (625)^2}$$

$$= \frac{432340 - 427500}{395790 - 390625}$$

$$= \frac{4840}{5165}$$

$$= .94$$

จากสูตร $a = \overline{Y} - b\overline{X}$

$$\overline{Y} = \frac{684}{10} = 68.4$$

$$\overline{X} = \frac{625}{10} = 62.5$$

แทนค่าในสูตร

$$a = 68.40 - (.94)\,(62.5)$$

$$= 68.40 - 58.75$$

$$= 9.65$$

ฉะนั้นสมการถดถอยคือ

$$\hat{Y} = .94\,X + 9.65$$

2) ถ้านักเรียนคนหนึ่งสอบคัดเลือกได้คะแนน 60 ให้พยากรณ์ว่าเขาจะได้คะแนนคณิตศาสตร์เท่าไร

จากสมการพยากรณ์

$$\hat{Y} = .94X + 9.65$$

ถ้า $X = 60$

$$\hat{Y} = (.94)(60) + 9.65$$

$$= 66.05$$

$\cong$ 66 คะแนน (โดยประมาณ)

เพราะฉะนั้น เขาจะได้คะแนนคณิตศาสตร์ 66 คะแนน

สรุปเนื้อหาบทที่ 11

1. ค่าแสดงความสัมพันธ์ระหว่างตัวแปร X และ Y เขียนแทนด้วยสัญลักษณ์ r_{xy} ค่า r_{xy} จะมีค่าอยู่ระหว่าง + 1.00 ถึง − 1.00

2. ถ้า $r_{xy} = +1.00$ แสดงว่าตัวแปร X และ Y มีความสัมพันธ์อย่างใกล้ชิด นั่นคือถ้า X มีค่ามาก Y ก็จะมีค่ามากตามไปด้วย และถ้า X มีค่าน้อย Y ก็จะมีค่าน้อยด้วยเช่นกัน

3. ถ้า $r_{xy} = -1.00$ แสดงว่าตัวแปร X และ Y มีความสัมพันธ์ในทางกลับกัน นั่นคือถ้า X มีค่ามาก Y จะมีค่าน้อย แต่ถ้า X มีค่าน้อย Y จะมีค่ามาก

4. ถ้า $r_{xy} = 0.00$ แสดงว่าตัวแปร X และ Y **ไม่มี**ความสัมพันธ์กันในเชิงเส้นตรง

5. การหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร 2 ตัว (r_{xy}) อาจหาได้จากสูตรใดสูตรหนึ่งข้างล่างนี้ แล้วแต่ลักษณะข้อมูล คือ

   5.1 ถ้าตัวแปร X และ Y เป็นตัวแปรสุ่มแบบต่อเนื่อง ให้หาค่า r_{xy} จากสูตร

$$r_{xy} = \frac{N\Sigma XY - \Sigma X\Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

   5.2 ถ้าตัวแปร X และ Y เป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง (นั่นคืออยู่ในมาตราจัดอันดับ) ให้หาค่า r_{xy} จากสูตร

$$r_{xy} = 1 - \frac{6\sum d^2}{N^3 - N}$$

6. การทดสอบความมีนัยสำคัญทางสถิติของค่า r_{xy} ให้ใช้สูตร

$$t = r_{xy}\sqrt{\frac{N-2}{1-r^2_{xy}}} \quad , df = N-2$$

7. ความสัมพันธ์เชิงเส้นตรงของตัวแปร X และ Y อาจแสดงให้เห็นได้โดยใช้สมการถดถอยเชิงเส้นตรงดังนี้ : $\hat{Y} = bX + a$

คำถามท้ายบทที่ 11

1. จงหาค่าสหสัมพันธ์ของคะแนน 2 ชุดข้างล่างนี้

นักเรียนคนที่	X	Y
1	15	40
2	18	42
3	22	50
4	17	45
5	19	43
6	20	46
7	16	41
8	21	41

2. ผลการทดสอบวิชาคณิตศาสตร์และวิทยาศาสตร์ของนักเรียน 12 คน แสดงในตารางข้างล่าง จงหาค่าสหสัมพันธ์ของคะแนน 2 ชุดนี้

นักเรียนคนที่	คะแนนคณิตศาสตร์	คะแนนวิทยาศาสตร์
1	50	22
2	54	25
3	56	34
4	59	28
5	60	26
6	62	30
7	61	32
8	65	30
9	67	28
10	71	34
11	71	36
12	74	40

3. สมมุติว่าทำการทดสอบข้อสอบ 4 ฉบับ คือ A, B, C และ D กับนักเรียน 5 คนและนำผลของการทดสอบข้อสอบฉบับ A ไปเปรียบเทียบกับข้อสอบฉบับ B, C และ D ผลของคะแนนของแบบทดสอบ 3 คู่ แสดงไว้ในตารางข้างล่าง จงหาค่าสหสัมพันธ์ของข้อสอบ 3 คู่ ที่กล่าวมาแล้วด้วยวิธี Spearman rank correlation Coefficient

	กรณีที่ 1		กรณีที่ 2		กรณีที่ 3	
นักเรียนคนที่	A	B	A	C	A	D
1	20	102	20	50	20	67
2	18	100	18	54	18	60
3	17	104	17	55	17	59
4	14	103	14	58	14	55
5	12	101	12	60	12	53

4. จงสร้าง scattergram จากตาราง 2 ทางข้างล่างนี้ และวิเคราะห์ว่าค่าสหสัมพันธ์ควรจะเป็นไปในทางบวก หรือทางลบ

คะแนนคณิตศาสตร์ (X)

คะแนนการอ่าน (Y)	12-13	14-15	16-17	18-19	20-21	22-23	24-25
35-37					1		1
32-34					6	3	
29-31		1	2	6	8	1	
26-28		4	4	6	11	4	1
23-25	2	1	6	5	4	1	
20-22	3	2	1	1			

5. จากข้อมูลดังตาราง

X	12	11	15	11	12	8
Y	19	17	24	19	20	14

ก. จงหาค่าสหสัมพันธ์

ข. เขียน scrattergram

ค. สร้างสมการถดถอยเชิงเส้น

6. คะแนนสอบกลางเทอม (X) และคะแนนสอบปลายเทอม (Y) ของนักศึกษา 10 คนเป็นดังนี้

X	50	75	49	71	70	81	94	90	66
Y	52	80	66	78	35	46	85	90	67

ก. จงสร้างสมการถดถอยเชิงเส้น เพื่อทำนายคะแนนสอบปลายเทอม จากคะแนนสอบกลางเทอม

ข. ถ้านักศึกษาคนหนึ่งสอบได้คะแนน 83 แต่ไม่ได้สอบปลายเทอม จงประมาณว่าเขาจะได้คะแนนปลายเทอมเท่าไร

7. จากข้อ 1. จงทดสอบว่าคะแนน 2 ชุดมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติหรือไม่ที่ระดับความเชื่อมั่น .05

8. จงทดสอบความมีนัยสำคัญทางสถิติของค่าสหสัมพันธ์ในข้อ 2 ถ้ากำหนดสมมุติฐานเพื่อเลือกว่า $\rho > O$ ที่ระดับความเชื่อมั่น .01

9. จงทดสอบความมีนัยสำคัญทางสถิติของค่าสหสัมพันธ์ในข้อ 6 ถ้ากำหนดสมมุติฐานเพื่อเลือกว่า $\rho \neq O$ ที่ระดับความเชื่อมั่น .05