บทที่ 10
สหสัมพันธ์

การวัดสหสัมพันธ์จะต้องบรรยายถึงระดับของความสัมพันธ์ระหว่าง 2 ตัวแปร นักวิจัยมักสนใจความสัมพันธ์ระหว่างกลุ่มบุคคล โรงเรียน ชุมชน ตัวอย่างเช่น การขาดเรียนสัมพันธ์กับสถานภาพทางเศรษฐกิจหรือไม่ ขนาดของห้องเรียนสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนหรือไม่ เพื่อตอบคำถามเหล่านี้ ก็ต้องวัดระดับความสัมพันธ์หรือสหสัมพันธ์

คนส่วนมากเข้าใจว่าสหสัมพันธ์ ต้องเกิดจากตัวแปร 2 ตัว ที่ไปด้วยกัน ถ้า X มีค่ามากก็จะทำให้ Y มีค่ามากไปด้วย จึงจะเรียกว่า X และ Y มีความสัมพันธ์กัน แท้ที่จริงแล้วระดับของความสัมพันธ์ ระหว่าง 2 ตัวแปร สามารถบรรยายได้ถึง ความสัมพันธ์ที่สูง ต่ำ หรือ กลางๆ โดยมีทิศทางของความสัมพันธ์ด้วย

1. สัมประสิทธิ์สหสัมพันธ์ของ Pearson

ในหัวข้อนี้จะแนะนำสูตรสำหรับคำนวณสัมประสิทธิ์สหสัมพันธ์ของ Pearson เมื่อนักวิจัยพูดถึงสัมประสิทธิ์สหสัมพันธ์โดยไม่จำเพาะเจาะจงในสิ่งที่เขาหมายถึงเราอาจตั้งสมมุติฐานว่าเขาอ้างอิงถึงสหสัมพันธ์ของเพียร์สัน ซึ่งมันคือ product-moment correlation coefficient

สหสัมพันธ์แสดงความสัมพันธ์ระหว่างคู่ของคะแนน ในรูปที่ง่ายที่สุด สัมประสิทธิ์สหสัมพันธ์ของเพียร์สันนิยามดังนี้

สูตรคะแนน z สำหรับสัมประสิทธิ์สหสัมพันธ์

$$r = \frac{\sum(z_x z_y)}{n} \qquad \text{..........(10.1)}$$

เมื่อ n เป็นจำนวนของคู่ของคะแนน ในการหา r เราต้องเปลี่ยนแต่ละคะแนนดิบเป็นคะแนน z ดังนั้นสำหรับแต่ละคู่ของคะแนน z เราคูณคะแนน บวกผลลัพธ์ เรียก cross-product และหารด้วยจำนวนของคู่ของคะแนน ดังนั้น r เป็นค่าเฉลี่ยของผลคูณไขว้ cross-product ของคู่ของคะแนน z

สูตร 10.1 แสดงง่ายๆแต่ยังไม่สะดวกในการคำนวณ เราสามารถนิยามสัมประสิทธิ์ของเพียร์โดยตรงในพจน์ของคะแนนการเบี่ยงเบน

สูตรคะแนนการเบี่ยงเบนสำหรับสัมประสิทธิ์สหสัมพันธ์

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{n s_x s_y} \qquad \text{..........10.2)}$$

เมื่อ n เป็นจำนวนคู่ของคะแนน และ s_X, s_y เป็นส่วนเบี่ยงเบนมาตรฐานของสองกลุ่มตัวอย่าง

เราจะคำนวณ Rของเพียร์สันโดยใช้สูตรนี้เพราะว่าจะช่วยให้เราเห็นว่าองค์ประกอบอะไรทำให้สัมประสิทธิ์เป็นบวกหรือลบและองค์ประกอบอะไรทำให้ค่าสูงหรือต่ำ เครื่องคำนวณจำนวนมากมีสูตรคำนวณ r อย่างไรก็ตาม สำหรับผู้ที่ต้องคำนวณ r ด้วยมือ วิธีคำนวณจากคะแนนดิบทำได้โดยง่าย ดังตัวอย่างต่อไปนี้

ตัวอย่าง ผลการสอบวิชาคณิตศาสตร์ และวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 10 คนมีคะแนนดังนี้

ตาราง 10.1

คนที่	x	Y
1	15	7
2	6	11
3	6	5
4	18	16
5	12	9
6	5	0
7	19	3
8	20	13
9	17	24
10	28	26

จงหาสัมประสิทธิ์สหสัมพันธ์ r ของคะแนนคณิตศาสตร์และวิทยาศาสตร์

วิธีทำ คำนวณหา r โดยสูตร 10.2 ดังนี้

		(3)	(4)	(5)	(6)	(11)
X	Y	X - $\bar{X}$	Y - $\bar{Y}$	$(X-\bar{X})^2$	$(Y-Y)^2$	$(X-\bar{X})(Y-\bar{Y})$
15	7	0.4	-4.4	0.16	19.36	-1.76
6	11	-8.6	-0.4	73.96	0.16	3.44
6	5	-8.6	-6.4	73.96	40.96	55.04
18	16	3.4	4.6	11.56	21.16	15.64
12	9	-2.6	-2.4	6.76	5.76	6.24
5	0	-9.6	-11.4	92.16	129.96	109.44
19	3	4.4	-8.4	19.16	70.56	-36.96
20	13	5.4	1.6	29.16	2.56	8.64
17	24	2.4	12.6	5.76	158.76	30.24
28	26	13.4	14.6	79.56	213.16	195.64
146	114	0	0	492.40	662.40	385.60
				(7)	(8)	(12)

(1) $\bar{X} = \dfrac{146}{10} = 14.6$

(2) $\bar{Y} = \dfrac{114}{10} = 11.4$

(9) $s_x = \sqrt{\sum(X-\bar{X})^2/n} = \sqrt{490.40/10} = 7.017$

(10) $s_Y = \sqrt{\sum(Y-\bar{Y})^2/n} = \sqrt{662.40/10} = 8.193$

(13) r $= \sum(X-\bar{X})(Y-\bar{Y})/nS_XS_Y$ = 385.60/(10)(7.017)(8.139) = 0.675

ซึ่งสรุปเป็นขั้นตอนการหาได้ดังนี้

ขั้นที่ 1 แจงคู่ของคะแนนในสองสดมภ์

ขั้นที่ 2 หาค่าเฉลี่ยของX (แสดงในสดมภ์ 1ในตาราง) และค่าเฉลี่ยของ y (สดมภ์ 2)

ขั้นที่ 3 เปลี่ยนคะแนนในแต่ละแถวเป็นคะแนนการเปลี่ยนแปลง (3) และ (4)

ขั้นที่ 4 คำนวณส่วนเบี่ยงเบนมาตรฐานสำหรับ X (9) และ Y (10) เนื่องจากเรามีคะแนนส่วนเบี่ยงเบน ดังนั้นการคำนวณโดยคะแนนส่วนเบี่ยงเบนจึงเป็นวิธีที่ง่ายกว่านั่นคือต้องหากำลังสองของการเบี่ยงเบน (5) และ (6) และคำนวณผลบวกของกำลังสองของการเบี่ยงเบน (7) และ (8) หารด้วย n (ขนาดของกลุ่มตัวอย่าง) และใส่รากที่สอง ซึ่งจะได้ค่าต่าง ๆที่นำไปใช้คำนวณ r

ขั้นที่ 5 คำนวณตัวเศษของสูตร 10.2 โดยคูณคะแนนความเบี่ยงเบนของแต่ละคน (11) และหาผลบวกของผลคูณไขว้ (cross-product) (12)

ขั้นที่ 6 หาตัวส่วนของ 10.2 โดยหาผลคูณของ n s_X s_y

ขั้นที่ 7 หาผลหาร (13) จะได้ค่า r

การคำนวณ r จากคะแนนดิบ

สูตรคะแนน z และสูตรคะแนนความเบี่ยงเบนสำหรับ r เป็นเรื่องยุ่งยาก ถ้าจะคำนวณ r ด้วยมือ แต่จะเป็นเรื่องง่ายขึ้นที่จะแปลงสูตรคะแนนเป็นความเบี่ยงเบนเพื่อให้สะดวกในการคำนวณ

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{n s_x s_y} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{n\sqrt{\frac{\sum (X-\bar{X})^2}{n}}\sqrt{\frac{\sum (Y-\bar{Y})^2}{n}}}$$

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

ดังนั้น สูตรคำนวณหาสัมประสิทธิ์สหสัมพันธ์

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2 (Y-\bar{Y})^2}} \qquad \text{..........10.3)}$$

นิพจน์ในตัวส่วนเป็นผลบวกของกำลังสองของคะแนนความเบี่ยงเบนสำหรับ X และ Y โดยที่

$$SS_X = \sum(X-\bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$SS_Y = \sum(Y-\bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

ดังนั้นสูตร 10.3 อาจเขียนได้เป็น

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{SS_X SS_Y} \qquad \text{..........} (10.4)$$

สำหรับตัวเศษของสูตร 10.3 หรือ 10.4 $\sum(X-\bar{X})(Y-\bar{Y})$ เป็นผลบวกของผลคูณไขว้ของคะแนนความเบี่ยงเบนและสามารถแสดงในพจน์ของคะแนนดิบ

คะแนนดิบที่สมมูลกันของ $\sum(X-\bar{X})(Y-\bar{Y})$

$$\sum(X-\bar{X})(Y-\bar{Y}) = \sum XY - \frac{(\sum X)(\sum Y)}{n} \qquad \text{..........} (10.5)$$

เราสามารถคำนวณ r ของเพียร์สันจากคะแนนดิบ ในตาราง 10.2 โดยใช้ข้อมูลจากตาราง 10.1 อีกครั้ง ซึ่งสรุปขั้นตอนการทำดังต่อไปนี้

ขั้นที่ 1 ลงรายละเอียดเกี่ยวกับ n, $\sum X$, $\sum Y$, $\sum XY$, ผลลัพธ์แสดงดัง (1) ในตาราง

ขั้นที่ 2 หา SS_X, SS_y และ $\sum(X-\bar{X})(Y-\bar{Y})$ ดังแสดงที่ (2), (3) และ (4)

ขั้นที่ 3 แทนค่าสามค่าในขั้นที่ 2 ในสูตร 10.4 ตาม (5)

ข้อควรทราบ

จดจำว่า ข้อมูลที่ค่อนข้างจะเป็นเส้นตรงมากกว่าจะให้ค่า r สูงกว่า สำหรับข้อมูลซึ่งจุดต่าง ๆไม่ใกล้เคียงกับเส้นตรง ค่าสัมบูรณ์ ของ r จะมีค่าต่ำ หรือกล่าวได้ว่าความสัมพันธ์ระหว่าง X และ Y มีมาก การกระจายของจุดต่าง ๆ จะห่างจากเส้นไม่มาก

ตาราง 10.2 คำนวณค่า r จากคะแนนดิบ

	X	Y	X^2	Y^2	XY
	15	7	225	49	105
	6	11	36	121	66
	6	5	36	25	30
	18	16	324	256	288
	12	9	144	81	108
	5	0	25	0	0
	19	3	361	9	57
	20	13	400	169	260
	17	24	289	576	408
	28	26	784	676	728
ผลบวก	146	114	2624	1962	2050

(1)

(2) $SS_X = \sum X^2 - (\sum X)^2/n = 2624 - 146^2/10 = 2624 - 2131.6 = 492.4$

(3) $SS_Y = \sum Y^2 - (\sum Y)^2/n = 1962 - 114^2/10 = 1962 - 1299.6 = 662.4$

(4) $\sum (X-\bar{X})(Y-\bar{Y}) = \sum XY - \dfrac{(\sum X)(\sum Y)}{n}$

$= 2050 - \dfrac{(146)(114)}{10}$

$= 2050 - 1664.4$

$= 385.6$

(5) $r = \dfrac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{(SS_X)(SS_Y)}} = \dfrac{385.6}{\sqrt{(492.4)(662.4)}} = \dfrac{385.6}{571.11} = 0.675$

ปกติเราต้องการค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานด้วย

$$\bar{X} = \frac{\sum X}{n} = \frac{146}{10} = 14.6 \qquad \bar{Y} = \frac{\sum Y}{n} = \frac{114}{10} = 11.4$$

$$s_X = \sqrt{\frac{SS_X}{n}} = \sqrt{\frac{492.4}{10}} = 7.107 \qquad s_Y = \sqrt{\frac{SS_Y}{n}} = \sqrt{\frac{662.4}{10}} = 8.139$$

ข้อสังเกตทางคณิตศาสตร์

1. ความเบี่ยงเบนของสูตรคะแนนความเบี่ยงเบนสำหรับ r จากคะแนน z

$$r = \frac{\sum(z_X z_Y)}{n}$$

$$= \frac{\sum\left(\frac{X-\bar{X}}{s_X}\right)\left(\frac{Y-\bar{Y}}{s_Y}\right)}{n}$$

$$= \frac{\sum(X-\bar{X})(Y-\bar{Y})}{n s_X s_Y}$$

2. คะแนนดิบที่สมมูลกับ $\sum(X-\bar{X})(Y-\bar{Y})$

$$\sum(X-\bar{X})(Y-\bar{Y}) = \sum(XY - \bar{X}Y - \bar{Y}X + \bar{X}\bar{Y})$$

$$= \sum XY - \bar{X}\sum Y - \bar{Y}\sum X + n\bar{X}\bar{Y})$$

$$= \sum XY - \frac{(\sum X)(\sum Y)}{n} - \frac{(\sum Y)(\sum X)}{n} + \frac{n(\sum X)(\sum Y)}{(n)(n)}$$

$$= \sum XY - \frac{(\sum X)(\sum Y)}{n}$$

2. สัมประสิทธิ์สหสัมพันธ์ของ Spearman Rank

สหสัมพันธ์ Pearson ที่กล่าวมาแล้ว เป็นการหาสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อมูลของตัวแปรชุด X กับข้อมูลของตัวแปรชุด Y ที่ต่างก็เป็นตัวแปรที่เป็นข้อมูล

แบบต่อเนื่อง แต่ถ้าข้อมูลอยู่ในระดับอันดับ (ranks) หรือแปลงข้อมูลต่อเนื่องให้เป็นข้อมูลในระดับอันดับทั้ง 2 ชุด นั่นคือ X ก็เป็นข้อมูลของตัวแปรอันดับ Y ก็เป็นข้อมูลของตัวแปรอันดับ ชุดของตัวแปรทั้ง 2 นี้ ที่มีจำนวน n คน ก็สามารถคำนวนวณหาความสัมพันธ์ในแบบเดียวกับ Pearsonได้ โดยพัฒนามาจากสูตรดั้งเดิมของ Pearson

ตาราง 10.3 แสดงจำนวนค่าสหสัมพันธ์ Spearman Rank ระหว่างจำนวนอัตราความเร็วที่จำกัดบนถนนทางด่วน กับอัตราการตายบนเส้นทางจราจรของประเทศต่าง ๆ 10 ประเทศ

ประเทศ	ความเร็วจำกัด	อัตราการตาย	จัดอันดับ		ผลต่างในการจัดอันดับ	
			X	Y	D	D^2
West Germany	none	7.9	1	8	7	49
Italy	87 mph	6.4	2	9	7	49
France	81	8.0	3	7	4	16
Hungary	75 [a]	14.5	5 [a]	3	2	4
Belgium	75	10.5	5	6	1	1
Portugal	75	22.5	5	2	3	9
Britain	70	4.5	7	14	7	49
Spain	62	12.4	10	5	5	25
Denmark	62	4.8	10	11	1	1
Netherlands	62	6.0	10	10	0	0
Greece	62	12.9	10	4	6	36
Japan	62	4.7	10	12	2	4
Norway	56	4.2	13.5	13	.5	.25
Turkey	56	32.2	13.5	1	12.5	156.25
United States	55	3.3	15	15	0	0
					$\sum D^2 = 399.5$	

[a] When ties occur on X or Y,assign the average of the ranks involved to each score,e.g.,(4+5+6)/3 = 5 for X = 75. (ที่มา Glass & Hopkins,1984 P 96)

$$r = \frac{\ddot{E}_{xy}}{s_x s_y}$$

$$r = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\left(\sum X_i^2 - n\overline{X}^2\right)\left(\sum Y_i^2 - n\overline{Y}^2\right)}$$

มาเป็น $r_{ranks} = 1 - \frac{6\sum D_i^2}{n(n^2-1)}$

$$r_{ranks} = 1 - \frac{6(399.5)}{15(225-1)}$$

$$= 1 - \frac{2{,}397}{3{,}360}$$

$$= 1 - .7134$$

$$= .2866 \text{ หรือ } .29$$

เมื่อ D คือ ค่าความแตกต่างระหว่างข้อมูลชุด X กับ ชุด Y เป็นรายบุคคล โดยที่ X และ Y ถูกจัดไว้เป็นลำดับในชุดของ X และในชุดของ Y ดังตัวอย่าง ในตาราง 10.3

3. สัมประสิทธิ์สหสัมพันธ์ Phi ทั้ง x และ Y เป็นตัวแปรคู่ (Dichotomous variables)

เมื่อ X และ Y ต่างก็เป็นตัวแปรคู่ คือ มีค่าเพียง 2 ค่า หรือ 2 ระดับ แล้วกำหนดให้มีลักษณ์เป็น 0 กับ 1 เช่น X เป็นตัวแปรบอกการเป็นโรคไข้หวัด 2009 ก็กำหนดลักษณ์ 1 คือ เป็นโรค 0 คือ ไม่เป็นโรค และY เป็นตัวแปรบอกเพศ ก็กำหนดให้ 1 เป็นเพศชาย 0 เป็นเพศหญิง

ตัวอย่างต่อไปนี้แสดงการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ Phi (r_ϕ) ระหว่างตัวแปร X ที่มีค่าเป็น 1 หมายถึง “แต่งงานแล้ว” และ 0 หมายถึง “ยังไม่แต่งงาน” กับตัวแปร Y ที่มีค่าเป็น 1 หมายถึง "ออกกลางคัน" และ 0 หมายถึง .ยังคงอยู่ในโรงเรียน" การ

คำนวณความสัมพันธ์ระหว่างXและYนี้ จะใช้หลักการง่ายๆของ Pearson ดังแสดงในตาราง 10.4ตาราง 10.4 แสดงการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ ϕ

นักเรียนคนที่	X สถานะภาพ (แต่งงานแล้ว = 1; (ยังไม่แต่งงาน = 0)	Y (ออกกลางคัน =1; (ยังอยู่ในโรงเรียน = 0	การคำนวณ
1	0	0	$p_x = \frac{4}{10} = .4 \quad q_x = \frac{6}{10} = .6$
2	1	1	
3	0	1	$p_y = \frac{5}{10} = .5 \quad q_y = 0.5$
4	0	0	
5	1	1	
6	1	0	$p_{xy} = \frac{3}{10} = .3$
7	0	0	
8	1	1	$r_{\phi} = \frac{p_{xy} - p_x p_y}{\sqrt{p_x q_x p_y q_y}}$
9	0	0	
10	0	1	
	$\sum X = 4$	$\sum Y = 5$	$r_{\phi} = \frac{.3 - (.4)(.5)}{\sqrt{(.4)(.6)(.5)(.5)}} = .408$

$$r_{\phi} = \frac{p_{xy} - p_x p_y}{\sqrt{p_x q_x p_y q_y}}$$

และเมื่อแปลงสูตร r_{ϕ} ข้างต้นนี้ ให้เป็นสัญลักษณ์แทนจำนวนของค่า 1 และ 0 ของทั้ง 2 ตัวแปร X และ Y ในตาราง (Contingency Table) ดังตัวอย่างตาราง 10.5

ตาราง 10.5 ตารางแสดงถึงความสัมพันธ์ระหว่าง เพศ กับการรับประทานอาหารเช้า

X

	อาหารเช้า?	เพศ: เพศหญิง(0)	เพศ: เพศชาย(1)	ผลรวม
Y	ทานอาหารเช้า (1)	8 = a	6 = b	14 = a+b
	ไม่ทานอาหารเช้า (0)	9 = c	4 = d	13 = c+d
	ผลรวม	17 = a+c	10 = b+d	27 = n

จำนวนคนที่เป็นเพศหญิง (X=0) และรับประทานอาหารเช้า (Y=1) มี a คน และจำนวนคนที่เป็นเพศหญิง (X=0) ทั้งหมดมี (a+e)คน โดยจำนวนคนทั้งหมดเป็น n

เมื่อแทนค่า P_X = (b+d)/n

P_Y = (a+b)/n

P_{XY} = b/n

$$r_{XY} = \frac{bc - ad}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

$$= \frac{(6)(9)-(8)(4)}{\sqrt{(17)(10)(14)(13)}}$$

$$= \frac{22}{\sqrt{30{,}940}} = .125$$

4. สัมประสิทธิ์สหสัมพันธ์ Point-Biserial

เมื่อตัวแปรคู่ (one dichotomous variable) เช่น เพศ สอบได้สอบตก สัมพันธ์กับตัวแปรต่อเนื่อง (continuous variable) สูตรการหาค่าสัมประสิทธิ์สหสัมพันธ์ของ Pearson ก็สามารถขยายไปสู่การคำนวณหาค่าสหสัมพันธ์ที่เรียกว่า สัมประสิทธิ์ Point-Biserial ใช้สัญลักษณ์ r_{pb} ซึ่งมีสูตรดังนี้

$$r_{pb} = \frac{\overline{Y}_1 - \overline{Y}_0}{s_Y}\sqrt{\frac{n_1 n_0}{n(n-1)}}$$

เมื่อตัวแปรคู่ (X) มี 2 ลักษณะ คือ ลักษณะที่เป็น 0 กับลักษณะที่เป็น 1 ดังนั้นจึงสามารถหาค่าเฉลี่ย Y ซึ่งเป็นตัวแปรต่อเนื่องได้ 2 กลุ่ม คือ $\overline{Y}_1$ คือค่าเฉลี่ยของตัวแปร Y ในกลุ่มที่มี X เป็น 1 และ $\overline{Y}_0$ คือค่าเฉลี่ยของตัวแปร Y ในกลุ่มที่มี X เป็น 0 โดยที่ n_1 คือจำนวนคนที่มีค่า X เป็น 1 และ n_0 จำนวนคนที่มีค่า X เป็น 0

ตัวอย่าง หาความสัมพันธ์ระหว่างคะแนนผลสัมฤทธิ์ทางการเรียนกับเพศของนักเรียน โดยที่นักเรียนชาย (X=1) มีค่าเฉลี่ยของผลสัมฤทธิ์ทางการเรียน ($\overline{Y}_1$) เท่ากับ 69.90 และนักเรียนหญิง (X=0) มีค่าค่าเฉลี่ยของผลสัมฤทธิ์ทางการเรียน ($\overline{Y}_0$) เท่ากับ 64.35 จำนวนนักเรียนชาย (n_1) เท่ากับ 10 คน จำนวนนักเรียนหญิง (n_0) เท่ากับ 7 คน และ $n = n_1 + n_0$ ตัวอย่างการคำนวณแสดงอยู่ในตาราง 10.6

ตาราง 10.6 แสดงการคำนวณค่าสหสัมพันธ์ Point-Biserial

เพศหญิง: $n_0 = 17$, $\overline{Y}_0 = 64.35$ in.

$s_Y = 3.75$ in. $n = 17+10 = 27$

เพศชาย: $n_1 = 10$, $\overline{Y}_1 = 69.90$ in.

$$r_{pb} = \frac{\overline{Y}_1 - \overline{Y}_0}{s_Y}\sqrt{\frac{n_1 n_0}{n(n-1)}} = \frac{69.90 - 64.35}{3.75}\sqrt{\frac{(10)(17)}{(27)(27-1)}} = (1.48)\sqrt{.2422} = .728$$

---

5. สัมประสิทธิ์ สหสัมพันธ์ Biserial

สมมุติว่า ตัวแปรคู่ (X) ได้มาจากข้อมูลที่มีการแจกแจงเป็น โค้งปกติ แล้วมาถูกจัดกระทำให้เป็นข้อมูล 2 พวก นั่นคือ มีการตั้งเกณฑ์จุดตัด แบ่งตัวแปร X ออกเป็น 2 พวก เช่น พวกที่ทำข้อสอบ IQ เสร็จทันเวลากับพวกที่ทำข้อสอบ IQ ไม่เสร็จในช่วงเวลาที่กำหนด ตัวแปรคู่ X นี้ ถูกกำหนดด้วยช่วงเวลา ส่วนตัวแปร Y ซึ่งเป็นตัวแปรต่อเนื่องคือคะแนน IQ โดยผู้วิจัย ได้ตั้งคำถามวิจัยว่า "ช่วงเวลาที่ทำข้อสอบมีความสัมพันธ์กับ

คะแนน IQ หรือไม่" ค่าสหสัมพันธ์ที่จะตอบคำถามนี้ คือ สูตรค่าสัมประสิทธิ์สหสัมพันธ์ Biserial

$$r_{bis} = \frac{\overline{Y}_1 - \overline{Y}_0}{S_Y}\left(\frac{n_1 n_2}{u n_\bullet^2}\right)$$

ตัวอย่างแสดงการคำนวณอยู่ในตาราง 10.7

ตาราง 10 7 แสดงการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ Biserial

ไม่เสร็จในช่วงที่กำหนด : $n_0 = 12$, $\overline{Y}_0 = 105.1$

$S_Y = 16.0$, $n_\bullet = 40$

เสร็จในช่วงที่กำหนด : $n_1 = 28$, $\overline{Y}_1 = 113.3$

$u = .3477$ (u จากตาราง A, จะได้ $p = n_0 / n_\bullet$ ใน สดมภ์ "พื้นที่ใต้โค้งของตาราง z ," และเลือก u ใน สดมภ์.ติดกัน)

$$r_{bis} = \frac{\overline{Y}_1 - \overline{Y}_0}{S_Y}\left(\frac{n_1 n_2}{u n_\bullet^2}\right) = \frac{113.3 - 105.1}{16.0}\left(\frac{(28)(12)}{(.3477)(40)^2}\right) = (.5125)(.6040) = .310$$

ค่า u คือค่า ordinate ที่เปิดจากตาราง A ในภาคผนวก ที่มีพื้นที่ใต้โค้งเป็น P ($P = \frac{n_0}{n_\bullet}$) ในที่นี้พื้นที่ใต้โค้งของตาราง z เป็น 0.300 ค่า ordinate u จึงเท่ากับ .3477

6. ความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์ Biserial กับ Point Biserial

บางครั้งก็เกิดการถกเถียงระหว่างนักวัดผลกับนักสถิติในการกำหนดค่า X ที่เป็นตัวแปรคู่ว่า ค่า X ที่มีค่าเป็น 0 และ 1 นั้นเป็น 0 ,1 แท้ (True dichotomous) หรือ ไม่เป็น 0 ,1 แท้ (Artificial dichotomous) ดังนั้นการตัดสินใจใช้สูตร สำหรับ Point Biserial (r_{pb}) หรือ Biserial (r_{bis}) จึงอยู่บนรากฐานแนวคิด อย่างไรก็ตามได้มีผู้หาความสัมพันธ์ของ 2 สูตรนี้ ได้เป็น

$$r_{bis} = r_{pb}\sqrt{\frac{n_1 n_0 (n_\bullet - 1)}{u^2 n_\bullet^3}}$$

การเปรียบเทียบ r_{bis} และ r_{pb} จากสมการข้างต้นนี้ เมื่อ ค่า $n_1 = n_0$ และ $n_{\square}$ มีค่าใหญ่มาก ทำให้ค่าที่เป็นรากที่ 2 ในสมการนี้จะมีค่า น้อยที่สุดประมาณ 1.25 ดังนั้น ค่า r_{bis} □1.25 r_{pb}

7. สัมประสิทธิ์สหสัมพันธ์ Tetrachoric Coefficient

เมื่อทั้งตัวแปร X และตัวแปร Y ต่างก็เป็น 0, 1 ไม่แท้ และต่างก็มีการแจกแจงของข้อมูลเป็นโค้งปกติ เช่น มีคน 400คน ให้ตอบคำถามเจตคติ 2 ข้อ โดยคำตอบแต่ละข้อเป็น “เห็นด้วย" หรือ “ไม่เห็นด้วย"

X : คุณเห็นด้วยหรือไม่กับการให้บุคคลเพศเดียวกันจดทะเบียนสมรสกันได้

Y : คุณเห็นด้วยหรือไม่กับการมีกฎหมายให้ทำแท้งได้อย่างเสรี

$$r_{tet} = \frac{bc - ad}{u_x u_y n_{\square}^2}$$

แสดงการคำนวณสัมประสิทธิ์ Tetrachoric Coefficient ในตาราง 10.8

ตาราง 10.8 แสดงการคำนวณสัมประสิทธิ์ Tetrachoric Coefficient

		คำถาม X			
		ไม่เห็นด้วย	เห็นด้วย	ผลรวม	สัดส่วน
คำถาม Y	เห็นด้วย	90 (a)	110 (b)	200	.50 (p_y)
	ไม่เห็นด้วย	190 (c)	10 (d)	200	.50 (q_y)
	ผลรวม	280	120	400 ($n_{\square}$)	
	สัดส่วน	.70 (p_x)	.30 (q_x)		

$$r_{tet} = \frac{bc - ad}{u_x u_y n_{\square}^2} = \frac{(110)(190) - (90)(10)}{(3477)(3989)(400)^2}$$

$$r_{tet} = .901$$

8. สัมประสิทธิ์สหสัมพันธ์ Rank Biserial

Cureton (1956) and Glass (1966b) ได้ศึกษาหาค่าสหสัมพันธ์ระหว่างตัวแปร X ที่เป็นตัวแปรคู่ (Dichotomous) กับ ตัวแปร Yที่เป็นอันดับ (Ordinal) เรียกว่า Rank Biserial ใช้สัญลักษณ์ r_{rb} ตัวอย่างการคำนวณในตาราง 10.9

$$r_{rb} = \frac{2}{n_{\bullet}}\left(\overline{Y_1} - \overline{Y_0}\right)$$

เมื่อ $\overline{Y_1}$ เป็นค่าเฉลี่ยของอันดับของผู้ที่ได้คะแนน X เป็น 1

$\overline{Y_0}$ เป็นค่าเฉลี่ยของอันดับของผู้ที่ได้คะแนน X เป็น 0

ตาราง10.9 แสดงการคำนวณค่าสัมประสิทธิ์ Rank Biserial

อันดับของ Y		
X = 1	X = 0	การคำนวณ
10	8	
9	6	$r_{rb} = \frac{2}{n_{\bullet}}\left(\overline{Y_1} - \overline{Y_0}\right)$
7	5	
2	4	$r_{rb} = \frac{2}{10}(7.00 - 4.50) = \frac{2.50}{5} = .50$
	3	
	1	
$\sum Y_1 = 28$	$\sum Y_0 = 27$	
$(n_1 = 4)$	$(n_0 = 6)$	
$\overline{Y_1} = \frac{28}{4} = 7.00$	$\overline{Y_0} = \frac{27}{6} = 4.50$	

แบบฝึกหัด 10

1. นักเรียน 7 คน ทดสอบวิชาภาษาไทยและวิชาภาษาอังกฤษ ได้คะแนนดังนี้

    วิชาภาษาไทย(X) : 4, 10, 8, 9, 5, 7, 8

    วิชาภาษาอังกฤษ(Y) : 3, 8, 9, 7, 7, 6, 8

    1.1 จงเขียนแผนภาพการกระจาย

    1.2 จงหาสัมประสิทธิ์สหสัมพันธ์(r)

2. ในการทดสอบสองกระบวนวิชาของนักเรียนกลุ่มเดียวกันจำนวน 10 คนได้คะแนนดังนี้

    วิชาคณิตศาสตร์(X) : 32, 36, 37, 38, 41, 41, 42, 42, 44, 48

    วิชาวิทยาศาสตร์(Y) : 74, 72, 78, 82, 75, 81, 78, 85, 87, 83

    2.1 จงเขียนแผนภาพการกระจาย

    2.2 จงหาสัมประสิทธิ์สหสัมพันธ์(r)

3. ในชั้นเรียนสถิติ ผู้สอนได้ทดสอบก่อนเรียน (X) ในวันแรก และทดสอบในวันสุดท้ายของการสอน (Y) สำหรับนักเรียนกลุ่มเดียวกัน ข้อมูลกำหนดในตาราง

นักเรียนคนที่	X	Y	นักเรียนคนที่	X	Y	นักเรียนคนที่	X	Y
1	27	54	21	17	40	41	27	43
2	7	24	22	10	37	42	25	44
3	12	41	23	7	34	43	13	34
4	26	36	24	8	31	44	29	55
5	19	56	25	17	34	45	16	27
6	8	34	26	32	40	46	7	30
7	14	31	27	21	32	47	25	51
8	9	36	28	32	54	48	17	20
9	16	46	29	17	36	49	12	33
10	25	40	30	16	42	50	32	39

ตารางต่อ

นักเรียนคนที่	X	Y	นักเรียนคนที่	X	Y	นักเรียนคนที่	X	Y
11	21	42	31	4	27	51	13	49
12	9	35	32	16	54	52	12	26
13	10	35	33	14	50	53	21	42
14	14	32	34	4	18	54	19	53
15	21	47	35	20	47	55	18	21
16	12	35	36	30	36	56	33	42
17	19	39	37	25	51	57	25	51
18	12	23	38	1	32	58	37	46
19	20	42	39	5	26	59	19	50
20	14	22	40	27	43	60	22	38

3.1 จงสร้างแผนภาพการกระจายสำหรับคะแนน 60 คู่

3.2 หา r สำหรับ30 คู่แรกของคะแนนในตาราง คำนวณS_X และ S_Y

3.3 หา r สำหรับ30 คู่หลังของคะแนนในตาราง คำนวณS_X และ S_Y

3.4 คำตอบในข้อ 2. เหมือนกับในข้อ 3. หรือไม่

3.5 หา r สำหรับ 60 คู่ นักศึกษาคิดว่าคำตอบในข้อ4. เกี่ยวข้องกับข้อ 2. และข้อ 3 หรือไม่