สหสัมพันธ์และการทำนาย 4

วัตถุประสงค์

เมื่อผู้อ่านได้ศึกษาเนื้อหาบทที่ 4 แล้ว ควรจะมีความสามารถดังนี้

1. อธิบายความหมายของสหสัมพันธ์ได้
2. คำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบต่าง ๆ ได้
3. พิสูจน์ที่มาของสูตรต่าง ๆ ในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ได้

เนื้อหา

4.1 สหสัมพันธ์ (Correlation)

ในบทนี้มุ่งให้ผู้อ่านเข้าใจความคิดรวบยอด (concept) เกี่ยวกับสหสัมพันธ์อย่างง่าย ๆ ทั้งนี้ก็เพื่อเป็นการช่วยเหลือผู้ที่อ่อนทางด้านสถิติ ดังนั้นหากผู้ใดมีความรู้ทางด้านสถิติเป็นอย่างดีแล้ว ก็อาจที่จะข้ามบทนี้ไปได้เลย

ถ้าตัวแปรสุ่ม X และ Y มีความสัมพันธ์กัน จะหาความสัมพันธ์นั้นได้ และค่าความสัมพันธ์นี้เรียกว่าสัมประสิทธิ์สหสัมพันธ์ เขียนแทนด้วย r หรือ r_{xy}

ค่า r จะมีค่าระหว่าง -1 ถึง $+1$

ถ้า r เป็นบวก หมายความว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์กันในทางตรง คือถ้า X มีค่าน้อย Y จะมีค่าน้อย และถ้า X มีค่ามาก Y จะมีค่ามาก

ถ้า r เป็นลบ หมายความว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์ในทางกลับกัน คือถ้า X มีค่าน้อย Y จะมีค่ามาก และถ้า X มีค่ามาก Y จะมีค่าน้อย

ถ้า r มีค่าอยู่ใกล้ศูนย์ (0) หมายความว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์กันน้อยมากหรือเกือบไม่มีเลย

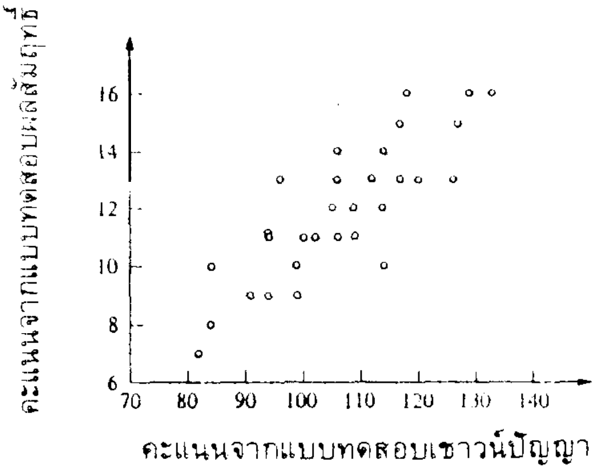
ตาราง 4-1 แสดงผลการสอบเชาวน์ปัญญาและผลสัมฤทธิ์ทางการเรียนของนักเรียน จำนวน 30 คน

แบบทดสอบเชาวน์ปัญญา x	**แบบทดสอบผลสัมฤทธิ์** y	**แบบทดสอบเชาวน์ปัญญา** x	**แบบทดสอบผลสัมฤทธิ์** y
114	14	94	II
84	8	117	13
114	10	126	13
84	10	102	11
106	13	127	15
129	16	133	16
106	14	105	12
94	9	109	II
114	12	118	16
8 2	7	96	13
109	12	99	10
106	II	120	13
117	15	99	9
94	II	91	9
112	13	100	II

M_x = 106.700 M_y = II.933

S_x = 13.372 S_y = 2.353

จากตารางจะเห็นได้ทันทีว่า ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นไปในทิศทางบวก นั่นคือนักเรียนที่ได้คะแนนในแบบทดสอบเชาวน์ปัญญาสูง มักจะได้คะแนนในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนสูงด้วย ความสัมพันธ์ระหว่างตัวแปรทั้งสองนี้สามารถนำมาเขียน Scatter diagram ได้ดังภาพ 4-1



ภาพ 4-1 แสดง Scatter diagram ของความสัมพันธ์ระหว่างตัวแปรทั้งสอง ในตาราง 4-1

จากภาพ 4-1 จุดแต่ละจุดในภาพนี้ แสดงถึงคะแนนของนักเรียนแต่ละคนในแบบทดสอบทั้ง 2 ฉบับ โดยแกน X แทนคะแนนที่ได้จากแบบทดสอบเชาวน์ปัญญา และแกน Y แทนคะแนนจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน จากภาพจะเห็นได้ว่าในจำนวนนักเรียน 12 คน ที่ได้คะแนนเชาวน์ปัญญาเกิน 110 นั้น มีเพียงคนเดียวเท่านั้นที่ได้คะแนนผลสัมฤทธิ์ทางการเรียนต่ำกว่า 12 คะแนน และในจำนวนนักเรียน 10 คน ที่ได้คะแนนเชาวน์ปัญญาน้อยกว่า 100 ก็มีเพียงคนเดียวเช่นกันที่ได้คะแนนผลสัมฤทธิ์ทางการเรียนเกิน 12 คะแนน จากข้อมูลนี้จึงพอจะสรุปได้ว่า คะแนนเชาวน์ปัญญากับผลสัมฤทธิ์ทางการเรียนมีความสัมพันธ์กันในทางบวกค่อนข้างสูง

ถ้าหากรู้ว่าตัวแปรทั้งสองมีความสัมพันธ์กันมากน้อยเพียงใด สามารถใช้ข้อมูลจากตัวแปรหนึ่งทำนายอีกตัวแปรหนึ่งได้ จากตัวอย่างข้างต้นก็เช่นกัน สามารถทำนายได้ว่าคนที่ได้คะแนนเชาวน์ปัญญาสูงน่าจะได้คะแนนผลสัมฤทธิ์ทางการเรียนสูง และคนที่ได้คะแนนเชาวน์ปัญญาต่ำก็น่าจะได้คะแนนผลสัมฤทธิ์ทางการเรียนต่ำด้วย

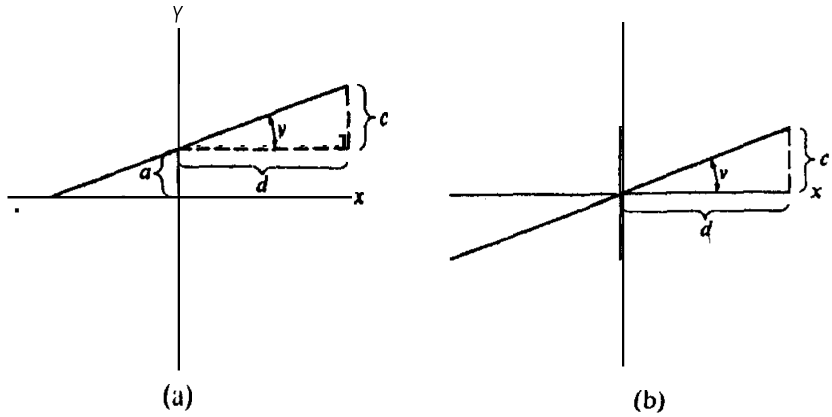
4.2 เส้นถดถอย (Regression line)

ถ้า X และ Y เป็นตัวแปรสุ่มที่มีความสัมพันธ์กัน ความสัมพันธ์นั้นอาจแสดงด้วยสมการเส้นตรง เช่น $Y = 5X+4$ หรือหากเขียนให้อยู่ในรูปของสมการทั่วไปของเส้นตรงจะได้ว่า

$$Y = a + bX \qquad \text{.......... (4-1)}$$

โดยที่ a แทนความยาวตั้งแต่จุด (0, 0) ถึงเส้นที่ตัดกับแกน Y

b แทนความลาดของเส้นตรง



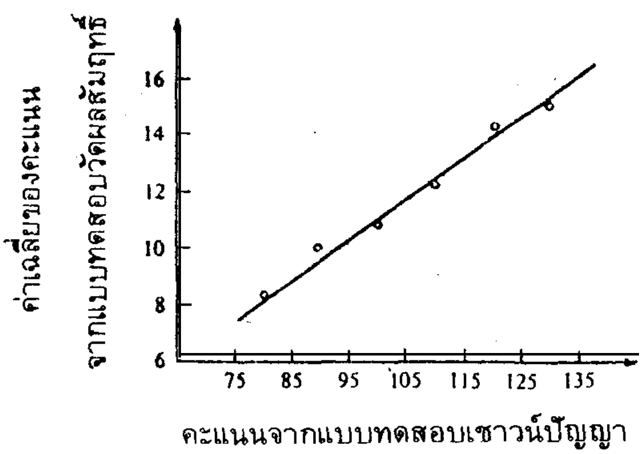
ภาพ 4-2 แสดงสมการของเส้นตรง

ในทางการศึกษามักจะไม่พบตัวแปรทั้งสองที่มีความสัมพันธ์กันอย่างสมบูรณ์ (r = 1.00) เหมือนในตัวอย่างข้างต้น แต่อย่างไรก็ตาม สามารถที่จะหาเส้นตรงที่เหมาะสมมาแทนความสัมพันธ์นั้น ๆ ได้ เช่น ความสัมพันธ์ระหว่างเชาวน์ปัญญา และผลสัมฤทธิ์ทางการเรียน ดังตาราง 4-2

ตาราง 4-2 แสดงค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางการเรียนของนักเรียน ซึ่งแบ่งตามช่วงระดับ I.Q.

I.Q.	75 – 84	85 – 94	95 – 104	105 – 114	115 – 124	125 – 134
M	8.33	10.00	10.80	12.20	14.25	15.00

ความสัมพันธ์นี้ สามารถนำมาเขียนกราฟได้ดังภาพ 4-3



ภาพ 4-3 แสดงความสัมพันธ์ระหว่างเชาวน์ปัญญากับผลสัมฤทธิ์ทางการเรียน

เส้นตรงที่เหมาะสมที่สุดที่จะอธิบายความสัมพันธ์ของตัวแปรทั้งสองนั้น เรียกว่า เส้นถดถอย (Regression line) ซึ่งอยู่ในรูปสมการ

$$Y' = a + bX$$

โดยที่ Y' เป็นค่า Y ที่ได้จากการทำนาย

a, b เป็นค่าคงที่

การทำนายโดยอาศัยสมการเส้นตรงที่เหมาะสม นิยมใช้หลักการของวิธีกำลังน้อยที่สุด (The method of least squares) ซึ่งทำให้ผลบวกทั้งหมดยกกำลังสอง (Sum of square) ของระยะทางระหว่างจุด (x, y) กับเส้นตรงเส้นหนึ่งมีค่าน้อยที่สุด หรืออาจจะสรุปหลักของวิธีกำลังสองน้อยที่สุด ได้ดังนี้

1. ผลรวมของส่วนเบี่ยงเบนจากเส้นตรงนี้ยกกำลังสอง $(\Sigma(Y - Y'))^2$ จะมีค่าน้อยกว่าส่วนเบี่ยงเบนที่เกิดจากเส้นตรงอื่นใด

2. โดยเฉลี่ยแล้วส่วนเบี่ยงเบนที่อยู่เหนือเส้นจะเท่ากับส่วนเบี่ยงเบนที่อยู่ใต้เส้น สรุปแล้วจะมีเส้นตรงเพียงเส้นเดียวที่เป็นไปตามเงื่อนไขดังกล่าว

ค่า a และ b ในสมการของเส้นถดถอย คำนวณได้จากสูตรดังนี้

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

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

ตาราง 4-3 แสดงคะแนนที่ได้จากการสอบคัดเลือก (X) และเปอร์เซ็นต์ที่ได้ (Y) ของนักเรียน จำนวน 12 คน

คนที่	X	Y	X^2	XY
I	63	87	3969	5481
2	50	74	2500	3700
3	55	76	3025	4180
4	65	90	4225	5850
5	55	85	3025	4675
6	70	87	4900	6090
7	64	92	4096	5888
8	70	98	4900	6860
9	58	82	3364	4756
10	68	91	4624	6188
II	52	77	2704	4011
12	60	78	3600	4680
N = 12	$\Sigma x = 730$	$\Sigma Y = 1017$	$\Sigma X^2 = 44932$	$\Sigma XY = 62352$

$\overline{X} = 60.83$

$\overline{Y} = 84.75$

N = 12

แทนค่าจากสูตร

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

$$= 0.925$$

$$a = 84.75 - (0.925)(60.83)$$

$$= 28.482$$

∴ สมการถดถอย คือ

$$Y' = 0.925X + 28.482$$

สมการนี้ทำนายเปอร์เซ็นต์ได้ ถ้าทราบคะแนนสอบคัดเลือก

เช่น นักเรียนคนหนึ่งได้คะแนนสอบคัดเลือก (X) เท่ากับ 40 คะแนน อยากทราบว่านักเรียนคนนี้จะได้เปอร์เซ็นต์ (Y) เท่าไร

$$Y' = 0.925\,(40) + 28.482$$
$$= 37 + 28.482$$
$$= 65.482$$

ในการทำนายแต่ละครั้ง ย่อมเกิดความคลาดเคลื่อนได้ ทั้งนี้เพราะตัวแปรทั้งสองมักไม่ค่อยมีความสัมพันธ์กันอย่างสมบูรณ์ ความคลาดเคลื่อนในการทำนาย เรียกว่า Standard error of estimate เขียนแทนด้วย $S.E._{est.}$

ถ้า $S.E._{est.}$ มีค่ามาก แสดงว่าการทำนายครั้งนี้มีความคลาดเคลื่อนเกิดขึ้นมาก และถ้า $S.E._{est.}$ มีค่าน้อย แสดงว่าการทำนายครั้งนี้มีความคลาดเคลื่อนเกิดขึ้นน้อย

ถ้าค่า Y ที่ได้จากการทำนาย (Y') มีค่าต่างไปจากค่า Y_i ซึ่งได้จากการวัดจริง ๆ มาก ค่า $S.E._{est.}$ ก็จะมากด้วย

ถ้าค่า Y ที่ได้จากการทำนาย (Y') มีค่าต่างไปจากค่า Y_i ซึ่งได้จากการวัดจริง ๆ น้อย ค่า $S.E._{est.}$ ก็จะน้อยด้วย

และถ้าค่า Y ที่ได้จากการทำนาย (Y') ไม่ต่างจากค่า Y_i ซึ่งได้จากการวัดจริง ๆ เลย $(Y_i - Y')^2 = 0$ ค่า $S.E._{est.}$ ก็จะเป็น 0 นั่นคือไม่มีความคลาดเคลื่อนในการทำนายเลย

ค่าความคลาดเคลื่อนในการทำนายคำนวณได้จากสูตร

$$S.E._{est.} = \sqrt{\frac{\Sigma(Y_i - Y')^2}{N-2}}$$

จากภาพ 4-2(b) ถ้าเส้นตรงตัดกับแกน X และ Y ที่จุดกำเนิด (0,0) ค่าของ a จะเป็น 0 จะได้สมการเป็น

$$Y = bX \qquad \ldots\ldots\ldots (4\text{-}2)$$

ในรูปคะแนนดิบอาจมีปัญหาเรื่องการแจกแจงของคะแนนทั้ง 2 ชุดนั้น ฉะนั้น เพื่อตัดปัญหาเรื่องนี้ออกไปก็สามารถทำได้โดยแปลงคะแนนดิบให้เป็นคะแนนมาตรฐาน ทั้งนี้ เพื่อให้ค่าเฉลี่ยของการแจกแจงทั้งสองเป็นศูนย์ นั่นคือถ้าเส้นตรงผ่านจุดกำเนิด และค่า X อยู่ในรูปของคะแนนมาตรฐาน จะได้

$$Z_y' = bZ_x \qquad \text{.......... (4-3)}$$

เนื่องจากคะแนนทุก ๆ ตัวในที่นี้ได้แปลงเป็นคะแนนมาตรฐาน ดังนั้นค่า b (ค่า tangent ของมุม v) จะเท่ากับค่าสหสัมพันธ์ระหว่างตัวแปรทั้งสอง (r_{xy})

$$Z_y' = r_{xy}Z_x \qquad \text{.......... (4-4)}$$

ในทำนองเดียวกันก็สามารถทำนายค่า X จากค่า Y ได้เช่นกัน

$$Z_x' = r_{xy}Z_y \qquad \text{(4-5)}$$

4.3 สัมประสิทธิ์สหสัมพันธ์ (The correlation coefficient)

ค่าสัมประสิทธิ์สหสัมพันธ์ (r) จะมีค่าอยู่ระหว่าง – 1.00 ถึง 1.00 ค่า r จะมีค่าเป็น 1.00 เมื่อตัวแปรทั้งสองมีความสัมพันธ์กันทางบวกอย่างสมบูรณ์ นั่นคือนักเรียนแต่ละคนจะได้ตำแหน่งเดียวกันในทั้งสองตัวแปร และค่า r จะเป็น – 1.00 เมื่อตัวแปรทั้งสองมีความสัมพันธ์กันทางลบอย่างสมบูรณ์ นั่นคือตำแหน่งของนักเรียนแต่ละคนในตัวแปรทั้งสองจะต้องตรงกันข้าม และถ้าตัวแปรทั้งสองไม่มีความสัมพันธ์กันเลย ค่าของ r จะเป็น 0

ความถูกต้องของการทำนายจะขึ้นอยู่กับค่าสัมประสิทธิ์สหสัมพันธ์ นั่นคือ ถ้าตัวแปรทั้งสองมีความสัมพันธ์กันมากเท่าไร ก็จะยิ่งทำให้การทำนายถูกต้องมากขึ้นเท่านั้น

ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทั้งสองเป็น 1.00 หรือ – 1.00 การทำนายก็จะสามารถทำนายได้ถูกต้อง 100% แต่ถ้าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรทั้งสองเป็น 0 และถ้าต้องการจะทำนายค่า Y กันจริง ๆ ก็ทำได้โดยให้เอาค่าเฉลี่ยของข้อมูลในตัวแปร Y มาเป็นค่าในการทำนาย Y ซึ่งจะเห็นได้ว่าไม่ว่าตัวแปร X จะมีค่าเท่าใดก็ตาม จะทำนายค่า Y ออกมาเป็นค่าเดียว คือค่าเฉลี่ยของข้อมูลในตัวแปร Y ดังนั้นสรุปได้ว่า ถ้าค่าสหสัมพันธ์ระหว่างตัวแปรทั้งสองเป็น 0 ก็ไม่มีประโยชน์ที่จะทำนาย

ค่าสัมประสิทธิ์สหสัมพันธ์ที่จะทำให้ $\Sigma(Z_y - Z_y')^2$ มีค่าน้อยที่สุด คือ

$$r_{xy} = \frac{\Sigma Z_xZ_y}{N} \qquad \text{.......... (4-6)}$$

จะเห็นได้ว่าค่าสัมประสิทธิ์สหสัมพันธ์จะเท่ากับค่าเฉลี่ยของผลคูณของคะแนนมาตรฐานของนักเรียนในตัวแปรทั้งสอง

ตารางที่ 4-4 คะแนนมาตรฐานในแบบทดสอบเชาวน์ปัญญาและผลสัมฤทธิ์ทางการเรียนของนักเรียนหญิงจำนวน 30 คน

แบบทดสอบ เชาวน์ปัญญา Z_x	แบบทดสอบ ผลสัมฤทธิ์ Z_y	Z_xZ_y	แบบทดสอบ เชาวน์ปัญญา Z_x	แบบทดสอบ ผลสัมฤทธิ์ Z_y	Z_xZ_y
0.546	0.880	0.4805	-0.950	0.395	0.3753
1.697	-1.670	2.8340	0.770	0.455	0.3504
0.546	-0.820	0.4477	1.443	0.455	0.6566
1.697	0.820	1.3915	-0.351	0.395	0.1386
0.052	0.455	0.0237	1.518	1.305	1.9810
1.668	1.730	2.8856	1.967	1.730	3.4029
0.052	0.880	0.0458	-0.127	0.030	-0.0038
0.950	1.245	1.1828	0.172	-0.395	0.0679
0.546	0.030	0.0164	0.845	1.730	1.4619
1.847	2.095	3.8695	-0.800	0.455	0.3640
0.172	0.030	0.0052	-0.576	-0.820	0.4723
0.052	0.395	0.0205	0.995	0.455	0.4527
0.770	1.305	1.0049	-0.576	-1.245	0.7171
0.950	-0.395	0.3753	-1.174	-1.245	1.4616
0.396	0.455	0.1802	-0.501	-0.395	-0.395

$\Sigma Z_xZ_y = 24.9618, \quad r = \frac{\Sigma Z_xZ_y}{N} = \frac{24.9618}{30} = 0.832$

จากตาราง 4-4 จะได้ค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0.832

จาก $Z_x = \frac{X - M_x}{S_x}$ และ $Z_y = \frac{Y - M_y}{S_y}$

ดังนั้นสมการทั่วไปในรูปคะแนนดิบ

$$r_{xy} = \frac{\Sigma (X - M_x)(Y - M_y)}{NS_xS_y}$$

ถ้าให้ x แทน $X - M_x$

y แทน $Y - M_y$

จะได้ว่า $r_{xy} = \dfrac{\Sigma xy}{NS_xS_y}$ .... (4-7)

ในทางปฏิบัตินิยมแปลงสูตรนี้ให้อยู่ในรูปคะแนนดิบเพื่อง่ายแก่การคำนวณ ซึ่งจะได้ดังนี้

$$r_{xy} = \frac{\Sigma (X-M_x)(Y-M_y)}{\sqrt{\Sigma (X-M_x)^2}\sqrt{\Sigma (Y-M_y)^2}}$$

$$= \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} \quad \text{.......... (4-8)}$$

ตาราง 4-5 แสดงคะแนนผลการสอบคณิตศาสตร์ (X) และวิทยาศาสตร์ (Y) ของนักเรียน จำนวน 10 คน

คนที่	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	63
10	10	8	100	64	80
Σ	50	50	340	290	303

แทนค่าจากสูตร

$$r = \frac{10(303) - (50)(50)}{\sqrt{[10(340) - (50)^2][10(290) - (50)^2]}}$$

$$= \frac{530}{\sqrt{(900)(400)}}$$

530
6 0 0

$$= 0.883$$

4.4 The Ø – coefficient

ในกรณีที่คะแนนที่ได้จากตัวแปร X และ Y เป็น Dichotomus ทั้งคู่ คือมีค่าเป็น 0 กับ 1 จะสามารถหาค่าสหสัมพันธ์ได้โดยใช้ Ø – coefficient

ตาราง 4-6 แสดงคะแนนของนักเรียน 20 คน ในข้อสอบข้อที่ 5 และข้อที่ 6

ข้อที่ \ นักเรียน	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	p	q
5	0	0	0	1	0	0	1	0	1	1	1	1	1	1	0	1	0	1	0	1	0.55	0.45
6	0	0	0	1	0	0	1	0	1	1	1	0	1	0	0	0	0	1	1	0	0.40	0.60

จากตาราง 4-6 จะเห็นได้ว่ามีนักเรียน 4 คน ที่ตอบข้อ 5 ถูก แต่ตอบข้อ 6 ผิด มีนักเรียน 7 คน ที่ตอบถูกทั้งสองข้อ มีนักเรียน 8 คน ที่ตอบผิดทั้งสองข้อ และมีนักเรียนเพียง 1 คนเท่านั้นที่ตอบข้อ 5 ผิด แต่ตอบข้อ 6 ถูก จากข้อมูลนี้สามารถสรุปเป็นตารางได้ดังนี้

ตาราง 4-7 สรุปจำนวนคนตอบถูกและผิดในแต่ละข้อ

ข้อ 5 \ ข้อ 6	**ผิด**	**ถูก**	**รวม**
ถูก	4	7	11
ผิด	8	1	9
รวม	12	8	20

สามารถสรุปเป็นตารางทั่วไปได้ดังนี้

ตาราง 4-8 ตารางสรุป

ข้อ k / ข้อ i	ผิด	ถูก	รวม
ถูก	A (a)	B (b)	$a + b = p_i$
ผิด	C (c)	D (d)	$c + d = q_i$
รวม	$a + c = q_k$	$b + d = p_k$	1.00

จากตาราง 4-8 ให้ A, B, C, D แทนความถี่ของคะแนนดิบในแต่ละช่อง และ a, b, c, d แทนสัดส่วนในแต่ละช่อง ซึ่ง $a + b + c + d = 1.00$

จากสมการ 4-7 จะได้ว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อ i และ k เป็น

$$r_{ik} = \frac{\Sigma (X_i - M_i)(X_k - M_k)}{N S_i S_k}$$

พิจารณาทางด้านขวามือโดยไม่พิจารณาค่าความเบี่ยงเบนมาตรฐาน

$$\frac{\Sigma X_i X_k}{N} - \frac{M_k \Sigma X_i}{N} - \frac{M_i \Sigma X_k}{N} + \frac{\Sigma M_i M_k}{N}$$

แต่ $\Sigma X_i X_k / N = p_{ik}$

(p_{ik} คือสัดส่วนของจำนวนคนที่ตอบถูกทั้งสองข้อ ซึ่งก็คือ b ในตาราง 4-8)

และ $\frac{\Sigma x_i}{N} = M_i$, $\frac{\Sigma X_k}{N} = M_k$

ดังนั้นจะได้เทอมสุดท้าย $NM_iM_k/N = M_iM_k$

นั่นคือถ้าจะเขียนให้อยู่ในรูปง่าย ๆ จะได้เป็น

$$p_{ik} - M_iM_k - M_iM_k + M_iM_k = p_{ik} - M_iM_k$$

เนื่องจาก $M_i = p_i$ และ $M_k = p_k$

ดังนั้น $$r_{ik} = \frac{p_{ik} - p_i p_k}{S_i S_k} \qquad \text{.......... (4-9)}$$

ค่าความเบี่ยงเบนมาตรฐานของข้อ i คือ $\sqrt{p_i q_i}$

$$r_{phi} = \frac{p_{ik} - p_i p_k}{\sqrt{p_i q_i p_k q_k}} \qquad (4\text{-}10)$$

จากตาราง 4-7 จะได้ว่า

$$r_{phi} = \frac{\frac{7}{20} - (\frac{11}{20})(\frac{8}{20})}{\sqrt{(\frac{11}{20})(\frac{9}{20})(\frac{8}{20})(\frac{12}{20})}}$$

$$= \frac{0.35 \quad 0.22}{\sqrt{0.0594}}$$

$$= 0.53$$

หรืออาจคำนวณค่า r_{phi} ได้โดยตรงจากความถี่ ดังสูตร

$$r_{phi} = \frac{BC - AD}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \qquad \text{.........} (4\text{-}11)$$

ดังนั้นค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อ 5 และข้อ 6

$$r_{phi} = \frac{7 \times 8 - 4 \times 1}{\sqrt{11 \times 9 \times 12 \times 8}}$$

$$\frac{56 - 4}{\sqrt{9504}}$$

$$= 0.53$$

สรุปเนื้อหาบทที่ 4

1. ถ้าตัวแปรสองชุดมีความสัมพันธ์กัน สามารถใช้ข้อมูลจากตัวแปรหนึ่งทำนายอีกตัวแปรหนึ่งได้ และในทำนองกลับกันหากตัวแปรทั้งสองไม่มีความสัมพันธ์กัน ก็ไม่สามารถจะทำนายอีกตัวแปรหนึ่งได้
2. ถ้าตัวแปรสองชุดมีความสัมพันธ์กันมาก ความคลาดเคลื่อนในการทำนายก็จะน้อย แต่หากตัวแปรทั้งสองมีความสัมพันธ์กันน้อย ความคลาดเคลื่อนในการทำนายก็จะมีค่ามาก
3. ในกรณีที่คะแนนชุดหนึ่งได้แปลงเป็นคะแนนมาตรฐานแล้ว ค่าความชันของเส้นถดถอยจะมีค่าเท่ากับค่าสหสัมพันธ์ระหว่างตัวแปรทั้งสอง
4. ในกรณีที่ข้อมูลจากตัวแปรทั้งสองเป็นข้อมูลต่อเนื่อง (Continuous) สามารถคำนวณหาค่าสหสัมพันธ์โดยใช้ Pearson Product Moment Correlation และถ้าข้อมูลจากตัวแปรทั้งสองเป็นข้อมูลไม่ต่อเนื่อง (Discrete) ชนิด Dichotomus สามารถหาค่าสหสัมพันธ์โดยใช้ Ø-coefficient

คำถามท้ายบทที่ 4

1. ในการนำข้อสอบฉบับหนึ่งซึ่งมี 2 ข้อไปทดสอบกับนักเรียน 12 คน ปรากฏผลดังนี้

ข้อที่ \ คนที่	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	0	0	0	0	0	0	1	1	1
2	0	0		0		111	0		0	0		111

ใช้ข้อมูลจากตารางตอบคำถามต่อไปนี้

1.1 S_1^2 มีค่าเท่าไร

1.2 S_2^2 จะมีค่าเท่ากับ 1 ได้ในกรณีใด

1.3 r_{12} มีค่าเท่าไร

1.4 C_{12} มีค่าเท่าไร

1.5 $S^2_{\text{ทั้งฉบับ}}$ มีค่าเท่าไร

2. จงแสดงวิธีการสร้างสมการต่อไปนี้

$$r_{ik} = \frac{p_{ik} - p_i p_k}{\sqrt{p_i q_i p_k q_k}}$$

3. ในกรณีที่คะแนนชุดหนึ่งเป็นคะแนนมาตรฐานแล้ว ค่าความชันของเส้นถดถอย (b) จะมีค่าเท่ากับอะไร