

เทคนิคการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร

รองศาสตราจารย์ ดร.สุภมาส อังคุโษิต

บทคัดย่อ

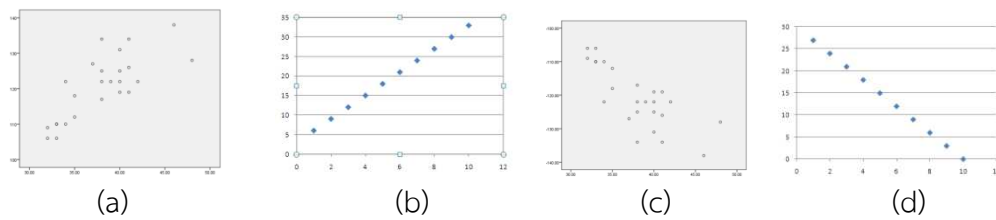
การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เป็นการวิเคราะห์ว่าตัวแปรสองตัวหรือชุดของตัวแปรมีความผันแปรเกี่ยวเนื่องกันมากน้อยเพียงใด เรียกค่าที่แสดงขนาดความสัมพันธ์ว่าค่าสัมประสิทธิ์สหสัมพันธ์ ซึ่งมีค่าระหว่าง 0 ถึง ± 1 ค่าใกล้ 0 หมายถึงสัมพันธ์กันน้อย ค่าใกล้ 1 หมายถึงสัมพันธ์กันมาก ส่วนเครื่องหมายจะแสดงทิศทางของความสัมพันธ์ การแปลความหมายความสัมพันธ์จึงต้องพิจารณาทั้งขนาดและทิศทางของความสัมพันธ์ สถิติสำหรับการวิเคราะห์ความสัมพันธ์มีหลายชนิด ขึ้นอยู่กับมาตรวัดของตัวแปรและจำนวนตัวแปรที่นำมาหาความสัมพันธ์

แนวคิดเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร

การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเริ่มจากการสังเกตการเปลี่ยนแปลงของตัวแปรสองตัว เพื่อหา รูปแบบความสัมพันธ์ของตัวแปรเหล่านั้น Natalia Ospina-Alvarez และ Francese Piferrer (2008) นักชีววิทยาชาวสเปนได้ศึกษาทางห้องปฏิบัติการพบว่า อุณหภูมิของน้ำเพิ่มขึ้นจะทำให้ปริมาณประชากรปลาเพศผู้เพิ่มขึ้นอย่างชัดเจน เช่น เมื่อเพิ่มอุณหภูมิของน้ำเพิ่มขึ้น 4 องศาเซลเซียส จะเพิ่มปริมาณปลา South American pejerrey เพศผู้ได้ถึงร้อยละ 98 ของปริมาณปลาทั้งสองเพศรวมกัน จากผลการศึกษานี้แสดงให้เห็นความสัมพันธ์ของตัวแปรสองตัวคือ อุณหภูมิของน้ำ กับ จำนวนปลา South American pejerrey เพศผู้ เป็นความสัมพันธ์เชิงบวก กล่าวคือถ้าอุณหภูมิของน้ำเพิ่มขึ้นปริมาณประชากรปลาเพศผู้เพิ่มขึ้น เรียกว่า **สหสัมพันธ์เชิงบวก** (positive correlation) ยังมีการเปลี่ยนแปลงอีกลักษณะหนึ่งคือเมื่อตัวแปรตัวหนึ่งมีค่าเพิ่มขึ้นตัวแปรอีกตัวหนึ่งจะมีค่าลดลง เรียกว่า **สหสัมพันธ์เชิงลบ** (negative correlation) เช่น จากการสำรวจของวารสารคลินิกไอเสีย พบว่า “รถตุ๊กตุ๊กที่มีเครื่องยนต์ 4 จังหวะ จะมีเสียงดังโดยเฉลี่ย 84.9 เดซิเบลเอ (dBA) รถตุ๊กตุ๊กที่มีเครื่องยนต์ 2 จังหวะ จะมีเสียงดังโดยเฉลี่ย 93.9 เดซิเบลเอ (dBA)” แสดงให้เห็นความสัมพันธ์ว่าเมื่อเครื่องยนต์มีการจุดระเบิดด้วยจังหวะลดลงจะมีเสียงดังมากขึ้น

นักวิจัยสามารถคำนวณหาตัวชี้เพื่อใช้วัดระดับความสัมพันธ์ของตัวแปร เรียกตัวชี้นี้ว่า **สัมประสิทธิ์สหสัมพันธ์** (correlation coefficient) ซึ่งสัมประสิทธิ์สหสัมพันธ์ ก็คือค่าสัมประสิทธิ์ที่แสดงขนาดของความสัมพันธ์ของตัวแปร เครื่องหมายบวก (+) หรือลบ (-) จะแสดงทิศทางของความสัมพันธ์ ค่าสัมประสิทธิ์สหสัมพันธ์จะมีพิสัยระหว่าง +1 ถึง -1 ค่า +1 แสดงถึงความสัมพันธ์เชิงบวกอย่างสมบูรณ์ ค่า 0 แสดงว่าไม่มีความสัมพันธ์กัน และค่า -1 แสดงถึงความสัมพันธ์ผกผันอย่างสมบูรณ์ (Hair and others, 2010: 156-157)

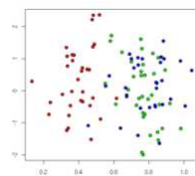
การตรวจสอบรูปแบบความสัมพันธ์ระหว่างตัวแปรสองตัวอย่างง่าย ๆ ทำได้โดยใช้แผนภาพการกระจาย (scatter diagram หรือ scatterplot หรือ scattergram) แผนภาพการกระจายก็คือการแสดงตำแหน่งของค่าสังเกตด้วยจุด ซึ่งแต่ละจุดจะแสดงค่าร่วมกันของตัวแปรทั้งสอง ดังภาพที่ 1-3



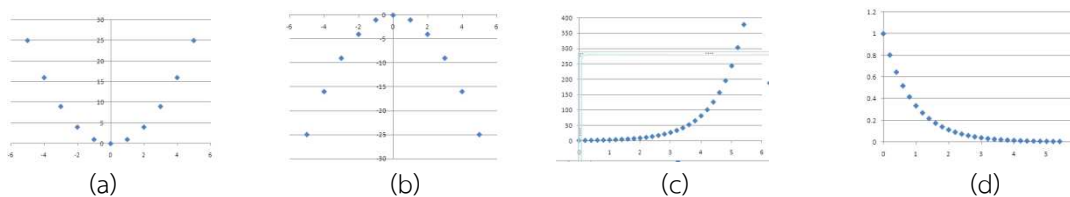
ภาพที่ 1 แผนภาพการกระจายแสดงความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรสองตัว

(a) และ (b) แสดงความสัมพันธ์เชิงเส้นตรงแบบความสัมพันธ์เชิงบวก

(c) และ (d) แสดงความสัมพันธ์เชิงเส้นตรงแบบความสัมพันธ์เชิงลบ



ภาพที่ 2 แผนภาพการกระจายแสดงว่าตัวแปรสองตัวไม่มีความสัมพันธ์ต่อกัน



ภาพที่ 3 แผนภาพการกระจายแสดงความสัมพันธ์เชิงเส้นโค้งระหว่างตัวแปรสองตัว

การสร้างแผนภาพการกระจายด้วยโปรแกรม SPSS ใช้คำสั่ง

Graphs --> Legacy Dialogs --> Scatter/Dot --> Simple Scatter --> Define.....

การเลือกใช้สถิติเพื่อวิเคราะห์ความสัมพันธ์

แผนภาพการกระจายทำให้ทราบถึงลักษณะและทิศทางของความสัมพันธ์ระหว่างตัวแปร แต่ถ้าต้องการคำนวณขนาดความสัมพันธ์จำเป็นต้องใช้สถิติในการวิเคราะห์ ซึ่งมีให้เลือกมากมาย ผู้วิจัยต้องเลือกให้เหมาะสมโดยพิจารณาจากมาตรวัดและจำนวนตัวแปรที่นำมาหาความสัมพันธ์ ในที่นี้จะนำเสนอการวิเคราะห์ความสัมพันธ์เป็น 2 ประเภทคือ ความสัมพันธ์ระหว่างตัวแปร (ตารางที่ 1) และความสัมพันธ์ระหว่างชุดของตัวแปร (ตารางที่ 2)

ตารางที่ 1 แนวทางในการเลือกสถิติเพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร

มาตรวัดของตัวแปร		สถิติที่ใช้วิเคราะห์ความสัมพันธ์
มาตรวัดของตัวแปรตัวที่ 1	มาตรวัดของตัวแปรตัวที่ 2	
Interval scale หรือ Ratio scale	Interval scale หรือ Ratio scale	Pearson Correlation (linear correlation)
Ordinal scale	Ordinal scale	Spearman Rank Correlation
Ordinal scale (สองอันดับ)	Ordinal scale (สองอันดับ)	Tetrachoric correlation
Ordinal scale (มากกว่าสองอันดับ)	Ordinal scale (มากกว่าสองอันดับ)	Polychoric correlation Polyserial correlation
Nominal scale (สองกลุ่ม)	Nominal scale (สองกลุ่ม)	Phi Correlation
Nominal scale (มากกว่าสองกลุ่ม)	Nominal scale (มากกว่าสองกลุ่ม)	Cramer's V, Contingency Coefficient
Nominal scale (สองกลุ่มตามธรรมชาติ เช่น เพศ)	Interval scale หรือ Ratio scale	Point-biserial correlation
Nominal scale (สองกลุ่มที่เกิดจากการจัดกระทำ เช่น วยแบ่งตามช่วงอายุ)	Interval scale หรือ Ratio scale	Biserial correlation
Nominal scale (สองกลุ่ม)	Ordinal scale (มากกว่าสองอันดับ)	Rank biserial correlation

ตารางที่ 2 แนวทางในการเลือกสถิติเพื่อวิเคราะห์ความสัมพันธ์ระหว่างชุดของตัวแปร

จำนวนตัวแปร		สถิติที่ใช้วิเคราะห์ความสัมพันธ์
ตัวแปรชุดที่ 1	ตัวแปรชุดที่ 2	
หลายตัว (X1, X2, X3,...,Xn) Interval scale หรือ Ratio scale	1 ตัว (Y) Interval scale หรือ Ratio scale	Multiple Correlation
หลายตัว (X1, X2, X3, ..., Xn) Interval scale หรือ Ratio scale	หลายตัว (Y1, Y2, Y3, ..., Ym) Interval scale หรือ Ratio scale	Canonical Correlation

1. สหสัมพันธ์แบบเพียร์สัน (Pearson correlation)

สหสัมพันธ์แบบเพียร์สัน (Pearson correlation หรือ Pearson product-moment correlation) เป็นเทคนิคสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรที่มีมาตรวัดแบบช่วง (Interval scale) หรืออัตราส่วน (Ratio scale) สองตัว หรือเรียกตัวแปรประเภทนี้ว่าตัวแปรเชิงปริมาณ (metrics variables) ชื่อสหสัมพันธ์แบบเพียร์สัน ให้ชื่อตาม คาร์ล เพียร์สัน (Karl Pearson) ผู้คิดสถิตินี้ขึ้น โดยใช้ชื่อว่า product-moment coefficient of correlation ประมาณ ค.ศ. 1900 สหสัมพันธ์แบบเพียร์สัน ได้ถูกนำไปใช้กันอย่างแพร่หลายในสาขาสังคมศาสตร์ทุกแขนง (Glass & Hopkins, 1984: 80) สัญลักษณ์ที่ใช้

แทนค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน คือ r สำหรับตัวอย่าง และ ρ สำหรับประชากร กำลังสองของ r (r^2) แทนค่าสัมประสิทธิ์ของการทำนาย (coefficient of determination) ซึ่งเป็นค่าที่แสดงว่าตัวแปรหนึ่ง อธิบายความแปรปรวนของตัวแปรอีกตัวหนึ่งได้ประมาณร้อยละเท่าไร สูตรในการคำนวณ r มาจาก

$$r = \frac{\text{cov}_{xy}}{s_x s_y}$$
 (Glass & Hopkins, 1984: 85, Howell, 2007: 238) แต่สูตรที่นิยมใช้กันเป็นการคำนวณ จากค่าเบี่ยงเบนของตัวแปร X และ Y มีสูตรการคำนวณ ดังนี้ (Forthofer and others, 2007: 356)

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

การทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันใช้สถิติทดสอบที (t - test) ซึ่งมีสูตรว่า

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

โดยมีสมมติฐานศูนย์ว่า $H_0: \rho = 0$ (ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน)

ดังนั้น เมื่อการทดสอบมีนัยสำคัญ แปลความหมายได้ว่า ตัวแปรทั้งสองมีความสัมพันธ์กัน (ค่า สัมประสิทธิ์สหสัมพันธ์ต่างจากศูนย์) มิได้หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์กันมาก การจะแปล ความหมายว่าตัวแปรทั้งสองสัมพันธ์กันมากน้อยเพียงใดต้องนำค่าสัมประสิทธิ์สหสัมพันธ์ไปเทียบกับตารางการ แปลความหมายซึ่งนักสถิติกำหนดเป็นช่วงของค่าสัมประสิทธิ์สหสัมพันธ์ไว้หลายแบบ ดังตัวอย่างของ Best (1977:240)

ตารางที่ 3 การแปลความหมายค่าสัมประสิทธิ์สหสัมพันธ์ ของ Best

ค่าสัมประสิทธิ์สหสัมพันธ์	ระดับความสัมพันธ์
0.81 - 1.00	มีความสัมพันธ์กันมาก
0.51 - 0.80	มีความสัมพันธ์กันปานกลาง
0.21 - 0.50	มีความสัมพันธ์กันน้อยหรือต่ำ
0.01 - 0.20	มีความสัมพันธ์กันน้อยมาก
0.00	ไม่มีความสัมพันธ์กัน

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ด้วยโปรแกรม SPSS ใช้คำสั่ง

Analyze --> Correlate --> Bivariate เลือก Pearson

ตัวอย่าง การหาความสัมพันธ์ระหว่างคะแนนด้านความรู้หนังสือและความสามารถในการอ่าน (X1) ความสามารถในการคิดคำนวณ (X2) และความสามารถในการแก้ปัญหาภายใต้สภาพแวดล้อมที่ใช้ประโยชน์ จากเทคโนโลยี (X3) ของคนไทย 2,000 คน ที่มีอายุระหว่าง 15-65 ปี ได้ค่าสัมประสิทธิ์สหสัมพันธ์ ดังตารางที่ 4

ตารางที่ 4 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง คะแนนด้านการรู้หนังสือและความสามารถในการอ่าน (X1) ความสามารถในการคิดคำนวณ (X2) และความสามารถในการแก้ปัญหาภายใต้สภาพแวดล้อมที่ใช้ประโยชน์จากเทคโนโลยี (X3)

คะแนน	X1	X2	X3
การรู้หนังสือและความสามารถในการอ่าน (X1)	1.000		
ความสามารถทางการคิดคำนวณ (X2)	0.661**	1.000	
ความสามารถในการแก้ปัญหาภายใต้ IT ENV (X3)	0.063*	0.048*	1.000

*p<0.05 **p<0.01

ผลการวิเคราะห์พบว่า คะแนนการรู้หนังสือและความสามารถในการอ่าน (X1) มีความสัมพันธ์ทางบวก ขนาดปานกลางกับคะแนนความสามารถทางการคิดคำนวณ (X2) ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ 0.661 มีนัยสำคัญทางสถิติที่ระดับ 0.01 ส่วนคะแนนการรู้หนังสือและความสามารถในการอ่าน (X1) และคะแนนความสามารถทางการคิดคำนวณ (X2) มีความสัมพันธ์ทางบวกอย่างน้อยกับคะแนนความสามารถในการแก้ปัญหาภายใต้สภาพแวดล้อมที่ใช้ประโยชน์จากเทคโนโลยี (X3) ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ 0.063 และ 0.048 ตามลำดับ ความสัมพันธ์นี้มีนัยสำคัญทางสถิติที่ระดับ 0.05

2. สหสัมพันธ์เชิงอันดับของสเปียร์แมน (Spearman Rank Correlation)

สหสัมพันธ์เชิงอันดับของสเปียร์แมน (Spearman Rank Correlation) เป็นเทคนิคทางสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์ระหว่างการจัดอันดับของหลายสิ่งด้วยคนสองคนว่ามีการจัดอันดับที่สอดคล้องสัมพันธ์กัน เช่น การประเมินภาพวาด 10 ภาพ โดยผู้ประเมิน 2 คนจัดอันดับ **Spearman Rank Correlation** จะให้ค่าสัมประสิทธิ์ที่บอกว่าผู้ประเมินทั้งสองคนจัดอันดับภาพสอดคล้องกันมากน้อยเพียงใด สูตรในการหาขนาดและทิศทางของสหสัมพันธ์เชิงอันดับของสเปียร์แมน ดังนี้

$$r_{sy} = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

โดยที่ d คือ ความแตกต่างระหว่างอันดับของสองตัวแปร

n คือ ขนาดของกลุ่มตัวอย่าง

การทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์เชิงอันดับของสเปียร์แมน ใช้สถิติทดสอบที (t - test) ซึ่งใช้สูตรเดียวกับการทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

การแปลความหมายขนาดและทิศทางความสัมพันธ์ใช้หลักการเดียวกับสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์เชิงอันดับของสเปียร์แมน ด้วยโปรแกรม SPSS ใช้คำสั่ง

Analyze --> Correlate --> Bivariate เลือก Spearman

ตัวอย่าง กรรมการตัดสินประกวดร้องเพลง 2 คน จัดอันดับความสามารถในการร้องเพลง ของผู้เข้าประกวด 10 คน ดังตารางที่ 5

ตารางที่ 5 ผลการจัดอันดับของผู้เข้าประกวด 10 คน ของกรรมการ 2 คน

ผู้เข้าประกวด	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5	คนที่ 6	คนที่ 7	คนที่ 8	คนที่ 9	คนที่ 10
ผู้ประเมิน 1	7	8	4	3	1	2	5	6	10	9
ผู้ประเมิน 2	5	9	7	2	1	3	4	6	8	10

ผลการวิเคราะห์ได้ค่าสัมประสิทธิ์สหสัมพันธ์เชิงอันดับของสเปียร์แมน 0.867 มีนัยสำคัญทางสถิติที่ระดับ 0.01 หมายความว่า กรรมการตัดสินประกวดร้องเพลง 2 คน ให้คะแนนสอดคล้องกันสูงมาก

3. Tetrachoric correlation

Tetrachoric correlation เป็นเทคนิคสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์ระหว่างตัวแปรที่มีมาตรวัดเป็นเรียงอันดับ (Ordinal scale) แบบแบ่งเป็น 2 อันดับ (binary variables) ทั้งสองตัวแปร ซึ่ง Holgado-Tello, et al (2010) และ Garrido et al (2013) แนะนำให้ใช้

สูตรการคำนวณ ดังนี้
$$r_{tet} = \frac{\gamma - 1}{\gamma + 1}$$

โดยที่
$$\gamma = \left(\frac{a_{00}a_{11}}{a_{10}a_{01}} \right)^{\frac{\pi}{4}}$$

ตัวอย่าง จากการสอบถามกลุ่มตัวอย่าง 80 คน ให้จัดอันดับความชอบช้อปปิ้ง (X) กับ ความชอบคอนเสิร์ต (Y) โดยให้ตอบว่าชอบมาก หรือชอบน้อย ทั้งสองตัวแปร ค่าตอบปรากฏ ดังตารางที่ 6

ตารางที่ 6 การหาความสัมพันธ์ระหว่างความชอบช้อปปิ้งกับความชอบคอนเสิร์ต

ความชอบคอนเสิร์ต (Y)	ความชอบช้อปปิ้ง (X)		รวม
	ชอบน้อย	ชอบมาก	
ชอบน้อย	15 (a_{00})	32 (a_{01})	47
ชอบมาก	21 (a_{10})	12 (a_{11})	33
รวม	36	44	80

ผลการวิเคราะห์ $r_{tet} = -0.476$

หมายความว่า ความชอบช้อปปิ้งกับความชอบคอนเสิร์ตมีความสัมพันธ์ในทิศทางตรงข้ามคือคนที่ชอบช้อปปิ้งจะไม่ชอบคอนเสิร์ต และในทางกลับกัน คนที่ชอบคอนเสิร์ตจะไม่ชอบช้อปปิ้ง ความสัมพันธ์นี้มีขนาดปานกลาง (ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.476)

การคำนวณค่า Tetrachoric correlation ด้วยโปรแกรม SPSS ขณะนี้ยังไม่สามารถใช้คำสั่ง Analyze หากต้องการคำนวณต้องไป download คำสั่ง SPSSINC_HETCOR จากเว็บไซต์

<https://www.ibm.com/developerworks/mydeveloperworks/files/app?lang=en#/person/270002VCWN/file/dd26db1f-a481-45a9-bb26-ce8ceab65f4b> .

ปัจจุบันโปรแกรมสำเร็จรูปที่ใช้คำนวณ Tetrachoric correlation ได้คือ โปรแกรม PRELIS ใน LISREL หรือโปรแกรม Mplus

ข้อมูลชุดเดียวกันนี้ผลการวิเคราะห์ด้วยโปรแกรม PRELIS และ โปรแกรม Mplus ดังภาพ

Correlation Matrix		
	X	Y
X	1.000	
Y	-0.476	1.000

PRELIS

SAMPLE TETRACHORIC CORRELATIONS		
	X	Y
X		
Y	-0.476	

Mplus

4. Polychoric correlation และ Polyserial correlation

Polychoric correlation และ Polyserial correlation เป็นเทคนิคสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์ระหว่างตัวแปรที่มีมาตรวัดเป็นเรียงอันดับ (Ordinal scale) แบบแบ่งมากกว่า 2 อันดับ

ตัวอย่าง ผู้วิจัยต้องการหาความสัมพันธ์ระหว่างความพึงพอใจในภาพรวมของการจัดอบรม กับความพึงพอใจเกี่ยวกับวิทยากร ซึ่งทั้งสองตัวแปรวัดจากมาตราประมาณค่า (rating scale) 5 ระดับ 2 ข้อ ซึ่งตัวแปรสองตัวนี้วัดในมาตรวัดเรียงอันดับ (Ordinal scale) ซึ่ง Holgado-Tello, et al (2010) และ Garrido et al (2013) แนะนำให้ใช้ Polychoric correlation เนื่องจากมีความเหมาะสมกับมาตรวัดของตัวแปร

ปัจจุบันโปรแกรมสำเร็จรูปที่ใช้คำนวณ Polychoric correlation คือ โปรแกรม PRELIS ใน LISREL

OPIN1		
Category 1	0.195	247.000
Category 2	0.301	381.000
Category 3	0.279	354.000
Category 4	0.177	224.000
Category 5	0.048	61.000
OPIN2		
Category 1	0.166	210.000
Category 2	0.294	373.000
Category 3	0.340	431.000
Category 4	0.164	208.000
Category 5	0.036	45.000

CORRELATION MATRIX (WITH VARIANCES ON THE DIAGONAL)		
	OPIN1	OPIN2
OPIN1		
OPIN2	0.825	

5. Phi Correlation, Contingency Coefficient และ Cramer's V

Phi Correlation (Φ) เป็นเทคนิคสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์ระหว่างตัวแปรที่มีมาตรวัดเป็นนามบัญญัติ (Nominal scale) ที่ตัวแปรทั้งสองตัวแบ่งได้เป็น 2 กลุ่ม

Contingency Coefficient (C) และ Cramer's V (V) เป็นเทคนิคสถิติที่ใช้หาขนาดความสัมพันธ์ระหว่างตัวแปรที่มีมาตรวัดเป็นนามบัญญัติ (Nominal scale) ที่ตัวแปรตัวหนึ่งหรือทั้งสองตัวแบ่งได้มากกว่า 2 กลุ่ม

สถิติทั้งสามถ้าจะคำนวณอย่างง่ายให้เริ่มต้นด้วยการคำนวณค่าไค-สแควร์ก่อน ดังนั้นการจะเลือกใช้สถิติตัวใดให้ดูที่ขนาดของเมทริกซ์ ดังนี้

phi correlation ใช้เมื่อเมทริกซ์มี ขนาด 2X2

Contingency Coefficient ใช้เมื่อเมทริกซ์มีจำนวนแถวและจำนวนสดมภ์เท่ากัน (เมทริกซ์จัตุรัส) และมีขนาดตั้งแต่ 3X3 หรือมากกว่า

Cramer's V ใช้เมื่อเมทริกซ์มีจำนวนแถวและจำนวนสดมภ์ไม่เท่ากัน เช่น 2X3 3X2 4X5
เมื่อคำนวณค่า ไค-สแควร์ได้แล้วสามารถคำนวณ Phi Correlation, Contingency Coefficient และ Cramer's V จากสูตรต่อไปนี้

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}$$

เมื่อ k คือค่าที่น้อยกว่าระหว่างแถวกับสดมภ์

ตัวอย่าง ผู้วิจัยต้องการหาความสัมพันธ์ระหว่างการสูบบุหรี่ กับการเป็นถุงลมโป่งพอง ผู้วิจัยต้องนำข้อมูลที่เก็บรวบรวมได้มาแจกแจงความถี่แบบสองทาง ดังตารางที่ 7

ตารางที่ 7 การหาความสัมพันธ์ระหว่างการสูบบุหรี่ กับการเป็นถุงลมโป่งพอง

ตัวแปรที่ 2 (การเป็นถุงลมโป่งพอง)		ตัวแปรที่ 1 (การสูบบุหรี่)		รวม
		สูบ	ไม่สูบ	
เป็น	(O)	40 (a)	12 (b)	52
	(E)	23.21	28.79	
ไม่เป็น	(O)	10 (c)	50 (d)	60
	(E)	26.79	33.21	
รวม		50	62	112

หมายเหตุ O คือค่าสังเกต หรือความถี่ที่เก็บข้อมูลมา

E คือค่าคาดหวัง ซึ่งมีสูตร $E_{ij} = \frac{\text{ผลรวมความถี่แถวที่ } i \times \text{ผลรวมความถี่สดมภ์ที่ } j}{\text{ผลรวมของความถี่ทั้งหมด}}$

$$\chi^2 = \sum_{i=1}^k \frac{(O - E)^2}{E} \quad df = k - 1$$

$$\phi = \sqrt{\frac{40.93}{112}} = 0.6045$$

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

ทดสอบความมีนัยสำคัญโดยใช้ $\chi^2 = N\phi^2$ หรือ $t = \frac{\phi\sqrt{n-2}}{\sqrt{1-\phi^2}}$, df = n-2

การคำนวณค่า Phi Correlation, Contingency Coefficient และ Cramer's ด้วยโปรแกรม SPSS ใช้คำสั่ง

Analyze --> Descriptive statistics --> Crosstabs --> Statistics เลือก Phi and Cramer's V และ Contingency Coefficient

6. Point-biserial correlation

Point-biserial correlation เป็นเทคนิคสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งที่มีมาตรวัดนามบัญญัติ (Nominal scale) แบบสองกลุ่มตามธรรมชาติ (dichotomous) กับตัวแปรอีกตัวหนึ่งที่มีมาตรวัดเป็นอันตรภาค หรืออัตราส่วน (Interval scale หรือ Ratio scale) เช่น ความสัมพันธ์ระหว่างเพศกับรายได้ ความสัมพันธ์ระหว่างการสูบบุหรี่กับปริมาณคอเรสเตอรอลในเลือด

การวิเคราะห์ Point-biserial correlation ด้วยโปรแกรม SPSS ใช้คำสั่ง

Analyze --> Correlate --> Bivariate --> Pearson

โดยข้อมูลต้องเป็นไปตามข้อตกลงเบื้องต้น 3 ข้อ ดังนี้

1) ตัวแปรตัวหนึ่งที่วัดในมาตรวัดอันตรภาค หรืออัตราส่วน (Interval scale หรือ Ratio scale) เป็นตัวแปรต่อเนื่อง (Continuous Variable)

2) ตัวแปรอีกตัวหนึ่งวัดในมาตรวัดนามบัญญัติ (Nominal scale) แบบสองกลุ่มตามธรรมชาติ (dichotomous)

3) เมื่อแยกตัวแปรแปรต่อเนื่องออกเป็น 2 กลุ่มตามตัวแปรที่วัดในมาตรวัดนามบัญญัติ จะต้อง

3.1 มีการแจกแจงปกติทั้งสองกลุ่ม (ทดสอบโดยใช้คำสั่ง Analyze --> Descriptive Statistics --> Explore)

3.2 ไม่มีข้อมูลสุดโต่ง (Outliers) (ทดสอบโดยใช้ scatterplots ในคำสั่ง Graphs --> Legacy Dialogs --> scatterplots) (Hair et.al (2010: 67) แนะนำให้แปลงคะแนนเป็นค่ามาตรฐาน ถ้ากลุ่มตัวอย่างขนาดเล็ก (ไม่เกิน 80) ค่ามาตรฐาน ข้อมูลสุดโต่งมีค่ามาตรฐานตั้งแต่ 2.5 ถ้ากลุ่มตัวอย่างขนาดใหญ่ (มากกว่า 80) ข้อมูลสุดโต่ง ค่ามาตรฐานตั้งแต่ 4)

3.3 มีความแปรปรวนเท่ากันทั้งสองกลุ่ม (ทดสอบโดยใช้ Levene's test of equality of variances)

7. Biserial correlation

เป็นเทคนิคสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งที่มีมาตรวัดนามบัญญัติ (Nominal scale) แบบสองกลุ่มที่เกิดจากเกณฑ์ กับตัวแปรอีกตัวหนึ่งที่มีมาตรวัดอันดับหรืออัตราส่วน (Interval scale หรือ Ratio scale) เช่น ความสัมพันธ์ระหว่างการตอบข้อสอบข้อ 5 (ถูก/ผิด) กับคะแนนรวมของวิชา AA

$$\text{สูตร } r_{bi} = \frac{(\bar{y}_1 - \bar{y}_0)pq}{s_y}$$

โดยที่ $\bar{y}_1$ = ค่าเฉลี่ยของข้อมูลระหว่างตัวแปร y จากกลุ่มตัวแปร x = 1

$\bar{y}_0$ = ค่าเฉลี่ยของข้อมูลระหว่างตัวแปร y จากกลุ่มตัวแปร x = 0

p = สัดส่วนของคนที่อยู่ในกลุ่มตัวแปร x = 1

q = สัดส่วนของคนที่อยู่ในกลุ่มตัวแปร x = 0

s_y = ส่วนเบี่ยงเบนมาตรฐานของข้อมูลจากตัวแปร y ทั้งหมด

8. Rank biserial correlation

Rank-biserial correlation เป็นเทคนิคทางสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งที่มีมาตรวัดนามบัญญัติ (Nominal scale) แบบสองกลุ่มตามธรรมชาติ (dichotomous) กับตัวแปรที่มีมาตรวัดเรียงอันดับ เช่น การทำการบ้าน (ทำ/ไม่ทำ) กับอันดับที่ของคะแนนที่ได้

$$\text{สูตร } r_{rb} = \frac{2}{n}(\bar{y}_1 - \bar{y}_0)$$

เมื่อ $\bar{y}_1$ คือ ค่าเฉลี่ยอันดับของตัวแปร y จากกลุ่มตัวแปร x=1

$\bar{y}_0$ คือ ค่าเฉลี่ยอันดับของตัวแปร y จากกลุ่มตัวแปร x=0

โปรแกรม SPSS ไม่มีคำสั่งคำนวณค่า Rank-biserial correlation coefficient โดยตรง แต่สามารถคำนวณค่า $\bar{y}_1$ และ $\bar{y}_0$ ได้โดยใช้คำสั่ง

Analyze --> Nonparametric Test --> 2 Independent Samples --> Mann-Whitney U
แล้วนำค่า $\bar{y}_1$ และ $\bar{y}_0$ ที่ได้แทนค่าในสูตร เพื่อหา Rank-biserial correlation coefficient

ตัวอย่าง นักเรียน 10 คน ทำการบ้าน 7 คน ไม่ทำการบ้าน 3 คน นำคะแนนสอบมาเรียงลำดับจากน้อยไปหามาก ได้ลำดับที่ ดังตาราง ผู้วิจัยต้องการหาความสัมพันธ์ระหว่าง การทำการบ้านกับลำดับที่ของคะแนนที่ได้

การทำการบ้าน (X)*	คะแนน	ลำดับที่ (Y)
1	1	1
0	6	2
1	8	3
1	10	4
1	11	5
0	20	6
0	25	7
1	26	8
1	34	9
1	35	10

หมายเหตุ * 0 = ไม่ทำการบ้าน 1 = ทำการบ้าน

Mann-Whitney Test

Ranks				
	dicho	N	Mean Rank	Sum of Ranks
ordinal	.00	3	5.00	15.00
	1.00	7	5.71	40.00
Total		10		

$$r_{rb} = \frac{2}{n} (5.71 - 5) = 0.142$$

9. Multiple Correlation

Multiple Correlation เป็นเทคนิคทางสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์ระหว่างตัวแปรชุดหนึ่งที่วัดในมาตราอันดับหรืออัตราส่วนกับตัวแปรหนึ่งตัวที่วัดในมาตราอันดับหรืออัตราส่วน เช่น การหาความสัมพันธ์ระหว่าง คะแนน IQ (X1) คะแนนเจตคติ (X2) และคะแนนเฉลี่ยสะสม (X3) กับผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ (Y)

การวิเคราะห์ **Multiple Correlation** สามารถใช้โปรแกรม SPSS ด้วยคำสั่ง

Analyze --> Regression --> Linear

Multiple Correlation Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.330 ^a	.109	.102	.841

10. Canonical Correlation

Canonical correlation เป็นเทคนิคที่เหมาะสมสำหรับการระบุความสัมพันธ์ระหว่างตัวแปร สองชุด (Sharma, 1996, p. 391) เช่น ความสัมพันธ์ระหว่าง คะแนน IQ (X1) คะแนนเจตคติ (X2) และคะแนนเฉลี่ยสะสม (X3) กับ ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ (Y1) วิชาวิทยาศาสตร์ (Y2) วิชาสังคมศึกษา (Y3) วิชาภาษาอังกฤษ (Y4)

การวิเคราะห์ Canonical correlation ด้วยโปรแกรม SPSS ต้องเขียนคำสั่งใน **SYNTAX** ดังนี้

File → New → SYNTAX

MANOVA X1 X2 X3 WITH Y1 Y2 Y3 Y4

/DISCRIM ALL ALPHA(1)

/PRINT=SIG(EIGEN DIM)

ตัวอย่างผลการวิเคราะห์

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	.14155	91.84495	91.84495	.35214	.12400
2	.00871	5.64904	97.49399	.09290	.00863
3	.00386	2.50601	100.00000	.06203	.00385

Canonical correlation

สรุป

สิ่งสำคัญในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร หรือกลุ่มของตัวแปรคือการเลือกใช้เทคนิคทางสถิติให้เหมาะสมกับมาตรวัดของตัวแปรที่ต้องการหาความสัมพันธ์ นักวิจัยต้องระบุมาตรวัดของตัวแปรให้ถูกต้อง เมื่อวิเคราะห์แล้วจะต้องแปลความหมายผลการวิเคราะห์ทั้งความมีนัยสำคัญของขนาดของความสัมพันธ์ และทิศทางของความสัมพันธ์

บรรณานุกรม

- จุฬาลงกรณ์มหาวิทยาลัย. (ม.ป.ป.). *ความสัมพันธ์ระหว่างตัวแปรและการทำนายตัวแปร*.
<http://pioneer.netserv.chula.ac.th/~jaimorn/b4st.htm> ค้นเมื่อ 1 มีนาคม 2559.
(ออนไลน์)
- รัชนิกุล ภิญโญภาณุวัฒน์ และคณะ. (2558). รายงานการวิจัยเรื่อง การวิเคราะห์ปัจจัยเชิงสาเหตุเพื่อพัฒนา
สมรรถนะกำลังคนรองรับโลกศตวรรษที่ 21. เอกสารอัดสำเนา.
- ลัดดาวลย์ เพชรโรจน์ สุภมาสอังสุโชติ อัจฉรา ชำนิประศาสน์. (2555). สถิติสำหรับการวิจัย และเทคนิคการใช้
SPSS. (ฉบับปรับปรุง ครั้งที่ 2) กรุงเทพฯ: เจริญดีมั่นคงการพิมพ์.
- อังคณา จันอุไร. (บรรณาธิการ) (2552). รณรงค์ตุ๊กตักตักไม่เกิน 95 เดซิเบลเอ *คลินิกไอเสีย* ปีที่ 12(3)
กันยายน, 5.
- Best, John W. (1977). *Research in Education*. (3rd ed). Englewood cliffs: N.J.: Prentice-Hall Inc.
- Forthofer, Ronald N., Lee, Eun Sul. and Hernandez, Michael. (2007). *Biostatistics: A Guide to
Design, Analysis, and Discovery*. 2nd.ed California: Elsevier Academic Press.
- Garrido, L.E., Abad, J.J. and Ponsoda, V. (2013). A new look at Horn's parallel analysis with
ordinal variables. *Psychological Methods*. 18(4), 454-474.
- Glass, Gene V., and Hopkins, Kenneth D. (1984) *Statistical Methods in Education and
Psychology*. (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Hair, Joseph F., Black, William C., Babin, Barry J. and Anderson, Rolph E. (2010). *Multivariate
Data Analysis*. 7th.ed. NJ: Pearson.
- Holgado-Tello, F.P., Chacón-Moscoso, S., Barbero-García, I. and Vila-Abad E. (2010).
Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis
of ordinal variables. *Quality and Quantity*. 44(1), 153-166.
- Howell, David C. (2007) *Statistical Method for Psychology*. 6th.ed. CA:Thomson Wadsworth.
- Natalia Ospina-Alvarez, Francesc Piferrer. (2008). "Temperature-dependent sex determination
in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of
climate change." *PLoS One*. Published: July 30.
- Sharma. (1996). *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc.
<http://www.harding.edu/sbreezeel/460%20files/statbook/chapter15.pdf> (25/2/16)
[PDF] CHAPTER 15 NOMINAL MEASURES OF CORRELATION ...
www.harding.edu/sbreezeel/460%20files/.../chapter15.pdf