

ตัวแปรดัมมี่ (Dummy Variable)

ฉัตรศิริ ปิยะพิมลสิทธิ์*

ในงานวิจัยทางศึกษาศาสตร์ สังคมศาสตร์ หรือพฤติกรรมศาสตร์ ผู้วิจัยจะต้องคำนึงถึงระดับการวัดของตัวแปรที่ศึกษาเพื่อจะได้เลือกใช้สถิติอ้างอิงที่เหมาะสม ซึ่งสถิติมีอยู่ด้วยกัน 2 ประเภท คือ

1. สถิติพารามิเตอร์ (Parametric Statistics) เป็นวิธีการทางสถิติที่จะต้องเป็นไปตามข้อตกลงเบื้องต้น 3 ประการดังนี้

1) ข้อมูลที่เก็บรวบรวมได้จะต้องอยู่ในระดับอันตรภาคชั้นขึ้นไป (Interval Scale)

2) ข้อมูลที่เก็บรวบรวมได้จากกลุ่มตัวอย่างจะต้องมีการแจกแจงเป็นโค้งปกติ

3) กลุ่มประชากรแต่ละกลุ่มที่นำมาศึกษาจะต้องมีความแปรปรวนเท่ากัน สถิติที่อยู่ในประเภทนี้ เช่น t-test, Z-test, ANOVA, Regression Analysis ฯลฯ

2. สถิติไร้พารามิเตอร์ (Nonparametric Statistics) เป็นวิธีการทางสถิติที่สามารถนำมาใช้ได้โดยปราศจากข้อตกลงเบื้องต้นทั้ง 3 ประการข้างต้น สถิติที่อยู่ในประเภทนี้ เช่น Chi-square, Median test, Sign test ฯลฯ

จากข้อตกลงเบื้องต้นข้อแรกของสถิติพารามิเตอร์ที่ว่า ข้อมูลที่เก็บรวบรวมได้จะต้องอยู่ในระดับอันตรภาคชั้นขึ้นไป นั่นคือตัวแปรจะต้องถูกวัดในเชิงปริมาณ แต่มีงานวิจัยบางเล่มที่ใช้สถิติพารามิเตอร์ทั้งที่ตัวถูกวัดในระดับนามบัญญัติ ซึ่งเป็นการฝ่าฝืนข้อตกลงเบื้องต้น

ตัวแปรนามบัญญัติหรือตัวแปรเชิงคุณลักษณะ เป็นตัวแปรที่แสดงให้เห็นถึงความแตกต่างในเชิงคุณลักษณะ เช่น เพศ, ระดับการศึกษา, อาชีพ, สถานภาพสมรส, ศาสนา ฯลฯ ซึ่งตัวแปรนามบัญญัติเหล่านี้จะประกอบไปด้วยกลุ่มจำนวน g กลุ่ม เมื่อ g แทนจำนวนของกลุ่มย่อยในตัวแปรนามบัญญัติ เช่น ตัวแปรเพศ ประกอบไปด้วยเพศชายและเพศหญิง ดังนั้นจำนวนกลุ่มของตัวแปรเพศคือ 2 ($g = 2$) วิธีการนำตัวแปรนามบัญญัติเหล่านี้ไปวิเคราะห์ด้วยสถิติพารามิเตอร์เช่น สหสัมพันธ์เพียร์สัน หรือการวิเคราะห์การถดถอย ก็ย่อมเป็นการละเมิดข้อตกลงเบื้องต้น แต่ก็ไม่ใช่ว่าจะวิเคราะห์ไม่ได้ วิธีการนำตัวแปรเชิงคุณลักษณะเหล่านี้ไปวิเคราะห์ด้วยสถิติพารามิเตอร์นั้นจำเป็นที่จะต้องมีการแปลงให้เป็นตัวแปรเชิงปริมาณเสียก่อน กรณีที่ต้องการวัดตัวแปรศาสนา ปัญหาอยู่ตรงที่ว่า “เราจะให้คะแนนตัวแปรศาสนาได้อย่างไร”

* กศ.ม. (การวัดผลการศึกษา) <http://www.watpon.com>

คำตอบสำหรับปัญหานี้ก็คือการนำตัวแปรเชิงคุณลักษณะมาดำเนินการลงรหัสดัมมี่ (Dummy Coding) ตัวแปรที่ได้ใหม่จะเรียกว่า ตัวแปรดัมมี่ (Dummy Variable) ตำราบางเล่มใช้ว่า “ตัวแปรหุ่น” สมมติว่าได้ดำเนินการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างที่มีศาสนาแตกต่างกัน g ศาสนา ดังนั้นจะสามารถแปลงเป็นตัวแปรดัมมี่ที่แตกต่างกันได้ $g - 1$ ตัวแปร ซึ่งแต่ละตัวแปรจะแสดงความแตกต่างระหว่างกลุ่ม g กลุ่ม เช่น ตัวแปรศาสนา ประกอบด้วย 4 กลุ่มย่อย ($g = 4$) คือศาสนาโปรเตสแตนต์ (Protestant) ศาสนาคาทอลิก (Catholic) ศาสนายิว (Jewish) และศาสนาอื่น ๆ (Other) ดังนั้นจะได้ตัวแปรดัมมี่ k_g ตัว $k_g = g - 1 = 3$ ตัวแปร ในที่นี้สมมติว่าเป็นตัวแปร X_1, X_2 และ X_3 ซึ่งตัวแปรแต่ละตัวจะแสดงข้อมูลในแต่ละกลุ่มศาสนา

การลงรหัสดัมมี่

การลงรหัสดัมมี่ เป็นการลงรหัสตัวแปร 1 ตัวแปรที่แบ่งเป็น g กลุ่ม ซึ่งจะได้ตัวแปรดัมมี่ $g - 1$ ตัวแปร ในกรณีของตัวแปรศาสนาที่แบ่งเป็น 4 ศาสนา สามารถลงรหัสดัมมี่ได้ 4 วิธี ดังแสดงในตาราง 1 เป็นตารางแสดงวิธีการลงรหัสดัมมี่ พิจารณาที่ตัวแปร X_1 จะแทนศาสนาโปรเตสแตนต์ ถ้ากลุ่มตัวอย่างนับถือศาสนาโปรเตสแตนต์ (เป็นสมาชิกของกลุ่ม G1) จะให้ 1 คะแนน ถ้าไม่ใช่ (ไม่เป็นสมาชิกของกลุ่ม G1) จะให้ 0 คะแนน ในตัวแปร X_1 นี้ความแตกต่างของทั้ง 2 คุณลักษณะจะถูกใช้แค่ 0 และ 1 ซึ่งตัวแปร X_2 แทนศาสนา катоลิก และ X_3 แทนศาสนายิว ก็ให้คะแนนเป็น 0 และ 1 เช่นเดียวกัน

ตาราง 1 แสดงการลงรหัสดัมมี่

กลุ่ม	ศาสนา	X_1	X_2	X_3
G1	โปรเตสแตนต์	1	0	0
G2	คาทอลิก	0	1	0
G3	ยิว	0	0	1
G4	อื่น ๆ	0	0	0

$Y =$ เจตคติต่อการทำแท้ง

ตาราง 2 แสดงข้อมูลที่ได้จากการลงรหัสดัมมี่และคะแนนตัวแปร Y ของกลุ่มตัวอย่าง

คนที่	กลุ่ม	Y	X ₁	X ₂	X ₃
1	G ₂	61	0	1	0
2	G ₄	78	0	0	0
3	G ₁	47	1	0	0
4	G ₂	65	0	1	0
5	G ₂	45	0	1	0
6	G ₄	106	0	0	0
7	G ₁	120	1	0	0
8	G ₂	49	0	1	0
9	G ₄	45	0	0	0
10	G ₄	62	0	0	0
11	G ₂	79	0	1	0
12	G ₄	54	0	0	0
13	G ₁	140	1	0	0
14	G ₂	52	0	1	0
15	G ₁	88	1	0	0
16	G ₂	70	0	1	0
17	G ₂	56	0	1	0
18	G ₃	124	0	0	1
19	G ₄	98	0	0	0
20	G ₂	69	0	1	0
21	G ₁	56	1	0	0
22	G ₃	135	0	0	1
23	G ₁	64	1	0	0
24	G ₁	130	1	0	0
25	G ₃	74	0	0	1
26	G ₄	58	0	0	0
27	G ₁	116	1	0	0
28	G ₄	60	0	0	0
29	G ₃	84	0	0	1
30	G ₁	68	1	0	0
31	G ₁	90	1	0	0
32	G ₁	112	1	0	0
33	G ₃	94	0	0	1
34	G ₁	80	1	0	0
35	G ₃	110	0	0	1
36	G ₁	102	1	0	0

ดังนั้นเราจะมี $k_g = 3$ ($k_g = g - 1 = 4 - 1 = 3$) จากตัวแปรศาสนาที่เป็นนอกเป็น 4 กลุ่มศาสนา ในกลุ่มศาสนาอื่น ๆ นั้นอาจจะแสดงในตัวแปร X_4 แต่ก็ไม่จำเป็น เพราะเป็นการฟุ่มเฟือย ซึ่งกลุ่มศาสนาอื่น ๆ จะถูกรวมแสดงในตัวแปร X_1 , X_2 และ X_3 อยู่แล้ว คือคะแนน 0, 0 และ 0 ตามลำดับ มีความหมายว่า ไม่ใช่โปรเตสแตนต์ ไม่ใช่คาทอลิก และไม่ใช่นิกายอื่น

จะยกตัวอย่างของตัวแปรดัมมี่ให้ชัดเจนขึ้น เช่น ตัวแปรเพศที่มี 2 กลุ่ม ($g = 2$) ถ้ากลุ่มตัวอย่างเป็นเพศหญิงให้คะแนน 1 และถ้าเป็นเพศชายให้คะแนน 0 ดังนั้น X_1 จะถูกนิยามให้เป็นเพศหญิง และไม่มีตัวแปร X_2 แทนเพศชาย แต่ถ้ามีตัวแปร X_2 ก็จะต้องตรงข้ามกับ X_1 และสหสัมพันธ์ระหว่าง X_1 กับ X_2 (r_{12}) ควรจะได้ค่า -1.00 และความสัมพันธ์ของตัวแปรอื่น ๆ กับตัวแปร X_2 ก็ควรจะได้เครื่องหมายตรงกันข้ามเมื่อตัวแปรนั้นมาสัมพันธ์กับ X_1 นั่นคือ $r_{y1} = -r_{y2}$

ตาราง 2 เป็นข้อมูลที่เก็บรวบรวมได้จากกลุ่มตัวอย่าง 36 คน เมื่อแต่ละคนในกลุ่มตัวอย่างอยู่ในกลุ่มใดกลุ่มหนึ่งของตัวแปรศาสนา ขนาดของกลุ่มตัวอย่างในแต่ละกลุ่มคือ $n_1 = 13$, $n_2 = 9$, $n_3 = 6$ และ $n_4 = 8$ ขนาดของกลุ่มตัวอย่างแต่ละกลุ่มจึงจะไม่เท่ากัน ซึ่งตรงกับสภาพจริงของการทำวิจัย จากตาราง 2 จะเป็นชุดของตัวแปรดัมมี่ X_1 , X_2 , X_3 ที่มีการลงรหัสคะแนนตามแบบแผนของตาราง 1 สังเกตกลุ่มตัวอย่างที่ 3, 7, 13 และกลุ่มตัวอย่างอื่น ๆ จะมีแบบแผนเหมือน G1 นั่นคือ ตัวแปร X_1 , X_2 และ X_3 มีคะแนนเป็น 1, 0 และ 0 ตามลำดับ สังเกตกลุ่มตัวอย่างที่ 1, 4, 5 และกลุ่มตัวอย่างอื่น ๆ จะมีแบบแผนเหมือน G2 คือมีคะแนนเป็น 0, 1 และ 0 ตามลำดับ สังเกตกลุ่มตัวอย่างที่ 18, 22, 25 และกลุ่มตัวอย่างอื่น ๆ จะมีแบบแผนเหมือน G3 คือมีคะแนนเป็น 0, 0 และ 1 ตามลำดับ และกลุ่มตัวอย่างที่ 2, 6, 9 และกลุ่มตัวอย่างอื่น ๆ มีแบบแผนเหมือน G4 คือมีคะแนนเป็น 0, 0 และ 0 ตามลำดับ

ตัวแปรที่ปรากฏในตาราง 2 ถือว่าเป็นตัวแปรเชิงปริมาณ ผู้วิจัยสามารถนำมาคำนวณหาสถิติมีพารามิเตอร์ต่าง ๆ ได้ เช่น หาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน สหสัมพันธ์เพียร์สัน และยังสามารถใช้คำนวณสหสัมพันธ์พหุคูณ สหสัมพันธ์แยกส่วน สหสัมพันธ์ครึ่งส่วน ตลอดจนการวิเคราะห์การถดถอยได้ด้วย

ความสัมพันธ์ของตัวแปรดัมมี่กับตัวแปร Y

ตัวแปรดัมมี่ X_i แต่ละตัวเป็นตัวแปรแบ่ง 2 (dichotomous) เช่น ตัวแปร X_1 ให้คะแนน 1 ถ้าเป็นโปรเตสแตนต์ และให้ 0 ถ้าไม่ใช่โปรเตสแตนต์ สามารถใช้ในการคำนวณหาสหสัมพันธ์แบบเพียร์สัน (รวมทั้งพอยท์ไบซีเรียล) ระหว่างตัวแปรดัมมี่กับตัวแปรตาม ค่า r_{yi} จะแสดงในสดมภ์ที่ 2 ของตาราง 3 ดังนั้นศาสนาโปรเตสแตนต์จะมีความสัมพันธ์กับตัวแปรตาม $Y = .318$ ในทำนองเดียวกัน ศาสนาคาทอลิก มีความสัมพันธ์กับตัวแปรตาม $Y = -.442$ และศาสนานิกายอื่นมีความสัมพันธ์กับตัวแปรตาม $Y = .355$ เครื่องหมายของค่าสหสัมพันธ์เป็นการบอกทิศทางของความสัมพันธ์ คะแนนเฉลี่ยของตัวแปร Y คือค่า $m_y = 81.69$ และส่วนเบี่ยงเบน

มาตรฐานคือ 27.49

ตาราง 3 แสดงค่าสหสัมพันธ์ คะแนนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของตัวแปรดัมมี่ และตัวแปร Y

	r				r_{yi}^2	t_i (df = 34)	
	Y	X_1	X_2	X_3			
Y	1.000	.318	-.442	.355	-	-	
X_1	.318	1.000	-.434	-.336	.1011	1.955	
X_2	-.442	-.434	1.000	-.258	.1954	-2.874**	
X_3	.355	-.336	-.258	1.000	.1260	2.214*	
m	81.69	.361	.250	.167			
sd	27.49	.480	.433	.373	n = 36		
$R_{Y.123}^2 = .354944$		$F = 5.869^{**}$ (df = 3, 32)					
$\tilde{R}_{Y.123}^2 = .294470$							

*P < .05, **P < .01

ค่าเฉลี่ยของตัวแปรดัมมี่ X_i คือ สัดส่วนของสมาชิกใน G_i กับกลุ่มตัวอย่างทั้งหมด

$$\text{ดังนั้น } m_i = \bar{X}_i = \frac{\sum X_i}{n} = \frac{n_i}{n} = P_i$$

$n_1 = 13, n = 36$ ดังนั้น $n_1/n = P_1 = 13/36 = .361$ และ sd_i จะได้

$$sd_i = \sqrt{P_i - P_i^2} = \sqrt{P_i(1 - P_i)}$$

$$m_1 = \bar{X}_i = P_1 = .361$$

$$sd_1 = \sqrt{.361(1-.361)} = .480$$

จะเห็นได้ว่า sd_i ของตัวแปรดัมมี่ขึ้นอยู่กับสัดส่วนของกลุ่มตัวอย่าง ค่าของ sd_i มีค่าสูงสุดแค่ .50 เมื่อ $P_i = .50$;

$$sd_{\max} = \sqrt{.5(1-.5)} = \sqrt{.25} = .5$$

จากข้อมูลที่ได้มาจากกลุ่มตัวอย่าง $n = 36$ จากประชากรทั้งหมด โดยดำเนินการสอบวัดตัวแปรศาสนา และวัดเจตคติต่อการทำแท้ง สมมติข้อมูลเก็บรวบรวมได้ดังตาราง 2 P_i แทนคะแนนเฉลี่ยของตัวแปรศาสนาในกลุ่มตัวอย่างแต่ละกลุ่ม เช่น กลุ่มศาสนายิว (G_3) มี $P_3 = .167$ ในตาราง 3 ได้ค่า $r_{y3}^2 = .1260$ แปลความหมายได้ว่า การนับถือศาสนายิวจะมีผลต่อความแปรปรวนในเจตคติต่อการทำแท้ง 12.6% เปอร์เซนต์ สำหรับตัวแปรอื่น ๆ ก็แปลความหมายเช่นเดียวกัน

เราสามารถทดสอบนัยสำคัญของ r_{yi} โดยใช้ t-test ดังสมการ

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

ทดสอบนัยสำคัญของตัวแปร X_2 กับตัวแปร Y จะได้

$$t = \frac{-.442\sqrt{34}}{\sqrt{1-(-.442)^2}} = -2.874$$

กับ $df = n - 2 = 34$ มีนัยสำคัญทางสถิติที่ .01

นั่นคือความสัมพันธ์ของศาสนาคาทอลิกกับคะแนนเจตคติต่อการทำแท้งไม่เป็น 0

ในตาราง 3 จะเสนอความสัมพันธ์ระหว่างตัวแปร Y กับตัวแปรดัมมี่ X_1, X_2, X_3 เท่านั้น แต่ยังมีอีกกรณีหนึ่งที่ยังไม่ได้หา ก็คือกรณีที่ไม่นับถือศาสนาทั้ง 3 หรือศาสนาอื่น ๆ สามารถคำนวณหาความสัมพันธ์ระหว่างตัวแปร Y กับศาสนาอื่น ๆ ได้โดยใช้สูตร

$$r_{yg} = -\frac{\sum r_{yi}sd_i}{sd_g} = -\frac{\sum r_{yi}\sqrt{P_i(1-P_i)}}{\sqrt{P_g(1-P_g)}}$$

($P_g = n_g/n = 8/36 = .222$)

$$\begin{aligned} r_{yg} &= -\frac{.318(.480) + (-.442)(.433) + .355(.373)}{\sqrt{.222(1-.222)}} \\ &= -\frac{.09367}{.4157} \\ &= -.225 \end{aligned}$$

และอาจทดสอบระดับนัยสำคัญโดยใช้สูตร t-test ที่ให้ไว้ข้างต้น

ความสัมพันธ์ระหว่างตัวแปรดัมมี่

ในหัวข้อที่แล้วเป็นการหาความสัมพันธ์ระหว่างตัวแปร Y กับตัวแปรดัมมี่ X_1, X_2, X_3 ได้ค่าเป็น .318, .442 และ .355 ตามลำดับ ดังตาราง 3 ในหัวข้อนี้จะขอเสนอการหาความสัมพันธ์ระหว่างตัวแปรดัมมี่ สมมติว่าตัวแปร X_1, X_2, X_3 เป็นอิสระจากกันแล้ว $r_{12} = r_{13} = r_{23} = 0$ แต่มันก็เป็นไปไม่ได้ ดังนั้นขนาดของความสัมพันธ์ระหว่างตัวแปรดัมมี่ 2 ตัว (X_i, X_j) ซึ่งมีจำนวนของกลุ่มตัวอย่างในแต่ละกลุ่มคือ n_i และ n_j และมีกลุ่มตัวอย่างรวมทั้งหมด n กลุ่ม สามารถคำนวณหาความสัมพันธ์ได้จากสูตร

$$\begin{aligned} r_{ij} &= -\sqrt{\frac{n_i n_j}{(n - n_i)(n - n_j)}} \\ &= -\sqrt{\frac{P_i P_j}{(1 - P_i)(1 - P_j)}} \end{aligned}$$

ดังนั้นความสัมพันธ์ระหว่างศาสนาโปรเตสแตนต์และคาทอลิกจากข้อมูลในตาราง 2

คือ

$$r_{12} = -\sqrt{\frac{(.361)(.250)}{(1-.361)(1-.250)}} = -.434$$

นั่นคือศาสนาโปรเตสแตนต์มีความสัมพันธ์กับศาสนาคาทอลิก = -.434 เครื่องหมายลบของค่าสหสัมพันธ์เป็นเครื่องหมายที่บ่งบอกถึงทิศทางของความสัมพันธ์ สำหรับความสัมพันธ์ระหว่างศาสนาอื่น ๆ ก็คำนวณเช่นเดียวกัน

สหสัมพันธ์พหุคูณของตัวแปรดัมมี่

ในตาราง 3 $R^2_{Y.G} = R^2_{Y.123} = .3549$ ดังนั้นค่าสหสัมพันธ์พหุคูณ $R_{Y.G} = R_{Y.123} = .596$ แปลความหมายได้ว่า 35.5 เปอร์เซ็นต์ของความแปรปรวนในคะแนนเจตคติต่อการทำแท้ง มีความสัมพันธ์กับศาสนาของกลุ่มตัวอย่าง หรือค่าสหสัมพันธ์ของเจตคติต่อการทำแท้งกับศาสนาคือ .596

เราสามารถทดสอบนัยสำคัญของ R^2 ได้โดยใช้ F-test

$$F = \frac{R^2(n-k-1)}{(1-R^2)k}$$

เมื่อ k คือจำนวนของตัวแปรอิสระ $k = k_g = g - 1$

$$F = \frac{R^2(n-g)}{(1-R^2)(g-1)} = \frac{.354944(36-4)}{(1-.354944)(3)} = 5.869$$

$df = 3, 32$, F มีนัยสำคัญที่ระดับ .01 นั่นคือมีความสัมพันธ์กันจริงระหว่างตัวแปรเจตคติต่อการทำแท้งกับศาสนา

เนื่องจากค่า R^2 มีความลำเอียงที่ขึ้นอยู่กับขนาดของ k และ n ดังนั้นความสัมพันธ์ที่แท้จริงระหว่างตัวแปร Y กับตัวแปรดัมมี่ที่ถูกต้องจะต้องใช้สูตรหาสหสัมพันธ์ในการประมาณค่าที่ไม่ลำเอียง คือสูตร

$$\tilde{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

จะได้ $\tilde{R}^2 = 1 - (1 - .3549) \frac{35}{32} = .2945$

ดังนั้นสัดส่วนของความแปรปรวนในเจตคติต่อการทำแท้ง อธิบายได้ด้วยตัวแปรศาสนา 29.45%

สหสัมพันธ์แยกส่วน (Partial Correlation)

เราจะใช้สัญลักษณ์ pr_i แทนค่าสหสัมพันธ์ของตัวแปร Y กับ X_i ที่ขจัดอิทธิพลของตัวแปรอื่น ๆ ออก ดังตาราง 4 pr_1 มีค่า .363 ซึ่งก็คือความสัมพันธ์ระหว่างตัวแปร Y กับ X_1 ที่ขจัดอิทธิพลของตัวแปร X_2 และ X_3 ออก $pr_2 = -.145$ คือความสัมพันธ์ของตัวแปร Y กับ X_2 เมื่อ

ขจัดอิทธิพลของ X_1 กับ X_3 ออก และ $pr_3 = .423$ คือความสัมพันธ์ของตัวแปร Y กับ X_3 เมื่อขจัดอิทธิพลของ X_1 และ X_2 ออก

ตาราง 4 แสดงค่าสหสัมพันธ์แยกส่วน ค่าสหสัมพันธ์ครึ่งส่วนและสัมประสิทธิ์การถดถอยของตัวแปรดัมมี่

X_i	pr_i	pr_i^2	sr_i	sr_i^2	B_i	$t_i(df=32)$
X_1	.363	.1318	.313	.0980	23.18	2.203*
X_2	-.145	.0210	-.118	.0139	-9.46	-0.831
X_3	.423	.1789	.375	.1406	33.37	2.639*

$Y' = B_1X_1 + B_2X_2 + B_3X_3 + A$ $A = 70.13$
 $= 23.18 X_1 - 9.46 X_2 + 33.37 X_3 + 70.13$
 $Y_1' = 23.18(1) - 9.46(0) + 33.37(0) + 70.13 = 93.31 = \bar{Y}_1$
 $Y_2' = 23.18(0) - 9.46(1) + 33.37(0) + 70.13 = 60.67 = \bar{Y}_2$
 $Y_3' = 23.18(0) - 9.46(0) + 33.37(1) + 70.13 = 103.50 = \bar{Y}_3$
 $Y_4' = 23.18(0) - 9.46(0) + 33.37(0) + 70.13 = 70.13 = \bar{Y}_4$
 $sd_{Y-Y'}^2 = sd_Y^2(1 - R^2) \frac{n}{n-k-1} = 27.49^2(1-.354944) \frac{36}{36-3-1} = 548.41$

*P < .05

สหสัมพันธ์ครึ่งส่วน (Semipartial Correlation)

เป็นความสัมพันธ์ระหว่างตัวแปร Y กับตัวแปร X_i ซึ่งตัวแปร X_i ถูกขจัดอิทธิพลของตัวแปรอื่น ๆ ออก ใช้สัญลักษณ์ sr_i มีสูตรการคำนวณดังนี้

$$sr_i^2 = R_{Y.123\dots k}^2 - R_{Y.123\dots(i)\dots k}^2$$

(i) เป็นตัวแปรที่ i ซึ่งเราจะไม่นำเข้ากระบวนการคำนวณ นั่นคือ

$$sr_1^2 = R_{Y.123}^2 - R_{Y.23}^2$$

$$sr_2^2 = R_{Y.123}^2 - R_{Y.13}^2$$

$$sr_3^2 = R_{Y.123}^2 - R_{Y.12}^2$$

ผลลัพธ์ที่ได้แสดงในตาราง 4

$sr_1^2 = .0980$ หมายถึง ความสัมพันธ์ระหว่างตัวแปร Y กับตัวแปร X_1 ซึ่งตัวแปร X_1

ถูกขจัดอิทธิพลของตัวแปร X_2 และ X_3 ออกแล้ว มีค่า 9.8 เปอร์เซนต์ที่จะใช้อธิบายความแปรปรวนของตัวแปร Y

$sr_2^2 = .0139$ หมายถึงความสัมพันธ์ระหว่างตัวแปร Y กับตัวแปร X_2 ซึ่งตัวแปร X_2 ถูกขจัดอิทธิพลของตัวแปร X_1 และ X_3 ออกแล้ว มีค่า 1.39 เปอร์เซนต์ที่จะใช้อธิบายความแปรปรวนของตัวแปร Y

$sr_3^2 = .1406$ หมายถึงความสัมพันธ์ระหว่างตัวแปร Y กับตัวแปร X_3 ซึ่งตัวแปร X_3 ถูกขจัดอิทธิพลของตัวแปร X_1 และ X_2 ออกแล้ว มีค่า 14.06 เปอร์เซนต์ที่จะใช้อธิบายความแปรปรวนของตัวแปร Y

สัมประสิทธิ์การถดถอยและสมการถดถอย (The Regression Coefficients and The Regression Equation)

ในความหมายโดยทั่วไปแล้ว สัมประสิทธิ์การถดถอยของคะแนนดิบ (B_i) คือค่าที่บ่งบอกถึงปริมาณในการเปลี่ยนแปลงของตัวแปร Y แสดงให้เห็นว่า เมื่อ X_i เปลี่ยนแปลงไป 1 หน่วยแล้ว ตัวแปร Y จะเปลี่ยนแปลงไปเท่าไร? เนื่องจาก X_i เป็นตัวแปรดัมมี่มีค่าเป็น 0 และ 1 เช่น ไม่ใช่ค่าทอลิค (ให้คะแนน 0) กับใช้คาทอลิค (ให้คะแนน 1) ดังนั้น $B_2 = -9.46$ จากตาราง 4 หมายความว่า ถ้า X_2 เป็นคาทอลิคแล้ว ($X_2 = 1$) ตัวแปร Y จะเปลี่ยนแปลงไป -9.46

พิจารณาจากสมการถดถอย

$$Y' = B_1X_1 + B_2X_2 + B_3X_3 + A$$

เมื่อแทนค่า B_i , X_i และ A แล้ว ผลที่ได้ก็คือ \bar{Y}_i คือค่าเฉลี่ยของตัวแปรเจตคติต่อการทำแท้งในแต่ละกลุ่มศาสนา

ในกลุ่มศาสนาอื่น ๆ ที่ให้เป็น 0, 0, 0 ในตัวแปรดัมมี่แต่ละตัว จะมีคะแนนเฉลี่ย Y เท่ากับ 70.13 หรือ

$$\bar{Y}_g = A$$

ดังนั้นคะแนนเฉลี่ย Y ของกลุ่มศาสนาอื่น ๆ เขียนเป็นสมการได้ว่า

$$\begin{aligned}\bar{Y}_i &= A + B_i \\ &= \bar{Y}_g + B_i\end{aligned}$$

ถ้าย้ายข้างสมการจะเห็นความแตกต่างระหว่างคะแนนเฉลี่ย Y ของกลุ่ม G_i กับกลุ่ม G_g

$$B_i = Y_i - \bar{Y}_g$$

ดังนั้นศาสนายิว $B_3 = 33.37$ ซึ่งก็คือคะแนนเฉลี่ยของตัวแปรเจตคติต่อการทำแท้งในกลุ่มคนศาสนายิว (G_3) สูงกว่าคนกลุ่มศาสนาอื่น ๆ (G_4) 33.37 คะแนน

ต่อมาดูความแตกต่างของ B_i 2 กลุ่ม กับความแตกต่างของค่าเฉลี่ย Y

$$B_i = Y_i' - Y_g' \quad \text{และ} \quad B_j = Y_j' - Y_g'$$

$$\text{ดังนั้น} \quad \bar{Y}_i - \bar{Y}_j = B_i - B_j$$

ตัวอย่างความแตกต่างของคะแนนเฉลี่ยเจตคติต่อการทำแท้งระหว่างศาสนาโปรเตสแตนต์และคาทอลิค คือ $23.18 - (-9.46) = 32.64$

การทดสอบความแตกต่างระหว่างค่าเฉลี่ย

$$\text{จากสมการ} \quad \bar{Y}_i - \bar{Y}_j = B_i - B_j$$

การทดสอบความแตกต่างระหว่างค่าเฉลี่ย $Y_i' - Y_j'$ เท่ากับการทดสอบความแตกต่างระหว่าง $B_i - B_j$ สามารถทดสอบได้โดยใช้ t-test ดังสูตร

$$t = \frac{B_i - B_j}{\sqrt{\tilde{sd}_{Y-Y'}^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \quad ; \quad df = n - k - 1$$

$$\text{เมื่อ} \quad \tilde{sd}_{Y-Y'}^2 = sd_Y^2 (1 - R^2) \frac{n}{n - k - 1}$$

$\tilde{sd}_{Y-Y'}^2$ เป็นการประมาณค่าที่ไม่ลำเอียงของความแปรปรวน

ดังนั้นการทดสอบความแตกต่างของคะแนนเจตคติต่อการทำแท้งระหว่างกลุ่มศาสนา

โปรเตสแตนต์กับคาทอลิก จะได้

$$\begin{aligned} \tilde{sd}_{Y-Y'}^2 &= (27.49)^2 (1 - .354944) \frac{36}{36 - 3 - 1} \\ &= 548.41 \end{aligned}$$

และแทนค่าในสูตร t-test จะได้

$$\begin{aligned} t &= \frac{23.18 - (-9.46)}{\sqrt{548.41 \left(\frac{1}{13} + \frac{1}{9} \right)}} \\ &= \frac{32.64}{10.16} \\ &= 3.213 \end{aligned}$$

■ $df = 32$ มีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นคะแนนเจตคติต่อการทำแท้งระหว่างกลุ่มโปรเตสแตนต์กับคาทอลิกมีความแตกต่างกันจริง

ตาราง 5 แสดงผลการวิเคราะห์ความแปรปรวน

แหล่งความแปรปรวน	SS	df	MS
ระหว่างกลุ่ม	9656.49	$g - 1 = 3$	3218.83
ภายในกลุ่ม	17549.15	$n - g = 32$	548.41
รวม	27205.64	$n - 1 = 35$	

G_i	n_i	\bar{Y}_i
G_1	13	93.31
G_2	9	60.67
G_3	6	103.50
G_4	8	70.13

$$F = \frac{BSS / (g - 1)}{WSS / (n - g)} = \frac{BMS}{WMS} = \frac{3218.83}{548.41} = 5.869 = F \text{ ของการทดสอบ } R^2$$

$$\eta_{Y.G}^2 = \frac{BSS}{TSS} = \frac{9656.49}{27205.64} = .354944 = R^2$$

การวิเคราะห์ความแปรปรวน

ในโครงสร้างของข้อมูลมีตัวแปรอิสระ ซึ่งแบ่งเป็น g ระดับ กับตัวแปรตาม Y สามารถนำมาวิเคราะห์ความแปรปรวนโดยใช้ One-way ANOVA ได้ จากข้อมูลในตาราง 2 นำมาวิเคราะห์หา Sums of Squares (SS) ของระหว่างกลุ่ม (BSS) ของภายในกลุ่ม (WSS) และรวม (TSS) และคำนวณหา Mean Squares (MS) ของระหว่างกลุ่ม (B) และภายในกลุ่ม (W) โดยหาร SS ด้วย df ผลที่ได้จะแสดงในตาราง 5 และหาค่า F เพื่อทดสอบสมมติฐาน

ผลลัพธ์จากตาราง จะเห็นได้ว่า $BSS = R^2(TSS)$ และ $WSS = (1-R^2)(TSS)$ และค่า F -test ที่คำนวณได้จะเท่ากับ F -test ของ R^2

อัตราส่วนความสัมพันธ์ (Correlation Ratio) คำนวณได้ด้วยสูตร

$$\eta_{Y.G}^2 = \frac{BSS}{TSS}$$

จากตาราง 5 $\eta_{Y.G}^2$ (Eta squared) = .354944 เท่ากับค่าของ $R_{Y.123\dots(g-1)}^2$

(จากตาราง 3 ได้ค่า $R_{Y.123}^2 = .354944$)

จะเห็นว่าผลลัพธ์ของการวิเคราะห์ความสัมพันธ์ทั้ง Regression Analysis และ ANOVA โดยใช้ตัวแปรดัมมี่ มีค่าสอดคล้องกัน 2 ประการคือ ประการแรก เราจะเห็นว่าค่าสถิติต่าง ๆ ของข้อมูลระดับนามบัญญัติที่ได้รับการลงรหัสดัมมี่ เมื่อวิเคราะห์ด้วยสถิติทั้งสองแล้วจะได้ผลสอดคล้องกัน ประการที่สอง รูปแบบของ ANOVA อื่น ๆ (เช่น Factorial, ANCOVA ฯลฯ) เมื่อนำมาเปรียบเทียบกับผลที่ได้จาก Regression Analysis ก็จะได้ผลสอดคล้องกัน



หนังสืออ้างอิง

- Cohen, Jacof and Cohen, Patricia. ***Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences***. Second Edition. London : Lawrence Erlbaum Associates, Publishers, 1983.
- Pedhazur, Elazar J. ***Multiple Regression in Behavioral Research : Explanation and Prediction***. Third Edition. U.S.A. Holt, Rinchart and Winston, Inc., 1997.