



บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

การวิเคราะห์ความสัมพันธ์ (Correlation analysis) เป็นการศึกษาเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป ว่ามีมากน้อยเพียงใด และมีทิศทางของความสัมพันธ์ไปทิศทางใด

การวัดความสัมพันธ์โดยวิธีทางสถิติมีอยู่หลายวิธี แตกต่างกันไป ขึ้นอยู่กับลักษณะของตัวแปรหรือมาตรวัดค่าตัวแปร ลักษณะข้อมูลเป็นแบบต่อเนื่องหรือแบบไม่ต่อเนื่อง และคุณลักษณะของการแจกแจงของข้อมูล ฯลฯ ซึ่งมาตรวัดความสัมพันธ์ระหว่างตัวแปร เรียกว่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร 2 ตัว ที่ใช้กันอย่างกว้างขวางคือ สัมประสิทธิ์สหสัมพันธ์เชิงเส้นแบบ Pearson (Pearson Product Moment Correlation Coefficient) ที่ใช้สัญลักษณ์ ρ ซึ่งคำนวณได้ดังนี้

$$\rho = E \left[\frac{(X - \mu_x)(Y - \mu_y)}{\sigma_x \sigma_y} \right]$$
$$= \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

โดยมีตัวประมาณค่าของ ρ คือ r ซึ่งคำนวณได้จาก

$$r = \frac{1}{n-1} E \left[\left(\frac{X - \bar{X}}{S_x} \right) \left(\frac{Y - \bar{Y}}{S_y} \right) \right]$$

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

เมื่อ ρ คือ สัมประสิทธิ์สหสัมพันธ์ของประชากรระหว่างตัวแปร X และ Y

r คือ สัมประสิทธิ์สหสัมพันธ์ของตัวอย่างระหว่างตัวแปร X และ Y

ΣX คือ ผลรวมของค่าตัวแปร X

ΣY คือ ผลรวมของค่าตัวแปร Y

ΣXY คือ ผลรวมของผลคูณระหว่างค่าตัวแปร X และตัวแปร Y

ΣX^2 คือ ผลรวมของกำลังสองของค่าตัวแปร X

ΣY^2 คือ ผลรวมของกำลังสองของค่าตัวแปร Y

N คือ จำนวนคู่ของค่าตัวแปร X และ Y จากประชากร

n คือ จำนวนคู่ของค่าตัวแปร X และ Y จากตัวอย่าง

σ_x คือ ส่วนเบี่ยงเบนมาตรฐานของประชากรของตัวแปร x

σ_y คือ ส่วนเบี่ยงเบนมาตรฐานของประชากรของตัวแปร y

S_x คือ ส่วนเบี่ยงเบนมาตรฐานของตัวอย่างของตัวแปร x

S_y คือ ส่วนเบี่ยงเบนมาตรฐานของตัวอย่างของตัวแปร y

โดยค่าของ ρ หรือ r จะอยู่ตั้งแต่ -1 และ $+1$ เสมอ ถ้า ρ หรือ r มีค่าเป็นบวกแสดงว่า ตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นในทิศทางเดียวกัน นั่นคือ ถ้าตัวแปรตัวหนึ่งมีค่าเพิ่มขึ้น (หรือลดลง) ตัวแปรอีกตัวหนึ่งจะมีค่าเพิ่มขึ้น (หรือลดลง) ด้วย แต่

ถ้ามีค่าเป็นลบ แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์กันในทิศทางตรงข้าม นั่นคือ ถ้าตัวแปรตัวหนึ่งมีค่าเพิ่มขึ้น (หรือลดลง) ตัวแปรอีกตัวหนึ่งจะมีค่าลดลง (หรือเพิ่มขึ้น) ด้วย ถ้าค่า ρ หรือ r เท่ากับศูนย์ แสดงว่าตัวแปรทั้งสองไม่มีความสัมพันธ์เชิงเส้นต่อกัน แต่ถ้าค่า ρ หรือ r เท่ากับ ± 1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์กันมาก หรือมีความสัมพันธ์กันอย่างสมบูรณ์ (Perfect correlation) ถ้าค่า ρ หรือ r มีค่าเข้าใกล้ $+1$ หรือ -1 แสดงว่าขนาดของความสัมพันธ์มีค่าสูง และถ้าค่า ρ หรือ r เข้าใกล้ 0 แสดงว่าขนาดของความสัมพันธ์มีค่าต่ำ

ในการวิเคราะห์ความสัมพันธ์ บางครั้งเราสามารถเก็บข้อมูลข้ามตัวแปรตัวหนึ่งได้ ทำให้เกิดตัวแปรที่มีการวัดซ้ำและตัวแปรที่ไม่มีการวัดซ้ำ เช่น เมื่อต้องการหาความสัมพันธ์ระหว่างน้ำหนักแม่หมูก่อนคลอดและน้ำหนักลูกหมูแรกเกิด โดยปกติแม่หมูแต่ละตัวจะคลอดลูกหมูมากกว่าหนึ่งตัวและอาจมีจำนวนไม่เท่ากัน ความสัมพันธ์ที่สนใจศึกษาจึงเป็นความสัมพันธ์ระหว่างกลุ่มตัวแปรที่มีการวัดซ้ำคือน้ำหนักลูกหมูแรกเกิด และกลุ่มตัวแปรที่ไม่มีการวัดซ้ำคือน้ำหนักแม่หมูก่อนคลอด ความสัมพันธ์ระหว่างกลุ่มนี้มีความสำคัญมากสำหรับการศึกษาด้านพันธุศาสตร์ (Genetics) ซึ่งมีความเกี่ยวข้องกับการถ่ายทอดลักษณะบางประการจากรุ่นหนึ่งไปสู่อีกรุ่นหนึ่ง เช่น ความดันเลือด โครงสร้างทางร่างกาย และน้ำหนัก เป็นต้น จึงมีผู้สนใจสร้างตัวประมาณค่าความสัมพันธ์ระหว่างกลุ่มในหลายรูปแบบ รวมทั้งมีการสร้างตัวสถิติที่ใช้ทดสอบความสัมพันธ์ดังกล่าว ดังนั้นผู้วิจัยจึงสนใจศึกษาและเปรียบเทียบสถิติทดสอบ เพื่อเลือกตัวสถิติทดสอบที่เหมาะสม เนื่องจาก การเลือกใช้สถิติทดสอบที่เหมาะสม จะทำให้ผลลัพธ์ของการวิจัยเป็นไปอย่างถูกต้อง ซึ่งมีผลให้งานวิจัยนั้นมีคุณภาพสูง ซึ่งในการพิจารณาความเหมาะสมของสถิติทดสอบนั้น สิ่งที่ต้องพิจารณาก็คือ อำนาจของการทดสอบ (Power of the test) และความแกร่ง (Robustness) ดังคำกล่าวของ Neyman (1950:265 อ้างโดย Direk Srisukho 1974:38) ซึ่งกล่าวว่า "เมื่อต้องการใช้สถิติทดสอบ เราต้องเริ่มพิจารณาถึงความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ก่อน แล้วจึงพิจารณาถึงความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2 โดยมีขั้นตอนดังนี้คือ ให้ความน่าจะเป็นที่จะยอมให้เกิดความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน α ที่กำหนดไว้ และเมื่อเป็นไปตามเงื่อนไขดังกล่าวแล้ว สิ่งที่จะใช้ในการพิจารณาเลือกสถิติทดสอบอีกก็คือ เลือกสถิติทดสอบที่มีโอกาสน้อยที่สุดที่จะยอมรับสมมติฐานว่าง H_0 เมื่อสมมติฐานว่าง H_0 นั้นไม่จริง ซึ่ง

หมายความว่าให้อำนาจการทดสอบสูงที่สุด" ดังนั้นการวิจัยครั้งนี้จึงได้ทำการศึกษาโดยอาศัยวิธีการจำลองแบบที่เรียกว่า วิธีมอนติคาร์โล (Monte Carlo methods) ซึ่งเป็นเทคนิคที่ทำให้ได้ผลสรุปจากสถานการณ์ที่ต้องการศึกษา

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบตัวสถิติที่ใช้ทดสอบความสัมพันธ์ระหว่างกลุ่ม 3 วิธี คือ

1.2.1 The Adjusted Pairwise Test (PAIR)

1.2.2 The Sib-Mean Test (MEAN)

1.2.3 The Ensemble Test (ENSE)

โดยการเปรียบเทียบความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 (Type I error) และอำนาจการทดสอบ (Power of the Test) ของตัวสถิติทดสอบทั้ง 3 วิธี

1.3 สมมติฐานของการวิจัย

1.3.1 เมื่อความสัมพันธ์ภายในกลุ่มมีค่าต่ำ ตัวสถิติทดสอบ PAIR จะมีอำนาจการทดสอบสูงกว่าสถิติทดสอบอื่น ๆ

1.3.2 เมื่อความสัมพันธ์ภายในกลุ่มมีค่าสูง ตัวสถิติทดสอบ ENSE จะมีอำนาจการทดสอบสูงกว่าสถิติทดสอบอื่น ๆ

1.4 ข้อตกลงเบื้องต้น

1.4.1 คำสั่งเกิดของตัวอย่างที่ α $X_{\alpha} = (X_{0\alpha}, X_{1\alpha}, \dots, X_{k\alpha, \alpha})'$
; $\alpha = 1, 2, \dots, N$ เป็นค่าต่อเนื่อง และมีการแจกแจงปกติพหุ (Multivariate Normal distribution)

1.4.2 ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นแบบเส้นตรง (Linear relationship)

1.4.3 $X_{\alpha} = (X_{0\alpha}, X_{1\alpha}, \dots, X_{k\alpha, \alpha})'$ เป็นตัวแปรเชิงสุ่ม

1.5. ขอบเขตของการวิจัย

1.5.1 ความสัมพันธ์ระหว่างกลุ่ม (Interclass correlation ; ρ_{ms})

$\rho_{ms} = 0.0, 0.2, 0.3, 0.4$ และ 0.7

1.5.2 ความสัมพันธ์ภายในกลุ่ม (Intraclass correlation ; ρ_{ss})

$\rho_{ss} = 0.1, 0.3, 0.35, 0.36, 0.365, 0.4, 0.5$ และ 0.7

1.5.3 จำนวนครั้งการวัดซ้ำของแต่ละตัวอย่างที่ α มีการแจกแจงแบบนินลัสทวินาม (Negative Binomial distribution ; $N(k,p)$; mean = kq/p) และมีค่าเฉลี่ยเท่ากับ 3, 5 และ 7

1.5.4 ขนาดตัวอย่าง $N = 15, 30$ และ 50

1.5.5 ระดับนัยสำคัญ $\alpha = 0.05$ และ 0.10

1.5.6 สร้างแบบจำลองข้อมูลให้มีสถานการณ์ตามต้องการศึกษาโดยใช้เทคนิคมอนติคาร์โล จากเครื่องคอมพิวเตอร์ IBM 370/3031 และทดลองซ้ำ 1,000 ครั้งในแต่ละสถานการณ์

1.6. คำจำกัดความ

1.6.1 ความคลาดเคลื่อนประเภทที่ 1 (Type I error) เป็นความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่าง (H_0) เมื่อสมมติฐานว่างจริง

1.6.2 ความคลาดเคลื่อนประเภทที่ 2 (Type II error) เป็นความผิดพลาดที่เกิดจากการยอมรับสมมติฐานว่าง เมื่อสมมติฐานว่างไม่จริง

1.6.3 อำนาจการทดสอบ (Power of the test) คือความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่างเมื่อสมมติฐานว่างไม่จริง

1.6.4 ความแกร่ง (Robustness) ของการทดสอบ หมายถึง คุณสมบัติของการทดสอบที่ไม่แสดงถึงความไว (Sensitive) ต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น การฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบนั้น และสิ่งที่ใช้พิจารณาความแกร่งของการทดสอบ คือความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

1.6.5 ความสัมพันธ์ระหว่างกลุ่ม (Interclass correlation) เป็นความสัมพันธ์ระหว่างกลุ่มตัวแปรที่มีการวัดซ้ำ และกลุ่มตัวแปรที่ไม่มีการวัดซ้ำ

1.6.6 ความสัมพันธ์ภายในกลุ่ม (Intraclass correlation) เป็นความสัมพันธ์ภายในกลุ่มของตัวแปรที่มีการวัดซ้ำ

1.7 ประโยชน์ของการวิจัย

1.7.1 เป็นแนวทางให้ผู้วิจัยได้เลือกใช้สถิติทดสอบที่เหมาะสม ในการทดสอบความสัมพันธ์ระหว่างกลุ่ม

1.7.2 เป็นแนวทางในการวิจัยเปรียบเทียบสถิติทดสอบอื่น ๆ ต่อไป