บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการวิเคราะห์ความถดถอยพหุคูณ (multiple regression analysis) เป็นการศึกษาความสัมพันธ์ของตัวแปร 2 กลุ่ม กลุ่มหนึ่งเรียกว่า ตัวแปรตาม (dependent variable) และอีกกลุ่มหนึ่งเรียกว่า ตัวแปรอิสระ (independent variables) โดยที่ตัวแปรอิสระมีมากกว่าหนึ่งตัวแปร และความสัมพันธ์ระหว่างตัวแปรทั้งสองกลุ่มมีลักษณะความสัมพันธ์เชิงเส้น ซึ่งมีรูปแบบทั่วไปดังนี้

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ คือ เวกเตอร์ของตัวแปรตามที่มีขนาด $n \times 1$ โดยที่ n คือขนาดตัวอย่าง

X คือ เมทริกซ์ของตัวแปรอิสระที่มีขนาด $n \times (p+1)$ โดยที่ p คือจำนวนตัวแปรอิสระ

$\underset{\sim}{\beta}$ คือ เวกเตอร์ของพารามิเตอร์ที่มีขนาด $(p+1) \times 1$

$\underset{\sim}{\varepsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนที่มีขนาด $n \times 1$

และ $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$, $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $\text{cov}(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

จุดประสงค์อย่างหนึ่งที่สำคัญของการวิเคราะห์ความถดถอยคือการพยากรณ์หรือการศึกษาเกี่ยวกับแนวโน้มของตัวแปรตาม และการที่จะสามารถพยากรณ์ค่าของตัวแปรตามได้นั้นจำเป็นต้องทราบถึงอิทธิพลของตัวแปรอิสระที่มีต่อตัวแปรตาม ซึ่งก็คือการประมาณค่าพารามิเตอร์หรือเวกเตอร์สัมประสิทธิ์การถดถอย $\underset{\sim}{\beta}$ และวิธีการประมาณเวกเตอร์ค่าพารามิเตอร์ $\underset{\sim}{\beta}$ วิธีหนึ่งที่นิยมใช้กันมากคือวิธีกำลังสองน้อยสุด (Least Square method) โดยเป็นที่ทราบกันทั่วไปว่าวิธีกำลังสองน้อยสุดนี้เป็นวิธีที่ให้ตัวประมาณที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (Best Linear Unbiased Estimator : BLUE) ทั้งนี้ต้องเป็นไปตามข้อสมมติที่ตัวแปรอิสระต้องไม่มีความสัมพันธ์กัน

ในทางปฏิบัติกรณีที่ข้อมูลมีจำนวนตัวแปรอิสระมากและมีความสัมพันธ์กันหรือที่เรียกว่าตัวแปรอิสระมีพหุสัมพันธ์กัน (multicolinearity) มีผลทำให้เกิดปัญหาในการวิเคราะห์ความถดถอยดังนี้

1. เราหาเมทริกซ์ของ $(X'X)^{-1}$ ได้ยาก เมื่อตัวแปรอิสระมีจำนวนมากและมีความสัมพันธ์กันหรือตัวแปรอิสระหนึ่งตัวเป็นผลบวกเชิงเส้นของตัวแปรอิสระตัวอื่นๆจะไม่สามารถหาเมทริกซ์ผกผัน

ของเมทริกซ์ได้โดยง่าย หรืออาจจะหาไม่ได้

2. ความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์การถดถอยมีค่าเท่ากับผลคูณของความคลาดเคลื่อนมาตรฐานของการประมาณค่าและสมาชิกแนวทแยงมุมของเมทริกซ์ $(X'X)^{-1}$ ซึ่งสมาชิกในเมทริกซ์ผกผันนี้มีค่าเท่ากับ $1/\left[(n-1)s_i^2(1-R_i^2)\right]$ เมื่อ R_i เป็นสัมประสิทธิ์สหสัมพันธ์พหุคูณระหว่าง X_i และตัวแปรอิสระอื่นๆ และ s_i^2 เป็นความแปรปรวนของตัวแปรอิสระ X_i ถ้า R_i มีค่ามากจะทำให้ค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์การถดถอยมีค่ามากด้วยและทำให้เกิดปัญหาในการอธิบายความหมายของสัมประสิทธิ์การถดถอยเหล่านี้

3. เมทริกซ์สหสัมพันธ์ของสัมประสิทธิ์การถดถอยเป็นฟังก์ชันของ $\underset{\sim}{b}'X'X\underset{\sim}{b}$ ถ้าจำนวนตัวแปรอิสระมากขึ้นและมีความสัมพันธ์กันมากขึ้นจะทำให้จำนวนสัมประสิทธิ์การถดถอยมีมากขึ้นและมีความสัมพันธ์กันมากขึ้น ซึ่งทำให้เกิดปัญหาในการอธิบายความหมายของสัมประสิทธิ์การถดถอย

จากปัญหาเหล่านี้นักสถิติหรือนักวิเคราะห์หลายคนได้เสนอการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยโดยใช้วิธีการประมาณที่มีความเอนเอียง แต่สามารถลดจำนวนตัวแปรอิสระลงและมีความแปรปรวนของตัวประมาณสัมประสิทธิ์การถดถอยน้อยกว่าตัวประมาณวิธีกำลังสองน้อยสุดแต่เพิ่มส่วนของความเอนเอียง เนื่องจากความแปรปรวนของตัวประมาณบวกกับกำลังสองของความเอนเอียงเป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error : MSE) เป็นเกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยว่าตัวประมาณสัมประสิทธิ์การถดถอยมีค่าใกล้เคียงกับค่าจริงมากน้อยเพียงใด ซึ่งถ้าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณสัมประสิทธิ์การถดถอยมีค่าน้อยแสดงว่าตัวประมาณสัมประสิทธิ์การถดถอยเป็นตัวประมาณที่ดี ซึ่งมีผลต่อการพยากรณ์ค่าตัวแปรตามในสมการการถดถอยโดยจะได้ค่าพยากรณ์ที่ใกล้เคียงความเป็นจริงด้วย ผู้วิจัยจึงสนใจว่าในการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณในกรณีที่เกิดพหุสัมพันธ์กันระหว่างตัวแปรอิสระด้วยวิธีกำลังสองน้อยสุดแบ่งส่วน (Parital Least Square : PLS) วิธีการถดถอยองค์ประกอบหลัก (Principal Component Regression : PCR) และวิธีกำลังสองน้อยสุด (Ordinary Least square : OLS) วิธีใดจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการพยากรณ์ (Mean Square Error of Prediction : MSEP) ต่ำสุด

1.2 วัตถุประสงค์ของการวิจัย

ผู้วิจัยต้องการเปรียบเทียบวิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติด้วยวิธีกำลังสองน้อยสุดแบ่งส่วน (PLS) วิธีการถดถอยองค์ประกอบหลัก (PCR)

และวิธีกำลังสองน้อยสุด (OLS) ในกรณีที่ตัวแปรอิสระเกิดพหุสัมพันธ์ในรูปแบบของความสัมพันธ์และระดับของความสัมพันธ์ที่แตกต่างกัน ความแปรปรวนของความคลาดเคลื่อนมีค่าสูงและเมื่อตัวแปรอิสระมีจำนวนมากขึ้น

1.3 สมมติฐานการวิจัย

ในการวิเคราะห์ความถดถอยพหุกรณีที่ตัวแปรอิสระมีพหุสัมพันธ์ ตัวแบบถดถอยที่ได้จากวิธีกำลังสองน้อยสุดแบ่งส่วนให้ค่าพยากรณ์ที่ดีกว่าตัวแบบถดถอยที่ได้จากวิธีการถดถอยองค์ประกอบหลักและวิธีกำลังสองน้อยสุดเมื่อจำนวนตัวแปรอิสระมีความสัมพันธ์กันมากและความแปรปรวนของความคลาดเคลื่อนในตัวแบบถดถอยมีค่าสูง และตัวแบบถดถอยที่ได้จากวิธีการถดถอยองค์ประกอบหลักจะให้ค่าพยากรณ์ที่ดีกว่าตัวแบบที่ได้จากวิธีกำลังสองน้อยสุดและกำลังสองน้อยสุดแบ่งส่วนเมื่อตัวแปรอิสระมีความสัมพันธ์กันทุกตัว

1.4 ข้อตกลงเบื้องต้น

1. สมการการถดถอยที่ใช้ในการศึกษานี้เป็นสมการการถดถอยพหุกูณ มีรูปแบบดังนี้

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ คือ เวกเตอร์ของตัวแปรตามที่มีขนาด $n \times 1$ โดยที่ n คือขนาดตัวอย่าง

X คือ เมทริกซ์ของตัวแปรอิสระที่มีขนาด $n \times p$ โดยที่ p คือจำนวนตัวแปรอิสระ

$\underset{\sim}{\beta}$ คือ เวกเตอร์ของพารามิเตอร์ที่มีขนาด $p \times 1$

$\underset{\sim}{\varepsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนที่มีขนาด $n \times 1$

โดยที่ $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$, $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $\text{cov}(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

2. ตัวแปรอิสระ $\underset{\sim}{X} = (X_1, X_2, ..., X_m)'$ มีการแจกแจงร่วม คือ การแจกแจงปกติของหลายตัวแปร (multivariate normal distribution) เขียนเป็นสัญลักษณ์ $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ โดยที่ $\underset{\sim}{\mu} = E(\underset{\sim}{X})$ และ $\Sigma = \text{cov}(\underset{\sim}{X})$

3. ตัวแปรอิสระ X และตัวแปรตาม y เป็นตัวแปรเชิงปริมาณ

1.5 ขอบเขตการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยกำหนดขอบเขตของตัวแปรและค่าคงที่ต่างๆ สำหรับเปรียบเทียบการประมาณค่าพารามิเตอร์ด้วยวิธีต่างๆ ทั้ง 3 วิธี ที่กล่าวมาแล้วข้างต้น ดังนี้

1. จำนวนตัวแปรอิสระในสมการการถดถอยมี 3 ขนาดคือ 5, 8 และ 12

2. กำหนดขนาดตัวอย่างของแต่ละตัวแปรมากกว่าจำนวนตัวแปรอยู่ 40 ตัว ดังนั้น ขนาดตัวอย่างจะเป็น 40+p เมื่อ p คือจำนวนตัวแปรอิสระ ทั้งนี้เพื่อให้ข้อมูลสามารถใช้กับการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยสุดได้เป็นอย่างดี

3. ความแปรปรวนของความคลาดเคลื่อน σ^2 มีค่าเป็น 1.0, 3.0, 5.0, 7.0 และ 10.0

4. กำหนดจำนวนตัวแปรอิสระที่มีพหุสัมพันธ์และระดับความสัมพันธ์ของตัวแปร เป็น 3 กรณีดังนี้

4.1 <u>กรณีที่จำนวนตัวแปรอิสระเท่ากับ 5</u> ผู้วิจัยสนใจศึกษาเมื่อตัวแปรอิสระต่างๆ มีความสัมพันธ์ดังนี้

ก. พหุสัมพันธ์ระหว่างกลุ่มตัวแปรอิสระ 1 กลุ่ม ผู้วิจัยทำการศึกษาที่ระดับความสัมพันธ์ของกลุ่มตัวแปรมีค่าเท่ากับ (0.60), (0.70), (0.80), (0.95), (0.975) และ (0.999) โดยแบ่งออกเป็น

1) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 3 ตัวแปร ได้แก่ ตัวแปร X_1, X_2 และ X_3 มีพหุสัมพันธ์กัน

2) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 4 ตัวแปร ได้แก่ ตัวแปร X_1, X_2, X_3 และ X_4 มีพหุสัมพันธ์กัน

3) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 5 ตัวแปร ได้แก่ ตัวแปร X_1, X_2, X_3, X_4 และ X_5 มีพหุสัมพันธ์กัน

ข. พหุสัมพันธ์ระหว่างกลุ่มตัวแปรอิสระ 2 กลุ่ม ผู้วิจัยทำการศึกษาที่ระดับความสัมพันธ์ของตัวแปรมีค่าเท่ากับ (0.60,060), (0.60,0.70), (0.60,0.80), (0.60,0.95), (0.60,0.975), (0.60,0.999), (0.70,0.60), (0.70,0.70), (0.70,0.80), (0.70,0.95), (0.70,0.975), (0.70,0.999) , (0.80,0.60), (0.80,0.70), (0.80,0.80), (0.80,0.95), (0.80,0.975), (0.80,0.999), (0.95,0.60), (0.95,0.70), (0.95,0.80), (0.95,0.95), (0.95,0.975), (0.95,0.999), (0.975,0.60), (0.975,0.70), (0.975,0.80), (0.975,0.95), (0.975,0.975), (0.975,0.999), (0.999,0.60), (0.999,0.70), (0.999,0.80), (0.999,0.95), (0.999,0.975), (0.999,0.999)* โดยแบ่งออกเป็น

1) พหุสัมพันธ์ระหว่าง 2 ตัวแปร 2 กลุ่ม ได้แก่ ตัวแปร X_1, X_2 มีพหุ

---

*ตัวเลขในวงเล็บแสดงถึงระดับความสัมพันธ์ของตัวแปรอิสระของ 2 กลุ่มตัวแปรโดยที่ตัวเลขตัวแรกหมายถึงระดับความสัมพันธ์ของตัวแปรกลุ่มที่หนึ่งและตัวเลขตัวที่สองหมายถึงระดับความสัมพันธ์ของตัวแปรอิสระในกลุ่มที่สอง

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สัมพันธ์กันและ X_3, X_4 มีพหุสัมพันธ์กัน

2) พหุสัมพันธ์ระหว่าง 3 ตัวแปร 1 กลุ่ม และ 2 ตัวแปร 1 กลุ่ม ได้แก่ ตัวแปร X_1, X_2, X_3 มีพหุสัมพันธ์กัน และ X_4, X_5 มีพหุสัมพันธ์กัน

4.2 <u>กรณีที่จำนวนตัวแปรอิสระเท่ากับ 8</u> ผู้วิจัยสนใจศึกษาเมื่อตัวแปรอิสระต่างๆ มีความสัมพันธ์กัน ดังนี้

ก. พหุสัมพันธ์ะหว่างตัวแปรอิสระ 1 กลุ่ม ผู้วิจัยทำการศึกษาที่ระดับความสัมพันธ์ของตัวแปรมีค่าเท่ากับ (0.60), (0.70), (0.80), (0.95), (0.975) และ (0.999) โดยแบ่งออกเป็น

1) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 5 ตัวแปร ได้แก่ X_1, X_2, X_3, X_4 และ X_5 มีพหุสัมพันธ์กัน

2) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 6 ตัวแปร ได้แก่ X_1, X_2, X_3, X_4, X_5 และ X_6 มีพหุสัมพันธ์กัน

3) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 7 ตัวแปร ได้แก่ X_1, X_2, X_3, X_4, X_5, X_6 และ X_7 มีพหุสัมพันธ์กัน

4) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 8 ตัวแปร ได้แก่ X_1, X_2, X_3, X_4, X_5, X_6, X_7 และ X_8 มีพหุสัมพันธ์กัน

ข. พหุสัมพันธ์ระหว่างกลุ่มตัวแปรอิสระ 2 กลุ่ม ผู้วิจัยทำการศึกษาที่ระดับความสัมพันธ์ของตัวแปรมีค่าเท่ากับ (0.60,060),(0.60,0.70), (0.60,0.80), (0.60,0.95), (0.60,0.975), (0.60,0.999), (0.70,0.60), (0.70,0.70), (0.70,0.80), (0.70,0.95), (0.70,0.975), (0.70,0.999) ,(0.80,0.60), (0.80,0.70), (0.80,0.80), (0.80,0.95), (0.80,0.975), (0.80,0.999), (0.95,0.60), (0.95,0.70), (0.95,0.80), (0.95,0.95), (0.95,0.975), (0.95,0.999), (0.975,0.60), (0.975,0.70), (0.975,0.80), (0.975,0.95), (0.975,0.975), (0.975,0.999), (0.999,0.60), (0.999,0.70), (0.999,0.80), (0.999,0.95), (0.999,0.975), (0.999,0.999) โดยแบ่งออกเป็น

1) พหุสัมพันธ์ระหว่าง 3 ตัวแปร 2 กลุ่ม ได้แก่ ตัวแปร X_1, X_2, X_3 มีพหุ-สัมพันธ์กัน และ X_4, X_5, X_6 มีพหุสัมพันธ์กัน

2) พหุสัมพันธ์ระหว่าง 4 ตัวแปร 1 กลุ่ม และ 3 ตัวแปร 1 กลุ่ม ได้แก่ ตัวแปร X_1, X_2, X_3, X_4 มีพหุสัมพันธ์กัน และ X_5, X_6, X_7 มีพหุสัมพันธ์กัน

3) พหุสัมพันธ์ระหว่าง 4 ตัวแปร 2 กลุ่ม ได้แก่ ตัวแปร X_1, X_2, X_3, X_4 มีพหุสัมพันธ์กัน และ X_5, X_6, X_7, X_8 มีพหุสัมพันธ์กัน

4.3 <u>กรณีจำนวนตัวแปรอิสระเท่ากับ 12</u> ผู้วิจัยสนใจศึกษาเมื่อตัวแปรอิสระต่างๆ มีความสัมพันธ์กันดังนี้

ก. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 1 กลุ่ม ผู้วิจัยทำการศึกษาที่ระดับความสัมพันธ์ของตัวแปรเท่ากับ (0.60), (0.70), (0.80), (0.95), (0.975) และ (0.999) โดยแบ่งออกเป็น

1) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 7 ตัวแปร ได้แก่ ตัวแปร X_1, X_2, X_3, X_4, X_5, X_6 และ X_7 มีพหุสัมพันธ์กัน

2) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 8 ตัวแปร ได้แก่ ตัวแปร X_1, X_2, X_3, X_4,X_5, X_6, X_7 และ X_8 มีพหุสัมพันธ์กัน

3) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 9 ตัวแปร ได้แก่ ตัวแปร X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8 และ X_9 มีพหุสัมพันธ์กัน

4) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 10 ตัวแปร ได้แก่ ตัวแปร X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9 และ X_{10} มีพหุสัมพันธ์กัน

5) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 11 ตัวแปร ได้แก่ ตัวแปร X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}และ X_{11} มีพหุสัมพันธ์กัน

6) พหุสัมพันธ์ระหว่างตัวแปรอิสระ 12 ตัวแปร ได้แก่ ตัวแปร X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11} และ X_{12} มีพหุสัมพันธ์กัน

ข. พหุสัมพันธ์ระหว่างตัวแปรอิสระ 2 กลุ่มตัวแปร ผู้วิจัยทำการศึกษาระดับความสัมพันธ์ของตัวแปรมีค่าเท่ากับ (0.60,060),(0.60,0.70), (0.60,0.80), (0.60,0.95), (0.60,0.975), (0.60,0.999), (0.70,0.60), (0.70,0.70), (0.70,0.80), (0.70,0.95), (0.70,0.975), (0.70,0.999) ,(0.80,0.60), (0.80,0.70), (0.80,0.80), (0.80,0.95), (0.80,0.975), (0.80,0.999), (0.95,0.60), (0.95,0.70), (0.95,0.80), (0.95,0.95), (0.95,0.975), (0.95,0.999), (0.975,0.60), (0.975,0.70), (0.975,0.80), (0.975,0.95), (0.975,0.975), (0.975,0.999), (0.999,0.60), (0.999,0.70), (0.999,0.80), (0.999,0.95), (0.999,0.975), (0.999,0.999) โดยแบ่งออกเป็น

1) พหุสัมพันธ์ระหว่าง 4 ตัวแปร 2 กลุ่ม ได้แก่ ตัวแปร X_1, X_2, X_3, X_4 มีพหุสัมพันธ์กัน และ X_5, X_6, X_7, X_8 มีพหุสัมพันธ์กัน

2) พหุสัมพันธ์ระหว่าง 5 ตัวแปร 2 กลุ่ม ได้แก่ ตัวแปร X_1, X_2, X_3, X_4, X_5 มีพหุสัมพันธ์กัน และ X_6, X_7, X_8, X_9, X_{10}มีพหุสัมพันธ์กัน

3) พหุสัมพันธ์ระหว่าง 6 ตัวแปร 2 กลุ่ม ได้แก่ ตัวแปร X_1, X_2, X_3, X_4, X_5, X_6 มีพหุสัมพันธ์กัน และ X_7, X_8, X_9, X_{10}, X_{11}, X_{12}มีพหุสัมพันธ์กัน

1.6 เกณฑ์ที่ใช้ในการเปรียบเทียบ

ผู้วิจัยกำหนดเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณโดยใช้ค่าเฉลี่ยความผิดพลาดกำลังสองของการพยากรณ์ (Average Mean Square Error of - Prediction : AMSEP) ซึ่งมีวิธีการคำนวณดังนี้

1.6.1 <u>การหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการพยากรณ์</u>

ผู้วิจัยสามารถคำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการพยากรณ์ (MSEP) จาก

$$MSEP = E\left[(y_j - \hat{y}_j)(y_j - \hat{y}_j)'\right]$$

$$= \frac{1}{n_2}\sum_{j=n_1+1}^{n_1+n_2}(y_j - \hat{y}_j)^2 \quad ; j=n_1+1,\ldots,n_1+n_2$$

เมื่อ y_j คือ ค่าจริงของตัวแปรตาม

$\hat{y}_j$ คือ ค่าประมาณของตัวแปรตามที่ได้จากสมการความถดถอยในแต่ละวิธี

n_1 คือ ขนาดตัวอย่างที่ใช้ในการหาค่าสัมประสิทธิ์การถดถอยในสมการของแต่ละวิธี

n_2 คือ ขนาดตัวอย่างที่ใช้ในการหาค่าพยากรณ์ตัวแปรตาม y

และ $n_1 + n_2$ คือ ขนาดตัวอย่างของข้อมูลทั้งหมดที่สร้างให้มีการแจกแจงตามที่ต้องการศึกษา

1.6.2 <u>การหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการพยากรณ์ของการทำซ้ำ</u>

เนื่องจากการวิจัยครั้งนี้กระทำซ้ำ 500 ครั้ง ในแต่ละสถานการณ์ ดังนั้น ค่าเฉลี่ยของ MSEP ของแต่ละวิธี คือ

$$AMSEP(OLS) = \frac{1}{500}\sum_{i=1}^{500} MSEP(OLS)_i$$

$$AMSEP(PCR) = \frac{1}{500}\sum_{i=1}^{500} MSEP(PCR)_i$$

$$AMSEP(PLS) = \frac{1}{500}\sum_{i=1}^{500} MSEP(PLS)_i$$

1.6.3 <u>การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการพยากรณ์ของแต่ละวิธี</u>

คำนวณเปอร์เซ็นต์ของความแตกต่างระหว่าง AMSEP ทั้ง 3 วิธี โดยการเปรียบเทียบกับวิธีที่ให้ค่า MSEP น้อยที่สุด ดังนี้

$$DIFF = \frac{AMSEP_{(i)} - AMSEP_{(min)}}{AMSEP_{(min)}} \times 100 \quad ;i = 1,2,3$$

เมื่อ DIFF คือ เปอร์เซ็นต์ของอัตราส่วนผลต่างค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการพยากรณ์

$AMSEP_{(min)}$ คือ AMSEP ของวิธีที่ให้ค่าน้อยที่สุด

และ $AMSEP_{(i)}$ คือ AMSEP ของแต่ละวิธี

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่ได้รับจากงานวิจัยครั้งนี้คือเราสามารถเลือกใช้วิธีการประมาณค่าในการวิคราะห์ความถดถอยในกรณีที่ตัวแปรอิสระเกิดพหุสัมพันธ์ได้อย่างเหมาะสมและเป็นตัวประมาณสัมประสิทธิ์การถดถอยในตัวแบบที่ดีกว่าตัวประมาณอื่นๆ ซึ่งจะให้ค่าพยากรณ์ที่ดีและใกล้เคียงกับค่าจริง