



### 1.1 ความเป็นมา และความสำคัญของปัญหา

ในการวิเคราะห์ความถดถอยพหุ (Multiple regression analysis) ถือว่าการใช้ตัวแปรอิสระที่เหมาะสมมากกว่าหนึ่งตัว โดยทั่ว ๆ ไปย่อมทำให้ผลของการประมาณค่าตัวแปรตามมีความถูกต้องมากกว่าการใช้ตัวแปรอิสระเพียงตัวเดียว โดยที่ตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์กันในลักษณะใดลักษณะหนึ่ง

สำหรับตัวแบบทั่วไป (General model) ที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระ และตัวแปรตามแบบเชิงเส้น (linear relationship) จะมีลักษณะดังนี้

$$Y = X\beta + \epsilon ; \quad \epsilon \sim N(0, \sigma^2 I)$$

โดยที่ Y คือ เมตริกซ์ของตัวแปรตามขนาด  $n \times 1$

X คือ เมตริกซ์ของตัวแปรอิสระขนาด  $n \times p$

$\beta$  คือ เมตริกซ์ของสัมประสิทธิ์ความถดถอยพหุขนาด  $p \times 1$

$\epsilon$  คือ เมตริกซ์ของความคลาดเคลื่อนขนาด  $n \times 1$

n คือ ขนาดตัวอย่าง

p คือ จำนวนตัวแปรอิสระ

ในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุ จากรูปแบบดังกล่าวนี้ วิธีที่นิยมใช้กันมากที่สุดคือ วิธีกำลังสองน้อยที่สุด (Least square method) จะได้ตัวประมาณ  $\hat{\beta} = (X'X)^{-1}X'Y$  เป็นตัวประมาณที่ไม่เอนเอียง และให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุดในบรรดาตัวประมาณที่ไม่เอนเอียงทั้งหลาย แต่ในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุด มีข้อสมมติที่จำเป็นข้อหนึ่งคือ ตัวแปรอิสระจะต้องไม่มีความสัมพันธ์กันในลักษณะเชิงเส้น ซึ่งในทางปฏิบัติเป็นไปได้น้อยมาก เพราะตัวแปรต่าง ๆ ที่นำมาศึกษาอาจมีความสัมพันธ์กัน ตัวแปรอิสระบางตัว อาจเป็นฟังก์ชันของตัวแปรอิสระตัวอื่น ๆ ซึ่งกรณีเช่นนี้กล่าวได้ว่าตัวแปรอิสระมีพหุ

สัมพันธ์ (Multicollinearity) กัน ค่าประมาณตัวแปรตามที่ได้ อาจไม่เหมาะสม และอาจทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์ความถดถอยพหุ มีค่ามาก นั้นหมายถึงว่าค่าประมาณสัมประสิทธิ์ความถดถอยพหุที่ได้ขาดความแม่นยำ (accuracy) ถ้าหากตัวแปรอิสระ มีพหุสัมพันธ์กัน อาจแก้ไขได้โดยการตัดตัวแปรบางตัว ซึ่งมีความสัมพันธ์กับตัวแปรอื่นออกจากตัวแบบ ถ้าหากความสัมพันธ์ระหว่างตัวแปรอิสระนั้นชัดเจนพอ แต่ในบางครั้งการตัดตัวแปรอิสระตัวใดตัวหนึ่งออกจากตัวแบบทำได้ยาก เนื่องจากลักษณะความสัมพันธ์ระหว่างตัวแปรอิสระที่เกิดขึ้นไม่ชัดเจนพอ และถือว่าตัวแปรอิสระทุกตัวมีผลต่อการเปลี่ยนแปลงของตัวแปรตามมากพอสมควร

Hoerl and Kennard (1970: 69-82) ได้ศึกษาวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยพหุที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่า วิธีการกำลังสองน้อยที่สุด โดยให้ชื่อว่า ริตจ์ รีเกรสชัน (Ridge regression) ซึ่งวิธีนี้ ไม่ต้องตัดตัวแปรอิสระออกจากตัวแบบ ถึงแม้จะเกิดพหุสัมพันธ์ในระหว่างตัวแปรอิสระก็ตาม แต่ค่าประมาณสัมประสิทธิ์ความถดถอยพหุที่ได้จากวิธี ริตจ์ รีเกรสชันมีอคติ (Bias)

เนื่องจาก ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีการกำลังสองน้อยที่สุด เป็นฟังก์ชันของ  $(XX)^{-1}$  ฉะนั้น การที่จะพยายามลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ให้ต่ำลง จึงต้องพยายามลดค่า  $(XX)^{-1}$  ให้ต่ำลง ซึ่ง ริตจ์ รีเกรสชัน อาจทำได้ โดยการบวกค่าคงที่มากกว่าศูนย์ กับสมาชิกทุกตัวบนเส้นทแยงมุมของเมตริกซ์  $(XX)$  ได้ตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุโดยวิธี ริตจ์ รีเกรสชันดังนี้

$$\hat{\beta}_R = (XX + kI)^{-1}XY ; k > 0$$

Hoerl and Kennard ได้กล่าวว่าตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุที่ได้จากวิธี ริตจ์ รีเกรสชัน จะมีลักษณะค่อนข้างคงที่ ค่าสัมบูรณ์ของตัวประมาณมีค่าสัมเหตุสมผล และเครื่องหมายของค่าประมาณสัมประสิทธิ์ความถดถอยพหุ จะถูกต้อง

Conniffe and Stone (1973: 181-187) ได้วิจารณ์เกี่ยวกับการใช้ ริตจ์ รีเกรสชันว่า การเปลี่ยนแปลงของเครื่องหมายของค่าประมาณสัมประสิทธิ์ความถดถอยพหุนี้ ควรจะได้รับการพิจารณาอย่างถี่ถ้วน ในกรณีที่  $\hat{\beta}_R$  มีเครื่องหมายตรงกันข้ามกับค่าประมาณสัมประสิทธิ์ความถดถอยพหุที่ได้จากวิธีการกำลังสองน้อยที่สุด การที่ได้เครื่องหมายต่างกันนี้ ไม่ได้หมายความว่า เครื่องหมายที่ได้จากวิธีการกำลังสองน้อยที่สุด (เมื่อ  $k = 0$ ) จะผิด

Gunst and Mason (1977 : 616-628) ได้กล่าวว่า ถ้า  $\sigma^2$  (ความแปรปรวนของความคลาดเคลื่อน) มีค่าน้อย ตัวประมาณค่าสัมประสิทธิ์ความถดถอยพหุที่ได้จากวิธีกำลังสองน้อยที่สุดจะเป็นตัวประมาณที่ดี ถึงแม้ว่า ตัวแปรอิสระจะมีพหุสัมพันธ์กันก็ตาม

นอกจากจะใช้ วิธีรีดักต์ รีเกรสชั่น ในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุ เมื่อตัวแปรอิสระมีพหุสัมพันธ์แล้ว ก็อาจใช้วิธีรีเกรสชั่นพริ้นซิเปิ้ลคอมโพเนนท์ (Regression principal component) ซึ่งจะได้ตัวประมาณที่เอนเอียง และให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยที่สุด เช่นเดียวกัน

สิ่งที่น่าสนใจว่า ในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุ และการประมาณค่าตัวแปรตามเมื่อตัวแปรอิสระมีพหุสัมพันธ์กัน ด้วยวิธีกำลังสองน้อยที่สุด วิธีรีดักต์ รีเกรสชั่นและวิธีรีเกรสชั่นพริ้นซิเปิ้ลคอมโพเนนท์ วิธีใดจึงจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด

ดังนั้น ในการวิจัยเรื่องนี้ สิ่งสนใจที่จะเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอยพหุ และการประมาณค่าตัวแปรตามโดยใช้วิธีกำลังสองน้อยที่สุด วิธีรีดักต์ รีเกรสชั่น และวิธีรีเกรสชั่นพริ้นซิเปิ้ลคอมโพเนนท์ โดยจะศึกษาเปรียบเทียบที่ดัชนีพหุสัมพันธ์ ระดับต่าง ๆ กัน

## 1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบการประมาณค่าสัมประสิทธิ์ความถดถอยพหุ และการประมาณค่าตัวแปรตาม เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยใช้วิธีกำลังสองน้อยที่สุด วิธีรีดักต์ รีเกรสชั่น และวิธีรีเกรสชั่นพริ้นซิเปิ้ลคอมโพเนนท์

## 1.3 สมมติฐานของการวิจัย

เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันมาก วิธีรีดักต์ รีเกรสชั่น และวิธีรีเกรสชั่นพริ้นซิเปิ้ลคอมโพเนนท์ จะให้ค่าประมาณที่มีความถูกต้อง เชื่อถือ ได้มากกว่าวิธีกำลังสองน้อยที่สุด ไม่ว่าขนาดตัวอย่างหรือความแปรปรวนของข้อมูลจะมีค่ามากหรือน้อยก็ตาม

## 1.4 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัยนี้

เพื่อช่วยให้นักวิจัยมีผลสรุป และหลักฐานในการเลือกใช้วิธีการประมาณค่าจากรูปแบบความสัมพันธ์ที่กำหนดขึ้น เมื่อตัวแปรอิสระมีพหุสัมพันธ์กันได้อย่างเหมาะสม