



## บทที่ 1

### บทนำ

#### 1.1 ความเป็นมาและความสำคัญของปัญหา

การใช้เทคนิคการวิเคราะห์ความถดถอยเพื่อใช้เป็นเครื่องมือช่วยในการหาคำตอบในงานวิจัยโดยส่วนใหญ่ผู้วิจัยมักเลือกใช้วิธีกำลังสองน้อยที่สุดเนื่องจากเป็นวิธีการที่ง่ายและเป็นวิธีการที่ให้ตัวประมาณที่ดีคือได้ตัวประมาณเชิงเส้นที่ไม่เอนเอียงและมีความแปรปรวนต่ำที่สุดภายใต้ข้อตกลงเบื้องต้นของการวิเคราะห์ความถดถอย ถ้าหากข้อมูลที่นำมาวิเคราะห์ เป็นข้อมูลที่ไม่สมบูรณ์ (Incomplete Data) คือไม่ทราบค่าสังเกตที่แท้จริงของตัวแปรที่สนใจหรือค่าสังเกตนั้นมีลักษณะเป็นข้อมูลที่ถูกตัดทิ้ง (Censored Data) ทำให้ต้องบันทึกค่าเท่าที่สังเกตได้ เมื่อทำการวิเคราะห์ข้อมูลอาจจะได้ตัวประมาณที่เอนเอียง ส่งผลกระทบต่อการวิเคราะห์และสรุปผล นอกจากนี้ถ้าหากมีข้อมูลถูกตัดทิ้งเป็นจำนวนมากอาจส่งผลให้ความแม่นยำในการประมาณค่าลดน้อยลง ความคลาดเคลื่อนที่เกิดจากการประมาณก็จะมากขึ้น ดังนั้นจึงควรทำการประมาณค่าที่ถูกตัดทิ้งโดยใช้วิธีการที่มีความเหมาะสมที่จะทำให้ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง

ข้อมูลถูกตัดทิ้งจะพบมากในด้านการแพทย์ ด้านอุตสาหกรรม ด้านการประกันภัย เช่น ในโครงการทดลองเปลี่ยนหัวใจคนไข้<sup>1</sup> Stanford Heart Transplantation Program เมื่อต้องการทดลองว่าจำนวนวันที่มีชีวิตรอดของคนไข้หลังการผ่าตัดหัวใจขึ้นอยู่กับอายุของคนไข้ที่เข้ารับการรักษาหรือไม่ ซึ่งในการทดลองนี้ตัวแปรอิสระคืออายุของคนไข้ที่ได้รับการรักษา ตัวแปรตามคือจำนวนวันที่มีชีวิตรอดของคนไข้หลังการผ่าตัดหัวใจ โดยเริ่มทำการทดลองเมื่อวันที่ 1 ตุลาคม 1967 และสิ้นสุดการทดลองในวันที่ 1 เมษายน 1974 ในระหว่างการทดลองมีคนไข้ 69 คน เข้ารับการเปลี่ยนหัวใจและสังเกตว่าคนไข้จะมีชีวิตรอดหลังจากเปลี่ยนหัวใจ โดยนับจำนวนวันที่มีชีวิตรอด ดังนั้นคนไข้ที่เสียชีวิตในระยะเวลาที่ทำการทดลองจะเป็นข้อมูลที่ไม่ถูกตัดทิ้ง และคนไข้ที่มีชีวิตรอดเมื่อสิ้นสุดการทดลองจะเป็นข้อมูลถูกตัดทิ้ง เนื่องจากเราไม่สามารถทราบได้เลยว่าคนไข้กลุ่มนี้จะมีชีวิตอีกนานเท่าใด ทราบเฉพาะจำนวนวันที่มีชีวิตรอดเมื่อสิ้นสุดการทดลองเท่านั้น ตัวอย่างข้อมูลถูกตัดทิ้งด้านอุตสาหกรรม เช่น ศึกษาเกี่ยวกับอายุการใช้งาน

<sup>1</sup> Rupert G. Miller. 'Least Squares Regression with Censored Data'. *Biometrika* 63(3) : (1976) : 456-458.



ของเครื่องจักรชนิดหนึ่ง โดยศึกษาในระยะเวลา 8000 ชม. เริ่มทำการทดลองโดยให้เครื่องจักรเครื่องนี้ทำงานแล้วบันทึกเวลาไว้ตั้งแต่เริ่มทำงานจนกระทั่งเครื่องจักรเสื่อมสภาพ ในระหว่างการทดลองเครื่องจักรใดเสื่อมสภาพ จะเป็นค่าสังเกตที่ไม่ถูกตัดทิ้ง (Uncensored data หรือ Survival Time) และเมื่อสิ้นสุดการทดลอง เครื่องจักรเครื่องใดที่ยังคงอยู่ในสภาพที่ใช้งานได้ก็จะเป็นเครื่องที่ไม่ทราบอายุการใช้งานที่แน่นอน จะบันทึกไว้ว่ามีอายุการใช้งาน 8000 ชม. ซึ่งข้อมูลนี้จะเป็นค่าสังเกตที่ถูกตัดทิ้ง (Censored data)

ในการวิจัยครั้งนี้จะประมาณค่าพารามิเตอร์  $\beta$  จากสมการถดถอยเชิงเส้นตรงพหุ โดยใช้วิเคราะห์กับข้อมูลที่ค่าสังเกตของตัวแปรตามมีค่าถูกตัดทิ้งทางขวาและเป็นค่าถูกตัดทิ้งประเภทที่ 1 (Type I Censoring) โดยทั่วไปแล้วการนำข้อมูลที่ถูกตัดทิ้งมาใช้ในการประมาณค่าพารามิเตอร์นั้นอาจจะทำได้ใน 2 กรณีคือ

กรณีแรก จะถือว่าค่าที่ถูกตัดทิ้งเสมือนเป็นค่าที่ไม่ถูกตัดทิ้งจะทำการวิเคราะห์ข้อมูลทั้งหมด

กรณีที่สอง ไม่สนใจข้อมูลที่ถูกลบทิ้งจะทำการวิเคราะห์เฉพาะข้อมูลที่ไม่ถูกตัดทิ้งเท่านั้น ดังนั้นจำนวนข้อมูลที่นำมาวิเคราะห์จึงมีน้อยกว่าวิธีแรก

เมื่อทำการวิเคราะห์ข้อมูลด้วยวิธีกำลังสองน้อยที่สุดจากข้อมูลทั้งสองกรณีดังกล่าว จะทำให้ได้ตัวประมาณที่เอนเอียงและโดยเฉลี่ยแล้วการประมาณค่าจะต่ำกว่าความเป็นจริง หรือจะทำให้ได้ช่วงความเชื่อมั่นที่แคบกว่าความเป็นจริง<sup>2</sup> ด้วยเหตุนี้จึงได้มีผู้คิดวิธีการในการประมาณค่าพารามิเตอร์เมื่อเกิดปัญหาข้อมูลถูกตัดทิ้งในตัวแปรตามขึ้นมาหลายวิธี เช่น วิธีการของบัคเลย์และเจมส์ (Jonathan Buckley & James : 1979) วิธีการของปีเตอร์ เจมส์ สมิธ (Peter James Smith : 1986) วิธีการของมิลเลอร์ (Rupert G. Miller : 1976) เป็นต้น

มิลเลอร์และเจอร์รี่ (Rupert Miller and Jerry Halpern : 1982) ได้ทำการเปรียบเทียบวิธีการของมิลเลอร์กับวิธีการของบัคเลย์และเจมส์ พบว่าวิธีการของบัคเลย์และเจมส์ จะให้ค่าประมาณพารามิเตอร์ที่น่าเชื่อถือกว่าวิธีการของมิลเลอร์ จงดี โรจนประศาสน์ (2536) ได้ทำการเปรียบเทียบพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อข้อมูลค่าสังเกตของตัวแปรตามเป็นค่าที่ถูกตัดทิ้งทางขวา โดยประมาณค่าพารามิเตอร์จากวิธีการกำลังสองน้อยที่สุด วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการของบัคเลย์และเจมส์ พบว่าวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด เป็นวิธีการที่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ ในทุกสถานการณ์ที่ทำการศึกษา

<sup>2</sup> Josef Schmee and Gerald J.Hahn. 'A Simple Method fir Regression Analysis with Censored Data.' *Technometrics*, 21(4) : (1979) ; 417-418.



ดังนั้นจึงเป็นที่น่าสนใจว่า เมื่อเกิดปัญหาตัวแปรตามบางค่าเป็นค่าที่ถูกตัดทิ้งและทราบค่าของการถูกตัดทิ้งในการวิเคราะห์การถดถอย วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด วิธีการของสมิท และวิธีการโมดิไฟด์แอกชูวเรียล วิธีการใดจะเป็นวิธีที่เหมาะสมในปัญหานี้

## 1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอยเชิงเส้นตรงพหุ เมื่อตัวแปรตามบางค่ามีค่าถูกตัดทิ้งทางขวาและเป็นการตัดทิ้งประเภทที่ 1 ด้วยวิธี

1.2.1 วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation Method)

1.2.2 วิธีการของสมิท (Smith Method)

1.2.3 วิธีการโมดิไฟด์แอกชูวเรียล (Modified Acturial Method)

## 1.3 สมมติฐานทางการวิจัย

ในการประมาณค่าตัวแปรตามในสมการถดถอยเชิงเส้นตรงพหุ เมื่อตัวแปรตามมีค่าถูกตัดทิ้งทางขวาและเป็นการตัดทิ้งประเภทที่ 1 ในแต่ละวิธีจะให้ค่าประมาณที่แตกต่างกัน วิธีการของสมิท จะให้ค่าประมาณที่ใกล้เคียงกับค่าจริงมากกว่าวิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด และวิธีการโมดิไฟด์แอกชูวเรียล

## 1.4 ข้อตกลงเบื้องต้น

1.4.1 ศึกษารูปแบบสมการถดถอยเชิงเส้นตรงพหุ

$$y = X\beta + \varepsilon$$

เมื่อ  $y$  เป็นเวกเตอร์ของตัวแปรตามขนาด  $n \times 1$  โดยที่  $n$  เป็นจำนวนค่าสังเกต

$X$  เป็นเมทริกซ์ของตัวแปรอิสระขนาด  $n \times p$  ( $p < n$ ) และมี  $\text{full rank} = p$

$\beta$  เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด  $p \times 1$

$\varepsilon$  เป็นความคลาดเคลื่อนที่มีขนาด  $n \times 1$



$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1} \quad X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{12} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix}_{n \times p} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}_{p \times 1} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

1.4.2 ตัวแปรตามเท่านั้นที่เป็นค่าที่ถูกตัดทิ้ง

1.4.3 ศึกษากรณีที่ค่าที่ถูกตัดทิ้งเป็นการถูกตัดทิ้งทางขวาและเป็นการตัดทิ้งประเภทที่ 1

## 1.5 ขอบเขตของการวิจัย

1.5.1 ในการวิจัยครั้งนี้ จะทำการเปรียบเทียบการประมาณค่าตัวแปรตามในสมการถดถอยเชิงเส้นตรงพหุ เมื่อตัวแปรตามบางค่ามีค่าถูกตัดทิ้งทางขวาและเป็นการตัดทิ้งประเภทที่ 1 ด้วยวิธีประมาณด้วยภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation Method) วิธีการของสมิท (Smith Method) และวิธีการโมดิไฟด์แอคชูวเรียล (Modified Actuarial Method)

1.5.2 ศึกษาเมื่อค่าที่ถูกตัดทิ้งเป็นค่าที่ถูกตัดทิ้งประเภทที่ 1 (Type I Censoring)

$t_c$  เป็นค่าที่กำหนดไว้ล่วงหน้า

ค่าที่ไม่ถูกตัดทิ้ง  $T_i$  มีรูปแบบเป็น

$$T = \beta X + \varepsilon$$

กำหนด

$$\beta = 1$$

จำนวนตัวแปรอิสระที่ทำการศึกษามี 3 ระดับ คือ 3, 5, 7

$$X \sim N(0, 25)^3$$

$$\varepsilon \sim N(0, 25)^3$$

โดยที่ตัวแปรใหม่ที่นำมาใช้ในการวิเคราะห์มาจาก

$$Y_i = \begin{cases} T_i & \text{เมื่อ } T_i \leq t_c \quad (\text{เป็นข้อมูลที่ไม่ถูกตัดทิ้ง}) \\ t_c & \text{เมื่อ } T_i > t_c \quad (\text{เป็นข้อมูลที่ถูกตัดทิ้ง}) \end{cases}$$

<sup>3</sup> การวิจัยครั้งนี้กำหนดเป็นค่าดังกล่าวเนื่องจากทำการทดลองเปลี่ยนค่าเป็นค่าต่าง ๆ แล้ว ก็ยังทำให้ผลการเปรียบเทียบในแต่ละวิธีการนั้นไม่เปลี่ยนแปลง

1.5.3 ศึกษาเมื่อกรณีสัดส่วนของข้อมูลที่ถูกตัดทิ้งเป็น 0.1,0.2,0.3,0.4,0.5

1.5.4 ศึกษาเมื่อขนาดตัวอย่างเป็น 10,20,30,50,100

### 1.6 เกณฑ์การตัดสินใจ

เกณฑ์การตัดสินใจว่าการประมาณค่าพารามิเตอร์ด้วยวิธีใดจะเป็นวิธีที่ใช้ได้ดี จะพิจารณาโดยการเปรียบเทียบค่าความคลาดเคลื่อน ระหว่างค่าประมาณของตัวแปรตามกับค่าจริงในรูปของค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Square Root of Mean Square Error : RMSE) วิธีการใดให้ค่า RMSE ต่ำกว่าจะเป็นวิธีการประมาณค่าที่ดี

### 1.7 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการตัดสินใจเลือกวิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นตรงพหุ เมื่อค่าสังเกตของตัวแปรตามมีค่าที่ถูกตัดทิ้งทางขวาและเป็นการตัดทิ้งประเภทที่ 1 (Type I Censoring)

ศูนย์วิจัยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย