

ประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบ  
ถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตาม  
ระยะยาวที่มีโครงสร้างความสัมพันธ์ที่แตกต่างกัน

**Efficiency of Parameter Estimation by the Weighted Least Square  
Method in Logistic Regression Model for Longitudinal Data  
with Various Types of Correlation Structure**

ชูเกียรติ ผุดพรมราช และ ลีลี อิงศรีสว่าง

Chookait Pudprommarat and Lily Ingsrisawang

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพฯ

**บทคัดย่อ**

การศึกษามุ่งเน้นที่จะเปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาวภายใต้โครงสร้างความสัมพันธ์ทั้ง 4 แบบ ได้แก่ แบบอิสระ แบบแลกเปลี่ยน แบบอัตโนมัติสัมพันธ์อันดับที่ 1 และแบบไม่อิสระ โดยที่ความสัมพันธ์ระหว่างค่าสังเกตที่มีการวัดซ้ำของตัวแปรตาม ที่ระดับ 0.1, 0.3, 0.5, 0.7 และ 0.9 และขนาดตัวอย่างที่ศึกษาเท่ากับ 10, 30, 50, 100 และ 200 ในการจำลองแต่ละสถานการณ์การทดลอง ได้มาจากการจำลองข้อมูลให้เป็นตามลักษณะของข้อมูลผู้ป่วยโรคไตวายเรื้อรังที่ทำการฟอกเลือดด้วยเครื่องไตเทียมใช้โปรแกรม SAS และทำการทดลองซ้ำสถานการณ์ละ 1,000 ครั้ง พร้อมทั้งคำนวณค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) ผลการศึกษาพบว่า ประสิทธิภาพของการประมาณค่าพารามิเตอร์ในทุกระดับความสัมพันธ์ของโครงสร้างความสัมพันธ์แบบต่างๆ และในแต่ละขนาดตัวอย่าง ไม่แตกต่างกัน ส่วนประสิทธิภาพของการประมาณค่าพารามิเตอร์ ภายใต้โครงสร้างความสัมพันธ์เหมือนกันและขนาดเท่ากัน พบว่าระดับความสัมพันธ์ 0.1 มีประสิทธิภาพดีที่สุด และประสิทธิภาพของการประมาณค่าพารามิเตอร์ภายใต้โครงสร้างความสัมพันธ์เหมือนกันและระดับความสัมพันธ์เท่ากัน พบว่าขนาดตัวอย่าง 200 มีประสิทธิภาพดีที่สุด

**คำสำคัญ :** การถดถอยโลจิสติก, วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก, โครงสร้างความสัมพันธ์

## Abstract

The purpose of this research was to compare the efficiency of parameter estimation by the weighted least square method in logistic regression model for longitudinal data with various types of correlation structure. The data collected by simulating the effect of coagulation system of chronic hemodialysis patients with 3 period repeated measures, level of heparin, and patient age. The data analyzed by using SAS to calculate mean square error and simulate 1,000 times with sample sizes of 10, 30, 50, 100 and 200 in the level of correlation of 0.1, 0.3, 0.5, 0.7, and 0.9 within four correlation structures such as independence, exchangeable, first-order autoregressive (AR1) and dependent. It was found that there was no statistically difference in the efficiency of parameter in all types of correlation structure and all sample sizes. At 0.1 level of correlation with 50 or more sample sizes was the most efficiency of parameter estimation. Moreover, the parameter estimation with 200 sample sizes was the best efficiency in case of the same types of correlation structure and level of correlation.

**Keywords:** Logistic regression, Weighted least square method, Correlation structure

## 1. บทนำ

จำนวนข้อมูลถือว่ามีความสำคัญมากในการพยากรณ์เพราะหากมีข้อมูลมากขึ้นจะทำให้โอกาสที่จะพยากรณ์ได้ถูกต้องมากขึ้น และเป็นประโยชน์ต่อการนำไปใช้ในการปฏิบัติงานจริง โดยเฉพาะข้อมูลทางด้าน การแพทย์ การเกษตร การศึกษา และอื่นๆ ที่ต้องใช้เวลายาวนานในการศึกษาติดตามผล ดังนั้นจึงมีการเก็บข้อมูลซ้ำจากตัวอย่าง โดยข้อมูลที่ได้จากการเก็บรวบรวมลักษณะนี้ เรียกว่า ข้อมูลที่มีการติดตามระยะยาว (Longitudinal data) ซึ่งข้อมูลที่ได้เก็บรวบรวมมานั้นมักจะได้รับปัจจัยใดปัจจัยหนึ่งตลอดเวลาที่ทำการศึกษา และพบว่าข้อมูลที่เก็บซ้ำในหน่วยทดลองมักมีความสัมพันธ์กัน ในรูปแบบโครงสร้างความสัมพันธ์ (Correlation Structure) ซึ่งมีรูปแบบต่างๆ [1-2] เช่น โครงสร้างความสัมพันธ์แบบอิสระ (Independence) โครงสร้างความสัมพันธ์แบบแลกเปลี่ยน (Exchangeable) โครงสร้างความสัมพันธ์แบบอัตโนมัติสหสัมพันธ์ (Autocorrelation : AR) และโครงสร้างความสัมพันธ์แบบไม่อิสระ (Dependence) เป็นต้น

ในการศึกษาวิจัยเกี่ยวกับข้อมูลที่มีการติดตามระยะยาวที่ใช้ตัวแบบการถดถอยแบบโลจิสติก ปัญหาอย่างหนึ่งของการวิเคราะห์งานลักษณะนี้ คือ การเลือกโครงสร้างความสัมพันธ์ที่เหมาะสมกับข้อมูล ในทางปฏิบัติ ผู้วิจัยมักไม่ทราบ โครงสร้างความสัมพันธ์ที่แท้จริงของข้อมูลว่ามีโครงสร้างความสัมพันธ์เป็นแบบใด ดังนั้นผู้วิจัยจึงสนใจศึกษาเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว ภายใต้โครงสร้างความสัมพันธ์ของข้อมูลที่แตกต่างกัน โดยใช้ข้อมูล

จากการจำลองสถานการณ์ให้เป็นไปตามลักษณะข้อมูลผู้ป่วยโรคไตวายเรื้อรังที่ทำการฟอกเลือดด้วยเครื่องไตเทียม (hemodialysis) เพื่อเป็นแนวทางในการเลือกโครงสร้างความสัมพันธ์ที่เหมาะสมกับข้อมูล

## 2. วิธีการวิจัย

### 2.1 ขอบเขตการวิจัย

การศึกษาประสิทธิภาพการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว ภายใต้โครงสร้างความสัมพันธ์ที่แตกต่างกัน โดยข้อมูลที่ใช้ในการศึกษาได้มาจากการจำลองข้อมูลตามข้อมูลผู้ป่วยโรคไตวายเรื้อรังที่ทำการฟอกไตด้วยเครื่องไตเทียมจากโรงพยาบาลจุฬาลงกรณ์และโรงพยาบาลบำรุงราษฎร์ [3] มีคำชี้แจงว่าในกระบวนการฟอกเลือดของผู้ป่วยที่ได้รับการฟอกไตด้วยเครื่องไตเทียมจำเป็นต้องใช้สารต่อต้านการแข็งตัวของเลือด ระหว่างส่งผ่านเลือดของผู้ป่วยเข้าเครื่องไตเทียมเพื่อกรองของเสียและน้ำในร่างกาย และนำเลือดกลับเข้าสู่ร่างกายผู้ป่วยไตวายเรื้อรังซึ่งสารต่อต้านการแข็งตัวของเลือดที่นิยมใช้คือ เฮพาริน (heparin) โดยทั่วไปแพทย์จะใช้เฮพารินกับผู้ป่วยทุกคนในปริมาณที่เท่ากัน โดยการฉีดเข้าหลอดเลือดดำครั้งแรก ขนาด 2,000 ยูนิต เมื่อเริ่มต้นทำการฟอกไตด้วยเครื่องไตเทียม และหลังจากนั้นจะฉีดเฮพารินขนาด 1,000 – 1,600 ยูนิต ต่อชั่วโมง เข้าหลอดเลือดดำตลอดเวลา การใช้เฮพารินปริมาณมากเกินไป ก็มีโอกาสเสี่ยงต่อการมีเลือดออกมากขึ้น โดยเฉพาะในผู้ป่วยไตวายเรื้อรังซึ่งมีภาวะเลือดออกง่ายผิดปกติอยู่แล้ว และในทางตรงกันข้ามการใช้เฮพารินปริมาณน้อยไป ก็อาจทำให้เกิดมีลิ่มเลือด ดังนั้นการใช้เฮพารินจึงมีความจำเป็นสำหรับผู้ป่วยโรคไตเรื้อรัง

ผู้วิจัยได้จำลองสถานการณ์ที่มีผู้ป่วยโรคไตวายเรื้อรังที่ทำการฟอกไตด้วยเครื่องไตเทียม มีจำนวนเท่ากับ 10, 30, 50, 100 และ 200 คน แต่ละคนจะเข้ารับการฟอกเลือดด้วยเครื่องไตเทียม 1 ครั้งต่อสัปดาห์ เป็นระยะเวลาทั้งหมด 3 สัปดาห์ โดยศึกษาการเกิดการแข็งตัวของเลือดในการฟอกเลือดด้วยเครื่องไตเทียม เมื่อผู้ป่วยโรคไตวายเรื้อรังมีอายุเริ่มต้นเข้ารับการฟอกเลือดที่แตกต่างกัน และกลุ่มระดับปริมาณเฮพารินที่ฉีดเข้าหลอดเลือดดำของผู้ป่วยในปริมาณที่แตกต่างกัน โดยตัวแปรตามเป็นตัวแปรทวิคือ การเกิดภาวะการแข็งตัวของเลือดของผู้ป่วยโรคไตวายเรื้อรังที่ฟอกเลือดด้วยเครื่องไตเทียม ซึ่งมีการฟอกไต 3 ครั้งจากผู้ป่วยแต่ละราย และตัวแปรอิสระ 2 ตัวคือ ระดับปริมาณเฮพารินที่ฉีดเข้าหลอดเลือดดำของผู้ป่วยหลังฉีดเฮพารินเข้าหลอดเลือดดำครั้งแรก ขนาด 2,000 ยูนิตเป็นตัวแปรหุ่น (Dummy Variable) และ อายุเริ่มต้นของผู้ป่วยที่เป็นโรคไตวายเรื้อรังที่เข้ารับการฟอกเลือดด้วยเครื่องไตเทียม เป็นตัวแปรแบบต่อเนื่องที่มีการแจกแจงแบบยูนิฟอร์ม (Uniform) โดยทำการเปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ภายใต้เงื่อนไขต่าง ๆ ของการจำลองสถานการณ์

## 2.2 ตัวแบบที่ใช้ในการศึกษา

### 2.2.1 ตัวแบบการถดถอยโลจิสติก (Logistic Regression Modeling)

ตัวแบบการถดถอยโลจิสติก คือ ตัวแบบการถดถอยที่แสดงความสัมพันธ์เชิงเส้นระหว่างความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ (P) กับตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพ โดยค่าความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจมีค่าอยู่ระหว่าง 0 ถึง 1 ซึ่งต้องทำการแปลงให้มีค่าอยู่ในช่วง  $(-\infty, \infty)$  เมื่อตัวแปรอิสระมีค่าอยู่ในขอบเขต  $(-\infty, \infty)$  โดยใช้การแปลงแบบโลจิท (logit transformation) ซึ่งการแปลงแบบโลจิทแสดงอยู่ในรูปโลจิทของ P [2,4-5] ดังนี้

$$\text{logit}(P) = \log\left(\frac{P}{1-P}\right)$$

จากการแปลงแบบโลจิท ทำให้ได้ตัวแบบการถดถอยโลจิสติกเชิงเส้น (linear logistic regression modelling) [5] ดังนี้

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} = x_i' \beta$$

และจากตัวแบบการถดถอยโลจิสติกเชิงเส้น สามารถหาค่า  $P_i$  ซึ่งเป็นตัวแบบการถดถอยโลจิสติก [5] ดังนี้

$$P_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)} = \frac{1}{1 + \exp(-x_i' \beta)}$$

### 2.2.2 สมการของการประมาณค่าวงนัยทั่วไป (Generalized estimating equation)

การประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว จะใช้วิธีสมการของการประมาณค่าวงนัยทั่วไป (GEE) เป็นวิธีที่ใช้ในการประมาณค่าพารามิเตอร์โดยที่มีตัวแปรตามมีการแจกแจงแบบทวินาม โดยวิธีนี้เป็นวิธีที่อิสระในการกำหนดโครงสร้างความสัมพันธ์ของตัวแปรตาม เพราะในการวิเคราะห์ตัวแบบการถดถอยทั่วไป จะกำหนดโครงสร้างความสัมพันธ์แบบอิสระ ซึ่งเป็นข้อดกลงเบื้องต้นของการวิเคราะห์ ถ้าข้อมูลที่ศึกษาไม่ได้อิสระกัน จะทำให้ผลการประมาณค่าพารามิเตอร์ไม่เหมาะสม และวิธีการประมาณค่าพารามิเตอร์ของ GEE จะใช้วิธีการประมาณที่ภาวะน่าจะเป็นสูงสุด (Quasi-Maximum Likelihood Estimator) แต่เมื่อที่ข้อมูลมีขนาดตัวอย่างที่ใหญ่จะทำให้ได้ค่าการประมาณพารามิเตอร์เท่ากับค่าประมาณพารามิเตอร์จากวิธีภาวะน่าจะเป็นสูงสุดและวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก [1,6] ซึ่งวิธี GEE จะต้องมีการกำหนดโครงสร้างความสัมพันธ์ก่อนการประมาณค่าพารามิเตอร์ทุกครั้ง

## 2.3 การจำลองข้อมูลในการศึกษา

### 2.3.1 กำหนดลักษณะการแจกแจงของตัวแปรอิสระ $X$ ค่าพารามิเตอร์ และขนาดตัวอย่าง

ในขั้นตอนนี้ ผู้วิจัยได้กำหนดค่าพารามิเตอร์ให้  $\beta_0 = -6.57, \beta_{11} = -0.54, \beta_{12} = -0.55, \beta_2 = 0.24$   
ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ  $n = 10, n = 30, n = 50, n = 100$  และ  $n = 200$

### 2.3.2 สร้างข้อมูลตัวแปรอิสระ $X$

ผู้วิจัยสร้างตัวแปรอิสระ 2 ตัวคือ

1) ตัวแปรอิสระที่ 1 แทน ปริมาณสุขภาพารินที่ฉีดเข้าหลอดเลือดดำของผู้ป่วยหลังฉีดสุขภาพารินเข้าหลอดเลือดดำครั้งแรก ขนาด 2,000 ยูนิท ประกอบด้วย 3 กลุ่มคือ กลุ่มสูง กลุ่มปานกลาง และกลุ่มต่ำ มีตัวแปร  $X_{11ij}$  และ  $X_{12ij}$  โดยที่  $i=1,2,\dots,n$  (คน) และ  $j=1,2,3$  (ครั้ง) ให้แต่ละตัวแปร

การแจกแจงแบบเบอร์นูลลี (Bernoulli Distribution) ด้วยพารามิเตอร์  $p = \frac{1}{3}$  เพราะผู้วิจัยให้ความสำคัญน่าจะเป็นของการใช้ปริมาณสุขภาพารินเท่ากันทั้ง 3 กลุ่ม โดยมีวิธีการจำลองข้อมูลดังนี้

(1) จำลองเลขสุ่ม  $R_1 \sim U(0,1)$  และ  $R_2 \sim U(0,1)$

(2) ถ้าเลขสุ่ม  $R_1 \leq p$  ให้  $x_{11ij} = 1$  และ  $x_{12ij} = 0$

$R_1 > p$  และ  $R_2 \leq p$  ให้  $x_{11ij} = 0$  และ  $x_{12ij} = 1$

$R_1 > p$  และ  $R_2 > p$  ให้  $x_{11ij} = 0$  และ  $x_{12ij} = 0$

2) ตัวแปรอิสระที่ 2 แทน อายุเริ่มต้นของผู้ป่วยที่เป็นโรคไตวายเรื้อรังที่เข้ารับการฟอกเลือดด้วยเครื่องไตเทียม มีตัวแปร  $X_{21}$  โดยที่  $i=1,2,\dots,n$  (คน) ให้ตัวแปรมีการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ในช่วงอายุ 20 ถึง 70 ปี เพราะผู้วิจัยได้ศึกษาจากอายุของผู้ป่วยโรคไตวายเรื้อรังที่ฟอกเลือดด้วยเครื่องไตเทียม ชัยชนะ [3] พบว่ามีอายุตั้งแต่ 20 ถึง 70 ปี ดังนั้นจึงกำหนดพารามิเตอร์เป็น 20 และ 70 โดยมีทำการจำลองเลขสุ่ม  $U \sim U(0,1)$  และให้  $X_{21} \sim U(a, a+b) \sim U(20, 70)$  จะได้  $x_{21} = b \times U(0,1) + a = 50U + 20$

### 2.3.3 สร้างตัวแปรตาม $Y$

เมื่อได้ค่าตัวแปรอิสระ  $X_{11ij}, X_{12ij}$  และ  $X_{21}$  ที่มีลักษณะตามที่ผู้วิจัยศึกษาและกำหนด  $\beta_0, \beta_{11}, \beta_{12}, \beta_2$  แล้ว การสร้างค่าตัวแปร  $Y_{ij}$  ที่แทนการเกิดภาวะการแข็งตัวของเลือดของผู้ป่วยโรคไตวายเรื้อรังที่ฟอกเลือดด้วยเครื่องไตเทียม คนที่  $i$  ครั้งที่  $j$  โดยที่  $i=1,2,\dots,n$  (คน) และ  $j=1,2,3$  (ครั้ง) และ  $Y_{ij}$  มีค่าเพียง 2 ค่า คือ 1 และ 0 มีวิธีการจำลองดังนี้

(1) จำลองเลขสุ่ม  $Z, Z_1, Z_2$  และ  $Z_3$  มีการแจกแจงแบบปกติมาตรฐาน

(2) นำค่า  $Z, Z_1, Z_2$  และ  $Z_3$  มาสร้างค่า  $R_1, R_2$  และ  $R_3$  ที่มีความสัมพันธ์ ( $\rho$ ) ที่ระดับ 0.1, 0.3, 0.5, 0.7 และ 0.9 [8] ดังนี้

$$R_1 = \sqrt{1-\rho}Z_1 + \sqrt{\rho}Z, R_2 = \sqrt{1-\rho}Z_2 + \sqrt{\rho}Z \text{ และ } R_3 = \sqrt{1-\rho}Z_3 + \sqrt{\rho}Z$$

(3) ค่า  $R_1, R_2$  และ  $R_3$  มีค่าอยู่ในขอบเขตระหว่าง  $(-\infty, \infty)$  ซึ่งต้องการทำการแปลงให้มามีค่าอยู่ในช่วง  $(0, 1)$  โดยใช้การแปลงแบบโลจิส (logit) [5] จะได้

$$P_1 = \frac{\exp(R_1)}{1 + \exp(R_1)}, \quad P_2 = \frac{\exp(R_2)}{1 + \exp(R_2)} \quad \text{และ} \quad P_3 = \frac{\exp(R_3)}{1 + \exp(R_3)}$$

(4) นำค่า  $\beta_0, \beta_{11}, \beta_{12}, \beta_2$  และค่าตัวแปรอิสระที่จำลองข้อมูลได้มานั้น มาคำนวณหาค่า  $\pi(x_{ij})$  ดังนี้

$$\pi(x_{ij}) = \frac{\exp(\beta_0 + \beta_{11}x_{11ij} + \beta_{12}x_{12ij} + \beta_2x_{2i})}{1 + \exp(\beta_0 + \beta_{11}x_{11ij} + \beta_{12}x_{12ij} + \beta_2x_{2i})}$$

(5) นำค่า  $P_i$  และ  $\pi(x_{ij})$  สร้างค่าตัวแปรตาม โดยมีเกณฑ์ดังนี้

ถ้าค่า  $P_1 \leq \pi(x_{11})$  ให้  $y_{11} = 1$ , ถ้า  $P_1 > \pi(x_{11})$  ให้  $y_{11} = 0$

ถ้าค่า  $P_2 \leq \pi(x_{12})$  ให้  $y_{12} = 1$ , ถ้า  $P_2 > \pi(x_{12})$  ให้  $y_{12} = 0$

ถ้าค่า  $P_3 \leq \pi(x_{13})$  ให้  $y_{13} = 1$ , ถ้า  $P_3 > \pi(x_{13})$  ให้  $y_{13} = 0$

### 2.3.4 การประมาณค่าพารามิเตอร์

เมื่อได้ข้อมูลตัวแปรอิสระและตัวแปรตามแล้ว จะทำการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว ภายใต้โครงสร้างความสัมพันธ์ที่แตกต่างกัน โดยให้  $\beta_0, \beta_{11}, \beta_{12}, \beta_2$  แทนค่าประมาณพารามิเตอร์ของ  $\beta_0, \beta_{11}, \beta_{12}, \beta_2$  ตามลำดับ ซึ่งใช้โครงสร้างความสัมพันธ์ 4 แบบ [2] ดังนี้

#### แบบที่ 1 อิสระ (Independence)

โครงสร้างความสัมพันธ์แบบอิสระ มีความแปรปรวนร่วม  $\text{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j=k \\ 0 & j \neq k \end{cases}$

ซึ่งมีลักษณะดังนี้

$$R(p) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

**แบบที่ 2 แลกเปลี่ยน (Exchangeable)**

โครงสร้างความสัมพันธ์แบบแลกเปลี่ยน มีความแปรปรวนร่วม  $\text{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j=k \\ \rho & j \neq k \end{cases}$  ซึ่งมีลักษณะดังนี้

$$R(\rho) = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{bmatrix}$$

**แบบที่ 3 อัตตสหสัมพันธ์ อันดับที่ 1 (First-order autoregressive :AR1)**

โครงสร้างความสัมพันธ์แบบอัตตสหสัมพันธ์ อันดับที่ 1 มีความแปรปรวนร่วม  $\text{Corr}(Y_{ij}, Y_{i(j+t)}) = \rho^t$  เมื่อ  $t = 0, 1, 2, \dots, n_i - j$  ซึ่งมีลักษณะดังนี้

$$R(\rho) = \begin{bmatrix} 1 & \rho_{12} = \rho^{|1-2|} & \dots & \rho_{1j} = \rho^{|1-j|} \\ \rho_{21} = \rho^{|2-1|} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{(i-1)j} = \rho^{|(i-1)-j|} \\ \rho_{i1} = \rho^{|i-1|} & \dots & \rho_{i(i-1)} = \rho^{|i-(i-1)|} & 1 \end{bmatrix}$$

**แบบที่ 4 ไม่อิสระ (Dependence)**

โครงสร้างความสัมพันธ์แบบไม่อิสระ มีความแปรปรวนร่วม  $\text{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j=k \\ \rho_{jk} & j \neq k \end{cases}$  ซึ่งมีลักษณะดังนี้

$$R(\rho) = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1j} \\ \rho_{21} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{(i-1)j} \\ \rho_{i1} & \dots & \rho_{i(i-1)} & 1 \end{bmatrix}$$

**2.3.5 การคำนวณค่าเฉลี่ย ความแปรปรวน ค่าเอนเอียง และค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์**

จากการประมาณค่าพารามิเตอร์ภายใต้โครงสร้างความสัมพันธ์ทั้ง 4 แบบเรียบร้อยแล้ว หลังจากนั้นจะไปเริ่มขั้นตอนที่ 1 ใหม่ จนครบ 1,000 รอบ แล้วจะหาค่าเฉลี่ย ค่าความแปรปรวน และความเอนเอียงของค่าประมาณพารามิเตอร์ของแต่ละโครงสร้างความสัมพันธ์

ให้  $\overline{\text{Est}}$  แทนค่าเฉลี่ยของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดในการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว

$$\text{โดยที่ } \overline{\text{Est}} = \bar{\beta} = \sum_{i=1}^{1,000} \frac{\beta_i}{1,000}$$

ให้  $\text{Var}$  แทนค่าความแปรปรวนของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดในการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว

$$\text{โดยที่ } \text{Var} = \text{Var}(\overline{\text{Est}}) = \sum_{i=1}^{1,000} \frac{(\beta_i - \bar{\beta})^2}{1,000 - 1}$$

ให้  $\text{Bias}$  แทน ค่าเอนเอียงของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดในการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว

$$\text{โดยที่ } \text{Bias} = \text{Bias}(\overline{\text{Est}}) = \bar{\beta} - \beta$$

ให้  $\text{MSE}$  แทน ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดในการถดถอยโลจิสติก

$$\text{โดยที่ } \text{MSE} = \text{Var}(\overline{\text{Est}}) + \text{Bias}^2(\overline{\text{Est}})$$

### 2.3.6 สรุปผลการวิจัย

เมื่อได้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดในการถดถอยโลจิสติกในแต่ละโครงสร้างความสัมพันธ์เรียบร้อยแล้ว หากโครงสร้างความสัมพันธ์ใดที่มีค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์ต่ำที่สุด กล่าวได้ว่าโครงสร้างความสัมพันธ์นั้นมีประสิทธิภาพของการประมาณค่าพารามิเตอร์ที่ดีที่สุด และนำผลการวิจัยไปสรุป

## 3. ผลการวิจัย

การประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว โดยมีความสัมพันธ์ที่ระดับ 0.1, 0.3, 0.5, 0.7 และ 0.9 พร้อมทั้งขนาดตัวอย่างเท่ากับ 10, 30, 50, 100 และ 200 และมีการวัดซ้ำ 3 ครั้งในแต่ละหน่วยตัวอย่าง โดยใช้โครงสร้างความสัมพันธ์ แบบอิสระ (Independence) แบบแลกเปลี่ยน (Exchangeable) แบบอัตตสหสัมพันธ์ อันดับที่ 1 (AR1) และแบบไม่อิสระ (Dependence) จากนั้นนำค่าประมาณพารามิเตอร์ที่ได้แต่ละสถานการณ์มาหาค่าเฉลี่ย ค่าความแปรปรวน ค่าเอนเอียง และค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน ซึ่งผลการวิจัยนำเสนอในตารางที่ 1

ตารางที่ 1 ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน ของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดในการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว จำแนกตามโครงสร้างความสัมพันธ์และระดับความสัมพันธ์ พร้อมทั้งขนาดตัวอย่างเท่ากับ 10, 30, 50, 100 และ 200 และมีการวัดซ้ำ 3 ครั้งในแต่ละหน่วยตัวอย่าง

Parameter	Rho	Correlation structure	MSE				
			n = 10	n = 30	n = 50	n = 100	n = 200
$\beta_0$	0.1	Independence	4944350.19	445572.80	2445.84	44.31	34.86
		Exchangeable	4944350.23	445572.90	2445.98	44.42	35.03
		AR1	4944350.27	445572.65	2445.73	44.36	34.82
		Dependence	4944350.24	445573.34	2445.90	44.37	34.82
	0.3	Independence	770479.78	191453.93	38934.99	49.42	34.65
		Exchangeable	770479.80	191454.06	38935.07	49.57	34.83
		AR1	770479.80	191453.82	38934.58	49.41	34.68
		Dependence	770479.81	191455.94	38934.64	49.45	34.69
	0.5	Independence	307711756.43	2802025.00	37981.92	1478.44	37.83
		Exchangeable	307711757.01	2802025.01	37982.02	1478.49	38.02
		AR1	307711756.51	2802025.10	37981.88	1478.39	37.84
		Dependence	307711756.45	2802063.12	37981.98	1478.41	37.85
0.7	Independence	1057318.83	839911.88	39734.70	494.10	39.87	
	Exchangeable	1057318.83	839912.09	39734.79	494.26	40.07	
	AR1	1057318.78	839911.96	39734.49	494.00	39.87	
	Dependence	6338340.54	839913.70	39735.02	494.02	39.88	
0.9	Independence	2626268.92	40232628.56	344826.11	3591.00	47.95	
	Exchangeable	2626268.94	40232628.83	344826.31	3591.23	48.18	
	AR1	2626268.59	40232628.18	344825.95	3590.95	47.89	
	Dependence	2626268.58	40234490.70	344826.32	3591.37	47.89	
$\beta_{11}$	0.1	Independence	650893.70	6615.12	41.86	0.93	0.53
		Exchangeable	650893.61	6615.13	41.86	0.93	0.53
		AR1	650894.68	6615.11	41.87	0.94	0.53
		Dependence	650895.68	6615.18	41.88	0.94	0.53
	0.3	Independence	40258.28	19854.47	536.74	1.06	0.55
		Exchangeable	40258.20	19854.43	536.75	1.06	0.55
		AR1	40258.36	19854.45	536.73	1.06	0.55
		Dependence	40258.36	19855.83	536.73	1.06	0.55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1 (ต่อ)

Parameter	Rho	Correlation structure	MSE				
			n = 10	n = 30	n = 50	n = 100	n = 200
$\beta_{11}$	0.5	Independence	91084974.34	146871.68	1281.78	14.89	0.61
		Exchangeable	91084986.16	146871.67	1281.79	14.89	0.61
		AR1	91084974.75	146871.70	1281.78	14.89	0.61
		Dependence	91084974.44	241455.74	1281.78	14.89	0.61
	0.7	Independence	166372.87	41513.34	2121.56	17.74	0.71
		Exchangeable	166384.69	41513.27	2121.57	17.74	0.71
		AR1	166373.24	41518.99	2121.56	17.73	0.71
		Dependence	15316914.08	41518.27	2121.58	17.73	0.71
	0.9	Independence	401844.27	412587.50	37014.60	43.26	0.91
		Exchangeable	401858.84	412587.53	37014.61	43.27	0.92
		AR1	401871.15	412587.59	37014.59	43.28	0.91
		Dependence	401867.28	412675.83	37014.62	43.28	0.91
$\beta_{12}$	0.1	Independence	402068.49	13065.61	66.29	1.70	0.60
		Exchangeable	402068.51	13065.61	66.35	1.70	0.60
		AR1	402068.52	13065.56	66.31	1.70	0.60
		Dependence	402068.53	13065.60	66.31	1.70	0.60
	0.3	Independence	80446.72	23145.64	5169.16	2.71	0.65
		Exchangeable	80446.73	23145.59	5169.19	2.72	0.65
		AR1	80446.79	23146.15	5169.19	2.72	0.65
		Dependence	80446.77	23154.39	5169.18	2.72	0.65
	0.5	Independence	2785915.43	120911.41	3408.93	65.17	0.77
		Exchangeable	2785915.43	121000.23	3408.92	65.17	0.77
		AR1	2785915.46	120911.48	3408.95	65.17	0.76
		Dependence	2785915.48	120948.73	3408.96	65.17	0.76
0.7	Independence	150503.07	89646.12	5891.83	8.37	0.91	
	Exchangeable	150503.07	89646.15	5891.95	8.37	0.92	
	AR1	150503.07	90493.09	5891.85	8.37	0.92	
	Dependence	150503.07	89646.16	5891.96	8.38	0.92	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1 (ต่อ)

Parameter	Rho	Correlation structure	MSE				
			n = 10	n = 30	n = 50	n = 100	n = 200
$\beta_{12}$	0.9	Independence	3033862.94	999778.06	46499.92	190.70	1.13
		Exchangeable	3033862.94	999778.29	46499.89	190.72	1.13
		AR1	3033862.95	999778.06	46512.84	190.70	1.12
		Dependence	3033862.95	1000220.78	46511.80	190.84	1.12
$\beta_2$	0.1	Independence	2596.43	428.33	3.01	0.06	0.05
		Exchangeable	2596.43	428.34	3.01	0.06	0.05
		AR1	2596.43	428.33	3.01	0.06	0.05
		Dependence	2596.43	428.34	3.01	0.06	0.05
	0.3	Independence	437.46	268.07	45.44	0.07	0.05
		Exchangeable	437.46	268.07	45.44	0.07	0.05
		AR1	437.46	268.07	45.44	0.07	0.05
		Dependence	437.46	268.07	45.44	0.07	0.05
	0.5	Independence	224698.41	2035.32	44.53	1.97	0.05
		Exchangeable	224698.40	2035.31	44.53	1.97	0.05
		AR1	224698.41	2035.32	44.53	1.97	0.05
		Dependence	224698.40	2035.32	44.53	1.97	0.05
0.7	Independence	1422.57	1067.38	53.44	0.73	0.05	
	Exchangeable	1422.58	1067.38	53.44	0.73	0.05	
	AR1	1422.57	1067.39	53.44	0.73	0.05	
	Dependence	36848.89	1067.39	53.44	0.73	0.05	
0.9	Independence	2732.50	50067.71	3956.71	4.48	0.06	
	Exchangeable	2732.50	50067.71	3956.71	4.48	0.06	
	AR1	2732.49	50067.71	3956.70	4.47	0.06	
	Dependence	2732.50	50068.61	3956.70	4.48	0.06	

#### 4. สรุปและอภิปรายผล

จากการศึกษาการเปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว ภายใต้โครงสร้างความสัมพันธ์ที่แตกต่างกัน สรุปผลได้ดังนี้

1) ประสิทธิภาพของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักในตัวแบบการถดถอยโลจิสติกสำหรับข้อมูลที่มีการติดตามระยะยาว ในทุกระดับความสัมพันธ์ของโครงสร้างความสัมพันธ์แบบต่าง ๆ และในแต่ละขนาดตัวอย่าง ไม่แตกต่างกัน

2) ประสิทธิภาพของการประมาณค่าพารามิเตอร์ในระดับความสัมพันธ์ที่แตกต่างกัน ภายใต้โครงสร้างความสัมพันธ์เหมือนกัน พบว่า ระดับความสัมพันธ์ 0.1 มีประสิทธิภาพดีที่สุด ด้วยค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) ที่มีค่าต่ำที่สุด รองลงมาคือ ระดับความสัมพันธ์ 0.3, 0.5, 0.7 และ 0.9 ตามลำดับ เมื่อขนาดตัวอย่างตั้งแต่ 50 ตัวอย่างขึ้นไป

3) ประสิทธิภาพของการประมาณค่าพารามิเตอร์ในขนาดตัวอย่างที่แตกต่างกัน ที่มีการวัดซ้ำ 3 ครั้ง ภายใต้โครงสร้างความสัมพันธ์เหมือนกันและระดับความสัมพันธ์เท่ากัน พบว่าขนาดตัวอย่าง 200 มีประสิทธิภาพดีที่สุด รองลงมาคือ ขนาดตัวอย่าง 100, 50, 30 และ 10 ตามลำดับ ด้วยค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) ของการประมาณค่าพารามิเตอร์ที่มีค่าต่ำที่สุด

จากผลการวิจัยพบว่าประสิทธิภาพของการประมาณค่าพารามิเตอร์ในทุกระดับความสัมพันธ์ของโครงสร้างความสัมพันธ์แบบต่าง ๆ ในทุกระดับความสัมพันธ์ และในแต่ละขนาดตัวอย่าง ไม่แตกต่างกัน ซึ่งสอดคล้องกับงานวิจัยของ Liang และ Zeger [6], Lipsitz et al. [7] และ Chaganty [9] ที่ได้กล่าวว่า ประสิทธิภาพของตัวประมาณพารามิเตอร์ที่ได้เพิ่มขึ้นมากจากกำหนดโครงสร้างความสัมพันธ์ที่ถูกต้อง มีค่าเพิ่มขึ้นเพียงเล็กน้อยเท่านั้น หรือไม่แตกต่างกันกับตัวประมาณพารามิเตอร์ที่กำหนดโครงสร้างความสัมพันธ์ไม่ถูกต้อง เพราะเนื่องวิธีการของการประมาณค่าวงนัยทั่วไป (GEE) มีตัวประมาณพารามิเตอร์ที่ Robust หรือ empirical ที่ทำให้ประสิทธิภาพไม่แตกต่างกัน ถึงแม้เลือกโครงสร้างความสัมพันธ์ไม่ถูกต้อง

## เอกสารอ้างอิง

- [1] Hardin, J.W. and Hilbe, J.M. 2002. Generalized Estimating Equations. Chapman and Hall., New York.
- [2] Nick R.P. and other. 2008. Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R. Computational Statistics and Data Analysis, 53, 632-641.
- [3] ชัญญา บุญญไกร. 2542. ประสิทธิภาพของการใช้เกณฑ์งานศาสตร์ในการปรับขนาดสหพารินระหว่าง การทำการฟอกเลือดด้วยเครื่องไตเทียมในผู้ป่วยไตวายเรื้อรัง. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาอายุรศาสตร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- [4] Grace, Y., Wenqing, H. and Hua, L. 2009. Analysis of correlated binary data under partially linear single-index logistic models. Journal of Multivariate Analysis, 100, 278-290.

- [5] McCullagh, P. and Nelder, J.A. 1989. *Generalized Linear Models*. 2nd ed. Chapman and Hall, London.
- [6] Liang, K.Y. and Zeger, S.L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- [7] Lipsitz, S.R., Kim, K. and Zhao, L. 1994. Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, 13, 1149-1163.
- [8] Patrick, J.F. and Katrina, R. 2008. Methods for generating longitudinally correlated binary data. *International Statistical Review*, 76, 28-38.
- [9] Chaganty, N.R. 2004. Efficiency of generalized estimating equations for binary responses. *J.R. Statist. Soc.*, 66, 851-860.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้