การประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติคทวินาม

นางสาว ทัศนาพร จงเกตุกรณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2546

ISBN 974-17-4347-5

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

18 ต.ค. 254

I 21512358

PARAMETER - ESTIMATION METHODS FOR BINOMIAL LOGISTIC REGRESSION MODEL

Miss Tatsanaporn Jongketkorn

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2003

ISBN 974-17-4347-5

หัวข้อวิทยานิพนธ์	การประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติคทวินาม
โดย	นางสาวทัศนาพร จงเกตุกรณ์
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(ผู้ช่วยศาสตราจารย์ ดร.ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร. ธีระพร วีระถาวร)

..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

..... กรรมการ
(รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา)

ทัศนาพร จงเกตุกรณ์ : การประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติคทวินาม
(PARAMETER-ESTIMATION METHODS FOR BINOMIAL LOGISTIC REGRESSION MODEL)
อ.ที่ปรึกษา : รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา , 117 หน้า. ISBN 974-17-4347-5

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบถดถอยโลจิสติคทวินาม โดยวิธีการประมาณค่าพารามิเตอร์ที่ใช้ในการวิจัยครั้งนี้คือ วิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method : ML) วิธีการถ่วงน้ำหนัก (Weighting Method : WE) และ วิธีปรับแก้เบื้องต้น (Prior Correction Method : PC) เมื่อตัวแบบถดถอยโลจิสติคมีรูปแบบดังนี้ $\pi(x_i)=\frac{e^{\beta_0+\beta_1x_1+...+\beta_px_p}}{1+e^{\beta_0+\beta_1x_1+...+\beta_px_p}}$ โดยที่ $\pi(x_i)$ แทนความน่าจะเป็นเมื่อเกิดเหตุการณ์ที่สนใจของตัวแปรตาม (Y) $X_1, X_2,...,X_p$ แทนตัวแปรอิสระ $\beta_0,\beta_1,...,\beta_p$ แทนค่าสัมประสิทธิ์ของการถดถอย p แทนจำนวนตัวแปรอิสระ สำหรับข้อมูลตัวแปรตามที่ใช้ในการวิจัยครั้งนี้มีการแจกแจงแบบทวินามด้วยพารามิเตอร์ $n_i = n$ และ $\pi(x_i)$ การเปรียบเทียบกระทำภายใต้สถานการณ์ของจำนวนตัวแปรอิสระแต่ละตัวแบบเท่ากับ 3 5 และ 7 ตัว จำนวนกลุ่ม (m) เท่ากับ 30 90 150 และ 210 ค่าพารามิเตอร์ n เท่ากับ 10 20 และ 30 ค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากรเท่ากับ 0.1 0.3 0.5 และ 0.8 เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าระยะทางมาหาโลบิสเฉลี่ย (Average Mahalanobis distance : AMH) ข้อมูลที่ใช้ในวิจัยครั้งนี้ได้จากการจำลองด้วยเทคนิคมอนติคาร์โลโดยการกระทำซ้ำ 500 ครั้ง ในแต่ละสถานการณ์โดยใช้โปรแกรม S-Plus 2000 ผลการวิจัยสรุปดังนี้

เมื่อค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากรเท่ากับ 0.1 และ 0.3 จะพบว่า วิธีความควรจะเป็นสูงสุดให้ค่า AMH ต่ำที่สุด รองลงมาคือ วิธีปรับแก้เบื้องต้น และ วิธีการถ่วงน้ำหนัก ตามลำดับ สำหรับตัวอย่างทุกจำนวนกลุ่มและทุก n แต่ในกรณีที่ค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากรเท่ากับ 0.5 และ 0.8 จะพบว่า วิธีปรับแก้เบื้องต้น ให้ค่า AMH ต่ำที่สุด รองลงมาคือ วิธีการถ่วงน้ำหนัก และ วิธีความควรจะเป็นสูงสุด ตามลำดับ สำหรับตัวอย่างทุกจำนวนกลุ่มและทุก n จะเห็นได้ว่ากรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจระหว่างประชากรและตัวอย่างใกล้เคียงกันมากขึ้นวิธีการถ่วงน้ำหนักและวิธีการปรับแก้เบื้องต้นจะทำให้ประมาณค่าพารามิเตอร์มีประสิทธิภาพมากขึ้น

ภาควิชา.....สถิติ..... ลายมือชื่อนิสิต.....ทัศนาพร จงเกตุกรณ์.....
สาขาวิชา.....สถิติ..... ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา.....2546.....

4482253426 : MAJOR STATISTICS

KEY WORD : LOGISTIC REGRESSION/MAXIMUM LIKELIHOOD/ WEIGHTING MAXIMUM LIKELIHOOD /PRIOR CORRECTION

TATSANAPORN JONGKETKORN : PARAMETER ESTIMATION METHODS OF BINOMIAL LOGISTIC REGRESSION MODEL .

THESIS ADVISOR : PROF. SUPOL DURONGWATANA ,Ph.D., 117 pp. ISBN 974-17-4347-5

The objective of this research is to compare the parameter estimation methods of binomial logistic regression model .The methods of estimating parameter under consideration in this research are Maximum Likelihood Method (ML) Weighting Method (WE) and Prior Correction Method (PC) . The model for logistic regression is as follows : $\pi(x_i)=\frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}$ where $\pi(x_i)$ is the probability of interested events of dependent variable ; $X_1, X_2,\ldots,X_p$ are the independent variable ; $\beta_0,\beta_1,\ldots,\beta_p$ are the regression coefficients ; p is the number of independent variable .In addition the data of dependent variable of this research has been binomial distribution with n_i= n and $\pi(x_i)$ parameters. The comparison is done under conditions of the numbers of independent variable are equal to 3 5 and 7 , sample group are equal to 30 90 150 and 210 , the determination of n parameter value is 10 20 and 30 and the average probability of success in population are equal to 0.1 0.3 0.5 and 0.8 .The criteria employed for the comparison are Average Mahalanobis distance (AMH) . The data for this research is simulated by using the Monte Carlo simulation technique with 500 repetitions for each situation by S-plus 2000 package . The results of this research are as follows:

According to the comparison of AMH from three referred methods, it is found that when the average probability of success in population are equal to 0.1 and 0.3 , ML method would give the lowest AMH ,while PC and WE are the second and the third lowest , respectively , for all groups and values of n . In case the average probability of success in population are equal to 0.5 and 0.8 , PC method would give the lowest AMH ,while WE and ML are the second and the third lowest , respectively , for all groups and values of n . It is found that the closer the average probability of success between population and sample are , the more efficiency that parameter estimation from WE and PC methods will be.

Department .....Statistics..... Student's signature.....

Field of study .....Statistics..... Advisor's signature.....

Academic year .....2003.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำ ปรึกษา ตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่าง ๆ จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณและสำนึกในพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ธีระพร วีระถาวร และ รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา ในฐานะประธานกรรมการและกรรมการการสอบวิทยานิพนธ์ ที่กรุณาตรวจแก้วิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น ขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ให้โอกาสทางการศึกษา และประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ท้ายนี้ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา ซึ่งสนับสนุนในด้านการเงินและให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา รวมทั้ง น้องสาว น้องชาย ญาติๆ และเพื่อนๆ ทุกคนที่ส่งเสริมและให้กำลังใจให้แก่ผู้วิจัยมาตลอด และเนื่องจากทุนวิจัยครั้งนี้บางส่วนได้รับมาจากทุนอุดหนุนการวิจัยของบัณฑิตวิทยาลัย จึงขอกราบขอบพระคุณบัณฑิตวิทยาลัยมา ณ ที่นี้ด้วย

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ญ

สารบัญภาพ..... ฏ

บทที่

1 บทนำ..... 1

1.1 ความสำคัญและความเป็นมาของปัญหา..... 1

1.2 วัตถุประสงค์ของการวิจัย..... 4

1.3 ขอบเขตการวิจัย..... 4

1.4 เกณฑ์การตัดสินใจ..... 5

1.5 ประโยชน์ของการวิจัย..... 5

2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง..... 6

2.1 ข้อสมมติทั่วไปของตัวแบบถดถอยโลจิสติค..... 6

2.2 การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด..... 8

2.2.1 วิธี Newton – Raphson สำหรับตัวแปรอิสระ 3 ตัว..... 9

2.2.2 วิธี Newton – Raphson สำหรับตัวแปรอิสระ 5 ตัว..... 14

2.2.3 วิธี Newton – Raphson สำหรับตัวแปรอิสระ 7 ตัว..... 19

2.3 การประมาณค่าพารามิเตอร์ด้วยวิธีการถ่วงน้ำหนัก..... 25

2.3.1 วิธี Newton – Raphson สำหรับตัวแปรอิสระ 3 ตัว..... 27

2.3.2 วิธี Newton – Raphson สำหรับตัวแปรอิสระ 5 ตัว..... 33

2.3.3 วิธี Newton – Raphson สำหรับตัวแปรอิสระ 7 ตัว..... 38

2.4 การประมาณค่าพารามิเตอร์ด้วยวิธีปรับแก้เบื้องต้น..... 45

3 วิธีดำเนินการวิจัย..... 48

3.1 แผนการดำเนินการวิจัย..... 48

3.2 ขั้นตอนในการดำเนินการวิจัย..... 49

บทที่ หน้า

4 ผลการวิเคราะห์ข้อมูล..... 58

4.1 ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบถดถอย..... 61
โลจิสติคด้วยวิธี ML วิธี WE และวิธี PC กรณีตัวแปรอิสระ 3 ตัว

4.1.1 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 61
$(\bar{\pi}(x_i))$ เท่ากับ 0.1

4.1.2 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 64
$(\bar{\pi}(x_i))$ เท่ากับ 0.3

4.1.3 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 67
$(\bar{\pi}(x_i))$ เท่ากับ 0.5

4.1.4 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 70
$(\bar{\pi}(x_i))$ เท่ากับ 0.8

4.2 ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบถดถอย..... 74
โลจิสติคด้วยวิธี ML วิธี WE และวิธี PC กรณีตัวแปรอิสระ 5 ตัว

4.2.1 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 74
$(\bar{\pi}(x_i))$ เท่ากับ 0.1

4.2.2 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 77
$(\bar{\pi}(x_i))$ เท่ากับ 0.3

4.2.3 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 80
$(\bar{\pi}(x_i))$ เท่ากับ 0.5

4.2.4 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 83
$(\bar{\pi}(x_i))$ เท่ากับ 0.8

4.3 ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบถดถอย..... 87
โลจิสติคด้วยวิธี ML วิธี WE และวิธี PC กรณีตัวแปรอิสระ 7 ตัว

4.3.1 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 87
$(\bar{\pi}(x_i))$ เท่ากับ 0.1

4.3.2 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 90
$(\bar{\pi}(x_i))$ เท่ากับ 0.3

4.3.3 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 93
$(\bar{\pi}(x_i))$ เท่ากับ 0.5

บทที่ หน้า

4.3.4 กรณีค่าเฉลี่ยความน่าจะเป็นของเหตุการณ์ที่สนใจของประชากร..... 96
$(\bar{\pi}(x_i))$ เท่ากับ 0.8

5 สรุปผลการวิจัยและข้อเสนอแนะ..... 101

5.1 สรุปผลการวิจัย..... 102

5.1.1 ผลการเปรียบเทียบค่าระยะทางมาหาลาโนบิสเฉลี่ย (AMH)..... 102
ของแต่ละวิธีการ

5.1.2 ผลการวิจัยปัจจัยที่มีผลต่อค่า AMH..... 102

5.1.3 ผลการสรุปการเลือกวิธีการประมาณค่าพารามิเตอร์ของตัวแบบ..... 103
ถดถอยโลจิสติค

5.2 ข้อเสนอแนะ..... 104

รายการอ้างอิง..... 105

ภาคผนวก..... 107

ประวัติผู้เขียนวิทยานิพนธ์..... 117

ตาราง หน้า

ตารางที่ 4.1.1 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 61
ตัวแปรอิสระเท่ากับ 3 ตัว กรณี $\bar{\pi}(x_i)$ = 0.1

ตารางที่ 4.1.2 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 64
ตัวแปรอิสระเท่ากับ 3 ตัว กรณี $\bar{\pi}(x_i)$ = 0.3

ตารางที่ 4.1.3 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 67
ตัวแปรอิสระเท่ากับ 3 ตัว กรณี $\bar{\pi}(x_i)$ = 0.5

ตารางที่ 4.1.4 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 70
ตัวแปรอิสระเท่ากับ 3 ตัว กรณี $\bar{\pi}(x_i)$ = 0.8

ตารางที่ 4.2.1 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 74
ตัวแปรอิสระเท่ากับ 5 ตัว กรณี $\bar{\pi}(x_i)$ = 0.1

ตารางที่ 4.2.2 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 77
ตัวแปรอิสระเท่ากับ 5 ตัว กรณี $\bar{\pi}(x_i)$ = 0.3

ตารางที่ 4.2.3 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 80
ตัวแปรอิสระเท่ากับ 5 ตัว กรณี $\bar{\pi}(x_i)$ = 0.5

ตารางที่ 4.2.4 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 83
ตัวแปรอิสระเท่ากับ 5 ตัว กรณี $\bar{\pi}(x_i)$ = 0.8

ตารางที่ 4.3.1 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 87
ตัวแปรอิสระเท่ากับ 7 ตัว กรณี $\bar{\pi}(x_i)$ = 0.1

ตารางที่ 4.3.2 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 90
ตัวแปรอิสระเท่ากับ 7 ตัว กรณี $\bar{\pi}(x_i)$ = 0.3

ตารางที่ 4.3.3 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 93
ตัวแปรอิสระเท่ากับ 7 ตัว กรณี $\bar{\pi}(x_i)$ = 0.5

ตารางที่ 4.3.4 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 96
ตัวแปรอิสระเท่ากับ 7 ตัว กรณี $\bar{\pi}(x_i)$ = 0.8

สารบัญภาพ

ภาพประกอบ หน้า

รูปที่ 4.1.1 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 62
ตัวแปรอิสระเท่ากับ 3 ตัว กรณี $\bar{\pi}(x_i)$ = 0.1
รูปที่ 4.1.2 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 65
ตัวแปรอิสระเท่ากับ 3 ตัว กรณี $\bar{\pi}(x_i)$ = 0.3
รูปที่ 4.1.3 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 68
ตัวแปรอิสระเท่ากับ 3 ตัว กรณี $\bar{\pi}(x_i)$ = 0.5
รูปที่ 4.1.4 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 71
ตัวแปรอิสระเท่ากับ 3 ตัว กรณี $\bar{\pi}(x_i)$ = 0.8
รูปที่ 4.2.1 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 75
ตัวแปรอิสระเท่ากับ 5 ตัว กรณี $\bar{\pi}(x_i)$ = 0.1
รูปที่ 4.2.2 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 78
ตัวแปรอิสระเท่ากับ 5 ตัว กรณี $\bar{\pi}(x_i)$ = 0.3
รูปที่ 4.2.3 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 81
ตัวแปรอิสระเท่ากับ 5 ตัว กรณี $\bar{\pi}(x_i)$ = 0.5
รูปที่ 4.2.4 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 84
ตัวแปรอิสระเท่ากับ 5 ตัว กรณี $\bar{\pi}(x_i)$ = 0.8
รูปที่ 4.3.1 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 88
ตัวแปรอิสระเท่ากับ 7 ตัว กรณี $\bar{\pi}(x_i)$ = 0.1
รูปที่ 4.3.2 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 91
ตัวแปรอิสระเท่ากับ 7 ตัว กรณี $\bar{\pi}(x_i)$ = 0.3
รูปที่ 4.3.3 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 94
ตัวแปรอิสระเท่ากับ 7 ตัว กรณี $\bar{\pi}(x_i)$ = 0.5
รูปที่ 4.3.4 แสดงค่า AMH ของวิธี ML วิธี WE และวิธี PC เมื่อจำนวน..... 97
ตัวแปรอิสระเท่ากับ 7 ตัว กรณี $\bar{\pi}(x_i)$ = 0.8