การเปรียบเทียบการวิเคราะห์ข้อมูลระหว่างการวิเคราะห์ข้อมูลจำแนกประเภท และการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค สำหรับข้อมูลระยะยาว

นางสาวศุภรัตน์ ดิษบรรจง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2552

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

COMPARISON BETWEEN DISCRIMINANT ANALYSIS AND MULTINOMIAL LOGISTIC REGRESSION WITH LONGITUDINAL DATA

Miss Suparat Ditbanjong

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2009

Copyright of Chulalongkorn University

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หัวข้อวิทยานิพนธ์	การเปรียบเทียบการวิเคราะห์ข้อมูล ระหว่างการวิเคราะห์ข้อมูลจำแนกประเภทและการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค สำหรับข้อมูลระยะยาว
โดย	นางสาวศุภรัตน์ ดิษบรรจง
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร. อรรณพ ตันละมัย)

คณะกรรมการสอบวิทยานิพนธ์

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

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์)

..... กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร. บุญอ้อม โฉมที)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ศุภรัตน์ ดิษบรรจง : การเปรียบเทียบการวิเคราะห์ข้อมูล ระหว่างการวิเคราะห์ข้อมูลจำแนกประเภทและการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค สำหรับข้อมูลระยะยาว. (Comparison between Discriminant Analysis and Multinomail Logistic Regression with Longitudinal Data) **อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ.ดร.สุพล ดุรงค์วัฒนา, 111หน้า.**

งานวิจัยนี้มีวัตถุประสงค์เพื่อทำการเปรียบเทียบวิธีการวิเคราะห์ข้อมูลที่มีลักษณะของตัวแปรตามเป็นแบบพหุสำหรับข้อมูลระยะยาว วิธีการที่ใช้ในการวิเคราะห์มี 2 วิธี ได้แก่ การวิเคราะห์จำแนกประเภทแบบพหุ (multiple discriminant method: MDA) ซึ่งใช้หลักของเบส์ในการจำแนกกลุ่ม และวิธีการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค (multiple logistic regression method: MLR) ซึ่งใช้สมการประมาณค่าโดยนัยทั่วไป (generalized estimating equation: GEE) ในการจำแนกกลุ่ม ข้อมูลที่นำมาวิเคราะห์ประกอบด้วยตัวแปรอิสระและตัวแปรตาม ซึ่งตัวแปรตามเป็นตัวแปรเชิงคุณภาพ ที่มีจำนวนกลุ่มมากกว่า 2 กลุ่ม และตัวแปรอิสระประกอบด้วยตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพ โดยการวิจัยครั้งนี้ใช้ข้อมูลจริงทางด้านการแพทย์จำนวน 3 ชุด สำหรับการวิเคราะห์ทั้ง 2 วิธี (การติดตามผลโรคนอนไม่หลับ การผ่าตัดเปลี่ยนผิวข้อสะโพก และการรักษาอาการเอ็นร้อยหวายฉีก ตามลำดับ) เกณฑ์ที่ใช้ในการตัดสินใจ คือ การวัดค่าความถูกต้องในการจัดกลุ่ม ผลการวิเคราะห์ข้อมูล สรุปได้ดังนี้

ข้อมูลชุดที่ 1 เมื่อกำหนดให้โครงสร้างของความแปรปรวนและความแปรปรวนร่วมมีลักษณะเป็นแบบยูนิฟอร์ม โดยใช้ฟังก์ชันสำเร็จรูป ORDGEE พบว่า วิธีการของ MLR มีค่าค่าความถูกต้องในการจัดกลุ่มสูงสุด

ข้อมูลชุดที่ 2 เมื่อกำหนดให้โครงสร้างของความแปรปรวนและความแปรปรวนร่วมมีลักษณะเป็นแบบอิสระและยูนิฟอร์ม โดยใช้ฟังก์ชันสำเร็จรูป REPOLR พบว่า ทั้ง 2 วิธีการวิเคราะห์ มีค่าค่าความถูกต้องในการจัดกลุ่มเท่ากัน

ข้อมูลชุดที่ 3 เมื่อกำหนดให้โครงสร้างของความแปรปรวนและความแปรปรวนร่วมมีลักษณะเป็นแบบอิสระและยูนิฟอร์ม โดยใช้ฟังก์ชันสำเร็จรูป REPOLR พบว่า ทั้ง 2 วิธีการวิเคราะห์ มีค่าค่าความถูกต้องในการจัดกลุ่มเท่ากัน

เมื่อพิจารณาถึงความเหมาะสมในการวิเคราะห์ข้อมูล ผู้วิจัยแนะนำให้ใช้วิธี MLR โดยใช้เทคนิค GEE เนื่องจากเป็นวิธีง่ายและสะดวกในการประยุกต์ใช้กับข้อมูลระยะยาว และสอดคล้องกับเงื่อนไขของข้อมูล สำหรับวิธีการของ MDA มีเงื่อนไขที่ใช้ในการวิเคราะห์มาก เช่น เงื่อนไขของตัวแปรอิสระต้องมีการแจกแจงแบบปกติหลายตัวแปร สำหรับฟังก์ชันสำเร็จรูปผู้วิจัยแนะนำให้ใช้ฟังก์ชัน REPOLR เนื่องจากมีประสิทธิภาพในการวิเคราะห์ดีกว่าฟังก์ชัน ORDGEE

ภาควิชา สถิติ ลายมือชื่อนิสิต ศุภรัตน์ ดิษบรรจง

สาขาวิชา สถิติ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก

ปีการศึกษา 2552

4982230426 : MAJOR STATISTICS

KEYWORDS : DISCRIMINANT ANALYSIS / MULTINOMIAL LOGISTIC REGRESSION / LONGITUDINAL / GENERALIZED ESTIMATING EQUATION / MAXIMUM A POSTERIORI

SUPARAT DITBANJONG : COMPARISON BETWEEN DISCRIMINANT ANALYSIS AND MULTINOMIAL LOGISTIC REGRESSION WITH LONGITUDINAL DATA.

THESIS ADVISOR : ASSOC.PROF.SUPOL DURONGWATANA,Ph.D., 111 pp.

The objective of this study is to compare two methods of multiple classification for longitudinal data. These two methods are multiple discriminant method (MDA) using Bayes' rule as a classification rule and multiple logistic regression method (MLR) using generalized estimating equation (GEE) as a classification rule. The data consist of independent variables and a dependent variable. The dependent variable is categorical one which there are multiple categories. The independent variables are either quantitative variable or categorical variable. Three actual data sets from medical studies were analyzed using both methods (e.g. insomnia study, hip replacement study and achillies tendon rupture study, respectively). The correct classification rate is used as a criterion for comparing these two methods. The results are summarized as follows.

From the first data set, when variance-covariance structure is uniform, and ORDGEE built-in function is used, it is found that the MLR method yields the higher correct classification rate.

From the second data set, when variance-covariance structure is independent and uniform, and REPOLR built-in function is applied, it is found that both methods yield equal .correct classification rate.

From the third data set, when variance-covariance structure is independent and uniform, and REPOLR built-in function is applied, it is found that both methods yield equal correct classification rate.

For appropriateness of data analysis, it is recommended that the MLR method with GEE technique should be used because it is easily and conveniently applied to longitudinal data. In contrast, the MDA method is required several assumptions such as multivariate normality of independent variables. For built-in function selection, it is recommended REPOLR should be used because it is more effective than ORDGEE function in all cases of variance-covariance structure.

Department : Statistics Student's Signature Suparat Ditbanjong

Field of Study : Statistics Advisor's Signature

Academic Year : 2009

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำตลอดจนกระทั่งความช่วยเหลือ แก้ไขข้อบกพร่องต่าง ๆ ด้วยดีเสมอมา จนวิทยานิพนธ์นี้เสร็จสมบูรณ์ จึงใคร่ขอขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ธีระพร วีระถาวร ในฐานะประธานกรรมการ ผู้ช่วยศาสตราจารย์ ดร. บุญอ้อม โฉมที กรรมการภายนอกมหาวิทยาลัย ผู้ช่วยศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์ กรรมการสอบวิทยานิพนธ์ที่ช่วยตรวจสอบแก้ไขวิทยานิพนธ์นี้ให้สมบูรณ์ยิ่งขึ้น และขอกราบขอบพระคุณอาจารย์ภาควิชาสถิติที่ได้ให้ความรู้แก่ผู้วิจัย

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณยิ่งในความเมตตาจากบิดา มารดา น้า พี่ชาย และเพื่อน ๆ ผู้ซึ่งเป็นกำลังแรงใจ ให้ผู้วิจัยได้มีกำลังในการทำวิทยานิพนธ์ฉบับนี้ให้เสร็จสิ้น จึงขอกราบขอบพระคุณมา ณ ที่นี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญภาพ..... ฐ

บทที่ 1 บทนำ..... 1

1.1 ความเป็นมาและความสำคัญของปัญหา..... 1

1.2 วัตถุประสงค์ของการวิจัย..... 5

1.3 สมมติฐานของการวิจัย..... 6

1.4 ขอบเขตของการวิจัย..... 6

1.5 ข้อตกลงเบื้องต้น..... 8

1.6 คำจำกัดความ..... 9

1.7 ขั้นตอนการดำเนินงาน..... 9

1.8 เกณฑ์ที่ใช้ในการตัดสินใจ..... 10

1.9 ประโยชน์ที่คาดว่าจะได้รับ..... 10

บทที่ 2 ทฤษฎีและวิธีการทางสถิติ..... 11

2.1 ข้อมูลระยะยาว..... 11

2.2 การวิเคราะห์จำแนกประเภท..... 12

2.2.1 ตัวแบบที่ศึกษา..... 12

2.2.2 คุณสมบัติ..... 13

2.2.3 วิธีการประมาณค่าฟังก์ชันจำแนกประเภทและการจัดกลุ่ม..... 17

2.3 การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค..... 22

2.3.1 ตัวแบบที่ศึกษา..... 22

2.3.2 คุณสมบัติ..... 22

2.3.3 การประมาณค่าพารามิเตอร์..... 27

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.4 ลักษณะการทำงานของฟังก์ชันสำเร็จรูปในการวิเคราะห์ข้อมูล..... 32
2.4.1 การวิเคราะห์จำแนกประเภท..... 32
2.4.2 การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค..... 33
บทที่ 3 วิธีดำเนินการวิจัย..... 35
3.1 แผนการทดลอง..... 35
3.2 ขั้นตอนในการดำเนินงาน..... 37
3.2.1 การตรวจสอบลักษณะข้อมูล..... 39
3.2.2 การประมาณค่าพารามิเตอร์และการจัดกลุ่ม..... 41
บทที่ 4 ผลการวิเคราะห์ข้อมูล..... 47
4.1 กรณีศึกษาที่ 1: การติดตามผลของโรคนอนไม่หลับ..... 48
4.2 กรณีศึกษาที่ 2: การผ่าตัดเปลี่ยนผิวข้อสะโพก..... 57
4.3 กรณีศึกษาที่ 3: การรักษาอาการเอ็นร้อยหวายฉีก..... 65
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ..... 73
5.1 สรุปผลการวิจัย..... 74
5.2 อภิปรายผลการวิจัย..... 79
5.3 ข้อเสนอแนะ..... 81
รายการอ้างอิง..... 82
ภาคผนวก..... 85
ภาคผนวก ก ลักษณะโดยทั่วไปของข้อมูลที่ทำการศึกษา..... 86
ภาคผนวก ข ผลการวิเคราะห์ข้อมูล..... 93
ภาคผนวก ค รายละเอียดโปรแกรมคำสั่งที่ใช้ในการวิจัย..... 108
ประวัติผู้เขียนวิทยานิพนธ์..... 111

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่		หน้า
4.1	ค่าความน่าจะเป็นก่อนที่ใช้ในการวิเคราะห์ข้อมูล กรณีศึกษาที่ 1.....	49
4.2	สัมประสิทธิ์ของฟังก์ชันจำแนกในแต่ละกรณี กรณีศึกษาที่ 1.....	50
4.3	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนองเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 1.....	51
4.4	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์เปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 1.....	51
4.5	ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน ordgee กรณีศึกษาที่1.....	52
4.6	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน ordgee กรณีศึกษาที่ 1.....	53
4.7	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์มเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน ordgee กรณีศึกษาที่ 1.....	54
4.8	ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน repolr กรณีศึกษาที่ 1.....	54
4.9	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล สำหรับฟังก์ชัน repolr กรณีศึกษาที่ 1.....	55
4.10	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์มเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน repolr กรณีศึกษาที่ 1.....	56

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่		หน้า
4.11	ค่าความน่าจะเป็นก่อนที่ใช้ในการวิเคราะห์ข้อมูล กรณีศึกษาที่ 2.....	58
4.12	สัมประสิทธิ์ของฟังก์ชันจำแนกในแต่ละกรณี กรณีศึกษาที่ 2.....	58
4.13	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนองเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 2.....	59
4.14	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์เปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 2.....	60
4.15	ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน ordgee กรณีศึกษาที่ 2.....	61
4.16	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน ordgee กรณีศึกษาที่ 2.....	62
4.17	ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน repolr กรณีศึกษาที่ 2.....	62
4.18	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล สำหรับฟังก์ชัน repolr กรณีศึกษาที่ 2.....	63
4.19	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์มเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน repolr กรณีศึกษาที่ 2.....	64
4.20	ค่าความน่าจะเป็นก่อนที่ใช้ในการวิเคราะห์ข้อมูล กรณีศึกษาที่ 3.....	66
4.21	สัมประสิทธิ์ของฟังก์ชันจำแนกในแต่ละกรณี กรณีศึกษาที่ 3.....	66
4.22	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนองเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 3.....	67

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่		หน้า
4.23	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์เปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 3	68
4.24	ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน ordgee กรณีศึกษาที่ 3	69
4.25	ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน repolr กรณีศึกษาที่ 3	70
4.26	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล สำหรับฟังก์ชัน repolr กรณีศึกษาที่ 3	71
4.27	ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์มเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน repolr กรณีศึกษาที่ 3	71
ก.1	ตารางแจกแจงความถี่ของตัวแปรตอบสนองของกรณีศึกษาที่ 1 จำแนกตามระยะเวลาและประเภทของยาที่ใช้ในการรักษา	86
ก.2	ความถี่และความน่าจะเป็นเชิงประจักษ์สำหรับตัวแปรตอบสนองในแต่ละกลุ่ม.	87
ก.3	ตารางแจกแจงความถี่ของตัวแปรตอบสนองของกรณีศึกษาที่ 2 จำแนกตามระยะเวลาและเพศ	88
ก.4	ความถี่และความน่าจะเป็นเชิงประจักษ์สำหรับตัวแปรตอบสนองในแต่ละกลุ่ม.	88
ก.5	ตารางแจกแจงความถี่ของตัวแปรตอบสนองของกรณีศึกษาที่ 3 จำแนกตามระยะเวลาและประเภทของผ้าที่ใช้พัน	89
ก.6	ความถี่และความน่าจะเป็นเชิงประจักษ์สำหรับตัวแปรตอบสนองในแต่ละกลุ่ม.	90
ก.7	การทดสอบคุณสมบัติการแจกแจงปกติของตัวแปรเวลา สำหรับกรณีศึกษาที่ 1	90
ก.8	การทดสอบคุณสมบัติการเท่ากันของเมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่ม สำหรับกรณีศึกษาที่ 1	90
ก.9	การทดสอบคุณสมบัติการแจกแจงปกติของตัวแปรเวลา สำหรับกรณีศึกษาที่ 2	91
ก.10	การทดสอบคุณสมบัติการเท่ากันของเมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่ม สำหรับกรณีศึกษาที่ 2	91

ตารางที่		หน้า
ก.11	การทดสอบคุณสมบัติการแจกแจงปกติของตัวแปรเวลา สำหรับกรณีศึกษาที่ 3	91
ก.12	การทดสอบคุณสมบัติการเท่ากันของเมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่ม สำหรับกรณีศึกษาที่ 3.....	92

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพที่		หน้า
2.1	ลักษณะของข้อมูลวัดซ้ำ จำนวน n หน่วย เป็นเวลา t ช่วงระยะเวลา.....	11
2.2	ตัวอย่างรูปแบบขอบเขตของแต่ละกลุ่มในการจำแนกกลุ่ม.....	17
2.3	แผนผังการการจัดกลุ่ม.....	19
2.4	ภาพเปรียบเทียบการจำแนกกลุ่มระหว่างการใช้วิธีการ MAP กับวิธีการแบบอื่น	20
3.1	แสดงสถานการณ์ที่จะทำการวิเคราะห์ สำหรับข้อมูล 1 ชุด.....	35
4.1	แผนผังแสดงค่า ACC ที่ได้จากการคำนวณ กรณีศึกษาที่ 1.....	56
4.2	แผนผังแสดงค่า ACC ที่ได้จากการคำนวณ กรณีศึกษาที่ 2.....	64
4.3	แผนผังแสดงค่า ACC ที่ได้จากการคำนวณ กรณีศึกษาที่ 3.....	72
5.1	แสดงโครงสร้างของสถานการณ์ที่ทำการวิเคราะห์ สำหรับข้อมูล 1 ชุด.....	73

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1
บทนำ

งานวิจัยนี้เป็นการเปรียบเทียบการวิเคราะห์ข้อมูลจำแนกประเภทกับการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค สำหรับข้อมูลระยะยาว โดยผู้วิจัยจะทำการเปรียบเทียบการวิเคราะห์ข้อมูลโดยใช้ข้อมูลทุติยภูมิ ซึ่งเป็นข้อมูลทางด้านการแพทย์จำนวน 3 ชุดข้อมูล โดยนำข้อมูลที่ได้มาทำการวิเคราะห์ เพื่อเปรียบเทียบประสิทธิภาพ ความเหมือน และความแตกต่างสำหรับแต่ละวิธีการวิเคราะห์ข้อมูล ในบทนำนี้ผู้วิจัยจะกล่าวถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ในการวิจัย สมมติฐาน ขอบเขตการวิจัย ข้อตกลงเบื้องต้น ข้อจำกัดในการวิจัย นิยามศัพท์ที่ใช้ในการวิจัยครั้งนี้ และประโยชน์ที่ได้รับ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันนี้ได้มีการนำความรู้ทางสถิติมาประยุกต์ใช้กันเป็นอย่างมาก ในบางครั้งวิธีการทางสถิติจะถูกนำไปใช้ในการตอบคำถามที่เกิดจากผู้ที่ทำการศึกษา ในสิ่งที่เราสนใจศึกษานั้น มีสาเหตุเกิดจากสิ่งใด ปัจจัยใดมีอิทธิพลต่อการเกิดหรือส่งผลกระทบต่อปัญหานั้น หรือแม้กระทั่งผลลัพธ์ที่เกิดขึ้นนั้นมีลักษณะเป็นอย่างไร ซึ่งจะพบว่าข้อมูลที่สนใจนั้นอาจเกิดจากการสังเกตสิ่งที่อยู่รอบตัว หรืออาจเป็นการทดลอง ดังนั้น ผู้วิจัยมีความเห็นว่าความรู้ทางสถิติมีความเกี่ยวข้องกับสิ่งที่อยู่รอบตัวเราเสมอ ขึ้นอยู่กับว่า สิ่งที่เราสนใจนั้น จะนำความรู้ทางสถิติไปประยุกต์ใช้กับศาสตร์ในด้านต่าง ๆ ได้อย่างไร ตัวอย่างเช่น ทางด้านการแพทย์ เศรษฐศาสตร์ พฤติกรรมศาสตร์ สังคมศาสตร์ อุทกศาสตร์ ดาราศาสตร์ เป็นต้น

สาเหตุที่ผู้วิจัยมีความสนใจในการศึกษาหัวข้อนี้ เนื่องจาก เหตุผลส่วนตัวหลายประการ แต่จะขอยกมา 3 ประการที่เห็นได้ชัดจากประสบการณ์ของผู้วิจัย ได้แก่ ประการแรก ผู้วิจัยเคยมีปัญหาด้านสุขภาพและอาการป่วยนั้นส่งผลกระทบต่อการดำเนินชีวิตเป็นอย่างมาก ทำให้ผู้วิจัยต้องหยุดเรียนไป 1 ปีการศึกษา ในช่วงอายุ 15 ปี ในตอนนั้นผู้วิจัยต้องไปพบแพทย์ทุก 2 อาทิตย์ ซึ่งแพทย์จะทำการตรวจเลือด ให้ยา รวมทั้งประเมินความรุนแรงของโรคตลอดเวลาที่ทำการรักษา ผู้วิจัยจึงเห็นความสำคัญของการติดตามผลการรักษา และมีความรู้สึกว่าในบางครั้ง การวิเคราะห์วินิจฉัยโรคไม่สามารถประเมินได้จากเหตุการณ์ ณ เวลาใดเวลาหนึ่งได้ จะต้องมีการติดตามผลไปเรื่อย ๆ เพื่อที่จะสามารถรักษาโรคได้ตรงจุด ประการที่สอง ผู้วิจัยได้ทำงานทางด้านวิจัยตลาดในขณะที่กำลังศึกษาในระดับบัณฑิตศึกษา ทำให้ผู้วิจัยได้มีโอกาสได้เห็นข้อมูล ลักษณะการเก็บ

ข้อมูลในรูปแบบต่าง ๆ สิ่งหนึ่งที่ผู้วิจัยมีความรู้สึกว่าข้อมูลระยะยาวสามารถนำมาใช้ประโยชน์ได้ในทางด้านการตลาด เช่น การวิเคราะห์ถึง ความภักดีในตัวผลิตภัณฑ์ (Royality of brand) โดยการติดตามพฤติกรรมของผู้บริโภค เช่น การเลือกยี่ห้อของผลิตภัณฑ์ที่ใช้ อาจทำการศึกษาหาว่าปัจจัยใดบ้างที่มีผลต่อการเปลี่ยนยี่ห้อผลิตภัณฑ์ของผู้บริโภค และประการสุดท้าย คือ ในขณะที่ผู้วิจัยได้ทำวิทยานิพนธ์หัวข้อนี้ มารดาของผู้วิจัยได้ป่วยเป็นโรคมะเร็งเต้านม ในระหว่างที่ผู้วิจัยได้ดูแลมารดาของผู้วิจัย ผู้วิจัยได้สังเกตว่า การรักษาด้วยเคมีบำบัด ตัวยาที่ใช้ในการรักษาไม่ได้มีเพียงชนิดเดียว และให้ผลในการรักษาแตกต่างกัน เมื่อเปลี่ยนยาที่ทำการรักษา อาจจะมีผลกระทบต่อการลุกลามของเนื้อร้ายได้เมื่อตัวยาไม่ตอบสนองต่อโรค ด้วยสาเหตุสามประการนี้ทำให้ผู้วิจัยมีความสนใจที่จะศึกษาการวิเคราะห์ข้อมูลระยะยาวเพื่อที่จะได้เป็นแนวทางและเป็นประโยชน์ต่อผู้ที่จะต้องการศึกษาต่อไปในอนาคต

โดยทั่วไปแล้ววิธีการวิเคราะห์ข้อมูลจะมีเงื่อนไขและข้อกำหนดในการวิเคราะห์ข้อมูลที่แตกต่างกันออกไป ซึ่งการวิเคราะห์ข้อมูลโดยใช้เทคนิคในการพยากรณ์นั้น มีวัตถุประสงค์ 2 ประการ คือ ประการแรก เพื่อสร้างสมการพยากรณ์ที่เหมาะสมสำหรับข้อมูลนั้น ประการที่สอง คือ การนำสมการพยากรณ์ที่ได้มาใช้พยากรณ์ข้อมูลในอนาคต ซึ่งสมการที่ใช้ในการพยากรณ์ จะส่งผลต่อความถูกต้องแม่นยำในการพยากรณ์ได้มากน้อยเพียงใด อาจขึ้นอยู่กับวิธีการวิเคราะห์ข้อมูลในการประมาณค่าพารามิเตอร์ที่เหมาะสม สำหรับการวิเคราะห์ข้อมูลที่มีตัวแปรตามเป็นข้อมูลเชิงคุณภาพ พบว่า การวิเคราะห์การถดถอยโลจิสติค (Logistic Regression) ได้รับความนิยมมากขึ้นในปัจจุบัน โดยการวิเคราะห์การถดถอยโลจิสติคนี้ มีวัตถุประสงค์เพื่อหาสาเหตุว่ามีตัวแปรหรือปัจจัยใดบ้างที่ทำให้ตัวแปรตามในแต่ละกลุ่มมีความแตกต่างกัน ลักษณะของตัวแปรตามจะอยู่ในรูปตัวแปรเชิงกลุ่ม (Categories Variable) โดยงานวิจัยนี้จะทำการวิเคราะห์ข้อมูลที่ลักษณะของตัวแปรตามเป็นตัวแปรเชิงกลุ่มมากกว่า 2 กลุ่ม ซึ่งเรียกว่า การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค นอกจากนี้ การวิเคราะห์อีกประเภทหนึ่งที่มีวัตถุประสงค์เช่นเดียวกับการวิเคราะห์โลจิสติค คือ การวิเคราะห์จำแนกประเภท (Discriminant Analysis) ถึงแม้ว่าจะมีวัตถุประสงค์ในการวิเคราะห์ข้อมูลเดียวกัน แต่สิ่งที่มีความแตกต่างกันคือ เงื่อนไข (Assumption) ในการวิเคราะห์ข้อมูล ด้วยเหตุนี้ จึงเป็นสิ่งที่น่าศึกษาว่าการวิเคราะห์ข้อมูลด้วยวิธีการใดจะมีประสิทธิภาพมากกว่ากันภายใต้สถานการณ์ข้อมูลจริง

เมื่อพิจารณาถึงลักษณะของข้อมูลที่ผู้วิจัยได้ทำการศึกษา พบว่า เมื่อมีข้อมูลวัดซ้ำ (Repeated Measures) ซึ่งเป็นลักษณะหนึ่งของข้อมูลระยะยาว (Longitudinal Data) จะหมายถึง ข้อมูลที่เก็บรวบรวมจากหน่วยสังเกตเดียวกันหลายช่วงระยะเวลา (Multiple Time Period) โดยข้อมูลวัดซ้ำแตกต่างจากข้อมูลตัดขวางมากกว่า 1 ช่วงเวลา เพราะข้อมูลตัดขวางมากกว่า 1 ช่วงเวลา ในแต่ละช่วงเวลา หน่วยตัวอย่าง (Sampling Unit) หรือหน่วยสังเกต (Observation Unit) และจำนวนรายการของข้อมูลไม่จำเป็นต้องเท่ากัน นั่นคือ $n_1 \neq n_2 \neq ... \neq n_t$ ซึ่งแตกต่างจากข้อมูลวัดซ้ำซึ่งต้องทำการเก็บข้อมูลทั้งหมด t ช่วงระยะเวลา ที่หน่วยตัวอย่างหรือหน่วยสังเกตเดิมทุกช่วงระยะเวลา ดังนั้น จำนวนหน่วยตัวอย่างหรือหน่วยสังเกตเท่ากับ n หน่วย และด้วยสัญลักษณ์เดียวกัน ข้อมูลซ้ำก็มีความแตกต่างจากข้อมูลอนุกรมเวลา (Time-series Data) ที่ว่าข้อมูลอนุกรมเวลามีการเก็บข้อมูลเป็นแบบอนุกรม หมายความว่าเก็บข้อมูลต่อเนื่องตามระยะเวลา ขาดช่วงไม่ได้ แต่ข้อมูลวัดซ้ำไม่จำเป็นต้องต่อเนื่อง[1]

ตัวอย่างผลงานวิจัยสำหรับข้อมูลระยะยาวในประเทศไทยยังไม่ค่อยมีการทำวิจัยกันอย่างแพร่หลาย ซึ่งโดยส่วนใหญ่การวิเคราะห์ข้อมูลระยะยาวตัวแปรตามจะเป็นตัวแปรเชิงปริมาณหรือเป็นตัวแปรเชิงคุณภาพที่มีจำนวนกลุ่ม 2 กลุ่ม ดังตัวอย่าง หัวข้องานวิจัยเกี่ยวกับโมเดลการเปลี่ยนแปลงภายในบุคคลและสาเหตุของการเปลี่ยนแปลง กรณีสาเหตุของการเปลี่ยนแปลงความสุขใจของนักเรียน ซึ่งวิเคราะห์การเปลี่ยนแปลงความสุขใจของนักเรียน โดยเริ่มต้นทำการศึกษานักเรียน เรียนอยู่ในชั้นมัธยมศึกษาปีที่ 1 เทอมต้น จนกระทั่งจบมัธยมศึกษาปีที่ 3 ตอนปลาย พบว่าในช่วงภาคต้นของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ครอบครัวมีอิทธิพลต่อความสุขใจมากกว่าโรงเรียนเล็กน้อย แต่เมื่อศึกษาค่าสัมประสิทธิ์ของตัวแปรดัชนีทุกตัวพบว่า ดัชนีสภาพแวดล้อมในโรงเรียน ได้แก่ เจตคติต่อครู และการรับรู้ความสัมพันธ์ระหว่างนักเรียนกับครูเป็นตัวแปรเชิงสาเหตุที่สำคัญของการเปลี่ยนแปลงความสุขใจ[1]

---

[1] สุพล ดุรงค์วัฒนา, ตัวแบบและการวิเคราะห์ความถดถอยสำหรับการวิจัยขั้นสูง, (กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย, 2549), หน้า 753-754.

[2] ดุษฎี โยเหลา, "โมเดลการเปลี่ยนแปลงภายในบุคคลและสาเหตุของการเปลี่ยนแปลง," เอกสารประกอบการอบรมเรื่อง โมเดลสมการโครงสร้างเชิงเส้นโดยใช้โปรแกรม LISREL สำหรับข้อมูลช่วงยาว เสนอที่สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ 8 พฤษภาคม 2551. (เอกสารไม่ตีพิมพ์เผยแพร่)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ในการวิจัยครั้งนี้ ได้ทำการเปรียบเทียบการวิเคราะห์ทั้ง 2 ประเภท ได้แก่ การวิเคราะห์จำแนกประเภท และการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ซึ่งตัวแบบที่ทำการวิเคราะห์ มีลักษณะดังนี้

1) <u>ตัวแบบการวิเคราะห์จำแนกประเภท</u>

$$z_j = w_1 x_m + w_2 (time)_m \quad ; j = \min\{p, K-1\} \quad ; m = 1,\ldots,nT$$

โดยที่ p คือ จำนวนตัวแปรอิสระ มีค่าเท่ากับ 2

K คือ จำนวนกลุ่มของตัวแปรตาม

z_j คือ ฟังก์ชันจำแนก

w_1 , w_2 คือ น้ำหนักการจำแนกประเภทของตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2 ตามลำดับ

x คือ ตัวแปรกำหนด

$time$ คือ ตัวแปรเวลา

n คือ ขนาดตัวอย่าง (Sample Size)

T คือ ระยะเวลาที่เก็บข้อมูลซ้ำ (Preiod)

2) <u>ตัวแบบการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค</u>

$$logit\left[P(Y_{itk} \le k)\right] = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time)_{it} \quad , k = 1,\ldots,K-1 \;; t = 1,\ldots,T \;; i = 1,\ldots,n$$

โดยที่ $\left[P(Y_{itk} \le k)\right]$ คือ ความน่าจะเป็นที่ค่า Y_{it} จะน้อยกว่าหรือเท่ากับกลุ่มที่ k

β_{0k} คือ ค่า cut-point ของกลุ่มที่ k

β_1 , β_2 คือ สัมประสิทธิ์การถดถอย

K คือ จำนวนกลุ่มของตัวแปรตาม

x_{it} คือ ตัวแปรกำหนด สำหรับค่าสังเกตที่ i ระยะเวลาที่ t

$time_{it}$ คือ ตัวแปรเวลาสำหรับค่าสังเกตที่ i ระยะเวลาที่ t

n คือ ขนาดตัวอย่าง (Sample Size)

T คือ ระยะเวลาที่เก็บข้อมูลซ้ำ (Preiod)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

โดยปกติทั่วไปนั้น แนวทางในการเลือกวิธีการวิเคราะห์ข้อมูลระหว่างวิธีการวิเคราะห์จำแนกประเภท กับการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ผู้วิจัยส่วนมากมักจะใช้วิธีการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค เนื่องจากข้อจำกัดของวิธีการวิเคราะห์จำแนกประเภท ซึ่งได้แก่

1) ตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร

2) เมทริกซ์ค่าความแปรปรวนร่วมของตัวแปรของข้อมูล k กลุ่ม จะต้องมีค่าเท่ากัน

อย่างไรก็ตามบ่อยครั้งที่ข้อมูลไม่เป็นไปตามข้อกำหนดดังกล่าว ซึ่งในงานวิจัยครั้งนี้ได้ทำการศึกษาข้อมูลจริงทางด้านการแพทย์ ซึ่งเกิดสถานการณ์ข้อกำหนดของตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปรไม่เป็นจริง ซึ่งอาจส่งผลกระทบต่อความถูกต้องในการประมาณค่าฟังก์ชันการจำแนกที่ได้ [1] ถึงแม้ว่าจะส่งผลต่อความไม่ถูกต้องในการประมาณค่า แต่ก็ได้มีวิธีการประมาณค่าโดยใช้ทฤษฎีเบส์มาใช้ในการวิเคราะห์จำแนกประเภทโดยจะพิจารณาถึงการถ่วงน้ำหนักโดยใช้ความน่าจะเป็นก่อนในการจัดกลุ่ม และอีกวิธีการหนึ่งคือ การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ผู้วิจัยได้นำเสนอการใช้เทคนิค Generalized Estimating Equations (GEE) หรือเรียกว่า วิธีการสมการประมาณค่าโดยนัยทั่วไป เป็นวิธีที่ได้รับความนิยมในทางด้านการแพทย์ และชีวสถิติ ซึ่งวิธีการนี้เป็นวิธีวิเคราะห์ข้อมูลระยะยาว โดยศึกษาถึงโครงสร้างความสัมพันธ์ของข้อมูลในแต่ละช่วงเวลาในรูปแบบต่าง ๆ

จากสาเหตุที่ได้กล่าวมาข้างต้น ผู้วิจัยมีความเห็นว่า ถ้ามีการศึกษาข้อมูลระยะยาว และทำการวิเคราะห์หาสมการพยากรณ์ เมื่อตัวแปรตามเป็นตัวแปรเชิงคุณภาพที่มีจำนวนกลุ่มมากกว่า 2 กลุ่ม จะเป็นประโยชน์ต่อการศึกษาและพัฒนาความรู้ทางสถิติได้เป็นอย่างมาก เนื่องจากในปัจจุบัน วิธีการวิเคราะห์ดังกล่าวในประเทศไทย ยังไม่เป็นที่แพร่หลายมากนัก

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่ม สำหรับการวิเคราะห์จำแนกประเภท เมื่อกำหนดให้ความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตาม และความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์

---

[1] ศิริชัย กาญจนวาสี, การวิเคราะห์พหุระดับ MULTI-LEVAL ANALYSIS, (กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย : 2550), หน้า 51.

2. เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่มระหว่างฟังก์ชัน repolr และ ordgee ในการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

3. เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่ม สำหรับการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติคเมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระและเป็นแบบ ยูนิฟอร์ม

4. เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่มระหว่างวิธีการวิเคราะห์ข้อมูลจำแนกประเภท และการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

1.3 สมมติฐานของการวิจัย

การวิเคราะห์ข้อมูลจำแนกประเภทในกรณีที่ใช้วิธีการของเบส์ในการจัดกลุ่ม ความสามารถในการจัดกลุ่มข้อมูลได้ถูกต้อง จะไม่แตกต่างจากการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค โดยใช้วิธีการสมการประมาณค่าโดยนัยทั่วไป

1.4 ขอบเขตของการวิจัย

ขอบเขตการวิจัยในวิทยานิพนธ์ฉบับนี้จะครอบคลุมถึงการศึกษาในเรื่องต่อไปนี้

1. ลักษณะตัวแปรตาม ตัวแปรตามที่ทำการศึกษามีลักษณะเป็นข้อมูลเชิงกลุ่ม (categorical variables) ที่มีจำนวนกลุ่มมากกว่า 2 กลุ่ม โดยใช้ข้อมูลจริงทางด้านการแพทย์ 3 กลุ่มข้อมูล และลักษณะตัวแปรตามของข้อมูลมีลักษณะเป็นแบบอันดับ (ordinal scale)

1.1 ข้อมูลชุดที่ 1 เป็นข้อมูลเกี่ยวกับการติดตามผลของโรคนอนไม่หลับ (Insomnia) ตัวแปร ตอบสนองมี 4 ระดับในการใช้เวลานอน จำนวนผู้ทดลอง 239 คน ระยะเวลาในการเก็บข้อมูล 2 ครั้ง [1]

1.2 ข้อมูลชุดที่ 2 เป็นข้อมูลเกี่ยวกับการผ่าตัดเปลี่ยนผิวข้อสะโพก(Hip resurfacing) ตัวแปร ตอบสนอง มี 4 ระดับความเจ็บปวด จำนวนผู้ทดลอง 58 คน ระยะเวลาในการเก็บข้อมูล 3 ครั้ง [1]

---

[1] แหล่งที่มา : http://www.stat.ufl.edu/~aa/. [20/04/2552]

1.3 ข้อมูลชุดที่ 3 เป็นข้อมูลเกี่ยวกับการอักเสบของเอ็น Achilles (Achilles tendon rupture) ตัวแปรตาม คือ ความสามารถในการทำกิจกรรม จำนวนผู้ทดลอง 48 คน ระยะเวลาในการเก็บข้อมูล 3 ครั้ง [2]

2. การวิเคราะห์ข้อมูลจำแนกประเภท ขอบเขตที่ทำการศึกษาสำหรับวิธีการจำแนกประเภท มีดังนี้

2.1 จะใช้หลักการของเบส์ (Bayes solution)

2.2 กำหนดให้ตัวแปรในแต่ละจุดระยะเวลา (time point) เป็นอิสระกัน

2.3 การวิเคราะห์จำแนกประเภทจะทำการศึกษา โดยกำหนดให้ค่าความน่าจะเป็นก่อน (prior probability) ที่ใช้ในการถ่วงน้ำหนัก แบ่งเป็น 2 รูปแบบ ดังนี้

1) กำหนดให้มีค่าเท่ากันทุกกลุ่มในแต่ละตัวแปรตาม

2) ให้มีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ในแต่ละกลุ่มของข้อมูลที่ทำการศึกษา

3. การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ขอบเขตที่ทำการศึกษาสำหรับวิธีการจำแนกประเภท มีดังนี้

3.1 ใช้เทคนิคการวิเคราะห์ GEE

3.2 การวิเคราะห์มัลติโนเมียลโลจิสติค จะพิจารณาฟังก์ชันที่ใช้สำหรับโปรแกรม R ที่ได้รับการนำเสนอสำหรับการวิเคราะห์นี้จำนวน 2 ฟังก์ชัน คือ repolr [3] และ ordgee [4] ในแต่ละรูปแบบความสัมพันธ์ดังกล่าว ซึ่งฟังก์ชันสำเร็จรูปทั้ง 2 ฟังก์ชัน ใช้เทคนิคการวิเคราะห์ แบบเดียวกัน

---

[1] แหล่งที่มา : http://www.warwick.ac.uk/go/repolr. [20/04/2552]

[2] แหล่งที่มา : http://www.warwick.ac.uk/go/repolr. [25/04/2552]

[3] Parsons N.R. and Matthew L. Costa, "Repeated measure proportional odds logistic regression analysis of ordinal score data in the statistical software package R," Journal of Science Direct, 53 (2008): 632.

[4] Laura A. Thompson, R (and S-plus) Manual to Accompany Agresti's Categorical Data Analysis (2008), p. 219.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

แต่รูปแบบในการเขียนฟังก์ชัน รวมทั้งคำสั่งที่ใช้ในการเขียนฟังก์ชันมีความแตกต่างกัน รายละเอียดของทั้ง 2 ฟังก์ชันมีดังนี้

1) ฟังก์ชัน ordgee ถูกสร้างขึ้นโดย Jun Yan พบว่า สร้างขึ้นโดยใช้สัญลักษณ์เหมือนกับฟังก์ชัน glm ซึ่งฟังก์ชัน glm พัฒนามาจากฟังก์ชัน lm และมีการนำฟังก์ชัน gee มาประยุกต์ใช้ในการทำงาน ในส่วนของการตรวจสอบค่าสูญหาย พบว่า โปรแกรมจะตรวจสอบเมื่อมีการเรียกใช้คำสั่ง gee เท่านั้น

2) ฟังก์ชัน repolr ได้ถูกสร้างขึ้นโดย Parson และคณะ พบว่า พัฒนาจากฟังก์ชัน GEE ซึ่งขั้นตอนในการแก้ปัญหาจะใช้การประมาณค่าพารามิเตอร์ของความสัมพันธ์โดยให้มีค่าต่ำที่สุด และมีการปรับค่าความแปรปรวนในแต่ละขั้นตอนการดำเนินงาน ซึ่งรูปแบบความสัมพันธ์ในการประมาณค่าสามารถมีได้หลายรูปแบบ

3.3 ความสัมพันธ์ของความคลาดเคลื่อนในแต่ละข้อมูลที่ทำการศึกษา มีรูปแบบ ดังนี้

1) ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระ

2) ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม

1.5 ข้อตกลงเบื้องต้น

1. ศึกษาภายใต้ข้อมูลจริงทางด้านการแพทย์จำนวน 3 กลุ่มข้อมูล ดังรายละเอียดที่กล่าวมาในข้อที่ 1. ในส่วนของขอบเขตการวิจัย

2. ตัวแปรอิสระแต่ละตัวถือเป็นค่าคงที่

3. กำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนระหว่างหน่วยตัวอย่างข้อมูลเป็นอิสระกัน

4. สำหรับการวิเคราะห์มัลติโนเมียลโลจิสติค กำหนดให้ภายในหน่วยตัวอย่างเดียวกันโครงสร้างของความสัมพันธ์ของข้อมูลในแต่ละช่วงเวลาที่ทำการศึกษามีความสัมพันธ์กับตามรูปแบบที่กำหนด คือ มีรูปแบบอิสระ และยูนิฟอร์ม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.6 คำจำกัดความ

1. Longitudinal data คือ ข้อมูลที่มีการเก็บรวบรวมซ้ำ ๆ กันมากกว่า 1 ครั้ง

2. Maximum Posteriori Estimator คือ วิธีการประมาณค่าพารามิเตอร์ทางสถิติโดยใช้ทฤษฎีเบส์และใช้ความน่าจะเป็นก่อนในการประมาณค่าพารามิเตอร์

3. Generalized Estimating Equations (GEE) คือ วิธีการวิเคราะห์ข้อมูลระยะยาวโดยพัฒนาจากวิธีการของ Generalized Linear Model

4. Marginal Model คือ ตัวแบบที่พยายามหาค่าเฉลี่ยระดับประชากรว่าลักษณะที่เปลี่ยนไปตามกาลเวลานั้นมีผลอย่างไรต่อผลลัพธ์ที่ได้

5. Working correlation matrix คือ เมทริกซ์ความสัมพันธ์ของข้อมูลที่ถูกกำหนดขึ้นให้มีรูปแบบที่แตกต่างกันไป ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอย

1.7 ขั้นตอนในการดำเนินการ

1. ศึกษารูปแบบของตัวแปรตามและตัวแปรอิสระของข้อมูลที่ทำการศึกษา

2. สร้างตัวแบบโดยมีวิธีการวิเคราะห์หลัก 2 วิธี ได้แก่

1) <u>การวิเคราะห์จำแนกประเภท</u> สร้างตัวแบบโดยใช้วิธีตัวแบบของเบส์ ซึ่งในตัวแบบนี้นี้จะทำการสร้างตัวแบบย่อยโดยมีลักษณะดังนี้

1.1) ค่าที่ใช้ในการถ่วงน้ำหนักกำหนดให้ค่าความน่าจะเป็นเท่ากันทุกกลุ่ม

1.2) ค่าที่ใช้ในการถ่วงน้ำหนักกำหนดให้ค่าความน่าจะเป็นเท่ากับค่าความน่าจะเป็นที่แท้จริงของข้อมูล

2) <u>การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค</u> สร้างตัวแบบโดยใช้เทคนิคการวิเคราะห์ GEE จะทำการสร้างตัวแบบย่อยตามลักษณะความสัมพันธ์ของความคลาดเคลื่อน โดยมีลักษณะดังนี้

2.1) โครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบอิสระ

2.2) โครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบยูนิฟอร์ม

การวิเคราะห์มัลติโนเมียลโลจิสติค พบว่า ในโปรแกรม R ได้มีการเสนอฟังก์ชันสำเร็จรูป 2 ฟังก์ชัน คือ repolr และ ordgee ซึ่งผู้วิจัยพบว่ามีลักษณะการเขียนฟังก์ชันที่ใช้ในการวิเคราะห์

ข้อมูลมีความแตกต่างกันและให้ผลลัพธ์ที่แตกต่างกัน ดังนั้น ผู้วิจัยจึงได้ทำการทดสอบความแตกต่างในการทำงานและผลลัพธ์ที่ได้ของฟังก์ชันทั้ง 2 ฟังก์ชันนี้ด้วย

3. หาค่าพยากรณ์และนำมาเปรียบเทียบความถูกต้องที่ได้จากการพยากรณ์

4. สรุปผลในรูปแบบตารางและรูปภาพ

1.8 เกณฑ์ที่ใช้ในการตัดสินใจ

เกณฑ์ในการตัดสินใจ คือ ความสามารถในการพยากรณ์ข้อมูลได้ถูกต้อง (Accuracy :ACC)

$$\mathbf{ACC(\%)} = \frac{\text{จำนวนข้อมูลที่พยากรณ์ได้ถูกต้อง}}{(\text{จำนวนข้อมูลทั้งหมด})} \times 100$$

1.9 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถเปรียบเทียบวิธีการวิเคราะห์ข้อมูลในกรณีที่เป็นข้อมูลระยะยาว ซึ่งลักษณะของตัวแปรตามเป็นข้อมูลเชิงคุณภาพที่มีจำนวนกลุ่มมากกว่า 2 กลุ่ม และพิจารณาว่าวิธีการใดมีความเหมาะสมที่สุดในแต่ละสถานการณ์

2. เป็นแนวทางให้ผู้ที่ต้องการศึกษาสามารถเลือกวิธีการวิเคราะห์ข้อมูลได้เหมาะสม และมีประสิทธิภาพในการวิจัย เพื่อให้ได้สมการในการพยากรณ์ที่มีความถูกต้องมากที่สุด

3. เป็นแนวทางในการศึกษาวิธีการประมาณค่าพารามิเตอร์แบบอื่น ในการวิเคราะห์ข้อมูล

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2
ทฤษฎีและวิธีการทางสถิติ

การวิจัยครั้งนี้ได้ทำการศึกษาการเปรียบเทียบการวิเคราะห์ข้อมูลระยะยาว ด้วยวิธีการจำแนกประเภทและวิธีการถดถอยมัลติโนเมียลโลจิสติค ซึ่งในแต่ละวิธีการวิเคราะห์จะมีความแตกต่างในการประมาค่าพารามิเตอร์ ในบทนี้จะกล่าวถึงคุณสมบัติทั่วไปของวิธีการวิเคราะห์ในแต่ละวิธีและรายละเอียดในการประมาณค่าพารามิเตอร์ และความแตกต่างของการใช้ฟังก์ชันสำเร็จรูปในการวิเคราะห์มัลติโนเมียลโลจิสติค โดยใช้โปรแกรม R ในการวิเคราะห์ข้อมูล

2.1 ข้อมูลระยะยาว (Longitudinal Data)

เมื่อพิจารณาถึงหัวข้อที่ผู้วิจัยได้ทำการศึกษา พบว่า เมื่อมีข้อมูลวัดซ้ำ (Repeated Measure) ซึ่งเป็นลักษณะหนึ่งของข้อมูลระยะยาว (Longitudinal Data) จะหมายถึง ข้อมูลที่เก็บรวบรวมจากหน่วยสังเกตเดียวกันหลายช่วงระยะเวลา (Multiple Time Period) โดยข้อมูลวัดซ้ำแตกต่างจากข้อมูลตัดขวางมากกว่า 1 ช่วงเวลา เพราะข้อมูลตัดขวางมากกว่า 1 ช่วงเวลา ในแต่ละช่วงเวลา หน่วยตัวอย่าง (Sampling Unit) หรือหน่วยสังเกต (Observation Unit) และจำนวนรายการของข้อมูลไม่จำเป็นต้องเท่ากัน นั่นคือ $n_1 \neq n_2 \neq ... \neq n_t$ ซึ่งแตกต่างจากข้อมูลวัดซ้ำซึ่งต้องทำการเก็บข้อมูลทั้งหมด t ช่วงระยะเวลา ที่หน่วยตัวอย่างหรือหน่วยสังเกตเดิมทุกช่วงระยะเวลา ดังนั้น จำนวนหน่วยตัวอย่างหรือหน่วยสังเกตเท่ากับ n หน่วย และด้วยสัญลักษณ์เดียวกัน ข้อมูลซ้ำก็มีความแตกต่างจากอนุกรมเวลา (Time-series Data) ที่ว่าข้อมูลอนุกรมเวลามีการเก็บข้อมูลเป็นแบบอนุกรม หมายความว่า เก็บข้อมูลต่อเนื่องตามระยะเวลา ขาดช่วงไม่ได้ แต่ข้อมูลวัดซ้ำไม่จำเป็นต้องต่อเนื่อง ซึ่งมีลักษณะโครงสร้างดังนี้ [1]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

---

[1] สุพล ดุรงค์วัฒนา, ตัวแบบและการวิเคราะห์ความถดถอยสำหรับการวิจัยขั้นสูง, (กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย : 2549), หน้า 753-754

รูปภาพที่ 2.1 ลักษณะของข้อมูลวัดซ้ำ จำนวน n หน่วย เป็นเวลา t ช่วงระยะเวลา

ระยะเวลาที่ 1

หน่วย	X_{1i}	Y_{1i}
1	x_{11}	y_{11}
2	x_{12}	y_{12}
…	…	…
n	x_{1n}	y_{1n}

ระยะเวลาที่ 2

หน่วย	X_{2i}	Y_{2i}
1	X_{21}	Y_{21}
2	X_{22}	Y_{22}
…	…	…
n	X_{2n}	Y_{2n}

ระยะเวลาที่ t

หน่วย	X_{ti}	Y_{ti}
1	X_{t1}	Y_{t1}
2	X_{t2}	Y_{t2}
…	…	…
n	X_{tn}	Y_{tn}

2.2 การวิเคราะห์จำแนกประเภท

2.2.1 **ตัวแบบที่ศึกษา** มีลักษณะดังนี้

$$z_j = w_1 x_m + w_2 (time)_m \quad ; j = \min\{p, K-1\} \quad ; m = 1,2,\ldots,nT$$

โดยที่ p คือ จำนวนตัวแปรอิสระ มีค่าเท่ากับ 2

K คือ จำนวนกลุ่มของตัวแปรตาม

n คือ ขนาดตัวอย่าง (Sample Size)

T คือ ระยะเวลาที่ทำการเก็บข้อมูลซ้ำ (Period)

z_j คือ ฟังก์ชันจำแนก

w_1 , w_2 คือ น้ำหนักการจำแนกประเภทของตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2

x คือ ตัวแปรกำหนด

$time$ คือ ตัวแปรเวลา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.2.2 **คุณสมบัติ**

การวิเคราะห์จำแนกประเภท (Discriminant Analysis) เป็นเทคนิคการวิเคราะห์ความสัมพันธ์หรือการหาสาเหตุเทคนิคหนึ่ง โดยที่มีตัวแปรตาม 1 ตัว (Y) ซึ่งเป็นตัวแปรเชิงคุณภาพ และตัวแปรอิสระ p ตัว ($X_1,\ldots,X_p$) โดย $p \geq 1$ การวิเคราะห์จำแนกประเภทเป็นเทคนิคที่มีการแบ่งกลุ่มคน สัตว์ สิ่งของ ฯลฯ ออกเป็นกลุ่มอย่างน้อย 2 กลุ่มขึ้นไป ดังนั้นตัวแปรตามจะระบุกลุ่มที่หน่วยอยู่ โดยผู้ศึกษาจะต้องเป็นผู้แบ่งกลุ่มมาก่อนว่าจะให้แต่ละหน่วยอยู่ในกลุ่มใดและจำนวนกลุ่มต้องมีจำนวนอย่างน้อย 2 กลุ่มขึ้นไป และในแต่ละกลุ่มจะไม่มีการซ้ำซ้อนกัน โดยมีหลักเกณฑ์ในการแบ่ง คือ ให้หน่วยที่อยู่ในกลุ่มเดียวกันมีความคล้ายกันในปัจจัยหรือตัวแปรที่ใช้ในการแบ่งกลุ่ม ส่วนหน่วยที่อยู่ต่างกลุ่มกันจะมีความแตกต่างกันในตัวแปรที่ใช้แบ่งกลุ่ม

ข้อสมมติหรือเงื่อนไขของการวิเคราะห์จำแนกประเภท

1) ตัวแปรอิสระต้องมีการแจกแจงแบบปกติหลายตัวแปร

2) เมทริกซ์ค่าความแปรปรวนร่วมของตัวแปรของข้อมูล k กลุ่ม จะต้องมีค่าเท่ากัน

1. <u>ตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร</u> (Multivariate normal)

ในกรณีที่ต้องการตรวจสอบการแจกแจงแบบปกติหลายตัวแปร มีวิธีการตรวจสอบไม่มากนักและค่อนข้างยุ่งยากเมื่อตัวอย่างมีขนาดใหญ่ ในขณะที่การทดสอบการแจกแจงแบบปกติของตัวแปรหนึ่งตัวทำได้หลายวิธี อย่างไรก็ตามจะพบว่า ถ้าตัวแปรแต่ละตัว X_i (i=1,2,…,p) มีการแจกแจงแบบปกติแล้ว จะไม่สามารถสรุปได้ว่า เวกเตอร์ของตัวแปร $\underset{\sim}{x}$ ที่มีขนาด px1 จะมีการแจกแจงแบบปกติหลายตัวแปรด้วย นั่นหมายความว่า ถ้า X_i ทุกตัวมีการแจกแจงแบบปกติแล้ว เวกเตอร์ $\underset{\sim}{x}$ อาจจะมีการแจกแจงแบบปกติหลายตัวแปรหรือไม่ก็ได้ แต่ถ้าตรวจพบว่าตัวแปรใดตัวแปรหนึ่งไม่ได้มีการแจกแจงแบบปกติจะทำให้สรุปได้ว่าเวกเตอร์ของตัวแปร $\underset{\sim}{x}$ จะไม่มีการแจกแจงแบบปกติ แต่ถ้าสามารถตรวจสอบได้ว่า เวกเตอร์ $\underset{\sim}{x}$ มีการแจกแจงแบบปกติหลายตัวแปรจะทำให้สรุปได้เลยว่าตัวแปรแต่ละตัวมีการแจกแจงแบบปกติด้วย [1]

[1] กัลยา วานิชย์บัญชา, <u>การวิเคราะห์ข้อมูลหลายตัวแปร</u>, (ธรรมสาร: พฤศจิกายน 2548), หน้า 375.

วิธีตรวจสอบข้อกำหนด ผู้วิจัยใช้วิธีการ Shapiro-Wilk W Test ใช้ในการทดสอบการแจกแจงแบบปกติ ถ้าสถิติ W มีนัยสำคัญ สมมติฐานที่กำหนดว่าข้อมูลมีการแจกแจงแบบปกติจะถูกปฏิเสธ ซึ่งพบว่าวิธีการนี้มีคุณสมบัติที่ดีในการวิเคราะห์เมื่อเปรียบเทียบกับวิธีการตรวจสอบวิธีการอื่น เนื่องจากงานวิจัยนี้มีตัวแปรอิสระเป็นตัวแปรเชิงปริมาณเพียง 1 ตัว คือ ตัวแปรของเวลา ดังนั้น สามารถตรวจสอบเงื่อนไขการแจกแจงแบบปกติ 1 ตัวแปรได้

<u>การทดสอบสมมติฐาน</u>

H_0 : ตัวแปรเวลามีการแจกแจงแบบปกติ

H_1 : ตัวแปรเวลาไม่มีการแจกแจงแบบปกติ

สถิติทดสอบ

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

ค่า a_i คำนวณได้จาก

$$a_i = \frac{\sum_{j=1}^{n} m_j \underset{\sim}{\upsilon}_{ij}}{C}$$

โดย $\underset{\sim}{\upsilon}_{ij} = \mathrm{cov}(x_i, x_j) \quad (i, j = 1,2,...,n)$

$E(x_i) = m_i \quad (i = 1,2,...,n)$

$C^2 = \underset{\sim}{m}^T \mathbf{V}^{-1}\mathbf{V}^{-1} \underset{\sim}{m}$

เมื่อ $\underset{\sim}{m} = (m_1,...,m_n)^T$

ซึ่ง $m_1,...,m_n$ คือ ค่าคาดหวังของค่าสถิติที่เรียงลำดับที่มีการแจกแจงที่เหมือนกันและเป็นอิสระกันของตัวแปรสุ่ม โดยสุ่มจากการแจกแจงแบบปกติมาตรฐาน

$\mathbf{V}$ คือ เมทริกซ์ความแปรปรวนร่วม

จะปฏิเสธสมมติฐาน เมื่อ ค่า W ที่ได้มีค่าต่ำ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. <u>เมทริกซ์ค่าความแปรปรวนร่วมตัวแปรอิสระ k กลุ่ม จะต้องมีค่าเท่ากัน</u>

การทดสอบเมทริกซ์ค่าความแปรปรวนร่วมของข้อมูล k กลุ่ม จะใช้เทคนิคของ MANOVA ในการตรวจสอบวาเมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่มเท่ากันหรือไม่ มีสมมติฐานดังนี้

$H_0 \quad : \quad \mathbf{\Sigma_1} = \mathbf{\Sigma_2} = \mathbf{...} = \mathbf{\Sigma_k}$

$H_1 \quad : \quad \mathbf{\Sigma}_i \neq \mathbf{\Sigma}_j$ อย่างน้อย 1 คู่ ; i ≠ j

โดย $\mathbf{\Sigma}_i$ คือ เมทริกซ์ความแปรปรวนร่วมของตัวแปรตาม p ตัว ของกลุ่มที่ i ; i = 1,2,...,k

สถิติในการทดสอบ Box's M โดยที่

$$Box'sM \quad = \quad M \quad = \quad \frac{\left|\mathbf{S_1}\right|^{\frac{v_1}{2}} \cdot \left|\mathbf{S_2}\right|^{\frac{v_2}{2}} \cdot ... \cdot \left|\mathbf{S_k}\right|^{\frac{v_k}{2}}}{\left|\mathbf{S}_p\right|^{\frac{1}{2}\sum_{i=1}^{k} v_i}}$$

โดยที่ $0 < M < 1$ และ $v_i = n_i - 1$

S_i คือ เมทริกซ์ความแปรปรวนร่วมของกลุ่มตัวอย่างที่ i

S_p คือ เมทริกซ์ความแปรปรวนร่วมของกลุ่มทุกกลุ่ม

$$\mathbf{S}_p = \frac{\sum_{i=1}^{k} v_i \mathbf{S}_i}{\sum_{i=1}^{k} v_i} = \frac{\mathbf{W}}{\sum_{i=1}^{k} n_i - k}$$

โดยที่ $v_i > p$ มิฉะนั้นจะทำให้ค่า M ในสมการเป็น ศูนย์ เนื่องจากค่าดิเทอร์มิแนนท์ของตัวเศษในสมการจะเป็นศูนย์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Box ได้เสนอวิธีการแปลงสถิติทดสอบ Box's M ให้มีการแจกแจงแบบ F [1] ดังนี้

กำหนดให้ $$C_1 = \left[\sum_{i=1}^{k}\frac{1}{v_i^2} - \frac{1}{\sum_{i=1}^{k} v_i^2}\right]\left[\frac{2p^2+3p-1}{6(p+1)(k-1)}\right]$$

$$\ln M = \frac{1}{2}\sum_{i=1}^{k} v_i \ln|S_i| - \frac{1}{2}\sum_{i=1}^{k} v_i \ln|S_p|$$

$$C_2 = \frac{(p-1)(p+2)}{6(k-1)}\left[\sum_{i=1}^{k}\frac{1}{v_i^2} - \frac{1}{\sum_{i=1}^{k} v_i^2}\right]$$

กำหนดให้ $$a_1 = \frac{1}{2}(k-1)p(p+1) \quad , \quad a_2 = \frac{a_1+2}{\left|C_2 - C_1^2\right|}$$

$$b_1 = \frac{1-C_1-a_1/a_2}{a_1} \quad , \quad b_2 = \frac{1-C_1-2/a_2}{a_2}$$

สถิติทดสอบ F จะแบ่งเป็น 2 กรณี คือ

1) ถ้า $C_2 > C_1^2$ สถิติทดสอบ F จะเป็น $F = -2b_1 \ln M$

2) ถ้า $C_2 < C_1^2$ สถิติทดสอบ F จะเป็น $F = -\frac{a_2 b_2 \ln M}{a_1(1+2b_2 \ln M)}$

ค่าสถิติ F ทั้งสองสมการข้างต้น จะมีการแจกแจงโดยประมาณแบบ F ที่องศาอิสระ a_1 และ a_2 ตามลำดับ การสรุปผลการทดสอบ จะปฏิเสธสมมติฐานว่างถ้า $F > F_\alpha$ ที่องศาอิสระ a_1 และ a_2 ตามลำดับ

[1] Box G.E.P, "A general Distribution Theory for a class of likelihood Criteria," Biometrica 36 (1949): 317-346.

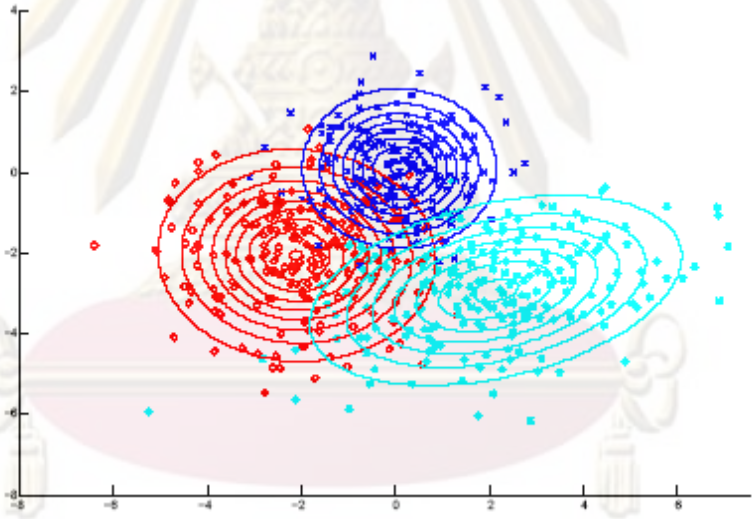
2.2.3 วิธีการประมาณค่าฟังก์ชันจำแนกประเภทและการจัดกลุ่ม

ทฤษฎีเบส์ จะมีรูปแบบดังนี้

$$P\left(\omega_C \mid \underset{\sim}{x}\right) = \frac{p\left(\underset{\sim}{x} \mid \omega_C\right)P(\omega_C)}{p(\underset{\sim}{x})}$$

เมื่อต้องการจัดกลุ่มของค่าสังเกต $\underset{\sim}{x}$ จะพิจารณาถึงค่าความน่าจะเป็นหลัง $P\left(\omega_C \mid \underset{\sim}{x}\right)$ ที่มีค่าสูงที่สุด ดังนั้นการพิจารณาถึง x-space ในการกำหนดให้อยู่ในขอบเขตของ $\Re_C$ ซึ่งจุดที่ตกอยู่ใน $\Re_C$ จะถูกกำหนดให้อยู่ในกลุ่ม ω_C

รูปภาพที่ 2.2 ตัวอย่างรูปแบบขอบเขตของแต่ละกลุ่มในการจำแนกกลุ่ม [1]



พิจารณาในกรณีของ 1-dimensional เฉพาะ space (x) และมีจำนวนกลุ่ม 2 กลุ่ม คือ ω_1 และ ω_2 เมื่อต้องการจัดกลุ่มสามารถทำได้โดยหาค่าความน่าจะเป็นที่น้อยที่สุดของความผิดพลาดในการจัดกลุ่มผิด

[1] แหล่งที่มา : http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn10-lec-2x3.pdf [15/08/2552]

ลักษณะของความคลาดเคลื่อนในการจำแนกกลุ่มผิด

1. กำหนดให้ x อยู่ในกลุ่มที่ ω_1 แต่ในความเป็นจริง x อยู่ในกลุ่มที่ ω_2

2. กำหนดให้ x อยู่ในกลุ่มที่ ω_2 แต่ในความเป็นจริง x อยู่ในกลุ่มที่ ω_1

จะได้ว่าความน่าจะเป็นของความคลาดเคลื่อน

$$\begin{aligned} P(error) &= P(x \in \Re_2, \omega_1) + P(x \in \Re_1, \omega_2) \\ &= P(x \in \Re_2 \mid \omega_1)P(\omega_1) + P(x \in \Re_1 \mid \omega_2)P(\omega_2) \\ &= \int_{\Re_2} p(x \mid \omega_1)P(\omega_1)dx + \int_{\Re_1} p(x \mid \omega_2)P(\omega_2)dx \end{aligned}$$

ถ้า $p(x \mid \omega_1)P(\omega_1) > p(x \mid \omega_2)P(\omega_2)$ จะกำหนดให้ x อยู่ในพื้นที่ของ $\Re_1$

ค่าความน่าจะเป็นของความคลาดเคลื่อนจะมีค่าต่ำที่สุด โดยกำหนดให้ในแต่ละจุดที่ถูกจัดอยู่ในกลุ่มนั้นมีค่าความน่าจะเป็นหลังสูงที่สุด[1] ซึ่งเหตุผลที่ใช้ค่าความน่าจะเป็นหลังที่มีค่าสูงที่สุด เนื่องจากามารถประยุกต์กับเหตุการณ์เมื่อมี d-dimensional สำหรับเวกเตอร์ที่ต้องการศึกษา และมีจำนวนกลุ่ม C กลุ่ม

ในการวิเคราะห์จะพิจารณาถึงฟังก์ชันความหนาแน่น (probability density function) และการแจกแจงของความน่าจะเป็น แต่เกณฑ์การตัดสินใจในการจัดกลุ่มจะใช้ความน่าจะเป็นหลังที่มากที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

[1] แหล่งที่มา : http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn10-lec-2x3.pdf [17/08/2552]

การกำหนดกลุ่มในรูปแบบของฟังก์ชันจำแนกประเภท ในที่นี้จะเขียนอยู่ในรูป $y_1(\underset{\sim}{x}), y_2(\underset{\sim}{x})..., y_C(\underset{\sim}{x})$ ซึ่งกำหนดให้ค่าสังเกตของ x อยู่ในกลุ่มที่ ω_C ถ้า

$$y_C(\underset{\sim}{x}) > y_i(\underset{\sim}{x}) \qquad \forall i \neq c$$

ดังนั้น ค่าความคลาดเคลื่อนของความน่าจะเป็นที่ต่ำที่สุด จะอยู่ในรูป

$$y_C(\underset{\sim}{x}) \quad = \quad P(\omega_c \mid \underset{\sim}{x}) \propto p(\underset{\sim}{x} \mid \omega_C) P(\omega_C)$$

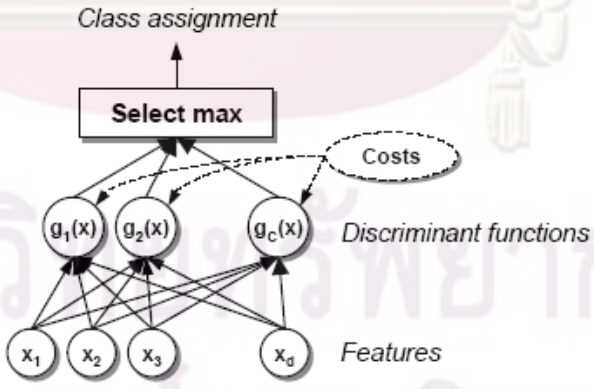
แต่ถ้าความสัมพันธ์ของข้อมูลมีขนาดใหญ่ เราสามารถนำค่าลอการิทึมลงในสมการ ฟังก์ชันจำแนกประเภทและการจัดกลุ่ม จะอยู่ในรูป

$$y_C(\underset{\sim}{x}) \quad = \quad \ln(p(\underset{\sim}{x} \mid \omega_C)) + \ln(P(\omega_C))$$

กำหนดให้เวกเตอร์ตัวแปรอิสระ $\underset{\sim}{x}$ จะอยู่ในกลุ่มที่ ω_C เมื่อค่าความน่าจะเป็นของความคลาดเคลื่อนจะมีค่าต่ำที่สุด โดยกำหนดให้ในแต่ละจุดที่ถูกจัดอยู่ในกลุ่มนั้นมีค่าความน่าจะเป็นหลังสูงที่สุด

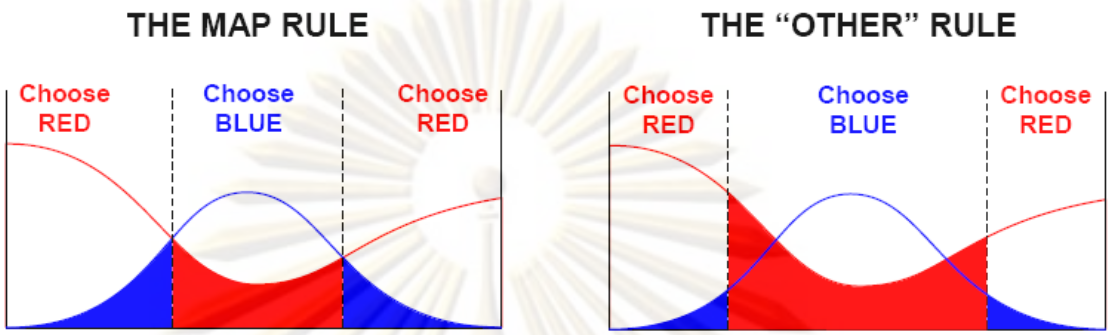
จากการจัดกลุ่มโดยใช้วิธี Maximum a posterior (MAP) สามารถเขียนเป็นแผนภาพการทำงานดังนี้

รูปที่ 2.2 แผนผังการการจัดกลุ่ม [1]



[1] แหล่งที่มา : http://research.cs.tamu.edu/prism/lectures/iss/iss_l12.pdf [17/082552]

รูปที่ 2.3 ภาพเปรียบเทียบการจำแนกกลุ่มระหว่างการใช้วิธีการ MAP กับวิธีการแบบอื่น[1]



2.2.2.2 การประมาณค่าฟังก์ชันจำแนกประเภท

การประมาณค่าฟังก์ชันจำแนกประเภทจะใช้ตัวแบบ Gaussian density เขียนอยู่ในรูป

$$P(\underset{\sim}{x}\,|\,\omega_C) = \frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_C\right|^{1/2}}\exp\left(-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim C}{\mu})^T\boldsymbol{\Sigma}_C^{-1}(\underset{\sim}{x}-\underset{\sim C}{\mu})\right)$$

จะได้ว่า ฟังก์ชันการจัดกลุ่ม คือ

$$y_C(\underset{\sim}{x}) = -\frac{1}{2}(\underset{\sim}{x}-\underset{\sim C}{\mu})^T\boldsymbol{\Sigma}_C^{-1}(\underset{\sim}{x}-\underset{\sim C}{\mu}) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_C\right| + \ln P(\omega_C)$$

พิจารณาในกรณีที่เมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่มมีค่าเท่ากัน $\boldsymbol{\Sigma}_C = \boldsymbol{\Sigma}$ จะพบว่า

$$y_C(\underset{\sim}{x}) = -\frac{1}{2}(\underset{\sim}{x}-\underset{\sim C}{\mu})^T\boldsymbol{\Sigma}^{-1}(\underset{\sim}{x}-\underset{\sim C}{\mu}) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}\right| + \ln P(\omega_C)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

[1] แหล่งที่มา : http://research.cs.tamu.edu/prism/lectures/iss/iss_l12.pdf [17/082552]

ซึ่งจะได้ว่า ค่าของ $-\frac{1}{2}\ln|\boldsymbol{\Sigma}|$ สามารถตัดค่านี้ออกได้ ซึ่งจะได้ว่า

$$y_C(\underset{\sim}{x}) = -\frac{1}{2}(\underset{\sim}{x} - \underset{\sim C}{\mu})^T \boldsymbol{\Sigma}^{-1}(\underset{\sim}{x} - \underset{\sim C}{\mu}) + \ln P(\omega_C)$$

สามารถขยายรูปแบบให้อยู่ในรูป Quadratic ได้ ดังนี้

$$y_C(\underset{\sim}{x}) = -\frac{1}{2}(\underset{\sim}{x}^T \boldsymbol{\Sigma}^{-1} \underset{\sim}{x} - \underset{\sim}{x}^T \boldsymbol{\Sigma}^{-1} \underset{\sim C}{\mu} - \underset{\sim C}{\mu}^T \boldsymbol{\Sigma}^{-1} \underset{\sim}{x} + \underset{\sim C}{\mu}^T \boldsymbol{\Sigma}^{-1} \underset{\sim C}{\mu} + \ln P(\omega_C)$$

หมายเหตุ 1) $\underset{\sim}{x}^T \boldsymbol{\Sigma}^{-1} \underset{\sim}{x}$ คือ กลุ่มที่เป็นอิสระกัน สามารถตัดค่านี้ออกได้ สำหรับการจำแนก

2) $\boldsymbol{\Sigma}$ เป็นเมทริกซ์สมมาตร จะทำให้ $\boldsymbol{\Sigma}^{-1}$ เป็นเมทริกซ์สมมาตรด้วย ซึ่งจะได้ว่า

$$\underset{\sim}{x}^T \boldsymbol{\Sigma}^{-1} \underset{\sim \mathbf{C}}{\mu} = \underset{\sim \mathbf{C}}{\mu} \boldsymbol{\Sigma}^{-1} \underset{\sim}{x}^T$$

ฟังก์ชันการจัดกลุ่มในการจัดกลุ่ม จะอยู่ในรูป

$$y_C(\underset{\sim}{x}) = (\underset{\sim C}{\mu}^T \boldsymbol{\Sigma}^{-1})\underset{\sim}{x} - \left(\frac{1}{2}\underset{\sim \mathbf{C}}{\mu}^{\mathbf{T}} \boldsymbol{\Sigma} \underset{\sim C}{\mu} + \ln P(\omega_C)\right)$$

$$= \underset{\sim \mathbf{C}}{w}^{\mathbf{T}} \underset{\sim}{x} + w_{C0}$$

$$\underset{\sim \mathbf{C}}{w}^{\mathbf{T}} = \underset{\sim C}{\mu}^T \boldsymbol{\Sigma}^{-1} \qquad w_{C0} = -\frac{1}{2}\underset{\sim \mathbf{C}}{\mu}^{\mathbf{T}} \boldsymbol{\Sigma} \underset{\sim C}{\mu} + \ln P(\omega_C)$$

$$y_C(\mathbf{x}) = \underset{\sim \mathbf{C}}{w}^{\mathbf{T}} \underset{\sim}{x} + w_{C0}$$

ฟังก์ชันการจัดกลุ่มของกลุ่มที่ c ค่าพารามิเตอร์จะอยู่ในรูปของเวกเตอร์ถ่วงน้ำหนัก คือ $\underset{\sim C}{w}$ และค่าความเอนเอียง คือ w_{C0}

จะพบว่า พารามิเตอร์ของ $\underset{\sim C}{w}$ และ w_{C0} จะเป็นค่าเฉลี่ยและเมทริกซ์ความแปรปรวนร่วมของตัวแปรตาม

ซึ่ง $\underset{\sim \mathbf{C}}{w}^{\mathbf{T}} \underset{\sim}{x}$ คือ ฟังก์ชันจำแนกประเภท $w_1x_1 + w_2x_2 + \ldots + w_px_p$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เมื่อฟังก์ชันจำแนกประเภทสำหรับ Gaussians เมื่อเมทริกซ์ความแปรปรวนร่วมมีค่าเท่ากันทุกกลุ่ม คือ ฟังก์ชันจำแนกประเภทเชิงเส้นตรง ซึ่งเป็นฟังก์ชันเชิงเส้นตรงของ **x** และถ้าขอบเขตในการตัดสินใจมีลักษณะ $y_k(\underset{\sim}{x}) = y_l(\underset{\sim}{x})$ จะเรียกว่า Hyperplanes

2.3 การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค (Multinomial Logistic Regression)

2.3.1 ตัวแบบที่ศึกษา มีลักษณะดังนี้

$$logit\left[P(Y_{itk} \le k)\right] = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time)_{it} \quad , k = 1,...,K-1 \;; t = 1,...,T \;; i = 1,...,n$$

โดยที่ $\left[P(Y_{itk} \le k)\right]$ คือ ความน่าจะเป็นที่ค่า Y_{itk} จะน้อยกว่าหรือเท่ากับกลุ่มที่ k

β_{0k} คือ ค่า cut-point ของกลุ่มที่ k

β_1 , β_2 คือ สัมประสิทธิ์การถดถอย

K คือ จำนวนกลุ่มของตัวแปรตาม

x_{it} คือ ตัวแปรกำหนด สำหรับค่าสังเกตที่ i ระยะเวลาที่ t

$time_{it}$ คือ ตัวแปรเวลาสำหรับค่าสังเกตที่ i ระยะเวลาที่ t

n คือ ขนาดตัวอย่าง (Sample Size)

T คือ ระยะเวลาที่เก็บข้อมูลซ้ำ (Preiod)

2.3.2 คุณสมบัติ

สมการประมาณค่าโดยนัยทั่วไป [1] (Generalized Estimating Equation: GEE) ผู้วิเริ่มวิธีการนี้ คือ Liang และ Zeger ในปี ค.ศ. 1986 ซึ่งเทคนิคการวิเคราะห์นี้เป็นส่วนขยายจาก Generalized Linear Model (GLM) สำหรับการวิเคราะห์ข้อมูลระยะยาว โดยใช้การประมาณค่าแบบ Quasi – likelihood วิธีการประมาณค่านี้เป็นวิธีการแบบกึ่งพารามิเตอร์ (semi - parametric)

[1] แหล่งที่มา : http://www.dmbn.net/ [20/06/2552]

โดยมีกำเนิดมาจากรูปแบบที่ปราศจากรูปแบบที่มีลักษณะจำเพาะเจาะจงของการแจกแจงร่วม (joint distribution) ของค่าสังเกตในแต่ละหน่วยตัวอย่าง โดยการแทนที่นี้จะใช้ในลักษณะการแจกแจงเดี่ยว (marginal distribution) สำหรับค่าสังเกตในหน่วยตัวอย่างที่ i ในช่วงเวลาที่ t เมทริกซ์สหสัมพันธ์ที่ถูกนำมาใช้จะเป็นลักษณะของเวกเตอร์ที่มีการวัดซ้ำของค่าสังเกตในแต่ละหน่วยตัวอย่าง สมการประมาณค่าโดยนัยทั่วไป จะมีความคงเส้นคงวาและมีการผสมผสานกับลักษณะที่ไม่มีที่สิ้นสุดของรูปแบบโครงสร้างความสัมพันธ์ สำหรับโครงสร้างของความแปรปรวนร่วมจะกำหนดให้เป็นตัวแปรรบกวนและหน่วยตัวอย่างในแต่ละค่าเป็นอิสระกัน

2.3.2.1 ข้อกำหนดของ Marginal model สำหรับข้อมูลระยะยาว

1) Marginal expectation ของตัวแปรตาม $E(z_{itk}) = \mu_{itk}$ โดยใช้ข้อมูลจาก $\mathbf{X}_{ij}$ ผ่าน link function $\mathbf{g(\mu_{itk})}$

2) Marginal variance จะขึ้นอยู่กับ Marginal mean ดังสมการ

$$V(y_{itk}) = V(\mu_{itk})\varphi$$

โดยที่ V เป็นฟังก์ชันของความแปรปรวนที่ทราบค่า ในที่นี้กำหนดให้ $V(\mu_{itk}) = \mu_{itk}(1-\mu_{itk})$ สำหรับตัวแปรตามที่มีข้อมูลลักษณะเป็นเชิงกลุ่ม

ϕ คือ ค่า scale parameter

3) ความสัมพันธ์ระหว่าง z_{ijk} และ $z_{ijk'}$ เป็นฟังก์ชันของ marginal mean และพารามิเตอร์ α

2.3.2.2 Working correlation matrix

การประมาณค่าพารามิเตอร์ โดยใช้สมการประมาณค่าโดยนัยทั่วไป มีลักษณะพิเศษคือ จะมีรูปแบบความสัมพันธ์ของความคลาดเคลื่อนเป็นรูปแบบต่าง ๆ ที่หลากหลาย ดังเช่น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1) Independence

2) Exchangeable หรือ Uniform

3) First auto-regressive (AR1)

สำหรับงานวิจัยครั้งนี้ ผู้วิจัยจะทำการวิเคราะห์โดยเลือกใช้รูปแบบความสัมพันธ์เพียง 2 รูปแบบ คือ รูปแบบ Independence และ Uniform สำหรับการวิเคราะห์การถดถอยมัลติโนเมียล โลจิสติค เมื่อมีรูปแบบความสัมพันธ์ของความคลาดเคลื่อนเป็นรูปแบบต่าง ๆ สามารถคำนวณหารูปแบบ Working Correlation Matrix ได้ดังนี้

1) หาค่าพารามิเตอร์ α การคำนวณจะรวมอยู่ในสมการประมาณค่าสำหรับรูปแบบของความสัมพันธ์ เมื่อค่าที่ต่ำที่สุดของ $f(\alpha) = \log\left|\mathbf{V}_{\boldsymbol{\beta}}(\alpha)\right|$ ซึ่งเป็นค่าลอการิทึมของการประมาณค่าความแปรปรวนของ $\boldsymbol{\beta}$ โดยจะสนใจค่า α ในแต่ละขั้นตอนของการหาค่าประมาณ

2) Working Correlation Matrix อยู่ในรูป $\mathbf{R}(\boldsymbol{\alpha}) = \mathbf{C}(\boldsymbol{\alpha}) \otimes \mathbf{S}$

เมื่อ $\mathbf{C}(\boldsymbol{\alpha})$ คือ เมทริกซ์ขนาด (TxT) ซึ่งเป็นความสัมพันธ์ระหว่างเวลาในแต่ละช่วง

$\mathbf{S}$ คือ เมทริกซ์ที่มีขนาด (K-1)x(K-1) เป็นความสัมพันธ์ระหว่างค่าของตัวแปรตอบสนอง 2 ค่าของ z_i ในแต่ละระยะเวลา

ตัวอย่างการคำนวณค่า Working Correlation Matrix

1. เมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระกัน

$$C(\alpha) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}_{TxT}$$

$$\mathbf{S} = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1(K-1)} \\ \vdots & \ddots & \vdots \\ \rho_{(K-1)1} & \cdots & \rho_{(K-1)(K-1)} \end{bmatrix}_{(K-1)x(K-1)}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. เมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม

$$C(\alpha)=\begin{bmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \ddots & \vdots \\ \vdots & \ddots & 1 & \alpha \\ \alpha & \dots & \alpha & 1 \end{bmatrix}_{TxT}$$

$$\mathbf{S}=\begin{bmatrix} \rho_{11} & \cdots & \rho_{1(K-1)} \\ \vdots & \ddots & \vdots \\ \rho_{(K-1)1} & \cdots & \rho_{(K-1)(K-1)} \end{bmatrix}_{(K-1)x(K-1)}$$

ตัวอย่างการคำนวณค่า Working Correlation Matrix จากผลการวิเคราะห์กรณีศึกษาที่ 3 การรักษาอาการเอ็นร้อยหวายฉีก

1. เมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระกัน

สมมติให้ เวลาที่ใช้ในการวิเคราะห์ 3 ช่วงเวลา โดยมีจำนวนกลุ่มของตัวแปรตาม 3 กลุ่ม และกำหนดให้ $\mathbf{S}=\begin{bmatrix} 1 & 0.270 \\ 0.270 & 1 \end{bmatrix}$

Working Correlation Matrix อยู่ในรูป $\mathbf{R(\alpha)=C(\alpha)\otimes S}$ จะได้ว่า

Working Correlation Matrix คือ

$$\begin{bmatrix} 1 & 0.207 & 0 & 0 & 0 & 0 \\ 0.207 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.207 & 0 & 0 \\ 0 & 0 & 0.207 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.207 \\ 0 & 0 & 0 & 0 & 0.207 & 1 \end{bmatrix}_{3(3-1)x3(3-1)}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. เมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม

สมมติให้ เวลาที่ใช้ในการวิเคราะห์ 3 ช่วงเวลา โดยมีจำนวนกลุ่มของตัวแปรตาม 3 กลุ่ม และกำหนดให้ $\mathbf{S} = \begin{bmatrix} 1 & 0.213 \\ 0.213 & 1 \end{bmatrix}_{(3-1)x(3-1)}$ ค่า α ที่คำนวณได้ = 0.05 จะได้ว่า

$$\mathbf{C(\alpha)} = \begin{bmatrix} 1 & 0.05 \\ 0.05 & 1 \end{bmatrix}_{3x3}$$

Working Correlation Matrix อยู่ในรูป $\mathbf{R(\alpha)} = \mathbf{C(\alpha)} \otimes \mathbf{S}$ จะได้ว่า

Working Correlation Matrix คือ

$$\begin{bmatrix} 1 & 0.213 & 0.05 & 0.011 & 0.05 & 0.011 \\ 0.213 & 1 & 0.011 & 0.05 & 0.011 & 0.05 \\ 0.05 & 0.011 & 1 & 0.213 & 0.05 & 0.011 \\ 0.011 & 0.05 & 0.213 & 1 & 0.011 & 0.05 \\ 0.05 & 0.011 & 0.05 & 0.011 & 1 & 0.213 \\ 0.011 & 0.05 & 0.011 & 0.05 & 0.213 & 1 \end{bmatrix}_{3(3-1)x3(3-1)}$$

ข้อแนะนำในการใช้ working correlation matrix R

- ควรเลือกเมทริกซ์สหสัมพันธ์ที่มีความคงเส้นคงวา โดยขึ้นอยู่กับรูปแบบความสัมพันธ์ของข้อมูลที่แท้จริงนั้น
- วิธีการประมาณค่าพารามิเตอร์โดยใช้สมการประมาณค่าโดยนัยทั่วไป จะให้ผลลัพธ์ในการประมาณค่าสัมประสิทธิ์การถดถอยและค่าความแปรปรวน โดยใช้โครงสร้างความสัมพันธ์ของเมทริกซ์ความแปรปรวนและความแปรปรวนร่วม
- การสูญเสียประสิทธิภาพในการประมาณค่าพารามิเตอร์ เมื่อมีการเลือกเมทริกซ์สหสัมพันธ์ผิดจะลดน้อยลงเมื่อขนาดตัวอย่างมีจำนวนเพิ่มมากขึ้น

2.3.3 การประมาณค่าพารามิเตอร์

2.3.3.1 Proportional odds model

จากข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลทางการแพทย์ซึ่งมีตัวแปรตามเป็นแบบอันดับในแต่ละช่วงเวลา T จากหน่วยตัวอย่าง N หน่วย ซึ่งมีจำนวนกลุ่ม K กลุ่ม กำหนดให้ Y_{it} คือ ค่าของตำแหน่งของหน่วยตัวอย่างที่ i ในช่วงเวลาที่ t และกำหนดให้ $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{iT})^T$ เป็นเวกเตอร์ที่เป็นข้อมูลเป็นเชิงอันดับ สำหรับหน่วยตัวอย่างที่ i ทั้งหมด T ช่วงเวลา กำหนดให้ $\mathbf{X}_{it}$ เป็นเมทริกซ์ตัวแปรอิสระของค่าสังเกต ดังนั้นความน่าจะเป็นเมื่อค่า Y_{it} ตกอยู่ในกลุ่มใดนั้นเกิดจากการวัดตัวแปรอิสระ โดยใช้ Proportional odds model ซึ่งอยู่ในรูป cumulative logit เป็นพื้นฐาน

$$h(\mu_{itk}) = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time) \quad ; i = 1,...,n \quad ; t = 1,...,T$$

เมื่อ $\mu_{itk} = \mathrm{P}(\mathrm{Y}_{itk} \le k)$ คือ ค่าความน่าจะเป็นที่ค่า Y_{itk} น้อยกว่าหรือเท่ากับกลุ่มที่ k

β_{0k} คือ ค่า cut-point ซึ่งมีค่าอยู่ระหว่าง $-\alpha < \beta_{01} < ... < \beta_{0(K-1)} < \alpha$ (k=1,2,..,K-1)

β_1 , β_2 คือ สัมประสิทธิ์การถดถอย

K คือ จำนวนกลุ่มของตัวแปรตาม

x_{it} คือ ตัวแปรกำหนด สำหรับค่าสังเกตที่ i ระยะเวลาที่ t

$time_{it}$ คือ ตัวแปรเวลาสำหรับค่าสังเกตที่ i ระยะเวลาที่ t

n คือ ขนาดตัวอย่าง (Sample Size)

T คือ ระยะเวลาที่เก็บข้อมูลซ้ำ (Preiod)

จากสมการ $h(\mu_{itk}) = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time)$ พบว่า มีประสิทธิภาพในการแปลงข้อมูลจากมาตรวัดแบบอันดับไปเป็นมาตรวัดแบบต่อเนื่อง [1] โดยใช้ลักษณะการพยากรณ์เชิงเส้นตรง $\beta_1 x_{it} + \beta_2 (time)$ ซึ่งพิสูจน์โดยตัวแบบธรรมดาสำหรับความสัมพันธ์ที่มีมาตรวัดแบบอันดับไปสู่รูปแบบที่แตกต่างกันออกไป สำหรับค่า cut-point ใช้อธิบายถึงขอบเขตที่มีลักษณะเป็นแบบต่อเนื่อง

---

[1] Agresti Alan, Categorical Data Analysis, (New York: Wiley, 2002), p.467

สำหรับการอธิบายโดยการพยากรณ์เชิงเส้นตรง ตัวแบบนี้เป็นที่รู้จักกันว่า Proportional odds model เนื่องจาก proportional odds ในเหตุการณ์ที่ $Y_{itk} \le k$ สำหรับในแต่ละคู่ของตัวแปร อธิบายที่เป็นอิสระกันของแต่ละทางเลือกที่ k การตรวจสอบข้อกำหนดของอัตราส่วนโอกาส สามารถใช้สถิติ Score test ในการทดสอบความเหมาะสมของค่า cut-point ได้

2.3.3.2 การประมาณค่าพารามิเตอร์

สำหรับวิธีการใช้สมการประมาณค่าโดยนัยทั่วไป โครงสร้างความสัมพันธ์จะถูกจัดประเภทโดยใช้ค่า Working correlation matrix R ซึ่งมักเป็นค่าที่กำหนดขึ้นมาก่อนตามความเหมาะสมของข้อมูล แต่โดยปกติมักทำการคำนวณและปรับปรุงในแต่ละขั้นตอนที่มีการทำซ้ำ ในขั้นตอนการหาค่าที่เหมาะสม โดยส่วนมากในการวิเคราะห์ข้อมูลทางการแพทย์มักจะมีปัญหาในเรื่องของการอนุมานปัจจัยของทรีทเมนต์เริ่มต้น และต้องสามารถพิสูจน์ให้ได้ว่า ในขั้นตอนการหาสมการในการประมาณค่า $Q_\beta(\beta;\alpha) = 0$ สำหรับ β เมื่อ α คือ พารามิเตอร์แฝง และในการประมาณค่าพารามิเตอร์ α บ่อยครั้งที่จะขึ้นอยู่กับความคลาดเคลื่อนของข้อมูล แต่อย่างไรก็ตาม ปัญหาที่เกิดขึ้น คือ เป็น non-convergence และการประมาณค่า α ซึ่งต่อมามีผู้เสนอทางเลือกในการประมาณค่า ซึ่งได้เสนอในส่วนของ robust method ในการประมาณค่า α ที่มีค่าลอการิทึมที่ต่ำที่สุด ในส่วนดีเทอร์มิแนนท์ของเมทริกซ์ความแปรปรวนร่วมในแต่ละขั้นตอนของการหาตัวแบบที่เหมาะสม

เมื่อพิจารณาถึงรูปแบบของ cumulative ของตัวแปรตามที่มี 2 ค่า เป็นสิ่งจำเป็นในการหาค่าวัดซ้ำที่เหมาะสมโดยใช้พื้นฐานของตัวแบบของสัดส่วนโอกาส สำหรับตัวแปรตามในแต่ละหน่วยตัวอย่าง สำหรับในแต่ละจุดช่วงเวลา ได้ทำการแปลงข้อมูลมาจาก K-1 กลุ่ม โดยกำหนดให้ $z_{itk} = 1$ ถ้า $Y_{it} \le k$ หรือ $z_{itk} = 0$ เมื่อ $Y_{it} > k$ สำหรับ $k = 1,2,...,K-1$ จะพบว่าอัตราส่วนของโอกาส อยู่ในรูปแบบของ

$$E\left(\underset{\sim}{z}_i\right) > h^{-1}(\mathbf{X}_i \underset{\sim}{\beta})$$

เมื่อ $\mathbf{X}_i$ คือ เมทริกซ์เต็มรูปของตัวแปรอิสระ

$\underset{\sim}{z}_i$ คือ ค่าของตัวแปรตาม โดยการจัดกลุ่ม Z_{itk} สำหรับหน่วยตัวอย่างที่ i

$\underset{\sim}{\beta}$ คือ เวกเตอร์ของค่า cut point และสัมประสิทธิ์การถดถอย

สำหรับข้อมูลในเมทริกซ์ $\mathbf{X}_i$ และค่าของตัวแปรตาม $\mathbf{Z}_i$ ในแต่ละหน่วยตัวอย่าง การหาค่า $\underset{\sim}{\beta}$ จะทำการประมาณค่าโดยใช้กระบวนการกำลังสองน้อยสุดที่ปรับค่าถ่วงน้ำหนักในแต่ละรอบของการประมาณค่า (re-weight least square) โดยใช้สมการประมาณค่าโดยนัยทั่วไป $Q_{\beta}(\underset{\sim}{\beta};\boldsymbol{\alpha})$ มีค่าเท่ากับ 0 เมื่อ

$$Q_{\beta}(\underset{\sim}{\beta};\alpha) = \sum_{i=1}^{N} \mathbf{D_i^T W(\alpha_i)^{-1}} (\underset{\sim}{z}_{\mathbf{i}} - \underset{\sim}{\mu}_i) \tag{21}$$

เมื่อ $\mathbf{D_i} = \boldsymbol{\mu}'(\underset{\sim}{\beta})_{\mathbf{i}}$ คือค่าอนุพันธ์อันดับแรก ของ $\boldsymbol{\mu}(\underset{\sim}{\beta})_{\mathbf{i}}$ ซึ่งขึ้นอยู่กับ $\boldsymbol{\beta}$

$$\mathbf{W(\alpha)}_i = \mathbf{V_i^{1/2} R(\alpha)_i V_i^{1/2}}$$

เมื่อ $\mathbf{R(\alpha)_i}$ คือฟังก์ชันที่ไม่ทราบความสัมพันธ์ของพารามิเตอร์ $\boldsymbol{\alpha}$ ซึ่งเป็นเมทริกซ์ของความสัมพันธ์ระหว่างองค์ประกอบของ $\underset{\sim}{z}_{\mathbf{i}}$

$\mathbf{V_i^{1/2}}$ คือ รากที่สองของเมทริกซ์ความแปรปรวนระหว่างค่า $\underset{\sim}{z}_{\mathbf{i}}$ ของตัวแปรที่ทำการเปรียบเทียบ โดยกำหนดให้มีค่าตามแนวเส้นทแยงมุมเท่ากับ $[\mu_{itk}(1-\mu_{itk})]^{1/2}$

จากการประมาณค่าประมาณค่าโดยนัยทั่วไป จะได้พารามิเตอร์การถดถอย $\underset{\sim}{\beta}$ ซึ่งมีลักษณะเข้าใกล้การแจกแจงแบบปกติหลายตัวแปรโดยเมทริกซ์ความแปรปรวนร่วมจะถูกเรียกว่า robust หรือ sandwich estimator

$$\mathbf{V_\beta(\alpha)} = \left[\sum_{i=1}^{N} \mathbf{D_i^T W(\alpha_i)^{-1} D_i}\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{D_i^T W(\alpha_i)^{-1} cov(z_i) W(\alpha_i)^{-1} D_i}\right] \left[\sum_{i=1}^{N} \mathbf{D_i^T W(\alpha_i)^{-1} D_i}\right]^{-1}$$

จากสมการข้างต้น มีคุณสมบัติเป็นตัวประมาณที่มีความคงเส้นคงวาของความแปรปรวนในพารามิเตอร์การถดถอย โดยไม่คำนึงถึง $\mathbf{W(\alpha_i)}$ สำหรับสมการข้างต้น ถ้ากำหนดให้การประมาณค่ามีความคงเส้นคงวาสำหรับพารามิเตอร์ในตัวแบบ เมื่อโครงสร้างความแปรปรวนร่วมของ $\underset{\sim}{z}_{\mathbf{i}}$ ไม่มีการปรับเปลี่ยนโครงสร้างใด ๆ อย่างไรก็ตามประสิทธิภาพในการประมาณค่าสำหรับเวกเตอร์ $\underset{\sim}{\beta}$ และความน่าเชื่อถือในการอนุมาน สามารถปรับปรุงได้ ถ้า working correlation matrix $\mathbf{R(\alpha_i)}$ ถูกเลือกขึ้นมาได้อย่างเหมาะสมกับค่าเมทริกซ์ความสัมพันธ์ของ $\underset{\sim}{z}_{\mathbf{i}}$ ที่เป็นไปได้

สำหรับวิธีการหาค่าพารามิเตอร์ α การคำนวณจะรวมอยู่ในสมการประมาณค่าสำหรับรูปแบบของความสัมพันธ์ เมื่อค่าที่ต่ำที่สุดของ $f(\alpha) = \log\left|\mathbf{V}_{\beta}(\alpha)\right|$ ซึ่งเป็นค่าลอการิทึมของการประมาณค่าความแปรปรวนของ $\underset{\sim}{\beta}$ โดยจะสนใจค่า α ในแต่ละขั้นตอนของการหาค่าประมาณ

Working Correlation Matrix อยู่ในรูป $\mathbf{R}(\boldsymbol{\alpha}) = \mathbf{C}(\boldsymbol{\alpha}) \otimes \mathbf{S}$

เมื่อ $\mathbf{C}(\boldsymbol{\alpha})$ คือ เมทริกซ์ขนาด (TxT) ซึ่งเป็นความสัมพันธ์ระหว่างเวลาในแต่ละช่วง

$\mathbf{S}$ คือ เมทริกซ์ที่มีขนาด (K-1)x(K-1) เป็นความสัมพันธ์ระหว่างค่าของตัวแปรตอบสนอง 2 ค่าของ $\mathbf{Z}_{\mathbf{i}}$ ในแต่ละระยะเวลา

ในงานวิจัยชิ้นนี้ผู้วิจัยกำหนดให้ โครงสร้างความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระ ดังนั้นจะได้เมทริกซ์ $\mathbf{C}(\boldsymbol{\alpha})$ และ $\mathbf{S}$ มี 2 ลักษณะ ดังนี้

1. เมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระกัน

$$C(\alpha) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}_{TxT}$$

$$\mathbf{S} = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1(K-1)} \\ \vdots & \ddots & \vdots \\ \rho_{(K-1)1} & \cdots & \rho_{(K-1)(K-1)} \end{bmatrix}_{(K-1)x(K-1)}$$

2. เมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม

$$C(\alpha) = \begin{bmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \ddots & \vdots \\ \vdots & \ddots & 1 & \alpha \\ \alpha & \dots & \alpha & 1 \end{bmatrix}_{TxT}$$

$$\mathbf{S} = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1(K-1)} \\ \vdots & \ddots & \vdots \\ \rho_{(K-1)1} & \cdots & \rho_{(K-1)(K-1)} \end{bmatrix}_{(K-1)x(K-1)}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การแก้ไขปัญหาที่เกิดขึ้นจากการประมาณค่า

การทำซ้ำระหว่างวิธีการ quasi-likelihood สำหรับเวกเตอร์ $\underset{\sim}{\beta}$ และวิธีการ robust สำหรับการประมาณค่า α ซึ่งเป็นฟังก์ชันของเวกเตอร์ $\underset{\sim}{\beta}$

1. นำค่าประมาณของเมทริกซ์ $\mathbf{R_i(\alpha)}$ และค่าคงที่ ϕ คำนวณหาค่าประมาณของเวกเตอร์ $\underset{\sim}{\beta}$ โดยกระบวนการกำลังสองน้อยสุดที่ปรับค่าถ่วงน้ำหนักในแต่ละรอบของการประมาณค่า (iterative re-weight least square)

2. นำค่าประมาณของเวกเตอร์ $\underset{\sim}{\beta}$ ที่เกิดจากการประมาณค่าของเวกเตอร์ $\underset{\sim}{\alpha}$ และ ϕ ทำการคำนวณเพื่อหาค่าความคลาดเคลื่อนมาตรฐาน (standardized residual)

$$r_{itk} = \frac{(z_{itk} - \hat{\mu}_{itk})}{\sqrt{V(\hat{\alpha})_{tt}}}$$

และใช้ค่าคลาดเคลื่อนมาตรฐานนั้น ในการประมาณค่า $\underset{\sim}{\alpha}$ และ ϕ

จากการวิเคราะห์จะได้ว่ารากที่สองของ $V(\underset{\sim}{\hat{\beta}})$ ที่มี diagonal เป็นองค์ประกอบส่งผลให้ ส่วนเบี่ยงเบนมาตรฐานของเวกเตอร์ $\underset{\sim}{\beta}$ มี 2 รูปแบบ คือ

1. Naïve หรือ Model-based

$$V(\underset{\sim}{\hat{\beta}}) = \left[\sum_{i=1}^{N} \mathbf{D_i^T \hat{V}_i^{-1} D_i}\right]^{-1}$$

2. Robust หรือ Empirical

$$V(\underset{\sim}{\hat{\beta}}) = \mathbf{M_0^{-1} M_1 M_0^{-1}}$$

เมื่อ $\mathbf{M_0} = \sum_{i=1}^{N} \mathbf{D_i^T V_i^{-1} D_i}$

$$\mathbf{M_1} = \sum_{i=1}^{N} \mathbf{D_i^T V_i^{-1} (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)^T V_i^{-1} D_i}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ในกรณีทั่วไปค่า robust หรือ sandwich จะเป็นค่าประมาณของ $V(\hat{\beta})$ ถ้าโครงสร้างของ working correlation $\mathbf{R_i(\alpha)}$ ไม่ใช่ค่าสหสัมพันธ์ที่แท้จริงของข้อมูล

2.4 ลักษณะการทำงานของฟังก์ชันสำเร็จรูปในการวิเคราะห์ข้อมูล

2.4.1 การวิเคราะห์จำแนกประเภท

การประยุกต์ใช้โปรแกรมสำเร็จรูป R ในการวิเคราะห์ข้อมูลจำแนกประเภทครั้งนี้ ผู้วิจัยเลือกใช้ฟังก์ชันสำเร็จรูป ใช้คำสั่ง lda() [1] ซึ่งเป็นการวิเคราะห์ข้อมูลจำแนกประเภทเชิงเส้นตรงโดยมีหลักการทำงาน ดังนี้

ฟังก์ชัน lda()

ฟังก์ชัน lda จะอยู่ package ของ MASS ในโปรแกรม R ซึ่งฟังก์ชัน lda จะทำหน้าที่ตรวจสอบอย่างละเอียด ในการค้นหาว่าเมทริกซ์ความแปรปรวนร่วมภายในข้อมูลชุดนั้นเป็นเมทริกซ์ singular หรือไม่ โดยกำหนดไว้ว่าถ้าความแปรปรวนภายในกลุ่ม มีค่าน้อยกว่า tol[2*] การคำนวณจะหยุดการวิเคราะห์และบันทึกข้อมูลโดยให้ตัวแปรนั้นเป็นค่าคงที่ ซึ่งรวมถึงเมื่อการประมาณค่าที่ได้มีคุณภาพที่ต่ำกว่ามาตรฐานที่กำหนดไว้ แต่ก็มีความเป็นไปได้ว่าผลลัพธ์ที่ได้นั้นเกิดจากตัวแปรเป็นค่าคงที่ การวิเคราะห์จะพิจารณาถึงค่าความน่าจะเป็นก่อนซึ่งมีผลกระทบต่อการจัดกลุ่มซึ่งแตกต่างจากรูปแบบทางสถิติอื่นโดยทั่วไป และนอกจากนี้ ยังมีผลกระทบต่อการหมุนของการจัดกลุ่มเชิงเส้นตรงที่มีระยะในแนวเดียวกัน ดังนั้น จึงมีการถ่วงน้ำหนักระหว่างกลุ่มรวมถึงพิจารณาเมทริกซ์ความแปรปรวนร่วมด้วย ดังนั้น ในการจำแนกประเภทเชิงเส้นตรงจะให้ความสำคัญต่อความแตกต่างระหว่างกลุ่มโดยกำหนดน้ำหนักโดยความน่าจะเป็นก่อน ซึ่งเปลี่ยนแปลงไปในแต่ละลักษณะของข้อมูลถ้ากลุ่มใดกลุ่มหนึ่งหรือหลาย ๆ กลุ่มมีค่าสูญหายเกิดขึ้น

ในการวิเคราะห์ข้อมูลโดยคำสั่งนี้ จุดสำคัญอยู่ที่การแบ่งกลุ่มเชิงเส้นตรงโดยการถ่วงน้ำหนักค่าเฉลี่ย ของจุดศูนย์กลางในแต่ละกลุ่ม การจัดกลุ่มของข้อมูลในแต่ละชุด จะใช้การจัดกลุ่มโดยวิธี Maximum a posteriori probability ในการจัดกลุ่ม

---

[*] ค่า tolerance เพื่อใช้ในการตรวจสอบว่าเมทริกซ์ความแปรปรวนร่วมเป็นเมทริกซ์ singular

[1] แหล่งที่มา : http://stat.ethz.ch/R-manual/R-patched/library/MASS/html/lda.html [20/06/2552]

2.4.2 การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

การประยุกต์ใช้โปรแกรมสำเร็จรูป R ในการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติคครั้งนี้ ผู้วิจัยได้สังเกตเห็นว่าการประมาณค่าพารามิเตอร์โดยใช้วิธีสมการประมาณค่านัยทั่วไป (GEE) ในการใช้โปรแกรม R มีผู้เสนอฟังก์ชันสำเร็จรูป 2 ฟังก์ชัน ในแต่ละฟังก์ชันค่อนข้างมีความแตกต่างกัน ดังรายละเอียดต่อไปนี้

2.4.2.1 ฟังก์ชัน ordgee

ฟังก์ชัน ordgee ได้ถูกสร้างขึ้นโดย Jun Yan [1] และผู้ที่นำฟังก์ชันนี้มาใช้ในการวิเคราะห์ข้อมูลระยะยาวที่มีตัวแปรตามเป็นข้อมูลเชิงกลุ่มที่มีมากกว่า 2 กลุ่ม คือ Laura A. Thompson ได้ทำการเปรียบเทียบกับโปรแกรม SAS โดยใช้คำสั่ง GENMOD กับตัวอย่างข้อมูล ผลการวิเคราะห์พบว่าการประมาณค่าของ cutpoint อย่างคร่าว ๆ คล้ายกันกับโปรแกรม SAS และสัมประสิทธิ์ของตัวแปรเวลาก็มีความคล้ายกันกับโปรแกรม SAS แต่สิ่งที่มีความแตกต่างกันมากที่สุดคือ ตัวแปรที่มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระ 2 ตัว พบว่าทิศทางเป็นไปในทางลบ ซึ่งในโปรแกรม SAS มีทิศทางในทางบวก ซึ่งได้ให้ข้อสันนิษฐานไว้ว่าอาจจะเป็นที่การกำหนดค่าตัวแปรที่แตกต่างกัน แต่ในความจริงก็ไม่พบข้อผิดปกติใด ๆ เกี่ยวกับการกำหนดค่าของตัวแปรที่แตกต่างกัน ค่าความคลาดเคลื่อนมาตรฐานมีค่าค่อนข้างสูง[2]

การทำงานของฟังก์ชัน ordgee พบว่า สร้างขึ้นโดยใช้สัญลักษณ์เหมือนกับฟังก์ชัน glm ซึ่งฟังก์ชัน glm พัฒนามาจากฟังก์ชัน lm อีกครั้งหนึ่ง และมีการนำฟังก์ชัน gee มาประยุกต์ใช้ในการทำงาน ในส่วนของการตรวจสอบค่าสูญหาย พบว่า โปรแกรมจะตรวจสอบเมื่อมีการเรียกใช้คำสั่ง gee เท่านั้น

---

[1] แหล่งที่มา : http://www.stat.ucl.ac.be/ISdidactique/Rhelp/library/geepack/html/ordgee.html [13/01/2552]

[2] Laura A. Thompson, R (and S-plus) Manual to Accompany Agresti's Categorical Data Analysis (2008), pp. 219-221.

2.4.2.2 ฟังก์ชัน repolr

ฟังก์ชัน repolr ได้ถูกสร้างขึ้นโดย Parson และคณะ [1] ฟังก์ชันนี้ถูกพัฒนาขึ้นมาสำหรับการวิเคราะห์โดยใช้ Proportional odds model โดยฟังก์ชันนี้พัฒนาจากฟังก์ชัน GEE ซึ่งขั้นตอนในการแก้ปัญหาจะใช้การประมาณค่าพารามิเตอร์ของความสัมพันธ์โดยให้มีค่าต่ำที่สุด และมีการปรับค่าความแปรปรวนในแต่ละขั้นตอนการดำเนินงาน ซึ่งรูปแบบความสัมพันธ์ในการประมาณค่าสามารถมีได้หลายรูปแบบ และมีการตรวจสอบค่าสูญหายในขั้นต้นของการวิเคราะห์ข้อมูล มีการตรวจสอบความเหมาะสมของตัวแบบ โดยใช้ สถิติทดสอบ Score test

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

[1] แหล่งที่มา : http://www2.warwick.ac.uk/fac/med/staff/nparsons/repolr/ [13/012552]

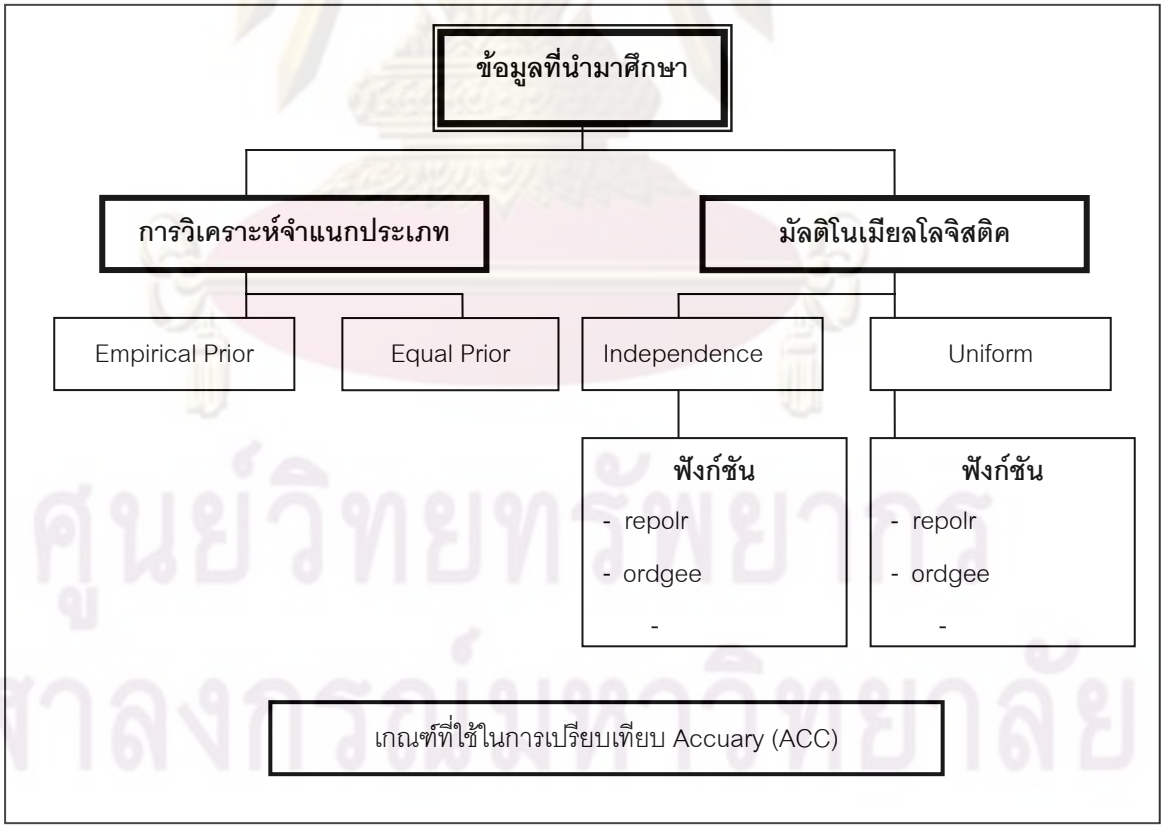
บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ เป็นการวิจัยเชิงปริมาณโดยใช้ข้อมูลจริงที่เกิดขึ้นภายใต้สถานการณ์ต่าง ๆ ที่กำหนด โดยใช้โปรแกรม R version 2.8.1 และ Microsoft Excel 2003 ซึ่งมีรายละเอียดของแผนการทดลองและขั้นตอนการวิจัย ดังต่อไปนี้

3.1 แผนการทดลอง

ในการวิจัยครั้งนี้ได้กำหนดสถานการณ์ต่าง ๆ เพื่อศึกษาเปรียบเทียบระหว่างวิธีการวิเคราะห์ข้อมูลจำแนกประเภทและวิธีการถดถอยมัลติโนเมียลโลจิสติค สำหรับข้อมูลระยะยาวทางด้านการแพทย์ โดยมีจำนวนตัวอย่างข้อมูลที่ใช้ทำการศึกษา 3 ชุด ซึ่งทำการทดลองภายใต้สถานการณ์ ดังต่อไปนี้

รูปที่ 3.1 แสดงสถานการณ์ที่จะทำการวิเคราะห์ สำหรับข้อมูล 1 ชุด



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.2 ขั้นตอนในการดำเนินงาน

ขั้นตอนในการดำเนินงานวิจัยมีดังนี้

1. ศึกษาวิธีการวิเคราะห์ข้อมูลจำแนกประเภท และวิธีการถดถอยมัลติโนเมียลโลจิสติคสำหรับข้อมูลระยะยาว

2. ศึกษาวิธีการประมาณค่าพารามิเตอร์ทั้ง 2 วิธี คือ วิธี Maximum a posteriori method และ วิธี Generalized Estimating Equation

3. ศึกษาวิธีการเขียนโปรแกรม และความแตกต่างของโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลมัลติโนเมียลโลจิสติคโดยใช้ วิธี Generalized Estimating Equation ตามสถานการณ์ที่กำหนด

4. สร้างข้อมูลตามสถานการณ์ที่นำมาใช้ในการวิเคราะห์ รายละเอียดกล่าวไว้ในภาคผนวก ก

4.1 ข้อมูลทางการแพทย์เกี่ยวกับการติดตามผลการรักษาโรคนอนไม่หลับ[1]

4.2 ข้อมูลทางการแพทย์เกี่ยวกับการผ่าตัดเปลี่ยนผิวข้อสะโพก [2]

4.3 ข้อมูลทางการแพทย์เกี่ยวกับการรักษาอาการเอ็นร้อยหวายฉีก [3]

5. ตรวจสอบข้อมูล และในกรณีที่มีตัวแปรอิสระเป็นเชิงคุณภาพ จะทำการปรับข้อมูลให้เป็นตัวแปรหุ่น (Dummy variable)

6. เขียนโปรแกรมตรวจสอบลักษณะเบื้องต้นของข้อมูล ได้แก่

6.1 ตัวแปรอิสระต้องมีการแจกแจงแบบปกติหลายตัวแปร

6.2 เมทริกซ์ค่าความแปรปรวนร่วมของตัวแปรของข้อมูล k กลุ่ม จะต้องมีค่าเท่ากัน

---

[1] แหล่งที่มา : http://www.stat.ufl.edu/~aa/cda/sas/sas.html [31/01/2552]

[2] แหล่งที่มา : http://www.warwick.ac.uk/go/repolr. [20/04/2552]

[3] แหล่งที่มา : http://www.warwick.ac.uk/go/repolr. [25/04/2552]

7. เขียนโปรแกรมสำหรับการคำนวณหาฟังก์ชันจำแนกประเภท โดยใช้วิธี Maximum a posteriori method โดยกำหนดสถานการณ์ 2 สถานการณ์ ดังนี้

7.1 กำหนดความน่าจะเป็นก่อนมีเท่ากันทุกระดับของตัวแปรตอบสนอง

7.2 กำหนดความน่าจะเป็นก่อนมีเท่ากับความน่าจะเป็นเชิงประจักษ์

8. เขียนโปรแกรมสำหรับการคำนวณหาค่าพารามิเตอร์ของตัวแบบ Proportion odds โดยใช้วิธี Generalized Estimating Equation โดยกำหนดสถานการณ์ 2 สถานการณ์

8.1 รูปแบบของ Working correlation matrix เป็นแบบอิสระ ซึ่งใช้โปรแกรมสำเร็จรูป 2 โปรแกรม ดังนี้

8.1.1 ฟังก์ชัน ordgee

8.1.2 ฟังก์ชัน repolr

8.2 รูปแบบของ Working correlation matrix เป็นแบบยูนิฟอร์ม

8.2.1 ฟังก์ชัน ordgee

8.2.2 ฟังก์ชัน repolr

9. ทำการวิเคราะห์ข้อมูลที่ได้จากการทดลองของข้อมูลแต่ละชุด และนำผลการวิเคราะห์มาเปรียบเทียบแต่ละตัวแบบภายใต้วิธีการประมาณค่า โดยใช้ค่าความถูกต้องในการจัดกลุ่ม (ACC) ซึ่งตัวแบบมีลักษณะดังต่อไปนี้

1) <u>ตัวแบบการวิเคราะห์จำแนกประเภท</u>

$$z_i = w_{i1}x + w_{i2}(time) \quad ; i = \min\{p, K-1\}$$

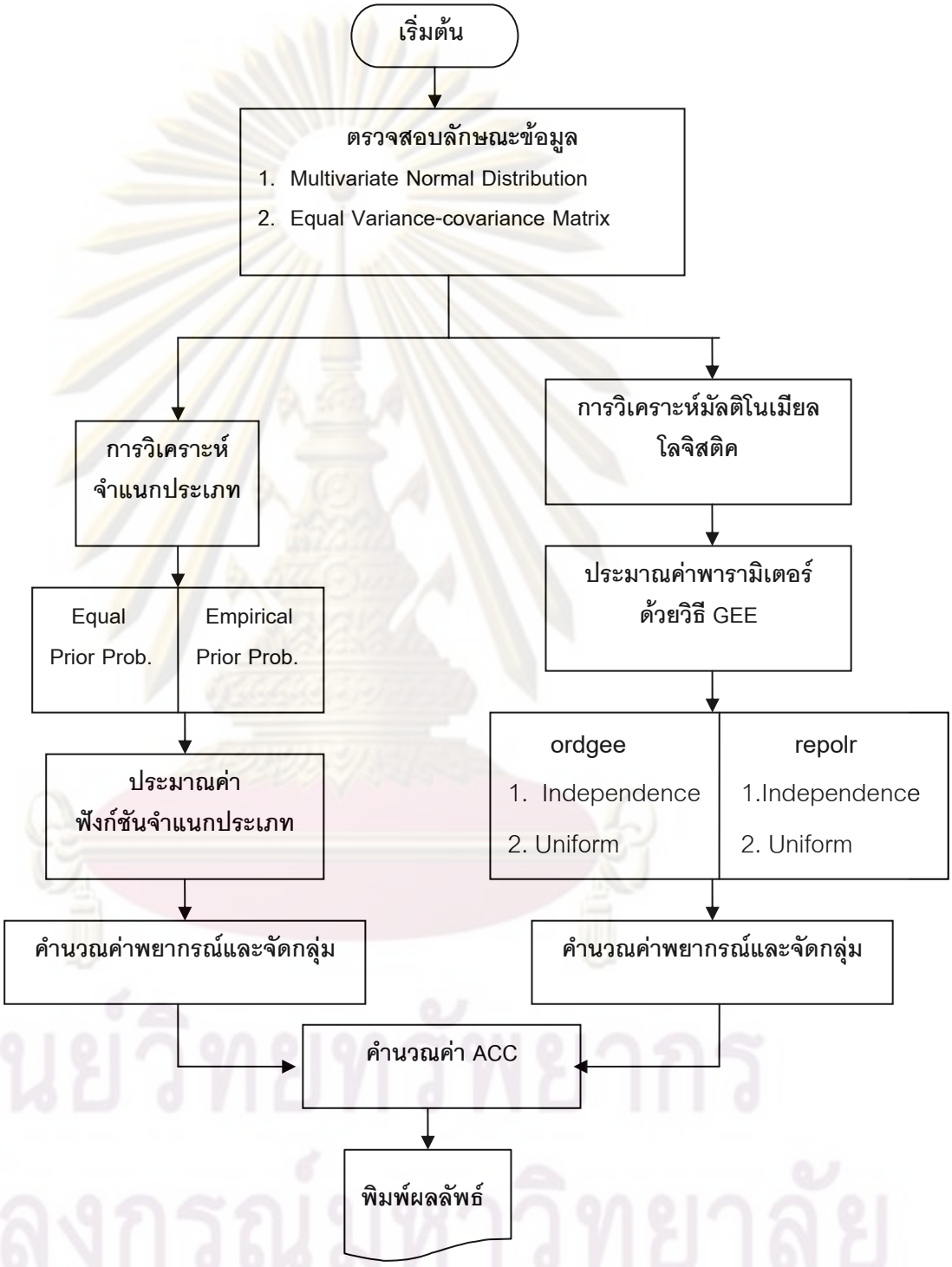
2) <u>ตัวแบบการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค</u>

$$logit\left[P(Y_{itk} \le k)\right] = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time)_{it} \quad , k = 1,...,K-1$$

10. ทำการเลือกตัวแบบ และวิธีการประมาณค่าที่เหมาะสมกับลักษณะของข้อมูลที่ทำการวิจัยและสรุปผล

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ผังแสดงขั้นตอนในการวิจัย



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.2.1 การตรวจสอบลักษณะข้อมูล

1) ตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร เนื่องจากตัวแปรเวลาเป็นตัวแปรเชิงประมาณเพียงตัวแปรเดียวและตัวแปรกำหนดในแต่ละข้อมูลเป็นตัวแปรเชิงคุณภาพ จึงทำการทดสอบข้อกำหนดของการแจกแจงแบบปกติเฉพาะตัวแปรเวลาเท่านั้น

การทดสอบสมมติฐาน

H_0 : ตัวแปรเวลามีการแจกแจงแบบปกติ

H_1 : ตัวแปรเวลาไม่มีการแจกแจงแบบปกติ

สถิติทดสอบ

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

ค่า a_i คำนวณได้จาก

$$a_i = \frac{\sum_{j=1}^{n} m_j \underset{\sim}{\upsilon}_{ij}}{C}$$

โดย $\underset{\sim}{\upsilon}_{ij} = \operatorname{cov}(x_i, x_j) \quad (i, j = 1,2,...,n)$

$E(x_i) = m_i \quad (i = 1,2,...,n)$

$C^2 = \underset{\sim}{m}^T \mathbf{V}^{-1}\mathbf{V}^{-1} \underset{\sim}{m}$

เมื่อ $\underset{\sim}{m} = (m_1,...,m_n)^T$

ซึ่ง $m_1,...,m_n$ คือ ค่าคาดหวังของค่าสถิติที่เรียงลำดับที่มีการแจกแจงที่เหมือนกันและเป็นอิสระกันของตัวแปรสุ่ม โดยสุ่มจากการแจกแจงแบบปกติมาตรฐาน

$\mathbf{V}$ คือ เมทริกซ์ความแปรปรวนร่วม

จะปฏิเสธสมมติฐาน เมื่อ ค่า W ที่ได้มีค่าต่ำ

2) <u>เมทริกซ์ค่าความแปรปรวนร่วมตัวตัวแปรอิสระ k กลุ่ม มีค่าเท่ากัน</u>

$$H_0 \quad : \quad \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$$

$$H_1 \quad : \quad \Sigma_i \neq \Sigma_j \quad \text{อย่างน้อย 1 คู่ ; } i \neq j$$

กำหนดให้

$$C_1 = \left[\sum_{i=1}^{k} \frac{1}{v_i^2} - \frac{1}{\sum_{i=1}^{k} v_i^2} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right]$$

$$\ln M = \frac{1}{2} \sum_{i=1}^{k} v_i \ln |S_i| - \frac{1}{2} \sum_{i=1}^{k} v_i \ln |S_p|$$

$$C_2 = \frac{(p-1)(p+2)}{6(k-1)} \left[\sum_{i=1}^{k} \frac{1}{v_i^2} - \frac{1}{\sum_{i=1}^{k} v_i^2} \right]$$

กำหนดให้

$$a_1 = \frac{1}{2}(k-1)p(p+1) \quad , \quad a_2 = \frac{a_1 + 2}{\left| C_2 - C_1^2 \right|}$$

$$b_1 = \frac{1 - C_1 - a_1 / a_2}{a_1} \quad , \quad b_2 = \frac{1 - C_1 - 2 / a_2}{a_2}$$

สถิติทดสอบ F จะแบ่งเป็น 2 กรณี คือ

1) ถ้า $C_2 > C_1^2$ สถิติทดสอบ F จะเป็น

$$F = -2b_1 \ln M$$

2) ถ้า $C_2 < C_1^2$ สถิติทดสอบ F จะเป็น

$$F = -\frac{a_2 b_2 \ln M}{a_1 (1 + 2b_2 \ln M)}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ค่าสถิติ F ทั้งสองสมการข้างต้น จะมีการแจกแจงโดยประมาณแบบ F ที่องศาอิสระ a_1 และ a_2 ตามลำดับ

การสรุปผลการทดสอบ จะปฏิเสธสมมติฐานว่างถ้า F > F_α ที่องศาอิสระ a_1 และ a_2 ตามลำดับ

3.2.2 การประมาณค่าพารามิเตอร์และการจัดกลุ่ม

3.2.2.1) <u>การวิเคราะห์จำแนกประเภท</u>

กำหนดค่าความน่าจะเป็นก่อน (prior probability) แบ่งเป็น 2 รูปแบบ ดังนี้

1) กำหนดให้มีค่าเท่ากันทุกกลุ่มในแต่ละตัวแปรตาม

2) ให้มีค่าเท่ากับความน่าจะเป็นที่แท้จริงในแต่ละกลุ่มของข้อมูลที่ทำการศึกษา

การกำหนดกลุ่มในรูปแบบของฟังก์ชันในการจัดกลุ่ม ในที่นี้จะเขียนอยู่ในรูป $y_1(\underset{\sim}{x}), y_2(\underset{\sim}{x})..., y_C(\underset{\sim}{x})$ ซึ่งกำหนดให้ค่าสังเกตของ x อยู่ในกลุ่มที่ ω_C ถ้า

$$y_C(\underset{\sim}{x}) > y_i(\underset{\sim}{x}) \qquad \forall i \neq c$$

ดังนั้น ค่าความคลาดเคลื่อนของความน่าจะเป็นที่ต่ำที่สุด จะอยู่ในรูป

$$y_C(\underset{\sim}{x}) \;=\; P(\omega_c \mid \underset{\sim}{x}) \propto p(\underset{\sim}{x} \mid \omega_C)P(\omega_C)$$

แต่ถ้าความสัมพันธ์ของข้อมูลมูลมีขนาดใหญ่ เราสามารถนำค่าลอการิทึมลงในสมการ ฟังก์ชันของการจัดกลุ่ม จะอยู่ในรูป

$$y_C(\underset{\sim}{x}) \;=\; \ln(p(\underset{\sim}{x} \mid \omega_C)) + \ln(P(\omega_C))$$

กำหนดให้เวกเตอร์ตัวแปรอิสระ **x** จะอยู่ในกลุ่มที่ ω_C เมื่อค่าความน่าจะเป็นของความคลาดเคลื่อนจะมีค่าต่ำที่สุด โดยกำหนดให้ในแต่ละจุดที่ถูกจัดอยู่ในกลุ่มนั้นมีค่าความน่าจะเป็นหลังสูงที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การประมาณค่าฟังก์ชันจำแนกประเภทจะใช้ตัวแบบ Gaussian density เขียนอยู่ในรูป

$$P(\underset{\sim}{x}\,|\,\omega_C) = \frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_C\right|^{1/2}}\exp\left(-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim C}{\mu})^T\boldsymbol{\Sigma}_C^{-1}(\underset{\sim}{x}-\underset{\sim C}{\mu})\right)$$

จะได้ว่า ฟังก์ชันจำแนกประเภท คือ

$$y_C(x) = -\frac{1}{2}(\underset{\sim}{x}-\underset{\sim C}{\mu})^T\boldsymbol{\Sigma}_C^{-1}(\underset{\sim}{x}-\underset{\sim C}{\mu}) - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_C\right| + \ln P(\omega_C)$$

พิจารณาในกรณีที่เมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่มมีค่าเท่ากัน $\boldsymbol{\Sigma}_C = \boldsymbol{\Sigma}$ จะพบว่า

$$y_C(\underset{\sim}{x}) = -\frac{1}{2}(\underset{\sim}{x}-\underset{\sim C}{\mu})^T\boldsymbol{\Sigma}^{-1}(\underset{\sim}{x}-\underset{\sim C}{\mu}) + \ln P(\omega_C)$$

หรือสามารถเขียนให้อยู่ในรูป

$$y_C(\underset{\sim}{x}) = (\underset{\sim C}{\mu}^T\boldsymbol{\Sigma}^{-1})\underset{\sim}{x} - \left(\frac{1}{2}\underset{\sim \mathbf{C}}{\mu}^{\mathbf{T}}\boldsymbol{\Sigma}\underset{\sim C}{\mu} + \ln P(\omega_C)\right)$$

$$= \underset{\sim \mathbf{C}}{w}^{\mathbf{T}}\underset{\sim}{x} + w_{C0}$$

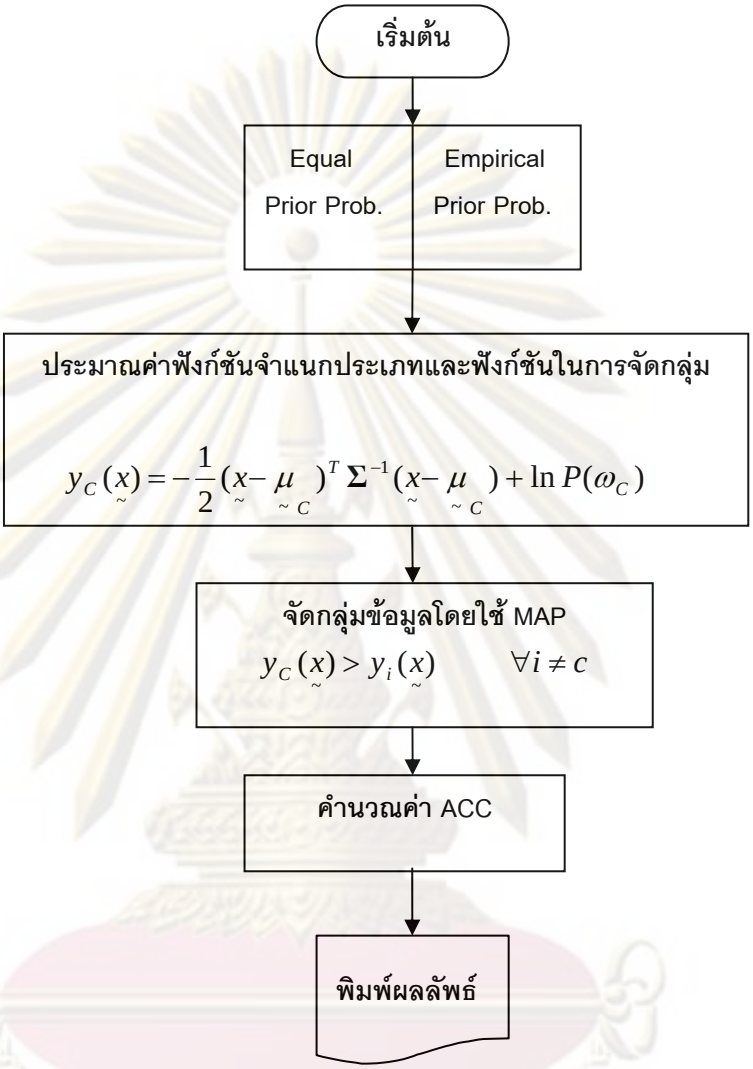
$$\underset{\sim \mathbf{C}}{w}^{\mathbf{T}} = \underset{\sim C}{\mu}^T\boldsymbol{\Sigma}^{-1} \qquad w_{C0} = -\frac{1}{2}\underset{\sim C}{\mu}^{\mathbf{T}}\boldsymbol{\Sigma}\underset{\sim C}{\mu} + \ln P(\omega_C)$$

$$y_C(\underset{\sim}{x}) = \underset{\sim \mathbf{C}}{w}^{\mathbf{T}}\underset{\sim}{x} + w_{C0}$$

ซึ่ง $\underset{\sim \mathbf{C}}{w}^{\mathbf{T}}\underset{\sim}{x}$ คือ ฟังก์ชันจำแนกประเภท $w_1x_1 + w_2x_2 + \ldots + w_px_p$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ผังแสดงขั้นตอนในการวิเคราะห์จำแนกประเภท



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.2.2.2) <u>การวิเคราะห์มัลติโนเมียลโลจิสติค</u>

ประมาณค่าโดยใช้กระบวนการกำลังสองน้อยสุดที่ปรับค่าถ่วงน้ำหนักในแต่ละรอบของการประมาณค่า (re-weight least square) โดยใช้สมการประมาณค่าโดยนัยทั่วไป $Q_{\beta}(\beta;\alpha)$ มีค่าเท่ากับ 0 เมื่อ

$$Q_{\beta}(\underset{\sim}{\beta};\alpha) = \sum_{i=1}^{N} \mathbf{D_i^T W(\alpha_i)^{-1}} (\underset{\sim}{z}_i - \underset{\sim}{\mu}_i) \qquad (21)$$

เมื่อ $\mathbf{D_i} = \boldsymbol{\mu}'(\underset{\sim}{\beta})_i$ คือค่าอนุพันธ์อันดับแรก ของ $\boldsymbol{\mu}(\underset{\sim}{\beta})_i$ ซึ่งขึ้นอยู่กับ $\boldsymbol{\beta}$

$$\mathbf{W(\alpha)}_i = \mathbf{V_i^{1/2} R(\alpha)_i V_i^{1/2}}$$

เมื่อ $\mathbf{R(\alpha)_i}$ คือฟังก์ชันที่ไม่ทราบความสัมพันธ์ของพารามิเตอร์ $\boldsymbol{\alpha}$ ซึ่งเป็นเมทริกซ์ของความสัมพันธ์ระหว่างองค์ประกอบของ $\underset{\sim}{z}_i$

$\mathbf{V_i^{1/2}}$ คือ รากที่สองของเมทริกซ์ความแปรปรวนระหว่างค่า $\underset{\sim}{z}_i$ ของตัวแปรที่ทำการเปรียบเทียบ โดยกำหนดให้มีค่าตามแนวเส้นทแยงมุมเท่ากับ $[\mu_{itk}(1-\mu_{itk})]^{1/2}$

จะได้ว่า

$$\mathbf{V_\beta(\alpha)} = \left[\sum_{i=1}^{N} \mathbf{D_i^T W(\alpha_i)^{-1} D_i}\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{D_i^T W(\alpha_i)^{-1} cov(z_i) W(\alpha_i)^{-1} D_i}\right] \left[\sum_{i=1}^{N} \mathbf{D_i^T W(\alpha_i)^{-1} D_i}\right]^{-1}$$

จากสมการข้างต้น มีคุณสมบัติเป็นตัวประมาณที่มีความคงเส้นคงวาของความแปรปรวนในพารามิเตอร์การถดถอย โดยไม่คำนึงถึง $\mathbf{W(\alpha_i)}$ สำหรับสมการข้างต้น ถ้ากำหนดให้การประมาณค่ามีความคงเส้นคงวาสำหรับพารามิเตอร์ในตัวแบบ เมื่อโครงสร้างความแปรปรวนร่วมของ $\underset{\sim}{z}_i$ ไม่มีการปรับเปลี่ยนโครงสร้าง

ในงานวิจัยชิ้นนี้ผู้วิจัยกำหนดให้ โครงสร้างความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระและยูนิฟอร์ม ดังนั้นจะได้เมทริกซ์ $\mathbf{C(\alpha)}$ และ $\mathbf{S}$ มี 2 ลักษณะ ดังนี้

1. <u>เมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระกัน</u>

$$C(\alpha)=\begin{bmatrix}1 & 0 & \dots & 0\\ 0 & 1 & \ddots & \vdots\\ \vdots & \ddots & 1 & 0\\ 0 & \dots & 0 & 1\end{bmatrix}_{TxT}$$

$$\mathbf{S}=\begin{bmatrix}\rho_{11} & \cdots & \rho_{1(K-1)}\\ \vdots & \ddots & \vdots\\ \rho_{(K-1)1} & \cdots & \rho_{(K-1)(K-1)}\end{bmatrix}_{(K-1)x(K-1)}$$

2. <u>เมื่อกำหนดให้ความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม</u>

$$C(\alpha)=\begin{bmatrix}1 & \alpha & \dots & \alpha\\ \alpha & 1 & \ddots & \vdots\\ \vdots & \ddots & 1 & \alpha\\ \alpha & \dots & \alpha & 1\end{bmatrix}_{TxT}$$

$$\mathbf{S}=\begin{bmatrix}\rho_{11} & \cdots & \rho_{1(K-1)}\\ \vdots & \ddots & \vdots\\ \rho_{(K-1)1} & \cdots & \rho_{(K-1)(K-1)}\end{bmatrix}_{(K-1)x(K-1)}$$

เมื่อหาความน่าจะเป็นสะสมสามารถหาค่าความน่าจะเป็นในแต่ละกลุ่มได้ ดังนี้

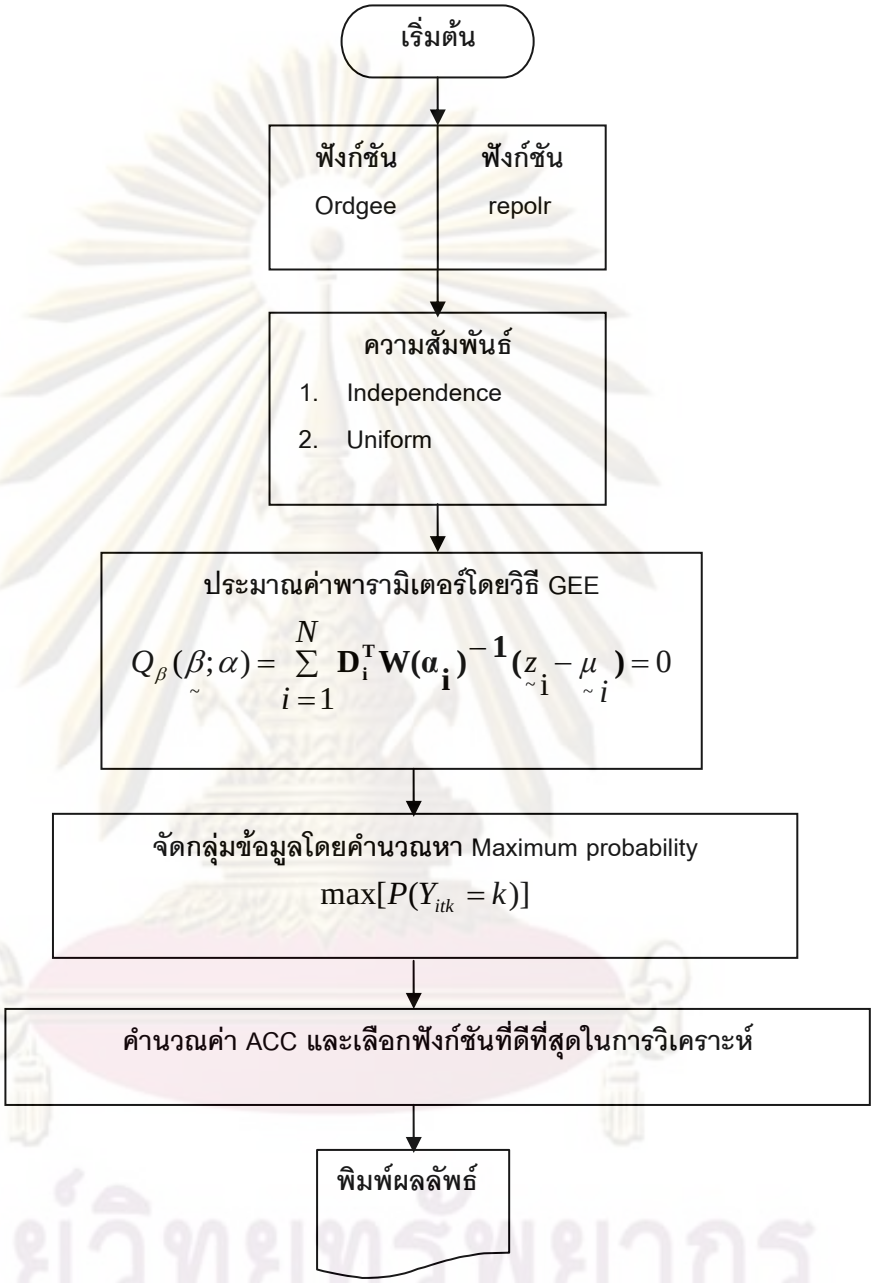
$$P(Y_{itk}\le k) = \frac{\exp(\beta_{0k}+\beta_1 x_{it}+\beta_2(time)_{it})}{1+\exp(\beta_{0k}+\beta_1 x_{it}+\beta_2(time)_{it})} = \frac{1}{1+\exp\{-(\beta_{0k}+\beta_1 x_{it}+\beta_2(time)_{it})\}}$$

$$P(Y_{itk}=k) = P(Y_{itk}\le k)-P(Y_{itk}\le j) \quad \text{เมื่อ } j \le k$$

เมื่อได้ความน่าจะเป็นในแต่ละกลุ่ม พิจารณาหาค่ากลุ่มที่มีค่าความน่าจะเป็นสูงสุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ผังแสดงขั้นตอนในการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบวิธีการวิเคราะห์ข้อมูล ในกรณีที่การเก็บข้อมูลมีการเก็บซ้ำกันตั้งแต่ 2 ช่วงเวลาขึ้นไป เมื่อตัวแปรตอบสนองเป็นตัวแปรเชิงกลุ่มที่มีมากกว่า 2 กลุ่ม โดยทำการศึกษาจากข้อมูลจริงทางด้านการแพทย์ ซึ่งเปรียบเทียบวิธีการวิเคราะห์ 2 วิธี ได้แก่ การวิเคราะห์จำแนกประเภทกับการวิเคราะห์มัลติโนเมียลโลจิสติค และทำการเปรียบเทียบประสิทธิภาพในการจัดกลุ่มข้อมูลได้ถูกต้อง การวิเคราะห์วิธีใดที่สามารถจำแนกกลุ่มได้ถูกต้องมากที่สุด

ตัวแบบที่ทำการวิเคราะห์ มีลักษณะดังนี้

1) <u>ตัวแบบการวิเคราะห์จำแนกประเภท</u>

$$z_i = w_{i1}x + w_{i2}(time) \quad ; i = \min\{p, K-1\}$$

โดยที่ p คือ จำนวนตัวแปรอิสระ มีค่าเท่ากับ 2

K คือ จำนวนกลุ่มทั้งหมด

z_i คือ ฟังก์ชันจำแนก

w_1 , w_2 คือ น้ำหนักการจำแนกประเภทของตัวแปรตัวที่ 1 และตัวแปรตัวที่ 2 ตามลำดับ

x คือ ตัวแปรกำหนด

$time$ คือ ตัวแปรเวลา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2) ตัวแบบการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

$$logit\left[P(Y_{itk} \le k)\right] = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time)_{it} \quad , k = 1,...,K-1 \;\; ; t = 1,...,T \quad ; i = 1,...,n$$

โดยที่ $\left[P(Y_{itk} \le k)\right]$ คือ ความน่าจะเป็นที่ค่า Y_{it} จะน้อยกว่าหรือเท่ากับกลุ่มที่ k

β_{0k} คือ ค่า cut-point ของกลุ่มที่ k

β_1 , β_2 คือ สัมประสิทธิ์การถดถอย

K คือ จำนวนกลุ่มของตัวแปรตาม

x_{it} คือ ตัวแปรกำหนด สำหรับค่าสังเกตที่ i ระยะเวลาที่ t

$time_{it}$ คือ ตัวแปรเวลาสำหรับค่าสังเกตที่ i ระยะเวลาที่ t

n คือ ขนาดตัวอย่าง (Sample Size)

T คือ ระยะเวลาที่เก็บข้อมูลซ้ำ (Preiod)

4.1 กรณีศึกษาที่ 1: การติดตามผลของโรคนอนไม่หลับ[1]

ข้อมูลชุดนี้เป็นการศึกษาทดลองทางการแพทย์ โดยทำการศึกษาการเปรียบเทียบระหว่างยาที่ใช้ในการรักษาโรคนอนไม่หลับและยาหลอก โดยหน่วยตัวอย่าง คือ ผู้ป่วยที่มีอาการนอนไม่หลับจำนวน 239 คน มีตัวแปรตอบสนอง คือ ข้อมูลที่บันทึกระยะเวลาที่ผู้ป่วยใช้ในการนอนหลับ ซึ่งลักษณะของตัวแปรตอบสนองนี้แบ่งออกเป็น 4 กลุ่ม ตามระยะเวลาที่ใช้ในการนอนหลับ (น้อยกว่า 20 นาที, ระหว่าง 20-30 นาที, 30-60 นาที และมากกว่า 60 นาทีขึ้นไป) โดยมีตัวแปรอิสระ 2 ตัวแปร ได้แก่ ช่วงระยะเวลาที่ใช้เก็บข้อมูล ซึ่งผู้ป่วยทุกคนจะต้องมาทำการตรวจสอบ 2 ครั้ง (ระยะเริ่มต้น, ระยะติดตามผล) และประเภทของยาที่ใช้ในการทดลอง (ยาจริง, ยาหลอก) ที่ผู้ป่วยได้รับ [2] (รายละเอียดของข้อมูลแสดงไว้ในภาคผนวก ก)

จากการทดสอบคุณสมบัติเบื้องต้นของข้อมูล พบว่า ข้อมูลไม่มีคุณสมบัติการแจกแจงปกติหลายตัวแปร แต่มีเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมในแต่ละกลุ่มมีค่าเท่ากัน (รายละเอียดของข้อมูลแสดงไว้ในภาคผนวก ก)

---

[1] แหล่งที่มา : http://www.stat.ufl.edu/~aa/cda/sas/sas.html [31/01/2552]

[2] Agresti Alan, Categorical Data Analysis, (New York: Wiley, 2002), pp.462-463

4.1.1 การวิเคราะห์ข้อมูลจำแนกประเภท

ตัวแบบของฟังก์ชันจำแนก จะอยู่ในรูป

$$z_j = w_1 x_m + w_2 (time)_m \quad ; j = \min\{2, 4\text{-}1\} \quad ; m = 1,2,\ldots,478$$

จะได้ว่า ฟังก์ชันจำแนก จะแบ่งออกเป็น 2 สมการ ได้แก่

$$z_1 = w_{11} x + w_{12} (time)$$
$$z_2 = w_{21} x + w_{22} (time)$$

w_{i1} , w_{i2} คือ น้ำหนักการจำแนกประเภทของตัวแปรประเภทของยา และตัวแปรเวลา ตามลำดับ

x คือ ประเภทของยาที่ใช้ในการทดลอง (ยาหลอก=0, ยาจริง=1)

$time$ คือ ตัวแปรเวลา (ระยะเวลาเริ่มต้น=0, ระยะเวลาติดตามผล=1)

การวิเคราะห์ข้อมูลศึกษาภายใต้สถานการณ์ 2 สถานการณ์ ดังนี้

1) กำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง (equal prior probability)

2) กำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ (empirical prior probability)

ตารางที่ 4.1 ค่าความน่าจะเป็นก่อนที่ใช้ในการวิเคราะห์ข้อมูล กรณีศึกษาที่ 1

ระยะเวลาที่ใช้ในการหลับ	equal prior probability	empirical prior probability
น้อยกว่า 20 นาที	0.25	0.2
20 - 30 นาที	0.25	0.25
30 – 60 นาที	0.25	0.27
มากกว่า 60 นาที	0.25	0.28

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 4.1 แสดงค่าความน่าจะเป็นก่อนในแต่ละสถานการณ์ที่กำหนด พบว่า สถานการณ์ที่กำหนดมีค่าความน่าจะเป็นใกล้เคียงกัน เมื่อทำการวิเคราะห์ข้อมูล ผลการวิเคราะห์สัมประสิทธิ์ฟังก์ชันจำแนก ดังนี้

ตารางที่ 4.2 สัมประสิทธิ์ของฟังก์ชันจำแนกในแต่ละกรณี กรณีศึกษาที่ 1

ตัวแปรอิสระ	equal prior probability		empirical prior probability	
	ฟังก์ชัน 1	ฟังก์ชัน 2	ฟังก์ชัน 1	ฟังก์ชัน 2
ประเภทของยา	-0.6239369	-1.9095743	-0.6395478	-1.9044028
เวลา	-2.0671338	0.5740055	-2.0623657	0.5909078

จากตารางที่ 4.2 พบว่า ในแต่ละสถานการณ์ที่กำหนด สามารถคำนวณฟังก์ชันการจำแนกได้ 2 ฟังก์ชัน สามารถอธิบายได้ดังนี้

<u>กรณีที่ 1</u> เมื่อกำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง ฟังก์ชันจำแนก คือ

$$z_1 = -0.6239369(x) - 2.0671338(time)$$

$$z_2 = -1.9095743(x) + 0.5740055(time)$$

<u>กรณีที่ 2</u> เมื่อกำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ ฟังก์ชันจำแนก คือ

$$z_1 = -0.6395478(x) - 2.0623657(time)$$

$$z_2 = -1.9044028(x) + 0.5909078(time)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.3 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อน
มีค่าเท่ากันทุกระดับของตัวแปรตอบสนองเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 1

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	71	0	0	26
20 - 30 นาที	78	0	0	40
30 – 60 นาที	54	0	0	75
มากกว่า 60 นาที	36	0	0	98
ค่า ACC (%)	35.35565			

จากตารางที่ 4.3 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 71+98 = 169 จากจำนวนข้อมูลทั้งสิ้น 478 ค่า และค่า ACC เท่ากับ 35.3656 %

ตารางที่ 4.4 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อน
มีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์เปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 1

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	0	71	0	26
20 - 30 นาที	0	78	0	40
30 – 60 นาที	0	54	0	75
มากกว่า 60 นาที	0	36	0	98
ค่า ACC (%)	36.82008			

จากตารางที่ 4.4 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 78+98 = 176 จากจำนวนข้อมูลทั้งสิ้น 478 ค่า และค่า ACC เท่ากับ 36.82008 %

4.1.2 การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

ตัวแบบของสมการโลจิทจะอยู่ในรูป

$$logit\left[P(Y_{itk} \le k)\right] = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time)_{it} \quad , k = 1,...,4-1 \;; t = 1,...,2 \quad ; i = 1,...,239$$

โดยที่ $\left[P(Y_{itk} \le k)\right]$ คือ ความน่าจะเป็นที่ค่า Y_{it} จะน้อยกว่าหรือเท่ากับกลุ่มที่ k

β_{0k} คือ ค่า cut-point ของกลุ่มที่ k

β_1 , β_2 คือ สัมประสิทธิ์การถดถอย

x_{it} คือ ประเภทของยาที่ใช้ในการทดลอง (ยาหลอก=0, ยาจริง=1)

$time_{it}$ คือ ตัวแปรเวลา (ระยะเวลาเริ่มต้น=0, ระยะเวลาติดตามผล=1)

การวิเคราะห์ข้อมูลจะศึกษาภายใต้สถานการณ์ 2 สถานการณ์ ดังนี้

1) กำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ

2) กำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม

โดยมีฟังก์ชันสำเร็จรูปที่ใช้ในการวิเคราะห์ในแต่ละสถานการณ์จำนวน 2 ฟังก์ชัน ได้แก่ ฟังก์ชัน ordgee และ repolr

4.1.2.1 การวิเคราะห์โดยใช้ฟังก์ชันสำเร็จรูป ordgee

ตารางที่ 4.5 ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค
โดยใช้ฟังก์ชัน ordgee กรณีศึกษาที่ 1

ตัวแปรอิสระ	Working correlation matrix			
	Independence		Uniform	
	Estimate	Robust S.E.	Estimate	Robust S.E.
β_{01}	-1.6642250	0.4041138	-1.6510678	0.3952532
β_{02}	-0.4217746	0.4237817	-0.4113955	0.4158147
β_{03}	0.7972724	0.4555004	0.7959079	0.4434476
ประเภทของยา	-0.3881508	0.5806307	-0.4231898	0.5706544
เวลา	0.7909077	0.5428676	0.8056417	0.5195927

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ โดยใช้ฟังก์ชัน ordgee อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = -1.664225 - 0.3881508\ x + 0.79077\ (time)$$
$$logit\left[P(Y \le 2)\right] = -0.4217746 - 0.3881508\ x + 0.79077\ (time)$$
$$logit\left[P(Y \le 3)\right] = -0.7972724 - 0.3881508\ x + 0.79077\ (time)$$

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม โดยใช้ฟังก์ชัน ordgee อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = -1.6510678 - 0.4231898\ x + 0.8056417\ (time)$$
$$logit\left[P(Y \le 2)\right] = -0.4113955 - 0.4231898\ x + 0.8056417\ (time)$$
$$logit\left[P(Y \le 3)\right] = 0.7959079 - 0.4231898\ x + 0.8056417\ (time)$$

จากตารางที่ 4.5 พบว่า ส่วนเบี่ยงเบนมาตรฐานเมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม มีค่าต่ำกว่า รูปแบบของ Working correlation matrix ที่มีลักษณะเป็นแบบอิสระ

ตารางที่ 4.6 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน ordgee กรณีศึกษาที่ 1

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	0	0	97	0
20 - 30 นาที	0	0	118	0
30 – 60 นาที	0	0	129	0
มากกว่า 60 นาที	0	0	134	0
ค่า ACC (%)	26.987448			

จากตารางที่ 4.6 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 129 จากจำนวนข้อมูลทั้งสิ้น 478 ค่า และค่า ACC เท่ากับ 26.987448 %

ตารางที่ 4.7 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์มเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน ordgee กรณีศึกษาที่ 1

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	31	40	0	26
20 - 30 นาที	29	49	0	40
30 – 60 นาที	35	19	0	75
มากกว่า 60 นาที	25	11	0	98
ค่า ACC (%)	37.23849			

จากตารางที่ 4.7 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 31+49+98 = 178 จากจำนวนข้อมูลทั้งสิ้น 478 ค่า และค่า ACC เท่ากับ 37.23849 %

4.1.2.2 **การวิเคราะห์ใช้ฟังก์ชันสำเร็จรูป repolr**

ตารางที่ 4.8 ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน repolr กรณีศึกษาที่ 1

ตัวแปรอิสระ	Working correlation matrix			
	Independence		Uniform	
	Estimate	Robust S.E.	Estimate	Robust S.E.
β_{01}	-2.4324414	0.2133084	-2.4376109	0.2131552
β_{02}	-1.1333301	0.1799315	-1.1357016	0.1795666
β_{03}	0.1625974	0.1694227	0.1621352	0.1687309
ประเภทของยา	0.3938017	0.2005294	0.3921730	0.2005456
เวลา	1.3936014	0.1382672	1.3946864	0.1383649

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ โดยใช้ฟังก์ชัน repolr อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = -2.4324414 - 0.3938017\ x + 1.3936014\ (time)$$
$$logit\left[P(Y \le 2)\right] = -1.1333301 - 0.3938017\ x + 1.3936014\ (time)$$
$$logit\left[P(Y \le 3)\right] = -0.1625974 - 0.3938017\ x + 1.3936014\ (time)$$

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม โดยใช้ฟังก์ชัน repolr อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = -2.4376109 - 0.4231898\ x + 0.8056417\ (time)$$
$$logit\left[P(Y \le 2)\right] = -1.1357016 - 0.4231898\ x + 0.8056417\ (time)$$
$$logit\left[P(Y \le 3)\right] = 1.3946864 - 0.4231898\ x + 0.8056417\ (time)$$

จากตารางที่ 4.8 พบว่า ส่วนเบี่ยงเบนมาตรฐานเมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม มีค่าต่ำกว่า รูปแบบของ Working correlation matrix ที่มีลักษณะเป็นแบบอิสระ

ตารางที่ 4.9 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล สำหรับฟังก์ชัน repolr กรณีศึกษาที่ 1

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	40	31	0	26
20 - 30 นาที	49	29	0	40
30 – 60 นาที	19	35	0	75
มากกว่า 60 นาที	11	25	0	98
ค่า ACC (%)	34.93724			

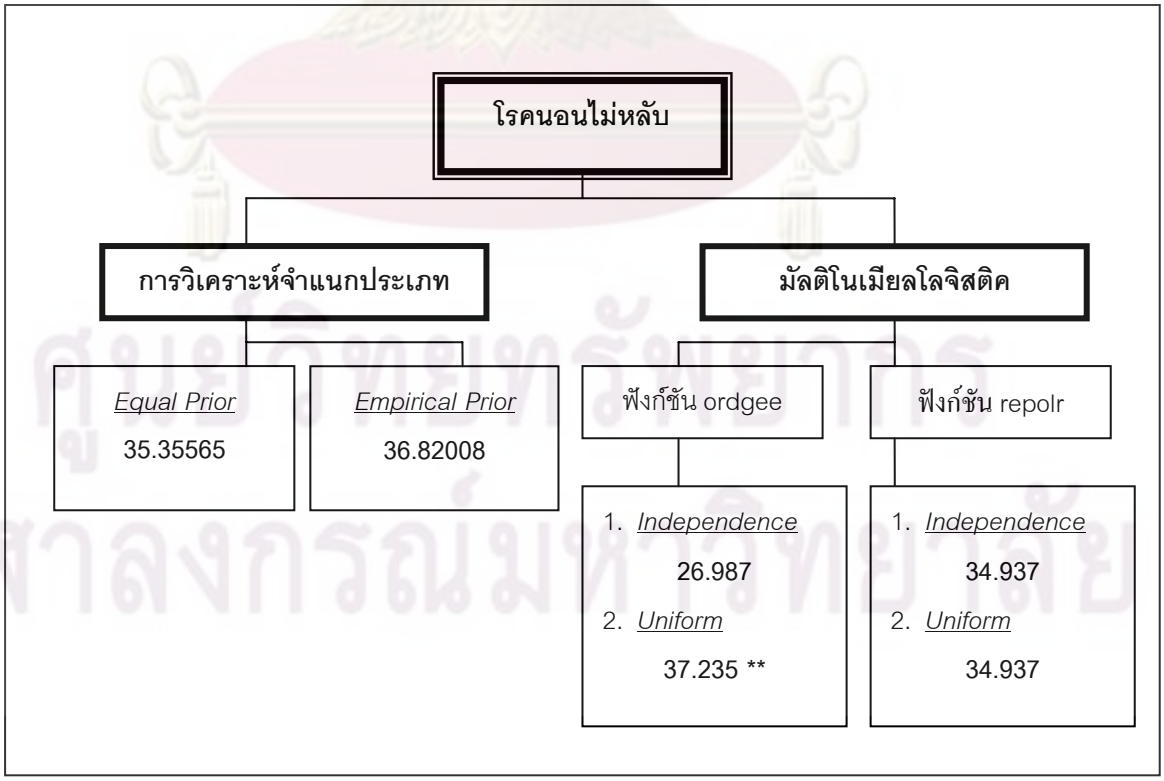
จากตารางที่ 4.9 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 40+29+98 = 167 จากจำนวนข้อมูลทั้งสิ้น 478 ค่า และค่า ACC เท่ากับ 34.96724 %

ตารางที่ 4.10 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์มเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน repolr กรณีศึกษาที่ 1

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	40	31	0	26
20 - 30 นาที	49	29	0	40
30 – 60 นาที	19	35	0	75
มากกว่า 60 นาที	11	25	0	98
ค่า ACC (%)	34.93724			

จากตารางที่ 4.10 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 40+29+98 = 167 จากจำนวนข้อมูลทั้งสิ้น 478 ค่า และค่า ACC เท่ากับ 37.23849%

รูปภาพที่ 4.1 แผนผังแสดงค่า ACC ที่ได้จากการคำนวณ กรณีศึกษาที่ 1



4.2 กรณีศึกษาที่ 2: การผ่าตัดเปลี่ยนผิวข้อสะโพก [1]

ข้อมูลชุดนี้เป็นการศึกษาทดลองทางการแพทย์ เป็นข้อมูลเกี่ยวกับการผ่าตัดเปลี่ยนผิวข้อสะโพก (Hip resurfacing) ตัวแปรตอบสนอง คือ ระดับความเจ็บปวดภายหลังจากการผ่าตัด ซึ่งมี 4 ระดับความเจ็บปวด (none, slight, mild, moderate) โดยมีหน่วยทดลอง คือ ผู้ป่วยที่เข้ารับการรักษาและทำการตอดตามผลการรักษา จำนวนผู้ทดลอง 58 คน ในการศึกษาครั้งนี้มีตัวแปรอิสระ 2 ตัวแปร ได้แก่ ช่วงระยะเวลาที่ใช้เก็บข้อมูล ซึ่งผู้ป่วยทุกคนจะต้องมาทำการตรวจสอบ 3 ครั้ง (ระยะเวลาเริ่มต้น, 2 ปี และ 5 ปี) และเพศ (ชาย, หญิง) (รายละเอียดข้อมูลแสดงไว้ในภาคผนวก ก)

จากการทดสอบคุณสมบัติเบื้องต้นของข้อมูล พบว่า ข้อมูลไม่มีคุณสมบัติการแจกแจงปกติหลายตัวแปร แต่มีเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมในแต่ละกลุ่มมีค่าเท่ากัน (รายละเอียดของข้อมูลแสดงไว้ในภาคผนวก ก)

4.2.1 การวิเคราะห์ข้อมูลจำแนกประเภท

ตัวแบบของฟังก์ชันจำแนก จะอยู่ในรูป

$$z_i = w_{i1}x + w_{i2}(time) \quad ; i = \min\{2, 4\text{-}1\} \quad ; m = 1,2,\ldots,174$$

จะได้ว่า ฟังก์ชันจำแนก จะแบ่งออกเป็น 2 สมการ ได้แก่

$$z_1 = w_{11}x + w_{12}(time)$$
$$z_2 = w_{21}x + w_{22}(time)$$

w_{i1} , w_{i2} คือ น้ำหนักการจำแนกประเภทของตัวแปรประเภทของเพศ และตัวแปรเวลาตามลำดับ

x คือ เพศ (ชาย = 1, หญิง = 0)

$time$ คือ ตัวแปรเวลา (ระยะเวลาเริ่มต้น = 1 , 2 ปี = 2 และ 5 ปี = 5)

---

[1] แหล่งที่มา : http://www.warwick.ac.uk/go/repolr. [20/04/2552]

[2] Parsons N.R. and Matthew L. Costa, “Repeated measure proportional odds logistic regression analysis of ordinal score data in the statistical software package R”, Journal of Science Direct 53 (2008): 632-641

การวิเคราะห์ข้อมูลศึกษาภายใต้สถานการณ์ 2 สถานการณ์ ดังนี้

1) กำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง (equal prior probability)

2) กำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ (empirical prior probability)

ตารางที่ 4.11 ค่าความน่าจะเป็นก่อนที่ใช้ในการวิเคราะห์ข้อมูล กรณีศึกษาที่ 2

ระดับความเจ็บปวด	equal prior probability	empirical prior probability
None	0.25	0.61
Slight	0.25	0.21
Mild	0.25	0.13
Moderate	0.25	0.06

จากตารางที่ 4.11 แสดงค่าความน่าจะเป็นก่อนในแต่ละสถานการณ์ที่กำหนด พบว่าสถานการณ์ที่กำหนดมีค่าความน่าจะเป็นแตกต่างกัน เมื่อทำการวิเคราะห์ข้อมูล ผลการวิเคราะห์สัมประสิทธิ์ฟังก์ชันจำแนก ดังนี้

ตารางที่ 4.12 สัมประสิทธิ์ของฟังก์ชันจำแนกในแต่ละกรณี กรณีศึกษาที่ 2

ตัวแปรอิสระ	equal prior probability		empirical prior probability	
	ฟังก์ชัน 1	ฟังก์ชัน 2	ฟังก์ชัน 1	ฟังก์ชัน 2
เพศ	0.5880606	0.1311794	-0.5621317	-0.216867
เวลา	-0.5014310	2.0155540	0.7945498	-1.919005

จากตารางที่ 4.2 พบว่า ในแต่ละสถานการณ์ที่กำหนด สามารถคำนวณฟังก์ชันการจำแนกได้ 2 ฟังก์ชัน สามารถอธิบายได้ดังนี้

กรณีที่ 1 เมื่อกำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง ฟังก์ชันจำแนก คือ

$$z_1 = 0.5880606(x) - 0.5014310(time)$$

$$z_2 = 0.1311794(x) + 2.0155540(time)$$

กรณีที่ 2 เมื่อกำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ ฟังก์ชันจำแนก คือ

$$z_1 = -0.5621317(x) + 0.7945498(time)$$

$$z_2 = -0.216867(x) - 1.919005(time)$$

ตารางที่ 4.13 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนองเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 2

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	None	Slight	Mild	Moderate
None	53	0	23	30
Slight	11	0	12	13
Mild	9	0	6	7
Moderate	1	0	1	8
ค่า ACC (%)	38.50575			

จากตารางที่ 4.13 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 53+6+8 = 67 จากจำนวนข้อมูลทั้งสิ้น 174 ค่า และค่า ACC เท่ากับ 38.50575 %

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.14 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อน
มีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์เปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	None	Slight	Mild	Moderate
None	106	0	0	0
Slight	36	0	0	0
Mild	22	0	0	0
Moderate	10	0	0	0
ค่า ACC (%)	60.91954			

จากตารางที่ 4.14 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 106 จากจำนวนข้อมูลทั้งสิ้น 174 ค่า และค่า ACC เท่ากับ 60.9154 %

4.1.2 **การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค**

ตัวแบบของสมการโลจิทจะอยู่ในรูป

$$logit\left[P(Y_{itk} \le k)\right] = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time)_{it} \quad , k = 1,...,4-1 \;; t = 1,...,3 \quad ; i = 1,...,58$$

โดยที่ $logit\left[P(Y \le k)\right]$ คือ โลจิทของความน่าจะเป็นที่ค่าสังเกต Y จะอยู่ในกลุ่มที่ k

β_{0k} คือ ค่า cut-point ของกลุ่มที่ k

β_1 , β_2 คือ สัมประสิทธิ์การถดถอย

x_{it} คือ เพศ (ชาย = 1, หญิง = 0)

$time_{it}$ คือ ตัวแปรเวลา (ระยะเวลาเริ่มต้น = 1 , 2 ปี = 2 และ 5 ปี = 5)

การวิเคราะห์ข้อมูลจะศึกษาภายใต้สถานการณ์ 2 สถานการณ์ ดังนี้

1) กำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ

2) กำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม

โดยมีฟังก์ชันสำเร็จรูปที่ใช้ในการวิเคราะห์ในแต่ละสถานการณ์จำนวน 2 ฟังก์ชัน ได้แก่ ฟังก์ชัน ordgee และ repolr

4.1.2.1 การวิเคราะห์ใช้ฟังก์ชัน ordgee

ตารางที่ 4.15 ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค
โดยใช้ฟังก์ชัน ordgee กรณีศึกษาที่ 2

ตัวแปรอิสระ	Working correlation matrix			
	Independence		Uniform	
	Estimate	Robust S.E.	Estimate	Robust S.E.
β_{01}	2.85696114	1.320559	NaN	NaN
β_{02}	3.89223616	1.286791	NaN	NaN
β_{03}	5.22516847	1.392501	NaN	NaN
เพศ	0.09079586	1.091562	NaN	NaN
เวลา	-0.70725367	0.270320	NaN	NaN

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ โดยใช้ฟังก์ชัน ordgee อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = 2.85696114 + 0.09079586\ x - 0.70725367\ (time)$$
$$logit\left[P(Y \le 2)\right] = 3.89223616 + 0.09079586\ x - 0.70725367\ (time)$$
$$logit\left[P(Y \le 3)\right] = 5.22516847 + 0.09079586\ x - 0.70725367\ (time)$$

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม พบว่าไม่สามารถประมาณค่าพารามิเตอร์ได้ จึงไม่มีสมการถดถอย

จากตารางที่ 4.15 สรุปได้ว่าข้อมูลชุดนี้ สามารถประมาณค่าพารามิเตอร์ได้เพียงรูปแบบเดียว คือ เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.16 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูลของฟังก์ชัน ordgee กรณีศึกษาที่ 2

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	106	0	0	0
20 - 30 นาที	36	0	0	0
30 – 60 นาที	22	0	0	0
มากกว่า 60 นาที	10	0	0	0
ค่า ACC (%)	60.91954			

จากตารางที่ 4.16 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 106 จากจำนวนข้อมูลทั้งสิ้น 174 ค่า และค่า ACC เท่ากับ 60.91954 %

เนื่องจากข้อมูลชุดนี้ เมื่อกำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์ม โปรแกรมสำเร็จรูป ordgee ไม่สามารถประมาณค่าพารามิเตอร์ได้ ส่งผลให้ไม่สามารถคำนวณค่า ACC ได้

4.1.2.2 **การวิเคราะห์โดยใช้ฟังก์ชัน** repolr

ตารางที่ 4.17 ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน repolr กรณีศึกษาที่ 2

ตัวแปรอิสระ	Working correlation matrix			
	Independence		Uniform	
	Estimate	Robust S.E.	Estimate	Robust S.E.
β_{01}	0.5875052	0.3536206	0.5671770	0.35216136
β_{02}	1.6640954	0.3599421	1.6414844	0.35665419
β_{03}	3.0029364	0.3940484	2.9199674	0.39489869
เพศ	-0.1936970	0.0549917	-0.1937062	0.05484294
เวลา	0.5993758	0.4377597	0.6237273	0.43918752

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ โดยใช้ฟังก์ชัน repolr อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = 0.5875052 - 0.1936970\ x + 0.5993758\ (time)$$
$$logit\left[P(Y \le 2)\right] = 1.6640954 - 0.1936970\ x + 0.5993758\ (time)$$
$$logit\left[P(Y \le 3)\right] = 3.0029364 - 0.1936970\ x + 0.5993758\ (time)$$

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม โดยใช้ฟังก์ชัน repolr อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = 0.5671770 - 0.1937062\ x + 0.6237273\ (time)$$
$$logit\left[P(Y \le 2)\right] = 1.6414844 - 0.1937062\ x + 0.6237273\ (time)$$
$$logit\left[P(Y \le 3)\right] = 2.9199674 - 0.1937062\ x + 0.6237273\ (time)$$

จากตารางที่ 4.17 พบว่า ส่วนเบี่ยงเบนมาตรฐานเมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม มีค่าต่ำกว่า รูปแบบของ Working correlation matrix ที่มีลักษณะเป็นแบบอิสระ

ตารางที่ 4.18 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล สำหรับฟังก์ชัน repolr กรณีศึกษาที่ 2

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	106	0	0	0
20 - 30 นาที	36	0	0	0
30 – 60 นาที	22	0	0	0
มากกว่า 60 นาที	10	0	0	0
ค่า ACC (%)	60.91954			

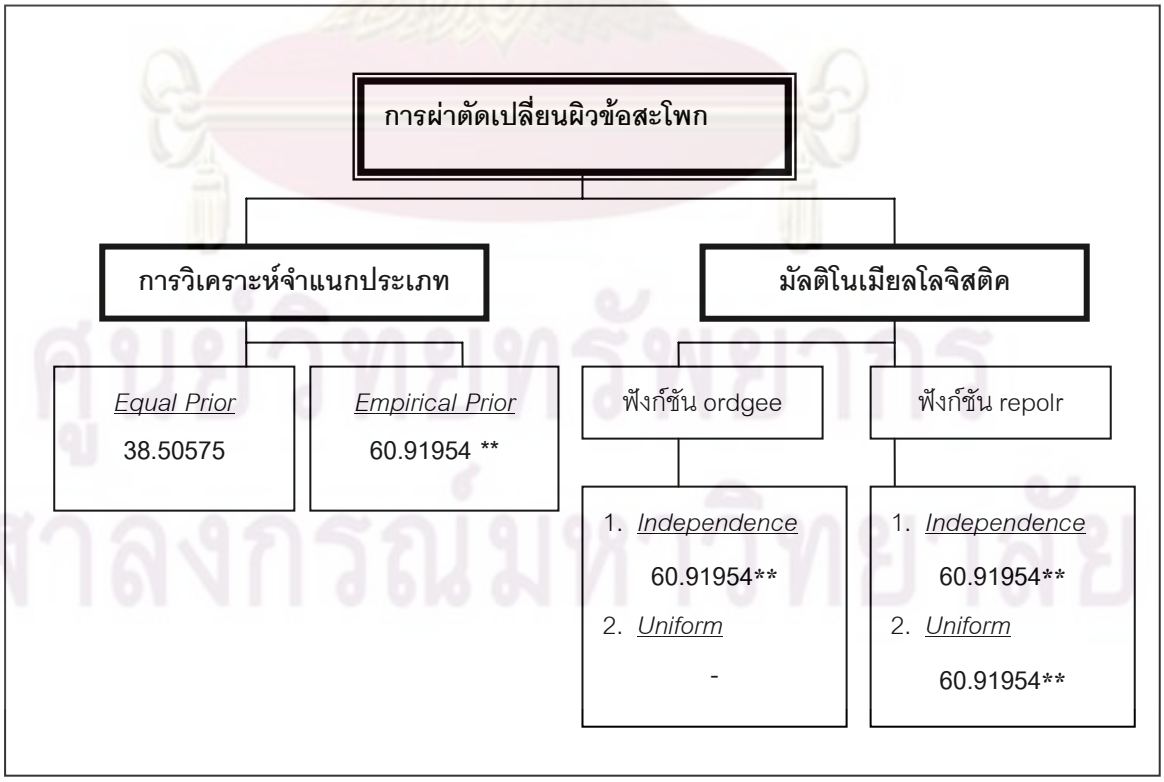
จากตารางที่ 4.18 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 106 จากจำนวนข้อมูลทั้งสิ้น 174 ค่า และค่า ACC เท่ากับ 60.91954 %

ตารางที่ 4.19 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์มเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล สำหรับฟังก์ชัน repolr กรณีศึกษาที่ 2

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท			
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที	มากกว่า 60 นาที
น้อยกว่า 20 นาที	106	0	0	0
20 - 30 นาที	36	0	0	0
30 – 60 นาที	22	0	0	0
มากกว่า 60 นาที	10	0	0	0
ค่า ACC (%)	60.91954			

จากตารางที่ 4.19 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 106 จากจำนวนข้อมูลทั้งสิ้น 174 ค่า และค่า ACC เท่ากับ 60.91954 %

รูปภาพที่ 4.2 แผนผังแสดงค่า ACC ที่ได้จากการคำนวณ กรณีศึกษาที่ 2



4.3 กรณีศึกษาที่ 3: การรักษาอาการเอ็นร้อยหวายฉีก [1]

ข้อมูลชุดนี้เป็นการศึกษาทดลองทางการแพทย์ เป็นข้อมูลเกี่ยวกับการรักษาอาการเอ็นร้อยหวายฉีก (Achilles tendon rupture) ตัวแปรตอบสนอง คือ ปัญหาจากการรักษา ซึ่งมี 3 ระดับของการเกิดปัญหาในการทำงาน (ไม่มีปัญหา, มีปัญหาบ้าง, ไม่สามารถทำงานได้) โดยมีหน่วยทดลอง คือ ผู้ป่วยที่เข้ารับการรักษาและทำการติดตามผลการรักษา จำนวนผู้ทดลอง 48 คน ในการศึกษาครั้งนี้มีตัวแปรอิสระ 2 ตัวแปร ได้แก่ ช่วงระยะเวลาที่ใช้เก็บข้อมูล ซึ่งผู้ป่วยทุกคนจะต้องมาทำการตรวจสอบ 3 ครั้ง (ระยะเวลาเริ่มต้น, 6 เดือน และ 12 เดือน) และลักษณะของผ้าที่ใช้พันแบ่งเป็น 2 กลุ่ม (กลุ่มทดลอง ใช้เส้นใยคาร์บอนมีขนาด 1.5 เซนติเมตร และกลุ่มควบคุม คือ ผู้ที่ใช้เส้นใยธรรมดา) [2]

จากการทดสอบคุณสมบัติเบื้องต้นของข้อมูล พบว่า ข้อมูลไม่มีคุณสมบัติการแจกแจงปกติหลายตัวแปร แต่มีเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมในแต่ละกลุ่มมีค่าเท่ากัน (รายละเอียดของข้อมูลแสดงไว้ในภาคผนวก ก)

4.1.1 การวิเคราะห์ข้อมูลจำแนกประเภท

ตัวแบบของฟังก์ชันจำแนก จะอยู่ในรูป

โดยที่ $z_i = w_{i1}x + w_{i2}(time)$; i = min{2, 3-1} ; m = 1,2,…,144

จะได้ว่า ฟังก์ชันจำแนก จะแบ่งออกเป็น 2 สมการ ได้แก่

$$z_1 = w_{11}x + w_{12}(time)$$
$$z_2 = w_{21}x + w_{22}(time)$$

w_{i1} , w_{i2} คือ น้ำหนักการจำแนกประเภทสำหรับสมการที่ i โดยมีตัวแปรประเภทของผ้าและตัวแปรเวลา ตามลำดับ

---

[1] แหล่งที่มา : http://www.warwick.ac.uk/go/repolr. [25/04/2552]

[2] Parsons N.R. and Matthew L. Costa, “Repeated measure proportional odds logistic regression analysis of ordinal score data in the statistical software package R”, Journal of Science Direct 53 (2008): 632-641.

x คือ ประเภทของผ้าที่ใช้พัน (กลุ่มทดลอง = 1 และกลุ่มควบคุม = 0)

time คือ ตัวแปรเวลา (ระยะเวลาเริ่มต้น = 0, 6 เดือน = 1 และ 12 เดือน = 2)

การวิเคราะห์ข้อมูลศึกษาภายใต้สถานการณ์ 2 สถานการณ์ ดังนี้

1) กำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง (equal prior probability)

2) กำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ (empirical prior probability)

ตารางที่ 4.20 ค่าความน่าจะเป็นก่อนที่ใช้ในการวิเคราะห์ข้อมูล กรณีศึกษาที่ 3

ปัญหาจากการรักษา	equal prior probability	empirical prior probability
ไม่มีปัญหา	0.33	0.608
มีปัญหาบ้าง	0.33	0.344
ไม่สามารถทำงานได้	0.33	0.048

จากตารางที่ 4.20 แสดงค่าความน่าจะเป็นก่อนในแต่ละสถานการณ์ที่กำหนด พบว่าสถานการณ์ที่กำหนดมีค่าความน่าจะเป็นใกล้เคียงกัน เมื่อทำการวิเคราะห์ข้อมูล ผลการวิเคราะห์สัมประสิทธิ์ฟังก์ชันจำแนก ดังนี้

ตารางที่ 4.21 สัมประสิทธิ์ของฟังก์ชันจำแนกในแต่ละกรณี กรณีศึกษาที่ 3

ตัวแปรอิสระ	equal prior probability		empirical prior probability	
	ฟังก์ชัน 1	ฟังก์ชัน 2	ฟังก์ชัน 1	ฟังก์ชัน 2
ประเภทของผ้า	0.1629043	-2.02222732	0.5017421	-1.9657559
เวลา	-1.4825677	0.01616402	-1.4640429	-0.2341938

จากตารางที่ 4.21 พบว่า ในแต่ละสถานการณ์ที่กำหนด สามารถคำนวณฟังก์ชันการจำแนกได้ 2 ฟังก์ชัน สามารถอธิบายได้ดังนี้

<u>กรณีที่ 1</u> เมื่อกำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง ฟังก์ชันจำแนก คือ

$$z_1 = 0.1629043(x) - 1.4825677(time)$$

$$z_2 = -2.02222732(x) + 0.01616402(time)$$

<u>กรณีที่ 2</u> เมื่อกำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ ฟังก์ชันจำแนก คือ

$$z_1 = 0.5017421(x) - 1.4640429(time)$$

$$z_2 = -1.9657559(x) - 0.2341938(time)$$

ตารางที่ 4.22 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนองเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 3

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท		
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที
ไม่มีปัญหา	52	16	8
มีปัญหาบ้าง	10	10	23
ไม่สามารถทำงานได้	0	3	3
ค่า ACC (%)	52.00		

จากตารางที่ 4.22 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 52+10+3 = 65 จากจำนวนข้อมูลทั้งสิ้น 144 ค่า ซึ่งมีค่าสูญหาย 19 ค่า และค่า ACC เท่ากับ 52.00 %

ตารางที่ 4.23 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดค่าความน่าจะเป็นก่อน มีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์เปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล กรณีศึกษาที่ 3

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท		
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที
ไม่มีปัญหา	68	8	0
มีปัญหาบ้าง	20	23	0
ไม่สามารถทำงานได้	0	6	0
ค่า ACC (%)	72.80		

จากตารางที่ 4.23 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 68+23 = 91 จากจำนวนข้อมูลทั้งสิ้น 144 ค่า ซึ่งมีค่าสูญหาย 19 ค่า และค่า ACC เท่ากับ 72.80 %

4.1.2 **การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค**

ตัวแบบของสมการโลจิทจะอยู่ในรูป

$$logit[P(Y_{itk} \le k)] = \beta_{0k} + \beta_1 x_{it} + \beta_2 (time)_{it} \quad , k = 1,...,4-1 \;; t = 1,...,2 \quad ; i = 1,...,239$$

โดยที่ $[P(Y_{itk} \le k)]$ คือ ความน่าจะเป็นที่ค่า Y_{it} จะน้อยกว่าหรือเท่ากับกลุ่มที่ k

β_{0k} คือ ค่า cut-point ของกลุ่มที่ k

β_1 , β_2 คือ สัมประสิทธิ์การถดถอย

x_{it} คือ ประเภทของผ้าที่ใช้พัน (กลุ่มทดลอง = 1 และกลุ่มควบคุม = 0)

$time_{it}$ คือ ตัวแปรเวลา (ระยะเวลาเริ่มต้น = 0, 6 เดือน = 1 และ 12 เดือน = 2)

การวิเคราะห์ข้อมูลจะศึกษาภายใต้สถานการณ์ 2 สถานการณ์ ดังนี้

1) กำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ

2) กำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม

โดยมีฟังก์ชันสำเร็จรูปที่ใช้ในการวิเคราะห์ในแต่ละสถานการณ์จำนวน 2 ฟังก์ชัน ได้แก่ ฟังก์ชัน ordgee และ repolr

4.1.2.1 การวิเคราะห์โดยใช้ฟังก์ชัน ordgee

ตารางที่ 4.24 ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน ordgee กรณีศึกษาที่ 3

ตัวแปรอิสระ	Working correlation matrix			
	Independence		Uniform	
	Estimate	Robust S.E.	Estimate	Robust S.E.
β_{01}	-1941.9685118	0	NaN	NaN
β_{02}	-1938.1028543	0	NaN	NaN
ประเภทของผ้า	0.5864554	0	NaN	NaN
เวลา	1940.4256377	0	NaN	NaN

จากตารางที่ 4.24 พบว่า ส่วนเบี่ยงเบนมาตรฐานเมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ ค่าที่คำนวณได้มีค่าเท่ากับ 0 และพบว่าค่าประมาณมีค่าสูงมาก ทำให้ไม่สามารถคำนวณหาค่าความน่าจะเป็นสะสมสำหรับค่าพยากรณ์ได้ สำหรับWorking correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม ไม่สามารถประมาณค่าพารามิเตอร์และส่วนเบี่ยงเบนมาตรฐานเลย ทำให้ ข้อมูลชุดนี้ไม่สามารถประมาณค่าพารามิเตอร์ และหาค่า ACC ได้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

4.1.2.2 การวิเคราะห์โดยใช้ฟังก์ชัน repolr

ตารางที่ 4.25 ค่าสัมประสิทธิ์ของสมการถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชัน repolr กรณีศึกษาที่ 3

ตัวแปรอิสระ	Working correlation matrix			
	Independence		Uniform	
	Estimate	Robust S.E.	Estimate	Robust S.E.
β_{01}	-2.5153596	0.6493466	-2.4931846	0.6458034
β_{02}	0.6343052	0.7697338	0.6040495	0.7624233
ประเภทของผ้า	-0.6730559	0.4440682	-0.6428807	0.4413975
เวลา	1.7172971	0.3096614	1.6991265	0.3085840

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบอิสระ โดยใช้ฟังก์ชัน repolr อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = -2.5153596 - 0.6730559\ x + 1.7172971\ (time)$$
$$logit\left[P(Y \le 2)\right] = 0.6343052 - 0.6730559\ x + 1.7172971\ (time)$$

สมการถดถอยมัลติโนเมียลโลจิสติค เมื่อกำหนดให้ Working correlation matrix มีลักษณะเป็นแบบยูนิฟอร์ม โดยใช้ฟังก์ชัน repolr อยู่ในรูป

$$logit\left[P(Y \le 1)\right] = -2.4931846 - 0.6428807\ x + 1.6991265\ (time)$$
$$logit\left[P(Y \le 2)\right] = 0.6040495 - 0.6428807\ x + 1.6991265\ (time)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.26 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบอิสระเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล สำหรับฟังก์ชัน repolr กรณีศึกษาที่ 3

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท		
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที
ไม่มีปัญหา	68	8	0
มีปัญหาบ้าง	20	23	0
ไม่สามารถทำงานได้	0	6	0
ค่า ACC (%)	72.80		

จากตารางที่ 4.26 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 68+23 = 91 จากจำนวนข้อมูลทั้งสิ้น 144 ค่า ซึ่งมีค่าสูญหาย 19 ค่า และค่า ACC เท่ากับ 72.80 %

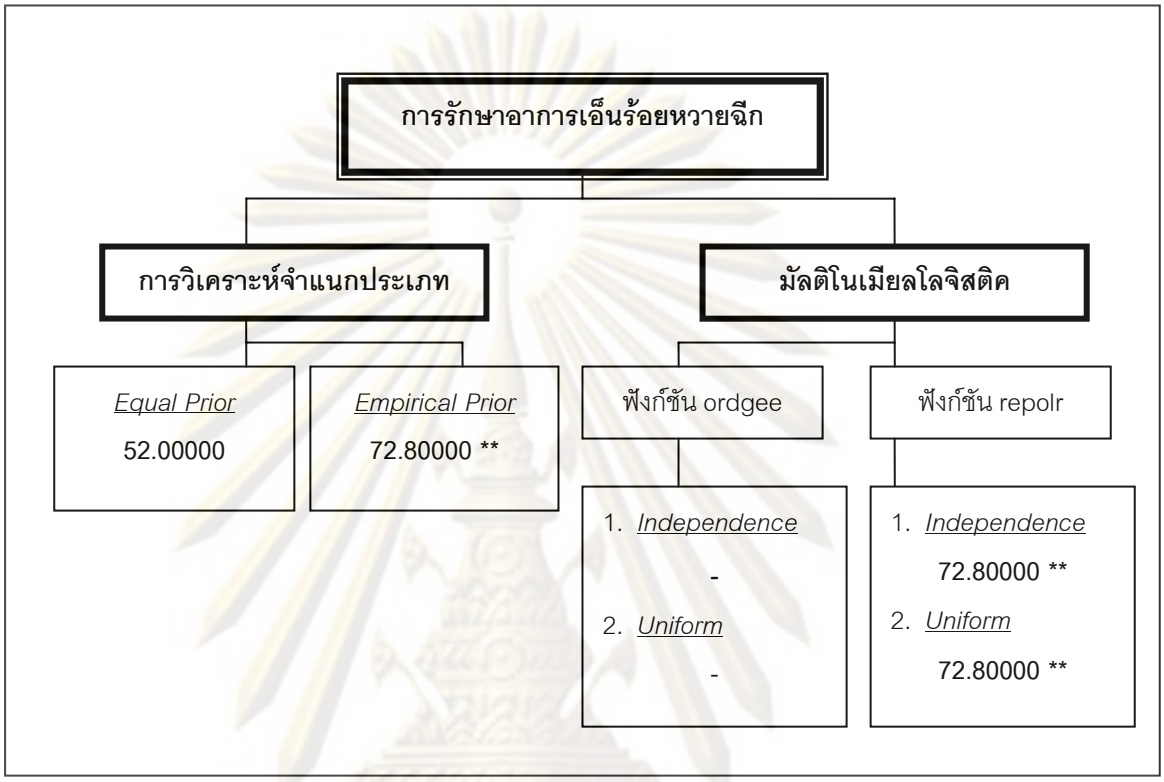
ตารางที่ 4.27 ผลการจำแนกกลุ่ม ในกรณีที่กำหนดให้ Working correlation matrix เป็นแบบยูนิฟอร์มเปรียบเทียบกับกลุ่มที่แท้จริงของข้อมูล สำหรับฟังก์ชัน repolr กรณีศึกษาที่ 3

กลุ่มจริง	การจัดกลุ่มใช้ฟังก์ชันจำแนกประเภท		
	น้อยกว่า 20 นาที	20 - 30 นาที	30 – 60 นาที
ไม่มีปัญหา	68	8	0
มีปัญหาบ้าง	20	23	0
ไม่สามารถทำงานได้	0	6	0
ค่า ACC (%)	72.80		

จากตารางที่ 4.27 ผลการเปรียบเทียบการจำแนกกลุ่มพบว่า จำนวนข้อมูลที่จัดกลุ่มได้ถูกต้องจำนวน 68+23 = 91 จากจำนวนข้อมูลทั้งสิ้น 144 ค่า ซึ่งมีค่าสูญหาย 19 ค่า และค่า ACC เท่ากับ 72.80 %

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูปภาพที่ 4.3 แผนผังแสดงค่า ACC ที่ได้จากการคำนวณ กรณีศึกษาที่ 3

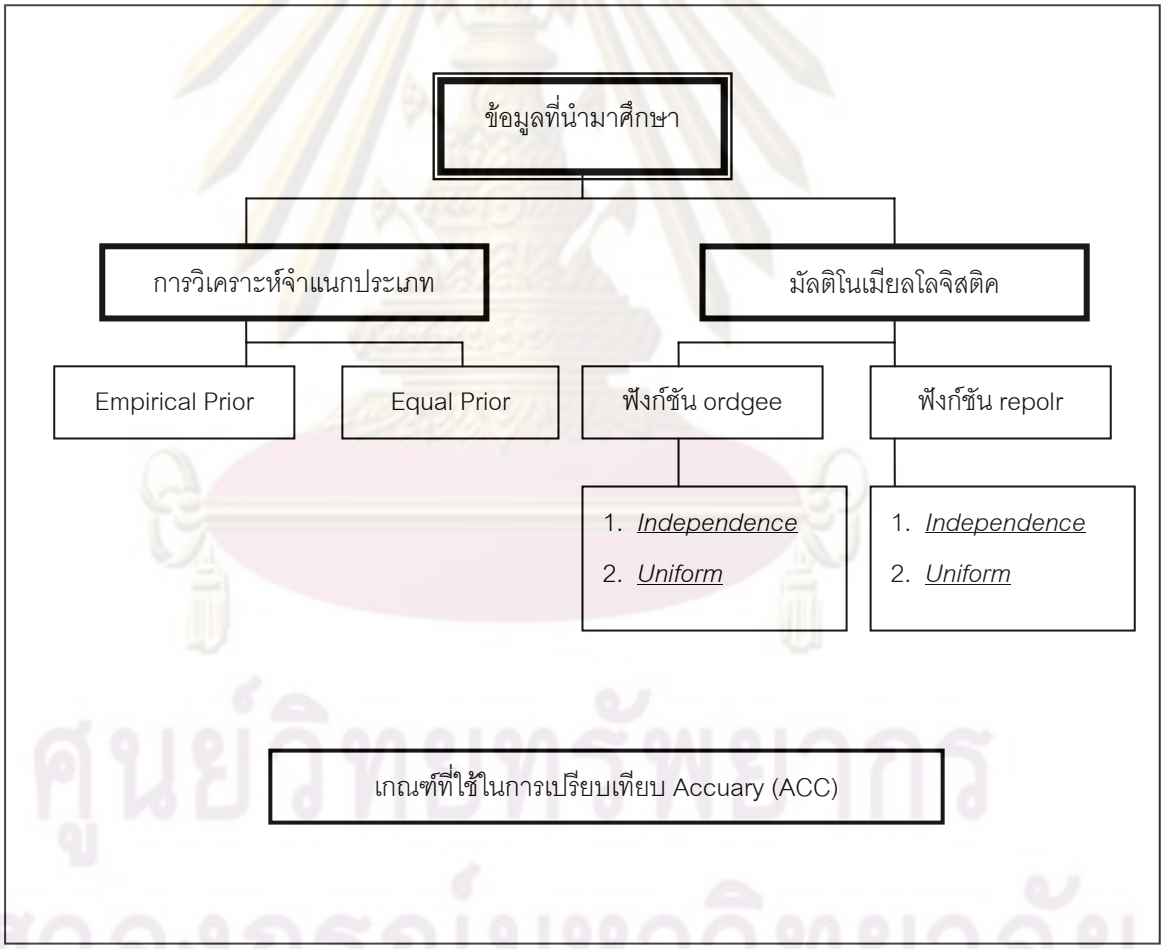


ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบการวิเคราะห์ข้อมูลระยะยาว ซึ่งในการวิเคราะห์จะใช้วิธีการวิเคราะห์ข้อมูล 2 แบบ คือ การวิเคราะห์จำแนกประเภท และการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค โดยใช้ข้อมูลจริงทางด้านการแพทย์ 3 ชุด และทำการศึกษาในแต่ละสถานการณ์ที่กำหนด สำหรับงานวิจัยครั้งนี้ ได้ทำการศึกษาภายใต้สถานการณ์ที่กำหนดขึ้นดังนี้

รูปที่ 5.1 แสดงโครงสร้างของสถานการณ์ที่ทำการวิเคราะห์ สำหรับข้อมูล 1 ชุด



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

5.1 สรุปผลการวิจัย

5.1.1 กรณีศึกษาที่ 1: การติดตามผลของโรคนอนไม่หลับ

1. จากการเปรียบเทียบค่า ACC ในแต่ละสถานการณ์ที่ทำการศึกษา สามารถสรุปผลการวิเคราะห์ได้ ดังนี้

- วิธีวิเคราะห์จำแนกประเภท

ทำการเปรียบเทียบโดยกำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับค่าความน่าจะเป็นเชิงประจักษ์ ผลการวิเคราะห์จะมีค่าความถูกต้องในการจัดกลุ่มสูงกว่าในกรณีที่กำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง

การวิเคราะห์ข้อมูลจำแนกประเภท ควรใช้ข้อมูลที่วิเคราะห์โดยกำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับค่าความน่าจะเป็นเชิงประจักษ์ จะมีความเหมาะสมที่สุด

- วิธีวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

สามารถแบ่งเป็น 2 สถานการณ์ย่อย ดังนี้

ก) กำหนดให้รูปแบบโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบอิสระ

เมื่อความสัมพันธ์ของข้อมูลเป็นแบบอิสระ ใช้ฟังก์ชัน repolr และ ordgee ในการวิเคราะห์ข้อมูล พบว่า ทั้ง 2 ฟังก์ชัน ค่าความถูกต้องในการจัดกลุ่มเท่ากันทุกตัวแบบที่ทำการวิเคราะห์

ข) กำหนดให้รูปแบบโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบยูนิฟอร์ม

เมื่อความสัมพันธ์ของข้อมูลเป็นแบยูนิฟอร์ม ใช้ฟังก์ชัน repolr และ ordgee ในการวิเคราะห์ข้อมูล พบว่า ทั้ง 2 ฟังก์ชัน มีค่าความถูกต้องในการจัดกลุ่มเท่ากันแตกต่างกัน ซึ่งฟังก์ชัน ordgee สามารถคำนวณค่าความถูกต้องในการจัดกลุ่มสูงกว่าฟังก์ชัน repolr

เมื่อพิจารณาภาพรวมของฟังก์ชันสำเร็จรูปทั้ง 2 ฟังก์ชัน พบว่า ถึงแม้ว่า ฟังก์ชัน ordgee ให้ค่าความถูกต้องในการจัดกลุ่มมากกว่า ฟังก์ชัน repolr แต่เมื่อพิจารณาถึงค่าความคลาดเคลื่อนมาตรฐานที่ได้จากการประมาณค่าพารามิเตอร์ พบว่า ฟังก์ชัน ordgee ให้ค่าความคลาดเคลื่อนมาตรฐานสูงกว่าฟังก์ชัน repolr ค่อนข้างมาก ซึ่งในการวิเคราะห์ข้อมูลจริงควรสังเกตสถานการณ์ดังกล่าวนี้ไว้ เนื่องจากจะส่งผลต่อการจัดกลุ่มได้

- <u>การเปรียบเทียบการวิเคราะห์จำแนกประเภทและการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค</u>

กรณีศึกษาที่ 1: การติดตามผลของโรคนอนไม่หลับ วิธีการวิเคราะห์จำแนกประเภท เมื่อกำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับค่าความน่าจะเป็นเชิงประจักษ์ พบว่าค่าความถูกต้องในการจัดกลุ่มมีค่าเท่ากับ 36.82 % และการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค โดยใช้เทคนิค GEE โดยใช้ฟังก์ชันสำเร็จรูป ordgee และโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบยูนิฟอร์ม ค่าความถูกต้องในการจัดกลุ่มมีค่าเท่ากับ 37.235 % ซึ่งมีค่ามากกว่าการวิเคราะห์จำแนกประเภท

สรุปได้ว่า กรณีศึกษาที่ 1: การติดตามผลของโรคนอนไม่หลับ วิธีการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ใช้เทคนิค GEE โดยใช้ฟังก์ชันสำเร็จรูป ordgee และโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบยูนิฟอร์ม จะมีค่าความถูกต้องในการจัดกลุ่มสูงสุด

5.1.2 **กรณีศึกษาที่ 2: การผ่าตัดเปลี่ยนผิวข้อสะโพก**

1. จากการเปรียบเทียบค่า ACC ในแต่ละสถานการณ์ที่ทำการศึกษา สามารถสรุปผลการวิเคราะห์ได้ ดังนี้

- <u>วิธีวิเคราะห์จำแนกประเภท</u>

เปรียบเทียบโดยกำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับค่าความน่าจะเป็นเชิงประจักษ์ ผลการวิเคราะห์จะมีค่าความถูกต้องในการจัดกลุ่มสูงกว่าเมื่อค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกตัวแปรตอบสนอง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การวิเคราะห์ข้อมูลจำแนกประเภทควรใช้ข้อมูลที่วิเคราะห์โดยกำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับค่าความน่าจะเป็นเชิงประจักษ์ จะมีความเหมาะสมที่สุด

- <u>วิธีวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค</u>

สามารถแบ่งเป็น 2 สถานการณ์ย่อย ดังนี้

ก) กำหนดให้รูปแบบโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบอิสระ

เมื่อความสัมพันธ์ของข้อมูลเป็นแบบอิสระ ใช้ฟังก์ชัน repolr และ ordgee ในการวิเคราะห์ข้อมูล พบว่า ทั้ง 2 ฟังก์ชัน ค่าความถูกต้องในการจัดกลุ่มเท่ากันทุกตัวแบบที่ทำการวิเคราะห์

ข) กำหนดให้รูปแบบโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบยูนิฟอร์ม

เมื่อความสัมพันธ์ของข้อมูลเป็นแบบยูนิฟอร์ม ใช้ฟังก์ชัน repolr และ ordgee ในการวิเคราะห์ข้อมูล พบว่า ทั้ง 2 ฟังก์ชัน มีค่าความถูกต้องในการจัดกลุ่มเท่ากันแตกต่างกัน โดยฟังก์ชัน repolr สามารถคำนวณค่าความถูกต้องในการจัดกลุ่มได้ทั้ง 2 รูปแบบความสัมพันธ์ แต่ฟังก์ชัน ordgee ไม่สามารถคำนวณค่าความถูกต้องในการจัดกลุ่มเมื่อกำหนดรูปแบบของความสัมพันธ์เป็นแบบยูนิฟอร์ม แต่ในตัวแบบที่ 1 ให้ค่าความถูกต้องในการจัดกลุ่ม เท่ากับฟังก์ชัน repolr

ดังนั้น การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติคสำหรับกรณีศึกษาที่ 2 ควรใช้ฟังก์ชัน repolr จะมีความเหมาะสมมากกว่า ฟังก์ชัน ordgee เมื่อพิจารณาถึงรูปแบบความสัมพันธ์เมื่อใช้ฟังก์ชัน repolr ทั้ง 2 ความสัมพันธ์ จะให้ค่าความถูกต้องในการจัดกลุ่มเท่ากันและความแตกต่างของความคลาดเคลื่อนมาตรฐานแตกต่างกันน้อยมาก ทำให้กรณีศึกษาที่ 2: การผ่าตัดเปลี่ยนผิวข้อสะโพก ควรใช้ฟังก์ชัน repolr และรูปแบบของความคลาดเคลื่อนสามารถเป็นได้ทั้งแบบอิสระและแบบยูนิฟอร์ม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

- <u>การเปรียบเทียบการวิเคราะห์จำแนกประเภทและการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค</u>

จากกรณีศึกษาที่ 2 เมื่อทำการเปรียบเทียบ ทั้ง 2 วิธีการวิเคราะห์ พบว่า วิธีการวิเคราะห์จำแนกประเภท โดยกำหนดให้ความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ ให้ค่าความถูกต้องในการจัดกลุ่มเท่ากับการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ซึ่งมีค่าเท่ากับ 60.92 % ดังนั้น ข้อมูลชุดนี้ควรเลือกใช้ การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชันสำเร็จรูป repolr และรูปแบบความสัมพันธ์ของข้อมูลสามารถเลือกได้ทั้งเป็นแบบอิสระและยูนิฟอร์ม เนื่องจาก การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค เป็นการวิเคราะห์ที่ศึกษาถึงความสัมพันธ์ของความคลาดเคลื่อนของข้อมูลในแต่ละช่วงเวลาโดยมีรูปแบบต่าง ๆ และข้อมูลที่ทำการศึกษาไม่เป็นไปตามข้อกำหนดในส่วนของการแจกแจงปกติหลายตัวแปรของสำหรับวิธีการวิเคราะห์จำแนกประเภท

5.1.3 **กรณีศึกษาที่** 3: **การฉีกของเอ็นร้อยหวาย**

1. จากการเปรียบเทียบค่า ACC ในแต่ละสถานการณ์ที่ทำการศึกษา สามารถสรุปผลการวิเคราะห์ได้ ดังนี้

- <u>วิธีวิเคราะห์จำแนกประเภท</u>

ทำการเปรียบเทียบโดยกำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับค่าความน่าจะเป็นเชิงประจักษ์ ผลการวิเคราะห์จะมีค่าความถูกต้องในการจัดกลุ่ม สูงกว่าเมื่อกำหนดให้ค่าความน่าจะเป็นก่อนมีค่าเท่ากันทุกตัวแปรตอบสนอง และเมื่อทำการทดสอบความแตกต่างในการจัดกลุ่มของทั้งสองวิธี พบว่ามีนัยสำคัญของความแตกต่าง

การวิเคราะห์ข้อมูลจำแนกประเภท ควรใช้ข้อมูลที่วิเคราะห์โดยกำหนดใช้ค่าความน่าจะเป็นก่อนมีค่าเท่ากับค่าความน่าจะเป็นเชิงประจักษ์ จะมีความเหมาะสมที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

- วิธีวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

วิธีวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค สามารถแบ่งเป็น 2 สถานการณ์ย่อย ดังนี้

ก) กำหนดให้รูปแบบโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบอิสระ

เมื่อความสัมพันธ์ของความคลาดเคลื่อนเป็นแบอิสระ ใช้ฟังก์ชัน repolr และ ordgee ในการวิเคราะห์ข้อมูล พบว่า ทั้ง 2 ฟังก์ชัน มีค่าความถูกต้องในการจัดกลุ่มเท่ากัน แตกต่างกัน โดยฟังก์ชัน repolr สามารถคำนวณค่าความถูกต้องในการจัดกลุ่มได้ แต่ฟังก์ชัน ordgee ทุกตัวแบบไม่สามารถคำนวณค่าความถูกต้องในการจัดกลุ่ม

ข) กำหนดให้รูปแบบโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบยูนิฟอร์ม

เมื่อความสัมพันธ์ของความคลาดเคลื่อนเป็นยูนิฟอร์ม ใช้ฟังก์ชัน repolr และ ordgee ในการวิเคราะห์ข้อมูล พบว่า ทั้ง 2 ฟังก์ชัน มีค่าความถูกต้องในการจัดกลุ่มเท่ากัน แตกต่างกัน โดยฟังก์ชัน repolr สามารถคำนวณค่าความถูกต้องในการจัดกลุ่มได้ทุกตัวแบบ แต่ฟังก์ชัน ordgee ทุกรูปแบบความสัมพันธ์ของข้อมูลไม่สามารถคำนวณค่าความถูกต้องได้ในการจัดกลุ่ม

การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ควรใช้ฟังก์ชัน repolr จะมีความเหมาะสมมากกว่า ฟังก์ชัน ordgee และเมื่อพิจารณาถึงรูปแบบความสัมพันธ์ของข้อมูลโดยใช้ฟังก์ชัน repolr ทั้ง 2 ความสัมพันธ์ จะให้ค่าความถูกต้องในการจัดกลุ่มเท่ากัน และความแตกต่างของความคลาดเคลื่อนมาตรฐานแตกต่างกันน้อยมาก ทำให้กรณีศึกษาที่ 3: การฉีกของเอ็นร้อยหวาย ควรใช้ฟังก์ชัน repolr และรูปแบบของความคลาดเคลื่อนสามารถเป็นได้ทั้งแบบอิสระและแบบยูนิฟอร์ม

- การเปรียบเทียบการวิเคราะห์จำแนกประเภทและการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

จากกรณีศึกษาที่ 3 เมื่อทำการเปรียบเทียบ ทั้ง 2 วิธีการวิเคราะห์ พบว่า วิธีการวิเคราะห์จำแนกประเภท เมื่อกำหนดให้ความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ ให้ค่าความถูกต้องในการจัดกลุ่มเท่ากับการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ซึ่งมีค่าเท่ากับ 72.8 % ดังนั้น ข้อมูลชุดนี้ควรเลือกใช้ การวิเคราะห์การถดถอยมัลติโนเมียลโล

จิสติค ใช้ฟังก์ชันสำเร็จรูป repolr ในส่วนของรูปแบบความสัมพันธ์ของข้อมูลสามารถเลือกได้ทั้งเป็นแบบอิสระและยูนิฟอร์ม เนื่องจาก การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค เป็นการวิเคราะห์ที่ศึกษาถึงความสัมพันธ์ของความคลาดเคลื่อนของข้อมูลในแต่ละช่วงเวลาโดยมีรูปแบบต่าง ๆ และข้อมูลที่ทำการศึกษาไม่เป็นไปตามข้อกำหนดในส่วนของการแจกแจงปกติหลายตัวแปรของสำหรับวิธีการวิเคราะห์จำแนกประเภท

5.2 อภิปรายผลการวิจัย

จากผลการวิเคราะห์ข้อมูลสำหรับกรณีศึกษาทั้ง 3 กรณี พบว่า

1. การวิเคราะห์จำแนกประเภท สำหรับกรณีที่มีการใช้ค่าความน่าจะเป็นเชิงประจักษ์มาถ่วงน้ำหนัก สามารถให้ผลลัพธ์ในการจัดกลุ่มแม่นยำเทียบเท่าการวิเคราะห์แบบมัลติโนเมียลโลจิสติค ที่ใช้วิธี Generalized Estimating Equation ซึ่งสอดคล้องกับบทความที่ได้กล่าวว่า การวิเคราะห์จำแนกประเภทสามารถให้ผลลัพธ์ในการจัดกลุ่มค่อนข้างดีและมีความแม่นยำได้ ถ้ามีขั้นตอนและกระบวนการที่เหมาะสม [1]

2. ข้อสังเกตจากข้อมูล 3 ชุดที่นำมาวิเคราะห์ พบว่า ถ้าพิจารณาถึงวิธีที่ดีที่สุดที่ใช้ในการวิเคราะห์ข้อมูลในแต่ละแบบ สังเกตได้ว่า เมื่อค่าความน่าจะเป็นที่แท้จริงของแต่ละกลุ่มมีค่าใกล้เคียงกัน ดังเช่น ในข้อมูลตัวอย่างของการติดตามผลของโรคนอนไม่หลับ การวิเคราะห์จำแนกประเภทให้ค่าความถูกต้องในการจัดกลุ่ม (ACC) ได้ดีกว่าการวิเคราะห์แบบมัลติโนเมียลโลจิสติค แต่ถ้าความน่าจะเป็นเชิงประจักษ์ ดังชุดข้อมูลการผ้าตัดเปลี่ยนผิวข้อสะโพกและการรักษาอาการเอ็นร้อยหวายฉีก การวิเคราะห์จำแนกประเภทให้ผลการจัดกลุ่มไม่แตกต่างจากการวิเคราะห์แบบมัลติโนเมียลโลจิสติค จากกรณีศึกษาทั้ง 3 กรณี แสดงให้เห็นว่าการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติคจะมีประสิทธิภาพลดลงเมื่อความน่าจะเป็นเชิงประจักษ์ในแต่ละกลุ่มใกล้เคียงกัน

3. ถ้าความแตกต่างระหว่างความน่าจะเชิงประจักษ์ของข้อมูลกับค่าความน่าจะเป็นก่อนที่กำหนดให้ มีความแตกต่างกันมากจะทำให้ความแม่นยำในการจัดกลุ่มจะลดลง

---

[1] Dunn O.J and Lachenbruch P.A. “Discriminant Analysis and Clustering: Panel on Discriminant Analysis, Classification, and Clustering,” Statistical Science, 4 (1989), pp.34-69.

4. การวิเคราะห์มัลติโนเมียลโลจิสติค เมื่อใช้ฟังก์ชัน ordgee ในการประมาณค่าพารามิเตอร์ จุดอ่อนของวิธีการนี้ อยู่ที่การวิเคราะห์โดยใช้ฟังก์ชัน ordgee ไม่สามารถประมาณค่าพารามิเตอร์ในบางสถานการณ์ได้สำหรับข้อมูลระยะยาว ดังเช่นข้อมูลการฉีกของเอ็นร้อยหวายในส่วนของตัวอย่างข้อมูลโรคนอนไม่หลับ ค่าพารามิเตอร์ที่ประมาณรวมทั้งส่วนเบี่ยงเบนมาตรฐานที่ได้จากการประมาณค่ามีความแตกต่างจากการใช้ฟังก์ชันสำเร็จรูป repolr และในโปรแกรมวิเคราะห์ทางสถิติอื่น ๆ ด้วย สอดคล้องกับข้อสังเกตของ Laura [1]

5. การวิเคราะห์มัลติโนเมียลโลจิสติค เมื่อพิจาณาถึงความสัมพันธ์ของข้อมูลในการวิเคราะห์ข้อมูล ถึงแม้ว่าจะไม่เห็นถึงความแตกต่างในแต่ละลักษณะของความสัมพันธ์ จากการใช้เกณฑ์ค่าความถูกต้องในการจัดกลุ่ม แสดงให้เห็นว่า ข้อมูลดังกล่าวอาจไม่มีความสัมพันธ์กันในแต่ละช่วงเวลาที่ทำการศึกษา

6. จากการวิเคราะห์ข้อมูลทั้ง 3 ชุด ผู้วิจัยเลือกการวิเคราะห์แบบมัลติโนเมียลโลจิสติค ใช้เทคนิคสมการประมาณค่าโดยนัยทั่วไป (GEE) เนื่องจากข้อกำหนดของวิธีการจำแนกประเภทไม่เป็นไปตามข้อตกลงเบื้องต้นในส่วนของการแจกแจงแบบปกติหลายตัวแปร และให้วิธีการวิเคราะห์จำแนกประเภทในแต่ละช่วงเวลาเป็นค่าสังเกต 1 ค่า นั่นคือ ในแต่ละช่วงเวลามีความเป็นอิสระกันภายในค่าสังเกตเดียวกันที่มีการวัดซ้ำ ซึ่งในความเป็นจริง ข้อกำหนดนี้อาจจะไม่เหมาะสมกับข้อมูลที่นำมาวิเคราะห์ และเทคนิคสมการประมาณค่าโดยนัยทั่วไป สามารถเลือกรูปแบบความสัมพันธ์ของความคลาดเคลื่อนที่มีความเหมาะสมที่สุดของข้อมูลแต่ละประเภทได้

7. สรุปวิธีที่ใช้ในการวิเคราะห์ข้อมูลระยะยาวสำหรับกรณีศึกษาทั้ง 3 กรณี มีดังนี้

7.1) กรณีศึกษาที่ 1: การติดตามผลของโรคนอนไม่หลับ เลือกวิธีการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค ใช้เทคนิค GEE โดยใช้ฟังก์ชันสำเร็จรูป ordgee และโครงสร้างความสัมพันธ์ของข้อมูลเป็นแบบ Uniform

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

[1] Laura A. Thompson. R (and S-plus) Manual to Accompany Agresti's Categorical Data Analysis (2008): 221

7.2) กรณีศึกษาที่ 2: การผ่าตัดเปลี่ยนผิวข้อสะโพก เลือกวิธีการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชันสำเร็จรูป repolr แต่รูปแบบความสัมพันธ์ของข้อมูลสามารถเลือกได้ทั้งเป็นแบบอิสระและยูนิฟอร์ม

7.3) กรณีศึกษาที่ 3: การฉีกของเอ็นร้อยหวาย เลือกวิธีการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค โดยใช้ฟังก์ชันสำเร็จรูป repolr แต่รูปแบบความสัมพันธ์ของข้อมูลสามารถเลือกได้ทั้งเป็นแบบอิสระและยูนิฟอร์ม

5.3 ข้อเสนอแนะ

จากผลการวิจัยครั้งนี้ ผู้วิจัย มีข้อเสนอแนะดังนี้

1. ควรตรวจสอบข้อกำหนดของวิธีการวิเคราะห์ข้อมูลในแต่ละข้อมูล
2. การเลือกใช้ฟังก์ชันสำเร็จรูปในการวิเคราะห์ข้อมูล ควรตรวจสอบถึงลักษณะในการเขียนฟังก์ชัน ว่ามีความถูกต้องหรือไม่ และข้อจำกัดที่ฟังก์ชันไม่สามารถทำงานได้
3. ควรศึกษาวิธีการประมาณค่าวิธีอื่นๆ เนื่องจากการวิเคราะห์ข้อมูลระยะยาวยังมีเทคนิคอีกในการประมาณค่าพารามิเตอร์อีกหลายรูปแบบ เช่น การใช้อิทธิพลสุ่ม (Random effect) Transition และ Mixed Model เป็นต้น
4. ควรศึกษาในกรณีที่ใช้ระยะเวลาในการเก็บรวบรวมข้อมูลมากกว่านี้
5. ควรศึกษาในกรณีที่ตัวแปรกำหนดเป็นตัวแปรเชิงปริมาณ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา . การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 1. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2548.

เครือข่ายจัดการข้อมูลและชีวสถิติ. สถิติ GEE.[ออนไลน์] แหล่งที่มา: http://www.dmbn.net/. [มิถุนายน 2552]

ดุษฎี โยเหลา. โมเดลการเปลี่ยนแปลงภายในบุคคลและสาเหตุของการเปลี่ยนแปลง. เอกสารประกอบการอบรมเรื่อง โมเดลสมการโครงสร้างเชิงเส้นโดยใช้โปรแกรม LISREL สำหรับข้อมูลช่วงยาว (พฤษภาคม 2551) : 30-36.

วีระศักดิ์ จงสู่วิวัศน์วงศ์ . กราฟ ตาราง และสมการ สำหรับการวิจัยทางสุขภาพ. พิมพ์ครั้งที่ 1. กรุงเทพฯ : สำนักพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย, 2550.

ศิริชัย กาญจนวาสี. การวิเคราะห์พหุระดับ MULTI-LEVEL ANALYSIS. กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2550.

สุชาต ประสิทธิ์รัฐสินธุ์ . เทคนิคการวิเคราะห์ตัวแปรหลายตัวสำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์ . พิมพ์ครั้งที่ 6. กรุงเทพฯ : สำนักพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย, 2551.

สุพล ดุรงค์วัฒนา. ตัวแบบและการวิเคราะห์ความถดถอยสำหรับการวิจัยขั้นสูง. กรุงเทพฯ : สำนักพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย, 2549.

ภาษาอังกฤษ

Agresti A. Categorical Data Analysis. 2nd ed. New york: Wiley, 2002.

Box, G.E.P. A general Distribution Theory for a class of likelihood Criteria. Biometrica Vol. 36 (1949) : 317-346.

Craiu R.V., Duchesne T. and Fortin D. Inference methods for the conditional logistic regression model with longitudinal data. Biometrical journal 50 (2008) : 97-109.

David C. Repeated ordinal measurements: a generalised estimating equation approach. (April 1992)

Don Hedeker. BSTT537 Longitudinal Data Analysis - Fall 2009.[online] Available from: http://www.uic.edu/classes/bstt/bstt513/. [2009, March 2]

Dunn O.J and Lachenbruch P.A. Discriminant Analysis and Clustering: Panel on Discriminant Analysis, Classification, and Clustering. Statistical Science 4 (1989) : 34-69.

Hans-George M. Functional Modelling and Classification of Longitudinal Data. Scandinavian Journal of Statistics Vol.32 (2005) : 223-240.

James Press S. and Sandra W. Choosing Between Logistic Regression and Discriminant Analysis. Journal of the American Statistics Association Vol. 73 (Dec., 1978) : 699-705.

Joao Gama. A Linear-Bayes Classifier.[online] Available from: http://www.springerlink.com/content/bwjwc1wl3k67amx4/ [2009, August 17].

Laura A. Thompson . R (and S-plus) Manual to Accompany Agresti's Categorical Data Analysis (2002), 2009.

Maja P., Mateja B. and Sandra T . Comparion of Logistic Regression and Linear Discrimiant Analysis: A Simulation Study. Metodoloski zevezki Vol. 1 No.1 (2004) : 143-161.

Mardina, K.V. Measure of multivariate skewness and kurtosis with applications. Biometrica Vol. 57 (1970) : 513-530.

Marshall G. Linear discriminant models for unbalanced longitudinal data. Statistics in Medicine 19 (2000) : 1969-1981.

McCullagh P. and Nelder J.A. Generalized Linear Models. 2nd ed. London : Chapman and Hall, 1989.

Paone G., De Angelis G. and Munno R. Discriminant analysis on small cell lung cancer and non-small cell lung cancer by means of NSE and CYFRA-21.1. European Respiratory Journal 8 (1995) : 1136-1140.

Parsons N.R., Matthew L. Costa, Juul Achten and Nigel Stallard. Repeated measure proportional odds logistic regression analysis of ordinal score data in the statistical software package R. Journal of Science Direct 53 (2008) : 632-641.

Peter J.D. and Patrick J.H. Analysis of Longitudinal Data. 2nd ed. New York : Oxford, 2005.

PRISM. Classification.[online] Available from: http://research.cs.tamu.edu/prism/lectures/iss/iss_l12.pdf. [2009, August 17]

Yen-Peng Li and Wenyaw Chan. Analysis of Longitudinal Multinomial Outcome Data. Biometrical Journal Vol.48 (2006) : 315-326.

University of Florida Department of Statistics.[online] SAS for GEE and Random Intercept Cumulative Logit Analyses of Insomnia data of Table 11.4. Available from: http://www.stat.ufl.edu/~aa/cda/sas/sas.html.[2009, January 31]

Warwick Medical School.[online] Repeated measures proportional odds logistic regression . [online] Available from: http://www.warwick.ac.uk/go/repolr. [2009, April]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก
ลักษณะโดยทั่วไปของข้อมูลที่ทำการศึกษา

เนื่องจากข้อมูลที่นำมาทำการศึกษาเป็นข้อมูลภายใต้สถานการณ์จริงทางด้านการแพทย์ ที่มีการวัดข้อมูลซ้ำ โดยมีข้อมูลที่นำมาทำการศึกษาทั้งสิ้น 3 ชุด ดังนี้

กรณีศึกษาที่ 1 การติดตามผลการรักษาโรคนอนไม่หลับ[1]

ข้อมูลชุดนี้เป็นการศึกษาทดลองทางการแพทย์ โดยทำการศึกษาการเปรียบเทียบระหว่างยาที่ใช้ในการรักษาโรคนอนไม่หลับและยาหลอก โดยหน่วยตัวอย่าง คือ ผู้ป่วยที่มีอาการนอนไม่หลับจำนวน 239 คน มีตัวแปรตอบสนอง คือ ข้อมูลที่บันทึกระยะเวลาที่ผู้ป่วยใช้ในการนอนหลับ ซึ่งลักษณะของตัวแปรตอบสนองนี้แบ่งออกเป็น 4 กลุ่ม ตามระยะเวลาที่ใช้ในการนอนหลับ (น้อยกว่า 20 นาที, ระหว่าง 20-30 นาที, 30-60 นาที และมากกว่า 60 นาทีขึ้นไป) โดยมีตัวแปรอิสระ 2 ตัวแปร ได้แก่ ช่วงระยะเวลาที่ใช้เก็บข้อมูล ซึ่งผู้ป่วยทุกคนจะต้องมาทำการตรวจสอบ 2 ครั้ง (ระยะเริ่มต้น, ระยะติดตามผล) และประเภทของยาที่ใช้ในการทดลอง (ยาจริง, ยาหลอก) ที่ผู้ป่วยได้รับ[2]

ตาราง ก.1 ตารางแจกแจงความถี่ของตัวแปรตอบสนองของกรณีศึกษาที่ 1 จำแนกตามระยะเวลาและประเภทของยาที่ใช้ในการรักษา

ระยะเวลาที่ใช้ในการหลับ	ระยะเวลาเริ่มต้น (t=0)		ระยะเวลาติดตามผล(t=1)	
	ยาหลอก(0)	ยาจริง(1)	ยาหลอก(0)	ยาจริง(1)
น้อยกว่า 20 นาที (1)	14	12	31	40
20 - 30 นาที (2)	20	20	29	49
30 – 60 นาที (3)	35	40	35	19
มากกว่า 60 นาที (4)	51	47	25	11
รวม	120	119	120	119

[1] แหล่งที่มา : http://www.stat.ufl.edu/~aa/cda/sas/sas.html [31/01/2552]

[2] Agresti Alan, Categorical Data Analysis, 2nd ed. (New York: Wiley, 2002), p.462-463

ตาราง ก.2 ความถี่และความน่าจะเป็นเชิงประจักษ์สำหรับตัวแปรตอบสนองในแต่ละกลุ่ม

ระยะเวลา ที่ใช้ในการหลับ	ระยะเวลา เริ่มต้น (t=0)	ระยะเวลา ติดตามผล(t=1)	รวม
น้อยกว่า 20 นาที	26	71	97 (0.20)
20 - 30 นาที	40	78	118 (0.25)
30 – 60 นาที	75	54	129 (0.27)
มากกว่า 60 นาที	98	36	134 (0.28)

(.) คือ ค่าความน่าจะเป็นเชิงประจักษ์

กรณีศึกษาที่ 2 การผ่าตัดเปลี่ยนผิวข้อสะโพก [1]

ข้อมูลชุดนี้เป็นการศึกษาทดลองทางการแพทย์ เป็นข้อมูลเกี่ยวกับการผ่าตัดเปลี่ยนผิวข้อสะโพก (Hip resurfacing) ตัวแปรตอบสนอง คือ ระดับความเจ็บปวดภายหลังจากการผ่าตัด ซึ่งมี 4 ระดับความเจ็บปวด (none, slight, mild, moderate) โดยมีหน่วยทดลอง คือ ผู้ป่วยที่เข้ารับการรักษาและทำการตอดตามผลการรักษา จำนวนผู้ทดลอง 58 คน ในการศึกษาครั้งนี้มีตัวแปรอิสระ 2 ตัวแปร ได้แก่ ช่วงระยะเวลาที่ใช้เก็บข้อมูล ซึ่งผู้ป่วยทุกคนจะต้องมาทำการตรวจสอบ 3 ครั้ง (ระยะเวลาเริ่มต้น, 2 ปี และ 5 ปี) และเพศ (ชาย, หญิง)[2]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

[1] แหล่งที่มา : http://www.warwick.ac.uk/go/repolr. [20/04/2552]

[2] Parsons N.R., Matthew L. Costa. “Repeated measure proportional odds logistic regression analysis of ordinal score data in the statistical software package R”, Journal of Science Direct 53 (2008): 632-641

ตาราง ก.3 ตารางแจกแจงความถี่ของตัวแปรตอบสนองของกรณีศึกษาที่ 2 จำแนกตามระยะเวลาและเพศ

อาการปวด	ระยะเวลาเริ่มต้น (t=0)		2 ปี (t=2)		5 ปี (t=5)	
	ชาย (1)	หญิง (0)	ชาย (1)	หญิง (0)	ชาย (1)	หญิง (0)
None (1)	25	13	28	10	20	10
Slight (2)	6	5	5	7	10	3
Mild (3)	6	3	3	3	3	4
Moderate (4)	0	0	1	1	4	4
รวม	37	21	37	21	37	21

ตาราง ก.4 ความถี่และความน่าจะเป็นเชิงประจักษ์สำหรับตัวแปรตอบสนองในแต่ละกลุ่ม

อาการปวด	ระยะเวลาเริ่มต้น (t=0)	2 ปี (t=2)	5 ปี (t=5)	รวม
None	38	38	30	106 (0.61)
Slight	11	12	13	36 (0.21)
Mild	9	6	7	22 (0.13)
Moderate	0	2	8	10 (0.06)

(.) คือ ค่าความน่าจะเป็นเชิงประจักษ์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

กรณีศึกษาที่ 3 การรักษาอาการเอ็นร้อยหวายฉีก [1]

ข้อมูลชุดนี้เป็นการศึกษาทดลองทางการแพทย์ เป็นข้อมูลเกี่ยวกับการรักษาอาการเอ็นร้อยหวายฉีก (Achilles tendon rupture) ตัวแปรตอบสนอง คือ ปัญหาจากการรักษา ซึ่งมี 3 ระดับของการเกิดปัญหาในการทำงาน (ไม่มีปัญหา, มีปัญหาบ้าง, ไม่สามารถทำงานได้) โดยมีหน่วยทดลอง คือ ผู้ป่วยที่เข้ารับการรักษาและทำการติดตามผลการรักษา จำนวนผู้ทดลอง 48 คน ในการศึกษาครั้งนี้มีตัวแปรอิสระ 2 ตัวแปร ได้แก่ ช่วงระยะเวลาที่ใช้เก็บข้อมูล ซึ่งผู้ป่วยทุกคนจะต้องมาทำการตรวจสอบ 3 ครั้ง (ระยะเวลาเริ่มต้น, 6 เดือน และ 12 เดือน) และลักษณะของผ้าที่ใช้พันแบ่งเป็น 2 กลุ่ม (กลุ่มทดลอง ใช้เส้นใยคาร์บอนมีขนาด 1.5 เซนติเมตร และกลุ่มควบคุม คือ ผู้ที่ใช้เส้นใยธรรมดา) [2]

ตาราง ก.5 ตารางแจกแจงความถี่ของตัวแปรตอบสนองของกรณีศึกษาที่ 2 จำแนกตามระยะเวลาและประเภทของผ้าที่ใช้พัน

ปัญหาในการทำงาน	ระยะเวลาเริ่มต้น (t=0)		6 เดือน (t=1)		12 เดือน (t=2)	
	กลุ่มควบคุม (0)	กลุ่มทดลอง (1)	กลุ่มควบคุม (0)	กลุ่มทดลอง (1)	กลุ่มควบคุม (0)	กลุ่มทดลอง (1)
ไม่มีปัญหา (1)	3	5	16	16	17	19
มีปัญหาบ้าง (2)	10	13	4	10	0	6
ไม่สามารถทำงานได้ (3)	3	3	0	0	0	0
รวม	16	21	20	26	17	25

[1] แหล่งที่มา : http://www.warwick.ac.uk/go/repolr. [25/04/2552]

[2] Parsons N.R., Matthew L. Costa. “Repeated measure proportional odds logistic regression analysis of ordinal score data in the statistical software package R”, Journal of Science Direct 53 (2008): 632-641

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตาราง ก.6 ความถี่และความน่าจะเป็นเชิงประจักษ์สำหรับตัวแปรตอบสนองในแต่ละกลุ่ม

ปัญหาในการทำงาน	ระยะเวลาเริ่มต้น (t=0)	6 เดือน (t=1)	12 เดือน (t=2)	รวม
ไม่มีปัญหา	8	32	36	76 (0.61)
มีปัญหาบ้าง	23	14	6	43 (0.34)
ไม่สามารถทำงานได้	6	0	0	6 (0.06)

(.) คือ ค่าความน่าจะเป็นเชิงประจักษ์

การทดสอบลักษณะของข้อมูล

กรณีศึกษาที่ 1 การติดตามผลการรักษาโรคนอนไม่หลับ

ตาราง ก.7 การทดสอบคุณสมบัติการแจกแจงปกติของตัวแปรเวลา สำหรับกรณีศึกษาที่ 1

	Shapiro-Wilk		
	Statistic	df	Sig.
time	.637	478	.000

จากตาราง ก.7 พบว่า ค่าสถิติ Shapiro-Wilk = 0.637 และ p-value = 0.000 ซึ่งมีค่าน้อยกว่า $\alpha = 0.05$ แสดงว่า ข้อมูลไม่มีการแจกแจงแบบปกติ

ตาราง ก.8 การทดสอบคุณสมบัติการเท่ากันของเมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่ม สำหรับกรณีศึกษาที่ 1

Box's M		11.128
F	Approx.	1.227
	df1	9
	df2	2019430.356
	Sig.	.273

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตาราง ก.8 พบว่า ค่าสถิติ Box's M = 11.128 ค่า p-value = 0.273 ซึ่งมีค่ามากกว่า $\alpha = 0.05$ แสดงว่า เมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่มของกรณีศึกษาที่ 1 มีค่าเท่ากัน

กรณีศึกษาที่ 2 การผ่าตัดเปลี่ยนผิวข้อสะโพก

ตาราง ก.9 การทดสอบคุณสมบัติการแจกแจงปกติของตัวแปรเวลา สำหรับกรณีศึกษาที่ 2

	Shapiro-Wilk		
	Statistic	df	Sig.
Time	.732	174	.000

จากตาราง ก.9 พบว่า ค่าสถิติ Shapiro-Wilk = 0.732 และ p-value = 0.000 ซึ่งมีค่าน้อยกว่า $\alpha = 0.05$ แสดงว่า กรณีศึกษาที่ 2 ไม่มีการแจกแจงแบบปกติ

ตาราง ก.10 การทดสอบคุณสมบัติการเท่ากันของเมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่ม สำหรับกรณีศึกษาที่ 2

Test Results

Box's M		1.483
F	Approx.	.484
	df1	3
	df2	11108.293
	Sig.	.693

Tests null hypothesis of equal population covariance matrices.

จากตาราง ก.10 พบว่า ค่าสถิติ Box's M = 1.483 ค่า p-value = 0.693 ซึ่งมีค่ามากกว่า $\alpha = 0.05$ แสดงว่า เมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่มของกรณีศึกษาที่ 2 มีค่าเท่ากัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

กรณีศึกษาที่ 3 การผ่าตัดเปลี่ยนผิวข้อสะโพก

ตาราง ก.11 การทดสอบคุณสมบัติการแจกแจงปกติของตัวแปรเวลา สำหรับกรณีศึกษาที่ 3

	Shapiro-Wilk		
	Statistic	df	Sig.
Time	.793	144	.000

จากตาราง ก.9 พบว่า ค่าสถิติ Shapiro-Wilk = 0.793 และ p-value = 0.000 ซึ่งมีค่าน้อยกว่า $\alpha = 0.05$ แสดงว่า กรณีศึกษาที่ 3 ไม่มีการแจกแจงแบบปกติ

ตาราง ก.12 การทดสอบคุณสมบัติการเท่ากันของเมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่ม สำหรับกรณีศึกษาที่ 3

Box's M		3.777
F	Approx.	1.233
	df1	3
	df2	235644.083
	Sig.	.296

จากตาราง ก.10 พบว่า ค่าสถิติ Box's M = 3.777 ค่า p-value = 0.296 ซึ่งมีค่ามากกว่า $\alpha = 0.05$ แสดงว่า เมทริกซ์ความแปรปรวนร่วมในแต่ละกลุ่มของกรณีศึกษาที่ 3 มีค่าเท่ากัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข
ผลการวิเคราะห์ข้อมูล

จากการวิเคราะห์ข้อมูลด้วยโปรแกรม R เวอร์ชัน 2.8.1 ทำการศึกษาข้อมูลทางการแพทย์ 3 ชุด ซึ่งแต่ละชุดจะทำการวิเคราะห์ 4 ตัวแบบ โดยแต่ละตัวแบบจะมีลักษณะของตัวแปรอิสระแตกต่างกันออกไป ดังผลลัพธ์ที่ได้ต่อไปนี้

กรณีศึกษาที่ 1 การติดตามผลการรักษาโรคนอนไม่หลับ

1. การวิเคราะห์จำแนกประเภท

1.1 กรณีความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง

สมการประกอบด้วยตัวแปรอิสระ ได้แก่ ประเภทของยา เวลา

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Prior probabilities of groups:

1	2	3	4
0.25	0.25	0.25	0.25

Group means:

	time	treat1
1	0.7319588	0.5360825
2	0.6610169	0.5847458
3	0.4186047	0.4573643
4	0.2686567	0.4328358

Coefficients of linear discriminants:

	LD1	LD2
time	-2.0671338	0.5740055
treat1	-0.6239369	-1.9095743

Proportion of trace:

LD1	LD2
0.9859	0.0141

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.2 กรณีความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์

สมการประกอบด้วยตัวแปรอิสระ ได้แก่ ประเภทของยา และเวลา

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Group means:

	time	treat1
1	0.7319588	0.5360825
2	0.6610169	0.5847458
3	0.4186047	0.4573643
4	0.2686567	0.4328358

Coefficients of linear discriminants:

	LD1	LD2
time	-2.0629952	0.5887063
treat1	-0.6375148	-1.9050843

Proportion of trace:

LD1	LD2
0.9868	0.0132

2. การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

2.1 กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระ

2.1.1 ฟังก์ชันสำเร็จรูป ordgee

สมการประกอบด้วยตัวแปรอิสระ ได้แก่ ประเภทของยา เวลา

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Warning message:

In binomial(link) : use of binomial(link=link) is deprecated

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Coefficients:

	estimate	san.se	wald	p
Inter:1	-1.6642250	0.4041138	16.9596480	3.818278e-05
Inter:2	-0.4217746	0.4237817	0.9905500	3.196080e-01
Inter:3	0.7972724	0.4555004	3.0636267	8.006279e-02
treat	-0.3881508	0.5806307	0.4468904	5.038154e-01
time	0.7909077	0.5428676	2.1225787	1.451420e-01

Scale is fixed.

Correlation Model:

Correlation Structure: independence

Returned Error Value: 0

Number of clusters: 239 Maximum cluster size: 2

2.1.2 ฟังก์ชันสำเร็จรูป repolr

สมการประกอบด้วยตัวแปรอิสระ ได้แก่ ประเภทของยา เวลา

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Score test = 5.081 : d.f. 4 and p-value 0.279

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
cuts1	-2.4324414	0.1880027	-12.938335	0.2133084	-11.4033993
cuts2	-1.1333301	0.1600122	-7.082773	0.1799315	-6.2986748
cuts3	0.1625974	0.1505721	1.079864	0.1694227	0.9597145
treat1	0.3938017	0.1669388	2.358959	0.2005294	1.9638108
time	1.3936014	0.1749412	7.966113	0.1382672	10.0790423

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1.0000000	0.5222778	0.2732087	0.0000000	0.0000000	0.0000000
[2,]	0.5222778	1.0000000	0.5231099	0.0000000	0.0000000	0.0000000
[3,]	0.2732087	0.5231099	1.0000000	0.0000000	0.0000000	0.0000000
[4,]	0.0000000	0.0000000	0.0000000	1.0000000	0.5222778	0.2732087
[5,]	0.0000000	0.0000000	0.0000000	0.5222778	1.0000000	0.5231099
[6,]	0.0000000	0.0000000	0.0000000	0.2732087	0.5231099	1.0000000

2.2 กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม

2.2.1 ฟังก์ชันสำเร็จรูป ordgee

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Warning message:

In binomial(link) : use of binomial(link=link) is deprecated

Coefficients:

	estimate	san.se	wald	p
Inter:1	-1.6510678	0.3952532	17.4493371	2.950672e-05
Inter:2	-0.4113955	0.4158147	0.9788574	3.224811e-01
Inter:3	0.7959079	0.4434476	3.2213735	7.268262e-02
factor(treat)1	-0.4231898	0.5706544	0.5499511	4.583377e-01
time	0.8056417	0.5195927	2.4041298	1.210154e-01

Scale is fixed.

Correlation Model:

Correlation Structure: exchangeable

Correlation Link: log

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Estimated Correlation Parameters:

	estimate	san.se	wald	p
alpha	1.979546	0.5608297	12.45859	0.0004160731

Returned Error Value: 0

Number of clusters: 239 Maximum cluster size: 2

2.2.2 ฟังก์ชันสำเร็จรูป repolr

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Score test = 5.149 : d.f. 4 and p-value 0.272

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
cuts1	-2.4376109	0.1901570	-12.818938	0.2131552	-11.4358528
cuts2	-1.1357016	0.1618548	-7.016793	0.1795666	-6.3246833
cuts3	0.1621352	0.1532533	1.057956	0.1687309	0.9609098
treat1	0.3921730	0.1730515	2.266221	0.2005456	1.9555310
time	1.3946864	0.1673665	8.333129	0.1383649	10.0797728

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1.00000000	0.52100090	0.27208613	0.08574562	0.04467355	0.02333020
[2,]	0.52100090	1.00000000	0.52223735	0.04467355	0.08574562	0.04477957
[3,]	0.27208613	0.52223735	1.00000000	0.02333020	0.04477957	0.08574562
[4,]	0.08574562	0.04467355	0.02333020	1.00000000	0.52100090	0.27208613
[5,]	0.04467355	0.08574562	0.04477957	0.52100090	1.00000000	0.52223735
[6,]	0.02333020	0.04477957	0.08574562	0.27208613	0.52223735	1.00000000

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

กรณีศึกษาที่ 2 การผ่าตัดเปลี่ยนผิวข้อสะโพก

1. การวิเคราะห์จำแนกประเภท

1.1 กรณีความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Prior probabilities of groups:

1	2	3	4
0.25	0.25	0.25	0.25

Group means:

	Time	Sex1
1	2.490566	0.6886792
2	2.777778	0.5833333
3	2.545455	0.5454545
4	4.400000	0.5000000

Coefficients of linear discriminants:

	LD1	LD2
Time	0.5880606	0.1311794
Sex1	-0.5014310	2.0155540

Proportion of trace:

LD1	LD2
0.9587	0.0413

1.2 กรณีความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Prior probabilities of groups:

1	2	3	4
0.60919540	0.20689655	0.12643678	0.05747126

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Group means:

	Time	Sex1
1	2.490566	0.6886792
2	2.777778	0.5833333
3	2.545455	0.5454545
4	4.400000	0.5000000

Coefficients of linear discriminants:

	LD1	LD2
Time	-0.5621317	-0.216867
Sex1	0.7945498	-1.919005

Proportion of trace:

LD1	LD2
0.8964	0.1036

2. การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

2.1 กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระ

2.1.1 ฟังก์ชันสำเร็จรูป ordgee

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Warning message:

In binomial(link) : use of binomial(link=link) is deprecated

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Coefficients:

	estimate	san.se	wald	p
Inter:1	2.85696114	1.320559	4.680507179	0.0305067404
Inter:2	3.89223616	1.286791	9.149178025	0.0024882982
Inter:3	5.22516847	1.392501	14.080224174	0.0001751759
Sex1	0.09079586	1.091562	0.006918868	0.9337086274
Time	-0.70725367	0.270320	6.845325222	0.0088873519

Correlation Model:

Correlation Structure: independence

Returned Error Value: 0

Number of clusters: 58 Maximum cluster size: 3

2.1.2 ฟังก์ชันสำเร็จรูป repolr

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Score test = -81.411 : d.f. 4 and p-value 1

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
cuts1	0.5875052	0.34340947	1.710801	0.3536206	1.661400
cuts2	1.6640954	0.36640871	4.541637	0.3599421	4.623230
cuts3	3.0029364	0.45089930	6.659883	0.3940484	7.620731
Time	-0.1936970	0.08817463	-2.196743	0.0549917	-3.522296
Sex1	0.5993758	0.31119621	1.926038	0.4377597	1.369189

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.0000	0.5837	0.2989	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[2,]	0.5837	1.0000	0.5120	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[3,]	0.2989	0.5120	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
[4,]	0.0000	0.0000	0.0000	1.0000	0.5837	0.2989	0.0000	0.0000	0.0000
[5,]	0.0000	0.0000	0.0000	0.5837	1.0000	0.5120	0.0000	0.0000	0.0000
[6,]	0.0000	0.0000	0.0000	0.2989	0.5120	1.0000	0.0000	0.0000	0.0000
[7,]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.5837	0.2989
[8,]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5837	1.0000	0.5120
[9,]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2989	0.5120	1.0000

2.2 กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม

2.2.2 ฟังก์ชันสำเร็จรูป repolr

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Score test = 7.993 : d.f. 4 and p-value 0.092

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
cuts1	0.5671770	0.3594707	1.577812	0.35216136	1.610560
cuts2	1.6414844	0.3842538	4.271875	0.35665419	4.602454
cuts3	2.9199674	0.4783558	6.104174	0.39489869	7.394219
Sex1	0.6237273	0.3595018	1.734977	0.43918752	1.420185
Time	-0.1937062	0.0802894	-2.412600	0.05484294	-3.532017

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.0000	0.5844	0.3084	0.1718	0.1004	0.0530	0.1718	0.1004	0.0530
[2,]	0.5844	1.0000	0.5276	0.1004	0.1718	0.0906	0.1004	0.1718	0.0906
[3,]	0.3084	0.5276	1.0000	0.0530	0.0906	0.1718	0.0530	0.0906	0.1718
[4,]	0.1718	0.1004	0.0530	1.0000	0.5844	0.3084	0.1718	0.1004	0.0530
[5,]	0.1004	0.1718	0.0906	0.5844	1.0000	0.5276	0.1004	0.1718	0.0906
[6,]	0.0530	0.0906	0.1718	0.3084	0.5276	1.0000	0.0530	0.0906	0.1718
[7,]	0.1718	0.1004	0.0530	0.1718	0.1004	0.0530	1.0000	0.5844	0.3084
[8,]	0.1004	0.1718	0.0906	0.1004	0.1718	0.0906	0.5844	1.0000	0.5276
[9,]	0.0530	0.0906	0.1718	0.0530	0.0906	0.1718	0.3084	0.5276	1.0000

2.2.2 ฟังก์ชันสำเร็จรูป ordgee

2.2.2.2 <u>ตัวแบบที่ 2</u> สมการประกอบด้วยตัวแปรอิสระ ได้แก่ เพศ และเวลา

Warning message:

In binomial(link) : use of binomial(link=link) is deprecated

Mean Model:

Mean Link: logit

Variance to Mean Relation: binomial

Coefficients:

	estimate	san.se	wald	p
Inter:1	NaN	NaN	NaN	NaN
Inter:2	NaN	NaN	NaN	NaN
Inter:3	NaN	NaN	NaN	NaN
factor(Sex)1	NaN	NaN	NaN	NaN
Time	NaN	NaN	NaN	NaN

Scale is fixed.

Correlation Model:

Correlation Structure: exchangeable

Correlation Link: log

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Estimated Correlation Parameters:

	estimate	san.se	wald	p
alpha	NaN	NaN	NaN	NaN

กรณีศึกษาที่ 3 การรักษาอาการเอ็นร้อยหวายฉีก

1. การวิเคราะห์จำแนกประเภท

1.1 กรณีความน่าจะเป็นก่อนมีค่าเท่ากันทุกระดับของตัวแปรตอบสนอง

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Prior probabilities of groups:

1	2	3
0.3333333	0.3333333	0.3333333

Group means:

	Time	Treat2
1	2.368421	0.5263158
2	1.604651	0.6744186
3	1.000000	0.5000000

Coefficients of linear discriminants:

	LD1	LD2
Time	-1.4825677	0.01616402
Treat2	0.1629043	-2.02222732

Proportion of trace:

LD1	LD2
0.9663	0.0337

1.2 กรณีความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Prior probabilities of groups:

1	2	3

0.608 0.344 0.048

Group means:

	Time	Treat2
1	2.368421	0.5263158
2	1.604651	0.6744186
3	1.000000	0.5000000

Coefficients of linear discriminants:

	LD1	LD2
Time	-1.4640429	-0.2341938
Treat2	0.5017421	-1.9657559

Proportion of trace:

LD1	LD2
0.9757	0.0243

2. การวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

2.1 กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบอิสระ

2.1.1 ฟังก์ชันสำเร็จรูป ordgee

Warning message:

In binomial(link) : use of binomial(link=link) is deprecated

Mean Model:

Mean Link: logit

Variance to Mean Relation: binomial

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Coefficients:

	estimate	san.se	wald	p
Inter:1	-1941.9685118	0	Inf	0
Inter:2	-1938.1028543	0	Inf	0
factor(Treat)2	0.5864554	0	Inf	0
Time	1940.4256377	0	Inf	0

Correlation Structure: independence

Returned Error Value: 1

Number of clusters: 48 Maximum cluster size: 3

2.1.2 ฟังก์ชันสำเร็จรูป repolr

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Score test = 2.276 : d.f. 2 and p-value 0.32

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
cuts1	-2.5153596	0.6340579	-3.9670816	0.6493466	-3.8736780
cuts2	0.6343052	0.6486955	0.9778165	0.7697338	0.8240578
Treat2	-0.6730559	0.4297157	-1.5662818	0.4440682	-1.5156588
Time	1.7172971	0.3118054	5.5075929	0.3096614	5.5457246

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1.0000000	0.2070423	0.0000000	0.0000000	0.0000000	0.0000000
[2,]	0.2070423	1.0000000	0.0000000	0.0000000	0.0000000	0.0000000
[3,]	0.0000000	0.0000000	1.0000000	0.2070423	0.0000000	0.0000000
[4,]	0.0000000	0.0000000	0.2070423	1.0000000	0.0000000	0.0000000
[5,]	0.0000000	0.0000000	0.0000000	0.0000000	1.0000000	0.2070423
[6,]	0.0000000	0.0000000	0.0000000	0.0000000	0.2070423	1.0000000

2.1 กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม

2.2.1 ฟังก์ชันสำเร็จรูป ordgee

Warning message:

In binomial(link) : use of binomial(link=link) is deprecated

Mean Model:

Mean Link: logit

Variance to Mean Relation: binomial

Coefficients:

	estimate	san.se	wald	p
Inter:1	NaN	NaN	NaN	NaN
Inter:2	NaN	NaN	NaN	NaN
factor(Treat)2	NaN	NaN	NaN	NaN
Time	NaN	NaN	NaN	NaN

Scale is fixed.

Estimated Correlation Parameters:

	estimate	san.se	wald	p
alpha	NaN	NaN	NaN	NaN

Returned Error Value: 1

Number of clusters: 48 Maximum cluster size: 3

2.2.2 ฟังก์ชันสำเร็จรูป repolr

จากการวิเคราะห์ข้อมูลได้ผลการวิเคราะห์ ดังนี้

Score test = 2.306 : d.f. 2 and p-value 0.316

Coefficients:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
cuts1	-2.4931846	0.6255705	-3.9854576	0.6458034	-3.8605935
cuts2	0.6040495	0.6478425	0.9324018	0.7624233	0.7922758
Treat2	-0.6428807	0.4426216	-1.4524386	0.4413975	-1.4564666
Time	1.6991265	0.3037854	5.5931811	0.3085840	5.5062049

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Working Correlation

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	1.00000000	0.21254505	0.05000000	0.01062725	0.05000000	0.01062725
[2,]	0.21254505	1.00000000	0.01062725	0.05000000	0.01062725	0.05000000
[3,]	0.05000000	0.01062725	1.00000000	0.21254505	0.05000000	0.01062725
[4,]	0.01062725	0.05000000	0.21254505	1.00000000	0.01062725	0.05000000
[5,]	0.05000000	0.01062725	0.05000000	0.01062725	1.00000000	0.21254505
[6,]	0.01062725	0.05000000	0.01062725	0.05000000	0.21254505	1.00000000

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค
รายละเอียดโปรแกรมคำสั่งที่ใช้ในการวิจัย

ใช้โปรแกรม R ในการวิเคราะห์ข้อมูลจำแนกประเภทและมัลติโนเมียลโลจิสติค ในส่วนนี้จะแสดงตัวอย่างในการเรียกใช้คำสั่ง ดังนี้

1. ฟังก์ชัน lda ในการวิเคราะห์จำแนกประเภท
2. ฟังก์ชัน repolr ในการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค
3. ฟังก์ชัน ordgee ในการวิเคราะห์การถดถอยมัลติโนเมียลโลจิสติค

1. ฟังก์ชัน lda

```
################  ตัวอย่างการเรียกใช้คำสั่ง  ################
### กรณีกำหนดให้ความน่าจะเป็นก่อนมีค่าเท่ากับความน่าจะเป็นเชิงประจักษ์ ####
insonia.dat<-read.table("C:/Documents and Settings/Administrator/Desktop/THESIS By
Suparat/insonia2.txt", col.names=c("case","treat","time","outcome"),header=T)
attach(insonia.dat)
library(MASS)
fit <- lda(outcome ~ time * factor(treat), insonia.dat, na.action="na.omit",
CV=FALSE,prior=c(0.2,0.25,0.27,0.28))

################  ตัวอย่างการเรียกใช้คำสั่ง  ################
### กรณีกำหนดให้ความน่าจะเป็นก่อนมีค่าเท่ากันทุกค่าของตัวแปรตอบสนอง ####
insonia.dat<-read.table("C:/Documents and Settings/Administrator/Desktop/THESIS By
Suparat/insonia2.txt", col.names=c("case","treat","time","outcome"),header=T)
attach(insonia.dat)
library(MASS)
fit <- lda(outcome ~ time * factor(treat), insonia.dat,
  na.action="na.omit", CV=FALSE,prior=c(0.25,0.25,0.25,0.25))
**********************************************************************************************************
**********************************************************************************************************
```

2. **ฟังก์ชัน** repolr

```
################  ตัวอย่างการเรียกใช้คำสั่ง  ################
############  กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นอิสระกัน  ############

insonia.dat<-read.table("C:/Documents and Settings/Administrator/Desktop/THESIS By
Suparat/insonia2.txt", col.names=c("case","treat","time","outcome"),header=T)
attach(insonia.dat)
library(gee)
source("C:/Documents and Settings/Administrator/Desktop/codefunction/repolr_functions.r")
fitted.mod <- repolr(formula = outcome~factor(treat) *(time), subjects = "case" ,
data = insonia.dat , categories = 4 ,  times = c(0,1) , corstr = "independence", tol = 0.001,
scalevalue = 1, alpha = 0.5,po.test=TRUE, fixed=FALSE)
summary(fitted.mod[["gee"]])

################  ตัวอย่างการเรียกใช้คำสั่ง  ################
############  กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม  ############
insonia.dat<-read.table("C:/Documents and Settings/Administrator/Desktop/THESIS By
Suparat/insonia2.txt", col.names=c("case","treat","time","outcome"),header=T)
attach(insonia.dat)
library(gee)
source("C:/Documents and Settings/Administrator/Desktop/codefunction/repolr_functions.r")
fitted.mod <- repolr(formula = outcome~factor(treat) *(time), subjects = "case" ,
data = insonia.dat , categories = 4 ,  times = c(0,1) , corstr = "uniform", tol = 0.001,
scalevalue = 1, alpha = 0.5,po.test=TRUE, fixed=FALSE)
summary(fitted.mod[["gee"]])

**********************************************************************************************************************
**********************************************************************************************************************
**********************************************************************************************************************
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. **ฟังก์ชัน** ordgee

```
################ ตัวอย่างการเรียกใช้คำสั่ง ################
########### กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นอิสระต่อกัน ###########
table.insonia<-insonia.dat<-read.table("C:/Documents and Settings/Administrator/Desktop/THESIS
By Suparat/insonia2.txt",header=T)
table.insonia_ORD<-table.insonia[order(table.insonia$case, table.insonia$time),]
table.insonia_ORD$outcome<-ordered(table.insonia_ORD$outcome, levels=1:4)
library(geepack)
fit.ordgee<-ordgee(outcome~factor(treat)*time,id=case,family=binomial,
data=table.insonia_ORD,corstr="independence",rev=TRUE,
control=geese.control(maxit=100))
summary(fit.ordgee)

################ ตัวอย่างการเรียกใช้คำสั่ง ################
########### กรณีความสัมพันธ์ของความคลาดเคลื่อนเป็นแบบยูนิฟอร์ม ###########
table.insonia<-insonia.dat<-read.table("C:/Documents and Settings/Administrator/Desktop/THESIS
By Suparat/insonia2.txt",header=T)
table.insonia_ORD<-table.insonia[order(table.insonia$case, table.insonia$time),]
table.insonia_ORD$outcome<-ordered(table.insonia_ORD$outcome, levels=1:4)
library(geepack)
fit.ordgee<-ordgee(outcome~factor(treat)*time,id=case,family=binomial,
data=table.insonia_ORD,corstr="exchangeable",rev=TRUE,
control=geese.control(maxit=100))
summary(fit.ordgee)
```

****

****

****

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศุภรัตน์ ดิษบรรจง เกิดเมื่อวันที่ 22 มิถุนายน พ.ศ. 2526 ที่จังหวัดลพบุรี สำเร็จการศึกษาปริญญาวิทยาศาสตร์บัณฑิต (สถิติ) จากภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ เมื่อปีการศึกษา 2548 จากนั้นได้เข้าศึกษาต่อในระดับมหาบัณฑิต หลักสูตรสถิติศาสตร์มหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ระหว่างศึกษาในระดับมหาบัณฑิต ได้ประกอบอาชีพอิสระ ตำแหน่ง โปรแกรมเมอร์ แผนก Information service บริษัท Ipsos (ประเทศไทย) ในปี พ.ศ. 2548

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย