

การศึกษาเปรียบเทียบตัวแบบในการวินิจฉัยโรคเลปโตสไปโรซิส

โดย

นางสาวพนิดา คงแจ่ม

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติประยุกต์

ภาควิชาสถิติ

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2548

ISBN 974-11-6146-8

ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

A COMPARATIVE STUDY OF LEPTOSPIROSIS DIAGNOSIS MODELS

By

Panida Kongjam

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

Department of Statistics

Graduate School

SILPAKORN UNIVERSITY

2005

ISBN 974-11-6146-8

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร อนุมัติให้วิทยานิพนธ์เรื่อง “การศึกษาเปรียบเทียบตัวแบบในการวินิจฉัยโรคเลปโตสไปโรสิส” เสนอโดย นางสาวพนิดา กงแจ่ม เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์

.....

(รองศาสตราจารย์ ดร.วิสาข์ จัติวัตร)

รองอธิการบดีฝ่ายวิชาการ รักษาราชการแทน

คณบดีบัณฑิตวิทยาลัย

วันที่ เดือน พ.ศ.

ผู้ควบคุมวิทยานิพนธ์

1. ผู้ช่วยศาสตราจารย์ ดร.กมลชนก พานิชการ
2. ศาสตราจารย์ แพทย์หญิง ยูพิน สุพุทธมงคล

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์
คณะกรรมการตรวจสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ วีรฉัตร พงศาภักดิ์)

...../...../.....

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.กมลชนก พานิชการ)

...../...../.....

..... กรรมการ

(ศาสตราจารย์ แพทย์หญิง ยูพิน สุพุทธมงคล)

...../...../.....

..... กรรมการ

(รองศาสตราจารย์ วัฒนา เกาศัลย์)

...../...../.....

..... กรรมการ

(อาจารย์ ดร. บุญอ้อม โนมทิ)

...../...../.....

กิตติกรรมประกาศ

ปัจจัยสำคัญที่ทำให้วิทยานิพนธ์สำเร็จเรียบร้อย เนื่องมาจากความกรุณาของคณาจารย์ ผู้ประสิทธิ์ประสาทความรู้ ในโอกาสที่ผู้วิจัยกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. กมลชนก พานิชการ สำหรับความเอาใจใส่และความช่วยเหลือทั้งหลายที่ช่วยให้ผ่านพ้นปัญหาต่างๆไปได้ และศาสตราจารย์แพทย์หญิง ยุพิน ศุพุทธมงคล สำหรับข้อมูลที่เป็นประโยชน์อย่างยิ่งในการศึกษา และข้อเสนอแนะอันเป็นประโยชน์ต่อผู้วิจัย ขอบพระคุณอาจารย์ผู้ควบคุมทั้งสองท่านอย่างยิ่ง ที่คอยให้คำแนะนำตรวจแก้ไขการเรียบเรียงวิทยานิพนธ์ รวมทั้งเสียสละเวลาทุกเมื่อเพื่อให้งานวิจัย เสร็จสมบูรณ์

กราบขอบพระคุณ รองศาสตราจารย์ วีรพันธ์ พงศาภักดี ประธานกรรมการสอบ วิทยานิพนธ์ รองศาสตราจารย์ วัฒนา เกาศัลย์ และ อาจารย์ ดร. บุญอ้อม โนมที กรรมการ ผู้ตรวจสอบวิทยานิพนธ์ ที่กรุณาชี้แนะส่วนที่ต้องแก้ไข ทำให้วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ด้วยความเรียบร้อย

กราบขอบพระคุณอาจารย์ทุกท่านในภาควิชาสถิติที่คอยประสาทความรู้ ส่งเสริมและเป็นกำลังใจให้อย่างสม่ำเสมอ ตลอดเวลาที่อยู่ในรั้วของมหาวิทยาลัยศิลปากร สถาบันที่ภาคภูมิใจ กราบขอบพระคุณ อาจารย์อุไรวรรณ อมรมนิมิตร มหาวิทยาลัยหอการค้าไทย สำหรับบทความที่ อาจารย์กรุณาส่งมาให้ศึกษาเพิ่มเติม ขอบพระคุณเจ้าหน้าที่ห้องปฏิบัติการโรคติดเชื้อ ดึกะอบ ชั้น 4 โรงพยาบาลศิริราชทุกท่าน สำหรับข้อมูลที่เป็นประโยชน์ในงานวิจัยและกำลังใจที่ดี ตลอดเวลาการทำงาน ขอบคุณ คุณหทัยชนก แจ่มถิ่น เพื่อนที่แสนดีจากบัณฑิตวิทยาลัย ที่อำนวยความสะดวกและเป็นกำลังใจให้เสมอ

ขอบคุณเพื่อน พี่ และน้องๆ สาขาวิชาสถิติประยุกต์ที่ให้กำลังใจ เป็นที่พึ่งที่ดี และขอบคุณสำหรับความรักความผูกพันอันเหนียวแน่นที่มีให้กันไม่เคยเปลี่ยนแปลง โดยเฉพาะเพื่อนร่วมชั้นปี อ.ภาณุพงศ์ พนมวัน , อ.อาฟีฟี ลาเต๊ะ , อ.พวงทิพย์ พูลสวัสดิ์ และ อ.สุมณฑา เขียวเสน สำหรับความช่วยเหลืออย่างดีเสมอมา และมิตรภาพที่แสนดีที่ไม่สามารถหาได้จากที่อื่น

สุดท้ายกราบขอบพระคุณ ผอ.ไพฑูริย์ คงแจ่ม และอาจารย์นภา คงแจ่ม คุณพ่อคุณแม่ ผู้อยู่เบื้องหลังความสำเร็จทั้งหมด ขอบคุณสำหรับการเป็นครูและเพื่อนที่ดีที่สุดในชีวิต ขอบคุณกำลังใจที่มีให้ตลอดเวลาที่ต้องการ และขอบคุณสำหรับการเป็นผู้สนับสนุนทุนวิจัยส่วนตัว

ผู้วิจัยขออาราธนาคุณพระศรีรัตนตรัย พระร่วงโรจนฤทธิ์และสิ่งศักดิ์สิทธิ์ทั้งหลาย คลบบันดาลให้ทุกท่านประสบแต่ความสุข ความเจริญ และประสบความสำเร็จตลอดไป

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ฅ
สารบัญรูป.....	ฎ
บทที่	
1 บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
วัตถุประสงค์การวิจัย	3
สมมติฐานของการวิจัย	4
ขอบเขตการวิจัย	4
ประโยชน์ที่คาดว่าจะได้รับ	5
นิยามศัพท์	6
2 ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง	9
การวิเคราะห์การจำแนก	9
ตัวแบบการถดถอยโลจิสติก	25
การประมาณอัตราความผิดพลาดในการจำแนกกลุ่ม	32
โรคเลปโตสไปโรซิส	34
งานวิจัยที่เกี่ยวข้อง	40
3 วิธีดำเนินการวิจัย	43
4 ผลการวิจัย	48
ผลการวิเคราะห์ข้อมูลกรณีที่ 1	50
ผลการวิเคราะห์ข้อมูลกรณีที่ 2	63
ผลการวิเคราะห์ข้อมูลกรณีที่ 3	76
ผลการตรวจสอบตัวแบบโดยใช้ข้อมูลชุดใหม่	96
ผลการเปรียบเทียบกับตัวแบบองค์การอนามัยโลก	100

บทที่	หน้า
5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	102
สรุปผลการวิจัย	102
อภิปรายผล	106
ข้อเสนอแนะจากงานวิจัย.....	107
บรรณานุกรม.....	108
ประวัติผู้วิจัย.....	110

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์

สารบัญตาราง

ตารางที่	หน้า
1 ข้อมูลคะแนนการผลิตและความทนทานสูงสุด ของล้อยี่ห้อที่ผลิตในอุณหภูมิต่างกัน 2 ระดับ	15
2 Discriminant function $z = -1.633y_1 + 1.819y_2$ คำนวณจากข้อมูลในตารางที่ 1.....	16
3 ผลการประมวลข้อมูลการแพร่กระจายของเชื้อโรคมะเร็งด้วย SPSS/FW	31
4 ผลการจำแนกค่าสังเกตเข้ากลุ่ม (Confusion Matrix)	33
5 จำนวนผู้ป่วยและผู้ตายด้วยโรคเลปโตสไปโรสิส แยกตามปี พ.ศ. 2531-2543	37
6 ค่าสถิติ Box'M ของการจำแนกกรณีที่ 1	50
7 ข้อมูลพื้นฐานของกลุ่มที่ศึกษากรณีที่ 1.....	53
8 ค่าเฉลี่ยของผลตรวจทางห้องปฏิบัติการสำหรับกรณีที่ 1.....	55
9 ผลการคัดเลือกตัวแปรเข้าสมการโดยวิธี Stepwise สำหรับกรณีที่ 1.....	56
10 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนกด้วยค่าไอเกน (Eigenvalue) สำหรับกรณีที่ 1.....	56
11 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนกด้วยค่า Wilks'Lambda สำหรับกรณีที่ 1	57
12 ค่าศูนย์กลางของกลุ่ม (Group centroids) ของข้อมูลทั้ง 2 กลุ่มสำหรับกรณีที่ 1.....	57
13 ค่าสัมประสิทธิ์คานอนิคัล (Unstandardized coefficients)ค่าสัมประสิทธิ์คานอนิคัลมาตรฐาน(Standardized Coefficients) สหสัมพันธ์กับฟังก์ชันการจำแนกของตัวแปรอิสระและสัมประสิทธิ์ฟังก์ชันจำแนกสำหรับกรณีที่ 1.....	58
14 ผลการจัดข้อมูลเข้ากลุ่ม เมื่อใช้ฟังก์ชันการวิเคราะห์จำแนกสำหรับกรณีที่ 1	59
15 ผลการวิเคราะห์ข้อมูลเพื่อสร้างสมการถดถอยโลจิสติกสำหรับกรณีที่ 1	59
16 ผลการตรวจสอบภาวะสารูปดี (Goodness-of-fit Test) และค่าR-Squareสำหรับกรณีที่ 1	60
17 จำนวนข้อมูลที่ถูกต้องจัดเข้ากลุ่มถูกต้อง เมื่อใช้จุดตัดการจำแนก (Classification Cut Point) ที่ระดับต่างๆสำหรับกรณีที่ 1.....	61
18 ผลการจำแนกกรณีที่ 1 โดยใช้วิธีการวิเคราะห์จำแนกและการถดถอยโลจิสติก.....	63
19 ค่าสถิติ Box's M ของการจำแนกกรณีที่ 2	63
20 ข้อมูลพื้นฐานของกลุ่มที่ศึกษากรณีที่ 2	66
21 ค่าเฉลี่ยของผลตรวจทางห้องปฏิบัติการสำหรับกรณีที่ 2	67

ตารางที่	หน้า
22 ผลการคัดเลือกตัวแปรเข้าสมการ โดยวิธี Stepwise สำหรับกรณีที่ 2	68
23 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนก ด้วยค่าไอเกนสำหรับกรณีที่ 2	69
24 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนก ด้วยค่า Wilks' Lambda สำหรับกรณีที่ 2	69
25 ค่าศูนย์กลางของกลุ่มของข้อมูลทั้ง 2 กลุ่มสำหรับกรณีที่ 2.....	70
26 ค่าสัมประสิทธิ์คานอนิคัล, ค่าสัมประสิทธิ์คานอนิคัลมาตรฐาน, สหสัมพันธ์กับ ฟังก์ชันการจำแนกของตัวแปรอิสระและสัมประสิทธิ์ฟังก์ชันจำแนกกรณีที่ 2	70
27 ผลการจัดข้อมูลเข้ากลุ่ม เมื่อใช้ฟังก์ชันการวิเคราะห์จำแนกสำหรับกรณีที่ 2.....	71
28 ผลการวิเคราะห์ข้อมูลเพื่อสร้างสมการถดถอยโลจิสติกสำหรับกรณีที่ 2.....	72
29 ผลการตรวจสอบภาวะสารถูดี (Goodness-of-fit Test) และ R-Square สำหรับกรณีที่ 2....	73
30 จำนวนข้อมูลที่ถูกต้องเข้ากลุ่มถูกต้อง เมื่อใช้จุดตัดการจำแนกที่ระดับต่างๆกรณีที่ 2.....	74
31 ผลการจำแนกกรณีที่ 2 โดยใช้วิธีการวิเคราะห์จำแนกและการถดถอยโลจิสติก.....	75
32 ค่าสถิติ Box's M ของการจำแนกกรณีที่ 3	76
33 ข้อมูลพื้นฐานของกลุ่มที่ศึกษา กรณีที่ 3	80
34 ค่าเฉลี่ยของผลตรวจทางห้องปฏิบัติการสำหรับกรณีที่ 3.....	81
35 ผลการคัดเลือกตัวแปรเข้าสมการการวิเคราะห์จำแนกกรณีที่ 3 โดยใช้วิธี Stepwise.....	83
36 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนก กรณีที่ 3 ด้วยค่าไอเกน	83
37 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนก กรณีที่ 3 ด้วยค่า Wilks' Lambda	84
38 ค่าสัมประสิทธิ์แคนอนิคัล และค่าสัมประสิทธิ์ แคนอนิคัลมาตรฐาน และ สหสัมพันธ์กับฟังก์ชันจำแนกของตัวแปรอิสระ สำหรับกรณีที่ 3	84
39 ค่าสัมประสิทธิ์ของฟังก์ชันการวิเคราะห์จำแนกกรณีที่ 3.....	85
40 ค่าศูนย์กลางของกลุ่ม สำหรับกรณีที่ 3	86
41 ผลการจัดข้อมูลเข้ากลุ่ม โดยใช้การวิเคราะห์การจำแนก สำหรับกรณีที่ 3	86
42 การทดสอบความเหมาะสมและ R-Square ของตัวแบบกรณีที่ 3 โดยใช้ตัวแปร อิสระ 19 ตัว	87

ตารางที่	หน้า
43 ค่าจากการทดสอบ Likelihood Ratio Test ของตัวแปรอิสระแต่ละตัวกรณีที่มี 3.....	88
44 ผลการวิเคราะห์ข้อมูลเพื่อสร้างสมการถดถอยโลจิสติกกรณีที่มี 3 ตัวแปรอิสระ 19 ตัว.....	89
45 ผลการจัดค่าข้อมูลเข้ากลุ่มเมื่อใช้การวิเคราะห์การถดถอยโลจิสติกกรณีที่มี 3 ใช้ 19 ตัวแปร..	91
46 การทดสอบความเหมาะสมและ R-Square ของตัวแบบกรณีที่มี 3 โดยใช้ตัวแปร อิสระ 10 ตัว	92
47 ค่าจากการทดสอบ Likelihood Ratio Test ของตัวแปรอิสระแต่ละตัวเมื่อใช้ การวิเคราะห์การถดถอยโลจิสติก กรณีที่มี 3 ตัวแปรอิสระ 10 ตัวแปร	92
48 ผลการวิเคราะห์ข้อมูลเพื่อสร้างสมการถดถอยโลจิสติกกรณีที่มี 3 ตัวแปรอิสระ 10 ตัว.....	93
49 ผลการจัดค่าข้อมูลเข้ากลุ่มโดยการวิเคราะห์การถดถอยโลจิสติกกรณีที่มี 3 ใช้ 10 ตัวแปร..	95
50 ผลการจำแนกข้อมูลเข้ากลุ่มกรณีที่มี 3 ใช้ 10 ตัวแปรเปรียบเทียบ 2 วิธี.....	96
51 ผลการจำแนกชุดใหม่ เข้ากลุ่ม โดยใช้การวิเคราะห์จำแนกจากกรณีที่มี 1	97
52 ผลการจำแนกข้อมูลชุดใหม่เข้ากลุ่มโดยใช้การวิเคราะห์การถดถอยโลจิสติก กรณีที่มี 1..	97
53 ผลการจำแนกชุดใหม่ เข้ากลุ่ม โดยใช้วิธีการวิเคราะห์จำแนกจากกรณีที่มี 2	98
54 ผลการจำแนกข้อมูลชุดใหม่เข้ากลุ่ม โดยใช้การวิเคราะห์การถดถอยโลจิสติกกรณีที่มี 2..	98
55 ผลการจำแนกข้อมูลชุดใหม่เข้ากลุ่ม โดยใช้การวิเคราะห์จำแนกสำหรับกรณีที่มี 3	99
56 ผลการจำแนกข้อมูลชุดใหม่เข้ากลุ่มโดยใช้การวิเคราะห์การถดถอยโลจิสติกกรณีที่มี 3....	100
57 ผลการจำแนกข้อมูลเข้ากลุ่ม โดยใช้วิธีการวิเคราะห์จำแนก , การถดถอยโลจิสติก และใช้เกณฑ์ขององค์การอนามัยโลก สำหรับการจำแนกข้อมูลเป็น 2 กลุ่ม ...	101
58 เปอร์เซนต์จัดเข้ากลุ่มถูกต้องโดยวิธีการวิเคราะห์การจำแนกและการวิเคราะห์การ ถดถอยโลจิสติกสำหรับทั้ง 3 กรณี	103
59 เปอร์เซนต์จัดข้อมูลชุดใหม่เข้ากลุ่มถูกต้องโดยวิธีการวิเคราะห์การจำแนกและการ วิเคราะห์การถดถอยโลจิสติกสำหรับทั้ง 3 กรณี	105

สารบัญรูป

รูปที่	หน้า
1 การวิเคราะห์การจำแนกสำหรับ 2 กลุ่ม	13
2 การจำแนกโดยใช้ฟังก์ชันการจำแนกกลุ่ม	14
3 กราฟระหว่างคะแนนการผลิตและความทนทานสูงสุดที่ระดับอนุหภูมิ 2 ระดับ	16
4 วิธีการของ Fisher สำหรับการจำแนกกลุ่ม 2 กลุ่ม	18
5 การกระจายของส่วนเหลือกับค่าพยากรณ์ความน่าจะเป็นของกรณีที่ 1	51
6 การกระจายของค่า Leverage กับหมายเลขประจำของค่าสังเกต กรณีที่ 1	52
7 การกระจายของค่า Cook's distance กับหมายเลขประจำของค่าสังเกต กรณีที่ 1	52
8 การกระจายของส่วนเหลือกับค่าพยากรณ์ความน่าจะเป็นของกรณีที่ 2	64
9 การกระจายของค่า Leverage กับหมายเลขประจำของค่าสังเกต กรณีที่ 2	65
10 การกระจายของค่า Cook's distance กับหมายเลขประจำของค่าสังเกต กรณีที่ 2	65
11 การกระจายของส่วนเหลือกับค่าพยากรณ์ความน่าจะเป็นของกรณีที่ 3	77
12 การกระจายของค่า Leverage กับหมายเลขประจำของค่าสังเกต กรณีที่ 3	78
13 การกระจายของค่า Cook's distance กับหมายเลขประจำของค่าสังเกต กรณีที่ 3	79

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

โรคเลปโตสไปโรซิสหรือเป็นที่รู้จักในนามโรคฉี่หนู เป็นโรคของสัตว์ที่ติดต่อมายังคนเกิดจากเชื้อก่อโรคใน genus pathogenic *Leptospira spp.* ผู้ป่วยมีอาการและอาการแสดงแตกต่างกันได้ตั้งแต่ไม่รุนแรงหายได้เองจนเสียชีวิตได้ เป็นโรคที่พบได้ทุกภูมิภาคของโลก ในประเทศไทยมีรายงานผู้ป่วยครั้งแรกจำนวน 4 ราย และเสียชีวิต 2 ราย เมื่อปี พ.ศ. 2486 ซึ่งเกิดน้ำท่วมกรุงเทพฯ ครั้งใหญ่ โดยนายแพทย์ใช้ ยูนิพันธ์ ที่โรงพยาบาลศิริราช หลังจากนั้นมียารายงานพบผู้ป่วยโรคนี้ได้ทุกภาคของประเทศโดยภาคตะวันออกเฉียงเหนือมีอัตราป่วยสูงสุด (Suputtamongkol ,et al. 2004 :1417-24)

ผู้ป่วยโรคเลปโตสไปโรซิสส่วนใหญ่มีอาการไข้เฉียบพลันซึ่งไม่รุนแรง ผู้ป่วยร้อยละ 10-30 จะมีภาวะแทรกซ้อน เช่น ตาเหลือง ปอดอักเสบ และผู้ป่วยจำนวนหนึ่งจะมีอาการที่รุนแรง เช่น ไตวาย เลือดออกผิดปกติที่รุนแรง โดยเฉพาะอย่างยิ่งอาการเลือดออกผิดปกติในปอดซึ่งเป็นสาเหตุสำคัญที่ทำให้ผู้ป่วยโรคนี้เสียชีวิตในขณะนี้ การวินิจฉัยโรคนี้อาศัยอาการทางคลินิกเป็นสำคัญ เนื่องจากยังไม่มี การตรวจทางห้องปฏิบัติการที่มีความไว ความจำเพาะสูงและสามารถนำมาใช้ได้อย่างแพร่หลายในปัจจุบัน ยาต้านจุลชีพซึ่งใช้ในการรักษาโรคนี้ในปัจจุบันได้แก่ ยาเพนิซิลลินและด็อกซิซัยคลิน โดยยาเพนิซิลลินเป็นยาซึ่งออกฤทธิ์จำเพาะต่อเชื้อกรัมบวกบางชนิดและเชื้อเลปโตสไปราเท่านั้น จึงควรใช้ในการรักษาผู้ป่วยในกรณีที่แพทย์สามารถยืนยันการวินิจฉัยว่าผู้ป่วยเป็นโรคเลปโตสไปโรซิสได้อย่างรวดเร็วเท่านั้น

ข้อมูลจากการศึกษาทางระบาดวิทยา ของผู้ป่วยที่มีภาวะไข้เฉียบพลันซึ่งไม่พบสาเหตุว่า เกิดจากการติดเชื้อที่อวัยวะใดอวัยวะหนึ่งชัดเจน (Acute Pyrexia of Unknown Origin, Acute PUO) ทั้งในประเทศต่างๆในภูมิภาคเอเชียตะวันออกเฉียงใต้และในประเทศไทยพบว่า โรคเลปโตสไปโรซิสและโรคสครับไทฟัสเป็นสาเหตุที่พบบ่อยที่สุดที่ทำให้เกิดภาวะนี้ ข้อมูลจากการศึกษาภาวะนี้ในภาคตะวันออกเฉียงเหนือของประเทศไทยเมื่อปี พ.ศ. 2544 นี้พบว่า โรคเลปโตสไปโรซิสและโรคสครับไทฟัสเป็นสาเหตุของภาวะ Acute PUO นี้ร้อยละ 29 และร้อยละ 20 ตามลำดับยิ่งไปกว่านั้นพบว่าอาการและอาการแสดงของโรคเลปโตสไปโรซิสและโรคสครับไทฟัสคล้ายคลึงกันเป็นอย่างมาก การตรวจพบแผลที่เกิดจากตัวไรอ่อนกัด (Eschar) เท่านั้นที่มีประโยชน์

ในการวินิจฉัยแยกโรคสคริปต์พิษจากโรคเลปโตสไปโรสิส และภาวะอื่นๆที่เป็นสาเหตุของภาวะ Acute PUO นี้ได้ อย่างไรก็ตามแผล Eschar ดังกล่าวพบได้ประมาณร้อยละ 30 ของผู้ป่วยโรคสคริปต์พิษเท่านั้น แม้ว่าผู้ป่วยโรคเลปโตสไปโรสิส และโรคสคริปต์พิษที่มีอาการไม่รุนแรงอาจหายได้เองโดยไม่ต้องได้รับการรักษาด้วยยาต้านจุลชีพ แต่การศึกษาทางคลินิกพบว่าการรักษาด้วยยาต้านจุลชีพที่เหมาะสมจะทำให้การดำเนินโรคลงอย่างชัดเจน อีกทั้งอาจป้องกันการเกิดภาวะแทรกซ้อนที่รุนแรงได้ นอกจากนี้โรคสคริปต์พิษโรคที่ผู้ป่วยมีอาการคล้ายโรคเลปโตสไปโรสิสมีด้วยกันหลายโรคได้แก่ ไข้หวัดใหญ่ เชื้อหุ้มสมองอักเสบ ไข้เลือดออกเด็งกี ไข้ไทฟอยด์ เป็นต้น (สำนักงานโครงการควบคุมโรคเลปโตสไปโรสิส 2544 : 28)

ในการวินิจฉัยโรคของผู้ป่วยแพทย์จำเป็นต้องอาศัยความรวดเร็วในการตัดสินใจว่าผู้ป่วยคนนั้นเป็นโรคเลปโตสไปโรสิส หรือโรคอื่นๆ เพื่อให้แพทย์สามารถเลือกวิธีการรักษาได้ทันเวลาที่และตรงกับโรค และเพื่อให้เกิดผลดีต่อตัวผู้ป่วยมากที่สุดทั้งในแง่ของการรักษาอาการของโรคที่เป็นอยู่ให้หาย , หลีกเลี่ยงการได้รับยาที่ไม่จำเป็น และป้องกันการเกิดภาวะแทรกซ้อนรุนแรงที่อาจเกิดจากการได้รับยา

ในปัจจุบันแพทย์จะวินิจฉัยโรคโดยอาศัยอาการที่ทางคลินิกที่เข้ากับโรค , ประวัติทางระบาดวิทยา และอาศัยผลตรวจทางห้องปฏิบัติการเพื่อยืนยันได้แก่ ผลเพาะเชื้อ (Culture) และผลตรวจซีรัม รวมทั้งอาศัยตัวแบบในการช่วยวินิจฉัยโรค ซึ่งเป็นตัวแบบที่องค์การอนามัยโลก (WHO) ซึ่งมีด้วยกันทั้งหมด 3 ส่วน ส่วนที่หนึ่งเป็นอาการแสดงและผลการตรวจทางห้องปฏิบัติการเบื้องต้นของผู้ป่วย ส่วนที่สองเป็นปัจจัยเสี่ยงต่อการเป็นโรคเลปโตสไปโรสิสของผู้ป่วย และส่วนที่สามเป็นผลตรวจที่สามารถยืนยันได้ผู้ป่วยเป็นโรคเลปโตสไปโรสิสหรือไม่ ในทางปฏิบัติพบว่าตัวแบบนี้ยังไม่สอดคล้องกับข้อมูลจริงที่พบในประเทศไทยมากนัก อาจเนื่องมาจากลักษณะอาการที่พบมีความแตกต่างกันและข้อมูลขององค์การอนามัยโลกสร้างตัวแบบมาจากข้อมูลของหลายประเทศ และการนำตัวแบบนี้ไปใช้หากจะพิจารณาเพียงสองส่วนแรกจะพบว่าตัวแบบยังไม่สามารถจำแนกการเป็นหรือไม่เป็นโรคของผู้ป่วยได้อย่างมีประสิทธิภาพดีพอ แต่ถ้าพิจารณาทั้งสามส่วนจะสามารถจำแนกได้อย่างมีประสิทธิภาพทั้งนี้เนื่องจากตัวแบบที่อิงเกณฑ์ขององค์การอนามัยโลกในส่วนที่สามจะเป็นส่วนที่เกี่ยวข้องกับผลการตรวจทางห้องปฏิบัติการซึ่งเป็นผลที่สามารถระบุการเป็นโรคได้ค่อนข้างแน่นอน แต่การจะได้ผลตรวจเหล่านั้นมาได้จะต้องรอผลนานถึง 3 สัปดาห์ (สำนักงานโครงการควบคุมโรคเลปโตสไปโรสิส 2544)

รวมทั้งผลตรวจทางห้องปฏิบัติการที่แพทย์ใช้เป็นมาตรฐานในการวินิจฉัยโรคนี้ต้องอาศัยระยะเวลาในการรอผลเป็นเวลานาน คือ ผลการเพาะเชื้อต้องรอผลประมาณ 3 สัปดาห์จึง

ตัดสินได้ว่าผู้ป่วยเป็นหรือไม่เป็นโรค และผลการตรวจทางซีโรโลยี (Serology) หากจะนำมาช่วยวินิจฉัยอย่างเชื่อถือได้จะแยกการพิจารณาออกเป็น 2 กรณีได้แก่ การทดสอบครั้งเดียว ไตเตอร์ (Titer) สูง คือมีค่าตั้งแต่ 200 ขึ้นไป (Single > 200) หรือทดสอบ 2 ครั้งโดยห่างกันประมาณ 14 วัน (Paired Serum) หากค่าของผลครั้งที่ 2 มีค่าไตเตอร์เพิ่มขึ้นมากกว่าหรือเท่ากับ 4 เท่าของครั้งที่ 1 (4 - Fold Rising) เช่น ครั้งแรกไตเตอร์ 100 ครั้งที่สองไตเตอร์ 400 เป็นต้น กรณีใดกรณีหนึ่งก็จะตัดสินว่าคนไข้นั้นเป็นโรคเลปโตสไปโรซิสจริง ในแง่การปฏิบัติจะพบว่าการรักษาผู้ป่วยหากรอผลตรวจเหล่านี้มายืนยันการเป็นโรคจะไม่สามารถรักษาผู้ป่วยได้ทันเวลาที่

ดังนั้น ผู้วิจัยจึงทำงานวิจัยนี้เพื่อใช้เทคนิคทางสถิติต่างๆที่สามารถนำมาใช้ในการจำแนกกลุ่มโรคของผู้ป่วยซึ่งคาดว่าจะมีประสิทธิภาพดีกว่าการใช้ตัวแบบที่อิงเกณฑ์ขององค์การอนามัยโลก โดยในการศึกษานี้จะเลือกทำการศึกษาเปรียบเทียบระหว่างเทคนิคการวิเคราะห์การจำแนก (Discriminant Analysis) และการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Model) และศึกษาเปรียบเทียบประสิทธิภาพกับตัวแบบที่อิงเกณฑ์ขององค์การอนามัยโลก โดยพิจารณาจากการจำแนกข้อมูลว่าเทคนิคใดสามารถทำการจำแนกได้ถูกต้องมากกว่า และใช้ตัวแปรต้นมาเป็นปัจจัยในการศึกษาน้อยกว่า เนื่องจากการวัดค่าตัวแปรต้นแต่ละตัวอาจมีค่าใช้จ่ายในการจะได้มาซึ่งค่าของตัวแปรนั้นๆแตกต่างกัน รวมทั้งตัวแปรบางตัวเช่นผลตรวจทางห้องปฏิบัติการต้องอาศัยระยะเวลาในการรอคอยผลเป็นเวลานาน อาจทำให้การรักษาโรคของผู้ป่วยล่าช้า รวมทั้งมีข้อจำกัดในการตรวจในโรงพยาบาลขนาดเล็ก ดังนั้นจึงทำงานวิจัยนี้ขึ้นเพื่อสร้างตัวแบบของโรคเลปโตสไปโรซิส ให้เกิดความสะดวกรวดเร็ว และแม่นยำเพื่อช่วยการตัดสินใจของแพทย์ในการวินิจฉัยโรค

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างตัวแบบในการจำแนกผู้ป่วย ระหว่างโรคเลปโตสไปโรซิสและโรคอื่นๆ เปรียบเทียบผลที่ได้จากการใช้เทคนิคการวิเคราะห์การจำแนก และการวิเคราะห์การถดถอยโลจิสติก เมื่อข้อมูลตัวแปรอิสระมีลักษณะผสม ทั้งต่อเนื่องและไม่ต่อเนื่อง
2. เพื่อสร้างตัวแบบในการจำแนกผู้ป่วย ระหว่างโรคเลปโตสไปโรซิสและโรคในกลุ่ม Rickettsiosis เปรียบเทียบผลที่ได้จากการใช้เทคนิคการวิเคราะห์การจำแนก และการวิเคราะห์การถดถอยโลจิสติก เมื่อข้อมูลตัวแปรอิสระมีลักษณะผสม ทั้งต่อเนื่องและไม่ต่อเนื่อง
3. เพื่อสร้างตัวแบบในการจำแนกผู้ป่วย ระหว่างโรคเลปโตสไปโรซิส โรคในกลุ่ม Rickettsiosis และโรคอื่นๆ เปรียบเทียบผลที่ได้จากการใช้เทคนิคการวิเคราะห์การจำแนก และการวิเคราะห์การถดถอยโลจิสติก เมื่อข้อมูลตัวแปรอิสระมีลักษณะผสม ทั้งต่อเนื่องและไม่ต่อเนื่อง

4. เพื่อเปรียบเทียบตัวแบบที่สร้างขึ้นกับตัวแบบที่อิงเกณฑ์อนามัยโลกโดยพิจารณาเฉพาะสองส่วนแรก เพื่อหาตัวแบบที่สามารถพยากรณ์ได้แม่นยำที่สุด โดยนำไปเทียบกับผลตรวจที่ได้จากห้องปฏิบัติการ

สมมติฐานของการวิจัย

สมมติฐานในการวิจัยนี้คือ ผลการจำแนกข้อมูลเข้ากลุ่มถูกต้องสำหรับกรณีที่ 1 เมื่อใช้การจำแนกกลุ่ม โดยวิธีการวิเคราะห์การจำแนก หรือวิธีการวิเคราะห์การถดถอยโลจิสติก สูงกว่าผลการจำแนกถูกโดยใช้ตัวแบบที่อิงเกณฑ์อนามัยโลกเมื่อพิจารณาเฉพาะสองส่วนแรก

ขอบเขตของการวิจัย

1. พิจารณาเทคนิคการวิเคราะห์ทางสถิติที่ใช้ในการจำแนกกลุ่ม 2 เทคนิค คือ การวิเคราะห์การจำแนก และการวิเคราะห์การถดถอยโลจิสติกเท่านั้น โดยนำผลการจัดข้อมูลเข้ากลุ่มถูกต้องที่ได้จาก 2 เทคนิคนี้ไปเปรียบเทียบกับผลการจัดข้อมูลเข้ากลุ่มของตัวแบบที่อิงเกณฑ์องค์การอนามัยโลก โดยในการศึกษาจะไม่พิจารณาถึงปฏิสัมพันธ์ (Interaction Between Variables) ของตัวแปรอิสระ

2. เกณฑ์ที่ใช้ในการพิจารณาว่าตัวแบบใดมีประสิทธิภาพสูงกว่า จะพิจารณาจากเปอร์เซ็นต์การจำแนกถูกรวมและเปอร์เซ็นต์การจำแนกถูกกรณีเฉพาะสำหรับคนไข้ที่เมื่อเสร็จสิ้นโครงการแล้วได้รับการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส และอัตราการจำแนกผิดที่เห็นชัดเจน

3. การวิจัยนี้จะทำการพยากรณ์การเป็นหรือไม่เป็นโรคเลปโตสไปโรซิส โดยอาศัยประวัติของคนไข้ อาการทางคลินิกจากประวัติ การตรวจร่างกายแรกรับและการตรวจทางห้องปฏิบัติการแรกรับ

4. ตัวแปรตามที่ใช้ในการศึกษา แยกเป็น 3 กรณี คือ (1) กลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กับกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ (2) กลุ่มคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กับกลุ่มที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis (3) กลุ่มคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส , กลุ่มที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ

5. ตัวแปรอิสระที่ใช้ มี 27 ตัวได้แก่ X_1 (อายุ), X_2 (เพศ), X_3 (การมีหรือไม่มีปัจจัยเสี่ยง), X_4 (ปวดศีรษะ), X_5 (อ่อนเพลีย), X_6 (เบื่ออาหาร), X_7 (เจ็บคอ), X_8 (ไอ), X_9 (เจ็บหน้าอก), X_{10} (มีเลือดออกผิดปกติ), X_{11} (ความรู้สึกตัวผิดปกติ), X_{12} (อุณหภูมิร่างกาย), X_{13}

(ความดันโลหิตต่ำ), X_{14} (อักเสบที่ผนังคอหอย), X_{15} (เยื่อตาอักเสบ), X_{16} (เลือดออกใต้ตาขาว), X_{17} (กดเจ็บกล้ามเนื้อท้อง), X_{18} (ตับโต), X_{19} (ต่อมน้ำเหลืองโต), X_{20} (มีจุดเลือดออกใต้ผิวหนัง), X_{21} ($WBC \geq 12,000/cumm.$), X_{22} ($PMN > 75\%$), X_{23} (Platelets $< 100,000/cumm.$), X_{24} (Abnormal Urine Examination), X_{25} (Total Bilirubin > 3 mg/dl), X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase) และ X_{27} (Abnormal BUN/Cr)

6. การเลือกตัวแปรอิสระขั้นแรกได้จากคำแนะนำของแพทย์ผู้เชี่ยวชาญ อ้างอิงจากคู่มือวิชาการโรคเลปโตสไปโรซิส ของสำนักงานโครงการควบคุมโรคโรคเลปโตสไปโรซิส กรมควบคุมโรคติดต่อ กระทรวงสาธารณสุขและศึกษาจากบทความอื่นๆที่เกี่ยวข้อง จากนั้นนำตัวแปรอิสระที่เลือกแล้วมาทำการตรวจสอบความแตกต่างอย่างมีนัยสำคัญของค่าเฉลี่ยของแต่ละตัวระหว่างกลุ่มตามกรณีที่ทำการศึกษา

7. ข้อมูลที่นำมาศึกษาเป็นข้อมูลคนไขซึ่งตัดข้อมูลสูญหายทิ้งแล้วจากคนไข้ที่เข้าร่วมโครงการวิจัย “การศึกษาเปรียบเทียบประสิทธิภาพของยาเซฟโทรเฟอซิม และ คีอ็อกซิซัยคลิน ในการรักษาผู้ป่วยด้วยโรคเลปโตสไปโรซิสที่มีอาการรุนแรง” ซึ่ง ศ.พญ. ยุพิน ศุพทวมงคล โรงพยาบาลศิริราชทำการศึกษาพร้อมกับแพทย์และคณะผู้วิจัยโรงพยาบาลอื่นๆ

8. สำหรับทุกการทดสอบสมมติฐาน พิจารณาที่ระดับนัยสำคัญ 0.05

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ตัวแบบสำหรับกรณีที่ 1 ที่มีประสิทธิภาพในการจำแนกดีกว่าตัวแบบที่ได้จากการอิงเกณฑ์องค์การอนามัยโลกสองส่วนแรก
2. ทำการศึกษาเพื่อเปรียบเทียบประสิทธิภาพของตัวแบบที่ได้จากเทคนิคการวิเคราะห์การจำแนกและตัวแบบการวิเคราะห์การถดถอยโลจิสติก ว่าความสามารถในการจำแนกของทั้งสองเทคนิคจำแนกกลุ่มได้เหมือนหรือแตกต่างกันอย่างไรเมื่อตัวแปรอิสระเป็นแบบผสม และตัวแปรตามเป็นแบบเชิงกลุ่ม
3. ตัวแบบที่ได้มีประโยชน์สำหรับแพทย์ที่รักษาโรคเลปโตสไปโรซิส เพื่อช่วยให้แพทย์สามารถวินิจฉัยได้เร็วขึ้นและทำให้การรักษาผู้ป่วยเป็นไปอย่างรวดเร็ว ลดความเสี่ยงของผู้ป่วย
4. เทคนิคที่ทำให้ได้ตัวแบบที่มีประสิทธิภาพดีในการศึกษานี้สามารถนำไปปรับ เพื่อสร้างตัวแบบเพื่อช่วยในการวินิจฉัยโรคอื่นๆได้เพิ่มเติม เมื่อทำการศึกษาในลักษณะเดียวกัน

นิยามศัพท์

1. การประมาณอัตราความผิดพลาด หมายถึง สัดส่วนของจำนวนตัวอย่างที่ถูกจัดเข้ากลุ่มไม่ถูกต้องมีหลายวิธีการ แต่สำหรับงานวิจัยนี้เลือกใช้ อัตราการจำแนกผิดที่เห็นชัดเจน Apparent Error Rate (APER) เท่านั้น

2. ผู้ป่วยที่เป็นโรคเลปโตสไปโรสิส หมายถึง ผู้ป่วยในโครงการวิจัยที่ผลตรวจทางห้องปฏิบัติการยืนยันว่าเป็นโรคเลปโตสไปโรสิสจริง โดยที่ผลตรวจทางห้องปฏิบัติการที่นำมาใช้เป็นมาตรฐานในการวินิจฉัยโรคได้แก่ ผลเพาะเชื้อ และผลตรวจซีรัม

3. ผู้ป่วยที่เป็นโรคในกลุ่ม Rickettsiosis หมายถึงผู้ป่วยในโครงการวิจัยที่ผลตรวจทางห้องปฏิบัติการยืนยันว่าเป็นโรคใดโรคหนึ่งในกลุ่มนี้ได้แก่ scrub typhus , murine typhus , Q Fever และ *R. Helvetica* Infection

4. ผู้ป่วยที่เป็นโรคอื่นๆ หมายถึง ผู้ป่วยในโครงการวิจัยที่ผลตรวจทางห้องปฏิบัติการไม่พบว่าโรคเลปโตสไปโรสิส หรือโรคในกลุ่ม Rickettsiosis โดยที่ผู้ป่วยอาจป่วยเป็นโรคอื่นๆนอกเหนือจากโรคนี้

5. ตัวแบบที่สร้างขึ้น หมายถึง ตัวแบบที่ใช้เทคนิคทางสถิติช่วยในการสร้างเพื่อนำไปใช้วินิจฉัยโรค โดยในงานวิจัยนี้พิจารณา 2 เทคนิคได้แก่ เทคนิคการวิเคราะห์การจำแนก และ เทคนิคการวิเคราะห์การถดถอยโลจิสติก

6. ตัวแบบอิงเกณฑ์อนามัยโลก หมายถึง ตัวแบบการวินิจฉัยโรคเลปโตสไปโรสิส ที่อิงจากเกณฑ์ขององค์การอนามัยโลก (WHO) ได้จากคู่มือวิชาการโรคเลปโตสไปโรสิส ของสำนักงานโครงการควบคุมโรคเลปโตสไปโรสิส กรมควบคุมโรคติดต่อ กระทรวงสาธารณสุข

7. ปัจจัยเสี่ยง หมายถึง ปัจจัยที่ทำให้คนไข้เสี่ยงต่อการเป็น โรคเลปโตสไปโรสิสได้แก่ ลุยน้ำหรือแช่น้ำโดยไม่มีกำบัง , เลี้ยงสัตว์ , มีประวัติดักจับหนู หรือบริเวณบ้านมีหนู เป็นต้น น่าจะมีโอกาสเป็นโรคเลปโตสไปโรสิสมากกว่าผู้ป่วยที่ไม่มีปัจจัยเสี่ยง เกณฑ์การพิจารณาคือหากผู้ป่วยมีอย่างน้อย 1 ข้อจากข้อย่อยต่อไปนี้จะถือว่ามีความเสี่ยง

7.1 สองสัปดาห์ที่ผ่านมาเคยลุยน้ำหรือแช่น้ำโดยไม่ใส่บู๊ท

7.2 สองสัปดาห์ที่ผ่านมาดักจับหนู หรือใช้ยาเบื่อหนู

7.3 เลี้ยงสุนัขในบ้าน

7.4 เลี้ยงสัตว์เศรษฐกิจในบริเวณบ้านเช่น วัว, ควาย, หมู, เป็ดไก่ เป็นต้น

8. อาการเลือดออกผิดปกติ หมายถึง มีประวัติเลือดออกผิดปกติ ซึ่งมักเกิดในผู้ป่วยโรคเลปโตสไปโรสิสโดยเฉพาะผู้ป่วยที่มีอาการรุนแรงและมีอาการตัวเหลือง เช่น ผื่นเลือดออก

(Purpuric Spots) หรือเสมหะเป็นเลือด บางครั้งอาจพบเพียงผื่นเลือดออก 2-3 แห่ง ที่หน้าอก ท้อง หรือแขน เชื่อว่าเกิดจากเส้นเลือดฝอยที่เปราะ

9. ความรู้สึกตัวผิดปกติ หมายถึง มีความผิดปกติเกี่ยวกับการรู้สึกตัว เช่น มีอาการ ลึบสน , ชัก หรือแขนขาอ่อนแรง

10. ความดันโลหิตต่ำ หมายถึง การมีค่า Systolic Blood Pressure ต่ำกว่า 90 mmHg.

11. อักเสบที่ผนังคอหอย (Injected Pharynx) หมายถึง แพทย์ตรวจพบอาการอักเสบ แดงที่ผนังคอหอย

12. เยื่อบุตาอักเสบ (Conjunctival Injection) หมายถึง แพทย์ตรวจพบอาการเยื่อบุตา ขาวอักเสบ ซึ่งเป็นอาการแสดงอย่างหนึ่งของโรค เกิดขึ้นในตาทั้งสองข้าง มักเกิดภายใน 3 วัน แรกของโรคและอยู่ได้นานตั้งแต่ 1 วันถึง 1 สัปดาห์ อาจพบร่วมกับอาการเลือดออกที่ตาขาวข้าง เดียวหรือสองข้างก็ได้ แต่ไม่ใช่ตาแดงที่เกิดจากการอักเสบชนิดเป็นหนอง

13. เลือดออกใต้ตาขาว (Subconjunctival hemorrhage) หมายถึง แพทย์ตรวจพบ เลือดออกใต้ตาขาว

14. กดเจ็บกล้ามเนื้อน่อง (Calf tenderness) หมายถึง แพทย์ตรวจพบการกดเจ็บที่ กล้ามเนื้อน่อง

15. ตับโต (Hepatomegaly) หมายถึง แพทย์ตรวจพบ ตับโต

16. ต่อม้ำเหลืองโต (Lymphadenopathy) หมายถึง แพทย์ตรวจพบ ต่อม้ำเหลืองโต ที่ตำแหน่งต่างๆ หรือหลายตำแหน่ง

17. มีจุดเลือดออกใต้ผิวหนัง (Petichiae) หมายถึง แพทย์ตรวจพบมีจุดเลือดออกใต้ ผิวหนัง

18. White Blood Cell (WBC) หมายถึง จำนวนเม็ดเลือดขาว

19. % PMN หมายถึง สัดส่วนที่พบเม็ดเลือดขาวชนิด neutrophils

20. Plateletes หมายถึง จำนวนเกล็ดเลือดค่าปกติคือ มากกว่า 100,000 หน่วยต่อ cumm.

21. ค่าผลตรวจปัสสาวะผิดปกติ (Abnormal Urine Examination) หมายถึง ผลตรวจที่ ผิดปกติอย่างใดอย่างหนึ่งในปัสสาวะ เช่น มีไข่ขาวในปัสสาวะ (Albuminuria) มากกว่า 1+ หรือมีตะกอน (Urine Sediment) ในปัสสาวะ เป็นต้น

22. ค่าบิลิรูบินรวม (Total Bilirubin) หมายถึง ค่าผลตรวจทางห้องปฏิบัติการที่บ่ง บอกรสภาวะการทำงานของตับ ค่าปกติคือไม่เกิน 3 mg/dl ผู้ป่วยที่มีค่าบิลิรูบินผิดปกติจะเกิด อาการดีซ่าน

23. ค่า SGOT/SGPT/Alkaline phosphatase ผิดปกติ หมายถึง ค่าผลตรวจทางห้องปฏิบัติการ SGOT หรือ SGPT หรือ Alkaline phosphatase ค่าใดค่าหนึ่งผิดปกติ ผลตรวจเหล่านี้บ่งบอกสภาพการทำงานของตับ ค่าปกติของ SGOT และ SGPT คือไม่เกิน 120 U/L ส่วนค่าปกติของ Alkaline phosphatase คือไม่เกิน 350 U/L

24. ค่า BUN/Cr ผิดปกติ หมายถึง ค่าผลตรวจทางห้องปฏิบัติการ BUN หรือ Cr ค่าใดค่าหนึ่งผิดปกติ ผลตรวจทั้งสองตัวนี้บ่งบอกสภาพการทำงานของไต โดยที่ค่าปกติของ BUN คือค่าไม่เกิน 40 mg/dl ส่วนค่าปกติของ Cr คือมีค่าไม่เกิน 2.5 mg/dl

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิเคราะห์การจำแนก (Discriminant Analysis)

การวิเคราะห์การจำแนก เป็นเทคนิคการวิเคราะห์ตัวแปรพหุ (Multivariate Analysis) ที่มุ่งสร้างสมการหรือตัวแปรมิติขึ้นมาใหม่จากตัวแปรอิสระหลายตัวที่เก็บข้อมูลมา เพื่อให้ตัวแปรใหม่สามารถจำแนกกลุ่มหรือระดับของตัวแปรตามให้ได้มากที่สุด มีส่วนคล้ายกับการสร้างสมการทำนายแบบถดถอยพหุคูณ (Multiple Regression Analysis) ที่สร้างสมการจากตัวแปรอิสระหลายตัว เพื่อทำนายค่าของตัวแปรตามตัวเดียว จุดต่างกันอยู่ตรงที่ระดับการวัดของตัวแปรตาม การวิเคราะห์การถดถอยตัวแปรตามเป็นตัวแปรแบบต่อเนื่อง แต่ตัวแปรตามของการวิเคราะห์การจำแนก เป็นตัวแปรเชิงกลุ่ม (Categorical Variable)

เป้าหมายของการใช้การวิเคราะห์การจำแนก ก็เพื่อหาตัวจำแนก (Discriminants) ซึ่งสามารถจำแนกค่าสังเกตให้ถูกจัดเข้ากลุ่มได้ถูกต้องมากที่สุด โดยจะพิจารณาจากการจัดเข้ากลุ่มผิดพลาดที่น้อยที่สุด วัดได้จากค่าอัตราความผิดพลาด (Error Rate) หรือความน่าจะเป็นในการจัดเข้ากลุ่มผิด (Misclassification Probability)

ในการวิเคราะห์การจำแนก มีข้อตกลง (Assumptions) ซึ่งต้องทำการทดสอบดังนี้

1. ตัวแปรตาม เป็นตัวแปรเชิงกลุ่ม โดยจะมีตั้งแต่ 2 กลุ่มขึ้นไป
2. ตัวอย่างในแต่ละกลุ่มมีจำนวนต่ำสุด 2 ค่าสังเกต
3. จำนวนตัวแปรจำแนก (p) ที่ใช้ในการวิเคราะห์มีจำนวนน้อยกว่าตัวอย่างทั้งหมดลบสองหรือ $0 < p < n-2$
4. ตัวแปรอิสระหรือตัวแปรจำแนก (Discriminating Variables) เป็นได้ทั้งตัวแปรต่อเนื่อง หรือตัวแปรเชิงกลุ่มโดยที่ถ้าเป็นตัวแปรเชิงกลุ่ม ต้องแปลงให้เป็นตัวแปรหุ่น (Dummy Variables) ก่อน
5. ตัวแปรอิสระแต่ละตัวต้องไม่มีความสัมพันธ์เชิงเส้นตรงซึ่งกันและกัน
6. เมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระแต่ละกลุ่มเท่ากัน
7. ตัวแปรจำแนกของตัวอย่างแต่ละกลุ่มสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติพหุ (Multivariate Normal Distribution)

ข้อตกลงดังกล่าวเป็นข้อกำหนดที่สำคัญที่ต้องทำการตรวจสอบก่อนใช้เทคนิคการวิเคราะห์การจำแนก แต่หากในการศึกษาต้องการเน้นที่การจำแนกจะมีข้อกำหนดทางสถิติน้อยกว่าการศึกษาแบบเน้นการนำไปอนุมาน ถ้าการจำแนกเข้ากลุ่มถูกต้องเป็นวัตถุประสงค์หลักของงานวิจัยและตัวอย่างที่ใช้มีขนาดใหญ่แล้วข้อกำหนดเกี่ยวกับการแจกแจงแบบปกติพหุก็เป็นสิ่งที่สามารถละเว้นได้ (Barbara and Linda 2001)

ส่วนข้อสมมติเกี่ยวกับความเท่ากันของเมทริกซ์ความแปรปรวนร่วมของกลุ่มที่จะจำแนกนั้น หากข้อมูลไม่สอดคล้องกับข้อสมมติดังกล่าว เทคนิคนี้ยังคงถูกนำมาใช้เพียงแต่ต้องทำการปรับเกี่ยวกับเมทริกซ์ความแปรปรวนร่วมที่นำมาใช้โดยการเปลี่ยนมาพิจารณาเมทริกซ์ความแปรปรวนร่วมแบบแยก (Separate variance-covariance matrices) แทนเมทริกซ์ความแปรปรวนร่วมแบบรวม (Pooled variance-covariance matrices) ได้ ซึ่งจะสอดคล้องกับลักษณะของข้อมูลมากกว่า (Barbara and Linda 2001 : 462-492)

วัตถุประสงค์ของการวิเคราะห์การจำแนก

1. เป็นการค้นหาว่าตัวแปรอิสระที่ใช้ทำนายกลุ่มนั้น มีการวัดร่วมกันอยู่ที่ด้าน แต่ละด้านเกิดจากการผสมระหว่างตัวแปรอิสระแต่ละตัวด้วยน้ำหนักที่เหมาะสมเป็นเท่าใดจึงจะทำให้เกิดความแตกต่างระหว่างกลุ่มตัวแปรตามได้สูงสุด ขั้นนี้คือการสร้างสมการหรือฟังก์ชันผสมเชิงเส้นตรงที่ทำให้ค่าเฉลี่ยของคะแนนในด้านที่จัดใหม่มีค่าแตกต่างกันมากที่สุด เรียกสมการผสมเชิงเส้นตรงด้วยน้ำหนักที่เหมาะสมว่า “ฟังก์ชันการจำแนก” (Discriminant Function) หรือ “สมการการจำแนกกลุ่ม” สมการการจำแนกกลุ่มมีทั้งสมการในรูปคะแนนดิบ และคะแนนในรูปแบบมาตรฐาน

2. ทำนายหรือจัดกลุ่มค่าสังเกตหรือสิ่งของว่าควรอยู่ในกลุ่มใด โดยอาศัยสมการจำแนกที่สร้างขึ้น

ฟังก์ชันการจำแนก สำหรับกรณีสองกลุ่ม

สมมติให้ประชากรสองกลุ่มที่จะนำมาเปรียบเทียบกันมีเมทริกซ์ความแปรปรวนร่วม (Covariance matrix) ที่เท่ากันคือ Σ โดยที่

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \Lambda & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2p} \\ M & M & O & M \\ \sigma_{p1} & \Lambda & \Lambda & \sigma_{pp} \end{bmatrix}$$

แต่มีเวกเตอร์ค่าเฉลี่ยแตกต่างกันอย่างชัดเจนเป็น μ_1 และ μ_2 ตัวอย่างที่ใช้ให้เป็น $x_{11}, x_{12}, \dots, x_{1n_1}$ และ $x_{21}, x_{22}, \dots, x_{2n_2}$ ซึ่งเป็นตัวอย่างจากประชากร 2 กลุ่ม โดยปกติแล้วแต่ละเวกเตอร์ x_{ij} ประกอบด้วยค่าตัวแปร p ตัว ฟังก์ชันการจำแนกกลุ่มเป็นผลรวมเชิงเส้นของตัวแปรทั้ง p ตัวนี้ ซึ่งทำให้ระยะห่างระหว่างเวกเตอร์ค่าเฉลี่ยของกลุ่มทั้งสองกลุ่ม(ซึ่งแปลงแล้ว) มีค่าสูงที่สุด ผลรวมเชิงเส้น $z = \mathbf{a}'\mathbf{x}$ จะแปลงให้เวกเตอร์ค่าสังเกตแต่ละเวกเตอร์เป็นตัวเลข

$$z_{1j} = \mathbf{a}'\mathbf{x}_{1j} = a_1x_{1j1} + a_2x_{1j2} + \dots + a_px_{1jp} \quad , \quad j=1,2,\dots,n_1$$

$$z_{2j} = \mathbf{a}'\mathbf{x}_{2j} = a_1x_{2j1} + a_2x_{2j2} + \dots + a_px_{2jp} \quad , \quad j=1,2,\dots,n_2$$

มหาวิทยาลัยศิลปากร ส่วนนลิขสิทธิ์

เพราะฉะนั้น เวกเตอร์ค่าสังเกต $n_1 + n_2$ ใน 2 ตัวอย่างคือ

$$\begin{array}{cc} \mathbf{x}_{11} & \mathbf{x}_{21} \\ \mathbf{x}_{12} & \mathbf{x}_{22} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \mathbf{x}_{1n_1} & \mathbf{x}_{2n_2} \end{array}$$

เมื่อแปลงเป็นตัวเลข จะเป็นดังนี้

$$\begin{array}{cc} z_{11} & z_{21} \\ z_{12} & z_{22} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ z_{1n_1} & z_{2n_2} \end{array}$$

จะหาค่าเฉลี่ย $\bar{z}_1 = \sum_{j=1}^{n_1} z_{1j} / n_1 = \mathbf{a}'\bar{\mathbf{x}}_1$ และ $\bar{z}_2 = \mathbf{a}'\bar{\mathbf{x}}_2$ และต้องการหา
 เวกเตอร์ \mathbf{a} ซึ่งทำให้ค่าความแตกต่างมาตรฐาน (Standardized difference) หรือ $(\bar{z}_1 - \bar{z}_2) / s_z$
 มีค่ามากที่สุด เนื่องจากค่า $(\bar{z}_1 - \bar{z}_2) / s_z$ อาจเป็นจำนวนลบได้เราจึงใช้ค่ากำลังสองของ
 ระยะทางแทน $(\bar{z}_1 - \bar{z}_2)^2 / s_z^2$ จะได้ว่า

$$\frac{(\bar{z}_1 - \bar{z}_2)^2}{s_z^2} = \frac{[\mathbf{a}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)]^2}{\mathbf{a}'\mathbf{S}_{pl}\mathbf{a}} \quad (1)$$

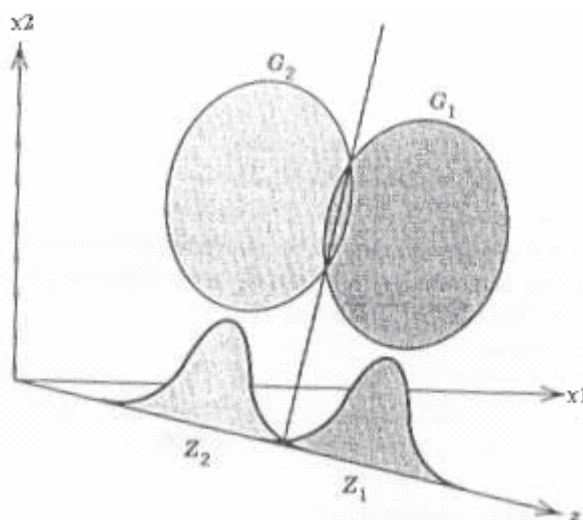
โดยที่ \mathbf{S}_{pl} คือเมทริกซ์ความแปรปรวนร่วมของตัวอย่าง ค่าในสมการที่ (1) จะมี
 ค่าสูงสุดเมื่อ

$$\mathbf{a} = \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (2)$$

หรือเมื่อ \mathbf{a} เป็นผลคูณ (Multiple) ใดๆของ $\mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ เนื่องจากผลคูณของ
 $\mathbf{a} = \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ จะทำให้สมการที่ (1) มีค่ามากที่สุด และจะสังเกตว่าสำหรับ \mathbf{S}_{pl}^{-1} ที่หาค่าได้
 เราจะได้ว่า $n_1 + n_2 - 2 > p$ เมื่อกำหนดให้ $\mathbf{a} = \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ จะได้เส้นที่ลากกลางระหว่าง
 $\bar{\mathbf{x}}_1$ และ $\bar{\mathbf{x}}_2$ เพราะวาระยะทางกำลังสอง $(\bar{z}_1 - \bar{z}_2)^2 / s_z^2$ เท่ากับเป็นระยะทางมาตรฐาน
 ระหว่าง $\bar{\mathbf{x}}_1$ และ $\bar{\mathbf{x}}_2$ จะเห็นว่าเมื่อแทนค่าสมการ (2) ใน สมการ (1) จะได้ว่า

$$\frac{(\bar{z}_1 - \bar{z}_2)^2}{s_z^2} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (3)$$

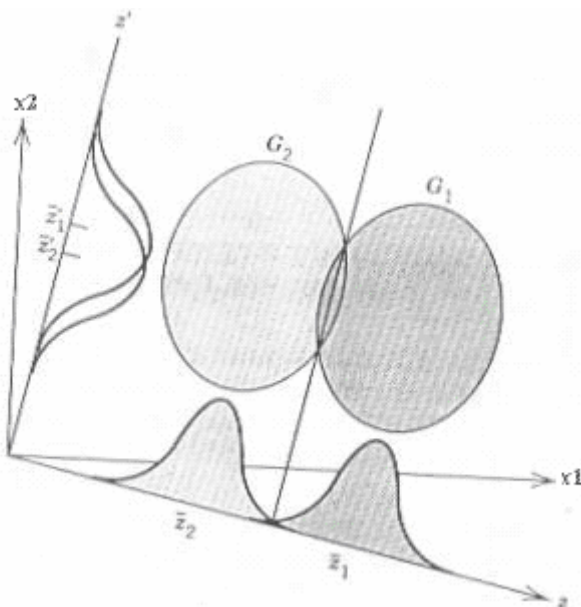
สำหรับ $z = \mathbf{a}'\mathbf{x}$ เมื่อ $\mathbf{a} = \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ ดังนั้น $(\bar{z}_1 - \bar{z}_2)^2 / s_z^2 = \mathbf{a}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$
 , และแนวทางอื่นๆนอกเหนือจากใช้ $\mathbf{a} = \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ ก็จะทำให้ผลความแตกต่างระหว่าง $\mathbf{a}'\bar{\mathbf{x}}_1$
 และ $\mathbf{a}'\bar{\mathbf{x}}_2$ ที่ต่ำกว่า



รูปที่ 1 การวิเคราะห์การจำแนกสำหรับ 2 กลุ่ม

รูปที่ 1 แสดงให้เห็นการแบ่งกลุ่มที่มีการแจกแจงแบบปกติของสองตัวแปร ($p=2$) โดยใช้ Discriminant Function $z = \mathbf{a}'\mathbf{x}$ แสดงตัวอย่างให้เห็นในรูปของ 1 มิติ เมื่อ \mathbf{a} แสดงดังสมการที่ (2) ในตัวอย่างนี้เมทริกซ์ความแปรปรวนร่วมของประชากรทั้งสองกลุ่มเท่ากัน เมื่อผลรวมเชิงเส้น $z_{ij} = \mathbf{a}'\mathbf{x}_{ij} = a_1x_{ij1} + a_2x_{ij2}$ ซึ่งปรับตามจุด \mathbf{x}_{ij} ซึ่งฉายลงบนเส้นของการแบ่งที่เหมาะสมที่สุดสำหรับกลุ่มทั้งสอง เนื่องจากตัวแปรทั้งสอง x_1 และ x_2 มีการแจกแจงแบบปกติของสองตัวแปร (Bivariate Normal) ผลรวมเชิงเส้นจึงมีการแจกแจงแบบปกติของตัวแปรเดียว (Univariate Normal) ดังนั้นจึงแสดงให้เห็นด้วย 2 ฟังก์ชันความหนาแน่นแบบปกติของตัวแปรเดียว (Two Univariate Normal Densities) ระหว่างเส้นซึ่งแทน z

จุดที่เส้นตัด Intersection ของวงรีทั้งสอง ตัดเส้น z (Discriminant Function) เป็นจุดที่มีการแบ่งแยกมากที่สุด นั่นคือเป็นจุดที่ซ้อนกันน้อยที่สุด ของจุดที่ Project ลงบนเส้น ถ้าประชากรสองกลุ่มมีการแจกแจงแบบปกติของตัวแปรพหุ มีเมทริกซ์ความแปรปรวนร่วมเป็น Σ เหมือนที่แสดงตัวอย่างในรูปที่ 1 แสดงให้เห็นว่าผลรวมของจุดที่ซ้อนกันนี้คือฟังก์ชันของระยะทางระหว่างเวกเตอร์ค่าเฉลี่ย $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ แล้วยังสามารถแสดงได้ว่าทุกกลุ่มที่สามารถแยกได้จะแสดงออกมาในรูปของมิติใหม่



รูปที่ 2 การจำแนกโดยใช้ฟังก์ชันการจำแนก (discriminant function)

ในรูปที่ 2 จะแสดงตัวอย่างว่าการจำแนกกลุ่มที่เหมาะสมที่สุดได้มาจากฟังก์ชันการจำแนกกลุ่ม เปรียบเทียบโดยทำการฉายในแนวทางอื่น ๆ ซึ่งแสดงด้วย z' จะให้ค่าระยะทางมาตรฐานระหว่างค่าเฉลี่ยที่ผ่านการแปลง \bar{z}'_1 และ \bar{z}'_2 ที่น้อยกว่า และพื้นที่การซ้อนระหว่างจุดที่ Project ที่มากกว่า จะสังเกตว่าการกระจายของจุดที่ฉายในมิติ z' จะมากกว่า ดังนั้น $s_{z'}$ จะมีค่ามากและ $(\bar{z}'_1 - \bar{z}'_2)/s_{z'}$ จะมีค่าน้อย

ตัวอย่างการใช้การวิเคราะห์การจำแนก

ตัวอย่าง ข้อมูลลึ่ที่ผลิตในอุณหภูมิที่แตกต่างกัน 2 ระดับแสดงในตารางที่ 1 ตัวแปรได้แก่ x_1 = คะแนนการผลิต (yield point) และ x_2 = ความทนทานสูงสุด (ultimate strength) จากข้อมูลคำนวณค่าต่างๆ ได้ดังนี้ (Rencher 1995 : 299-301)

$$\bar{\mathbf{x}}_1 = \begin{pmatrix} 36.4 \\ 62.6 \end{pmatrix} \quad \bar{\mathbf{x}}_2 = \begin{pmatrix} 39.0 \\ 60.4 \end{pmatrix} \quad \mathbf{S}_{pl} = \begin{pmatrix} 7.92 & 5.68 \\ 5.68 & 6.29 \end{pmatrix}$$

ตารางที่ 1 ข้อมูลแสดงคะแนนการผลิตและความทนทานสูงสุด ของล้อที่ผลิตในอุณหภูมิที่แตกต่างกัน 2 ระดับ

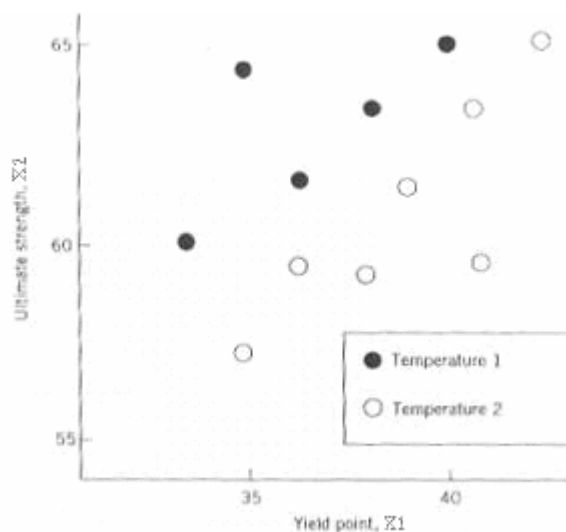
ระดับอุณหภูมิที่ 1		ระดับอุณหภูมิที่ 2	
x_1	x_2	x_1	x_2
33	60	35	57
36	61	36	59
35	64	38	59
38	63	39	61
40	65	41	63
		43	65
		41	59

นำข้อมูลจากตารางมาสร้างกราฟได้ดังรูปที่ 3 จะเห็นว่าถ้าจุด Project บนแกน x_1 หรือ x_2 แกนใดแกนหนึ่งแล้วจะเกิดการซ้อนค่อนข้างมาก การนำเสนอในรูปแบบของมิติเดียวซึ่งจุดถูก project นี้ก็คือฟังก์ชันจำแนกนั่นเอง

$$z = \mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 = -1.633x_1 + 1.819x_2$$

เมื่อ \mathbf{a} ได้มาจาก

$$\mathbf{a} = \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \begin{pmatrix} -1.633 \\ 1.819 \end{pmatrix}$$



รูปที่ 3 กราฟระหว่างคะแนนการผลิตและทนทานสูงสุดที่ระดับอุณหภูมิ 2 ระดับ

จากรูปที่ 3 เห็นได้ชัดเจนว่าข้อมูลทั้ง 2 กลุ่มสามารถแยกออกจากกันได้ ถ้า project ในทิศทางที่เหมาะสมดังที่เสนอในรูปที่ 1 คือจะไม่เกิดการซ้อนทับขึ้น ค่าของจุดที่ถูก Project หาได้จากการคำนวณค่า z สำหรับแต่ละเวกเตอร์ค่าสังเกต \mathbf{x} ใน 2 กลุ่ม ผลลัพธ์แสดงดังตารางต่อไปนี้ ซึ่งการแบ่งกลุ่มเห็นได้ที่ได้จากการใช้ฟังก์ชันจำแนกนี้ สามารถทำได้อย่างชัดเจน

ตารางที่ 2 แสดง Discriminant function $z = -1.633x_1 + 1.819x_2$ คำนวณจากข้อมูลในตารางที่ 1

ระดับอุณหภูมิที่ 1	ระดับอุณหภูมิที่ 2
55.24	46.52
52.17	48.53
59.24	45.26
52.54	47.27
52.91	47.64
	48.01
	40.37

การจัดหน่วยตัวอย่างเข้ากลุ่ม (Allocation of observations to groups)

ขั้นตอนต่อไปสำหรับการวิเคราะห์จำแนกก็คือการจัดหน่วยตัวอย่างเข้ากลุ่มที่ได้แบ่งไว้ หรือก็คือขั้นตอนการทำนายนั่นเอง ในขั้นตอนนี้หน่วยตัวอย่างซึ่งยังไม่ทราบกลุ่มจะถูกจัดให้เข้ากลุ่มโดยอาศัยหลักเกณฑ์ของเวกเตอร์ค่าวัดทั้ง p ตัวหรือเมทริกซ์ \mathbf{x} ที่สัมพันธ์กับค่าวัดนั้นๆ ในการจำแนกกลุ่มเราจำเป็นต้องมีเวกเตอร์ค่าสังเกตของตัวอย่างที่ได้ก่อนหน้าจากแต่ละกลุ่ม วิธีการหนึ่งก็คือการเปรียบเทียบ \mathbf{x} กับ $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_k$ ของตัวอย่างทั้ง k ตัว และจัดหน่วยตัวอย่างเข้ากลุ่มซึ่งมี $\bar{\mathbf{x}}_i$ ใกล้เคียงกับ \mathbf{x} มากที่สุด ด้วยเหตุนี้ถ้าเรามองว่าค่าวัดของ \mathbf{x} เปรียบเสมือน profile ของหน่วยตัวอย่างแล้วเราจะเปรียบเทียบ profile นี้กับ profile ของแต่ละกลุ่ม เราจะจัดหน่วยตัวอย่างเข้ากลุ่มที่ profile \mathbf{x} นี้คล้ายคลึงกันมากที่สุด

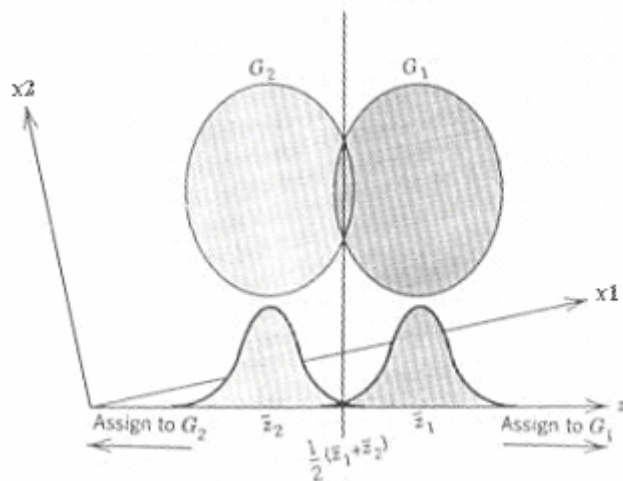
การจัดหน่วยตัวอย่างเข้ากลุ่มกรณี 2 กลุ่ม

สำหรับกรณีของ 2 ประชากร เมื่อเรามีหน่วยตัวอย่างที่ไม่ทราบว่าจะจัดเข้ากลุ่มไหนระหว่างประชากร 2 กลุ่มนั้น สารสนเทศที่เราถืออยู่ประกอบด้วยตัวแปรที่มีประโยชน์ p ตัวในเวกเตอร์ \mathbf{x} เมื่อมีประชากร 2 กลุ่มเราสามารถใช่วิธีการจำแนกของ Fisher ได้ วิธีการแบ่งกลุ่มของ Fisher นี้มีข้อสมมติเบื้องต้นว่าประชากรทั้งสองกลุ่มจะต้องมีเมทริกซ์ความแปรปรวนร่วมที่เท่ากัน แต่ไม่ได้กำหนดให้มีการแจกแจงแบบปกติ เมื่อเราได้ตัวอย่างจากทั้ง 2 ประชากรมาแล้วจะคำนวณหา $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$ และ \mathbf{S}_{pl} จากนั้นก็สามารถจำแนกได้โดยง่ายเมื่อใช้ฟังก์ชันจำแนก

$$z = \mathbf{a}'\mathbf{x} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1} \mathbf{x}, \quad (4)$$

เมื่อ \mathbf{x} เป็นเวกเตอร์ของค่าวัดหน่วยตัวอย่างใหม่ที่ต้องการทำการจัดเข้ากลุ่มใดกลุ่มหนึ่งระหว่าง 2 กลุ่มของประชากร

แทนกลุ่มทั้งสองด้วย G_1 และ G_2 หากค่าของสมการ (4) สำหรับแต่ละค่าสังเกต \mathbf{x}_{ij} จะได้ $z_{11}, z_{12}, \dots, z_{1n_1}$, จาก $\bar{z}_1 = \sum_{j=1}^{n_1} z_{1j} / n_1 = \mathbf{a}'\bar{\mathbf{x}}_1 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1} \bar{\mathbf{x}}_1$ และในทำนองเดียวกัน $\bar{z}_2 = \mathbf{a}'\bar{\mathbf{x}}_2$ วิธีการแบ่งกลุ่มเชิงเส้นของ Fisher (Fisher's linear classification procedure) จัดให้ \mathbf{x} อยู่ใน G_1 ถ้า $z = \mathbf{a}'\mathbf{x}$ เข้าใกล้ \bar{z}_1 มากกว่า \bar{z}_2 และจัดให้ \mathbf{x} อยู่ใน G_2 ถ้า z เข้าใกล้ \bar{z}_2 มากกว่า แสดงได้ดังรูปที่ 4



รูปที่ 4 วิธีของ Fisher สำหรับการจำแนกกลุ่ม 2 กลุ่ม

ดังที่แสดงตัวอย่างในรูปที่ 4 เราจะเห็นว่า z เข้าใกล้ \bar{z}_1 มากกว่าถ้า

$$z > \frac{1}{2}(\bar{z}_1 + \bar{z}_2) \quad (5)$$

มหาวิทยาลัยศิลปากร ส่วนวิจัยศิลปกรรมศาสตร์
 ซึ่งเป็นจริงในกรณีทั่วไป เนื่องจาก \bar{z}_1 มากกว่า \bar{z}_2 เสมอ แสดงให้เห็นได้ง่ายๆ
 ดังนี้

$$\bar{z}_1 + \bar{z}_2 = \mathbf{a}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) > 0 \quad (6)$$

เพราะว่า \mathbf{S}_{pl}^{-1} เป็นบวกอย่างแน่นอน (positive definite) ดังนั้น $\bar{z}_1 > \bar{z}_2$
 นอกจากนี้ \mathbf{a} อยู่ในรูป $\mathbf{a}' = (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)' \mathbf{S}_{pl}^{-1}$ แล้ว $\bar{z}_2 - \bar{z}_1$ จึงมีค่าเป็นบวก เนื่องจาก
 $\frac{1}{2}(\bar{z}_1 + \bar{z}_2)$ เป็นจุดกึ่งกลางดังนั้นเมื่อ $z > \frac{1}{2}(\bar{z}_1 + \bar{z}_2)$ จึงแสดงว่า z เข้าใกล้ \bar{z}_1 จะสังเกต
 ว่าระยะทางจาก \bar{z}_1 ถึง \bar{z}_2 เหมือนกับระยะทางระหว่าง $\bar{\mathbf{x}}_1$ กับ $\bar{\mathbf{x}}_2$

เพื่อแสดงสูตรของหลักการแบ่งกลุ่มในรูปของ \mathbf{x} ขั้นตอนแรกเราจะเขียน
 $\frac{1}{2}(\bar{z}_1 + \bar{z}_2)$ ในรูปของ

$$\frac{1}{2}(\bar{z}_1 + \bar{z}_2) = \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \quad (7)$$

แล้วหลักของการแบ่งกลุ่มจะกลายเป็น : จัดให้ \mathbf{x} เข้ากลุ่ม G_1 ถ้า

$$\mathbf{a}'\mathbf{x} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1}\mathbf{x} > \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \quad (8)$$

และจัดให้ \mathbf{x} เข้ากลุ่ม G_2 ถ้า

$$\mathbf{a}'\mathbf{x} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1} \mathbf{x} < \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pl}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \quad (9)$$

การที่ Fisher ใช้หลักตามสมการที่ (8) และ (9) จะพบว่าใกล้เคียงกับการไม่ใช้พารามิเตอร์ เนื่องจากไม่มีข้อสมมติเกี่ยวกับการแจกแจง อย่างไรก็ตามถ้าประชากรสองกลุ่มมีการแจกแจงแบบปกติที่มีเมทริกซ์ความแปรปรวนร่วมเท่ากัน แล้ววิธีการนี้จะเหมาะสมที่สุด นั่นคือความน่าจะเป็นของการจำแนกผิดจะมีค่าน้อยที่สุด

วิธีการอื่นๆสำหรับจัดข้อมูลเข้ากลุ่ม

นอกจากวิธีที่กล่าวมาแล้วข้างต้น โดยทั่วไปวิธีการที่นิยมใช้มี 3 วิธี (Klecka 1980 : 42-46 อ้างถึงใน กราบแก้ว โสภารายณ์ 2544 : 12) คือ

1. Simple Classification Function

วิธีการนี้ถูกเสนอครั้งแรกโดย Fisher R.A. เมื่อปี 1936 โดยการสร้าง ฟังก์ชันการจัดเข้ากลุ่ม (Classification Function) ซึ่งเป็นผลรวมเชิงเส้นของตัวแปรอิสระในแต่ละกลุ่ม กล่าวคือ จำนวนของฟังก์ชันการจัดเข้ากลุ่มมีเท่ากับจำนวนกลุ่มที่ต้องการจำแนก โดยฟังก์ชันการจัดเข้ากลุ่มมีรูปแบบดังนี้

$$h_k = b_{k0} + b_{k1} \mathbf{X}_1 + b_{k2} \mathbf{X}_2 + \dots + b_{kp} \mathbf{X}_p$$

เมื่อ h_k แทน คะแนนการจำแนกของกลุ่มที่ k

b_{k0} แทน ค่าคงที่สำหรับกลุ่มที่ k

b_{ki} แทน สัมประสิทธิ์ของตัวแปรอิสระที่ i สำหรับกลุ่มที่ k

โดยที่

$$b_{ki} = (N - g) \sum_{j=1}^p a_{ij} \bar{X}_{jk}$$

$$b_{k0} = -0.5 \sum_{j=1}^p b_{kj} \bar{X}_{jk}$$

a_{ij} แทน สมาชิกของอินเวอร์สเมทริกซ์ของผลต่างกำลังสองของตัวแปรอิสระภายในกลุ่ม \mathbf{W}^{-1} (element of inverse within – groups sum of crossproducts)

N แทน จำนวนตัวอย่างทั้งหมด

g แทน จำนวนกลุ่มที่นำมาศึกษา

ในการจัดเข้ากลุ่มให้พิจารณาจากค่าฟังก์ชันการจัดเข้ากลุ่ม (Classification Score) โดยจัดสมาชิกใดๆ เข้าเป็นสมาชิกของกลุ่มที่มีค่าฟังก์ชันการจัดเข้ากลุ่มสูงที่สุด

2. Generalized Distance Function

เทคนิคการจำแนกกลุ่มอีกวิธีหนึ่ง คือ การคำนวณค่า Mahalanobis distance: D^2 ซึ่งเป็นการวัดระยะทางระหว่าง 2 กลุ่ม

$$D^2(X | G_k) = (N - g) \sum_{i=1}^p \sum_{j=1}^p a_{ij} (\bar{X}_i - \bar{X}_{ik})(\bar{X}_j - \bar{X}_{jk})$$

เมื่อ $D^2(X | G_k)$ แทนระยะทางกำลังสองจากเวกเตอร์ \mathbf{X} ถึงค่าศูนย์กลางของกลุ่ม k ทำการคำนวณระยะทางกำลังสองจากจุด \mathbf{X} หรือหน่วยวิเคราะห์ใดๆไปยัง ศูนย์กลางของทุกๆกลุ่ม หลังจากนั้นเปรียบเทียบค่า D^2 ถ้าค่า D^2 ของกลุ่มใดมีค่าต่ำสุดก็ให้จัดเข้าเป็นสมาชิกของกลุ่มนั้น

3. Probabilities of Group Membership

โดยอาศัยหลักการพื้นฐานของ Bayes' Rule ความน่าจะเป็นที่สมาชิกใดๆ ซึ่งมีค่าคะแนนการจำแนก D เป็นสมาชิกของกลุ่ม k ซึ่งประมาณได้จาก

$$P(G_k | D) = \frac{P(D | G_k)P(G_k)}{\sum_{k=1}^g P(D | G_k)P(G_k)}$$

เมื่อ $P(G_k)$ แทน ความน่าจะเป็นเบื้องต้นที่หน่วยวิเคราะห์ใดๆจะเป็นสมาชิกของกลุ่ม k

$P(D | G_k)$ แทน ความน่าจะเป็นอย่างมีเงื่อนไข ที่จะได้คะแนนการจำแนก D เมื่อหน่วยวิเคราะห์นั้นเป็นสมาชิกกลุ่ม k

$P(G_k | D)$ แทน Posterior Probability หรือความน่าจะเป็นอย่างมีเงื่อนไข ที่หน่วยวิเคราะห์จะเป็นสมาชิกของกลุ่ม k เมื่อ คะแนนการจำแนกเท่ากับ D

ในการจัดสมาชิกเข้ากลุ่ม ให้พิจารณาจากค่า Posterior Probability สำหรับกลุ่ม k ใดๆ ซึ่งมีค่าสูงสุด ส่วนการกำหนดค่าความน่าจะเป็นเบื้องต้น ทำได้หลายวิธี เช่น

- ถ้าหากตัวอย่างเป็นตัวแทนของประชากร อาจใช้สัดส่วนของหน่วยวิเคราะห์ในแต่ละกลุ่ม เป็นค่าของโอกาสที่หน่วยวิเคราะห์แต่ละหน่วยจะเป็นสมาชิกของกลุ่มนั้น
- กำหนดความน่าจะเป็นเบื้องต้นจากงานวิจัยอื่นๆ ที่ศึกษาในเรื่องเดียวกันหรือจากแหล่งความรู้หรือบุคคลที่น่าเชื่อถือ
- เมื่อความเป็นสมาชิกของกลุ่มใดกลุ่มหนึ่งมีโอกาสเท่าๆกัน อาจให้ค่าของโอกาสความเป็นสมาชิกของแต่ละกลุ่มเท่าๆกันได้

ตัวอย่างการจัดหน่วยตัวอย่างเข้ากลุ่ม

จากตัวอย่างข้อมูลล้อย หากต้องการจัดหน่วยตัวอย่างที่เข้ามาใหม่เข้ากลุ่ม โดยใช้ฟังก์ชันจำแนกที่สร้างไว้ โดยที่หน่วยตัวอย่างนี้มี x_1 (yield point) = 34 และ x_2 (ultimate strength) = 62 แทนค่าใน $z = -1.633x_1 + 1.819x_2$ จะได้

$$z = -1.633(34) + 1.819(62)$$

$$= -55.522 + 112.778$$

$$= 57.256$$

จากตารางแสดงคะแนนการจำแนกของตัวอย่างล้อย คำนวณค่า \bar{z}_1 ได้เท่ากับ 54.42 และ \bar{z}_2 เท่ากับ 46.23 จะทำให้ได้ $\frac{1}{2}(\bar{z}_1 + \bar{z}_2) = \frac{1}{2}(54.42 + 46.23) = 50.325$ จากเกณฑ์การแบ่งกลุ่มเชิงเส้นของ Fisher จะได้ว่า $z > \frac{1}{2}(\bar{z}_1 + \bar{z}_2)$ นั่นคือ z เข้าใกล้ \bar{z}_1 มากกว่า และจัดหน่วยตัวอย่างที่เข้ามาใหม่นี้ให้อยู่ในกลุ่มที่ 1 หรือล้อยนี้ผลิตที่ระดับอุณหภูมิที่ 1

การวิเคราะห์การจำแนกกลุ่มสำหรับหลายกลุ่ม (Discriminant analysis for several groups)

การวิเคราะห์การจำแนกกลุ่มสำหรับหลายกลุ่ม ก็มีลักษณะเช่นเดียวกับการวิเคราะห์การจำแนกกลุ่มสำหรับ 2 กลุ่ม นั่นคือยังเกี่ยวข้องกับการหาผลรวมเชิงเส้นตัวแปรซึ่งสามารถใช้แบ่งกลุ่มของค่าสังเกตแบบพหุคูณได้ดีที่สุด การวิเคราะห์การจำแนกกลุ่มสำหรับหลายกลุ่มอาจตอบวัตถุประสงค์ต่อไปนี้ (Rencher 1995 : 303-304)

1. ตรวจสอบการแบ่งกลุ่มในรูปแบบกราฟ 2 มิติ และเมื่อเป็นกรณีที่มีมากกว่า 2 กลุ่มก็ต้องการมากกว่า 1 ฟังก์ชันจำแนกเพื่ออธิบายการแบ่งกลุ่ม ถ้าจุดใน p มิติ project บนพื้นที่ 2 มิติที่สร้างจากฟังก์ชันจำแนก 2 ฟังก์ชันแรก จะได้จุดสังเกตที่ดีที่สุดในการแบ่งกลุ่มนั้นๆ

2. หาเซตย่อยของตัวแปรอิสระที่สามารถแบ่งกลุ่มได้ดีเกือบจะเทียบเท่าเซตของตัวแปรอิสระทั้งหมด
3. จัดลำดับตัวแปรตามความสัมพันธ์ที่เกือหนุน (relative contribution) กับการแบ่งกลุ่ม
4. การแปลความหมายของมิติใหม่แสดงได้โดยใช้ฟังก์ชันจำแนก
5. ใช้ใน fixed-effects MANOVA

สมมติข้อมูลมี k กลุ่มและค่าสังเกต n_i ตัวสำหรับกลุ่มที่ i เราจะแปลงแต่ละเวกเตอร์ค่าสังเกต \mathbf{x}_{ij} จะได้ $z_{ij} = \mathbf{a}'\mathbf{x}_{ij}, i = 1, 2, \dots, k; j = 1, 2, \dots, n_i$ และหาค่าเฉลี่ย $\bar{z}_i = \mathbf{a}'\bar{\mathbf{x}}_i$ ลักษณะเดียวกับกรณี 2 กลุ่มคือเราจะหาเวกเตอร์ \mathbf{a} ที่สามารถจำแนก $\bar{z}_1, \bar{z}_2, \dots, \bar{z}_k$ ได้มากที่สุด การแสดงตัวอย่างการจำแนกระหว่าง $\bar{z}_1, \bar{z}_2, \dots$ และ \bar{z}_k เราจะขยายขอบเขตการจำแนกในสมการที่ (1) ไปเป็นกรณี k กลุ่ม เนื่องจาก $\mathbf{a}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{a}$ เราจึงสามารถแสดงสมการที่ (1) ได้ในรูปต่อไปนี้

$$\frac{(\bar{z}_1 - \bar{z}_2)^2}{s_z^2} = \frac{[\mathbf{a}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)]^2}{\mathbf{a}'\mathbf{S}_{pl}\mathbf{a}} = \frac{\mathbf{a}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{a}}{\mathbf{a}'\mathbf{S}_{pl}\mathbf{a}} \quad (10)$$

การขยายสมการที่ (10) ไปใช้สำหรับกรณี k กลุ่ม เราใช้เมทริกซ์ \mathbf{H} จาก MANOVA แทนที่ $(y_1 - y_2)(y_1 - y_2)'$ และ \mathbf{E} แทนที่ \mathbf{S}_{pl} จะได้

$$\lambda = \frac{\mathbf{a}'\mathbf{H}\mathbf{a}}{\mathbf{a}'\mathbf{E}\mathbf{a}} \quad (11)$$

สามารถเขียนให้อยู่ในรูป

$$\mathbf{a}'\mathbf{H}\mathbf{a} = \lambda\mathbf{a}'\mathbf{E}\mathbf{a}$$

หรือ

$$\mathbf{a}'(\mathbf{H}\mathbf{a} - \lambda\mathbf{E}\mathbf{a}) = 0 \quad (12)$$

ตรวจสอบค่า λ และ \mathbf{a} ที่หาได้จากสมการที่ (12) เพื่อตัดสินใจว่าค่าของ \mathbf{a} เท่ากับเท่าใดจึงจะทำให้ค่า λ มากที่สุด คำตอบที่ว่า $\mathbf{a}' = \mathbf{0}'$ นั้นไม่สามารถยอมรับได้เนื่องจากทำให้ $\lambda = 0/0$ ในสมการที่ (11) คำตอบอื่นๆสามารถหาได้จาก

$$\mathbf{H}\mathbf{a} - \lambda\mathbf{E}\mathbf{a} = \mathbf{0} \quad (13)$$

ซึ่งสามารถจะเขียนได้ในรูป

$$(\mathbf{E}^{-1}\mathbf{H} - \lambda\mathbf{I})\mathbf{a} = \mathbf{0} \quad (14)$$

คำตอบของสมการที่ (14) คือค่าไอเกน $\lambda_1, \lambda_2, \dots, \lambda_s$ และเวกเตอร์ไอเกนที่สัมพันธ์กัน $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s$ ของ $\mathbf{E}^{-1}\mathbf{H}$ เมื่อ $\lambda_1 > \lambda_2 > \dots > \lambda_s$ และจำนวนของเวกเตอร์ไอเกนที่ไม่เป็นศูนย์ (s) คือ rank ของเมทริกซ์ \mathbf{H} ซึ่งสามารถหาได้จากค่าที่น้อยกว่าระหว่าง $k-1$ และ p ค่าไอเกนที่มากที่สุด (λ_1) คือค่าที่มากที่สุดของ $\lambda = \frac{\mathbf{a}'\mathbf{H}\mathbf{a}}{\mathbf{a}'\mathbf{E}\mathbf{a}}$ จากสมการที่ (11) และสัมประสิทธิ์ของเวกเตอร์ที่ให้ค่าสูงสุดก็คือเวกเตอร์ไอเกนที่สัมพันธ์กัน (\mathbf{a}_1) ดังนั้นฟังก์ชันจำแนกที่ให้จำแนกค่าเฉลี่ยมากที่สุดก็คือ $z_1 = \mathbf{a}'_1\mathbf{x}$

จากเวกเตอร์ไอเกนทั้ง s เวกเตอร์ $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s$ ของ $\mathbf{E}^{-1}\mathbf{H}$ ซึ่งสัมพันธ์กับ $\lambda_1, \lambda_2, \dots, \lambda_s$ เราจะได้ฟังก์ชันจำแนก s ฟังก์ชัน $z_1 = \mathbf{a}'_1\mathbf{x}, z_2 = \mathbf{a}'_2\mathbf{x}, \dots, z_s = \mathbf{a}'_s\mathbf{x}$ ซึ่งไม่สัมพันธ์กัน ฟังก์ชันเหล่านี้จะแสดงมิติหรือทิศทางของความแตกต่างระหว่าง $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_k$,
เมื่อ $\bar{\mathbf{x}}_i = \sum_{j=1}^{n_i} \bar{\mathbf{x}}_{ij} / n_i$

ความสำคัญที่สัมพันธ์กันของแต่ละฟังก์ชันจำแนกสามารถกำหนดได้โดยพิจารณา
ค่าไอเกนของมันเป็นส่วนเดียวกับผลรวมดังนี้

$$\frac{\lambda_i}{\sum_j \lambda_j} \quad (15)$$

ด้วยเกณฑ์นี้ การใช้ฟังก์ชันจำแนก 2 หรือ 3 ฟังก์ชันก็มักจะเพียงพอที่จะอธิบายความแตกต่างระหว่างกลุ่ม ฟังก์ชันจำแนกที่สัมพันธ์กับค่าไอเกนที่เล็ก ๆ ก็จะถูกละทิ้งไป การทดสอบความมีนัยสำคัญของแต่ละฟังก์ชันจำแนกก็มีประโยชน์เช่นกัน

สถิติที่สำคัญของการวิเคราะห์จำแนกกลุ่ม

ในการนำเทคนิคการวิเคราะห์การจำแนกกลุ่มไปใช้ มีสถิติบางตัวที่ผู้ใช้จะต้องทำความเข้าใจ เพื่อจะใช้ตัดสินความสำคัญของสมการวิเคราะห์จำแนกกลุ่ม สถิติที่สำคัญมีด้วยกัน 6 ตัวคือ (เดือนเพ็ญ สนโต 2545 : 23)

1. Eigenvalue คือค่าอัตราส่วนการผันแปรระหว่างกลุ่มต่อการผันแปรภายในกลุ่ม ใช้วัดความสำคัญเชิงเปรียบเทียบของสมการว่า สมการที่ได้มีอำนาจในการแบ่งแยกการเป็นสมาชิกของกลุ่มได้ดีเพียงใด

2. Relative Percentage หรือ Cumulative Percentage เป็นค่าที่แสดงความสัมพันธ์ของ Eigenvalue กับสมการที่ได้ว่ามีอำนาจในการแบ่งแยกคิดเป็นร้อยละเท่าใดของอำนาจในการจำแนกรวม
3. Canonical Correlation เป็นค่าที่บ่งถึงความสัมพันธ์ระหว่างกลุ่มของตัวแปรซึ่งระบุการเป็นสมาชิกกลุ่มนั้นๆของตัวแปรตาม ที่ชี้ให้เห็นว่าการเป็นสมาชิกของกลุ่มมีความสัมพันธ์กับสมการที่หามาได้มากน้อยเพียงใดถ้าเทียบกับการวิเคราะห์ความแปรปรวนทางเดียว ค่าสหสัมพันธ์คาโนนิกอลนี้ก็คือค่า Correlation Ratio (Eta) ถ้าค่าสหสัมพันธ์คาโนนิกอลค่าที่แสดงว่าตัวแปรอิสระมีความสามารถน้อยในการคาดคะเนความสัมพันธ์กับตัวแปรตาม หรือ สมการนั้นไม่สามารถใช้ในการคาดคะเนการเป็นสมาชิกของกลุ่มนั้นได้ดี
4. Wilk's Lambda เป็นตัวทดสอบนัยสำคัญทางสถิติของวิธีการเลือกตัวแปรเข้าแบบ Stepwise เป็นค่าซึ่งจะอธิบายว่าตัวแปรที่เหลือจะมีความสามารถในการแบ่งแยกกลุ่มเพิ่มขึ้นอีกเท่าใดจากสมการที่ได้มาก่อนแล้ว โดยที่ค่า Wilk's Lambda คืออัตราส่วนของผลรวมกำลังสองของการผันแปรภายในกลุ่มต่อผลรวมกำลังสองของการผันแปรทั้งหมดที่ไม่อาจอธิบายได้ด้วยความแตกต่างระหว่างกลุ่ม ถ้าค่า Wilk's Lambda มีค่าน้อยหมายความว่าสมการที่ได้มีความผันแปรระหว่างกลุ่มมากกว่าภายในกลุ่ม หรือกล่าวได้ว่าสมการที่ได้มีความสามารถในการจำแนกกลุ่มได้ดี ถ้าค่าของ Wilk's Lambda เป็น 1 แสดงว่าไม่มีความแตกต่างระหว่างกลุ่ม
5. Chi-Square ค่า Wilk's Lambda ที่ได้มีการแจกแจงแบบ Chi-Square เมื่อคำนวณค่า Chi-Square จะเป็นตัวสถิติที่บอกระดับความมีนัยสำคัญ โดยทำการเปรียบเทียบกับคะแนนมาตรฐานที่ได้ เพื่อใช้ในการทดสอบว่าสมการที่สร้างได้มีอำนาจในการแบ่งกลุ่มได้อย่างมีนัยสำคัญหรือไม่
6. Exact F แทนค่าสถิติทดสอบ F ซึ่งใช้ในการทดสอบนัยสำคัญทางสถิติของวิธีการเลือกตัวแปรเข้าแบบ Stepwise โดยที่ค่าสถิติทดสอบ F ที่น้อยที่สุดสำหรับการนำตัวแปรเข้าสมการโดยใช้วิธีการคัดเลือกแบบ Stepwise มีค่าเท่ากับ 3.84 และค่าสถิติทดสอบ F ที่มากที่สุดสำหรับการคัดตัวแปรออกจากสมการมีค่าเท่ากับ 2.71

การวิเคราะห์การถดถอย (Regression Analysis)

การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติอย่างหนึ่งที่ใช้ในการศึกษาลักษณะของความสัมพันธ์ระหว่างตัวแปรตั้งแต่สองตัวขึ้นไป โดยมีวัตถุประสงค์เพื่ออธิบายลักษณะความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งกับตัวแปรอื่นๆ และคาดคะเนค่าของตัวแปรตัวนั้นโดยอาศัยความสัมพันธ์ข้างต้น การวิเคราะห์การถดถอยประกอบด้วยขั้นตอนต่างๆหลายขั้นตอนด้วยกัน เช่น ขั้นตอนการพิจารณาว่าตัวแปรใดบ้างเป็นตัวแปรที่มีอิทธิพลต่อความผันแปรของตัวแปรตาม ขั้นตอนการกำหนดตัวแบบความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระเหล่านั้น งานวิเคราะห์การถดถอยจะถูกต้องหรือไม่ และสามารถนำไปใช้ในการคาดคะเนค่าตัวแปรตามได้ดีเพียงใด ขึ้นอยู่กับว่าสามารถระบุตัวแปรและกำหนดตัวแบบความสัมพันธ์ถูกต้องมากน้อยแค่ไหน

ตัวแบบการถดถอยโลจิสติก (Logistic Regression Model) (วีรพันธ์ พงศาภักดิ์ 2541)

ตัวแบบการถดถอยโลจิสติก เป็นรูปแบบหนึ่งของการตัวแบบการถดถอย(Regression Model) กำหนดให้ $P(x) = E(Y | X)$ คือ ค่าเฉลี่ยแบบมีเงื่อนไขของตัวแปรตาม Y ซึ่งเป็นตัวแปรแบบเชิงกลุ่ม เช่น 2 กลุ่ม (Dichotomous) หรือหลายกลุ่ม (Polytomous) เมื่อกำหนดตัวแปรอิสระ X มาให้ (Conditional mean of y given x) การแจกแจงของตัวแปรตาม (Y) เป็นแบบทวินาม (Binomial Distribution) หรือ $Bin(1, P(x))$ เมื่อพิจารณาที่ยังไม่ได้จัดกลุ่ม (Ungrouped data) และเป็น $Bin(n_i, P_i(x))$ เมื่อพิจารณาข้อมูลแบบจัดกลุ่ม (Grouped data) โดย $P(x)$ แทนความน่าจะเป็นของตัวแปรตามในกลุ่มที่สนใจศึกษา ในกรณีที่ตัวแปรตามเป็นตัวแปรเชิงกลุ่ม 2 กลุ่ม คือ $Y=0$ และ $Y=1$ เมื่อ $Y=1$ ความน่าจะเป็นของตัวแปรตามคือ $P(x)$ และเมื่อ $Y=0$ ความน่าจะเป็นคือ $1 - P(x)$ ถ้าตัวแปรอิสระ (X) มีเพียง 1 ตัว จะได้ตัวแบบถดถอยโลจิสติกเป็นดังนี้

$$P(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

และเพื่อที่จะให้ค่าของ $P(x)$ หรือ $E(Y | X)$ มีค่าอยู่ระหว่าง 0 ถึง 1 และค่าของ x สามารถมีได้ไม่จำกัด จึงมีความจำเป็นจะต้องแปลง $P(x)$ ให้อยู่ในรูปแบบอื่นที่เข้าใจง่าย และมีคุณสมบัติตามต้องการ การแปลงที่ใช้นี้ใช้ Logit transformation หรือ $g(x)$

$$g(x) = \ln \left[\frac{P(x)}{1 - P(x)} \right]$$

และมีความสัมพันธ์เชิงเส้นในรูปแบบต่อไปนี้

$$\ln \left[\frac{P(x)}{1-P(x)} \right] = \beta_0 + \beta_1 x$$

ตัวแบบข้างต้นคือ logarithm ของ odds ในรูปความสัมพันธ์เชิงเส้นกับ X เรียกตัวแบบนี้ว่า ตัวแบบโลจิท (logit model)

$$\therefore \frac{P(x)}{1-P(x)} = e^{\beta_0 + \beta_1 x}$$

ตัวแบบนี้ มีรูปแบบเช่นเดียวกันกับ odds ของตัวแปรตาม เมื่อ Y=1 ด้วย

$$P(x) = [1-P(x)]e^{\beta_0 + \beta_1 x}$$

$$\therefore e^{\beta_0 + \beta_1 x} = P(x)[1 + e^{\beta_0 + \beta_1 x}]$$

นั่นคือ

$$P(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

หรือ

$$P(x) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$$

เรียกตัวแบบนี้ว่า ฟังก์ชันการถดถอยโลจิสติก (Logistic regression function) หรือ ตัวแบบโลจิสติก (Logistic regression model)

สำหรับการแจกแจงของ Y เมื่อพิจารณาแต่ละหน่วยของ Y ที่มีผลลัพธ์เป็น 0 หรือ 1 กรณีนี้คือ Y คือตัวแปรเชิงสุ่มเบอร์นูลลี (Bernoulli random variable) ที่มีค่าเฉลี่ย $E(Y)$ โดยที่

$$E(Y) = \sum YP(y)$$

$$E(Y) = 1 \times P(Y = 1) + 0 \times P(Y = 0)$$

$$E(Y) = P(Y = 1)$$

$$= P(x)$$

หรือ $E(Y | X) = P(x)$ โดยมีค่าขึ้นอยู่กับตัวแปร X

$$\begin{aligned} \text{ดังนั้น} \quad E(Y^2) &= \sum Y^2 P(y) \\ E(Y^2) &= 1^2 [P(x)] + 0^2 [1 - P(x)] \\ &= P(x) \end{aligned}$$

ฉะนั้นความแปรปรวนของ Y คือ $Var(Y)$ โดยที่

$$\begin{aligned} Var(Y) &= E(Y^2) - [E(Y)]^2 \\ &= P(x)[1 - P(x)] \end{aligned}$$

นอกจากนี้ เมื่อ Y คือตัวแปรสุ่มทวินาม (Binomial) $Bin[n_i, P_i(x)]$ โดยที่ $i = 1, 2, \dots, N$ ซึ่งมีค่าเฉลี่ยและความแปรปรวน คือ

$$\begin{aligned} E(Y_i) &= n_i P_i(x) \\ Var(Y_i) &= n_i P_i(x) [1 - P_i(x)] \end{aligned}$$

มหาวิทยาลัยศิลปากร ส่วนวนลิขสิทธิ์

จะเห็นได้ว่า $g(x)$ หรือ logit มีรูปแบบเหมือนกับการถดถอยเชิงเส้นตรง (Linear regression) ฉะนั้นในการคำนวณใดๆเกี่ยวกับการถดถอยโลจิสติก จึงสามารถใช้คุณสมบัติของการถดถอยเชิงเส้นตรงได้ กล่าวคือ หลักเกณฑ์ที่เป็นคู่ทางของการวิเคราะห์การถดถอยโลจิสติกสามารถใช้หลักเกณฑ์ของการวิเคราะห์การถดถอยเชิงเส้น ทั้งแบบง่ายและแบบพหุคูณ

การประมาณค่าสัมประสิทธิ์ถดถอยโลจิสติก (Estimating Logistic Coefficient)

ในการประมาณค่าสัมประสิทธิ์ถดถอยมักใช้วิธีกำลังสองน้อยที่สุด (Least squares) นั่นคือเลือก β_0 และ β_1 ที่ทำให้ผลรวมกำลังสองของส่วนเบี่ยงเบนระหว่างค่าสังเกตของตัวแปรตามกับค่าพยากรณ์ของตัวแปรตามมีค่าน้อยที่สุด การประมาณค่าโดยวิธีกำลังสองน้อยที่สุดนี้อยู่ภายใต้ข้อตกลงเบื้องต้นหลายประการ ถ้าใช้วิธีกำลังสองน้อยที่สุดนี้ไปประมาณค่าสัมประสิทธิ์ถดถอยโลจิสติก ข้อตกลงเบื้องต้นดังกล่าวจะไม่เป็นจริง

นอกจากนี้ ยังมีอีกวิธีที่นิยมใช้ทั่วไปและสามารถประมาณค่าสัมประสิทธิ์ถดถอยโลจิสติกได้อย่างเหมาะสม เรียกว่าวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood หรือ ML) ซึ่งให้ค่าประมาณสัมประสิทธิ์ที่มีค่าความน่าจะเป็นสูงสุด สำหรับเซตข้อมูลหรือค่าสังเกตที่ใช้ในการ

ประมาณดังกล่าว การประมาณค่าสัมประสิทธิ์โดยวิธีภาวะน่าจะเป็นสูงสุด เริ่มจากการสร้างฟังก์ชันภาวะน่าจะเป็น (Likelihood function) L ก่อน แล้วจากนั้นจึงหาสมการภาวะน่าจะเป็นสูงสุดต่อไป

ถ้ากำหนดให้ตัวแปรตาม Y มีค่า 0 หรือ 1 ดังนั้น $P(x)$ คือนิพจน์ (expression) ของค่าสัมประสิทธิ์ที่ต้องการประมาณค่า ของความน่าจะเป็นแบบมีเงื่อนไขที่ $Y=1$ เมื่อกำหนดค่าตัวแปรอิสระ X มาให้ นั่นคือ $P(x) = P(Y=1|X)$ และนิพจน์ $[1-P(x)]$ คือความน่าจะเป็นแบบมีเงื่อนไขที่ $Y=0$ เมื่อกำหนดค่าตัวแปรอิสระ X มาให้

การประมาณค่าสัมประสิทธิ์โดยวิธีภาวะน่าจะเป็นสูงสุดอาศัยการหาอนุพันธ์ (differentiation) ของฟังก์ชันภาวะน่าจะเป็น L เทียบกับค่าสัมประสิทธิ์ที่ต้องการประมาณ โดยที่ L มีรูปแบบคือ

$$L = \prod_{j=1}^n [P(x_j)]^{y_j} [1-P(x_j)]^{1-y_j}$$

$$= \prod_{j=1}^n \left[\frac{\exp(\beta_0 + \beta'X_j)^{y_j}}{1 + \exp(\beta_0 + \beta'X_j)} \right]$$

และหาอนุพันธ์เทียบกับ β_0 และ β ซึ่งถ้ามี X เพียง 1 ตัว $\beta = \beta_1$ โดยที่

$y=1$ เมื่อ X_j มาจากกลุ่ม 1 และ $y=0$ เมื่อ X_j มาจากกลุ่มที่ 2 ส่วน $n = \sum_{i=1}^2 n_i$ เมื่อแก้

สมการแล้วจะได้ค่าประมาณ $\hat{\beta}_0$ และ $\hat{\beta}$ ซึ่งเรียกว่า ตัวประมาณภาวะน่าจะเป็นสูงสุด

(Maximum Likelihood Estimator or MLE)

การทดสอบนัยสำคัญของค่าสัมประสิทธิ์

การทดสอบนัยสำคัญของค่าสัมประสิทธิ์ หรือการทดสอบนัยสำคัญของตัวแปรอิสระแต่ละตัวในตัวแบบ ว่ามีความสัมพันธ์กับตัวแปรตามหรือไม่ วิธีการที่นิยมใช้ (Hosmer and Lemeshow 1989:11-18 ; Agresti 1996:103-115 อ้างถึงใน กราบแก้ว โสภาราษฎร์ 2544:31) ได้แก่

1. การพิจารณาค่า Change of deviances หรือค่า Statistic G^2 แนวคิดพื้นฐานคือ การเปรียบเทียบค่าสังเกต (observed value) ของตัวแปรตาม กับค่าพยากรณ์ (predicted value) ซึ่งได้จากตัวแบบซึ่งมีและไม่มีตัวแปรอิสระที่กำลังสนใจ

$$\begin{aligned} G^2 &= D(\text{model without the variable}) - D(\text{model with the variable}) \\ &= \Delta D \end{aligned}$$

เมื่อ ΔD แทน change of deviances
 นอกจากค่าสถิติ G^2 หรือ change of deviances แล้วยังมีค่าสถิติ G^2 อีกรูปแบบ
 หนึ่ง คือ

$$G^2 = -2 \ln \left[\frac{(\text{maximum likelihood without the variable})}{(\text{maximum likelihood with the variable})} \right]$$

โดยที่ตัวสถิติ D มีชื่อเรียกอีกอย่างว่า Deviance และปริมาณที่อยู่ภายในวงเล็บ
 เรียกว่า Likelihood Ratio ส่วนการที่มีพจน์ $-2 \ln$ คูณกับ Likelihood ratio ก็เพื่อเหตุผลในการ
 ทดสอบสมมติฐาน กล่าวคือ ทำให้ตัวสถิติทดสอบมีการแจกแจงแบบไคส์แควร์ (Chi-Square)
 การทดสอบดังกล่าว เรียกว่า Likelihood Ratio Test

ในการทดสอบสมมติฐาน $\beta_i = 0; i = 1, 2, \dots, p$ จะใช้ตัวสถิติ G^2 ซึ่งมีการแจก
 แจกแจงแบบไคส์แควร์ ด้วยองศาอิสระเท่ากับ 1

2. การทดสอบ Wald test โดยที่เป็นการเปรียบเทียบค่า $\hat{\beta}_i$ กับค่าคาดเคลื่อน

มาตรฐาน (Standard error) ของค่า $\hat{\beta}_i$ นั่นก็คือ

$$W_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

โดยที่ตัวสถิติ W มีการแจกแจงแบบปกติเมื่อตัวอย่างมีขนาดใหญ่ (Large-sample
 normal distribution) แต่การใช้ Wald test มีข้อเสียคือ ถ้าค่าสัมบูรณ์ของค่าประมาณสัมประสิทธิ์
 มีค่ามาก ค่าคาดเคลื่อนมาตรฐานก็มักมีค่ามากด้วย ทำให้อัตราส่วน $\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$ หรือ Wald
 statistic มีค่าน้อย ซึ่งนำไปสู่การผิดพลาดในการปฏิเสธสมมติฐานว่าง (reject null hypothesis)
 ฉะนั้นเมื่อค่าสัมประสิทธิ์มีค่ามาก ควรจะใช้วิธีคำนวณตัวสถิติ G^2 จะเหมาะสมกว่า

การทดสอบความเหมาะสมของตัวแบบ

สถิติที่ใช้ในการทดสอบความเหมาะสมของตัวแบบได้แก่ค่าสถิติทดสอบ Hosmer-
 Lemeshow (กัลยา วานิชย์บัญชา 2546 : 93) ซึ่งใช้ในการตรวจสอบความเหมาะสมของตัวแบบ
 โดยจะแบ่งค่าสังเกต ออกเป็น 10 กลุ่มกลุ่มละพอกๆกัน โดยการแบ่งพิจารณาจากค่าประมาณของ
 โอกาสที่จะเกิดเหตุการณ์ที่เราสนใจ โดยใช้สถิติทดสอบไคสแควร์ โดยขนาดตัวอย่างต้องมาก

พอที่ทำให้จำนวนค่าสังเกต ที่คาดไว้ (Expected value : E_i ; $i=1, \dots, 10$) จะต้องมากกว่า 5 และไม่
มีกลุ่มใดที่มีค่า $E_i < 1$

$$\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i}$$

โดยใช้การทดสอบ H_0 : Model เหมาะสม

H_1 : Model ไม่เหมาะสม

ถ้าขนาดตัวอย่างใหญ่ ค่าสถิติทดสอบ Hosmer-Lemeshow จะมีค่ามากด้วย ทำให้
อาจสรุปว่าปฏิเสธ H_0 ทั้งที่ Model เหมาะสม จึงควรระวังในการแปลผล

ตัวอย่างการใช้ตัวแบบการถดถอยโลจิสติก

(วีรพันธ์ พงศาภักดี 2541 : 61-62) ในการศึกษาเกี่ยวกับการแพร่กระจายของเชื้อโรคมะเร็ง
หรือไม่ (Y) หลังทำการผ่าตัด ทำโดยการ พยากรณ์ค่า Y จากตัวแปรหลายๆตัว ที่สังเกตได้ก่อน
การผ่าตัด ดังต่อไปนี้

X_1	แทน	age of patients
X_2	แทน	serum acid phosphatase
X_3	แทน	x-ray results
X_4	แทน	grade of the tumor
X_5	แทน	stage of the disease

และ Y แทน ตัวแปรตามซึ่งมีค่า 1 (แพร่กระจาย) และ 0 (ไม่แพร่กระจาย) ข้อมูล
เกี่ยวกับตัวแปรข้างต้นของคนไข้จำนวน 53 คน เป็นเพศชายทั้งหมด ซึ่งเป็นข้อมูลจาก Brown
(1980) จากการประมวลผลด้วย SPSS/FW ผลลัพธ์ที่ได้จากการประมวลผลตัวแบบถดถอยโลจิสติกที่
สอดคล้องกับตัวแบบโลจิสติก ประกอบด้วยค่าดังตารางต่อไปนี้

ตารางที่ 3 ผลการประมวลข้อมูลการแพร่กระจายของเชื้อโรคมะเร็งด้วย SPSS/FW

ตัวแปร	ตัวประมาณ($\hat{\beta}$)	S.E.	Wald	d.f.	Sig	R	Exp($\hat{\beta}$)
X_1	-0.07	0.058	1.43	1	0.231	0.000	0.933
X_2	0.02	0.013	3.42	1	0.064	0.142	1.025
X_3	2.04	0.807	6.42	1	0.001	0.251	7.732
X_4	0.76	0.771	0.98	1	0.323	0.000	2.141
X_5	1.56	0.774	4.08	1	0.043	0.172	4.778
ค่าคงที่	0.06	3.460	0.00	1	0.986	-	-

จาก
$$P[Y = 1 | x] = P(x) = \frac{\exp(\hat{\phi})}{1 + \exp(\hat{\phi})}$$

และ
$$odds = \frac{P(x)}{1 - P(x)} = \exp(\hat{\phi})$$

โดยที่ $\hat{\phi} = 0.0618 - 0.0693X_1 + 0.0243X_2 + 2.0453X_3 + 0.7614X_4 + 1.5641X_5$

เช่นถ้าต้องการพยากรณ์ชายอายุ 60 ปี ที่มี $X_2 = 62$, $X_3 = 1$, $X_4 = X_5 = 0$

แล้ว จะคำนวณได้ค่าของ $\hat{\phi} = -0.54$ และความน่าจะเป็นที่เชื้อโรคมะเร็งจะแพร่กระจายคือ

$$P(x) = \frac{\exp(-0.54)}{1 + \exp(-0.54)}$$

$$= 0.37$$

$$1 - P(x) = 0.73$$

ดังนั้น
$$odds = \frac{0.37}{0.73} = 0.59$$

การประเมินผลโดยทั่วไป ถ้า $P(x) < 0.5$ หมายความว่าเชื้อโรคมะเร็งไม่กระจาย
 ถ้า $P(x) > 0.5$ หมายความว่าเชื้อโรคมะเร็งกระจาย
 ถ้า $P(x) = 0.5$ หมายความว่าเชื้อโรคมะเร็งอาจกระจายหรือไม่กระจายก็ได้ หากผลเป็นเช่นนี้ควรตัดสินใจด้วยการพิจารณาองค์ประกอบอื่น ๆ ร่วมด้วย หรือหากไม่มีข้อมูลอื่น ๆ อาจตัดสินใจด้วยการโยนเหรียญ

การประมาณอัตราการผิดพลาดในการจำแนกกลุ่ม

การประมาณอัตราการผิดพลาดในการจำแนกกลุ่ม (Estimating Misclassification Rates) เป็นการประเมินฟังก์ชันในการจำแนกกลุ่มว่าเหมาะสมหรือไม่ เช่นเดียวกับการดูว่าสมการที่ fit ในสมการวิเคราะห์การถดถอยว่า fit ได้ดีหรือไม่ พิจารณาจากค่า R^2 การพิจารณาว่าตัวแบบที่ fit จากตัวอย่างในการวิเคราะห์การจำแนกเหมาะสมหรือไม่ ทำได้โดยการคำนวณหาอัตราการผิดพลาดในการแบ่งกลุ่ม หรือความน่าจะเป็นในการจำแนกกลุ่มที่ไม่ถูกต้อง (Misclassification Probabilities) เพื่อพิจารณาว่าในตัวอย่างนั้นจำแนกกลุ่มได้ถูกต้องตามกลุ่มที่เป็นจริงเท่าไร แล้วคำนวณค่าจำแนกกลุ่มที่ถูกต้อง ในการคำนวณอัตราการผิดพลาดในการจำแนกกลุ่ม มีการเรียกชื่อแตกต่างกันออกไปเช่นอัตราการผิดพลาด (Error Rate) หรือความน่าจะเป็นในการจัดเข้ากลุ่มผิด ในการประมาณอัตราการผิดพลาดในการจำแนกกลุ่มโดยทั่วไป มีจุดมุ่งหมายหลัก 3 ประการ ได้แก่

ประการที่ 1 เพื่อศึกษาว่าการใช้กฎการจำแนก โดยอาศัยข้อมูลที่ได้จากประชากรเหมาะสมหรือไม่

ประการที่ 2 เพื่อศึกษาว่าการใช้สมการจำแนกโดยใช้ที่ได้จากตัวอย่างเหมาะสมที่จะนำไปประยุกต์ใช้กับตัวอย่างในอนาคตหรือไม่

ประการที่ 3 เป็นการศึกษาว่าสมการจำแนกมีความเที่ยงตรงเพียงพอ ที่จะนำไปประยุกต์กับตัวอย่างที่เป็นไปได้หรือไม่

การประมาณอัตราการจำแนกผิดที่เห็นชัดเจน

การประมาณอัตราการผิดพลาดมีหลายวิธี แต่ในงานวิจัยนี้จะกล่าวเฉพาะวิธีอัตราการจำแนกผิดที่เห็นชัดเจน Apparent error rate (APER) เท่านั้น

การคำนวณ APER เป็นวิธีที่ไม่ต้องทราบฟังก์ชันของประชากร วิธีการประมาณค่าด้วยวิธีนี้ทำได้โดยหาสัดส่วนการจำแนกผิดของ Substitution โดยที่ Substitution คือการนำตัวอย่างที่ใช้ในการสร้างฟังก์ชันจำแนก (Training Samples) มาทำการจัดเข้ากลุ่มอีกครั้ง แล้วนำผลของการจำแนกใส่ในตารางซึ่งเรียกว่า Confusion Matrix ดังนี้

ตารางที่ 4 ผลการจำแนกค่าสังเกตเข้ากลุ่ม (Confusion Matrix)

กลุ่มจริง (Actual membership)	กลุ่มจากสมการ (Predicted Membership)		จำนวนของค่าสังเกต (Number of observations)
	1	2	
1	n_{1c}	$n_{1m} = n_1 - n_{1c}$	n_1
2	$n_{2m} = n_2 - n_{2c}$	n_{2c}	n_2

เมื่อ n_{1c} = จำนวนกลุ่มตัวอย่างที่แบ่งกลุ่ม 1 ถูกต้อง
 n_{1m} = จำนวนกลุ่มตัวอย่าง 1 จริง แต่ถูกจัดอยู่กลุ่ม 2
 n_{2c} = จำนวนกลุ่มตัวอย่าง 2 จริง แต่ถูกจัดอยู่กลุ่ม 1
 n_{2m} = จำนวนกลุ่มตัวอย่างที่แบ่งกลุ่ม 2 ถูกต้อง

ดังนั้นการคำนวณอัตราความผิดพลาดของการจำแนกกลุ่มโดยวิธี APER คำนวณจาก

$$\begin{aligned}
 APER &= \frac{n_{1m} + n_{2m}}{n_1 + n_2} \\
 &= \frac{n_{1m} + n_{2m}}{n_{1c} + n_{1m} + n_{2c} + n_{2m}} \quad (16)
 \end{aligned}$$

ในทำนองเดียวกันเราสามารถจะหา อัตราการจำแนกถูกต้องที่เห็นชัดเจนได้จาก

$$\text{Apparent Correct Classification Rate} = \frac{n_{1c} + n_{2c}}{n_1 + n_2} \quad (17)$$

หรืออาจเขียนในรูป

$$APER = 1 - \text{Apparent Correct Classification Rate}$$

การคำนวณค่า APER สามารถทำได้ง่ายและส่วนมากโปรแกรมที่ใช้ในการจำแนกจะพบค่าสังเกตให้คำนวณค่านี้ แต่เนื่องจากการประมาณค่าความผิดพลาดของการจำแนกกลุ่มโดยวิธีดังกล่าว พบว่าเมื่อจำนวนตัวอย่างมีขนาดเล็กการประมาณค่าความผิดพลาดมีแนวโน้มจะให้ค่าประมาณที่ต่ำกว่าปกติ (Underestimate) ดังนั้นการนำไปใช้จึงต้องพิจารณาจำนวนตัวอย่างที่ใช้ในแต่ละการวิเคราะห์ร่วมด้วย การประมาณค่าความผิดพลาดวิธีดังกล่าว พบได้ในโปรแกรม BMDP7M และโปรแกรม SPSS และ SAS

โรคเลปโตสไปโรสิส

โรคเลปโตสไปโรสิสเป็นโรคของสัตว์ที่ติดต่อกันมาซึ่งคนเกิดจากเชื้อเลปโตสไปรา เชื้อถูกปล่อยออกมาทั้งปัสสาวะสัตว์ที่ติดเชื้อและปนเปื้อนอยู่ตามน้ำ ดินทรายเปียกชื้น หรือพืชผัก เชื้อสามารถไชเข้าสู่ผิวหนังปกติที่เปียกชุ่มเนื่องจากแช่น้ำอยู่นาน คนมักติดเชื้อโดยอ้อมขณะย่ำดินโคลน แช่น้ำท่วมหรือว่ายน้ำ หรืออาจติดโรคโดยตรงจากการสัมผัสเชื้อในปัสสาวะสัตว์หรือเนื้อสัตว์ที่ปนเปื้อนเชื้อ เชื้ออาจเข้าร่างกายโดยการกินอาหารหรือน้ำ หรือการหายใจเอาละอองนิวเคลียสจากของเหลวที่ปนเปื้อนเชื้อเข้าไป แต่พบได้น้อย ส่วนการติดจากคนถึงคน มีรายงานการติดต่อจากปัสสาวะผู้ป่วยเพียงรายเดียว แม้ว่าจะพบเชื้อในปัสสาวะของผู้ป่วยได้นาน 1 - 11 เดือนก็ตาม แต่การติดต่อจากแม่ไปทางรกทำให้ทารกตายในครรภ์นั้นมีรายงานเพียง 2 ราย นอกจากนั้นยังมีรายงานเด็กที่คลอดออกมา มีอาการป่วยเหมือนในผู้ใหญ่และรักษาได้ผลดี เชื้อเลปโตสไปราก่อให้เกิดโรคได้หลายระบบ ผู้ป่วยมีอาการและอาการแสดงแตกต่างกันได้ตั้งแต่ไม่รุนแรงหายได้เองจนเสียชีวิตได้ เป็นโรคที่พบได้ทุกภูมิภาคของโลก ในประเทศไทยมีรายงานผู้ป่วยครั้งแรกปี พ.ศ. 2486 ซึ่งเกิดน้ำท่วมกรุงเทพฯ ครั้งใหญ่ หลังจากนั้นมีการพบผู้ป่วยโรคนี้ได้ทุกภาคของประเทศ โดยภาคตะวันออกเฉียงเหนือมีอัตราป่วยสูงสุดและภาคกลางต่ำสุด ปัจจุบันการตรวจรักษาผู้ป่วยเป็นโรคนี้แพทย์จะวินิจฉัยเบื้องต้น โดยพิจารณาจากอาการแสดงของผู้ป่วยรายนั้นและทำการรักษาตามอาการ แล้วรอผลตรวจยืนยันทางห้องปฏิบัติการ หากผู้ป่วยติดเชื้อในปริมาณมากจะสามารถตรวจพบได้อย่างรวดเร็ว แต่หากผู้ป่วยติดเชื้อในปริมาณน้อยจะต้องใช้เวลาในการตรวจนานกว่าจึงจะสามารถสรุปได้ว่าผู้ป่วยติดเชื้อนี้จริง ดังนั้นในการรักษาหากรอผลตรวจทางห้องปฏิบัติการเพียงอย่างเดียวอาจไม่ทันท่วงทีและอาจเกิดอันตรายต่อผู้ป่วยได้

อาการและอาการแสดงของผู้ป่วยเลปโตสไปโรสิส

อาการในคนอาจแตกต่างกันออกไป (สำนักงานโครงการควบคุมโรคเลปโตสไปโรสิส 2544 : 14) ขึ้นกับชนิดและปริมาณของเชื้อ อาการที่พบบ่อยได้แก่ ไข้เฉียบพลัน ปวดศีรษะรุนแรง หนาวสั่น ปวดกล้ามเนื้ออย่างรุนแรงโดยเฉพาะบริเวณน่องโคนขา กล้ามเนื้อหลัง และท้อง ตาแดง อาจมีไข้ติดต่อกันหลายวันสลับกับระยะไข้ลด (Biphasic) และมีเยื่อหุ้มสมองอักเสบ มีผื่นที่เพดานปาก (Palatal Exanthem) โลหิตจาง มีจุดเลือดออกตามผิวหนังและเยื่อ ติบและไตวาย ดีซ่าน อาจมีเยื่อหุ้มสมองอักเสบ ทำให้รู้สึกสับสน เพื่อ ชิม กล้ามเนื้อหัวใจอักเสบ อาจมีอาการทางระบบทางเดินหายใจ ไอมีเสมหะ อาจมีเลือดปน (Hemoptysis) และเจ็บหน้าอก

แม้ว่าอาการของโรคจะค่อนข้างหลากหลาย โดยอาจมีอาการเด่นของอวัยวะใดอวัยวะหนึ่งที่ถูกทำลาย ไม่ว่าจะเป็นไต ตับ ระบบหายใจ หรือระบบไหลเวียนโลหิต แต่จากรายงานที่มีอยู่ในประเทศไทย อาการที่พบได้บ่อยมากคือ ไข้สูง (88.8-100%) ปวดศีรษะ (66-100%) ปวดกล้ามเนื้อ (76-100%) และตาแดง (74-100%) สำหรับอาการเหลือพบน้อยกว่าคือ 37-70% อาการอื่นๆ ได้แก่ ผื่น จุดเลือดออกตามผิวหนัง ไอเป็นเลือด ตับโต ม้ามโต เป็นต้น

อัตราป่วยตายโดยเฉลี่ยต่ำ แต่จะเพิ่มขึ้นในคนไข้สูงอายุ และอาจสูงถึง 20 เปอร์เซ็นต์หรือมากกว่าในคนไข้ที่มีไตช้ำและไตถูกทำลาย แต่ไม่ได้รับการรักษาที่รวดเร็วและเพียงพอ ซึ่งรวมทั้งการล้างไต (renal dialysis) ด้วย สาเหตุการตายมักมาจากตับและไตวาย ในปัจจุบันพบว่า การที่มีเลือดออกมากและกลุ่มอาการทางเดินหายใจล้มเหลว (adult respiratory distress syndrome) เป็นสาเหตุที่ทำให้ผู้ป่วยเสียชีวิตเร็ว สาเหตุการตายอาจมาจากการเด่นของหัวใจผิดปกติ เนื่องจากกล้ามเนื้อหัวใจอักเสบก็ได้

เกณฑ์การรายงานผู้ป่วยเพื่อเฝ้าระวังโรคเลปโตสไปโรสิส

1. ผู้ป่วยที่แพทย์วินิจฉัยจากอาการ (Clinical Diagnosis Cases) ได้แก่ ผู้ป่วยที่มีอาการไข้สูงเฉียบพลัน (39 องศาเซลเซียสหรือสูงกว่านั้น) ปวดศีรษะมาก ปวดกล้ามเนื้อโดยเฉพาะกล้ามเนื้ออ่อนๆ กดเจ็บ ร่วมกับอาการอย่างอื่น เช่น ตาแดงชนิดไม่มีขี้ตา (Suffusion Conjunctivitis) หนาวสั่น ตาเหลือง ตัวเหลือง ปัสสาวะขุ่นหรือปัสสาวะน้อย ไอเป็นเลือด ปวดท้อง ท้องเสียและมีประวัติที่เกี่ยวข้องทางระบาดวิทยาที่เสี่ยงต่อการเกิดโรคเลปโตสไปโรสิส เช่น ฝนตก น้ำท่วม น้ำท่วมขัง ปักดำ เกี่ยวข้าว หาดปลา ลงไปแช่น้ำนานๆ ล้างท่อน้ำโสโครกทำงานในโรงฆ่าสัตว์ ก่อนป่วย 2 วันถึงประมาณ 1 เดือน โดยไม่ได้เจาะเลือดผู้ป่วยตรวจทางห้องปฏิบัติการให้รายงานเป็นผู้ป่วยที่สงสัย (Clinical Diagnosis Cases) หรือถ้าหากเจาะเลือดผู้ป่วยในขณะแรกเริ่ม แล้วนำไปตรวจ Leptospiriosis Screening Test ชนิด Latex test (หรือ Screening Test วิธีอื่น) แล้วให้ผล Negative และไม่สามารถติดตามเจาะเลือดผู้ป่วยครั้งที่ 2 (10-14 วันหลังจากเจาะครั้งแรก) ได้ ก็ให้รายงานเป็น Clinical Diagnosis Cases เช่นเดียวกัน

2. ผู้ป่วยที่ตรวจยืนยันทางห้องปฏิบัติการ (Confirmed Cases) ได้แก่ ผู้ป่วยที่มีอาการตามข้อ 1 โดยมีประวัติเกี่ยวข้องกับทางระบาดวิทยาที่เสี่ยงต่อการเกิดโรคเลปโตสไปโรสิส และเจาะเลือดผู้ป่วยในช่วงแรกเริ่ม (เจาะครั้งที่ 1) ไปตรวจ Screening Test โดยวิธี Latex หรือตรวจโดยวิธีอื่นแล้วให้ผล Positive หรือผู้ป่วยที่มีผล Latex negative ครั้งแรกแต่สามารถติดตามเจาะเลือดครั้งที่ 2 ไปตรวจได้ ถ้าให้ผล Positive ก็ให้รายงานเป็น Confirmed Case แต่ถ้าผลการตรวจเลือดครั้งที่ 2 เป็น Negative ก็ให้ตัดจากโรคเลปโตสไปโรสิส เป็นไข้ไม่ทราบสาเหตุ

การวินิจฉัยโรค

1. การสอบประวัติ : โดยสอบถามเกี่ยวกับโอกาสสัมผัสกับสัตว์หรือสิ่งที่ปนเปื้อนกับปัสสาวะสัตว์ เช่น น้ำ ดิน โคลน ท่อระบายน้ำทิ้ง รวมทั้งสุขนิสัยในการบริโภคอาหาร เป็นต้น ซึ่งควรถามย้อนหลังให้ครอบคลุมระยะฟักตัวของโรค คือ ประมาณ 20 – 30 วัน

2. การวินิจฉัยทางคลินิก : โดยอาศัยอาการสำคัญของโรค ซึ่งการใช้แบบบันทึกจะช่วยในการตัดสินใจได้ง่ายขึ้น

องค์การอนามัยโลกแนะนำการกำหนดนิยามผู้ป่วยใน WHO Recommended Surveillance Standards, 1997 ไว้ดังนี้

ผู้ป่วยสงสัย (Suspected Case) ได้แก่ ผู้ที่มีอาการไข้เฉียบพลัน ปวดศีรษะ ปวดกล้ามเนื้อ อ่อนเพลียมาก ร่วมกับอาการใดอาการหนึ่ง คือ ตาแดง เยื่อหุ้มสมองอักเสบ ปัสสาวะน้อย (หรือปัสสาวะไม่ออก) มีโปรตีนในปัสสาวะ ดีซ่าน เลือดออก(ที่ลำไส้ ปอด) การเต้นของหัวใจผิดปกติ (หรือหัวใจล้มเหลว) หรือผื่นที่ผิวหนัง และมีประวัติสัมผัสสัตว์หรือสิ่งปนเปื้อนปัสสาวะสัตว์

ผู้ป่วยยืนยัน (Confirmed Case) ได้แก่ ผู้ป่วยสงสัยที่ได้รับการตรวจยืนยันทางห้องปฏิบัติการที่ได้มาตรฐาน

3. การวินิจฉัยทางห้องปฏิบัติการ กรณีผู้ป่วยมีอาการอย่างอ่อน การใช้การทดสอบทางห้องปฏิบัติการจะเป็นประโยชน์ต่อการดูแลรักษาและการพยากรณ์โรค และหากสามารถบอกชนิดของเชื้อได้ ก็จะช่วยในการป้องกันควบคุมโรคในชุมชนได้อีกด้วย

สถานการณ์ของโรคเลปโตสไปโรซิส

ตารางที่ 5 จำนวนผู้ป่วยและผู้ตายด้วยโรคเลปโตสไปโรซิส แยกตามปี พ.ศ. 2531-2543

พ.ศ.	จำนวนผู้ป่วย	จำนวนผู้ตาย
2531	272	12
2532	207	5
2533	190	8
2534	158	2
2535	166	0
2536	102	0
2537	100	7
2538	143	4
2539	358	19
2540	2,331	111
2541	2,230	102
2542	6,080	266
2543	14,267	362

ที่มา : กระทรวงสาธารณสุข, กรมควบคุมโรคติดต่อ, คู่มือการป้องกันและควบคุมโรค
 เลปโตสไปโรซิสสู่การมีส่วนร่วมขององค์การบริหารส่วนตำบล (กรุงเทพมหานคร: โรงพิมพ์
 ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย, 2544), 3.

เกณฑ์การวินิจฉัยผู้ป่วยโรคเลปโตสไปโรสิส
(อิงเกณฑ์การวินิจฉัยของ WHO)

- วิธีการบันทึก
1. ทำเครื่องหมาย ✓ ใน
 2. รวมคะแนนในแต่ละข้อ

การสรุปผลการวินิจฉัยทางคลินิก

เป็นผู้ป่วยโรคเลปโตสไปโรสิส : คะแนนรวมในส่วน ก. หรือ ก และ ข เท่ากับ 26
คะแนนขึ้นไป หรือส่วน ก ข และ ค รวมกันเท่ากับ 25 คะแนนขึ้นไป

น่าจะเป็นผู้ป่วยแต่ไม่เป็นการยืนยัน : คะแนนตั้งแต่ 20 ถึง 25

คำถาม	คำตอบ / คะแนน
-------	---------------

ก. อาการและอาการแสดงของผู้ป่วย

ปวดศีรษะเฉียบพลัน	มี <input type="checkbox"/> = 2	ไม่มี <input type="checkbox"/> = 0
มีไข้	มี <input type="checkbox"/> = 2	ไม่มี <input type="checkbox"/> = 0
ไข้สูงกว่า 39 องศาเซลเซียส	มี <input type="checkbox"/> = 4	ไม่มี <input type="checkbox"/> = 0
ตาแดงจัด (ทั้ง 2 ข้าง)	มี <input type="checkbox"/> = 4	ไม่มี <input type="checkbox"/> = 0
คอแข็ง	มี <input type="checkbox"/> = 4	ไม่มี <input type="checkbox"/> = 0
ปวดกล้ามเนื้อ	มี <input type="checkbox"/> = 4	ไม่มี <input type="checkbox"/> = 0
มีอาการทั้ง 3 อย่าง	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
(ตาแดงจัด ปวดกล้ามเนื้อและเชื้อหุ้มสมองอักเสบ)		
ดีซ่าน	มี <input type="checkbox"/> = 2	ไม่มี <input type="checkbox"/> = 0
มีไขขาวในปัสสาวะ	มี <input type="checkbox"/> = 2	ไม่มี <input type="checkbox"/> = 0
รวมคะแนนทั้งหมดของส่วน ก =		

ข. ปัจจัยทางระบาดวิทยา (คะแนนเต็ม 10 คะแนน ไม่ว่าจะมาก็ปัจจัยเสี่ยง)

เคยสัมผัสโดยตรงหรือคลุกเคล้าใกล้ชิดกับสัตว์ที่บ้าน	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
ที่ทำงาน สถานที่พักผ่อนหย่อนใจหรือที่ท่องเที่ยว หรือ		
ลงไปว่ายน้ำหรืออาบน้ำในภาวะปกติหรือมีน้ำท่วม หรือ	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
แช่อยู่ในน้ำเกินวันละ 6 ชั่วโมง หรือ	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0

คำถาม	คำตอบ / คะแนน	
ถางหญ้า เตรียมที่หว่านกล้า ปักดำ บำรุงรักษาไร่นา หรือ	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
ทำงานหรือย่ำในที่ชื้นแฉะโดยไม่สวมรองเท้าบู๊ท หรือ	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
ทำงานอยู่กับท่อน้ำโสโครก หรือ	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
ทำงานในโรงฆ่าสัตว์ หรือฟาร์มปศุสัตว์ หรือ	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
รับประทานอาหารประเภทเนื้อสัตว์ เครื่องในสัตว์ดิบ	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
หรือดิบๆสุกๆ หรืออาหารค้างคืนที่ไม่มีภาชนะปิด		

รวมคะแนนทั้งหมดของส่วน ข =

ค. ผลการตรวจชั้นสูตรเชื้อในห้องปฏิบัติการ

แยกเชื้อได้จากการเพาะเชื้อ	มี <input type="checkbox"/> = 25	ไม่มี <input type="checkbox"/> = 0
หรือ ผลทดสอบทางซีโร โลยีเป็นบวก+มีเลปโตสไปโรสิสเป็นโรคประจำถิ่น		
ทดสอบครั้งเดียว – ไตเตอร์ต่ำ	มี <input type="checkbox"/> = 2	ไม่มี <input type="checkbox"/> = 0
ทดสอบครั้งเดียว – ไตเตอร์สูง	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
ทดสอบ paired serum – ไตเตอร์เพิ่มขึ้น	มี <input type="checkbox"/> = 25	ไม่มี <input type="checkbox"/> = 0
หรือ ผลทดสอบทางซีโร โลยีเป็นบวก+ไม่ใช่โรคประจำถิ่น		
ทดสอบครั้งเดียว – ไตเตอร์ต่ำ	มี <input type="checkbox"/> = 5	ไม่มี <input type="checkbox"/> = 0
ทดสอบครั้งเดียว – ไตเตอร์สูง	มี <input type="checkbox"/> = 10	ไม่มี <input type="checkbox"/> = 0
ทดสอบ paired serum – ไตเตอร์เพิ่มขึ้น	มี <input type="checkbox"/> = 25	ไม่มี <input type="checkbox"/> = 0

รวมคะแนนทั้งหมดของส่วน ค =

การศึกษางานวิจัยที่เกี่ยวข้อง

ชาญชัยณรงค์ ทรงศาศรี และคณะ (2546) ทำการศึกษาการทำนายผู้ป่วยโรคเลปโตสไปโรซิสด้วยผลการตรวจทางห้องปฏิบัติการและลักษณะทางคลินิก โดยใช้เทคนิคการวิเคราะห์การถดถอยพหุโลจิสติก (Multiple logistic regression) และคัดเลือกตัวแปรอิสระด้วยวิธี Backward elimination ข้อมูลผู้ป่วยที่นำมาศึกษาเป็นผู้ป่วยที่เข้ามารักษาในโรงพยาบาลขอนแก่น กลุ่มตัวอย่างที่ศึกษา 160 ราย แบ่งเป็นกลุ่มผู้ป่วยโรคเลปโตสไปโรซิสและกลุ่มผู้สงสัยโรคเลปโตสไปโรซิส กลุ่มละ 80 ราย ได้สมการในการทำนายผู้ป่วยโรคเลปโตสไปโรซิส ดังนี้

$$P(W) = \frac{\text{Exp}(-3.62 + 2.34W_1 + 2.16W_2 + 1.51W_3 + 1.39W_4 + 0.02W_5 - 1.15W_6)}{1 + \text{Exp}(-3.62 + 2.34W_1 + 2.16W_2 + 1.51W_3 + 1.39W_4 + 0.02W_5 - 1.15W_6)}$$

เมื่อ $P(W)$ คือ โอกาสในการเป็นโรคเลปโตสไปโรซิส

Exp คือ ค่าเอ็กซ์โปเนนเชียลลอการิทึม

W_1 คือ ค่าจำนวนโปรตีนในปัสสาวะ ที่มีค่ามากกว่าหรือเท่ากับ 2 ขึ้นไป

W_2 คือ ค่าครีเอตินิน ที่มีค่ามากกว่าหรือเท่ากับ 2 mg%

W_3 คือ ค่าบิลิรูบินรวม ที่มีค่ามากกว่าหรือเท่ากับ 1 mg%

W_4 คือ การมีอาการปวดศีรษะ

W_5 คือ ค่ายูเรียไนโตรเจนในเลือด (mg%)

W_6 คือ ค่าจำนวนเกล็ดเลือด ที่มีค่าน้อยกว่า 150,000 cell/mm³

ตัวแปร W_1 , W_2 , W_3 , W_4 , W_6 เป็นตัวแปรเชิงกลุ่ม โดยพิจารณาว่ามีค่าผิดปกติแทนด้วย 1 ไม่ผิดปกติแทนด้วย 0 ส่วนตัวแปร W_5 แทนด้วยค่าที่วัดได้

จากการศึกษาพบว่า สมการนี้สามารถจำแนกกลุ่มได้ร้อยละ 80.6 และมีค่าความไว 81.2 ค่าความจำเพาะ 80.0 ผลการทดสอบความเหมาะสมทางสถิติของตัวแบบมีความเหมาะสม (Hosmer-Lemeshow chi-square test, p-value = 0.714)

กราบแก้ว โสภาราษฎร์ (2544) ทำการศึกษาเทคนิคการวิเคราะห์จำแนกแคนนอนนิตัลกับการวิเคราะห์ถดถอยโลจิสติกสำหรับเหตุการณ์ที่พบได้น้อย ประยุกต์ใช้กับข้อมูลผู้ป่วยโรคข้อเสื่อม เพื่อจำแนกผู้ป่วยปกติกับผู้ป่วยที่มีอาการไม่พึงประสงค์จากการใช้ยา ตัวแปรตามคือการมีและไม่มีอาการที่ไม่พึงประสงค์จากการใช้ยา ส่วนตัวแปรอิสระได้แก่ อายุ เพศ ดัชนีมวลกาย ความดันโลหิตซิสโตลิก ค่าBUN ค่าCreatinin ค่าอัตราส่วน SGOT/SGPT ratio จำนวนเกล็ดเลือด ประวัติการเป็นโรคแผลในกระเพาะอาหาร และประวัติการเป็นโรคเบาหวาน ข้อมูลที่ใช้ในการศึกษามีจำนวน 113 ราย เป็นผู้ป่วยปกติ 101 รายและผู้ป่วยที่มีอาการไม่พึงประสงค์จากการใช้ยา 12 ราย

ผลการศึกษาพบว่า วิธีการวิเคราะห์ถดถอยโลจิสติกจัดกลุ่มได้ถูกต้องมากกว่า และเมื่อพิจารณาค่าอัตราความผิดพลาดจริงที่เห็นชัดและค่าประมาณของค่าคาดหวังการจำแนกผิดที่แท้จริงที่ได้จากวิธีวิเคราะห์ถดถอยโลจิสติกจะต่ำกว่าวิธีการวิเคราะห์การจำแนกแคนนอนนิตัลเล็กน้อย ในงานวิจัยนี้ผู้วิจัยแนะนำให้ทำการศึกษาโดยนำค่าความเสียหายอันเนื่องมาจากการจัดกลุ่มผิดมาร่วมพิจารณาด้วยและควรเพิ่มขนาดตัวอย่างให้มากขึ้น

เดือนเพ็ญ สนโต (2545) ทำการศึกษาเปรียบเทียบความสามารถในการจำแนกกลุ่มถูกต้องระหว่างวิธีการวิเคราะห์จำแนกกลุ่มกับวิธีการวิเคราะห์การถดถอยลอจิสติกเมื่อใช้กลุ่มตัวอย่างขนาดต่างๆ เป็นงานวิจัยเชิงทดลองด้วยเทคนิคมอนติคาร์โล โดยใช้โปรแกรมเขียนด้วยภาษาเบสิกสุ่มเลขสุ่มสร้างประชากร 4 กลุ่มกลุ่มละ 10,000 จำนวน แต่ละกลุ่มประกอบด้วยตัวแปรอิสระ 4 ตัว ซึ่งมีการแจกแจงแบบปกติและเมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระแต่ละกลุ่มเท่ากันและตัวแปรตาม 1 ตัวแปร แบ่งเป็น 2, 3, 4 และ 5 กลุ่มโดยกำหนดให้ประชากรแต่ละกลุ่มมีความสามารถในการจำแนกกลุ่มถูกต้องประมาณ 80 เปอร์เซ็นต์ ใช้โปรแกรม SPSS สุ่มตัวอย่างขนาดต่างๆคือ 40 80 120 และ 160 จำนวนด้วยสถิติ 2 วิธีคือการวิเคราะห์จำแนกกลุ่มและการวิเคราะห์การถดถอยลอจิสติก โดยสุ่มตัวอย่างซ้ำ 500 ครั้ง เปรียบเทียบผลทั้ง 2 วิธี

ผลการศึกษาพบว่า สำหรับการจำแนก 2 กลุ่ม เมื่อขนาดตัวอย่างเท่ากับ 40 ทั้ง 2 วิธีสามารถจำแนกกลุ่มได้ถูกต้องไปแตกต่างกัน เมื่อขนาดตัวอย่างเพิ่มเป็น 80 120 และ 160 การวิเคราะห์การจำแนกกลุ่มสามารถจำแนกกลุ่มได้ถูกต้องมากกว่าการวิเคราะห์การถดถอยลอจิสติกและสำหรับการจำแนก 3, 4 และ 5 กลุ่มพบว่าเมื่อสุ่มตัวอย่างขนาด 40, 80, 120 และ 160 การวิเคราะห์การถดถอยลอจิสติกสามารถจำแนกกลุ่มได้ถูกต้องมากกว่าการวิเคราะห์จำแนกกลุ่ม

เมื่อพิจารณาจำนวนกลุ่มที่จำแนกพบว่าสำหรับกรณีจำแนกออกเป็น 2 กลุ่ม ผลการจำแนกได้ถูกต้องโดยใช้การวิเคราะห์จำแนกกลุ่มและการวิเคราะห์การถดถอยลอจิสติก ประมาณ

ร้อยละ 79.98 และ 79.79 สำหรับกรณีจำแนกออกเป็น 3 กลุ่ม ผลการจำแนกได้ถูกต้องโดยใช้การวิเคราะห์จำแนกกลุ่มและการวิเคราะห์การถดถอยลอจิสติก ประมาณร้อยละ 67.51 และ 76.98 สำหรับกรณีจำแนกออกเป็น 4 กลุ่ม ผลการจำแนกได้ถูกต้องโดยใช้การวิเคราะห์จำแนกกลุ่มและการวิเคราะห์การถดถอยลอจิสติก ประมาณร้อยละ 67.33 และ 77.43 สำหรับกรณีจำแนกออกเป็น 5 กลุ่ม ผลการจำแนกได้ถูกต้องโดยใช้การวิเคราะห์จำแนกกลุ่มและการวิเคราะห์การถดถอยลอจิสติก ประมาณร้อยละ 71.19 และ 79.39 โดยที่ผลการจำแนกได้ถูกต้องส่วนใหญ่ลดลงเมื่อขนาดตัวอย่างใหญ่ขึ้น

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยนี้มีวัตถุประสงค์ เพื่อสร้างตัวแบบที่ใช้วินิจฉัยการเป็นโรคเลปโตสไปโรซิส โดยอาศัยเทคนิคทางสถิติที่มีคุณสมบัติในการจำแนกกลุ่มตามสิ่งที่สนใจศึกษา ในที่นี้เลือกพิจารณา 2 เทคนิค ได้แก่ การสร้างตัวแบบโดยใช้การวิเคราะห์จำแนก และการสร้างตัวแบบการวิเคราะห์การถดถอยโลจิสติก โดยในการศึกษาจะใช้ข้อมูลคนไข้ที่เข้าร่วมโครงการวิจัย “การศึกษาเปรียบเทียบประสิทธิผลของยาเซฟโทรพ็อกซิม และ คีอซีซัยคลิน ในการรักษาผู้ป่วยด้วยโรคเลปโตสไปโรซิสที่มีอาการรุนแรง” ซึ่งเป็นโครงการที่ ศ.พญ. ยุพิน ศุพุททมงคล สาขาวิชาโรคติดเชื้อและอายุรศาสตร์เขตร้อน ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ ศิริราชพยาบาล มหาวิทยาลัยมหิดล ทำการศึกษาร่วมกับแพทย์และคณะผู้วิจัยโรงพยาบาลอื่นๆ 6 โรงพยาบาล ได้แก่ โรงพยาบาลเลย โรงพยาบาลมหาราชนครราชสีมา โรงพยาบาลบ้านใหม่ไชยพจน์ โรงพยาบาลอุดรธานี โรงพยาบาลชุมพร และโรงพยาบาลราชบุรี จำนวน 755 ราย สำหรับกรณีที่ 1 และ 3 ตัดข้อมูลสูญหายทิ้งแล้วเหลือข้อมูลที่นำมาวิเคราะห์จำนวน 559 ราย สำหรับกรณีที่ 2 เหลือข้อมูลที่นำมาวิเคราะห์จำนวน 315 ราย

สำหรับข้อมูลชุดใหม่ ซึ่งเป็นผู้ป่วยที่ทำการวิจัยและเก็บข้อมูลในโรงพยาบาลและช่วงเวลาเดียวกัน แยกออกมาเพื่อใช้ในการทดสอบโดยเฉพาะ เพื่อใช้ในการทดสอบประสิทธิภาพของตัวแบบที่สร้าง กรณีที่ 1 และ 3 จำนวน 149 ราย และกรณีที่ 2 จำนวน 105 ราย

ตัวแปรที่ใช้ในการวิจัย

ตัวแปรตาม ตัวแปรตามเป็นตัวแปรเชิงกลุ่ม ซึ่งถูกแยกพิจารณาเป็น 3 กรณี คือ กรณีที่ 1 การจำแนกเป็น 2 คือกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กับกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ

กรณีที่ 2 จำแนกข้อมูลออกเป็น 2 กลุ่มคือกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิสกับกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis

กรณีที่ 3 การจำแนกเป็น 3 กลุ่มคือกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส, กลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ

ตัวแปรอิสระ

ตัวแปรอิสระ ที่นำมาศึกษามีจำนวน 27 ตัว ได้แก่ X_1 (อายุ), X_2 (เพศ), X_3 (การมีหรือไม่มีปัจจัยเสี่ยง), X_4 (ปวดศีรษะ), X_5 (อ่อนเพลีย), X_6 (เบื่ออาหาร), X_7 (เจ็บคอ), X_8 (ไอ), X_9 (เจ็บหน้าอก), X_{10} (มีเลือดออกผิดปกติ), X_{11} (ความรู้สึกริวผิดปกติ), X_{12} (อุณหภูมิร่างกาย), X_{13} (ความดันโลหิตต่ำ), X_{14} (อัสเสบที่ผนังคอหอย), X_{15} (เยื่อตาอักเสบ), X_{16} (เลือดออกใต้ตาขาว), X_{17} (กดเจ็บกล้ามเนื้อ), X_{18} (ตับโต), X_{19} (ต่อมน้ำเหลืองโต), X_{20} (มีจุดเลือดออกใต้ผิวหนัง), X_{21} ($WBC \geq 12,000/\text{cumm.}$), X_{22} ($PMN > 75\%$), X_{23} ($\text{Platelets} < 100,000/\text{cumm.}$), X_{24} (Abnormal Urine Examination), X_{25} (Total Bilirubin $> 3 \text{ mg/dl}$), X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase) และ X_{27} (Abnormal BUN/Cr)

รายละเอียดของตัวแปรแต่ละตัวมีดังนี้

1. ตัวแปรตาม ได้แก่ การเป็นหรือไม่เป็นโรคเลปโตสไปโรซิส โดยยืนยันด้วยผลการตรวจทางห้องปฏิบัติการตามมาตรฐานการวินิจฉัยโรคเหล่านี้ แยกการพิจารณาออกเป็น 3 กรณีคือ

1.1 กรณีที่ 1 การจำแนกเป็น 2 กลุ่ม คือกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส ($\text{Diag1}=1$) กับกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ ($\text{Diag1}=2$)

1.2 กรณีที่ 2 จำแนกข้อมูลออกเป็น 2 กลุ่มคือกลุ่มคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส ($\text{Diag2}=1$) กับกลุ่มที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis ($\text{Diag2}=2$)

1.3 กรณีที่ 3 จำแนกเป็น 3 กลุ่มคือกลุ่มคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส ($\text{Diag3}=1$) , กลุ่มที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis ($\text{Diag3}=2$) และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ ($\text{Diag3}=3$)

2. ตัวแปรอิสระหรือปัจจัยที่มีอิทธิพลต่อการจำแนกกลุ่ม ได้แก่

2.1 อายุ (age) : X_1 อายุเป็นปัจจัยเสี่ยงอย่างหนึ่งเนื่องจากผู้ที่มียาอายุอยู่ในวัยทำงาน จะเสี่ยงต่อการเป็นโรคเลปโตสไปโรซิส มากกว่าเด็กและคนชรา เนื่องจากมีโอกาสที่จะเกี่ยวข้องกับปัจจัยเสี่ยงมากกว่า ข้อมูลที่เก็บเป็นตัวเลขหน่วยเป็นปี

2.2 เพศ (sex) : X_2 เนื่องจากเพศชายมีโอกาสเกี่ยวข้องกับปัจจัยเสี่ยงต่างๆ มากกว่าเพศหญิง เช่น การจับหนู หนูปล่า และเลี้ยงสัตว์ ดังนั้นข้อมูลที่เก็บจะเป็นเพศชายหรือหญิง ในการวิเคราะห์ข้อมูลจะแปลงเป็น ชายเท่ากับ 1 และ หญิงเท่ากับ 2

2.3 การมีหรือไม่มีปัจจัยเสี่ยง : X_3 ผู้ป่วยที่มีปัจจัยเสี่ยงอย่างน้อย 1 อย่างให้มีค่าเป็น 1 ไม่มีเป็น 0

- 2.4 ปวดศีรษะ (Headache) : X_4 ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.5 อ่อนเพลีย (Malaise) : X_5 ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.6 เบื่ออาหาร (Anorexia) : X_6 ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.7 เจ็บคอ (Sorethroat) : X_7 ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.8 ไอ (Cough) : X_8 ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.9 เจ็บหน้าอก (Chest pain) : X_9 ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.10 มีเลือดออกผิดปกติ (Bleeding) : X_{10} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.11 ความรู้สึกตัวผิดปกติ(CNS) : X_{11} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.12 อุณหภูมิร่างกาย (Temp) : X_{12} อุณหภูมิร่างกายของผู้ป่วยแรกเริ่ม หน่วยเป็น องศาเซลเซียส
- 2.13 ความดันโลหิตต่ำ (Hypotension) : X_{13} ความดันโลหิตต่ำให้มีค่าเป็น 1 ถ้าไม่ใช่ให้เป็น 0
- 2.14 อักเสบที่ผนังคอหอย (Injected Pharynx) : X_{14} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.15 เยื่อตาอักเสบ (Conjunctival Injection) : X_{15} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.16 เลือดออกใต้ตาขาว(Subconjunctival hemorrhage) : X_{16} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.17 กดเจ็บกล้ามเนื้อน่อง(Calf tenderness) : X_{17} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.18 ตับโต(Hepatomegaly) : X_{18} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.19 ต่อม้ำเหลืองโต(Lymphadenopathy) : X_{19} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.20 มีจุดเลือดออกใต้ผิวหนัง(Petichiae) : X_{20} ถ้ามีอาการให้มีค่าเป็น 1 ไม่มีอาการเป็น 0
- 2.21 $WBC \geq 12,000/cumm$: X_{21} ถ้าค่าผิดปกติให้เป็น 1 ปกติเป็น 0
- 2.22 $PMN > 75\%$: X_{22} ถ้าค่าผิดปกติให้เป็น 1 ปกติเป็น 0
- 2.23 $Platelets < 100,000/cumm.$: X_{23} ถ้าค่าผิดปกติให้เป็น 1 ปกติเป็น 0
- 2.24 Abnormal Urine Examination : X_{24} ถ้าผิดปกติให้เป็น 1 ปกติเป็น 0
- 2.25 Total Bilirubin > 3 mg/dl : X_{25} ถ้าผิดปกติให้เป็น 1 ปกติเป็น 0

2.26 Abnormal SGOT/SGPT/Alkaline Phosphatase : X_{26} ถ้ามีค่าผิดปกติให้เป็น 1 ปกติเป็น 0

2.27 Abnormal BUN/Cr : X_{27} ถ้ามีค่าผิดปกติให้เป็น 1 ปกติเป็น 0

ขั้นตอนการวิเคราะห์ข้อมูล

1. วิเคราะห์ข้อมูลเบื้องต้น ในทั้ง 3 กรณีที่ศึกษา จะทำการวิเคราะห์ข้อมูลตามขั้นตอนดังนี้ แบ่งกลุ่มข้อมูลตามกลุ่มของแต่ละกรณีนั้น หาค่าเฉลี่ย ค่าสูงสุด ค่าต่ำสุด สำหรับตัวแปรต่อเนื่อง และใช้วิธีการแจกแจงความถี่ โดยนำเสนอในรูปแบบของจำนวนและร้อยละ เมื่อตัวแปรเป็นตัวแปรเชิงกลุ่ม

2. เปรียบเทียบความแตกต่างอย่างมีนัยสำคัญ ของตัวแปรอิสระแต่ละตัวตามกลุ่มที่จะทำการจำแนกเพื่อเลือกตัวแปรที่จะนำเข้ามาสมการสำหรับแต่ละกรณี โดยใช้วิธีการต่อไปนี้

- สำหรับตัวแปรต่อเนื่องที่มีการแจกแจงแบบปกติใช้ตัวสถิติทดสอบ t ในกรณีที่ตัวอย่างถูกสุ่มมาอย่างเป็นอิสระกัน (Independent t-Test)

- สำหรับตัวแปรต่อเนื่องที่ไม่ได้มีการแจกแจงแบบปกติใช้การทดสอบสถิติไม่ใช้พารามิเตอร์ (Nonparametric Test)

- สำหรับตัวแปรเชิงกลุ่มใช้การทดสอบไคสแควร์ (Chi-Square Test)

3. วิเคราะห์และสร้างฟังก์ชันจำแนกสำหรับแต่ละกรณีที่ทำการศึกษา โดยใช้การวิเคราะห์การจำแนก คำนวณผลการจัดเข้ากลุ่มถูกต้องรวมและผลการจัดเข้ากลุ่มถูกต้องเฉพาะกรณีคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส

4. วิเคราะห์และสร้างสมการถดถอยโลจิสติกสำหรับแต่ละกรณีที่ทำการศึกษา โดยใช้การวิเคราะห์การถดถอยโลจิสติก คำนวณผลการจัดเข้ากลุ่มถูกต้องรวมและผลการจัดเข้ากลุ่มถูกต้องเฉพาะกรณีคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส

5. เปรียบเทียบเปอร์เซ็นต์การจำแนกถูกรวม , เปอร์เซ็นต์การจำแนกถูกเฉพาะกรณีคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส และอัตราความผิดพลาดในการจัดเข้ากลุ่ม ของการวิเคราะห์การจำแนกและการวิเคราะห์การถดถอยโลจิสติก

6. นำตัวแบบที่ได้จากทั้งสองเทคนิค สำหรับทั้ง 3 กรณี ไปทดสอบประสิทธิภาพการนำไปใช้กับข้อมูลชุดใหม่ ซึ่งเป็นผู้ป่วยที่ทำการวิจัยและเก็บข้อมูลในโรงพยาบาลและช่วงเวลาเดียวกัน สรุปเป็นเปอร์เซ็นต์การจำแนกถูกรวม , เปอร์เซ็นต์การจำแนกถูกเฉพาะกรณีคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส และอัตราความผิดพลาดในการจัดเข้ากลุ่ม

7. เปรียบเทียบการจัดข้อมูลเข้ากลุ่มถูกต้องเฉพาะกรณีที่ 1 คือการจำแนกเป็น 2 กลุ่ม กลุ่มคนไข้ในโครงการที่เมื่อเสร็จสิ้นโครงการแล้ววินิจฉัยเป็นโรคเลปโตสไปโรซิส กับกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ ของทั้ง 2 เทคนิค เทียบกับผลที่ได้กับผลการจัดเข้ากลุ่มโดยใช้ตัวแบบที่อิงเกณฑ์ขององค์การอนามัยโลก

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในงานวิจัยนี้ทำการวิเคราะห์ข้อมูลโดยใช้การวิเคราะห์จำแนก และการวิเคราะห์การถดถอยโลจิสติก ข้อมูลที่นำมาใช้เป็นข้อมูลคนไข้ที่เข้าร่วมโครงการวิจัย “การศึกษาเปรียบเทียบประสิทธิผลของยาเซฟโทรแทกซิม และ คีออสซัลคลิน ในการรักษาผู้ป่วยด้วยโรคเลปโตสไปโรซิสที่มีอาการรุนแรง” ซึ่ง ศ.พญ. ยุพิน ศุภุทธมงคล ทำการศึกษาร่วมกับแพทย์และคณะผู้วิจัยโรงพยาบาลอื่นๆ ในการศึกษาแบ่งการพิจารณาออกเป็น 3 กรณี คือ (1) กรณีที่ 1 การจำแนกเป็น 2 กลุ่มคือกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กับกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ (2) กรณีที่ 2 จำแนกข้อมูลออกเป็น 2 กลุ่มคือกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กับกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis (3) กรณีที่ 3 การจำแนกเป็น 3 กลุ่มคือกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส, กลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ

โดยพิจารณาตัวแปรอิสระประกอบด้วยข้อมูล 4 ส่วนคือ ประวัติของคนไข้ อาการทางคลินิกจากประวัติ ผลตรวจร่างกายแรกรับ และผลตรวจทางห้องปฏิบัติการแรกรับ มีทั้งหมด 27 ตัวแปร ได้แก่

1. X_1 (อายุ)
2. X_2 (เพศ)
3. X_3 (การมีหรือไม่มีปัจจัยเสี่ยง)
4. X_4 (ปวดศีรษะ)
5. X_5 (อ่อนเพลีย)
6. X_6 (เบื่ออาหาร)
7. X_7 (เจ็บคอ)
8. X_8 (ไอ)
9. X_9 (เจ็บหน้าอก)
10. X_{10} (มีเลือดออกผิดปกติ)
11. X_{11} (ความรู้สึกตัวผิดปกติ)
12. X_{12} (อุณหภูมิร่างกาย)

13. X_{13} (ความดันโลหิตต่ำ)
14. X_{14} (อักเสบที่ผนังคอหอย)
15. X_{15} (เยื่อตาอักเสบ)
16. X_{16} (เลือดออกใต้ตาขาว)
17. X_{17} (กดเจ็บกล้ามเนื้อ)
18. X_{18} (ตับโต)
19. X_{19} (ต่อมน้ำเหลืองโต)
20. X_{20} (มีจุดเลือดออกใต้ผิวหนัง)
21. X_{21} ($WBC \geq 12,000/cumm.$)
22. X_{22} ($PMN > 75\%$)
23. X_{23} ($Platelets < 100,000/cumm.$)
24. X_{24} (Abnormal Urine Examination)
25. X_{25} (Total Bilirubin > 3 mg/dl)
26. X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase)
27. X_{27} (Abnormal BUN/Cr)

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์

จากนั้นนำตัวแปรอิสระทั้งหมดมาผ่านกระบวนการเลือกจากแต่ละวิธีอีกชั้นหนึ่ง ก่อนทำการสร้างสมการวิเคราะห์จำแนก และสมการวิเคราะห์การถดถอยโลจิสติก สำหรับเทคนิคการวิเคราะห์การจำแนกคัดเลือกตัวแปรเข้าด้วยวิธีการแบบ Stepwise Method ส่วนการถดถอยโลจิสติก จะเลือกตัวแปรเข้าสมการ โดยวิธี Backward Stepwise (Likelihood Ratio)

จำนวนข้อมูลที่นำมาใช้ในการวิเคราะห์ 559 , 149 และ 559 ค่าสังเกต สำหรับกรณีที่ 1,2 และ 3 ตามลำดับ ซึ่งเป็นข้อมูลที่ตัดค่าที่สูญหาย (Missing Value) ทิ้งแล้ว ขั้นตอนในการวิเคราะห์ข้อมูล ประกอบด้วย

1. การสร้างตัวแบบโดยใช้การวิเคราะห์จำแนกและการวิเคราะห์การถดถอยโลจิสติก
 - 1.1 ผลการวิเคราะห์ข้อมูลกรณีที่ 1 หรือจำแนกข้อมูลออกเป็น 2 กลุ่ม คือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ
 - การตรวจสอบข้อสมมติของแต่ละเทคนิค
 - ผลการวิเคราะห์โดยใช้การวิเคราะห์การจำแนก
 - ผลการวิเคราะห์โดยใช้การวิเคราะห์การถดถอยโลจิสติก
 - 1.2 ผลการวิเคราะห์ข้อมูลกรณีที่ 2 หรือจำแนกข้อมูลออกเป็น 2 กลุ่ม คือกลุ่มที่วินิจฉัย

เป็นโรคเลปโตสไปโรซิสและกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis

- การตรวจสอบข้อสมมติของแต่ละเทคนิค
- ผลการวิเคราะห์โดยใช้การวิเคราะห์การจำแนก
- ผลการวิเคราะห์โดยใช้การวิเคราะห์การถดถอยโลจิสติก

1.3 ผลการวิเคราะห์ข้อมูลกรณี 3 หรือจำแนกข้อมูลออกเป็น 3 กลุ่ม คือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรซิส, กลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ

- การตรวจสอบข้อสมมติของแต่ละเทคนิค
- ผลการวิเคราะห์โดยใช้การวิเคราะห์การจำแนก
- ผลการวิเคราะห์โดยใช้การวิเคราะห์การถดถอยโลจิสติก

2. การตรวจสอบ(Validation) ตัวแบบโดยใช้ข้อมูลชุดใหม่
3. การเปรียบเทียบผลการจัดข้อมูลเข้ากลุ่มของตัวแบบที่สร้างและเกณฑ์ขององค์การอนามัยโลก

1. การสร้างตัวแบบโดยใช้การวิเคราะห์จำแนกและการวิเคราะห์การถดถอยโลจิสติก

1.1 ผลการวิเคราะห์สำหรับการจำแนกกรณี 1

การตรวจสอบข้อสมมติของเทคนิคการวิเคราะห์การจำแนก

- ข้อมูลมีการแจกแจงแบบปกติพหุ
ตัวแปรอิสระที่นำมาศึกษามีลักษณะเชิงผสม คือ มีทั้งตัวแปรแบบต่อเนื่องและตัวแปรเชิงกลุ่ม ซึ่งเป็นเรื่องยากที่จะมีการแจกแจงแบบปกติพหุ แต่เนื่องจากจำนวนค่าสังเกตที่ใช้ในงานวิจัยนี้มีขนาดใหญ่ รวมทั้งวัตถุประสงค์ของงานวิจัยคือต้องการสร้างตัวแบบที่มีประสิทธิภาพในการจำแนกข้อมูลเข้ากลุ่มได้ถูกต้องที่สุด ดังนั้นข้อสมมติที่ว่าข้อมูลต้องมีการแจกแจงแบบปกติพหุจึงไม่ใช่ข้อสมมติที่จำเป็น (Barbara and Linda 2001 :462-492)
- ข้อมูลมีเมทริกซ์ความแปรปรวน ความแปรปรวนร่วมเท่ากัน
เมื่อทำการตรวจสอบข้อสมมติข้อนี้โดยใช้สถิติ Box'M ผลแสดงดังตารางที่ 6

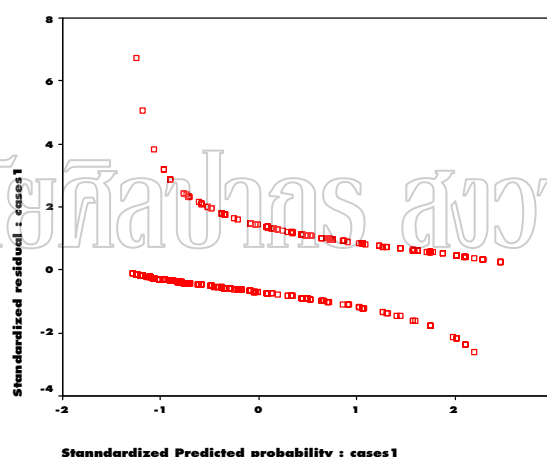
ตารางที่ 6 ค่าสถิติ Box's M ของการจำแนกกรณี 1

Box's M	F Approx.	df1	df2	Sig.
103.741	2.830	36	477485.2	0.000

พิจารณาค่า P-Value จากตารางเท่ากับ 0.000 นั่นคือ สำหรับการจำแนกกรณีที่ 1 มีเมทริกซ์ความแปรปรวนร่วม (Variance-Covariance Matrices) ไม่เท่ากัน ซึ่งไม่สอดคล้องกับข้อสมมติของเทคนิคการวิเคราะห์จำแนกแบบเชิงเส้น (Linear Discriminant Analysis) จึงพิจารณาเลือกการใช้เมทริกซ์ความแปรปรวนร่วมแบบแยกกลุ่ม (Separate-Groups Covariance Matrices) ในการสร้างตัวแบบการวิเคราะห์การจำแนก ซึ่งเหมาะสมกับข้อมูลมากกว่าการใช้เมทริกซ์ความแปรปรวนร่วมแบบรวม (Pooled Within-Group Covariance Matrices) (Barbara and Linda 2001 :461)

การตรวจสอบข้อสมมติของเทคนิคการวิเคราะห์การถดถอยโลจิสติก

- พล็อตของส่วนเหลือเทียบกับค่าประมาณความน่าจะเป็นจะมีการกระจายอย่างเป็นอิสระรอบค่า 0



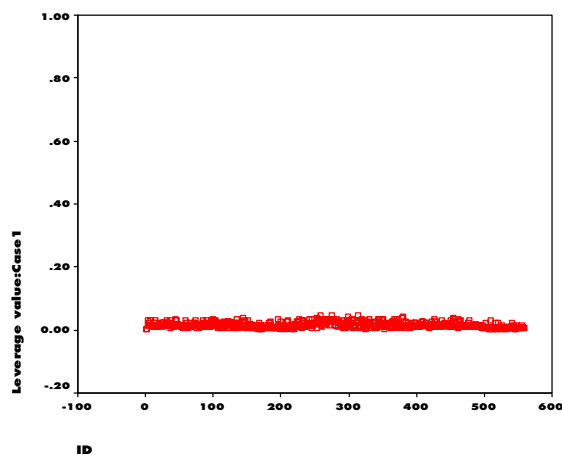
รูปที่ 5 การกระจายของส่วนเหลือ (standard residual) กับค่าพยากรณ์ความน่าจะเป็นของกรณีที่ 1

รูปที่ 5 เป็นการพลอตการกระจายของส่วนเหลือ เพื่อตรวจสอบว่าตัวแบบที่นำมาใช้กับข้อมูลตัวอย่างสอดคล้องกับข้อตกลงเบื้องต้นหรือไม่ จะเห็นได้ว่าค่าส่วนเหลือมาตรฐาน (Standard Residual) ของค่าสังเกตส่วนใหญ่จะมีค่าอยู่รอบๆค่าศูนย์ มีค่าไม่เกิน 3 หรือไม่น้อยกว่า -3 มีเพียง 5 ค่าสังเกตที่มากกว่า 3 ได้แก่ CHUP032, LL060, CHUP052, BL073 และ CHUP010 โดยมีค่าส่วนเหลือมาตรฐานเท่ากับ 6.72643, 5.05382, 3.82600, 3.17840 และ 3.17443 ตามลำดับ และเมื่อพิจารณาจากรูปแบบการกระจายของส่วนเหลือพบว่า การกระจายตัวยังไม่เป็นอิสระเท่าที่ควร

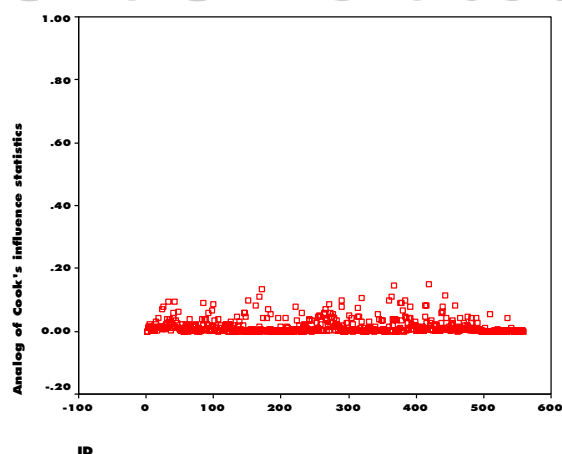
แต่เมื่อทำการตรวจสอบข้อมูลแล้วพบว่าค่าสังเกตที่มีส่วนเหลือมากกว่า 3 หรือน้อยกว่า -3 มีเพียง 5 ค่าสังเกตจากทั้งหมด 559 ค่าสังเกตซึ่งถือว่าเป็นค่าสังเกตส่วนน้อย รวมทั้ง

ข้อมูลดังกล่าวเป็นข้อมูลที่ถูกต้องแล้ว หากตัดออกอาจทำให้สูญเสียสารสนเทศบางอย่างเกี่ยวกับข้อมูลชุดนี้ จึงเก็บค่าสังเกตเหล่านี้ไว้วิเคราะห์ร่วมด้วย

- ตรวจสอบค่าผิดปกติ (outlier observation)



รูปที่ 6 การกระจายของค่า Leverage กับหมายเลขประจำของค่าสังเกต กรณีที่ 1



รูปที่ 7 การกระจายของค่า Cook's distance กับหมายเลขประจำของค่าสังเกต กรณีที่ 1

พิจารณารูปที่ 6 และ 7 เป็นการพลอตระหว่างค่า Leverage และ ค่า Cook's distance กับหมายเลขประจำของค่าสังเกตของกรณีที่ 1 เพื่อทำการตรวจสอบหาค่าผิดปกติของข้อมูล จากการสังเกตพบว่าไม่มีค่าที่ผิดปกติ

กรณีที่ 1 เป็นการสร้างตัวแบบที่ใช้จำแนกกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส กับกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ ในงานวิจัยนี้จะทำการสร้างตัวแบบจำแนกโดยใช้ 2 เทคนิคคือ การวิเคราะห์การจำแนกและการวิเคราะห์การถดถอยโลจิสติก ผลการวิจัยนำเสนอเป็นลำดับดังนี้

ขั้นตอนแรกจะทำการเลือกตัวแปรอิสระจากกลุ่มตัวแปรที่สนใจศึกษาทั้งหมด 27 ตัวแปรนำมาตรวจสอบความแตกต่างกันของข้อมูลในกลุ่มที่จะจำแนก โดยใช้การทดสอบสถิติไม่ใช้พารามิเตอร์ และการทดสอบไคสแควร์ ผลแสดงดังตารางที่ 7 พบว่าตัวแปรที่สมควรนำเข้ามาสมการสำหรับการจำแนกข้อมูลเป็น 2 กลุ่มคือกลุ่มที่วินิจฉัยเป็นโรค เลปโตสไปโรสิสและกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ มี 16 ตัวแปร

ผลการวิเคราะห์ข้อมูลเบื้องต้น

ตารางที่ 7 ข้อมูลพื้นฐานของกลุ่มที่ศึกษากรณีที่ 1

ตัวแปร	ผู้ป่วยที่วินิจฉัยเป็นโรคเลปโตสไปโรสิส (N=240)	ผู้ป่วยที่วินิจฉัยเป็นโรคอื่นๆ (N=515)	P-Value
ประวัติ			
X ₁ (อายุ) ; Median(Range)	35 (15-77)	38 (6-87)	0.142
X ₂ (เพศ) ; ชาย/หญิง	202/38	380/135	0.002
X ₃ (ประวัติเสี่ยง) ; มีประวัติเสี่ยง (%)	234 (97.5%)	458 (88.9%)	0.000
อาการทางคลินิกจากประวัติ ; n(%)			
X ₄ (ปวดศีรษะ)	222 (92.5%)	448 (87.0%)	0.026
X ₅ (อ่อนเพลีย)	233 (97.1%)	494 (97.1%)	0.432
X ₆ (เบื่ออาหาร)	194 (80.8%)	376 (73.0%)	0.020
X ₇ (เจ็บคอ)	140 (58.3%)	198 (38.4%)	0.000
X ₈ (ไอ)	128 (53.3%)	242 (47.0%)	0.104
X ₉ (เจ็บหน้าอก)	100 (41.7%)	113 (21.9%)	0.000
X ₁₀ (เลือดออกผิดปกติ)	27 (11.3%)	28 (5.4%)	0.004
X ₁₁ (ความรู้สึกรู้สึกตัวผิดปกติ)	57 (23.8%)	115 (22.3%)	0.665

ตารางที่ 7 (ต่อ)

ตัวแปร	ผู้ป่วยที่วินิจฉัยเป็น โรคเลปโตสไปโรซิส (N=240)	ผู้ป่วยที่วินิจฉัย โรคอื่นๆ (N=515)	P-Value
การตรวจร่างกายแรกเริ่ม ; n(%)			
X ₁₂ (อุณหภูมิร่างกาย) ; Median(Range)	38.0 (36.0-40.5)	38.1 (36.0-41.0)	0.089
X ₁₃ (ความดันโลหิตต่ำ)	25 (10.4%)	33 (6.4%)	0.054
X ₁₄ (อักเสบที่ผนังคอหอย)	52 (21.7%)	84 (16.3%)	0.075
X ₁₅ (เยื่อตาอักเสบ)	185 (77.1%)	228 (44.3%)	0.000
X ₁₆ (เลือดออกใต้ตาขาว)	32 (13.3%)	22 (4.3%)	0.000
X ₁₇ (กดเจ็บกล้ามเนื้อ)	101 (42.1%)	123 (23.9%)	0.000
X ₁₈ (ตับโต)	31 (12.9%)	63 (12.2%)	0.791
X ₁₉ (ต่อมน้ำเหลืองโต)	76 (31.7%)	139 (27.0%)	0.185
X ₂₀ (มีจุดเลือดออกใต้ผิวหนัง)	6 (2.5%)	8 (1.6%)	0.369
การตรวจทางห้องปฏิบัติการแรกเริ่ม ; n(%)			
CBC			
X ₂₁ (WBC≥12,000/cumm.)	72 (30.6%)	131 (26.1%)	0.204
X ₂₂ (PMN>75%)	183 (78.5%)	272 (54.9%)	0.000
X ₂₃ (Platelets < 100,000/cumm.)	74 (35.2%)	81 (18.5%)	0.000
X ₂₄ (Abnormal Urine Examination)	83 (36.7%)	119 (25.3%)	0.002
X ₂₅ (Total Bilirubin >3 mg/dl)	71 (30.6%)	68 (13.9%)	0.000
X ₂₆ (Abnormal SGOT/SGPT/Alkaline Phosphatase)	49 (21.4%)	160 (32.9%)	0.002
X ₂₇ (Abnormal BUN/Cr)	74 (31.0%)	78 (15.3%)	0.000
WHO score > 25	8 (3.3%)	10 (1.9%)	0.243

ตารางที่ 8 ค่าเฉลี่ยของผลตรวจทางห้องปฏิบัติการสำหรับกรณีที่ 1

ตัวแปร	ผู้ป่วยที่วินิจฉัยเป็น โรคเลปโตสไปโรสิส (N=240)	ผู้ป่วยที่วินิจฉัย เป็นโรคอื่นๆ (N=515)	P-Value
CBC ; Median(Range)			
WBC (/cu.mm)	9.3 (1.6-72.4)	8.8 (1.4-529)	0.038
%PMN	84.9 (26-98)	77.0 (14-98)	0.000
Platelets (x10 ³ /cu.mm)	147 (9-515)	192.00 (10-772)	0.000
Urine Examination ; n(%)			
Protein >1+	35 (18.9%)	62 (15.8%)	0.352
Abnormal Sediment	73 (32.3%)	82 (17.8%)	0.000
Liver Function Test; Median(Range)			
Total Bilirubin (mg/dl)	1.5 (0.1-37.8)	0.8 (0.1-47.7)	0.000
SGOT (U/L)	53 (6-519)	58 (5-3075)	0.329
SGPT (U/L)	43 (6-259)	42 (4-2750)	0.182
Alkaline Phosphatase (U/L)	139 (13-673)	137 (4-1690)	0.565
Albumin (g/dl)	3.2 (1.6-4.7)	3.3 (1.3-5.2)	0.007
Renal Function Test			
BUN (mg/dl)	20 (6-155)	15 (1-164)	0.000
Creatinine (mg/dl)	1.4 (0.1-19.5)	1.1 (0.05-12.6)	0.000
Serum Sodium (mEq/L)	137.8 (3.7-151.5)	136.0 (112.2-155.6)	0.026
Serum Potassium (mEq/L)	3.8 (2.1-7.7)	3.8 (2.5-8.5)	0.088
Serum Bicarbonate (mEq/L)	20.3 (3.2-27.9)	21.8 (4.6-34.0)	0.001

1.1.1 การสร้างตัวแบบโดยใช้เทคนิคการวิเคราะห์การจำแนกสำหรับการจำแนกกรณีที่1 ทำการสร้างสมการจำแนกโดยใช้โปรแกรม SPSS for Window Version 11.5 จำนวนค่าสังเกตที่ใช้ในงานวิจัย 559 ค่าสังเกต ตัวแปรที่นำมาพิจารณาสำหรับกรณีที่ 1 นี้มี 16 ตัวแปร คือ X_2 (เพศ), X_3 (การมีหรือไม่มีปัจจัยเสี่ยง), X_4 (ปวดศีรษะ), X_6 (เบื่ออาหาร), X_7 (เจ็บคอ), X_9 (เจ็บหน้าอก),

X_{10} (มีเลือดออกผิปกติ), X_{15} (เยื่อบุตาอักเสบ), X_{16} (เลือดออกใต้ตาขาว) X_{17} (กดเจ็บกล้ามเนื้อน่อง), X_{22} (PMN>75%), X_{23} (Platelets < 100,000/cumm.), X_{24} (Abnormal Urine Examination), X_{25} (Total Bilirubin >3 mg/dl), X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase), X_{27} (Abnormal BUN/Cr) ใช้การเลือกตัวแปรเข้าโดยวิธี Stepwise Method พบว่ามีตัวแปรที่เข้าสมการทั้งหมด 8 ตัว ผลการวิเคราะห์แสดงเป็นลำดับดังนี้

ตารางที่ 9 ผลการคัดเลือกตัวแปรเข้าสมการโดยวิธี Stepwise สำหรับกรณีที่ 1

Step	Entered	Wilk's Lambda		
		Statistic	Exact F	
			Statistic	Sig
1	X_{15}	0.891	68.340	0.000
2	X_{23}	0.844	51.340	0.000
3	X_{22}	0.810	43.477	0.000
4	X_{26}	0.786	37.749	0.000
5	X_{25}	0.768	33.426	0.000
6	X_{17}	0.756	29.764	0.000
7	X_9	0.745	26.984	0.000
8	X_6	0.738	24.388	0.000

จากตารางที่ 9 พบว่าผลการเลือกตัวแปรเข้าสมการวิเคราะห์จำแนกโดยใช้วิธี Stepwise มีตัวแปรเข้าสมการทั้งหมด 8 ตัว ได้แก่ X_{15} (เยื่อบุตาอักเสบ), X_{23} (Platelets < 100,000/cumm.), X_{22} (PMN>75%), X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase), X_{25} (Total Bilirubin >3 mg/dl), X_{17} (กดเจ็บกล้ามเนื้อน่อง), X_9 (เจ็บหน้าอก) และ X_6 (เบื่ออาหาร) ตามลำดับ

ตารางที่ 10 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนกด้วยค่าไอเกน (Eigenvalue) สำหรับกรณีที่ 1

Eigenvalue	Percent of Varince	Cumulative percent	Canonical correlation
0.355	100.0	100.0	0.512

ตารางที่ 11 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนก
ด้วยค่า Wilks' Lambda สำหรับกรณีที่ 1

Wilks' Lambda	Chi-square	df	Sig.
0.738	167.896	8	0.000

เมื่อพิจารณาค่าไอเกน จากตารางที่ 10 เท่ากับ 0.355 นั่นคือสมการวิเคราะห์จำแนกที่
สร้างนี้มีประสิทธิภาพในการจำแนกกลุ่มข้อมูลที่ใช้ในเกณฑ์ที่ไม่ดีนักเนื่องจากค่าไอเกน ยิ่งมาก
จะแสดงถึงประสิทธิภาพในการจำแนกที่สูงตามไปด้วย และค่าสหสัมพันธ์แคนนอนิคัล (Canonical
correlation) เท่ากับ 0.512 สรุปได้ว่า $(0.512)^2$ หรือฟังก์ชันที่ได้จากการวิเคราะห์การจำแนก
สามารถอธิบายความผันแปรของกลุ่มที่จำแนกได้ 26.21%

จากตารางที่ 11 เมื่อพิจารณาค่า Wilks' Lambda ซึ่งเท่ากับ 0.738 และมี P-Value
เท่ากับ 0.000 นั่นคือ สมการวิเคราะห์จำแนกที่สร้างได้นี้สามารถจำแนกกลุ่มได้อย่างมีนัยสำคัญ

ตารางที่ 12 ค่าศูนย์กลางของกลุ่ม (Group centroids) ของข้อมูลทั้ง 2 กลุ่มสำหรับกรณีที่ 1

Group	Function at Group centroids
Leptospirosis	0.849
Others	-0.416

จากตารางที่ 12 ซึ่งแสดงค่าศูนย์กลางของกลุ่มข้อมูลทั้ง 2 กลุ่ม โดยกลุ่มที่วินิจฉัยว่าเป็น
โรคเลปโตสไปโรสิสมีค่าศูนย์กลางของกลุ่มเท่ากับ 0.849 ส่วนกลุ่มที่วินิจฉัยว่าเป็นโรคอื่นๆ
มีค่าศูนย์กลางของกลุ่มเท่ากับ -0.416 ซึ่งเป็นค่าที่ห่างกันพอสมควร ดังนั้นการสร้างสมการ
วิเคราะห์จำแนกจากข้อมูลชุดนี้น่าจะมีประสิทธิภาพค่อนข้างดี

ตารางที่ 13 ค่าสัมประสิทธิ์คานอนิคล์(Unstandardized coefficients)ค่าสัมประสิทธิ์คานอนิคล์มาตรฐาน(Standardized Coefficients) สหสัมพันธ์กับฟังก์ชันการจำแนกของตัวแปรอิสระและสัมประสิทธิ์ฟังก์ชันจำแนกสำหรับกรณีที่ 1

Variables	Unstandardized Coefficients	Standardized Coefficients	Correlation	Discriminant Function Coefficient	
				Leptospirosis	Others
X ₆	0.440	0.185	0.179	4.828	4.272
X ₉	0.519	0.231	0.408	0.914	0.257
X ₁₅	1.206	0.568	0.588	3.504	1.978
X ₁₇	0.515	0.236	0.392	1.258	0.607
X ₂₂	0.755	0.358	0.375	3.422	2.467
X ₂₃	0.731	0.308	0.399	1.540	0.615
X ₂₅	0.782	0.309	0.348	1.096	0.106
X ₂₆	-0.865	-0.388	-0.209	0.202	1.296
Constant	-1.878			-6.391	-3.741

จากตารางที่ 13 เมื่อพิจารณาค่าสัมประสิทธิ์มาตรฐานจากตาราง พบว่าตัวแปร X₁₅ (เยื่อตาอักเสบ) เป็นตัวแปรที่มีความสำคัญต่อประสิทธิภาพในการจำแนกของสมการจำแนกกลุ่มมากกว่าตัวแปรอื่นๆ เนื่องจากมีค่าสัมประสิทธิ์มาตรฐานที่สูงกว่า

พิจารณาค่าสหสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 8 ตัวและฟังก์ชันการวิเคราะห์จำแนก พบว่าตัวแปรที่นำมาใช้ในการสร้างสมการวิเคราะห์จำแนก ตัวแปร X₁₅ (เยื่อตาอักเสบ) มีค่าสัมบูรณ์ของค่าสหสัมพันธ์สูงที่สุด

จากค่าสัมประสิทธิ์ฟังก์ชันจำแนก จะได้สมการการวิเคราะห์การจำแนกดังนี้

$$Z_1 = -6.391 + (4.828 X_6) + (0.914 X_9) + (3.504 X_{15}) + (1.258 X_{17}) + (3.422 X_{22}) + (1.540 X_{23}) + (1.096 X_{25}) + (0.202 X_{26})$$

$$Z_2 = -3.741 + (4.272 X_6) + (0.257 X_9) + (1.978 X_{15}) + (0.607 X_{17}) + (2.467 X_{22}) + (0.615 X_{23}) + (1.296 X_{25}) + (1.296 X_{26})$$

ตารางที่ 14 ผลการจัดข้อมูลเข้ากลุ่ม เมื่อใช้ฟังก์ชันการวิเคราะห์จำแนกสำหรับกรณีที่ 1

Original Diagnosis	Predicted Diagnosis		Percentage Correct	Overall Percentage Correct
	Leptospirosis	Others		
Leptospirosis	140	44	76.10	74.60
Others	98	277	73.90	

เมื่อทำการสร้างฟังก์ชันการวิเคราะห์จำแนกโดยกำหนดความน่าจะเป็นเบื้องต้น (Prior Probability) ของทั้ง 2 กลุ่มเท่ากัน คือ เท่ากับ 0.50 ทั้งสองกลุ่ม จะได้ค่าสัมประสิทธิ์ของตัวแปรที่ใช้ในการสร้างสมการวิเคราะห์จำแนกทั้ง 8 ตัว แสดงดังตารางที่ 13 ซึ่งเมื่อนำฟังก์ชันการวิเคราะห์จำแนกที่ได้มาทำการจัดข้อมูลเข้ากลุ่ม ได้ผลดังตารางที่ 14 แสดงให้เห็นว่าผลการจัดเข้ากลุ่มถูกต้องรวม (Overall Correctly Classified) เท่ากับ 74.6% ซึ่งมาจากการจำแนกเข้ากลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสถูกต้อง 76.1% และมาจากการจำแนกเข้ากลุ่มที่วินิจฉัยเป็นโรคอื่นๆได้ถูกต้อง 73.9% และเมื่อนำค่าที่ได้จากตารางที่ 14 มาคำนวณหาอัตราการจำแนกผิดที่เห็นชัดเจน (Apparent error rate ;APER) ได้เท่ากับ $(44+98)/559$ หรือมีค่าเท่ากับ 0.2540 ในการประมาณค่าความผิดพลาดของการจำแนกกลุ่มโดยวิธีดังกล่าวจะให้ค่าประมาณที่ต่ำกว่าปกติ (Underestimate) เมื่อจำนวนตัวอย่างมีขนาดเล็ก แต่เนื่องจากข้อมูลที่นำมาวิเคราะห์ในงานวิจัยนี้มีขนาดใหญ่พอจึงสามารถนำค่าการประมาณโดยวิธีนี้มาพิจารณาร่วมด้วย

- 1.1.2 การสร้างตัวแบบการวิเคราะห์การถดถอยโลจิสติกสำหรับการจำแนก กรณีที่ 1 นำข้อมูลชุดเดิมมาวิเคราะห์โดยใช้การวิเคราะห์การถดถอยโลจิสติก นำตัวแปร 16 ตัว มาทำการเลือกตัวแปรสมการเข้าโดยใช้ Backward Stepwise (Likelihood ratio) ผลเสนอในตารางตามลำดับต่อไปนี้

ตารางที่ 15 ผลการวิเคราะห์ข้อมูลเพื่อสร้างสมการถดถอยโลจิสติกสำหรับกรณีที่ 1

Variables	$\hat{\beta}$	S.E. ($\hat{\beta}$)	Wald	Sig.	Exp($\hat{\beta}$)	95% CI. For Exp($\hat{\beta}$)	
						Lower	Upper
X ₆	0.613	0.265	5.333	0.021	1.846	1.097	3.104
X ₉	0.562	0.229	6.045	0.014	1.755	1.121	2.748

ตารางที่ 15 (ต่อ)

Variables	$\hat{\beta}$	S.E.($\hat{\beta}$)	Wald	Sig.	Exp($\hat{\beta}$)	95% CI. For Exp($\hat{\beta}$)	
						Lower	Upper
X ₁₅	1.499	0.235	40.553	0.000	4.479	2.823	7.105
X ₁₇	0.572	0.222	6.639	0.010	1.771	1.147	2.737
X ₂₂	0.986	0.234	17.781	0.000	2.681	1.695	4.240
X ₂₃	0.786	0.270	8.477	0.004	2.194	1.293	3.725
X ₂₅	0.972	0.297	10.689	0.001	2.644	1.476	4.737
X ₂₆	-1.128	0.264	18.312	0.000	0.324	0.193	0.542
Constant	-3.296	0.360	83.852	0.000	0.037		

จากตารางที่ 15 ตัวสถิติ Wald ใช้ในการทดสอบว่าค่าประมาณสัมประสิทธิ์ของตัวแปรนั้นๆมีค่าเท่ากับศูนย์หรือไม่ พิจารณาจากค่า P-Value พบว่ามีตัวแปรที่มีค่าประมาณไม่เท่ากับศูนย์อย่างมีนัยสำคัญ มี 8 ตัว คือ X₆(เบื่ออาหาร), X₉(เจ็บหน้าอก), X₁₅(เยื่อตาอักเสบ), X₁₇(กดเจ็บกล้ามเนื้อ), X₂₂(PMN>75%), X₂₃(Platelets < 100,000/cumm.), X₂₅(Total Bilirubin >3 mg/dl) และ X₂₆(Abnormal SGOT/SGPT/Alkaline Phosphatase) ซึ่งสามารถเขียนลอจิสติกตัวอย่าง (Estimated Logit) สำหรับตัวแบบถดถอยโลจิสติก ได้ดังนี้

$$\hat{g}(x) = -3.296 + (0.613 * X_6) + (0.562 * X_9) + (1.499 * X_{15}) + (0.572 * X_{17}) + (0.986 * X_{22}) + (0.786 * X_{23}) + (0.972 * X_{25}) - (1.128 * X_{26})$$

การตีความจากค่าสัมประสิทธิ์ที่ได้ ทำได้โดยการพิจารณาจากค่า Exp(β) ของแต่ละตัวแปร เช่น ตัวแปร X₁₅ (เยื่อตาอักเสบ) จากตารางมีค่า Exp(β) เท่ากับ 4.479 หมายความว่าผู้ป่วยที่มีอาการเยื่อตาอักเสบ (Conjunctival Injection) มีความเสี่ยงต่อการเป็นโรคเลปโตสไปโรสิส มากกว่าผู้ป่วยที่ไม่มีอาการเยื่อตาอักเสบ 4.479 เท่า

ตารางที่ 16 ผลการตรวจสอบภาวะสารูปดี (Goodness-of-fit Test) และค่า R-Square สำหรับกรณีที่ 1

Method	Value	df	P-Value
Statistic G ²	164.172	8	0.000
Hosmer and Lemeshow	6.134	8	0.632
Nagelkerke R Square	0.354		

ตารางที่ 16 เป็นการตรวจสอบภาวะสารรูปดีของตัวแบบโลจิสติกจากค่า Statistic G^2 ที่ได้จากการทดสอบ Likelihood-ratio test ซึ่งเป็นการเปรียบเทียบตัวแบบ 2 ตัว ซึ่งมีจำนวนสัมประสิทธิ์ที่แตกต่างกันในที่นี้คือตัวแบบที่มีเฉพาะค่าคงที่เทียบกับตัวแบบที่มีตัวแปรอิสระ 8 ตัว เพื่อที่จะทดสอบว่าค่าสัมประสิทธิ์ของตัวแปรทั้ง 8 ตัวนี้มีค่าเท่ากับ 0 หรือไม่ จะเห็นได้ว่าค่า Statistic G^2 มีค่า 164.172 และ P-Value เท่ากับ 0.000 แสดงให้เห็นว่ามีค่าสัมประสิทธิ์บางตัวในจำนวนตัวแปร 8 ตัวนั้นที่มีค่าไม่เท่ากับ 0 อย่างมีนัยสำคัญทางสถิติ (ที่ระดับนัยสำคัญ 0.05) แสดงว่ามีตัวแปรอิสระอย่างน้อยหนึ่งตัวที่มีอิทธิพลต่อตัวแปรตาม ส่วนการทดสอบ Hosmer and Lemeshow เป็นการทดสอบความเหมาะสมของตัวแบบ จากตารางที่ 16 พบว่าค่าสถิติ Hosmer and Lemeshow มีค่าเท่ากับ 6.134 และมีค่า P-Value เท่ากับ 0.632 นั่นคือ ตัวแบบที่สร้างได้มีความเหมาะสม (fit) กับข้อมูลตัวอย่าง

เมื่อพิจารณาค่า R-Square จากตารางที่ 16 ได้ว่า ฟังก์ชันที่ได้จากการวิเคราะห์การถดถอยโลจิสติกสามารถอธิบายความผันแปรของกลุ่มที่จำแนกได้ 35.40%

ตารางที่ 17 จำนวนข้อมูลที่ถูกจัดเข้ากลุ่มถูกต้อง เมื่อใช้จุดตัดการจำแนก (Classification Cut Point) ที่ระดับต่างๆสำหรับกรณีที่ 1

Cut Point	Observed	Predicted		Percentage Correct	Overall Percentage Correct	APER
		Leptospirosis	Others			
0.2	Leptospirosis	163	21	88.6	66.4	0.336
	Others	167	208	55.5		
0.3	Leptospirosis	147	37	79.9	72.3	0.277
	Others	118	257	68.5		
0.4	Leptospirosis	125	59	67.9	75.7	0.243
	Others	77	298	79.5		
0.5	Leptospirosis	97	87	52.7	76.0	0.240
	Others	47	328	87.5		
0.6	Leptospirosis	57	127	31.0	73.9	0.261
	Others	19	356	94.9		
0.7	Leptospirosis	49	135	26.6	73.5	0.265
	Others	13	362	96.5		

ตารางที่ 17 (ต่อ)

Cut Point	Observed	Predicted		Percentage Correct	Overall Percentage Correct	APER
		Leptospirosis	Others			
0.8	Leptospirosis	27	157	14.7	70.8	0.292
	Others	6	369	98.4		

ตารางที่ 17 แสดงถึงผลการจัดข้อมูลเข้ากลุ่ม โดยใช้การสร้างสมการถดถอยโลจิสติก พิจารณาจากการจำแนกเข้ากลุ่มได้ถูกต้องรวมพบว่า เมื่อเลือกจุดตัดการจำแนกเท่ากับ 0.5 จะได้สมการถดถอยโลจิสติกที่จำแนกข้อมูลนี้ได้ถูกต้องมากที่สุด คือเท่ากับ 76.0%

เมื่อพิจารณาแยกเป็นกรณีผู้ป่วยที่เป็นโรคเลปโตสไปโรสิส ผลการใช้สมการถดถอยโลจิสติกจะจำแนกข้อมูลเข้ากลุ่มผู้ป่วยเลปโตสไปโรสิสถูกต้องมากที่สุดเมื่อเลือกจุดตัดการจำแนกเท่ากับ 0.2 และจำแนกถูกลดลงเรื่อยๆเมื่อเลือกจุดตัดการจำแนกที่มากขึ้น ส่วนผู้ป่วยที่วินิจฉัยว่าเป็นโรคอื่นๆ เมื่อใช้สมการถดถอยโลจิสติกจะจำแนกข้อมูลเข้ากลุ่มผู้ป่วยโรคอื่นๆ ได้ถูกต้องน้อยที่สุดเมื่อเลือกจุดตัดการจำแนกเท่ากับ 0.2 และอัตราการจำแนกถูกจะมากขึ้นเรื่อยๆเมื่อเลือกจุดตัดการจำแนกที่เพิ่มขึ้น สำหรับค่า APER พบว่ามีค่าต่ำที่สุดเมื่อกำหนดจุดตัดการจำแนกเท่ากับ 0.5 และที่จุดตัดการจำแนกจุดอื่นๆ APER มีค่าที่สูงขึ้น นั่นคือสมการถดถอยโลจิสติกที่สร้างขึ้นจัดข้อมูลเข้ากลุ่มได้ถูกต้องรวมน้อยลงไปด้วย

สำหรับการนำตัวแบบที่สร้างได้ไปเปรียบเทียบผลการจัดเข้ากลุ่มถูกต้อง โดยเทียบกับผลที่ได้จากตัวแบบการวิเคราะห์การจำแนกและตัวแบบที่อิงเกณฑ์ WHO จะเลือกใช้จุดตัดที่ทำให้ผลการจัดเข้ากลุ่มถูกต้อง กรณีผู้ป่วยที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกรณีผู้ป่วยวินิจฉัยเป็นโรคอื่นๆ มีค่าสูงทั้งสองกรณี ดังนั้นจึงเลือกใช้ที่จุด 0.3 ซึ่งมีเปอร์เซ็นต์การจำแนกถูกรวม จำแนกถูกกรณีผู้ป่วยที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกรณีผู้ป่วยวินิจฉัยเป็นโรคอื่นๆ เท่ากับ 72.3% , 79.9% และ 68.5% ตามลำดับ

เมื่อนำผลการจำแนกข้อมูลเข้ากลุ่มได้ถูกต้อง ทั้งกรณีผู้ป่วยเลปโตสไปโรสิสและผู้ป่วยอื่นๆ รวมทั้งผลการจำแนกรวมและค่า APER มาเปรียบเทียบกัน ระหว่างการใช้วิธีการวิเคราะห์จำแนกและวิธีวิเคราะห์การถดถอยโลจิสติก ผลแสดงดังตารางต่อไปนี้

ตารางที่ 18 ผลการจำแนกกรณีที่ 1 โดยใช้วิธีการวิเคราะห์จำแนกและการถดถอยโลจิสติก

Technique	Observed	Predicted		Percentage Correct	Overall Percentage Correct	APER
		Leptospirosis	Others			
Discriminant	Leptospirosis	140	44	76.10	74.6	0.254
	Others	98	277	73.90		
Logistic Regression	Leptospirosis	147	37	79.9	72.3	0.277
	Others	118	257	68.5		

จากผลการจำแนกข้อมูลเข้ากลุ่มถูกต้องในตารางที่ 18 พบว่า สำหรับกรณีที่ 1 นี้ การวิเคราะห์การจำแนก จะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่าวิธีการวิเคราะห์การถดถอยโลจิสติกเล็กน้อย เมื่อพิจารณาจากผลการจัดข้อมูลเข้ากลุ่มรวม แต่เมื่อแยกพิจารณาพบว่า กรณีคนไข้ที่เป็นโรคเลปโตสไปโรสิส พบว่า ตัวแบบที่ได้จากการวิเคราะห์การถดถอยโลจิสติกมีประสิทธิภาพสูงกว่า

1.2 ผลการวิเคราะห์สำหรับการจำแนกกรณีที่ 2

การตรวจสอบข้อสมมติของเทคนิคการวิเคราะห์การจำแนก

- ข้อมูลมีการแจกแจงแบบปกติพหุ
ตัวแปรอิสระที่นำมาศึกษามีลักษณะเชิงผสม คือ มีทั้งตัวแปรแบบต่อเนื่องและตัวแปรเชิงกลุ่ม ซึ่งเป็นเรื่องยากที่จะมีการแจกแจงแบบปกติพหุ แต่เนื่องจากจำนวนค่าสังเกตที่ใช้ในงานวิจัยนี้มีขนาดใหญ่ รวมทั้งวัตถุประสงค์ของงานวิจัยคือต้องการสร้างตัวแบบที่มีประสิทธิภาพในการจำแนกข้อมูลเข้ากลุ่มได้ถูกต้องที่สุด ดังนั้นข้อสมมติที่ว่าข้อมูลต้องมีการแจกแจงแบบปกติพหุจึงไม่ใช่ข้อสมมติที่จำเป็น
- ข้อมูลมีเมทริกซ์ความแปรปรวน ความแปรปรวนร่วมเท่ากัน
เมื่อทำการตรวจสอบข้อสมมติข้อนี้โดยใช้สถิติ Box's M ผลแสดงดังตารางที่ 17

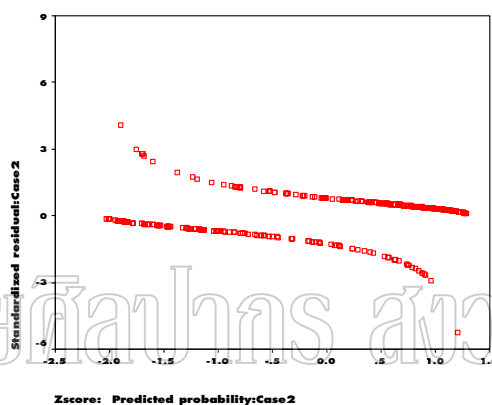
ตารางที่ 19 ค่าสถิติ Box's M ของการจำแนกกรณีที่ 2

Box's M	F Approx.	df1	df2	Sig.
107.506	2.894	36	196087.5	0.000

พิจารณาค่า P-Value จากตารางเท่ากับ 0.000 นั่นคือ สำหรับการจำแนกข้อมูลกรณี ที่ 2 มีเมตริกซ์ความแปรปรวนร่วมความแปรปรวนร่วม ไม่เท่ากัน ซึ่งไม่สอดคล้องข้อสมมติของ เทคนิคการวิเคราะห์จำแนกแบบเชิงเส้น จึงพิจารณาเลือกการใช้เมตริกซ์ความแปรปรวนร่วม แบบแยกกลุ่ม ในการสร้างตัวแบบการวิเคราะห์การจำแนก ซึ่งเหมาะสมกับข้อมูลมากกว่าการใช้ เมตริกซ์ความแปรปรวนร่วมแบบรวม (Barbara and Linda 2001 :461)

การตรวจสอบข้อสมมติของเทคนิคการวิเคราะห์การถดถอยโลจิสติก

- พล็อตของส่วนเหลือเทียบกับค่าประมาณความน่าจะเป็นมีการกระจายอย่างเป็นอิสระรอบค่า 0

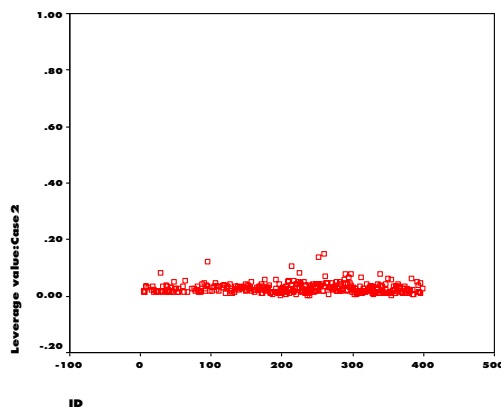


รูปที่ 8 การกระจายของส่วนเหลือ(standard residual) กับค่าพยากรณ์ความน่าจะเป็นของกรณีที่ 2

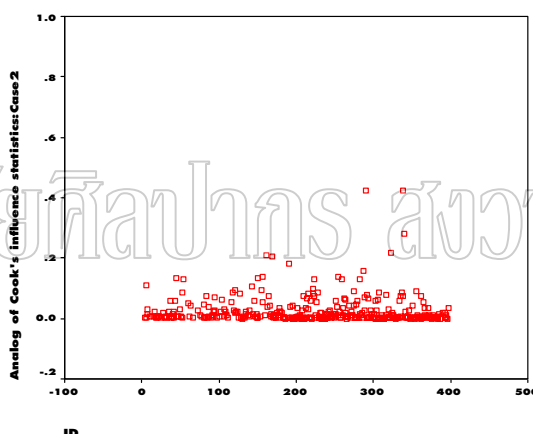
จะเห็นว่าค่าส่วนเหลือมาตรฐานของค่าสังเกตส่วนใหญ่จะมีค่าอยู่รอบๆค่าศูนย์ มีค่า ไม่เกิน 3 หรือไม่น้อยกว่า -3 มีเพียง 3 ค่าสังเกตที่ผิดปกติคือ น้อยกว่า -3 มี 2 ค่าสังเกตได้แก่ CHUP032 และ CHUP052 โดยมีค่าส่วนเหลือมาตรฐานเท่ากับ -4.05831 และ -3.01166 ตามลำดับ ส่วนค่าสังเกตที่มีค่าส่วนเหลือมากกว่า 3 มี 1 ค่าสังเกตได้แก่ LL086 โดยมีค่าส่วน เหลือเท่ากับ 5.25080 และเมื่อพิจารณาจากรูปแบบการกระจายของส่วนเหลือพบว่า การกระจาย ตัวยังไม่เป็นอิสระเท่าที่ควร

แต่เมื่อทำการตรวจสอบข้อมูลแล้วพบว่าค่าสังเกตที่มีส่วนเหลือมากกว่า 3 หรือน้อย กว่า -3 มีเพียง 3 ค่าสังเกตจากทั้งหมด 315 ค่าสังเกตซึ่งถือว่าเป็นค่าสังเกตส่วนน้อย รวมทั้ง ข้อมูลดังกล่าวเป็นข้อมูลที่ถูกต้องแล้ว หากตัดออกอาจทำให้สูญเสียสารสนเทศบางอย่างเกี่ยวกับ ข้อมูลชุดนี้ จึงเก็บค่าสังเกตเหล่านี้ไว้วิเคราะห์ร่วมด้วย

- ตรวจสอบค่าผิดปกติ (outlier observation)



รูปที่ 9 การกระจายของค่า Leverage กับหมายเลขประจำของค่าสังเกต กรณีที่ 2



รูปที่ 10 การกระจายของค่า Cook's distance กับหมายเลขประจำของค่าสังเกต กรณีที่ 2

กรณีที่ 2 เป็นการสร้างตัวแบบที่ใช้จำแนกกลุ่มคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส กับกลุ่มที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis ในงานวิจัยนี้จะทำการสร้างตัวแบบจำแนกโดยใช้ 2 เทคนิคคือ การวิเคราะห์การจำแนกและการวิเคราะห์การถดถอยโลจิสติก ผลการวิจัยนำเสนอเป็นลำดับดังนี้

ขั้นตอนแรกจะทำการเลือกตัวแปรอิสระจากกลุ่มตัวแปรที่สนใจศึกษาทั้งหมด 27 ตัวแปรนำมาตรวจสอบความแตกต่างกันของข้อมูลในกลุ่มที่จะจำแนกผลแสดงดังตารางที่ 34 พบว่าตัวแปรที่สมควรนำเข้ามาสำหรับการจำแนกข้อมูลเป็น 2 กลุ่มคือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis มี 15 ตัวแปร

ผลการวิเคราะห์ข้อมูลเบื้องต้น

ตารางที่ 20 ข้อมูลพื้นฐานของกลุ่มที่ศึกษากรณีที่ 2

ตัวแปร	ผู้ป่วยที่วินิจฉัยเป็น โรคเลปโตสไปโรสิส (N=240)	ผู้ป่วยที่วินิจฉัย เป็นRickettsiosis (N=160)	P-Value
ประวัติ			
X ₁ (อายุ) ; Median(Range)	35 (15-77)	44 (15-78)	0.000
X ₂ (เพศ) ; ชาย/หญิง	202/38	114/46	0.002
X ₃ (ประวัติเสี่ยง) ; มีประวัติเสี่ยง (%)	234 (97.5%)	151 (94.4%)	0.107
อาการทางคลินิกจากประวัติ ; n(%)			
X ₄ (ปวดศีรษะ)	222 (92.5%)	140 (87.5%)	0.095
X ₅ (อ่อนเพลีย)	233 (97.1%)	155 (96.9%)	0.905
X ₆ (เบื่ออาหาร)	194 (80.8%)	110 (68.8%)	0.006
X ₇ (เจ็บคอ)	140 (58.3%)	60 (37.5%)	0.000
X ₈ (ไอ)	128 (53.3%)	81 (50.6%)	0.595
X ₉ (เจ็บหน้าอก)	100 (41.7%)	40 (25.0%)	0.001
X ₁₀ (เลือดออกผิดปกติ)	27 (11.3%)	4 (2.5%)	0.001
X ₁₁ (ความรู้สึกตัวผิดปกติ)	57 (23.8%)	52 (32.5%)	0.054
การตรวจร่างกายแรกเริ่ม ; n(%)			
X ₁₂ (อุณหภูมิร่างกาย); Median(Range)	38 (31.8-41.2)	38.5 (36-41)	0.124
X ₁₃ (ความดันโลหิตต่ำ)	25 (10.4%)	4 (2.5%)	0.003
X ₁₄ (อักเสบที่ผนังคอหอย)	52 (21.7%)	23 (14.4%)	0.067
X ₁₅ (เยื่อตาอักเสบ)	185 (77.1%)	80 (50.0%)	0.000
X ₁₆ (เลือดออกใต้ตาขาว)	32 (13.3%)	10 (6.3%)	0.024
X ₁₇ (กดเจ็บกล้ามเนื้อ)	101 (42.1%)	30 (18.8%)	0.000
X ₁₈ (ตับโต)	31 (12.9%)	29 (18.1%)	0.153
X ₁₉ (ต่อมน้ำเหลืองโต)	76 (31.7%)	51 (31.9%)	0.965
X ₂₀ (มีจุดเลือดออกใต้ผิวหนัง)	6 (2.5%)	2 (1.3%)	0.382

ตารางที่ 20 (ต่อ)

ตัวแปร	ผู้ป่วยที่วินิจฉัยเป็น โรคเลปโตสไปโรสิส (N=240)	ผู้ป่วยที่วินิจฉัย เป็นRickettsiosis (N=160)	P-Value
การตรวจทางห้องปฏิบัติการแรกเก็บ ; n(%)			
CBC			
X ₂₁ (WBC \geq 12,000/cumm.)	72 (30.6%)	41 (26.3%)	0.352
X ₂₂ (PMN>75%)	183 (78.5%)	86 (55.1%)	0.000
X ₂₃ (Platelets < 100,000/cumm.)	74 (35.2%)	25 (19.1%)	0.001
X ₂₄ (Abnormal Urine Examination)	83 (36.7%)	44 (31.2%)	0.280
X ₂₅ (Total Bilirubin >3 mg/dl)	71 (30.6%)	27 (17.9%)	0.005
X ₂₆ (Abnormal SGOT / SGPT / Alkaline Phosphatase)	49 (21.4%)	76 (50.3%)	0.000
X ₂₇ (Abnormal BUN/Cr)	74 (31.0%)	30 (18.9%)	0.007

ตารางที่ 21 ค่าเฉลี่ยของผลตรวจทางห้องปฏิบัติการสำหรับกรณีที่ 2

ตัวแปร	ผู้ป่วยที่วินิจฉัยเป็น โรคเลปโตสไปโรสิส (N=240)	ผู้ป่วยที่วินิจฉัย เป็นRickettsiosis (N=160)	P-Value
CBC ; Median(Range)			
WBC (/cu.mm)	9.3 (1.6-72.4)	9.1 (2-38.2)	0.406
%PMN	84.9 (26-98)	77 (36-97)	0.000
Platelets (x10 ³ /cu.mm)	147 (9-515)	160 (12-417)	0.059
Urine Examination ; n(%)			
Protein >1+	35 (18.9%)	26 (20.5%)	0.734
Abnormal Sediment	73 (32.3%)	27 (19.4%)	0.007
Liver Function Test ; Median(Range)			
Total Bilirubin (mg/dl)	1.5 (0.1-37.8)	0.9 (0.1-47.7)	0.002
SGOT (U/L)	53 (6-519)	91 (10-507)	0.000

ตารางที่ 21 (ต่อ)

ตัวแปร	ผู้ป่วยที่วินิจฉัยเป็น โรคเลปโตสไปโรสิส (N=240)	ผู้ป่วยที่วินิจฉัยเป็น Rickettsiosis (N=160)	P-Value
SGPT (U/L)	43 (6-259)	65 (8-592)	0.000
Alkaline Phosphatase (U/L)	139 (13-673)	214 (12-1690)	0.000
Albumin (g/dl)	3.2 (1.6-4.7)	3.1 (1.9-4.3)	1.000
Renal Function Test			
BUN (mg/dl)	20 (6-155)	16 (4.8-129.6)	0.004
Creatinine (mg/dl)	1.4 (0.1-19.5)	1.1 (0.2-9.9)	0.000
Serum Sodium (mEq/L)	137.8 (3.7-151.5)	134.8 (112.2-155.6)	0.000
Serum Potassium (mEq/L)	3.8 (2.1-7.7)	3.8 (2.7-6.3)	0.237
Serum Bicarbonate (mEq/L)	20.3 (3.2-27.9)	21 (9.1-28.3)	0.091

มหาวิทยาลัยวลัยลักษณ์ สงขลา

1.2.1 การสร้างตัวแบบโดยใช้เทคนิคการวิเคราะห์การจำแนกสำหรับการจำแนกกรณีที่ 2

ทำการสร้างสมการจำแนกโดยใช้โปรแกรม SPSS for Window Version 11.5 จำนวนค่าสังเกตที่ใช้ในงานวิจัย 315 ค่าสังเกต ตัวแปรที่นำมาพิจารณาสำหรับกรณีที่ 1 นี้มี 15 ตัวแปร คือ X_1 (อายุ), X_2 (เพศ), X_6 (เบื่ออาหาร), X_7 (เจ็บคอ), X_9 (เจ็บหน้าอก), X_{10} (มีเลือดออกผิดปกติ), X_{13} (ความดันโลหิตต่ำ), X_{15} (เยื่อตาอักเสบ), X_{16} (เลือดออกใต้ตาขาว) X_{17} (กดเจ็บกล้ามเนื้ออ่อน), X_{22} (PMN > 75%), X_{23} (Platelets < 100,000/cumm.), X_{25} (Total Bilirubin > 3 mg/dl), X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase), X_{27} (Abnormal BUN/Cr) เลือกตัวแปรเข้าโดยวิธี Stepwise พบว่ามีตัวแปรที่เข้าสมการทั้งหมด 8 ตัว ผลการวิเคราะห์แสดงเป็นลำดับดังนี้

ตารางที่ 22 ผลการคัดเลือกตัวแปรเข้าสมการโดยวิธี Stepwise สำหรับกรณีที่ 2

Step	Entered	Wilk's Lambda		
		Statistic	Exact F	
			Statistic	Sig
1	X_{26}	0.888	39.542	0.000
2	X_{15}	0.808	37.087	0.000

ตารางที่ 22 (ต่อ)

Step	Entered	Wilk's Lambda		
		Statistic	Exact F	
			Statistic	Sig
3	X ₂₅	0.764	32.080	0.000
4	X ₁₇	0.726	29.298	0.000
5	X ₂₂	0.697	26.893	0.000
6	X ₁	0.677	24.526	0.000
7	X ₁₃	0.662	22.362	0.000
8	X ₉	0.623	20.286	0.000

จากตารางที่ 22 พบว่าผลการเลือกตัวแปรเข้าสมการวิเคราะห์จำแนกโดยใช้วิธี Stepwise มีตัวแปรเข้าสมการทั้งหมด 8 ตัว ได้แก่ X₂₆ (Abnormal SGOT/SGPT/Alkaline Phosphatase), X₁₅ (เยื่อตาอักเสบ), X₂₅ (Total Bilirubin >3 mg/dl), X₁₇ (กดเจ็บกล้ามเนื้อท้อง), X₂₂ (PMN>75%), X₁ (อายุ), X₁₃ (ความดันโลหิตต่ำ) และ X₉ (เจ็บหน้าอก) ตามลำดับ

ตารางที่ 23 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนก ด้วยค่าไอเกนสำหรับกรณีที่ 2

Eigenvalue	Percent of Varince	Cumulative percent	Canonical correlation
0.530	100.0	100.0	0.589

ตารางที่ 24 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนก ด้วยค่า Wilks' Lambda สำหรับกรณีที่ 2

Wilks' Lambda	Chi-square	df	Sig.
0.653	131.477	8	0.000

เมื่อพิจารณาค่าไอเกน จากตารางที่ 23 เท่ากับ 0.530 นั่นคือสมการวิเคราะห์จำแนกที่สร้างขึ้นมีประสิทธิภาพในการจำแนกกลุ่มข้อมูลที่ใช้ในเกณฑ์ที่ไม่ดีนักเนื่องจากค่าไอเกน ยิ่งมาก

จะแสดงถึงประสิทธิภาพในการจำแนกที่สูงตามไปด้วย และค่าสหสัมพันธ์แคนนอนิคัล เท่ากับ 0.589 สรุปได้ว่า หรือฟังก์ชันที่ได้จากการวิเคราะห์การจำแนกสามารถอธิบายความผันแปรของกลุ่มที่จำแนกได้ 34.69%

จากตารางที่ 24 เมื่อพิจารณาค่า Wilks' Lambda ซึ่งเท่ากับ 0.653 และมี P-Value เท่ากับ 0.000 นั่นคือ สมการวิเคราะห์จำแนกที่สร้างได้นี้สามารถจำแนกกลุ่มได้อย่างมีนัยสำคัญ ที่ระดับนัยสำคัญ 0.05

ตารางที่ 25 ค่าศูนย์กลางของกลุ่มของข้อมูลทั้ง 2 กลุ่มสำหรับกรณีที่ 2

Group	Function at Group Centroids
Leptospirosis	0.577
Rickettsiosis	-0.913

จากตารางที่ 23 ซึ่งแสดงค่าศูนย์กลางของกลุ่มข้อมูลทั้ง 2 กลุ่ม โดยกลุ่มที่วินิจฉัยว่าเป็นโรคเลปโตสไปโรสิสมีค่าศูนย์กลางของกลุ่มเท่ากับ 0.577 ส่วนกลุ่มที่วินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis มีค่าศูนย์กลางของกลุ่มเท่ากับ -0.913 ซึ่งเป็นค่าที่ห่างกันพอสมควร ดังนั้น การสร้างสมการวิเคราะห์จำแนกจากข้อมูลชุดนี้น่าจะมีประสิทธิภาพค่อนข้างดี

ตารางที่ 26 ค่าสัมประสิทธิ์คานอนิคัล, ค่าสัมประสิทธิ์คานอนิคัลมาตรฐาน, สหสัมพันธ์ กับฟังก์ชันการจำแนกของตัวแปรอิสระและสัมประสิทธิ์ฟังก์ชันจำแนก กรณีที่ 2

Variables	Unstandardized Coefficients	Standardized Coefficients	Correlation With Discriminant Function	Discriminant Function Coefficient	
				Leptospirosis	Rickettsiosis
X ₁	-0.019	-0.280	-0.333	0.207	0.236
X ₉	0.430	0.204	0.276	1.902	1.260
X ₁₃	1.033	0.261	0.196	1.462	-0.077
X ₁₅	0.946	0.427	0.448	4.493	3.083
X ₁₇	0.708	0.325	0.432	2.519	1.464
X ₂₂	0.692	0.313	0.330	3.673	2.643

ตารางที่ 26 (ต่อ)

Variables	Unstandardized Coefficients	Standardized Coefficients	Correlation With Discriminant Function	Discriminant Function Coefficient	
				Leptospirosis	Rickettsiosis
X ₂₅	0.919	0.405	0.181	0.437	-0.932
X ₂₆	-1.340	-0.594	-0.488	0.527	2.524
Constant	-0.613			-9.009	-8.346

เมื่อพิจารณาค่าสัมประสิทธิ์จากตารางที่ 26 พบว่าตัวแปร X₂₆ (Abnormal SGOT/SGPT/Alkaline Phosphatase) เป็นตัวแปรที่มีความสำคัญต่อประสิทธิภาพในการจำแนกของสมการจำแนกกลุ่มมากกว่าตัวแปรอื่นๆ เนื่องจากมีค่าสัมประสิทธิ์ที่สูงกว่า

พิจารณาค่าสหสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 8 ตัวและฟังก์ชันการวิเคราะห์จำแนก พบว่าตัวแปรที่นำมาใช้ในการสร้างสมการวิเคราะห์จำแนก ตัวแปรที่มีค่าสหสัมพันธ์สูงที่สุดเมื่อพิจารณาจากค่าสัมบูรณ์ คือตัวแปร X₂₆ (Abnormal SGOT/SGPT/Alkaline Phosphatase) จากค่าสัมประสิทธิ์ฟังก์ชันจำแนก จะได้สมการการวิเคราะห์การจำแนกดังนี้

$$Z_1 = -9.009 + (0.207 X_1) + (1.902 X_9) + (1.462 X_{13}) + (4.493 X_{15}) + (2.519 X_{17}) + (3.673 X_{22}) + (0.437 X_{25}) + (0.527 X_{26})$$

$$Z_2 = -8.346 + (0.236 X_1) + (1.260 X_9) - (0.077 X_{13}) + (3.083 X_{15}) + (1.464 X_{17}) + (2.643 X_{22}) - (0.932 X_{25}) + (2.524 X_{26})$$

ตารางที่ 27 ผลการจัดข้อมูลเข้ากลุ่ม เมื่อใช้ฟังก์ชันการวิเคราะห์จำแนกสำหรับกรณีที่ 2

Original Diagnosis	Predicted Diagnosis		Percentage Correct	Overall Percentage Correct
	Leptospirosis	Rickettsiosis		
Leptospirosis	158	35	81.9	79.0
Rickettsiosis	31	91	74.6	

เมื่อทำการสร้างฟังก์ชันการวิเคราะห์จำแนกโดยกำหนดความน่าจะเป็นเบื้องต้น ของ ทั้ง 2 กลุ่มเท่ากัน คือ เท่ากับ 0.50 ทั้งสองกลุ่ม จะได้ค่าสัมประสิทธิ์ของตัวแปรที่ใช้ในการสร้าง สมการวิเคราะห์จำแนกทั้ง 8 ตัว แสดงดังตารางที่ 26 ซึ่งเมื่อนำฟังก์ชันการวิเคราะห์จำแนกที่ ได้มาทำการจัดข้อมูลเข้ากลุ่ม ได้ผลดังตารางที่ 27 แสดงให้เห็นว่าผลการจัดเข้ากลุ่มถูกต้องรวม เท่ากับ 79.0% ซึ่งมาจากการจำแนกเข้ากลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสถูกต้อง 81.9% และมาจากการจำแนกเข้ากลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis ได้ถูกต้อง 74.6%

และเมื่อนำค่าที่ได้จากตารางที่ 27 มาคำนวณหาค่าอัตราการจำแนกผิดที่เห็นชัดเจน ได้ เท่ากับ $(35+31)/315$ หรือมีค่าเท่ากับ 0.210

1.2.2 การสร้างตัวแบบการวิเคราะห์การถดถอยโลจิสติกสำหรับการจำแนกกรณีที่ 2

นำข้อมูลชุดเดิมมาวิเคราะห์โดยใช้การวิเคราะห์การถดถอยโลจิสติก นำตัวแปร 15 ตัว มาทำการเลือกตัวแปรสมการเข้าโดยใช้ Backward Stepwise (Likelihood Ratio) ผลเสนอในตารางตามลำดับต่อไปนี้

ตารางที่ 28 ผลการวิเคราะห์ข้อมูลเพื่อสร้างสมการถดถอยโลจิสติกสำหรับกรณีที่ 2

Variables	$\hat{\beta}$	S.E. ($\hat{\beta}$)	Wald	Sig.	Exp ($\hat{\beta}$)	95% CI. For Exp ($\hat{\beta}$)	
						Lower	Upper
X ₁	-0.026	0.010	6.216	0.013	0.974	0.955	0.994
X ₉	0.642	0.316	4.140	0.042	1.900	1.024	3.527
X ₁₃	1.495	0.687	4.731	0.030	4.460	1.159	17.159
X ₁₅	1.257	0.312	16.255	0.000	3.516	1.908	6.479
X ₁₇	0.982	0.317	9.563	0.002	2.669	1.433	4.974
X ₂₂	0.881	0.311	8.011	0.005	2.414	1.311	4.442
X ₂₅	1.371	0.380	13.024	0.000	3.938	1.871	8.291
X ₂₆	-1.815	0.329	30.330	0.000	0.163	0.085	0.311
Constant	-0.209	0.581	0.130	0.718	0.811		

จากตารางที่ 28 ตัวสถิติ Wald ใช้ในการทดสอบว่าค่าประมาณสัมประสิทธิ์ของตัวแปร นั้นๆมีค่าเท่ากับศูนย์หรือไม่ พิจารณาจากค่า P-Value พบว่ามีตัวแปรที่มีค่าประมาณไม่เท่ากับ

คุณ้อย่างมีนัยสำคัญ มี 8 ตัว คือ X_1 (อายุ), X_9 (เจ็บหน้าอก), X_{13} (ความดันโลหิตต่ำ), X_{15} (เยื่อตาอักเสบ), X_{17} (กรดเจ็บบก้ามเนื้ออ่อน), X_{22} (PMN>75%), X_{25} (Total Bilirubin >3 mg/dl) และ X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase) ซึ่งสามารถเขียนลอจิสติกตัวอย่าง สำหรับตัวแบบถดถอยโลจิสติก ได้ดังนี้

$$\hat{g}(x) = -0.209 - (0.026 * X_1) + (0.642 * X_9) + (1.495 * X_{13}) + (1.257 * X_{15}) \\ + (0.982 * X_{17}) + (0.881 * X_{22}) + (1.371 * X_{25}) - (1.815 * X_{26})$$

การตีความจากค่าสัมประสิทธิ์ที่ได้ ทำได้โดยการพิจารณาจากค่า $Exp(\beta)$ ของแต่ละตัวแปร เช่น ตัวแปร X_{17} (กรดเจ็บบก้ามเนื้ออ่อน) จากตารางมีค่า $Exp(\beta)$ เท่ากับ 2.669 หมายความว่าผู้ป่วยอาการกรดเจ็บบก้ามเนื้ออ่อน มีความเสี่ยงต่อการเป็นโรคเลปโตสไปโรซิสมากกว่าผู้ป่วยที่ไม่มีอาการกรดเจ็บบก้ามเนื้ออ่อน 2.669 เท่า

ตารางที่ 29 ผลการตรวจสอบภาวะสารูปดี (Goodness-of-fit Test) และ R-Square สำหรับกรณีที่ 2

Method	Value	df	P-Value
Statistic G^2	126.163	8	0.000
Hosmer and Lemeshow	13.303	8	0.102
Nagelkerke R Square	0.448		

ตารางที่ 29 เป็นการตรวจสอบภาวะสารูปดีของตัวแบบโลจิสติกจากค่า Statistic G^2 ที่ได้จากการทดสอบ Likelihood-Ratio Test ซึ่งเป็นการเปรียบเทียบตัวแบบ 2 ตัว ซึ่งมีจำนวนสัมประสิทธิ์ที่แตกต่างกันในที่นี้คือตัวแบบที่มีเฉพาะค่าคงที่เทียบกับตัวแบบที่มีตัวแปรอิสระ 8 ตัว เพื่อที่จะทดสอบว่าค่าสัมประสิทธิ์ของตัวแปรทั้ง 8 ตัวนี้มีค่าเท่ากับ 0 หรือไม่ จะเห็นได้ว่าค่า Statistic G^2 มีค่า 126.163 และ P-Value เท่ากับ 0.000 แสดงให้เห็นว่ามีค่าสัมประสิทธิ์บางตัวในจำนวนตัวแปร 8 ตัวนั้นที่มีค่าไม่เท่ากับ 0 อย่างมีนัยสำคัญทางสถิติ (ที่ระดับนัยสำคัญ 0.05) แสดงว่ามีตัวแปรอิสระอย่างน้อยหนึ่งตัวที่มีอิทธิพลต่อตัวแปรตาม ส่วนการทดสอบ Hosmer and Lemeshow เป็นการทดสอบความเหมาะสมของตัวแบบ จากตารางที่ 29 พบว่าค่าสถิติ Hosmer and Lemeshow มีค่าเท่ากับ 13.303 และมีค่า P-value เท่ากับ 0.102 นั่นคือ ตัวแบบที่สร้างได้มีความเหมาะสม กับข้อมูลตัวอย่าง

เมื่อพิจารณาค่า R-Square จากตารางที่ 29 ได้ว่า ฟังก์ชันที่ได้จากการวิเคราะห์การถดถอยโลจิสติกสามารถอธิบายความผันแปรของกลุ่มที่จำแนกได้ 44.80%

ตารางที่ 30 จำนวนข้อมูลที่ถูกจัดเข้ากลุ่มถูกต้อง เมื่อใช้จุดตัดการจำแนกที่ระดับต่างๆกรณีที่ 2

Cut Point	Observed	Predicted		Percentage Correct	Overall Percentage Correct	APER
		Leptospirosis	Rickettsiosis			
0.2	Leptospirosis	187	6	96.9	71.7	0.283
	Rickettsiosis	83	39	32.0		
0.3	Leptospirosis	184	9	95.3	75.2	0.248
	Rickettsiosis	69	53	43.4		
0.4	Leptospirosis	175	18	90.7	77.8	0.222
	Rickettsiosis	52	70	57.4		
0.5	Leptospirosis	167	26	86.5	79.0	0.210
	Rickettsiosis	40	82	67.2		
0.6	Leptospirosis	156	37	80.8	78.7	0.213
	Rickettsiosis	30	92	75.4		
0.7	Leptospirosis	137	56	71.0	75.6	0.244
	Rickettsiosis	21	101	82.8		
0.8	Leptospirosis	103	90	53.4	67.6	0.324
	Rickettsiosis	12	110	90.2		

ตารางที่ 30 แสดงถึงผลการจัดข้อมูลเข้ากลุ่ม โดยใช้การสร้างสมการถดถอยโลจิสติกพิจารณาจากการจำแนกเข้ากลุ่มได้ถูกต้องรวมพบว่า เมื่อเลือกจุดตัดการจำแนกเท่ากับ 0.5 จะได้สมการถดถอยโลจิสติกที่จำแนกข้อมูลนี้ได้ถูกต้องมากที่สุด คือเท่ากับ 79.0%

เมื่อพิจารณาแยกเป็นกรณีผู้ป่วยที่เป็นโรคเลปโตสไปโรสิส ใช้สมการถดถอยโลจิสติกจะจำแนกข้อมูลเข้ากลุ่มผู้ป่วยเลปโตสไปโรสิสถูกต้องมากที่สุดเมื่อเลือกจุดตัดการจำแนก เท่ากับ 0.2 และจำแนกถูกลดลงเรื่อยๆเมื่อเลือกจุดตัดการจำแนกที่มากขึ้น ส่วนผู้ป่วยที่วินิจฉัยว่าเป็นโรคอื่นๆ เมื่อใช้สมการถดถอยโลจิสติกจะจำแนกข้อมูลเข้ากลุ่มผู้ป่วยโรคอื่นๆได้ถูกต้องน้อยที่สุด

เมื่อเลือกจุดตัดการจำแนกเท่ากับ 0.2 และอัตราการจำแนกจะมากขึ้นเรื่อยๆเมื่อเลือกจุดตัดการจำแนกที่มากขึ้น สำหรับค่า APER พบว่ามีค่าต่ำที่สุดเมื่อกำหนดจุดตัดการจำแนก เท่ากับ 0.5 และที่จุดตัดการจำแนกจุดอื่นๆ APER มีค่าที่สูงขึ้น นั่นคือสมการถดถอยโลจิสติกที่สร้างขึ้นจัดข้อมูลเข้ากลุ่มได้ถูกต้องรวมน้อยลงไปด้วย

สำหรับการนำตัวแบบที่สร้างได้ไปเปรียบเทียบผลการจัดเข้ากลุ่มถูกต้อง โดยเทียบกับผลที่ได้จากตัวแบบการวิเคราะห์การจำแนกและตัวแบบที่อิงเกณฑ์ WHO จะเลือกใช้จุดตัดที่ให้ผลการจัดเข้ากลุ่มถูกต้องกรณีผู้ป่วยที่วินิจฉัยเป็นโรคเลปโตสไปโรสิส และกรณีผู้ป่วยวินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis มีค่าสูงทั้งสองกรณี ดังนั้นจึงเลือกใช้จุด 0.5 ซึ่งมีเปอร์เซ็นต์การจำแนกถูกรวม, กรณีผู้ป่วยที่วินิจฉัยเป็นโรคเลปโตสไปโรสิส และกรณีผู้ป่วยวินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis เท่ากับ 86.5%, 67.2% และ 79.0% ตามลำดับ

เมื่อนำผลการจำแนกข้อมูลเข้ากลุ่มได้ถูกต้อง ทั้งกรณีผู้ป่วยเลปโตสไปโรสิสและผู้ป่วยอื่นๆ รวมทั้งผลการจำแนกรวม และค่า APER มาเปรียบเทียบกัน ระหว่างการใช้วิธีการวิเคราะห์จำแนกและวิธีวิเคราะห์การถดถอยโลจิสติก ผลแสดงดังตารางต่อไปนี้

ตารางที่ 31 ผลการจำแนกกรณีที่ 2 โดยใช้วิธีการวิเคราะห์จำแนกและการถดถอยโลจิสติก

Technique	Observed	Predicted		Percentage correct	Overall percentage correct	APER
		Leptospirosis	Rickettsiosis			
Discriminant	Leptospirosis	158	35	81.9	79.0	0.210
	Rickettsiosis	31	91	74.6		
Logistic Regression	Leptospirosis	167	26	86.5	79.0	0.210
	Rickettsiosis	40	82	67.2		

จากผลการจำแนกข้อมูลเข้ากลุ่มถูกต้องในตารางที่ 31 พบว่า สำหรับการจำแนกเป็น 2 กลุ่ม ของข้อมูลในงานวิจัยนี้ หากพิจารณาจากผลการจัดเข้ากลุ่มถูกต้องรวมพบว่าทั้งสองเทคนิคมีประสิทธิภาพเท่ากัน แต่เมื่อแยกพิจารณาเป็นกรณี กรณีที่คนไข้ที่เป็นโรคเลปโตสไปโรสิสพบว่า ตัวแบบที่ได้จากการวิเคราะห์การถดถอยโลจิสติกมีประสิทธิภาพสูงกว่า

1.3 ผลการวิเคราะห์สำหรับการจำแนกกรณีที่ 3

การตรวจสอบข้อสมมติของเทคนิคการวิเคราะห์การจำแนก

- ข้อมูลมีการแจกแจงแบบปกติพหุ
ตัวแปรอิสระที่นำมาศึกษามีลักษณะเชิงผสม คือ มีทั้งตัวแปรแบบต่อเนื่องและตัวแปรเชิงกลุ่ม ซึ่งเป็นเรื่องยากที่จะมีการแจกแจงแบบปกติพหุ แต่เนื่องจากจำนวนค่าสังเกตที่ใช้ในงานวิจัยนี้มีขนาดใหญ่ รวมทั้งวัตถุประสงค์ของงานวิจัยคือต้องการสร้างตัวแบบที่มีประสิทธิภาพในการจำแนกข้อมูลเข้ากลุ่มได้ถูกต้องที่สุด ดังนั้นข้อสมมติที่ว่าข้อมูลต้องมีการแจกแจงแบบปกติพหุจึงไม่ใช่ข้อสมมติที่จำเป็น (Barbara and Linda 2001 :462-492)
- ข้อมูลมีเมทริกซ์ความแปรปรวน ความแปรปรวนร่วมเท่ากัน
เมื่อทำการตรวจสอบข้อสมมติข้อนี้โดยใช้สถิติ Box'M ผลแสดงดังตารางที่ 30

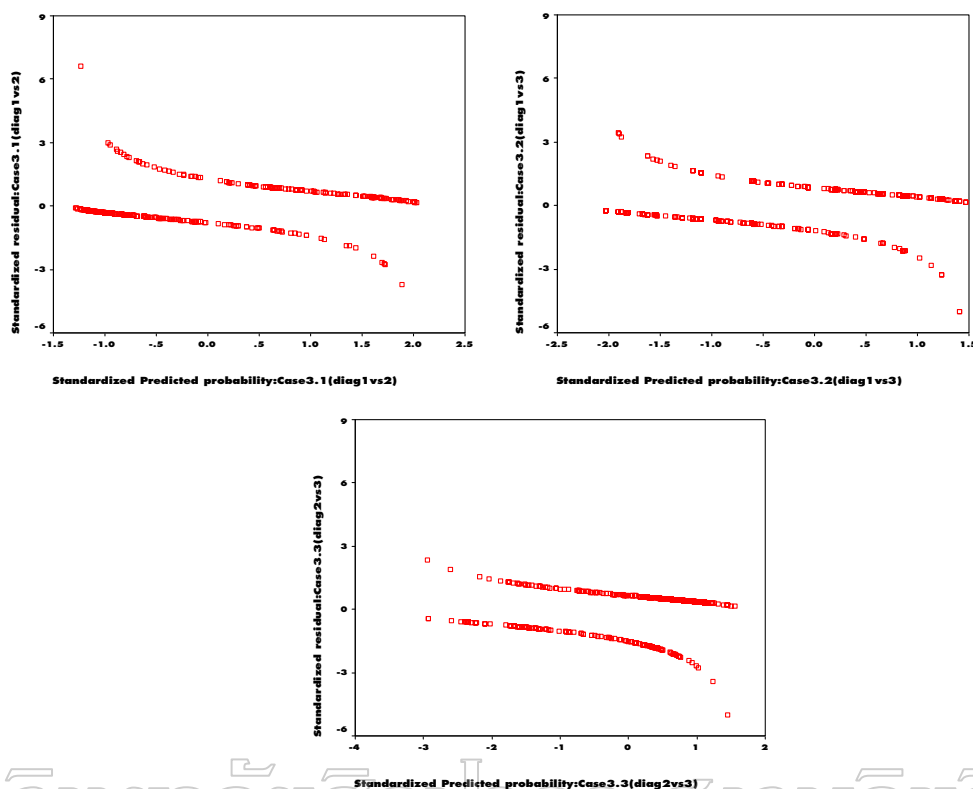
ตารางที่ 32 ค่าสถิติ Box's M ของการจำแนกกรณีที่ 3

Box's M	F Approx.	df1	df2	Sig.
208.092	2.252	90	416323.3	0.000

พิจารณาค่า P-Value จากตารางเท่ากับ 0.000 นั่นคือ สำหรับการจำแนกข้อมูลกรณีที่ 3 มีเมทริกซ์ความแปรปรวนความแปรปรวนร่วม ไม่เท่ากัน ซึ่งไม่สอดคล้องกับข้อสมมติของเทคนิคการวิเคราะห์จำแนกแบบเชิงเส้น จึงพิจารณาเลือกการใช้เมทริกซ์ความแปรปรวนร่วมแบบแยกกลุ่ม ในการสร้างตัวแบบการวิเคราะห์การจำแนกซึ่งเหมาะสมกับข้อมูลมากกว่าการใช้เมทริกซ์ความแปรปรวนร่วมแบบรวม (Barbara and Linda 2001 :461)

การตรวจสอบข้อสมมติของเทคนิคการวิเคราะห์การถดถอยโลจิสติก

- พลอตของส่วนเหลือเทียบกับค่าประมาณความน่าจะเป็นมีการกระจายอย่างเป็นอิสระรอบค่า 0



รูปที่ 11 การกระจายของส่วนเหลือ (standard residual) กับค่าพยากรณ์ความน่าจะเป็นของกรณีที่ 3

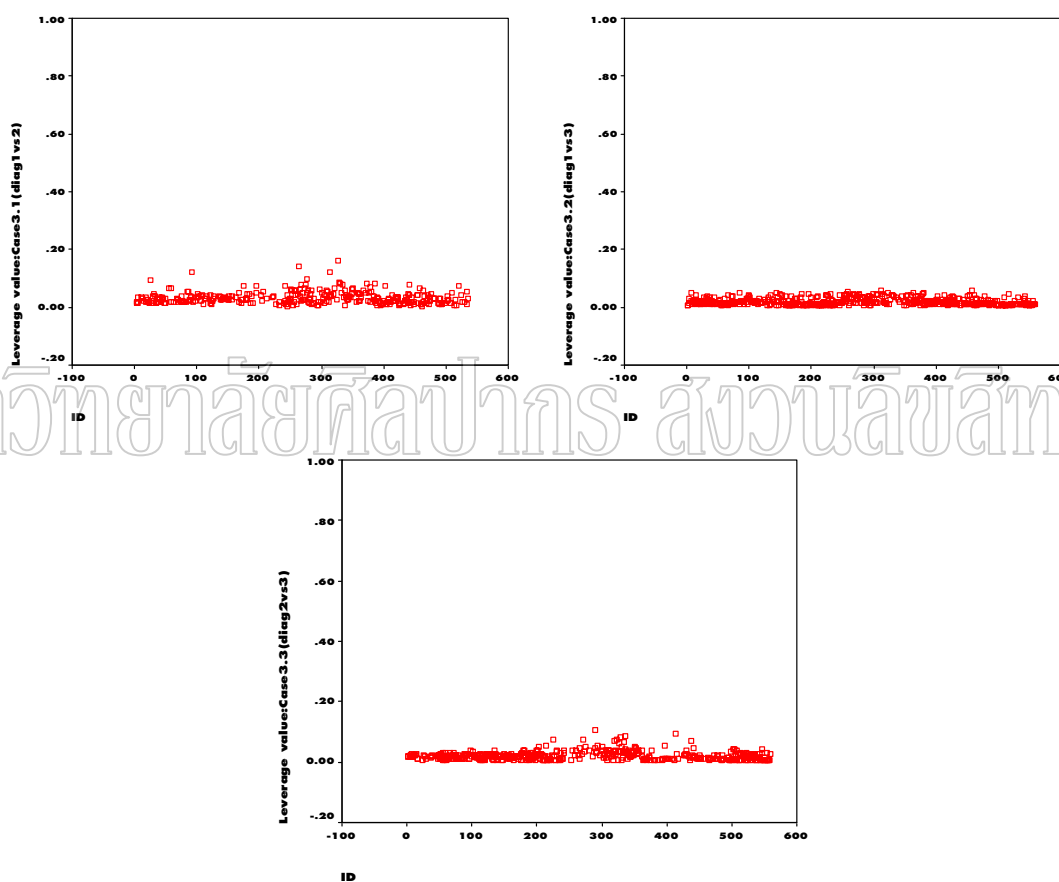
รูปที่ 11 เป็นการพลอตการกระจายของส่วนเหลือแต่ละคู่ของตัวแปรตาม เพื่อตรวจสอบว่าตัวแบบที่นำมาใช้กับข้อมูลตัวอย่างสอดคล้องกับข้อตกลงเบื้องต้นหรือไม่ จะเห็นได้ว่าค่าส่วนเหลือมาตรฐานของค่าสังเกตส่วนใหญ่จะมีค่าอยู่รอบๆค่าศูนย์

สำหรับรูปแรกเป็นการพลอตการกระจายของส่วนเหลือระหว่างกลุ่มแรกกับกลุ่มที่สอง หรือคือกลุ่มคนไข้ที่เป็นโรคเลปโตสไปโรสิสกับคนไข้ที่เป็นโรคในกลุ่ม Rickettsiosis พบว่าส่วนใหญ่มีค่าไม่เกิน 3 หรือไม่น้อยกว่า -3 มีเพียง 2 ค่าสังเกตที่มีค่าออกช่วงดังกล่าวคือ CHUP032 และ LL086 สำหรับรูปถัดมาเป็นการพลอตการกระจายของส่วนเหลือ ระหว่างกลุ่มแรกกับกลุ่มที่สามหรือคือกลุ่มคนไข้ที่เป็นโรคเลปโตสไปโรสิสกับคนไข้ที่เป็นโรคอื่นๆ พบว่าส่วนใหญ่มีค่าไม่เกิน 3 หรือไม่น้อยกว่า -3 มี 7 ค่าสังเกตที่มีค่าอยู่นอกช่วงดังกล่าวคือ KL070, KP035, LL051, LL060, LL082, CHUP032 และ CHUP052 สำหรับรูปสุดท้ายเป็นการพลอตการกระจายของส่วนเหลือระหว่างกลุ่มที่สองกับกลุ่มที่สามหรือคือกลุ่มคนไข้ที่เป็นโรคในกลุ่ม Rickettsiosis กับคนไข้ที่เป็นโรคอื่นๆ พบว่าส่วนใหญ่มีค่าไม่เกิน 3 หรือไม่น้อยกว่า -3 มีเพียง 2 ค่าสังเกตที่มีค่าอยู่

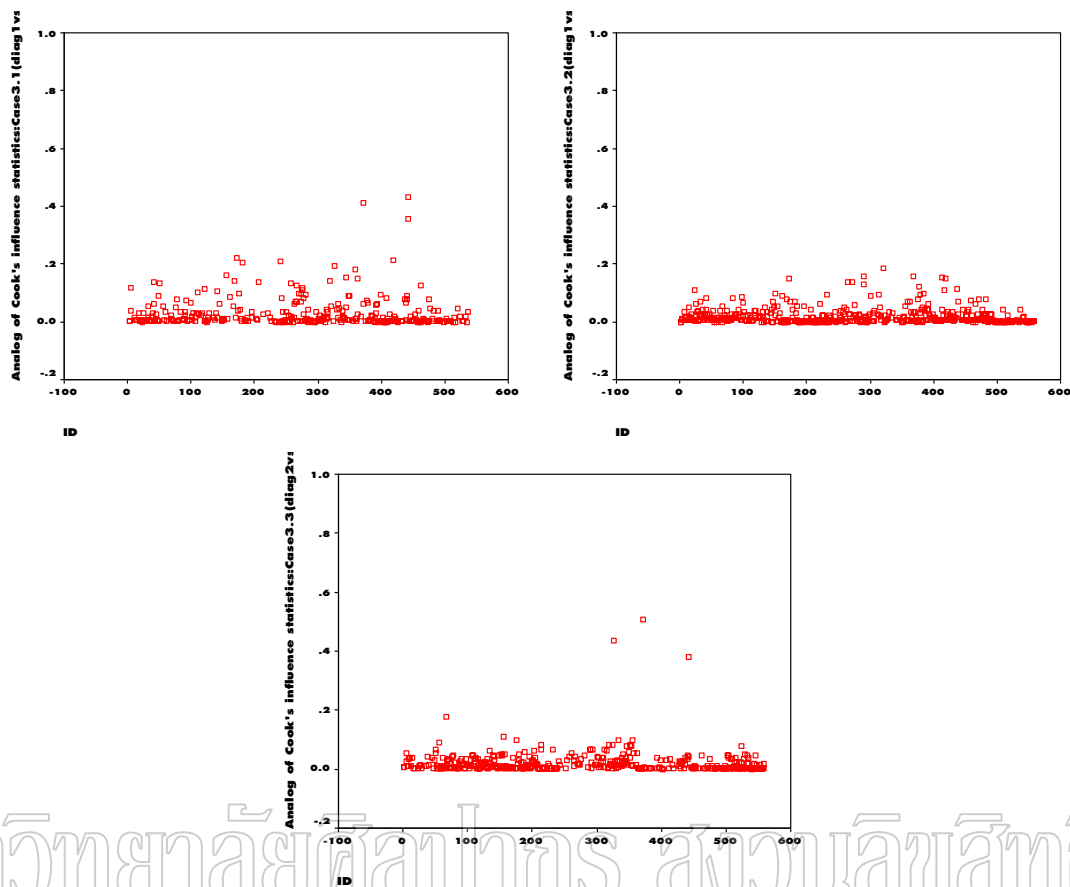
นอกช่วงดังกล่าวคือ KP045 และ LL008 เมื่อพิจารณาจากรูปแบบการกระจายของส่วนเหลือพบว่าการกระจายตัวยังไม่เป็นอิสระเท่าที่ควร

แต่เมื่อทำการตรวจสอบข้อมูลแล้วพบว่าค่าสังเกตที่มีส่วนเหลือมากกว่า 3 หรือน้อยกว่า -3 เป็นเพียงค่าสังเกตส่วนน้อย รวมทั้งข้อมูลดังกล่าวเป็นข้อมูลที่ถูกตัดออก อาจทำให้สูญเสียสารสนเทศบางอย่างเกี่ยวกับข้อมูลชุดนี้จึงเก็บค่าสังเกตเหล่านี้ไว้วิเคราะห์ร่วมด้วย

- ตรวจสอบค่าผิดปกติ (outlier observation)



รูปที่ 12 การกระจายของค่า Leverage กับหมายเลขประจำของค่าสังเกต กรณีที่ 3



รูปที่ 13 การกระจายของค่า Cook's distance กับหมายเลขประจำของค่าสังเกต กรณีที่ 3

พิจารณารูปที่ 12 และ 13 เป็นการพลอตระหว่างค่า Leverage และ ค่า Cook's distance กับหมายเลขประจำของค่าสังเกตของกรณีที่ 3 เพื่อทำการตรวจสอบหาค่าผิดปกติของข้อมูล จากการสังเกตพบว่าไม่มีค่าที่ผิดปกติ

กรณีที่ 3 เป็นการสร้างตัวแบบที่ใช้จำแนกกลุ่มคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส , กลุ่มที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ พิจารณาการสร้างตัวแบบจำแนกโดยใช้ 2 เทคนิคคือ การวิเคราะห์การจำแนกและการวิเคราะห์การถดถอยโลจิสติก ผลการวิจัยนำเสนอเป็นลำดับดังนี้

ขั้นตอนแรกจะทำการเลือกตัวแปรอิสระจากกลุ่มตัวแปรที่สนใจศึกษาทั้งหมด 27 ตัว นำมาตรวจสอบความแตกต่างกันของข้อมูลในกลุ่มที่จะจำแนกผลแสดงดังตารางที่ 31 พบว่าตัวแปรที่สมควรนำเข้ามาสมการสำหรับกรณีที่ 3 มี 19 ตัวแปร

ผลการวิเคราะห์ข้อมูลเบื้องต้น

ตารางที่ 33 ข้อมูลพื้นฐานของกลุ่มที่ศึกษา กรณีที่ 3

ตัวแปร	ผู้ป่วยที่วินิจฉัยเป็น โรคเลปโตสไปโรสิส (N=240)	ผู้ป่วยที่วินิจฉัย เป็นRickettsiosis (N=160)	ผู้ป่วยที่วินิจฉัย เป็นโรคอื่นๆ (N=355)	P-Value
ประวัติ				
X ₁ (อายุ) ; Median(Range)	35 (15-77)	44 (15-78)	35 (6-87)	0.000
X ₂ (เพศ) ; ชาย/หญิง	202/38	114/46	266/89	0.004
X ₃ (มีประวัติเสี่ยง) ; n(%)	234 (97.5%)	151 (94.4%)	307 (86.5%)	0.000
อาการทางคลินิกจากประวัติ ; n(%)				
X ₄ (ปวดศีรษะ)	222 (92.5%)	140 (87.5%)	308 (86.8%)	0.081
X ₅ (อ่อนเพลีย)	233 (97.1%)	155 (96.9%)	339 (95.5%)	0.547
X ₆ (เบื่ออาหาร)	194 (80.8%)	110 (68.8%)	266 (74.9%)	0.021
X ₇ (เจ็บคอ)	140 (58.3%)	60 (37.5%)	138 (38.9%)	0.000
X ₈ (ไอ)	128 (53.3%)	81 (50.6%)	161 (45.4%)	0.145
X ₉ (เจ็บหน้าอก)	100 (41.7%)	40 (25.0%)	73 (20.6%)	0.000
X ₁₀ (เลือดออกผิวดปกติ)	27 (11.3%)	4 (2.5%)	24 (6.8%)	0.004
X ₁₁ (ความรู้สึกตัวผิวดปกติ)	57 (23.8%)	52 (32.5%)	63 (17.7%)	0.001
การตรวจร่างกายแรกจับ ; n(%)				
X ₁₂ (อุณหภูมิ); Median(Range)	38.0 (31.8-41.2)	38.5 (36.0-41.0)	38.3 (35.5-41.3)	0.248
X ₁₃ (ความดันโลหิตต่ำ)	25 (10.4%)	4 (2.5%)	29 (8.2%)	0.013
X ₁₄ (อักเสบที่ผนังคอหอย)	52 (21.7%)	23 (14.4%)	61 (17.2%)	0.152
X ₁₅ (เยื่อตาอักเสบ)	185 (77.1%)	80 (50.0%)	148 (41.7%)	0.000
X ₁₆ (เลือดออกใต้ตาขาว)	32 (13.3%)	10 (6.3%)	12 (3.4%)	0.000
X ₁₇ (กดเจ็บกล้ามเนื้อ)	101 (42.1%)	30 (18.8%)	93 (26.2%)	0.000
X ₁₈ (ตับโต)	31 (12.9%)	29 (18.1%)	34 (9.6%)	0.024
X ₁₉ (ต่อมน้ำเหลืองโต)	76 (31.7%)	51 (31.9%)	88 (24.8%)	0.107
X ₂₀ (มีจุดเลือดออกใต้ผิวหนัง)	6 (2.5%)	2 (1.3%)	6 (1.7%)	0.630

ตารางที่ 33 (ต่อ)

ตัวแปร	ผู้ป่วยที่วินิจฉัย เป็นโรค เลปโตสไปโรสิส (N=240)	ผู้ป่วยที่วินิจฉัย เป็นโรคในกลุ่ม Rickettsiosis (N=160)	ผู้ป่วยที่ วินิจฉัยเป็น โรคอื่นๆ (N=355)	P -Value
การตรวจทางห้องปฏิบัติการแรกรับ ; n(%)				
CBC				
X ₂₁ (WBC ≥ 12,000/cumm.)	72 (30.6%)	41 (26.3%)	90 (26.1%)	0.446
X ₂₂ (PMN > 75%)	188 (78.5%)	86 (55.1%)	186 (54.9%)	0.000
X ₂₃ (Platelets < 100,000/cumm.)	74 (35.2%)	25 (19.1%)	56 (18.2%)	0.000
X ₂₄ (Abnormal Urine Examination)	83 (36.7%)	44 (31.2%)	75 (22.7%)	0.001
X ₂₅ (Total Bilirubin > 3 mg/dl)	71 (30.7%)	27 (17.9%)	41 (12.2%)	0.000
X ₂₆ (Abnormal SGOT/SGPT/Alk.)	49 (21.4%)	76 (50.3%)	84 (25.1%)	0.000
X ₂₇ (Abnormal BUN/Cr)	74 (31.1%)	30 (18.9%)	48 (13.7%)	0.000

ตารางที่ 34 ค่าเฉลี่ยของผลตรวจทางห้องปฏิบัติการสำหรับกรณี 3

ตัวแปร	ผู้ป่วยที่ วินิจฉัยเป็น โรคเลปโตสไปโรสิส (N=240)	ผู้ป่วยที่ วินิจฉัยเป็น Rickettsiosis (N=160)	ผู้ป่วยที่ วินิจฉัยเป็น โรคอื่นๆ (N=355)	P -Value
CBC ; Median(Range)				
WBC (/cu.mm)	10.9 (1.6-72.4)	11.2 (3.2-38.2)	9.2 (1.6-54.3)	0.049
%PMN	84.9 (26-98)	75 (42-93)	75 (19-96)	0.000
Platelets (x10 ³ /cu.mm)	100 (9-515)	140.5 (25-417)	190 (12-772)	0.000
Urine Examination				
Protein > 1+ (n)	35 (18.9%)	26 (20.5%)	36 (13.6%)	0.151
Abnormal Sediment (n)	73 (32.3%)	27 (19.4%)	55 (17.1%)	0.000

ตารางที่ 34 (ต่อ)

ตัวแปร	ผู้ป่วยที่วินิจฉัย เป็นโรค เลปโตสไปโรสิส (N=240)	ผู้ป่วยที่ วินิจฉัยเป็น Rickettsiosis (N=160)	ผู้ป่วยที่ วินิจฉัยเป็น โรคอื่นๆ (N=355)	P -Value
Liver Function Test ; Median(Range)				
Total Bilirubin (mg/dl)	2.1 (0.2-37.8)	1.2 (0.1-47.7)	0.8 (0.2-37.8)	0.000
SGOT (U/L)	58 (15-519)	97 (10-507)	44 (7-3075)	0.000
SGPT (U/L)	48 (6-259)	80 (8-592)	38 (6-2750)	0.000
Alkaline Phosphatase (U/L)	104 (21-629)	159 (12-1209)	89 (7-735)	0.000
Albumin (g/dl)	3.2 (1.6-4.7)	3.1 (2.2-4.3)	3.4 (1.6-5.2)	0.000
Renal Function Test				
BUN (mg/dl)	25 (7-155)	15.4 (6-129)	14 (3-164)	0.000
Creatinine (mg/dl)	1.8 (0.8-19.5)	1.3 (0.6-9.3)	1.2 (0.4-12.6)	0.000
Serum Sodium (mEq/L)	137 (103-151.5)	135 (112.2-155.6)	137 (118.8-148)	0.001
Serum Potassium (mEq/L)	3.9 (2.3-7.7)	3.8 (2.7-6.3)	3.9 (2.5-7.6)	0.226
Serum Bicarbonate (mEq/L)	20.3 (3.2-27.7)	20.9 (9.1-28.3)	22.1 (4.6-34)	0.001

1.3.1 การสร้างตัวแบบโดยใช้เทคนิคการวิเคราะห์การจำแนกสำหรับการจำแนกกรณีที่ใช้ทำการสร้างสมการจำแนกโดยมีจำนวนค่าสังเกตที่ใช้ในงานวิจัย 559 ค่าสังเกต ตัวแปรอิสระที่ใช้ 19 ตัวแปร X_1 (อายุ), X_2 (เพศ), X_3 (การมีหรือไม่มีปัจจัยเสี่ยง), X_6 (เบื่อกอาหาร), X_7 (เจ็บคอ), X_9 (เจ็บหน้าอก), X_{10} (มีเลือดออกผิดปกติ), X_{11} (ความรู้สึกตัวผิดปกติ), X_{13} (ความดันโลหิตต่ำ), X_{15} (เยื่อตาอักเสบ), X_{16} (เลือดออกใต้ตาขาว) X_{17} (กดเจ็บกล้ามเนื้ออ่อน), X_{18} (ตับโต), X_{22} (PMN > 75%), X_{23} (Platelets < 100,000/cumm.), X_{24} (Abnormal Urine Examination), X_{25} (Total Bilirubin >3 mg/dl), X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase), X_{27} (Abnormal BUN/Cr) วิธีการเลือกตัวแปรเข้าสมการการวิเคราะห์การจำแนกคือวิธี Stepwise Method ผลการวิเคราะห์แสดงเป็นลำดับดังนี้

ตารางที่ 35 ผลการคัดเลือกตัวแปรเข้าสมการการวิเคราะห์จำแนกกรณีที่ 3 โดยใช้วิธี Stepwise

Step	Entered	Wilk's Lambda		
		Statistic	Exact F	
			Statistic	Sig
1	X ₁₅	0.884	36.349	0.000
2	X ₂₆	0.828	27.479	0.000
3		0.773	25.322	0.000
4	X ₂₅	0.747	21.697	0.000
5	X ₂₂	0.724	19.304	0.000
6	X ₁₇	0.707	17.406	0.000
7	X ₂₃	0.691	15.955	0.000
8	X ₁	0.681	14.543	0.000
9	X ₁₈	0.671	13.467	0.000
	X ₉			

จากตารางที่ 35 พบว่าผลการเลือกตัวแปรเข้าสมการการวิเคราะห์จำแนกโดยวิธี Stepwise มีตัวแปรเข้าทั้งหมด 9 ตัว ได้แก่ X₁₅ (เยื่อตาอักเสบ), X₂₆ (Abnormal SGOT / SGPT / Alkaline Phosphatase), X₂₅ (Total Bilirubin >3 mg/dl), X₂₂ (PMN>75%), X₁₇ (กดเจ็บกล้ามเนื้ออ่อน), X₂₃ (Platelets < 100,000/cumm.), X₁ (อายุ), X₁₈ (ตับโต) และ X₉ (เจ็บหน้าอก) ตามลำดับ

ตารางที่ 36 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการการวิเคราะห์จำแนกกรณีที่ 3 ด้วยค่าไอเกน

Function	Eigenvalue	Percent of Varince	Cumulative Percent	Canonical Correlation
1	0.356	78.2	78.2	0.513
2	0.100	21.8	100.0	0.301

ตารางที่ 37 ผลการทดสอบความมีนัยสำคัญในการจำแนกกลุ่มของสมการวิเคราะห์จำแนก
กรณีที่ 3 ด้วยค่า Wilks' Lambda

Function	Wilks' Lambda	Chi-square	df	Sig.
1	0.671	220.590	18	0.000
2	0.909	52.384	8	0.000

เมื่อพิจารณาค่าไอเกนของฟังก์ชันที่ 1 และฟังก์ชันที่ 2 จากตารางที่ 36 เท่ากับ 0.356 และ 0.100 ตามลำดับ เนื่องจากค่าไอเกนยิ่งมากจะแสดงถึงประสิทธิภาพในการจำแนกที่สูงตามไปด้วย ดังนั้นสมการวิเคราะห์จำแนกที่สร้างนี้มีประสิทธิภาพในการจำแนกกลุ่มในเกณฑ์ที่ไม่ดีนัก และค่าสหสัมพันธ์คานอนิคัลเท่ากับ 0.513 และ 0.301 สรุปได้ว่าฟังก์ชันที่ได้จากการวิเคราะห์การจำแนกทั้งสองฟังก์ชันสามารถอธิบายความผันแปรของกลุ่มที่จำแนกได้ 35.38% และเมื่อพิจารณาค่า Wilks' Lambda จากตารางที่ 37 ซึ่งเท่ากับ 0.671 และ 0.909 ซึ่งมี P-Value เท่ากับ 0.000 ทั้งคู่ นั่นคือ สมการวิเคราะห์จำแนกที่สร้างได้นี้สามารถจำแนกกลุ่มได้อย่างมีนัยสำคัญ

ตารางที่ 38 ค่าสัมประสิทธิ์แคนนอนิคัล และค่าสัมประสิทธิ์ แคนนอนิคัลมาตรฐาน และสหสัมพันธ์กับฟังก์ชันจำแนกของตัวแปรอิสระ สำหรับกรณีที่ 3

Variables	Unstandardized Coefficients		Standardized Coefficients		Correlation With Discriminant Function	
	1	2	1	2	1	2
X ₁	-0.008	0.029	-0.120	0.461	-0.140	0.487
X ₉	0.553	0.136	0.246	0.060	0.398	0.191
X ₁₅	1.058	1.096	0.497	0.515	0.566	0.407
X ₁₇	0.603	-0.584	0.276	-0.267	0.410	-0.206
X ₁₈	-0.510	1.025	-0.174	0.350	0.010	0.409
X ₂₂	0.778	0.044	0.369	0.021	0.370	0.111
X ₂₃	0.775	0.174	0.327	0.073	0.387	0.225
X ₂₅	0.941	-0.284	0.372	-0.112	0.334	0.245
X ₂₆	-0.925	1.256	-0.406	0.551	-0.260	0.662
Constant	-1.171	-2.103				

เมื่อพิจารณาค่าสัมประสิทธิ์มาตรฐานจากตารางที่ 38 พบว่าตัวแปร X_{15} (เยื่อตาอักเสบ) เป็นตัวแปรที่มีความสำคัญต่อประสิทธิภาพในการจำแนกของสมการจำแนกกลุ่มมากกว่าตัวแปรอื่นๆ เนื่องจากมีค่าสัมประสิทธิ์มาตรฐานที่สูงกว่าตัวแปรอื่นๆ

พิจารณาค่าสหสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 9 ตัวและฟังก์ชันการวิเคราะห์จำแนก โดยพิจารณาจากค่าสัมบูรณ์ของค่าสหสัมพันธ์โดยไม่สนใจเครื่องหมายบวกหรือลบ พบว่าตัวแปรที่นำมาใช้ในการสร้างสมการวิเคราะห์จำแนกฟังก์ชันแรกมีตัวแปร X_{15} (เยื่อตาอักเสบ) ที่มีค่าสหสัมพันธ์สูงกว่าตัวอื่นๆคือเท่ากับ 0.566 และเมื่อพิจารณาฟังก์ชันที่ 2 พบว่าตัวแปร X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase) มีค่าสหสัมพันธ์สูงกว่าตัวแปรอื่นๆ และเป็นค่าสหสัมพันธ์ที่สูงที่สุดสำหรับทั้ง 2 ฟังก์ชันนี้ คือเท่ากับ 0.662

ตารางที่ 39 ค่าสัมประสิทธิ์ของฟังก์ชันการวิเคราะห์จำแนกกรณีที่ 3

Variable	Discriminant Function Coefficient		
	Leptospirosis	Rickettsiosis	Others
X_1	0.155	0.180	0.154
X_9	1.396	0.692	0.690
X_{15}	3.571	2.634	1.937
X_{17}	1.923	0.794	1.398
X_{18}	-0.031	1.185	0.233
X_{22}	2.828	1.763	1.882
X_{23}	1.434	0.437	0.448
X_{25}	0.556	-0.897	-0.473
X_{26}	0.007	1.915	0.689
(Constant)	-7.696	-7.042	-5.332

จากค่าสัมประสิทธิ์ฟังก์ชันจำแนก จะได้สมการการวิเคราะห์การจำแนกดังนี้

$$Z_1 = -7.696 + (0.155 X_1) + (1.396 X_9) + (3.571 X_{15}) + (1.923 X_{17}) - (0.031 X_{18}) \\ + (2.828 X_{22}) + (1.434 X_{23}) + (0.556 X_{25}) + (0.007 X_{26})$$

$$Z_2 = -7.042 + (0.180 X_1) + (0.692 X_9) + (2.634 X_{15}) + (0.794 X_{17}) + (1.185 X_{18}) \\ + (1.763 X_{22}) + (0.437 X_{23}) - (0.897 X_{25}) + (1.915 X_{26})$$

$$Z_3 = -5.332 + (0.154 X_1) + (0.690 X_9) + (1.937 X_{15}) + (1.398 X_{17}) + (0.233 X_{18}) \\ + (1.882 X_{22}) + (0.448 X_{23}) - (0.473 X_{25}) + (0.689 X_{26})$$

ตารางที่ 40 ค่าศูนย์กลางของกลุ่ม สำหรับกรณีที่ 3

Group	Function at Group centroids	
	1	2
Leptospirosis	0.843	0.057
Rickettsiosis	-0.553	0.549
Others	-0.353	-0.280

จากตารางที่ 40 ซึ่งแสดงค่าศูนย์กลางของกลุ่มข้อมูลทั้ง 3 กลุ่มในแต่ละฟังก์ชัน พบว่า ค่าศูนย์กลางของแต่ละกลุ่มห่างกันพอสมควร ดังนั้นการสร้างสมการวิเคราะห์จำแนกจากข้อมูลชุดนี้น่าจะแยกกลุ่มได้ค่อนข้างดี

ตารางที่ 41 ผลการจัดข้อมูลเข้ากลุ่มโดยใช้การวิเคราะห์การจำแนก สำหรับกรณีที่ 3

Original Diagnosis	Predicted Diagnosis			Percentage Correct	Overall Percentage Correct
	Leptospirosis	Rickettsiosis	Others		
Leptospirosis	129	23	32	70.1	57.8
Rickettsiosis	22	61	31	53.5	
Others	66	62	133	51.0	

เมื่อทำการสร้างฟังก์ชันการวิเคราะห์จำแนกโดยกำหนดความน่าจะเป็นเบื้องต้น ของ ทั้ง 3 กลุ่มเท่ากันคือ 0.33 นำฟังก์ชันการวิเคราะห์จำแนกจากตารางที่ 39 มาคำนวณสำหรับทุกค่าสังเกตจากนั้นทำการจัดข้อมูลเข้ากลุ่ม จะได้อัตราการจำแนกข้อมูลเข้ากลุ่มถูกต้องแสดงดัง

ตารางที่ 41 จะได้ว่าผลการจัดเข้ากลุ่มถูกต้องรวมเท่ากับ 57.8% ซึ่งมาจากการจำแนกเข้ากลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสถูกต้อง 70.1% , จำแนกเข้ากลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis ถูกต้อง 53.5% และจากการจำแนกเข้ากลุ่มที่วินิจฉัยเป็นโรคอื่นๆถูกต้อง 51.0%

เมื่อนำค่าที่ได้จากตารางที่ 41 มาคำนวณหาค่าอัตราการจัดจำแนกผิดที่เห็นชัดเจน ได้เท่ากับ $(23+32+22+31+66+62)/559$ หรือมีค่าเท่ากับ 0.422

- 1.3.2 การสร้างตัวแบบการวิเคราะห์การถดถอยโลจิสติกสำหรับการจำแนกกรณีที่ 3 เนื่องจากการใช้โปรแกรม SPSS ในการวิเคราะห์การถดถอยโลจิสติกกรณีจำแนกข้อมูลเป็น 3 กลุ่ม ไม่มีการคัดเลือกตัวแปรให้ดังนั้นจึงทำการคัดเลือกโดยพิจารณาจากค่าการทดสอบ Likelihood Ratio Test ร่วมกับการพิจารณาการมีนัยสำคัญของการทดสอบ Wald Test ได้ผลลัพธ์แสดงเป็นลำดับดังนี้

ตารางที่ 42 การทดสอบความเหมาะสมและ R-Square ของตัวแบบกรณีที่ 3 โดยใช้ตัวแปร

อิสระ 19 ตัว

Method	Value	df	Sig.
Chi-Square (Pearson)	1140.700	1072	0.072
Chi-Square (Deviance)	915.138	1072	1.000
Nagelkerke R Square	0.409		

ตารางที่ 42 แสดงผลการตรวจสอบภาวะสารูปดี ทำการเปรียบเทียบค่าสังเกตด้วยค่าคาดหวังของความถี่ โดยใช้ตัวแปรอิสระทั้งหมดในตัวแบบจะเห็นว่ามีความเหมาะสมดีมาก (Excellent Fit) คือมีค่า P-Value เท่ากับ 1.000 ด้วยเกณฑ์ Deviance Criterion และมีค่า P-Value เท่ากับ 0.072 เมื่อใช้เกณฑ์ Pearson Criterion

เมื่อพิจารณาค่า R-Square จากตารางที่ 42 ได้ว่า ฟังก์ชันที่ได้จากการวิเคราะห์การถดถอยโลจิสติกสามารถอธิบายความผันแปรของกลุ่มที่จำแนกได้ 40.90%

ตารางที่ 43 ค่าจากการทดสอบ Likelihood Ratio Test ของตัวแปรอิสระแต่ละตัวกรณีที่ 3

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig
Intercept	917.911	0.000	0	
X ₁	926.126	8.215	2	0.016
X ₂	919.108	1.197	2	0.550
X ₃	921.495	3.584	2	0.167
X ₆	924.546	6.635	2	0.036
X ₇	919.449	1.538	2	0.463
X ₉	921.368	3.458	2	0.177
X ₁₀	919.354	1.443	2	0.486
X ₁₁	922.760	4.850	2	0.088
X ₁₃	922.666	4.755	2	0.093
X ₁₅	952.264	34.354	2	0.000
X ₁₆	920.907	2.996	2	0.224
X ₁₇	925.137	7.227	2	0.027
X ₁₈	925.297	7.387	2	0.025
X ₂₂	935.494	17.584	2	0.000
X ₂₃	921.557	3.647	2	0.161
X ₂₄	922.544	4.634	2	0.099
X ₂₅	926.819	8.908	2	0.012
X ₂₆	944.255	26.344	2	0.000
X ₂₇	921.062	3.151	2	0.207

ตาราง Likelihood ratio test ให้ค่า -2Log Likelihood สำหรับแต่ละตัวแปรอิสระแต่ละตัว แสดงการปรับตัวแบบโดยใช้ตัวแปรอิสระทีละตัว ค่า P-Value ที่น้อยกว่าระดับนัยสำคัญที่กำหนด แสดงให้เห็นว่าตัวแบบแย่งอย่างมีนัยสำคัญเมื่อตัดตัวแปรอิสระแต่ละตัวออกไป ผลจากตารางที่ 43 จะได้ว่ามีตัวแปรอิสระ 8 ตัวที่จำแนกข้อมูลแต่ละกลุ่มออกได้อย่างน่าเชื่อถือ ได้แก่ X₁(อายุ), X₆(มื้ออาหาร), X₁₅(เชื่อบุตาอักเสบ), X₁₇(กดเจ็บกล้ามเนื้อ), X₁₈(ตับโต)

, X_{22} (PMN>75%), X_{25} (Total Bilirubin >3 mg/dl) และ X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase)

ตารางที่ 44 ผลการวิเคราะห์ข้อมูลเพื่อสร้างสมการถดถอยโลจิสติกกรณีที่มี 3 ตัวแปรอิสระ 19 ตัว

Variables	$\hat{\beta}$	S.E. ($\hat{\beta}$)	Wald	Sig.	Exp($\hat{\beta}$)	95% CI. For Exp($\hat{\beta}$)	
						Lower	Upper
Leptospirosis							
Intercept	2.940	0.893	10.844	0.001			
X_1	-0.004	0.008	0.225	0.636	0.996	0.981	1.012
X_2 (=1)	-0.127	0.308	0.170	0.680	0.881	0.482	1.610
X_3 (=0)	-0.616	0.507	1.476	0.224	0.540	0.200	1.459
X_6 (=0)	-0.636	0.288	4.867	0.027	0.530	0.301	0.932
X_7 (=0)	-0.302	0.245	1.526	0.217	0.739	0.458	1.194
X_9 (=0)	-0.424	0.268	2.497	0.114	0.654	0.387	1.107
X_{10} (=0)	0.269	0.415	0.420	0.517	1.308	0.580	2.949
X_{11} (=0)	-0.071	0.303	0.055	0.814	0.931	0.514	1.687
X_{13} (=0)	0.077	0.390	0.039	0.844	1.080	0.503	2.317
X_{15} (=0)	-1.464	0.261	31.559	0.000	0.231	0.139	0.385
X_{16} (=0)	-0.884	0.544	2.637	0.104	0.413	0.142	1.201
X_{17} (=0)	-0.352	0.248	2.019	0.155	0.703	0.433	1.143
X_{18} (=0)	0.422	0.400	1.115	0.291	1.525	0.697	3.338
X_{22} (=0)	-0.925	0.251	13.549	0.000	0.397	0.242	0.649
X_{23} (=0)	-0.519	0.324	2.561	0.110	0.595	0.315	1.124
X_{24} (=0)	-0.353	0.282	1.568	0.210	0.703	0.405	1.221
X_{25} (=0)	-0.823	0.360	5.212	0.022	0.439	0.217	0.890
X_{26} (=0)	0.729	0.290	6.303	0.012	2.073	1.173	3.662
X_{27} (=0)	-0.597	0.338	3.114	0.078	0.550	0.283	1.068

ตารางที่ 44 (ต่อ)

Variables	$\hat{\beta}$	S.E.($\hat{\beta}$)	Wald	Sig.	Exp($\hat{\beta}$)	95% CI. For Exp($\hat{\beta}$)	
						Lower	Upper
Rickettsiosis							
Intercept	-1.526	1.149	1.765	0.184			
X ₁	0.020	0.008	6.639	0.010	1.020	1.005	1.036
X ₂ (=1)	0.255	0.299	0.728	0.394	1.290	0.718	2.318
X ₃ (=0)	-0.823	0.544	2.288	0.130	0.439	0.151	1.276
X ₆ (=0)	0.164	0.289	0.324	0.569	1.179	0.669	2.076
X ₇ (=0)	-0.067	0.266	0.064	0.801	0.935	0.555	1.576
X ₉ (=0)	0.122	0.313	0.153	0.696	1.130	0.612	2.086
X ₁₀ (=0)	0.676	0.607	1.243	0.265	1.967	0.599	6.462
X ₁₁ (=0)	-0.650	0.302	4.639	0.031	0.522	0.289	0.943
X ₁₃ (=0)	1.231	0.651	3.582	0.058	3.426	0.957	12.259
X ₁₅ (=0)	-0.620	0.269	5.303	0.021	0.538	0.317	0.912
X ₁₆ (=0)	-0.239	0.645	0.137	0.712	0.788	0.222	2.790
X ₁₇ (=0)	0.524	0.314	2.781	0.095	1.689	0.912	3.126
X ₁₈ (=0)	-0.743	0.385	3.723	0.054	0.476	0.224	1.012
X ₂₂ (=0)	0.176	0.252	0.487	0.485	1.192	0.727	1.955
X ₂₃ (=0)	0.153	0.358	0.181	0.670	1.165	0.577	2.352
X ₂₄ (=0)	-0.609	0.295	4.276	0.039	0.544	0.305	0.969
X ₂₅ (=0)	0.360	0.401	0.808	0.369	1.434	0.654	3.143
X ₂₆ (=0)	-0.872	0.268	10.602	0.001	0.418	0.247	0.707
X ₂₇ (=0)	-0.309	0.370	0.696	0.404	0.734	0.355	1.517

ตารางที่ 44 ประกอบด้วย 2 ตารางย่อยซึ่งนำเสนอการพยากรณ์พารามิเตอร์ (Parameter Estimates) ตารางแรก ทำการเปรียบเทียบระหว่างกลุ่มผู้ป่วยที่เมื่อเสร็จสิ้นโครงการวิจัยแล้ววินิจฉัยว่าเป็นโรคเลปโตสไปโรสิสกับผู้ป่วยที่วินิจฉัยว่าเป็นโรคอื่นๆ และตารางย่อยที่ 2 ทำการ

เปรียบเทียบระหว่างผู้ป่วยที่เมื่อเสร็จสิ้นโครงการวิจัยแล้ววินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis กับผู้ป่วยที่วินิจฉัยว่าเป็นโรคอื่นๆ ใช้ระดับนัยสำคัญที่ 0.05 ด้วยเกณฑ์นี้มีเพียงตัวแปร X_6 (เบื้ออาหาร), X_{15} (เยื่อตาอักเสบ), X_{22} (PMN>75%), X_{25} (Total Bilirubin >3 mg/dl) และ X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase) เท่านั้นเป็นตัวแปร ที่สามารถแยกกลุ่มผู้ป่วยที่เมื่อเสร็จสิ้นโครงการวิจัยแล้ววินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กับผู้ป่วยที่วินิจฉัยว่าเป็นโรคอื่นๆ ได้อย่างน่าเชื่อถือ (Reliably Separates) และสำหรับกลุ่มผู้ป่วยที่เมื่อเสร็จสิ้นโครงการวิจัยแล้ววินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis กับผู้ป่วยที่วินิจฉัยว่าเป็นโรคอื่นๆ มีตัวแปร X_1 (อายุ), X_{11} (ความรู้สึกตัวผิดปกติ), X_{15} (เยื่อตาอักเสบ), X_{24} (Abnormal Urine Examination) และ X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase) ที่สามารถแยกกลุ่มผู้ป่วยได้อย่างน่าเชื่อถือ

ตารางที่ 45 ผลการจัดค่าข้อมูลเข้ากลุ่มเมื่อใช้การวิเคราะห์การถดถอยโลจิสติกกรณี 3 ใช้ 19 ตัวแปร

Original Diagnosis	Predicted Diagnosis			Percentage Correct	Overall Percentage Correct
	Leptospirosis	Rickettsiosis	Others		
Leptospirosis	123	6	55	66.8	63.0
Rickettsiosis	19	37	58	32.5	
Others	48	21	192	73.6	

จากผลลัพธ์ที่แสดงดังตารางที่ 45 พบว่า หากใช้ตัวแปรอิสระทั้งหมด 19 ตัว สามารถจำแนกค่าสังเกตเข้ากลุ่มได้ถูกต้องรวม 63.0% ซึ่งมาจากการกรณีที่คนไข้เป็นโรคเลปโตสไปโรซิสถูกต้อง 66.8% , คนไข้ที่เป็นโรคในกลุ่ม Rickettsiosis ถูกต้อง 32.5% และในกลุ่มอื่นๆถูกต้อง 73.6% คำนวณค่า APER เท่ากับ $(6+55+19+58+48+21)/559$ เท่ากับ 0.370

จากนั้นนำข้อมูลชุดเดิมมาทำการวิเคราะห์ โดยใช้การวิเคราะห์การถดถอยโลจิสติก โดยใช้ตัวแปรที่ผ่านการพิจารณาค่า P-value ของ Likelihood-Ratio Test จากตารางที่ 43 และค่า Parameter Estimated จากตารางที่ 44 จะได้ตัวแปรเข้าสมการ 10 ตัวแปร จากตัวแปรทั้งหมด 19 ตัว ได้แก่ X_1 (อายุ), X_6 (เบื้ออาหาร), X_{11} (ความรู้สึกตัวผิดปกติ), X_{15} (เยื่อตาอักเสบ), X_{17} (กดเจ็บกล้ามเนื้ออ่อน), X_{18} (ตับโต), X_{22} (PMN>75%), X_{24} (Abnormal Urine Examination), X_{25} (Total Bilirubin >3 mg/dl) และ X_{26} (Abnormal SGOT/SGPT/Alkaline Phosphatase) นำตัวแปรอิสระ 10 ตัวนี้มาสร้างตัวแบบวิเคราะห์การถดถอยโลจิสติก ผลเสนอตามลำดับต่อไปนี้

ตารางที่ 46 การทดสอบความเหมาะสมและ R-Square ของตัวแบบกรณีที่ 3 ใช้ตัวแปรอิสระ 10 ตัว

Method	Value	df	Sig.
Chi-Square (Pearson)	1080.600	1010	0.061
Chi-Square (Deviance)	905.722	1010	0.992
Nagelkerke R Square	0.364		

สถิติทดสอบ Goodness-of-fit ทำการเปรียบเทียบค่าสังเกตด้วยค่าคาดหวังของความถี่ โดยใช้ตัวแปรอิสระทั้งหมดในตัวแบบจะเห็นว่ามีความเหมาะสมดีมาก คือมีค่า P-Value เท่ากับ 0.992 ด้วยเกณฑ์ Deviance Criterion และมีค่า P-Value เท่ากับ 0.061 เมื่อใช้เกณฑ์ Pearson Criterion และเมื่อพิจารณาค่า R-Square จากตารางที่ 46 ได้ว่า ฟังก์ชันที่ได้จากการวิเคราะห์การถดถอยโลจิสติกสามารถอธิบายความผันแปรของกลุ่มที่จำแนกได้ 36.40%

ตารางที่ 47 ค่าจากการทดสอบ Likelihood Ratio Test ของตัวแปรอิสระแต่ละตัวเมื่อใช้

การวิเคราะห์การถดถอยโลจิสติก กรณีที่ 3 ตัวแปรอิสระ 10 ตัวแปร

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig
Intercept	929.525	0.000	0	
X ₁	940.623	11.098	2	0.004
X ₆	937.347	7.822	2	0.020
X ₁₁	936.031	6.507	2	0.039
X ₁₅	984.726	55.201	2	0.000
X ₁₇	940.614	11.089	2	0.004
X ₁₈	935.992	6.467	2	0.039
X ₂₂	949.839	20.314	2	0.000
X ₂₄	938.109	8.584	2	0.014
X ₂₅	957.828	28.303	2	0.000
X ₂₆	958.654	29.129	2	0.000

ตาราง Likelihood Ratio Test ให้ค่า -2Log Likelihood สำหรับแต่ละตัวแปรอิสระแต่ละตัว แสดงการปรับตัวแบบโดยใช้ตัวแปรอิสระทีละตัว พบว่าทั้ง 10 ตัวแปร มีค่า P-Value ที่น้อยกว่าระดับนัยสำคัญที่กำหนด (0.05) จึงแสดงให้เห็นว่าตัวแบบจะแย่ลงอย่างมีนัยสำคัญเมื่อตัดตัวแปรอิสระแต่ละตัวออกไป

ตารางที่ 48 ผลการวิเคราะห์ข้อมูลเพื่อสร้างสมการถดถอยโลจิสติกกรณีที่มี 3 ตัวแปรอิสระ 10 ตัว

Variables	$\hat{\beta}$	S.E.($\hat{\beta}$)	Wald	Sig.	Exp($\hat{\beta}$)	95% CI. For Exp($\hat{\beta}$)	
						Lower	Upper
Leptospirosis							
Intercept	2.319	0.589	15.520	0.000			
X ₁	-0.003	0.007	0.218	0.640	0.997	0.982	1.011
X ₆ (=0)	-0.624	0.277	5.064	0.024	0.536	0.311	0.923
X ₁₁ (=0)	-0.317	0.288	1.206	0.272	0.729	0.414	1.282
X ₁₅ (=0)	-1.720	0.245	49.455	0.000	0.179	0.111	0.289
X ₁₇ (=0)	-0.443	0.236	3.513	0.061	0.642	0.404	1.020
X ₁₈ (=0)	0.123	0.374	0.108	0.743	1.131	0.543	2.354
X ₂₂ (=0)	-0.966	0.243	15.816	0.000	0.381	0.236	0.613
X ₂₄ (=0)	-0.652	0.256	6.478	0.011	0.521	0.315	0.861
X ₂₅ (=0)	-1.347	0.319	17.854	0.000	0.260	0.139	0.486
X ₂₆ (=0)	0.717	0.280	6.571	0.010	2.049	1.184	3.547
Rickettsiosis							
Intercept	-0.095	0.609	0.024	0.877			
X ₁	0.022	0.007	8.978	0.003	1.023	1.008	1.038
X ₆ (=0)	0.242	0.278	0.761	0.383	1.274	0.739	2.195
X ₁₁ (=0)	-0.745	0.291	6.531	0.011	0.475	0.268	0.841
X ₁₅ (=0)	-0.635	0.257	6.117	0.013	0.530	0.320	0.876
X ₁₇ (=0)	0.585	0.305	3.684	0.055	1.795	0.988	3.263
X ₁₈ (=0)	-0.809	0.366	4.896	0.027	0.445	0.217	0.912
X ₂₂ (=0)	0.155	0.245	0.400	0.527	1.168	0.722	1.889

ตารางที่ 48 (ต่อ)

Variables	$\hat{\beta}$	S.E.($\hat{\beta}$)	Wald	Sig.	Exp($\hat{\beta}$)	95% CI. For Exp($\hat{\beta}$)	
						Lower	Upper
X ₂₄ (=0)	-0.603	0.275	4.827	0.028	0.547	0.319	0.937
X ₂₅ (=0)	0.429	0.367	1.368	0.242	1.536	0.748	3.154
X ₂₆ (=0)	-0.906	0.260	12.130	0.000	0.404	0.243	0.673

จากตารางที่ 48 พิจารณาตัวสถิติ Wald ซึ่งใช้ในการทดสอบว่าค่าประมาณสัมประสิทธิ์ของตัวแปรนั้นมีค่าเท่ากับศูนย์หรือไม่ ตารางที่ 48 มีตารางย่อย 2 ตารางที่นำเสนอการพยากรณ์พารามิเตอร์ (Parameter Estimates) ตารางแรกทำการเปรียบเทียบระหว่างกลุ่มผู้ป่วยที่เมื่อเสร็จสิ้นโครงการวิจัยแล้ววินิจฉัยว่าเป็นโรคเลปโตสไปโรซิสกับผู้ป่วยที่วินิจฉัยว่าเป็นโรคอื่นๆ และตารางย่อยที่ 2 ทำการเปรียบเทียบระหว่างผู้ป่วยที่เมื่อเสร็จสิ้นโครงการวิจัยแล้ววินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis กับผู้ป่วยที่วินิจฉัยว่าเป็นโรคอื่นๆ ใช้ระดับนัยสำคัญที่ 0.05 ด้วยเกณฑ์นี้พบว่ามีตัวแปร 6 ตัวคือ X₆ (เบื่ออาหาร), X₁₅ (เยื่อตาอักเสบ), X₂₂ (PMN>75%), X₂₄ (Abnormal Urine Examination), X₂₅ (Total Bilirubin >3 mg/dl) และ X₂₆ (Abnormal SGOT/SGPT/Alkaline Phosphatase) ที่สามารถแยกกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยเป็นโรคเลปโตสไปโรซิสและกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ ส่วนตัวแปรอิสระที่สามารถแยกกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ ได้อย่างน่าเชื่อถือคือตัวแปร X₁ (อายุ), X₁₁ (ความรู้สึกตัวผิดปกติ), X₁₅ (เยื่อตาอักเสบ), X₁₈ (ตับโต), X₂₄ (Abnormal Urine Examination) และ X₂₆ (Abnormal SGOT/SGPT/Alkaline Phosphatase)

สามารถเขียนลอจิสต์ตัวอย่างสำหรับตัวแบบถดถอยโลจิสติก ได้ดังนี้

$$\hat{g}_1(x) = 2.319 - (0.003 * X_1) - (0.624 * (X_6 = 0)) - (0.317 * (X_{11} = 0)) - (1.720 * (X_{15} = 0)) \\ - (0.443 * (X_{17} = 0)) + (0.123 * (X_{18} = 0)) - (0.966 * (X_{22} = 0)) \\ - (0.652 * (X_{24} = 0)) - (1.347 * (X_{25} = 0)) + (0.717 * (X_{26} = 0))$$

$$\hat{g}_2(x) = -0.095 + (0.022 * X_1) + (0.242 * (X_6 = 0)) - (0.745 * (X_{11} = 0)) - 0.635 * (X_{15} = 0) \\ + (0.585 * (X_{17} = 0)) - (0.809 * (X_{18} = 0)) + (0.155 * (X_{22} = 0)) \\ - (0.603 * (X_{24} = 0)) + (0.429 * (X_{25} = 0)) - (0.906 * (X_{26} = 0))$$

$\hat{g}_3(x) = 0$ เนื่องจาก กลุ่ม Others เป็น Reference Category หรือเป็นฐานในการถูกเปรียบเทียบ

นำค่าสัมประสิทธิ์ที่ได้ มาคำนวณความน่าจะเป็นของการเป็นสมาชิกของแต่ละกลุ่ม โดยใช้ตัวแบบมัลติโนเมียลโลจิสติกต่อไปนี้

$$P(x_i) = \frac{\exp(\hat{g}_i)}{\sum_{i=1}^3 \exp(\hat{g}_i)}$$

เมื่อนำความน่าจะเป็นค่า $P(x_i)$ สำหรับแต่ละกลุ่ม แล้วนำค่าที่ได้มาเปรียบเทียบกัน ค่าใดสูงที่สุดจะจัดให้ค่าสังเกตอยู่ในกลุ่มนั้น โดยที่กลุ่มที่ 1 แทนกลุ่มคนไข้ที่ได้รับการวินิจฉัยเป็นโรคเลปโตสไปโรสิส กลุ่มที่ 2 แทนกลุ่มคนไข้ที่ได้รับการวินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่ 3 แทนกลุ่มคนไข้ที่ได้รับการวินิจฉัยเป็นโรคอื่นๆ

ตารางที่ 49 ผลการจัดค่าข้อมูลเข้ากลุ่มโดยการวิเคราะห์การถดถอยโลจิสติกกรณี 3 ใช้ 10 ตัวแปร

Original	Predicted			Percentage Correct	Overall Percentage Correct
	Leptospirosis	Rickettsiosis	Others		
Leptospirosis	116	9	59	63.0	62.1
Rickettsiosis	16	38	60	33.3	
Others	50	18	193	73.9	

จากผลลัพธ์ที่แสดงดังตารางที่ 49 พบว่า หากใช้ตัวแปรอิสระ 10 ตัวนี้จะสามารถจำแนกค่าสังเกตเข้ากลุ่มได้ถูกต้องรวม 62.1% ซึ่งมาจากการจัดค่าสังเกตในกลุ่มคนไข้ที่เป็นโรคเลปโตสไปโรสิสถูกต้อง 63.0% , คนไข้เป็นโรคในกลุ่ม Rickettsiosis ถูกต้อง 33.3% และกลุ่มอื่นๆถูกต้อง 73.9% คำนวณค่า APER ได้เท่ากับ $(9+59+16+60+50+18)/559=0.379$

เมื่อนำผลการจำแนกข้อมูลเข้ากลุ่มได้ถูกต้องทั้งกรณีผู้ป่วยเลปโตสไปโรสิส , ผู้ป่วยเป็นโรคในกลุ่ม Rickettsiosis และผู้ป่วยอื่นๆ รวมทั้งผลการจำแนกรวม และค่า APER มาเปรียบเทียบกัน ระหว่างการใช้วิธีการวิเคราะห์จำแนกและวิธีการถดถอยโลจิสติก ผลแสดงดังตารางต่อไปนี้

ตารางที่ 50 ผลการจำแนกข้อมูลเข้ากลุ่มกรณีที่ 3 ใช้ 10 ตัวแปรเปรียบเทียบ 2 วิธี

Technique	Observed	Predicted			Percentage Correct	Overall Percentage Correct	APER
		Leptospirosis	Rickettsiosis	Others			
Discriminant	Leptospirosis	129	23	32	70.1	57.8	0.422
	Rickettsiosis	22	61	31	53.5		
	Others	66	62	133	51.0		
Logistic Regression	Leptospirosis	116	9	59	63.0	62.1	0.379
	Rickettsiosis	16	38	60	33.3		
	Others	50	18	193	73.9		

จากผลการจำแนกข้อมูลเข้ากลุ่มถูกต้องในตารางที่ 50 พบว่า สำหรับการจำแนกเป็น 3 กลุ่มของข้อมูลในงานวิจัยนี้ การวิเคราะห์จำแนกจะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่าการวิเคราะห์การถดถอยโลจิสติกทั้งการพิจารณาผลการจัดข้อมูลเข้ากลุ่มรวม และพิจารณาแยกแต่ละกรณี

2. การตรวจสอบตัวแบบโดยใช้ข้อมูลชุดใหม่

นำข้อมูลชุดใหม่ ซึ่งเป็นผู้ป่วยที่ทำการวิจัยและเก็บข้อมูลในโรงพยาบาลและช่วงเวลาเดียวกัน ที่แยกออกมาเพื่อใช้ในการทดสอบโดยเฉพาะ กรณีที่ 1 และ 3 จำนวน 149 ราย และกรณีที่ 2 จำนวน 105 รายมาใช้ในการทดสอบตัวแบบที่สร้างขึ้น ผลการจัดข้อมูลชุดใหม่เข้ากลุ่มแสดงเป็นลำดับดังนี้

กรณีที่ 1

นำตัวแบบที่ได้จากกรณีที่ 1 ในขั้นตอนที่ 1 มาทำการตรวจสอบ โดยนำมาใช้จัดค่าสังเกตในข้อมูลชุดใหม่เข้ากลุ่ม แล้วคำนวณผลการจัดเข้ากลุ่มได้ถูกต้อง จำนวนค่าสังเกตเมื่อตัดข้อมูลที่สูญหายทิ้งแล้ว มีจำนวน 149 ค่าสังเกต

2.1 ตัวแบบการวิเคราะห์การจำแนก สำหรับกรณีที่ 1

นำสัมประสิทธิ์ฟังก์ชันจำแนกที่ได้จากการวิเคราะห์การจำแนกในขั้นตอนที่ 1 ซึ่งจำแนกข้อมูลออกเป็น 2 กลุ่มคือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรซิสและกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ มาแทนค่าในชุดข้อมูลใหม่ ได้ผลการจัดค่าสังเกตเข้ากลุ่มถูกต้อง แสดงดังตารางที่ 51

ตารางที่ 51 ผลการจำแนกชุดใหม่ เข้ากลุ่ม โดยใช้การวิเคราะห์การจำแนกจากกรณีที่ 1

Original Diagnosis	Predicted Diagnosis		Percentage Correct	Overall Percentage Correct
	Leptospirosis	Others		
Leptospirosis	49	21	70.00	57.06
Others	43	36	45.60	

2.2 ตัวแบบการวิเคราะห์การถดถอยโลจิสติก สำหรับกรณีที่ 1

นำสัมประสิทธิ์การถดถอยโลจิสติก ที่ได้จากการจำแนกข้อมูลในขั้นตอนที่ 1 ซึ่งจำแนกข้อมูลออกเป็น 2 กลุ่มคือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรซิสและกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ มาแทนค่าในชุดข้อมูลชุดใหม่ จากนั้นคำนวณค่า $P(x)$ ใช้ Cut Point เท่ากับ 0.3 ได้ผลการจัดค่าสังเกตเข้ากลุ่มถูกต้อง แสดงดังตาราง

ตารางที่ 52 ผลการจำแนกข้อมูลชุดใหม่เข้ากลุ่มโดยใช้การวิเคราะห์การถดถอยโลจิสติก กรณีที่ 1

Original Diagnosis	Predicted Diagnosis		Percentage Correct	Overall Percentage Correct
	Leptospirosis	Others		
Leptospirosis	51	19	72.90	55.72
Others	47	32	40.50	

สำหรับกรณีที่ 1 พบว่าผลการตรวจสอบตัวแบบโดยใช้ข้อมูลชุดใหม่ เทคนิคการวิเคราะห์การจำแนกสูงกว่า การวิเคราะห์การถดถอยโลจิสติกเล็กน้อย เมื่อพิจารณาจากเปอร์เซ็นต์การจำแนกถูกรวม และพบว่าการวิเคราะห์การถดถอยโลจิสติกให้เปอร์เซ็นต์การจำแนกที่สูงกว่าเมื่อพิจารณาเฉพาะกรณีที่คนไข้ได้รับการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส

กรณีที่ 2

นำตัวแบบที่ได้จากกรณีที่ 2 ในขั้นตอนที่ 1 มาทำการตรวจสอบ โดยนำมาใช้จัดค่าสังเกตในข้อมูลชุดใหม่เข้ากลุ่ม แล้วคำนวณผลการจัดเข้ากลุ่มได้ถูกต้อง จำนวนค่าสังเกตเมื่อตัดข้อมูลที่สูญหายทิ้งแล้ว มีจำนวน 105 ค่าสังเกต

2.3 ตัวแบบการวิเคราะห์การจำแนก สำหรับกรณีที่ 2

นำสัมประสิทธิ์ฟังก์ชันจำแนกที่ได้จากการวิเคราะห์การจำแนกในขั้นตอนที่ 1 ซึ่งจำแนกข้อมูลออกเป็น 2 กลุ่มคือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis มาแทนค่าในชุดข้อมูลใหม่ ได้ผลการจัดค่าสังเกตเข้ากลุ่มถูกต้องดังนี้

ตารางที่ 53 ผลการจำแนกชุดใหม่ เข้ากลุ่ม โดยใช้วิธีการวิเคราะห์การจำแนกจากกรณีที่ 2

Original Diagnosis	Predicted Diagnosis		Percentage Correct	Overall Percentage Correct
	Leptospirosis	Rickettsiosis		
Leptospirosis	53	17	75.70	66.67
Rickettsiosis	18	17	48.60	

2.4 ตัวแบบการวิเคราะห์การถดถอยโลจิสติก สำหรับกรณีที่ 2

นำสัมประสิทธิ์การถดถอยโลจิสติก ที่ได้จากการจำแนกข้อมูลในขั้นตอนที่ 1 ซึ่งจำแนกข้อมูลออกเป็น 2 กลุ่มคือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis มาแทนค่าในชุดข้อมูลชุดใหม่ จากนั้นคำนวณค่า $P(x)$ ใช้ Cut Point เท่ากับ 0.5 ได้ผลการจัดค่าสังเกตเข้ากลุ่มถูกต้อง แสดงดังตารางที่ 54

ตารางที่ 54 ผลการจำแนกข้อมูลชุดใหม่เข้ากลุ่ม โดยใช้การวิเคราะห์การถดถอยโลจิสติกกรณีที่ 2

Original Diagnosis	Predicted Diagnosis		Percentage Correct	Overall Percentage Correct
	Leptospirosis	Rickettsiosis		
Leptospirosis	59	11	84.30	69.53
Rickettsiosis	21	14	40.00	

สำหรับกรณีที่ 2 พบว่าผลการตรวจสอบตัวแบบโดยใช้ข้อมูลชุดใหม่ เทคนิคการวิเคราะห์การถดถอยโลจิสติก มีประสิทธิภาพในการจำแนกสูงกว่าการวิเคราะห์การจำแนกเมื่อพิจารณาจากเปอร์เซ็นต์การจำแนกโดยรวม และผลการจำแนกถูกสำหรับกรณีที่ใช้ได้รับการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิสผลเป็นไปในทำนองเดียวกัน

กรณีที่ 3

นำตัวแบบที่ได้จากกรณีที่ 3 ในขั้นตอนที่ 1 มาทำการตรวจสอบ โดยนำมาใช้จัดค่าสังเกตในข้อมูลชุดใหม่เข้ากลุ่ม แล้วคำนวณผลการจัดเข้ากลุ่มได้ถูกต้อง จำนวนค่าสังเกตเมื่อตัดข้อมูลที่สูญหายทิ้งแล้ว มีจำนวน 149 ค่าสังเกต

2.5 ตัวแบบการวิเคราะห์การจำแนก สำหรับกรณีที่ 3

นำสัมประสิทธิ์ฟังก์ชันจำแนกที่ได้จากการวิเคราะห์การจำแนกข้อมูลในขั้นตอนที่ 1 ซึ่งจำแนกข้อมูลออกเป็น 3 กลุ่มคือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิส , กลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ มาแทนค่าในชุดข้อมูลชุดใหม่ ได้ผลการจัดค่าสังเกตเข้ากลุ่มถูกต้อง แสดงดังตาราง

ตารางที่ 55 ผลการจำแนกข้อมูลชุดใหม่เข้ากลุ่ม โดยใช้การวิเคราะห์จำแนกสำหรับกรณีที่ 3

Original Diagnosis	Predicted Diagnosis			Percentage Correct	Overall Percentage Correct
	Leptospirosis	Rickettsiosis	Others		
Leptospirosis	47	12	11	67.10	47.61
Rickettsiosis	15	11	9	31.40	
Others	21	10	13	29.50	

2.6 ตัวแบบการวิเคราะห์การถดถอยโลจิสติก สำหรับกรณีที่ 3

นำสัมประสิทธิ์การถดถอยโลจิสติก ที่ได้จากการจำแนกข้อมูลในขั้นตอนที่ 1 ซึ่งจำแนกข้อมูลออกเป็น 3 กลุ่มคือกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิส , กลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ มาแทนค่าในชุดข้อมูลชุดใหม่ จากนั้นคำนวณค่า $P(x)$ ใช้จุดตัดการจำแนกที่ 0.5 ได้ผลการจัดค่าสังเกตเข้ากลุ่มถูกต้องแสดงดังนี้

ตารางที่ 56 ผลการจำแนกข้อมูลชุดใหม่เข้ากลุ่มโดยใช้การวิเคราะห์การถดถอยโลจิสติกกรณีที่ 3

Original Diagnosis	Predicted Diagnosis			Percentage Correct	Overall Percentage Correct
	Leptospirosis	Rickettsiosis	Others		
Leptospirosis	11	38	21	15.7	29.52
Rickettsiosis	10	19	6	54.3	
Others	3	27	14	31.8	

สำหรับกรณีที่ 3 พบว่าผลการตรวจสอบตัวแบบโดยใช้ข้อมูลชุดใหม่ เทคนิคการวิเคราะห์การจำแนกมีประสิทธิภาพในการจำแนกสูงกว่าการวิเคราะห์การถดถอยโลจิสติกเมื่อพิจารณาจากเปอร์เซ็นต์การจำแนกโดยรวม และผลการจำแนกถูกสำหรับกรณีที่คนไข้ได้รับการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิสผลเป็นไปในทำนองเดียวกัน

3. การเปรียบเทียบผลการจัดข้อมูลเข้ากลุ่มของตัวแบบที่สร้างและเกณฑ์ขององค์การอนามัยโลก

เนื่องจากองค์การอนามัยโลก ได้เสนอเกณฑ์ที่ใช้ในการวินิจฉัยโรคเลปโตสไปโรสิส ซึ่งแยกเป็นกรณีที่คนไข้เป็นโรคเลปโตสไปโรสิสหรือไม่เป็นเท่านั้น ดังนั้นในการเปรียบเทียบประสิทธิภาพการจำแนกกับตัวแบบที่สร้างจากงานวิจัยนี้ จึงเปรียบเทียบได้เพียงผลที่ได้จากกรณีที่ 1 เท่านั้น ผลการจัดค่าสังเกตเข้ากลุ่มโดยใช้เกณฑ์ขององค์การอนามัยโลกพิจารณาเฉพาะ 2 ส่วนแรกซึ่งเกี่ยวข้องกับผลการตรวจร่างกายเบื้องต้น และประวัติเสี่ยงของคนไข้ เทียบกับผลเมื่อใช้การวิเคราะห์จำแนก , การวิเคราะห์การถดถอยโลจิสติก สำหรับการจำแนกข้อมูลกรณีที่ 1 แสดงดังตารางต่อไปนี้

ตารางที่ 57 ผลการจำแนกข้อมูลเข้ากลุ่ม โดยใช้วิธีการวิเคราะห์จำแนก , การถดถอยโลจิสติก และใช้เกณฑ์ขององค์การอนามัยโลก สำหรับการจำแนกข้อมูลเป็น 2 กลุ่ม

Technique	Observed	Predicted		Percentage Correct	Overall Percentage Correct	APER
		Leptospirosis	Others			
Discriminant	Leptospirosis	140	44	76.10	74.60	0.254
	Others	98	277	73.90		
Logistic Regression	Leptospirosis	147	37	79.9	72.30	0.277
	Others	118	257	68.5		
WHO Criterion	Leptospirosis	7	177	3.80	67.06	0.329
	Others	7	368	98.10		

เมื่อพิจารณาผลการจัดค่าสังเกตเข้ากลุ่มถูกต้องรวม และค่า APER พบว่า ตัวแบบที่ได้จากการวิเคราะห์การจำแนก มีประสิทธิภาพสูงกว่าตัวแบบจากการวิเคราะห์การถดถอยโลจิสติก และขององค์การอนามัยโลก โดยที่ผลการจัดค่าสังเกตเข้ากลุ่มถูกต้องรวม, ผลการจัดค่าสังเกตเข้ากลุ่มกรณีคนไข้ได้รับการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิสและค่า APER โดยใช้เกณฑ์ขององค์การอนามัยโลกเมื่อพิจารณาเฉพาะ 2 ส่วนแรกมีประสิทธิภาพต่ำที่สุด

หากพิจารณาผลการจัดเข้ากลุ่มถูกต้อง เฉพาะกลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิส พบว่าตัวแบบการวิเคราะห์การถดถอยโลจิสติกมีประสิทธิภาพสูงสุด โดยสูงกว่าตัวแบบการวิเคราะห์การจำแนกกลุ่มและตัวแบบอิงเกณฑ์องค์การอนามัยโลก โดยมีเปอร์เซ็นต์การจำแนกถูกในกลุ่มที่สนใจ เท่ากับ 79.9% , 76.1% และ 3.8% ตามลำดับ

ซึ่งผลการจัดข้อมูลเข้ากลุ่มถูกต้องตามที่นำเสนอไว้ข้างต้น เป็นไปตามสมมติฐานของการวิจัยที่ตั้งไว้

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยนี้ทำการศึกษาเปรียบเทียบ ประสิทธิภาพในการจำแนกของตัวแบบที่ได้จากการวิเคราะห์การจำแนกและตัวแบบการวิเคราะห์การถดถอยโลจิสติก โดยตัดสินจากเปอร์เซ็นต์การจำแนกถูก พิจารณาเฉพาะกรณีคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิสเป็นหลัก และตามด้วยการพิจารณาเปอร์เซ็นต์การจำแนกโดยรวม กรณีที่ทำการศึกษามี 3 กรณีคือ (1) กรณีที่ 1 การจำแนกเป็น 2 กลุ่มคือกลุ่มคนไข้ในโครงการที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กับกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ (2) กรณีที่ 2 จำแนกข้อมูลออกเป็น 2 กลุ่มคือกลุ่มคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กับกลุ่มที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis (3) กรณีที่ 3 การจำแนกเป็น 3 กลุ่มคือกลุ่มคนไข้ที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคเลปโตสไปโรซิส กลุ่มที่ได้รับการยืนยันการวินิจฉัยว่าเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ เมื่อสร้างตัวแบบและคำนวณผลการจัดข้อมูลเข้ากลุ่มถูกต้องแล้วทดลองนำตัวแบบที่ได้มาจัดข้อมูลชุดใหม่เข้ากลุ่ม เพื่อทดสอบความสามารถในการนำตัวแบบที่ได้ไปใช้งานจริง

ข้อมูลที่นำมาสร้างตัวแบบสำหรับงานวิจัยนี้ เป็นข้อมูลคนไข้ที่เข้าร่วมโครงการวิจัย “การศึกษาเปรียบเทียบประสิทธิภาพของยาเซฟโทรฟอกซิม และ คีอซีซัยคลิน ในการรักษาผู้ป่วยด้วยโรคเลปโตสไปโรซิสที่มีอาการรุนแรง” ซึ่งเป็นโครงการที่ ศ. พญ. ยุพิน ศุภทรมงคล สาขาโรคติดเชื้อและอายุรศาสตร์เขตร้อน ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ ศิริราชพยาบาล ทำการศึกษาร่วมกับแพทย์และคณะผู้วิจัยโรงพยาบาลอื่นๆ 6 โรงพยาบาล ตัดข้อมูลสูญหายทิ้งแล้วจำนวน 559 , 315 และ 559 รายสำหรับการจำแนกกรณีที่ 1,2 และ 3 ตามลำดับ สำหรับข้อมูลชุดใหม่ที่นำมาทดสอบตัวแบบเป็นข้อมูลคนไข้ในโครงการวิจัยเดียวกัน เก็บข้อมูลในโรงพยาบาลและช่วงเวลาเดียวกัน จำนวน 149 , 105 และ 149 รายสำหรับการจำแนกกรณีที่ 1 , 2 และ 3 ตามลำดับ

สรุปผลการวิจัย

ผลการวิเคราะห์สามารถสรุปได้ดังนี้

1. การเปรียบเทียบตัวแบบที่ใช้ในการจำแนก ผลการเปรียบเทียบแสดงดังตารางที่ 56 ผลสรุปสำหรับแต่ละกรณีแสดงเป็นลำดับดังนี้

ตารางที่ 58 เปอร์เซนต์จัดเข้ากลุ่มถูกต้องโดยวิธีการวิเคราะห์การจำแนกและการวิเคราะห์การถดถอยโลจิสติกสำหรับทั้ง 3 กรณี

Cases	Technique	Percentage Correct			Overall Percentage Correct	APER
		Leptospirosis	Rickettsiosis	Others		
กรณี 1	Discriminant	76.10		73.90	74.60	0.254
	Logistic Regression	79.9		68.5	72.30	0.277
	WHO Criterion	3.80		98.10	67.06	0.329
กรณี 2	Discriminant	81.9	74.6		79.0	0.210
	Logistic Regression	86.5	67.2		79.0	0.210
กรณี 3	Discriminant	70.1	53.5	51.0	57.8	0.422
	Logistic Regression	63.0	33.3	73.9	62.1	0.379

1.1 กรณีที่ 1 จำแนกเป็น 2 กลุ่มคือ กลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ ผลการจัดข้อมูลเข้ากลุ่มโดยใช้ตัวแบบที่สร้างและตัวแบบอิงเกณฑ์องค์การอนามัยโลกแสดงดังตารางที่ 58

สรุปสำหรับกรณีที่ 1

- พิจารณาแยกเฉพาะกรณีที่คนไข้ที่เป็นโรคเลปโตสไปโรสิส พบว่า ตัวแบบที่ได้จากการวิเคราะห์การถดถอยโลจิสติกมีประสิทธิภาพสูงกว่า การวิเคราะห์การจำแนก
- พิจารณาจากเปอร์เซ็นต์การจำแนกโดยรวม การวิเคราะห์การจำแนกจะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่า การวิเคราะห์การถดถอยโลจิสติก
- พิจารณาจากอัตราผิดพลาดที่เห็นชัดเจน (APER) การวิเคราะห์การจำแนกจะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่า การวิเคราะห์การถดถอยโลจิสติก
- ตัวแบบที่สร้างขึ้นมีประสิทธิภาพดีกว่าตัวแบบที่อิงเกณฑ์องค์การอนามัยโลกเมื่อพิจารณาจากทั้งสามเกณฑ์ข้างต้น

1.2 กรณีที่ 2 จำแนกเป็น 2 กลุ่มคือ กลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis ผลการจัดข้อมูลเข้ากลุ่มโดยใช้ตัวแบบที่สร้าง แสดงดังตารางที่ 58

สรุปสำหรับกรณีที่ 2

- พิจารณาแยกเฉพาะกรณีที่คนไข้ที่เป็นโรคเลปโตสไปโรสิส พบว่า ตัวแบบที่ได้จากการวิเคราะห์การถดถอยโลจิสติกมีประสิทธิภาพสูงกว่า
- พิจารณาจากเปอร์เซ็นต์การจำแนกถูกรวม การวิเคราะห์การจำแนกและการวิเคราะห์การถดถอยโลจิสติกให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพเท่ากัน
- พิจารณาจากอัตราผิดพลาดที่เห็นชัดเจน (APER) การวิเคราะห์การจำแนกและการวิเคราะห์การถดถอยโลจิสติกให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพเท่ากัน

1.3 กรณีที่ 3 จำแนกเป็น 3 กลุ่มคือ กลุ่มที่วินิจฉัยเป็นโรคเลปโตสไปโรสิสและกลุ่มที่วินิจฉัยเป็นโรคในกลุ่ม Rickettsiosis และกลุ่มที่วินิจฉัยเป็นโรคอื่นๆ ผลการจัดข้อมูลเข้ากลุ่มโดยใช้ตัวแบบที่สร้าง แสดงดังตารางที่ 58

สรุปสำหรับกรณีที่ 3

- พิจารณาแยกเฉพาะกรณีที่คนไข้ที่เป็นโรคเลปโตสไปโรสิส พบว่า ตัวแบบที่ได้จากการวิเคราะห์การจำแนกมีประสิทธิภาพสูงกว่า
- พิจารณาจากเปอร์เซ็นต์การจำแนกถูกรวม การวิเคราะห์การถดถอยโลจิสติกจะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่า
- พิจารณาจากอัตราผิดพลาดที่เห็นชัดเจน (APER) การวิเคราะห์การถดถอยโลจิสติกจะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่า

2. การนำตัวแบบไปตรวจสอบกับข้อมูลชุดใหม่

หลังจากได้ตัวแบบที่สร้างขึ้นโดยใช้ 2 เทคนิคดังที่กล่าวมาข้างต้นแล้ว นำตัวแบบที่ได้มาทำการตรวจสอบประสิทธิภาพการใช้งานได้จริง โดยนำมาจัดข้อมูลชุดใหม่ซึ่งเป็นผู้ป่วยที่ทำการวิจัยและเก็บข้อมูลในโรงพยาบาลและช่วงเวลาเดียวกัน ที่แยกออกมาเพื่อใช้ในการทดสอบโดยเฉพาะ เข้ากลุ่มแล้วพิจารณาเปอร์เซ็นต์การจำแนกถูก ผลการเปรียบเทียบแสดงดังตารางที่ 59 ผลแยกเป็นแต่ละกรณีแสดงเป็นลำดับดังนี้

ตารางที่ 59 เปอร์เซนต์จัดซื้อข้อมูลชุดใหม่เข้ากลุ่มถูกต้องโดยวิธีการวิเคราะห์การจำแนกและการวิเคราะห์การถดถอยโลจิสติกสำหรับทั้ง 3 กรณี

Cases	Technique	Percentage Correct			Overall Percentage Correct	APER
		Leptospirosis	Rickettsiosis	Others		
กรณี 1	Discriminant	70.00		45.60	57.60	0.424
	Logistic Regression	72.90		40.50	55.72	0.443
	WHO Criterion	4.29		97.47	53.69	0.463
กรณี 2	Discriminant	75.70	48.60		66.67	0.333
	Logistic Regression	84.30	40.00		69.53	0.305
กรณี 3	Discriminant	67.10	31.40	29.50	47.61	0.524
	Logistic Regression	15.70	54.30	31.80	29.52	0.705

2.1 กรณีที่ 1

- พิจารณาเฉพาะกรณีที่คนไข้ได้รับการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส พบว่าการวิเคราะห์การถดถอยโลจิสติกให้เปอร์เซนต์การจำแนกถูกต้องที่สูงกว่า
- พิจารณาจากเปอร์เซนต์การจำแนกโดยรวม พบว่าเทคนิคการวิเคราะห์การจำแนกสูงกว่าการวิเคราะห์การถดถอยโลจิสติกเล็กน้อย
- พิจารณาจากอัตราผิดที่เห็นชัดเจน การวิเคราะห์การจำแนกจะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่า
- ตัวแบบที่สร้างขึ้นมีประสิทธิภาพดีกว่าตัวแบบที่อิงเกณฑ์การอนามัยโลกเมื่อพิจารณาทั้งสามเกณฑ์ข้างต้น

2.2 กรณีที่ 2

- พิจารณาเฉพาะกรณีที่คนไข้ได้รับการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส พบว่าเทคนิคการวิเคราะห์การถดถอยโลจิสติกสูงกว่าการวิเคราะห์การจำแนก
- พิจารณาจากเปอร์เซนต์การจำแนกโดยรวม พบว่าเทคนิคการวิเคราะห์การถดถอยโลจิสติกสูงกว่าการวิเคราะห์การจำแนก
- พิจารณาจากอัตราผิดที่เห็นชัดเจน การวิเคราะห์การถดถอยโลจิสติกจะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่า

2.3 กรณีที่ 3

- พิจารณาเฉพาะกรณีที่คนไข้ได้รับการวินิจฉัยว่าเป็นโรคเลปโตสไปโรสิส พบว่าเทคนิคการวิเคราะห์การจำแนกสูงกว่าการวิเคราะห์การถดถอยโลจิสติก
- พิจารณาจากเปอร์เซ็นต์การจำแนกโดยรวม พบว่าเทคนิคการวิเคราะห์การจำแนกสูงกว่าการวิเคราะห์การถดถอยโลจิสติก
- พิจารณาจากอัตราผิดที่เห็นชัดเจน การวิเคราะห์การจำแนกจะให้ตัวแบบสำหรับการจำแนกที่มีประสิทธิภาพสูงกว่า

เมื่อพิจารณาเปอร์เซ็นต์การจำแนกกลุ่มถูกต้อง กรณีคนไข้เป็นโรคเลปโตสไปโรสิสเป็นหลัก พิจารณาเปอร์เซ็นต์การจำแนกกลุ่มถูกต้องรวมเป็นอันดับที่สอง และพิจารณาเปอร์เซ็นต์การจำแนกกลุ่มถูกต้องสำหรับกลุ่มอื่นๆ เป็นอันดับรองลงมา พบว่าตัวแบบที่สมควรนำไปใช้งานจริงมีเพียงตัวแบบสำหรับกรณีที่ 1 และกรณีที่ 2 เท่านั้น เนื่องจากมีเปอร์เซ็นต์การจำแนกกลุ่มถูกต้องค่อนข้างสูง ส่วนตัวแบบสำหรับการจำแนกกรณีที่ 3 ต้องทำการปรับก่อนจึงจะสามารถนำไปใช้ได้

และเมื่อพิจารณาตัวแบบที่สร้างขึ้นโดยรวม โดยพิจารณาจากเปอร์เซ็นต์การจำแนกกลุ่มกรณีที่คนไข้เป็นโรคเลปโตสไปโรสิสเป็นหลักแล้ว พบว่าตัวแบบที่สร้างโดยใช้เทคนิคการวิเคราะห์การถดถอยโลจิสติกน่าจะเป็นเทคนิคที่เหมาะสม และตรงกับความต้องการใช้งานสำหรับการวิเคราะห์ข้อมูลในงานวิจัยนี้มากกว่าเทคนิคการวิเคราะห์การจำแนก

อภิปรายผล

พิจารณาจากตัวแปรที่ถูกนำเข้าสู่สมการของทั้ง 3 กรณี พบว่า มีตัวแปรชุดหนึ่งซึ่งมักจะถูกนำเข้าสู่สมการในการสร้างตัวแบบการจำแนกสำหรับ 3 กรณีนี้ ประกอบด้วยตัวแปรทั้งหมด 5 ตัวแปร ได้แก่ เยื่อบุตาอักเสบ (X_{15}), กดเจ็บกล้ามเนื้อ (X_{17}), PMN>75% (X_{22}), Total Bilirubin>3 mg/dl (X_{25}) และ Abnormal SGOT/SGPT/Alkaline phosphatase (X_{26}) ซึ่งตัวแปรเหล่านี้เป็นอาการที่สามารถจำแนกกลุ่มคนไข้ที่เป็นโรคเลปโตสไปโรสิส ได้อย่างมีประสิทธิภาพ สำหรับทั้ง 3 กรณีที่ทำการศึกษา

พิจารณาเฉพาะเทคนิคการวิเคราะห์การถดถอยโลจิสติก ซึ่งมีผลการจัดข้อมูลเข้ากลุ่มถูกต้องดีกว่าการวิเคราะห์การจำแนก จากการนำส่วนเหลือไปพลอตเทียบกับค่าทำนายความน่าจะเป็นเพื่อตรวจสอบข้อสมมติ พบว่าส่วนเหลือยังไม่กระจายอย่างเป็นอิสระต่อกันหรือยังมีรูปแบบใน

ทุกกรณี ซึ่งเป็นลักษณะที่ไม่ตรงตามข้อสมมติของเทคนิคการวิเคราะห์การถดถอยโลจิสติก อาจเป็นสาเหตุสำคัญที่ทำให้ผลการจัดเข้ากลุ่มถูกต้องไม่ดีเท่าที่ควร

ข้อเสนอแนะจากงานวิจัย

ผลของการวิจัย พบว่าตัวแบบโลจิสติกมีประสิทธิภาพสูงที่สุดในการพยากรณ์การเป็นหรือไม่เป็นโรคเลปโตสไปโรสิส โดยสูงกว่าตัวแบบการวิเคราะห์การจำแนกและตัวแบบอิงเกณฑ์องค์การอนามัยโลก อย่างไรก็ตามจากผลของการวิจัยนี้ ยังพบว่ามีลักษณะที่ไม่ตรงตามข้อสมมติของเทคนิคการวิเคราะห์การถดถอยโลจิสติกในทุกกรณี ดังนั้นอาจสามารถปรับตัวแบบโลจิสติกโดยใช้ทางเลือกใกล้เคียง เช่น ตัวแบบโพรบิท ตัวแบบคอมพลีเมนต์ลอจิสติก หรือตัวแบบอื่นๆ ซึ่งอาจเพิ่มประสิทธิภาพของตัวแบบให้สูงขึ้นได้

ข้อมูลที่น่ามาใช้ในการวิเคราะห์นี้ใช้ข้อมูลทุติยภูมิ ซึ่งโครงการวิจัยที่มาของข้อมูลมุ่งเน้นที่การศึกษาเปรียบเทียบประสิทธิผลของยาที่คนไข้ในโครงการได้รับ ดังนั้นข้อมูลเกี่ยวกับอาการทางคลินิก การตรวจร่างกายแรกรับและผลตรวจทางห้องปฏิบัติการแรกรับจึงถูกนำไปใช้แค่ในภาพรวมเท่านั้น เนื่องจากวัตถุประสงค์ของงานวิจัยแตกต่างกันทำให้รายละเอียดบางอย่างอาจไม่ครบถ้วน ดังนั้นหากเป็นการเก็บข้อมูลเพื่อทำการสร้างตัวแบบในการจำแนกกลุ่ม โดยเฉพาะอาจให้ผลที่ดีขึ้น

บรรณานุกรม

- กราบแก้ว โสภาราชภูริ. “การพยากรณ์ผลโดยวิธีการวิเคราะห์การจำแนกแคนนอนัลกับการวิเคราะห์ถดถอยโลจิสติกสำหรับเหตุการณ์ที่พบได้น้อย.” วิทยานิพนธ์วิทยาศาสตร์ มหาวิทยาลัย สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร, 2544.
- กัลยา วานิชย์บัญชา. การวิเคราะห์สถิติขั้นสูงด้วย SPSS. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร : โรงพิมพ์ บริษัทธรรมสาร จำกัด, 2546.
- ชาญชัยณรงค์ ทรงศาศรีและคณะ. “การทำนายผู้ป่วยโรคเลปโตสไปโรสิสด้วยผลการตรวจทางห้องปฏิบัติการและลักษณะทางคลินิก.” วารสารวิชาการสาธารณสุข (กันยายน – ตุลาคม 2546) : 678-688.
- เดือนเพ็ญ สนโต. “การเปรียบเทียบความสามารถในการจำแนกกลุ่มได้ถูกต้องระหว่างวิธีการวิเคราะห์จำแนกกลุ่มกับวิธีการวิเคราะห์การถดถอยโลจิสติก เมื่อใช้กลุ่มตัวอย่างขนาดต่างๆ.” วิทยานิพนธ์ศึกษาศาสตร์มหาบัณฑิต สาขาวิชาวิจัยและสถิติการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่, 2545.
- ปราณี นิลกรณ์. “การจำแนกกลุ่มและการจัดเข้ากลุ่ม”. เอกสารประกอบการเรียนการสอนภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร, 2547.
- วีรานันท์ พงศาภักดี. การวิเคราะห์ข้อมูลเชิงกลุ่ม : ทฤษฎีและการประยุกต์ (กับ GLIM และ SPSS/FW). พิมพ์ครั้งที่ 2. นครปฐม : โรงพิมพ์มหาวิทยาลัยศิลปากร , 2541.
- สำนักงานโครงการควบคุมโรคเลปโตสไปโรสิส . คู่มือวิชาการ โรคเลปโตสไปโรสิส . กรุงเทพฯ: กรมควบคุมโรคติดต่อ กระทรวงสาธารณสุข, 2544.
- _____ . คู่มือการป้องกันและควบคุมโรคเลปโตสไปโรสิส สู่การมีส่วนร่วมขององค์การบริหารส่วนตำบล .กรุงเทพฯ: กรมควบคุมโรคติดต่อ กระทรวงสาธารณสุข , 2544.
- อุไรวรรณ อมรมนิมิตร. “เทคนิคการวิเคราะห์ความเสี่ยง : ศึกษาเปรียบเทียบระหว่าง Logistic Regression Analysis และ Discriminant Analysis.” วารสารสถิติประยุกต์ (สิงหาคม 2546) : 60-66.
- Barbara, G., and Linda S. Using multivariate statistics. New York: Allyn and Bacon , 2001.
- Hand, D.J. Discrimination and Classification. New York: John Wiley, 1981.
- Huberty, C.J. Applied Discriminant Analysis. New York: John Wiley, 1994.

Johnson, R.A. and D.W. Wichern W. Applied multivariate statistical analysis.4th . New Jersey : Prentice Hall International .Inc.,1988.

Klecka, W.R. Discriminant Analysis. Newbury Park : Sage Publications, Inc., 1980. อ้างถึงใน กราบแก้ว โสภารายณ์. “การพยากรณ์ผลโดยวิธีการวิเคราะห์การจำแนกแคนนอน นิคัลกับการวิเคราะห์ถดถอยโลจิสติกสำหรับเหตุการณ์ที่พบได้น้อย.” วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัย ศิลปากร, 2544.

Kleinbaum, David G. Logistic Regression : A Self –Learning Text. New York : Springer-Verlag Inc., 1994.

Morrison, F. Multivariate statistical methods . New York : McGraw-Hill Publishing Company., 1990.

Rencher , Alvin C. Methods of multivariate analysis. New York : John Wiley & Sons, Inc., 1995.

Suputtamongkol Y., et al. “An Open, Randomized, Controlled Trial of Penicillin, Doxycycline, and Cefotaxime for Patients with Severe Leptospirosis.” Clinical Infectious Diseases (39) 2004 :1417-24.

ประวัติผู้วิจัย

ชื่อ-สกุล
ที่อยู่

นางสาวพนิดา คองแจ่ม
44 ซอยไผ่เตย1 ถนนไผ่เตย ตำบลห้วยจรเข้ม
อำเภอเมือง จังหวัดนครปฐม 73000

ประวัติการศึกษา

พ.ศ. 2544

สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต(สถิติ)

พ.ศ. 2545

มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์
ศึกษาระดับปริญญาโท สาขาสถิติประยุกต์
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

พ.ศ. 2549

สำเร็จการศึกษาปริญญาวิทยาศาสตรมหาบัณฑิต
(สถิติประยุกต์)

มหาวิทยาลัยศิลปากร สงวนลิขสิทธิ์

ประวัติการทำงาน

พ.ศ. 2547 – พ.ศ. 2548

ผู้ช่วยนักวิจัย สาขาโรคติดเชื้อและอายุรศาสตร์เขตร้อน
ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล
มหาวิทยาลัยมหิดล

พ.ศ. 2548 – ปัจจุบัน

ผู้ช่วยนักวิจัยอิสระ