การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2
กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์

A COMPARATIVE EFFECTIVENESS OF CLASSIFICATION METHODS OF COMPLICATED KIDNEY DISEASE IN PATIENTS WITH TYPE II DIABETES CASE STUDY : A HOSPITAL IN BURIRAM PROVINCE

ทองสา บุตรงาม

THONGSA BUTNGAM

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2562

KMITL-2019-SC-M-050-076

การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์

A COMPARATIVE EFFECTIVENESS OF CLASSIFICATION METHODS OF COMPLICATED KIDNEY DISEASE IN PATIENTS WITH TYPE II DIABETES CASE STUDY : A HOSPITAL IN BURIRAM PROVINCE

ทองสา บุตรงาม

THONGSA BUTNGAM

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

KMITL-2019-SC-M-050-076

A COMPARATIVE EFFECTIVENESS OF CLASSIFICATION METHODS OF COMPLICATED KIDNEY DISEASE IN PATIENTS WITH TYPE II DIABETES CASE STUDY : A HOSPITAL IN BURIRAM PROVINCE

THONGSA BUTNGAM

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER IN APPLIED STATISTICS

DEPARTMENT OF STATISTICS FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2019
KMITL-2019-SC-M-050-076

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารนี้ทุกครั้งที่มีการนำไปใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2019
FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

หัวข้อวิทยานิพนธ์ การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์
ชื่อนักศึกษา ทองสา บุตรงาม
รหัสประจำตัว 57605098
ปริญญา วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)
ภาควิชา สถิติ
พ.ศ. 2562
อาจารย์ที่ปรึกษาวิทยานิพนธ์ ดร.บุญญสิทธิ์ วรจันทร์

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ด้วยวิธีการวิเคราะห์จำแนกประเภท การวิเคราะห์การถดถอยลอจิสติก และต้นไม้ตัดสินใจ โดยใช้ข้อมูลทุติยภูมิของผู้ป่วยที่เข้ามารับการรักษาในโรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ ตั้งแต่เดือนตุลาคม พ.ศ.2559 – กันยายน พ.ศ.2560 จำนวน 320 ราย จากการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มทั้ง 3 วิธี โดยใช้ค่าความถูกต้องและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ผลการวิจัยพบว่า วิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ที่มีประสิทธิภาพสูงที่สุด คือ การวิเคราะห์การถดถอยลอจิสติก ซึ่งให้ค่าความถูกต้องมากที่สุดและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด

คำสำคัญ : การวิเคราะห์จำแนกประเภท การวิเคราะห์การถดถอยลอจิสติก ต้นไม้ตัดสินใจ ภาวะแทรกซ้อนทางไต โรคเบาหวานชนิดที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	A Comparative Effectiveness of Classification Methods of Complicated Kidney Disease in Patients with Type II Diabetes Case Study : A Hospital in Buriram Province
Student Name	Thongsa Butngam
Student ID	57605098
Degree	Master of Science (Applied Statistics)
Department	Statistics
Year	2019
Thesis Advisor	Dr.Boonyasit Warachan

Abstract

This research aims to compare the effectiveness of classification methods of complicated kidney disease type II patients. The classification procedures are the discriminant analysis, logistic regression analysis and decision tree. The secondary data from a hospital in Buriram province consist of 320 cases during October 2016 to September 2017 will be analyzed. The results showed that the logistic regression analysis is the most effective classification with the highest accuracy and the lowest root mean square error.

Keywords : Classification analysis, Logistic regression analysis, Decision tree, Complication Kidney disease, Type II diabetes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี ด้วยความช่วยเหลือของอาจารย์ ดร.บุญญสิทธิ์ วรจันทร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่าง ๆ อันเป็นประโยชน์อย่างยิ่งในการทำวิจัย อีกทั้งยังช่วยแก้ปัญหาต่าง ๆ ที่เกิดขึ้นระหว่างการดำเนินงานวิจัยอีกด้วย ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.จุฑาภรณ์ สินสมบูรณ์ทอง ผู้ช่วยศาสตราจารย์ ดร.สิทธิชัย เจริญเศรษฐศิลป์ รองศาสตราจารย์สายชล สินสมบูรณ์ทอง และอาจารย์ภาควิชาสถิติทุกท่านที่ให้คำปรึกษาและคำแนะนำงานวิจัยฉบับนี้ทั้งหมด ตลอดจนข้อมูลทุกอย่างที่เป็นประโยชน์ในการทำวิจัย ขอขอบคุณ คุณอัจฉรา แผ้วบาง และเจ้าหน้าที่ภาควิชาสถิติทุกท่านสำหรับข้อแนะนำและความช่วยเหลือในทุก ๆ ด้านในการทำวิจัย นอกจากนี้ขอขอบคุณเพื่อน ๆ ทุกคนที่เป็นกำลังใจ และให้ความช่วยเหลือในการทำวิทยานิพนธ์ฉบับนี้

สุดท้ายนี้ ผู้วิจัยขอขอบพระคุณบิดามารดา และครอบครัว ซึ่งเปิดโอกาสให้ได้รับการศึกษาเล่าเรียน ตลอดจนคอยช่วยเหลือและให้กำลังใจผู้วิจัยเสมอมาจนสำเร็จการศึกษา

นางสาวทองสา บุตรงาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	จ
สารบัญรูป	ฉ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย	2
1.4 นิยามศัพท์เฉพาะ	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 โรคไตจากเบาหวาน	4
2.2 การวิเคราะห์จำแนกประเภท	9
2.3 การวิเคราะห์การถดถอยลอจิสติก	22
2.4 ต้นไม้ตัดสินใจ	30
2.5 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม	35
2.6 งานวิจัยที่เกี่ยวข้อง	36
บทที่ 3 วิธีการดำเนินงานวิจัย	42
3.1 การเก็บรวบรวมข้อมูล	42
3.2 เครื่องมือที่ใช้ในการวิจัย	43
3.3 ขั้นตอนในการดำเนินการวิจัย	43
3.4 วิธีการวิเคราะห์ข้อมูล	44
บทที่ 4 ผลการวิเคราะห์ข้อมูล	47
4.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีการวิเคราะห์จำแนกประเภท	47
4.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีการวิเคราะห์ถดถอยลอจิสติก	54
4.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ	65
4.1 ผลการเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์	68
บทที่ 5 สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ	69
5.1 สรุปผลการวิจัย	69
5.2 อภิปรายผล	70
5.3 ข้อเสนอแนะ	71

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม 72
ภาคผนวก ก 75
ภาคผนวก ข 90
ประวัติผู้เขียน 108

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 ตัวอย่างข้อมูลรายได้ครอบครัวและขนาดพื้นที่	19
2.2 ตัวอย่างข้อมูลรายได้และอายุ	29
2.3 ผลการประมาณการศึกษาการซื้อเครื่องสำอางยี่ห้อ A กับรายได้ของลูกค้า	29
2.4 ตัวอย่างชุดข้อมูลการเล่นเทนนิส	31
2.5 เมทริกซ์ความสับสน	35
3.1 เมทริกซ์ความสับสน	45
4.1 ผลการตรวจสอบการแจกแจงปรกติของตัวแปรอิสระหลายตัวแปร โดยใช้ตัวสถิติทดสอบ Kolmogorov-Smirnov	48
4.2 ผลการทดสอบด้วยวิธี Box's M	50
4.3 ค่าสัมประสิทธิ์การวิเคราะห์จำแจกประเภทของตัวแปรอิสระโดยใช้วิธีของ Fisher's	51
4.4 ผลการทดสอบสมการจำแนกกลุ่มข้อมูลโดยใช้วิธี Hotelling's T^2	52
4.5 ผลการประเมินความเหมาะสมของสมการการวิเคราะห์จำแนกประเภท	53
4.6 ค่า Tolerance และ Variance Inflation Factor ของตัวแปรอิสระทั้ง 23 ตัว	57
4.7 ค่าความสัมพันธ์ของตัวแปรอิสระทั้ง 5 ตัว โดยใช้วิธีของเพียร์สัน	58
4.8 ค่าความสัมพันธ์ของตัวแปรอิสระที่มีความสัมพันธ์กันมากทั้ง 5 ตัว กับตัวแปรตามโดยใช้การวิเคราะห์ความแปรปรวนแบบทางเดียว	58
4.9 ค่า Tolerance และ Variance Inflation Factor ของตัวแปรอิสระทั้ง 20 ตัว	59
4.10 ค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระทั้ง 20 ตัว	60
4.11 ผลทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระโดยใช้การทดสอบวอดล์	62
4.12 ผลการทดสอบค่าสัมประสิทธิ์การถดถอยลอจิสติก	63
4.13 ผลการตรวจสอบความเหมาะสมของสมการการวิเคราะห์การถดถอยลอจิสติก	63
4.14 ผลการประเมินความเหมาะสมของสมการการวิเคราะห์การถดถอยลอจิสติก	64
4.15 ผลการประเมินความเหมาะสมของสมการต้นไม้ตัดสินใจที่ใช้ในการพยากรณ์	65
4.16 ผลการเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์	68

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 ตัวอย่างต้นไม้ตัดสินใจจากชุดข้อมูลการเล่นเทนนิส	32
4.1 การกระจายของค่าส่วนเหลือมาตรฐาน	54
4.2 การกระจายของค่า Leverage	55
4.3 การกระจายของค่า Cook’s Distance	56
4.4 ต้นไม้ตัดสินใจเพื่อใช้ในการจำแนกกลุ่ม	66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันโรคเบาหวานเป็นหนึ่งในโรคเรื้อรังที่บั่นทอนคุณภาพชีวิตและคร่าชีวิตมนุษย์มากที่สุดโรคหนึ่ง และมีแนวโน้มอัตราการเกิดโรคเพิ่มมากขึ้นในอนาคต จากรายงานของสหพันธ์เบาหวานโลกปี พ.ศ. 2558 พบว่า มีผู้ป่วยเบาหวานทั่วโลกจำนวน 415 ล้านคน ณ ปีปัจจุบันสามารถพบผู้ป่วยเป็นโรคเบาหวาน 1 คนจากจำนวนผู้ใหญ่ 11 คน และทำนายว่า อีก 25 ปีข้างหน้าจะพบผู้ป่วยเป็นโรคเหวานเพิ่มขึ้นเป็น 1 คนจาก 10 คน และทั่วโลกจะมีผู้ป่วยโรคเบาหวานจากจำนวน 415 ล้านคนเป็น 642 ล้านคนในปี พ.ศ. 2583 (สมาคมโรคเบาหวานแห่งประเทศไทย, 2558) สำหรับประเทศไทยพบผู้ป่วยโรคเบาหวานจำนวน 670,664 คน หรือคิดเป็นอัตรา 1,033 คน ต่อประชากรหนึ่งแสนคน (สำนักโรคไม่ติดต่อ กรมควบคุมโรค, 2557)

สถานการณ์โรคเบาหวานในจังหวัดบุรีรัมย์ ในปี 2550 – 2558 พบว่า แนวโน้มของผู้ป่วยโรคเบาหวานสูงเพิ่มขึ้นทุกปี โดยพบว่าจำนวนผู้ป่วยโรคเบาหวาน ในปี 2558 มีอัตราความชุกร้อยละ 2.90 และในปี 2559 จังหวัดบุรีรัมย์มีผู้ป่วยโรคเบาหวาน จำนวน 54,341 คน หรือคิดเป็นอัตราป่วย 3,453 คน ต่อประชากรหนึ่งแสนคน (กลุ่มงานควบคุมโรค สำนักงานสาธารณสุขจังหวัดบุรีรัมย์, 2559) ซึ่งมีแนวโน้มที่จะมีจำนวนผู้ป่วยสูงขึ้นในปีต่อไป อันเนื่องมาจากโรคเบาหวานเป็นโรคที่ไม่สามารถรักษาให้หายขาดได้และยังสามารถถ่ายทอดทางพันธุกรรมได้ ซึ่งผู้ป่วยยังมีปัญหาการควบคุมระดับน้ำตาลได้น้อย การปรับเปลี่ยนพฤติกรรมค่อนข้างยาก จึงทำให้มีภาวะแทรกซ้อนทุกระบบโดยเฉพาะภาวะแทรกซ้อนทางไต ซึ่งเป็นภาวะแทรกซ้อนที่พบได้มากที่สุดในผู้ป่วยโรคเบาหวาน และถ้าหากผู้ป่วยไม่มีการดูแลอย่างถูกต้องเหมาะสม โอกาสเสี่ยงต่อการเกิดภาวะแทรกซ้อนทางไต ก็จะเพิ่มมากขึ้น ถ้าเกิดภาวะแทรกซ้อนทางไตแล้วจะส่งผลต่อระบบการทำงานต่างๆ ในร่างกาย ซึ่งก่อให้เกิดโรคแทรกซ้อนอื่น ๆ ตามมา ตลอดจนส่งผลกระทบต่าง ๆ หลายประการตามมา ทั้งในด้านตัวผู้ป่วยเอง ครอบครัวและสังคม

โรคไตจากเบาหวานเป็นปัญหาสาธารณสุขที่สำคัญของประเทศ ข้อมูลจากการรายงานผลการลงทะเบียนการรักษาด้วยการบำบัดทดแทนไตของสมาคมโรคไตแห่งประเทศไทย (Thailand Renal Replacement Therapy Registry Report) ในปี 2556 จากหน่วยไตเทียมในประเทศ พบว่าสาเหตุของโรคไตวายเรื้อรังระยะสุดท้ายของผู้ป่วยที่เข้ารับการบำบัดทดแทนไตจากสถานพยาบาลที่ให้บริการการฟอกเลือดด้วยเครื่องไตเทียมเกิดจากโรคเบาหวานมากที่สุด คิดเป็นร้อยละ 37.5 อันดับสองรองลงมา คือ โรคความดันโลหิตสูง คิดเป็นร้อยละ 26.8 ซึ่งผู้ป่วยโรคเบาหวานชนิดที่ 2 ร้อยละ 20 - 40 มีโอกาสเกิดภาวะโรคไตเรื้อรัง (Diabetic Kidney Disease, DKD) และมักมีการดำเนินโรคไปสู่ภาวะไตเรื้อรัง (Chronic Kidney Disease, CKD) และเป็นไตวายเรื้อรังระยะสุดท้าย (End Stage Renal Disease, ESRD) ในที่สุด เมื่อความชุกของโรคเบาหวานเพิ่มขึ้นโรคไตจากเบาหวานจะยิ่งเพิ่มมากขึ้นตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากความเป็นมาและความสำคัญของปัญหา ผู้วิจัยจึงได้มีความสนใจศึกษาการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ โดยใช้วิธีการจำแนกกลุ่ม 3 วิธี คือ การวิเคราะห์จำแนกประเภท (Discriminant Analysis) การวิเคราะห์การถดถอยลอจิสติก (Logistic Regression Analysis) และต้นไม้ตัดสินใจ (Decision Tree) ในการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ว่าผู้ป่วยจะอยู่ในกลุ่มเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไต ซึ่งงานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยใช้วิธีการจำแนกกลุ่มทั้ง 3 วิธี ว่าวิธีการจำแนกกลุ่มวิธีใดมีประสิทธิภาพสูงที่สุดและเหมาะสมสำหรับการจำแนกกลุ่มผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ว่าจะเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไต และสามารถนำผลการวิจัยที่ได้ไปใช้เพื่อสนับสนุนในทางการแพทย์เกี่ยวกับการวินิจฉัยการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ได้ เพื่อให้ทราบผลการวินิจฉัยที่รวดเร็วยิ่งขึ้น สามารถรักษาได้อย่างทันท่วงทีและเพื่อเป็นแนวทางในการป้องกันการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ได้

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยใช้วิธีการวิเคราะห์จำแนกประเภท การวิเคราะห์การถดถอยลอจิสติก และต้นไม้ตัดสินใจ

1.2.2 เพื่อสามารถนำวิธีการจำแนกกลุ่มที่มีประสิทธิภาพสูงที่สุดไปใช้ในการจำแนกกกลุ่มผู้ป่วยโรคเบาหวาน ชนิดที่ 2 รายใหม่ ในกรณีที่ผู้ป่วยยังไม่ได้รับการตรวจวินิจฉัยภาวะแทรกซ้อนทางไตได้

1.3 ขอบเขตของการวิจัย

การศึกษางานวิจัยในครั้งนี้ ข้อมูลที่ผู้วิจัยใช้ในการวิเคราะห์เป็นข้อมูลทุติยภูมิ (Secondary Data) ของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ที่เข้ามารับการรักษาในโรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ ตั้งแต่วันที่ 1 ตุลาคม พ.ศ.2559 – 31 กันยายน พ.ศ.2560 เป็นระยะเวลา 12 เดือน ในการศึกษางานวิจัยครั้งนี้เป็นการศึกษาการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเหวาน ชนิดที่ 2 ของผู้ป่วยที่มารับการรักษาในโรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์เท่านั้น

สำหรับประชากรที่ใช้ในการศึกษางานวิจัยครั้งนี้ คือ ผู้ป่วยโรคเบาหวานชนิดที่ 2 ที่มารับการรักษาในโรงพยาบาลแห่งหนึ่งในจังบุรีรัมย์ จำนวน 507 ราย ซึ่งเก็บรวบรวมโดยฝ่ายเวชระเบียนของโรงพยาบาล และตัวแปรที่ใช้ในการศึกษางานวิจัยในครั้งนี้ คือ

1) ตัวแปรอิสระ (Independent Variables) คือ ปัจจัยที่คาดว่าจะส่งผลต่อการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยตัวแปรอิสระที่ใช้ในการวิจัยครั้งนี้ ผู้วิจัยได้จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับปัจจัยที่ก่อให้เกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) ตัวแปรตาม (Dependent Variable) คือ การเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2

1.4 นิยามศัพท์เฉพาะ

การจำแนกกลุ่ม (Classification) หมายถึง กระบวนการสร้างโมเดลสำหรับการจัดกลุ่มหรือประเภทของข้อมูลโดยการวิเคราะห์ข้อมูลจากระบบเพื่อเรียนรู้ โดยแต่ละแถวประกอบด้วยข้อมูลเชิงคุณลักษณะที่มีลักษณะเป็นข้อมูลเชิงตัวเลขที่มีค่าต่อเนื่อง

โรคเบาหวานชนิดที่ 2 (Diabetes Mellitus Type II) คือ โรคเบาหวานชนิดไม่พึ่งอินซูลิน (Non - Insulin Dependent Diabetes) หรือโรคเบาหวานชนิดที่ร่างกายผลิตอินซูลินไม่เพียงพอและอินซูลินไม่สามารถนำน้ำตาลกลูโคสเข้าเซลล์ได้อย่างมีประสิทธิภาพ หรือเรียกว่า การดื้ออินซูลิน

ภาวะแทรกซ้อนทางไต (Diabetic Nephropathy) คือ การเกิดโรคไตในผู้ป่วยโรคเบาหวานชนิดที่ 2 และมีค่าการตรวจอัตราการกรองของไต (Glomerular Filtration Rate : GFR) ต่ำกว่า 60 มิลลิลิตรต่อนาทีต่อพื้นที่ผิวร่างกาย 1.73 ตารางเมตร

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ทราบถึงประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยใช้วิธีการวิเคราะห์จำแนกประเภท การวิเคราะห์การถดถอยลอจิสติกและต้นไม้ตัดสินใจ

1.5.2 ในทางการแพทย์สามารถนำวิธีการจำแนกกลุ่มที่มีประสิทธิภาพสูงที่สุด ไปใช้เพื่อเป็นแนวทางในการสนับสนุนการวินิจฉัยการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 รายใหม่ ในกรณีที่ผู้ป่วยยังไม่ได้รับการตรวจวินิจฉัยภาวะแทรกซ้อนทางไตได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ ผู้วิจัยได้กล่าวถึงทฤษฎีที่ใช้ในการศึกษาการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษาผู้ป่วยโรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ และงานวิจัยที่เกี่ยวข้อง ซึ่งมีรายละเอียดดังนี้

2.1 โรคไตจากเบาหวาน (Diabetic Nephropathy)

2.2 การวิเคราะห์จำแนกประเภท (Discriminant Analysis)

2.3 การวิเคราะห์การถดถอยลอจิสติก (Logistic Regression Analysis)

2.4 ต้นไม้ตัดสินใจ (Decision Tree)

2.5 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม

2.6 งานวิจัยที่เกี่ยวข้อง

2.1 โรคไตจากเบาหวาน (Diabetic Nephropathy)

โรคเบาหวาน (Diabetic) เป็นโรคเกี่ยวกับกระบวนการเมแทบอลิซึม (Metabolism) ซึ่งสามารถเกิดเป็นโรคเรื้อรัง สาเหตุเกิดจากตับอ่อนไม่สามารถสร้างฮอร์โมนอินซูลินได้ หรืออาจเกิดจากร่างกายไม่สามารถตอบสนองกับฮอร์โมนอินซูลินได้เหมือนคนปกติ ส่งผลให้ร่างกายของผู้ป่วยเบาหวานมีปริมาณน้ำตาลในกระแสเลือดเพิ่มสูงขึ้น ซึ่งถ้าผู้ป่วยไม่สามารถที่จะควบคุมความผิดปกติได้ จะส่งผลให้เกิดโรคแทรกซ้อนขึ้น เช่น โรคไตที่เกิดจากเบาหวาน (Diabetic Nephropathy) โรคจอประสาทตาเสื่อม (Diabetic Retinopathy) โรคระบบประสาทเสื่อม (Diabetic Neuropathy) และโรคหลอดเลือดหัวใจ (Cardiovascular Disease) เป็นต้น (อัจจิมา กาญจนาภา, 2558) ซึ่งโรคเบาหวานแบ่งออกเป็น 2 ชนิด ได้แก่

1) โรคเบาหวาน ชนิดที่ 1 หรือโรคเบาหวานชนิดพึ่งอินซูลิน (Insulin Dependent Diabetes) หรือโรคเบาหวานในเด็ก ผู้ป่วยที่เป็นโรคเบาหวานชนิดนี้เป็นผู้ที่ร่างกายขาดอินซูลินโดยสิ้นเชิง เนื่องจากตับอ่อนไม่สามารถสร้างอินซูลินได้ อาจเนื่องมาจากเบต้าเซลล์ในตับอ่อนมีน้อยหรือไม่มีเลย พบในผู้ป่วยอายุไม่เกิน 30 ปี ทั้งชายและหญิง อาการของโรคมักเป็นรุนแรง ส่วนใหญ่มีรูปร่างผอม น้ำหนักตัวลดลงอย่างรวดเร็ว ซึ่งพบจำนวนน้อย (บุญทิพย์ สิริธรังศรี, 2539)

2) โรคเบาหวาน ชนิดที่ 2 หรือโรคเบาหวานชนิดไม่พึ่งอินซูลิน (Non - Insulin Dependent Diabetes) มักพบในคนอายุมากกว่า 40 ปีขึ้นไป เพศหญิงเป็นมากกว่าเพศชาย มักพบในคนที่อ้วนมาก นอกจากนี้กรรมพันธุ์ยังมีส่วนเกี่ยวข้องกับการเกิดโรคเบาหวานเป็นอย่างมาก ผู้ที่มีประวัติสมาชิกในครอบครัว โดยเฉพาะญาติสายตรงเป็นโรคเบาหวาน มีแนวโน้มที่จะเป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้ง

โรคเบาหวานชนิดนี้ได้มาก อาการที่เกิดขึ้นมีได้ตั้งแต่ไม่แสดงอาการ แต่ตรวจพบโดยบังเอิญหรือมีอาการแบบค่อยเป็นค่อยไปจนถึงขั้นแสดงอาการรุนแรง ตับอ่อนของผู้ป่วยเบาหวานประเภทนี้ยังสามารถผลิตอินซูลินได้ตามปกติหรืออาจจะน้อยหรืออาจจะมากกว่าปกติได้ แต่อินซูลินที่มีอยู่ออกฤทธิ์ได้ไม่ดีจึงไม่ถึงกับขาดอินซูลินไปโดยสิ้นเชิงเหมือนคนที่เป็นโรคเบาหวาน ชนิดที่ 1 ผู้ป่วยจึงไม่เกิดภาวะกรดคั่งในเลือดจากสารคีโตน โดยผู้ป่วยเบาหวานส่วนใหญ่กว่าร้อยละ 95 พบว่าเป็นโรคเบาหวาน ชนิดที่ 2 (เทพ หิมะทองคำ, 2547)

2.1.1 โรคไตจากเบาหวาน (Diabetic Nephropathy)

โรคไตจากเบาหวานเป็นภาวะแทรกซ้อนเรื้อรังที่สำคัญของผู้ป่วยโรคเบาหวาน พบได้ประมาณร้อยละ 30-50 ของผู้ป่วยโรคเบาหวาน จากการสำรวจในประเทศไทยพบว่ามีผู้ป่วยเบาหวาน ชนิดที่ 2 ที่มีภาวะไมโครอัลบูมินูเรียร้อยละ 18.7-43.5 และภาวะ Overt Nephropathy ร้อยละ 1.6-5.1 นอกจากนี้โรคไตจากเบาหวานยังเป็นสาเหตุอันดับหนึ่งของผู้ป่วยโรคไตเรื้อรังระยะสุดท้าย คือ พบประมาณร้อยละ 30.1 ของผู้ป่วยที่ได้รับการรักษาทดแทนไตในประเทศไทย โดยลักษณะทางคลินิกของโรคไตจากเบาหวานสามารถตรวจพบได้ดังนี้ (อัจจิมา กาญจนาภา, 2558)

1) ภาวะโปรตีนชนิดอัลบูมินรั่วออกมาในปัสสาวะ (Albuminuria) ซึ่งในระยะแรกมีปริมาณเล็กน้อยหรือระยะที่เรียกว่า ไมโครอัลบูมินูเรีย (Micro Albuminuria) คือมี Urine Albumin Excretion Rate (UAE) 30-300 มิลลิกรัมต่อวัน ถ้าทิ้งไว้ไม่ได้รับการรักษาปริมาณโปรตีนชนิดอัลบูมินจะตรวจพบในปัสสาวะมากขึ้น (Macro Albuminuria หรือ Overt Proteinuria) คือมี UAE มากกว่า 300 มิลลิกรัมต่อวัน

2) อาการความดันโลหิตสูง คือ มีความดันโลหิตสูงกว่า 140/90 มิลลิเมตรปรอท เมื่อยังไม่ได้รับการรักษา หรือความดันโลหิตสูงกว่า 130/80 มิลลิเมตรปรอท เมื่อได้รับการรักษาแล้ว

3) การทำงานของไต (Glomerular Filtratio Rate ; GER) คือ อัตราการกรองของไต พบว่าในระยะแรกการทำงานจะปกติหรือสูงกว่าปกติ ต่อมาจะเริ่มเสื่อมและเสื่อมมากขึ้นจนเกิดโรคไตเรื้อรังระยะท้ายในที่สุด

2.1.2 ธรรมชาติการดำเนินโรคไตจากเบาหวาน

การตรวจพบการรั่วของโปรตีนในปัสสาวะของผู้ป่วยโรคเบาหวาน จะเป็นตัวบ่งบอกการเกิดโรคไตจากเบาหวาน ในระยะเริ่มแรกปริมาณของโปรตีนที่รั่วในปัสสาวะสามารถพยากรณ์การดำเนินโรคไตและความเสี่ยงในการเกิดโรคหัวใจและหลอดเลือดได้ เมื่อเริ่มวินิจฉัยโรคเบาหวาน ผู้ป่วยที่ตรวจไม่พบโปรตีนรั่วในปัสสาวะจะมีโอกาสตรวจพบปริมาณโปรตีนรั่วมากขึ้นในระดับ Micro Albuminuria ร้อยละ 2 ต่อปี เมื่อเกิด Micro Albuminuria แล้วจะมีโอกาสเพิ่มขึ้นเป็น Macro Albuminuria ร้อยละ 2.8 ต่อปี และจากระยะที่มี Macro Albuminuria แล้ว ต่อไปจะพบการเพิ่มขึ้นของซีรั่มครีเอตินินหรืออาจต้องได้รับการบำบัดทดแทนไตร้อยละ 2.3 ต่อปี หลังการวินิจฉัยเบาหวานชนิดที่ 2 นาน 10 ปี ผู้ที่มีการเพิ่มขึ้นของซีรั่มครีเอตินินมากกว่า 175 ไมโครโมลต่อลิตร หรือได้รับการบำบัดทดแทนไตแล้ว จะมีโอกาสตายคิดเป็น Annual Death Rate เท่ากับร้อยละ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

19.2 เมื่อเทียบกับร้อยละ 0.7 ในผู้ป่วยเบาหวานที่ไม่มีโรคไตร้อยละ 2 ในผู้ป่วยที่มี Micro Albuminuria และร้อยละ 3.5 ในผู้ป่วย Macro Albuminuria และในผู้ป่วยที่เกิด Macro Albuminuria แล้วจะมีโอกาสเสี่ยงต่อการตายมากกว่าที่จะดำเนินไปเป็นภาวะไตวายเรื้อรัง ดังนั้น การคัดกรองโรคไตด้วยการตรวจปัสสาวะในผู้ป่วยโรคเบาหวานในระยะเริ่มแรกจึงมีความสำคัญ

2.1.3 ระยะการดำเนินโรคไตจากเบาหวาน แบ่งออกเป็น 5 ระยะ คือ

ระยะที่ 1 Normoalbuminuria หมายถึง ภาวะที่ผู้ป่วยเบาหวานมีค่า Albumin/ Creatinine Ratio (ACR) < 30 mg/g Creatinine จากการตรวจตัวอย่างปัสสาวะถ่ายครั้งเดียว หากตรวจด้วยชุดตรวจสำหรับ Microalbuminuria จะให้ผลลบ

ระยะที่ 2 Microalbuminuria หมายถึง ระยะเริ่มแรกของโรคไตจากเบาหวานที่ตรวจไม่พบโปรตีนชนิดอัลบูมินในปัสสาวะด้วยชุดตรวจโปรตีนในปัสสาวะ (Dipstick) ทั่วไป แต่จะตรวจพบได้ด้วยชุดตรวจหาอัลบูมินปริมาณน้อย ๆ (Microalbuminuria Dipstick) หรือวัดปริมาณอัลบูมินในปัสสาวะได้ 20-200 ไมโครกรัมต่อนาทีหรือ 30-300 มิลลิกรัมต่อวัน อย่างน้อย 2 ใน 3 ครั้ง ภายในระยะเวลา 6 เดือน โดยไม่มีสาเหตุอื่น

ระยะที่ 3 Macroalbuminuria, Overt Proteinuria หรือ Overtnephropathy หมายถึง ระยะของโรคไตจากเบาหวานที่ตรวจพบโปรตีนชนิดอัลบูมินในปัสสาวะด้วยชุดตรวจโปรตีนในปัสสาวะ(Dipstick) ได้ตั้งแต่ Trace ขึ้นไป หรือวัดปริมาณอัลบูมินในปัสสาวะได้มากกว่า 200 ไมโครกรัมต่อนาที หรือ 300 มิลลิกรัมต่อวัน อย่างน้อย 2 ใน 3 ครั้ง ภายในระยะเวลา 6 เดือน โดยไม่มีสาเหตุอื่น

ระยะที่ 4 โรคไตเรื้อรัง (Chronic Kidney Disease) หมายถึง ภาวะที่ไตมีความบกพร่องในการกำจัดของเสีย โดยค่า Glomerular Filtratio Rate หรือ Creatinine Clearance ต่ำกว่า 60 มิลลิลิตรต่อนาที มีความผิดปกติทางพยาธิสภาพหรือโครงสร้างของไต ถึงแม้ว่า GFRปกติ เช่น ตรวจพบโปรตีนรั่วทางปัสสาวะเกินปกติ พบเม็ดเลือดแดงในปัสสาวะเกินปกติ เป็นต้น และความผิดปกติที่กล่าวมาข้างต้นต้องเป็นอย่างเรื้อรัง คือ เป็นติดต่อกันนานเกิน 3 เดือน

ระยะที่ 5 ไตวายเรื้อรังระยะสุดท้าย (End-Stage Kidney Disease) หมายถึง โรคไตเรื้อรังที่มีการทำงานของไตต่ำมาก คือ ค่า Creatinine Clearance ต่ำกว่า 10-15 มิลลิลิตรต่อนาที ผู้ป่วยมักจะมีอาการหรือภาวะแทรกซ้อนของการที่มีของเสียคั่งในร่างกาย และส่วนใหญ่จำเป็นต้องประทังชีวิตด้วยการรักษาทดแทนไต ได้แก่ การฟอกเลือดด้วยเครื่องไตเทียม (Hemodialysis) การล้างไตด้วยน้ำยาทางช่องท้อง (Continuous Ambulatory Peritoneal Dialysis, CAPD) หรือการปลูกถ่ายไต (Kidney Transplantation)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.4 ปัจจัยเสี่ยงต่อการเกิดโรคไตจากเบาหวาน ได้แก่

2.1.4.1 ระยะเวลาของการเป็นโรคเบาหวานมานาน

2.1.4.2 มีประวัติครอบครัวของโรคไตจากเบาหวาน หรือไตวายเรื้อรังระยะสุดท้าย หรือความดันโลหิตสูง

2.1.4.3 การควบคุมระดับน้ำตาลได้ไม่ดี

2.1.4.4 การควบคุมระดับความดันโลหิตสูงได้ไม่ดี

2.1.4.5 ภาวะไขมันในเลือดสูง

2.1.4.6 มีโปรตีนชนิดอัลบูมินรั่วออกทางปัสสาวะมากกว่าปกติ

2.1.4.7 การสูบบุหรี่

2.1.5 การป้องกันและการรักษาโรคไตจากเบาหวาน

2.1.5.1 ควบคุมระดับน้ำตาลในเลือดให้อยู่ในระดับดี คือ มี Fasting Plasma Glucose ระหว่าง 90 – 130 มิลลิกรัมต่อเดซิลิตร และระดับ Hemoglobin A1c น้อยกว่าร้อยละ 7

2.1.5.2 ควบคุมระดับความดันโลหิตให้ต่ำกว่า 130/90 มิลลิเมตรปรอท กลุ่มของยาลดความดันโลหิตที่สามารถเลือกใช้ตัวใดตัวหนึ่งหรือใช้ร่วมกัน หากไม่มีข้อห้าม (ตามหลักฐานการศึกษาต่าง ๆ ที่สามารถชะลอความเสื่อมของโรคไตหรือสามารถลดภาวะแทรกซ้อนหรืออัตราการเสียชีวิตจากโรคหัวใจและหลอดเลือดได้) ได้แก่

1) Angiotensin-Converting Enzyme Inhibitors
2) Angiotensin II Receptor Blockers
3) Calcium Channel Blockers
4) Beta-Blockers
5) Diuretics
6) ยาลดความดันโลหิตกลุ่มอื่น ๆ

ทั้งนี้แม้ว่าการศึกษาเพื่อประเมินบทบาทของยาลดความดันโลหิตในการชะลอการเสื่อมของไตในผู้ป่วยเบาหวานชนิดที่ 2 จะมีเพียงการศึกษาเฉพาะยาในกลุ่ม Angiotensin II Receptor Blockers ก็ตาม แต่การใช้ยาในกลุ่ม Angiotensin II Receptor Blockers มีข้อจำกัดคือมีราคายาสูง ซึ่งอาจทำให้ผู้ป่วยจำนวนหนึ่งไม่สามารถรับประทานยานี้ได้ กลุ่มผู้เชี่ยวชาญจึงมีความเห็นร่วมกันว่าเพื่อให้เหมาะสมกับสภาพเศรษฐกิจของประเทศไทยในปัจจุบัน ยากลุ่ม Angiotensin - Converting Enzyme Inhibitors ก็น่าจะเป็นทางเลือกที่ดีสำหรับป้องกันการเกิดโรคไตและชะลอการเสื่อมของไตในผู้ป่วยไทยที่มีโรคเบาหวาน ชนิดที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.5.3 ผู้ป่วยโรคเบาหวานที่มีอัลบูมินในปัสสาวะจาก Dipstick และ/หรือ มีภาวะ Microalbuminuria ที่มีความดันโลหิตสูงต้องได้รับการรักษา ควบคุมระดับความดันโลหิตให้ต่ำกว่า130/90 มิลลิเมตรปรอท ในกรณีที่ผู้ป่วยมีความดันโลหิตไม่สูงแต่มี Dipstick และ/หรือ

Microalbuminuria ควรพิจารณาให้ยากลุ่ม Angiotensin-Converting Enzyme Inhibitors หรือ Angiotensin II Receptor Blockers เพื่อช่วยชะลอการเสื่อมของไต

2.1.5.4 ควรให้รับคำแนะนำและการรักษาปัจจัยเสี่ยงอื่น ๆ อย่างเหมาะสม ได้แก่ การควบคุมระดับไขมันในเลือด การควบคุมน้ำหนัก การออกกำลังกายสม่ำเสมอ และงดสูบบุหรี่

2.1.5.5 ผู้ป่วยเบาหวานที่มีระดับซีรัมครีอะตินีนตั้งแต่ 1.5 มิลลิกรัมต่อเดซิลิตรขึ้นไป ควรได้รับการดูแลรักษาเช่นเดียวกับผู้ป่วยโรคไตเรื้อรัง เช่น จำกัดปริมาณโปรตีนในอาหารให้เหมาะสม

2.1.6 การชะลอความเสื่อมของโรคไตจากเบาหวาน ได้แก่

2.1.6.1 การควบคุมระดับน้ำตาลในเลือด คำแนะนำจาก The Kidney Disease Improving Global Outcomes (KDIGO) ปี 2012 และ ADA ปี 2014 แนะนำให้ควบคุมระดับน้ำตาลสะสม HbA1c อยู่ที่ร้อยละ 7 เพื่อลดการเกิดภาวะแทรกซ้อนด้าน Microvascular และไม่ควรให้ระดับน้ำตาล HbA1c ต่ำกว่าร้อยละ 7 ในผู้ที่มีความเสี่ยงเกิดน้ำตาลในเลือดต่ำหรือผู้ป่วย CKD ระยะ 4 หรือ 5 ส่วนผู้ป่วยที่มีโรคร่วมหลายโรค ผู้ที่มีภาวะแทรกซ้อนด้าน Microvascular หรือ Macrovascular อยู่ก่อนแล้ว หรือผู้สูงอายุที่ควบคุมระดับน้ำตาลในเลือดยาก ให้อนุโลมรักษาระดับน้ำตาล HbA1c มากกว่าร้อยละ 7 ได้

2.1.6.2 การควบคุมภาวะความดันโลหิตสูง เป้าหมายของการควบคุมระดับความดันโลหิตจากคำแนะนำของ KDIGO แนะนำให้น้อยกว่า 140/90 มิลลิเมตรปรอท การรักษาภาวะความดันโลหิตสูง แนะนำให้ยารักษารักษาภาวะความดันโลหิตสูงร่วมกับการเน้นปรับเปลี่ยนพฤติกรรม (Lifestyle Modifications)

2.1.6.3 การควบคุมระดับไขมันในเลือด ภาวะไขมันสูงในผู้ป่วยโรคเบาหวานเป็นปัจจัยเสี่ยงที่สำคัญต่อการเกิดโรคหัวใจและหลอดเลือดและเพิ่มอัตราการตาย ปัจจุบันคำแนะนำของ KDIGO ปี 2013 แนะนำให้ผู้ป่วยโรคไตเรื้อรังจากเบาหวานที่ยังไม่ได้รับการบำบัดทดแทนไต ควรได้รับยากลุ่ม Statin หรือ Statin ร่วมกับ Ezetimibe โดยไม่ขึ้นกับระดับ LDL Cholesterol เพื่อป้องกันการเกิดโรคจากภาวะหลอดเลือดแข็งตัว (Atherosclerosis) ส่วนผู้ป่วยเบาหวานที่ได้รับการบำบัดทดแทนไตแล้วไม่พบประโยชน์จากการให้ยากลุ่ม Statin และอาจเกิดผลข้างเคียงได้ จึงไม่ควรเริ่มยาในผู้ป่วยรายใหม่ ส่วนผู้ป่วยที่ได้รับยา Statin มาก่อนหน้านี้แล้ว ยังไม่มีการศึกษาที่ชัดเจน จึงยังไม่มีคำแนะนำว่าต้องปฏิบัติเช่นไร และไม่แนะนำให้การรักษาด้วยยา Statin ร่วมกับยาในกลุ่ม Fibrate หรือNiacin เนื่องจากไม่ได้ประโยชน์เพิ่มขึ้น อาจเกิดภาวะตับอักเสบ หรือเกิด Rhabdomyolysis ได้ เป้าหมายของการรักษาระดับไขมันในเลือดในผู้ป่วยโรคไตจากเบาหวาน คือ ระดับ LDL Cholesterol น้อยกว่า 100 มิลลิกรัมต่อเดซิลิตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของผลงานทุกครั้งที่มีการนำไปใช้

2.1.6.4 การควบคุมระดับไขมันในเลือด การจำกัดปริมาณโปรตีนผู้ป่วยโรคไตจากเบาหวาน ควรได้รับโปรตีนปริมาณ 0.8 กรัมต่อกิโลกรัมต่อวัน (คิดเป็นร้อยละ 10 ของปริมาณแคลอรี่ทั้งวัน) มีการศึกษาพบว่าว่าผู้ป่วยเบาหวานที่ได้รับอาหารโปรตีนสูง จะเพิ่มการรั่วของโปรตีนไข่ขาว

และทำให้การทำงานของไตลดลงได้ควรระวังภาวะการขาดสารอาหารจากการจำกัดโปรตีนมากจนเกินไปด้วย

2.1.6.5 การควบคุมระดับไขมันในเลือด การลดน้ำหนักภาวะอ้วนพบได้ร้อยละ 80 ในผู้ป่วยเบาหวานชนิดที่ 2 เป็นปัจจัยเสี่ยงในการเกิดโรคไตจากเบาหวานร่วมกับปริมาณ Visceral Adiposity สามารถพยากรณ์โรคที่สำคัญในการเกิด Microalbuminuria การลดน้ำหนักจึงอาจช่วยลดความเสี่ยงในการเกิดโรคไตจากเบาหวานได้ ยาลดระดับน้ำตาลในเลือดบางกลุ่มอาจมีผลกับน้ำหนักตัว เช่น Sulfonylureas, Thiazolidinediones และ Insulin อาจทำให้มีน้ำหนักเพิ่มได้ประมาณ 2.2 ถึง 3.4 กิโลกรัม ในขณะที่ Metformin และ DPP-4 Inhibitors ไม่มีผลกับน้ำหนักตัว ส่วน GLP-1 Receptor Agonists อาจทำให้น้ำหนักตัวลดลงได้

2.2 การวิเคราะห์จำแนกประเภท (Discriminant Analysis)

การวิเคราะห์จำแนกประเภทเป็นเทคนิคการวิเคราะห์ข้อมูลหลายตัวแปร (Multivariate Analysis) หรือเป็นวิธีการทางสถิติที่ใช้วิเคราะห์จำแนกกลุ่มข้อมูลตั้งแต่ 2 กลุ่มขึ้นไป ที่มีวัตถุประสงค์เพื่อสร้างสมการหรือตัวแบบสำหรับพยากรณ์กลุ่มของตัวแปรตามจากตัวแปรอิสระหลายตัว โดยที่ตัวแปรตามเป็นตัวแปรเชิงกลุ่ม (Categorical Variable) และตัวแปรอิสระอาจจะเป็นตัวแปรเชิงกลุ่มหรือตัวแปรต่อเนื่องก็ได้ ซึ่งวิธีการจำแนกประเภทนี้ถูกคิดค้นโดย R.A. Fisher ในปี ค.ศ. 1936

2.2.1 วัตถุประสงค์ของการวิเคราะห์จำแนกประเภท

2.2.1.1 เพื่อศึกษาปัจจัยหรือตัวแปรที่ทำให้กลุ่มมีความแตกต่างกัน ซึ่งเป็นการศึกษาว่าตัวแปรอิสระใดบ้างเป็นตัวแปรที่ทำให้กลุ่มต่างกัน โดยสร้างฟังก์ชันการจำแนกกลุ่มที่อยู่ในรูปเชิงเส้น

2.2.1.2 เพื่อพยากรณ์หน่วยใหม่ที่ยังไม่ทราบกลุ่มมาก่อนว่าควรจะอยู่กลุ่มใดในอนาคต เช่น พยากรณ์ลูกค้าสินเชื่อรายใหม่ว่าในอนาคตจะอยู่กลุ่มใด ถ้าพยากรณ์ว่านาย ก อยู่ในกลุ่มลูกหนี้ปกติ ธนาคารมีความเสี่ยงน้อยที่จะให้กู้ แต่ถ้าพยากรณ์ว่านาย ก อยู่ในกลุ่มลูกหนี้มีปัญหาธนาคารมีความเสี่ยงสูงที่จะให้กู้ เป็นต้น

2.2.2 ข้อสมมติเบื้องต้นของการวิเคราะห์จำแนกประเภท

2.2.2.1 ตัวแปรอิสระมีการแจกแจงปรกติหลายตัวแปร (Normality of Independent Variable)

2.2.2.2 เมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระของแต่ละกลุ่มเท่ากัน (Equal Dispersion Matrices)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.3 การตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์จำแนกประเภท

2.2.3.1 ตัวแปรอิสระมีการแจกแจงปรกติหลายตัวแปร ตรวจสอบได้จากการแจกแจงปรกติตัวแปรเดียว (Univariate Normal Distribution) โดยทำการแยกตรวจสอบตัวแปรอิสระทีละตัว หากพบว่าตัวแปรอิสระทุกตัวมีการแจกแจงปรกติ ก็มีความน่าจะเป็นสูงที่ตัวแปรอิสระจะมีการแจกแจงปรกติหลายตัวแปรด้วย หรืออาจจะตรวจการสอบการแจกแจงปรกติหลายตัวแปรด้วยข้อมูลค่านอกเกณฑ์หลายตัวแปร (Multivariate Outliers) โดยการวิเคราะห์ Mahalanobis Distances และถ้าพบว่าไม่มีข้อมูลค่านอกเกณฑ์หลายตัวแปร ก็มีความน่าจะเป็นสูงว่าตัวแปรอิสระมีการแจกแจงปรกติหลายตัวแปรด้วย (ทรงศักดิ์ ภูสีอ่อน, 2551) ซึ่งในงานวิจัยนี้ผู้วิจัยได้เลือกใช้วิธีการตรวจสอบการแจกแจงปรกติหลายตัวแปรด้วยการตรวจสอบการแจกแจงปรกติตัวแปรเดียว โดยทำการแยกตรวจสอบตัวแปรอิสระทีละตัว โดยใช้ตัวสถิติทดสอบ Kolmogorov-Smirnov หรือ Shapiro-Wilk กรณีที่ข้อมูลมีจำนวนมากกว่า 50 ชุด จะไม่มีการคำนวณค่าตัวสถิติสอบ Shapiro-Wilk ดังผู้วิจัยจึงเลือกใช้ตัวสถิติทดสอบ Kolmogorov-Smirnov ในการตรวจสอบจากการแจกแจงปรกติตัวของแปรอิสระในครั้งนี้ เนื่องจากข้อมูลที่ใช้ในงานวิจัยครั้งนี้มีจำนวนมากกว่า 50 ชุด

Barbara and Linda (2001) กล่าวว่า เมื่อตัวแปรอิสระที่นำมาศึกษามีทั้งตัวแปรเชิงกลุ่มและตัวแปรต่อเนื่อง จึงเป็นเรื่องยากที่ข้อมูลจะมีการแจกแจงปรกติหลายตัวแปร แต่ถ้าจำนวนตัวอย่างที่นำมาศึกษามีขนาดใหญ่ และวัตถุประสงค์หลักของการวิจัยคือ ต้องการสร้างสมการจำแนกประเภทที่มีประสิทธิภาพในการจำแนกกลุ่มข้อมูลให้ถูกต้องมากที่สุด และเพื่อเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่มข้อมูล โดยใช้วิธีการจำแนกกลุ่มข้อมูลทั้ง 3 วิธี คือ การวิเคราะห์จำแนกประเภท การวิเคราะห์การถดถอยลอจิสติก และต้นไม้ตัดสินใจ ดังนั้นข้อสมมติเบื้องต้นดังกล่าวจึงไม่ใช่ข้อสมมติเบื้องต้นที่จำเป็น (ศยามล ลำลองรัตน์, 2550)

2.2.3.2 เมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระของแต่ละกลุ่มเท่ากัน (Equal Dispersion Matrices) ตรวจสอบได้โดยใช้ตัวสถิติทดสอบ Box's M

สมมติฐานของการทดสอบ

H_0 : $\Sigma_1 = \Sigma_2 = ... = \Sigma_k$

H_1 : $\Sigma_i = \Sigma_j$ อย่างน้อย 1 คู่ ; $i \neq j$

หรือ H_0 : เมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระทุกกลุ่มเท่ากัน

H_1 : เมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระอย่างน้อย 1 คู่ ไม่เท่ากัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวสถิติทดสอบ คือ

$$B = (1-c)\left\{\left[\sum_{i=1}^{k}(n_i-1)\ln\left|S_p\right|\right] - \left[\sum_{i=1}^{k}(n_i-1)\ln\left|S_i\right|\right]\right\}$$

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

$$df = \frac{1}{2}p(p+1)(k-1)$$

p คือ จำนวนตัวแปรอิสระ

k คือ จำนวนกลุ่มข้อมูล

n_i คือ จำนวนข้อมูลในกลุ่มที่ i

S_i คือ เมทริกซ์ความแปรปรวนร่วมของกลุ่มตัวอย่างที่ i โดยที่ i = 1, 2,..., k

S_p คือ เมทริกซ์ความแปรปรวนร่วมของกลุ่มตัวอย่างทั้งหมด

จากนั้นนำค่าตัวสถิติทดสอบ Box's M ที่ได้จากการคำนวณมาเปรียบเทียบกับค่าไคกำลังสอง (χ^2) ถ้าค่าตัวสถิติทดสอบ Box's M มีค่ามากกว่าค่าไคกำลังสอง จะปฏิเสธ H_0 หรือสรุปได้ว่าเมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระอย่างน้อย 1 คู่ ไม่เท่ากัน ตัวสถิติทดสอบไคกำลังสองจะขึ้นอยู่กับขนาดตัวอย่าง ในกรณีที่ตัวอย่างมีขนาดใหญ่จะทำให้ค่าไคกำลังสองมีค่ามาก ซึ่งจะทำให้ปฏิเสธ H_0 เสมอ

Barbara and Linda (2001) กล่าวว่า หากข้อมูลไม่สอดคล้องกับข้อสมมติเบื้องต้นเกี่ยวกับเมทริกซ์ความแปรปรวนร่วมของกลุ่มตัวอย่างเท่ากัน สามารถใช้การปรับค่าเมทริกซ์ความแปรปรวนร่วมโดยใช้เมทริกซ์ความแปรปรวนร่วมแบบแยก (Separate Variance – Covariance matrices) แทนเมทริกซ์ความแปรปรวนร่วมแบบรวม (Pool Variance – Covariance matrices) ได้ หรือข้อมูลที่นำศึกษานี้ไม่เหมาะสมที่จะใช้ฟังก์ชันจำแนกประเภทในการจำแนกกลุ่ม และอาจจะแก้ปัญหาโดยใช้ฟังก์ชันจำแนกประเภทกำลังสอง (Quadratic Discriminant Function) ในการสร้างสมการจำแนกกลุ่มแทนฟังก์ชันจำแนกประเภทเชิงเส้นตรง (Linear Discriminant Function) (ศยามล ลำลองรัตน์, 2550)

2.2.4 ลักษณะข้อมูลและการเตรียมข้อมูลที่จะนำมาวิเคราะห์

2.2.4.1 แบ่งกลุ่มประชากรหรือกลุ่มตัวอย่างแบ่งออกเป็นกลุ่มอย่างน้อย 2 กลุ่ม และต้องทราบมาก่อนการวิเคราะห์ว่าจะแบ่งออกเป็นกี่กลุ่ม

2.2.4.2 เลือกตัวแปรอิสระที่คาดว่าจะทำให้กลุ่มที่แบ่งไว้ในข้อ 2.2.4.1 แตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.4.3 ตัวแปรตาม คือ ตัวแปรที่ถูกจำแนกหรือตัวแปรที่ถูกทำนาย เป็นตัวแปรที่แบ่งออกเป็นกลุ่มอย่างน้อย 2 กลุ่มขึ้นไป มีระดับการวัดข้อมูลในระดับมาตรานามบัญญัติ (Nominal Scale) หรือมาตราเรียงลำดับ (Ordinal Scale) และถ้าหากข้อมูลอยู่ในระดับการวัดแบบอื่นให้แปลงข้อมูลเป็นตัวแปรเชิงกลุ่มก่อนที่จะนำไปวิเคราะห์ต่อไป

2.2.4.4 ตัวแปรอิสระ คือ ตัวแปรที่ใช้ในการจำแนกหรือตัวแปรที่ใช้ในการทำนายเป็นตัวแปรต่อเนื่องที่มีระดับการวัดอยู่ในมาตราอันตรภาค (Interval Scale) หรือมาตราอัตราส่วน (Ratio Scale) และหากข้อมูลอยู่ในระดับการวัดแบบอื่นให้แปลงข้อมูลเป็นตัวแปรหุ่น (Dummy Coding) ก่อนที่จะนำไปวิเคราะห์ต่อไป

2.2.5 วิธีการวิเคราะห์จำแนกประเภทมีดังนี้

2.2.5.1 กรณีการวิเคราะห์จำแนกประเภท ออกเป็น 2 กลุ่ม

หลักการในการวิเคราะห์จำแนกประเภทโดยวิธีของ Fisher's เป็นการแบ่งตัวอย่างข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มที่ 1 และ 2 มีตัวแปรอิสระ p ตัว คือ X_1, X_2,..., X_p โดยที่ขนาดตัวอย่างคือ n อยู่ในกลุ่มที่ 1 จำนวน n_1 หน่วย และอยู่ในกลุ่มที่ 2 จำนวน n_2 หน่วย โดยที่ $n = n_1 + n_2$ และกำหนดให้

กลุ่มที่ 1 มีเวกเตอร์ค่าเฉลี่ย $\underline{\mu}_1$ และเมทริกซ์ความแปรปรวนร่วม Σ_1

กลุ่มที่ 1 มีเวกเตอร์ค่าเฉลี่ย $\underline{\mu}_2$ และเมทริกซ์ความแปรปรวนร่วม Σ_2

ถ้า $\Sigma_1 = \Sigma_2 = \Sigma_{p\times p}$ และ $\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$

ฟังก์ชันจำแนกประเภทจะอยู่ในรูปเชิงเส้น ดังนี้

$$\hat{Y} = b_1X_1 + b_2X_2 + ... + b_pX_p$$

หรือ $\hat{Y} = \underline{b}'\underline{X}$

โดยที่ $\underline{X}$ เป็นเวกเตอร์ของตัวแปรอิสระ p ตัว

$\underline{X}' = (X_1, X_2, ..., X_p)$

และ $\underline{b}$ เป็นเวกเตอร์ของสัมประสิทธิ์จำแนกประเภท

$\underline{b}' = (b_1, b_2, ..., b_p)$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารทุกครั้งที่มีการนำไปใช้

โดยสัมประสิทธิ์ b ที่ทำให้อัตราส่วนระหว่างความผันแปรระหว่างกลุ่ม (Between - Group Sum Square : SSB) กับความผันแปรภายในกลุ่ม (Within - Group Sum Square : SSW) มีค่ามากที่สุด หรือเป็นการหาค่า b ที่ทำให้ $L = \frac{SSB}{SSW}$ มีค่ามากที่สุด

โดยที่ Between - Group Sum Square (SSB) คือ ผลบวกกำลังสองของความแตกต่างระหว่างกลุ่ม

Within - Group Sum Square (SSW) คือ ผลบวกกำลังสองของความแตกต่างภายในกลุ่ม

จะได้ว่า ค่าเฉลี่ยของกลุ่มที่ 1 คือ $\overline{Y}_1 = \underline{b}'\underline{\overline{X}}_1$

ค่าเฉลี่ยของกลุ่มที่ 2 คือ $\overline{Y}_2 = \underline{b}'\underline{\overline{X}}_2$

โดยที่ X_1 เป็นเมทริกซ์ขนาด $p \times n_1$ ของข้อมูลในกลุ่มที่ 1

X_2 เป็นเมทริกซ์ขนาด $p \times n_2$ ของข้อมูลในกลุ่มที่ 2

$\underline{\overline{X}}_1$ เป็นเวกเตอร์ค่าเฉลี่ยของตัวแปรอิสระ p ตัว ของข้อมูลในกลุ่มที่ 1

$\underline{\overline{X}}_2$ เป็นเวกเตอร์ค่าเฉลี่ยของตัวแปรอิสระ p ตัว ของข้อมูลในกลุ่มที่ 2

$$X_1 = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_1,1} & X_{n_1,2} & \cdots & X_{n_1,p} \end{bmatrix} \quad \text{และ} \quad \underline{\overline{X}}_1 = \begin{bmatrix} \overline{X}_{11} \\ \overline{X}_{12} \\ \vdots \\ \overline{X}_{1p} \end{bmatrix}_{p\times 1}$$

$$X_2 = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_2,1} & X_{n_2,2} & \cdots & X_{n_2,p} \end{bmatrix} \quad \text{และ} \quad \underline{\overline{X}}_2 = \begin{bmatrix} \overline{X}_{21} \\ \overline{X}_{22} \\ \vdots \\ \overline{X}_{2p} \end{bmatrix}_{p\times 1}$$

เมื่อ $SSW = \sum_{i=1}^{n_1}\left(Y_{1i} - \overline{Y}_1\right) + \sum_{j=1}^{n_2}\left(Y_{2j} - \overline{Y}_2\right)$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อแทนค่า $Y_i = \underline{b}'\underline{X}_i$ และ $\overline{Y}_i = \underline{b}'\underline{\overline{X}}_i$ จะได้

$$SSW = \sum_{i=1}^{n_1} \underline{b}'\left(\underline{X}_{1i}-\overline{\underline{X}}_1\right)\underline{b}+\sum_{j=1}^{n_2} \underline{b}'\left(\underline{X}_{2j}-\overline{\underline{X}}_2\right)\underline{b}$$

$$= \underline{b}'\left[\sum_{i=1}^{n_1}\left(\underline{X}_{1i}-\overline{\underline{X}}_1\right)\left(\underline{X}_{1i}-\overline{\underline{X}}_1\right)'\right]\underline{b}+\underline{b}'\left[\sum_{j=1}^{n_2}\left(\underline{X}_{2j}-\overline{\underline{X}}_2\right)\left(\underline{X}_{2j}-\overline{\underline{X}}_2\right)'\right]\underline{b}$$

$$= \underline{b}'\left[\sum_{i=1}^{n_1}\left(\underline{X}_{1i}-\overline{\underline{X}}_1\right)\left(\underline{X}_{1i}-\overline{\underline{X}}_1\right)'+\sum_{j=1}^{n_2}\left(\underline{X}_{2j}-\overline{\underline{X}}_2\right)\left(\underline{X}_{2j}-\overline{\underline{X}}_2\right)'\right]\underline{b}$$

ในกรณีที่ $\Sigma_1 = \Sigma_2 = \Sigma_{p\times p}$ จะได้ว่า SSW แปรผันตรงกับ $\underline{b}'\Sigma\underline{b}$

และ $SSB = n_1\left(\overline{Y}_1-\overline{Y}\right)+n_2\left(\overline{Y}_2-\overline{Y}\right)$

เมื่อแทนค่า $\overline{Y} = \underline{b}'\overline{\underline{X}}$ และ $\overline{Y}_i = \underline{b}'\overline{\underline{X}}_i$ จะได้

$$SSB = \underline{b}'\left[n_1\left(\overline{\underline{X}}_1-\overline{\underline{X}}\right)\left(\overline{\underline{X}}_1-\overline{\underline{X}}\right)'+n_2\left(\overline{\underline{X}}_2-\overline{\underline{X}}\right)\left(\overline{\underline{X}}_2-\overline{\underline{X}}\right)'\right]\underline{b}$$

โดยที่ $\overline{\underline{X}} = \dfrac{n_1\overline{\underline{X}}_1+n_2\overline{\underline{X}}_2}{n_1+n_2}$

หรือ $\overline{\underline{X}}_1-\overline{\underline{X}} = \dfrac{n_2\left(\overline{\underline{X}}_1-\overline{\underline{X}}_2\right)}{n_1+n_2} = \dfrac{n_2\underline{d}}{n_1+n_2}$

$$\overline{\underline{X}}_2-\overline{\underline{X}} = \frac{n_1\left(\overline{\underline{X}}_2-\overline{\underline{X}}_1\right)}{n_1+n_2} = -\frac{n_1\underline{d}}{n_1+n_2}$$

โดยที่ $\underline{d} = \overline{\underline{X}}_1-\overline{\underline{X}}_2$ เป็นเวกเตอร์ของผลต่างของเวกเตอร์ค่าเฉลี่ยของ 2 กลุ่ม

แทนค่า $\overline{\underline{X}}_1-\overline{\underline{X}}$ และ $\overline{\underline{X}}_2-\overline{\underline{X}}$ ลงในสมการของ SSB จะได้

$$SSB = \underline{b}'\left[n_1\left(\frac{n_2\underline{d}}{n_1+n_2}\right)\underline{d}\underline{d}'+n_2\left(-\frac{n_2\underline{d}}{n_1+n_2}\right)\underline{d}\underline{d}'\right]\underline{b}$$

หรือ SSB ผันแปรตรงกับ $\underline{b}'\underline{d}\underline{d}'\underline{b}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้น $L = \dfrac{SSB}{SSW} = \dfrac{\underline{b}'\underline{d}\underline{d}'\underline{b}}{\underline{b}'\Sigma\underline{b}}$

ซึ่งวิธีของ Fisher คือ การหาเวกเตอร์สัมประสิทธิ์ของการจำแนกประเภท b ที่ทำให้ L มีค่าสูงสุด ดังนี้

$$\frac{\partial L}{\partial \underline{b}} = \frac{\left[\left(\underline{b}'\Sigma\underline{b}\right)\left(2\underline{d}\underline{d}'\underline{b}\right)-\left(\underline{b}'\underline{d}\underline{d}'\underline{b}\right)\left(2\Sigma\underline{b}\right)\right]}{\left(\underline{b}'\Sigma\underline{b}\right)^2} = 0$$

นำ $-\frac{1}{2}\underline{b}'\Sigma\underline{b}$ มาคูณทั้ง 2 ข้างของสมการ จะได้

$$\frac{\left(\underline{b}'\Sigma\underline{b}\right)^2\underline{d}\underline{d}'\underline{b}}{\left(\underline{b}'\Sigma\underline{b}\right)^2}-\frac{\underline{b}'\Sigma\underline{b}\left(\underline{b}'\underline{d}\underline{d}'\underline{b}\right)\Sigma\underline{b}}{\left(\underline{b}'\Sigma\underline{b}\right)^2} = 0$$

$$\underline{d}\underline{d}'\underline{b}-\frac{\left(\underline{b}'\underline{d}\underline{d}'\underline{b}\right)\Sigma\underline{b}}{\underline{b}'\Sigma\underline{b}} = 0$$

$$\underline{d}\underline{d}'\underline{b}-L\Sigma\underline{b} = 0$$

$$\underline{b} = \frac{\Sigma^{-1}\underline{d}\underline{d}'\underline{b}}{L}$$

หรือ $\underline{b}$ ผันแปรตรงกับ $\Sigma^{-1}\underline{d}$

โดยที่ L และ $\underline{d}'\underline{b}$ เป็นสเกลาร์

ในกรณีที่ไม่ทราบค่าเมทริกซ์ความแปรปรวนร่วมของประชากร Σ จะประมาณค่าด้วยเมทริกซ์ความแปรปรวนร่วมของตัวอย่าง S_p โดยที่ S_p เป็นความแปรปรวนร่วมของตัวอย่างแบบรวม (Pool Sample Variance – Covariance Matrix)

ดังนั้น $\underline{b} = S_p^{-1}\left(\overline{\underline{X}}_1-\overline{\underline{X}}_2\right)$

โดยที่ $S_p = \frac{1}{n_1+n_2-2}\left(\overline{\underline{X}}'_1\overline{\underline{X}}_1-\overline{\underline{X}}'_2\overline{\underline{X}}_2\right)$

การทดสอบสมมติฐานของการวิเคราะห์จำแนกประเภท

การวิเคราะห์จำแนกประเภท เบื้องต้นต้องทราบจำนวนกลุ่มข้อมูลที่จะทำการวิเคราะห์ก่อนว่าข้อมูลนั้นแบ่งออกเป็นกี่กลุ่ม ซึ่งกลุ่มที่แบ่งนี้อาจแตกต่างกันหรอไม่แตกต่างกันก็ได้ จึงต้องทำการตรวจสอบก่อนว่าข้อมูลทั้ง 2 กลุ่มนี้แตกต่างกันจริงหรือไม่ ก่อนจะวิเคราะห์ในขั้นตอนต่อไป ซึ่งวิธีในการตรวจสอบอาจมีหลายวิธีได้แก่ Wilk's Lamda, Hotelling's T^2, Roy's Largest Root และ Pillia's Trace เป็นต้น ซึ่งในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้วิธีการทดสอบ Hotelling's T^2 ในการตรวจสอบ

เอกสารนี้เป็น ... ไม่ว่ากรณีใด ...

สมมติฐานการทดสอบ

$H_0: \underline{\mu}_1 = \underline{\mu}_2$

$H_1: \underline{\mu}_1 \neq \underline{\mu}_2$

หรือ H_0 : ค่าเฉลี่ยของตัวแปรอิสระทุกตัวในกลุ่มที่ 1 เท่ากับกลุ่มที่ 2

H_1 : ค่าเฉลี่ยของตัวแปรอิสระทุกตัวในกลุ่มที่ 1 ไม่เท่ากับกลุ่มที่ 2

ตัวสถิติทดสอบ Hotelling's T^2

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\underline{\bar{X}}_1 - \underline{\bar{X}}_2)' S_p^{-1} (\underline{\bar{X}}_1 - \underline{\bar{X}}_2)$$

$$= \frac{n_1 n_2}{n_1 + n_2} D^2$$

โดยที่ $D^2 = (\underline{\bar{X}}_1 - \underline{\bar{X}}_2)' S_p^{-1} (\underline{\bar{X}}_1 - \underline{\bar{X}}_2)$

$$F = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

$$= \frac{n_1 n_2 (n_1 + n_2 - p - 1)}{(n_1 + n_2) p (n_1 + n_2 - 2)} D^2$$

df = $p, (n_1 + n_2 - p - 1)$

n_i คือ ขนาดตัวอย่างของกลุ่มที่ i

p คือ จำนวนตัวแปรอิสระ

จากนั้นนำค่าสถิติทดสอบ Hotelling's T^2 ที่ได้จากการคำนวณมาเปรียบเทียบกับค่า F ถ้าค่าสถิติทดสอบ Hotelling's T^2 มีค่ามากกว่าค่า F จะปฏิเสธ H_0 แสดงว่าค่าเฉลี่ยของตัวแปรอิสระทุกตัวในกลุ่มที่ 1 ไม่เท่ากับกลุ่มที่ 2 หรือตัวแบบที่ได้สามารถนำไปใช้ในการจำแนกกลุ่มได้

การจัดหน่วยตัวอย่างเข้ากลุ่ม (Allocation of Observations to Group)

การจัดหน่วยตัวอย่างเข้ากลุ่ม คือ การกำหนดกลุ่มให้กับหน่วยตัวอย่างใหม่ หรือหน่วยที่ยังไม่ได้จัดกลุ่ม สามารถทำได้โดยการนำสมการวิเคราะห์จำแนกประเภทมาใช้ในการพยากรณ์กลุ่มให้กับหน่วยใหม่ โดยมีเทคนิคการจัดกลุ่มหลายวิธีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) Simple Classification Function เป็นวิธีการนำสมการจำแนกกลุ่มของ Fisher's มาแทนค่าของตัวแปรอิสระหรือตัวแปรจำแนกกลุ่มของหน่วยใหม่ แล้วคำนวณค่าตัวแปรตาม ถ้าค่าของ

ตัวแปรตามกลุ่มใดมีค่ามากที่สุด จะจัดหน่วยตัวอย่างใหม่ให้อยู่ในกลุ่มนั้น ซึ่งเทคนิคนี้เป็นอีกวิธีอิงพารามิเตอร์

2) Probabilities of Group Membership เป็นวิธีการคำนวณหาค่าความน่าจะเป็นหรือโอกาสที่หน่วยใหม่จะอยู่ในกลุ่มที่ i โดยที่ i = 1, 2, 3,..., k แล้วนำมาเปรียบเทียบกัน ถ้าพบว่าความน่าจะเป็นของกลุ่มใดมีค่ามากกว่าก็จะจัดหน่วยตัวอย่างใหม่ให้อยู่ในกลุ่มนั้น แต่มีข้อสมมติเบื้องต้นว่าตัวแปรอิสระต้องมีการแจกแจงปรกติหลายตัวแปร (Multivariate Normal Distribution) และความน่าจะเป็นที่คำนวณจะเป็นความน่าจะเป็นภายหลัง (Posterior Probability)

3) Generalized Distance Function เป็นวิธีการคำนวณหาระยะทางจากหน่วยใหม่ที่ต้องการจัดกลุ่มไปยังค่ากลางของกลุ่ม (Group Centroid) ถ้าระยะห่างที่คำนวณได้ห่างจากค่ากลางของกลุ่มใดน้อยที่สุด จะจัดให้หน่วยใหม่อยู่ในกลุ่มนั้น โดยที่ใช้วิธีการคำนวณระยะห่างของ Mahalanobis Distance นั่นคือคำนวณหาค่า Mahalanobis Distance ของหน่วยใหม่กับค่ากลางของกลุ่ม และจะจัดให้หน่วยใหม่อยู่ในกลุ่มที่มีระยะห่างจากหน่วยนั้นถึงจุดกึ่งกลางของกลุ่มน้อยที่สุด ซึ่งเทคนิคนี้เป็นอีกวิธีอิงนอนพารามิเตอร์

ซึ่งในงานวิจัยครั้งนี้ ผู้วิจัยเลือกใช้การจัดหน่วยตัวอย่างเข้ากลุ่มโดยใช้เทคนิค Simple Classification Function ในการจัดกลุ่มของหน่วยตัวอย่างใหม่ เนื่องจากงานวิจัยนี้มีการทดสอบสมมติฐานเบื้องต้นเกี่ยวกับตัวแปรอิสระมีการแจกแจงปรกติหลายตัวแปร ซึ่งเทคนิคนี้มีรายละเอียดดังต่อไปนี้

1) ฟังก์ชันการจำแนกกลุ่ม (Classification Function) ซึ่งเป็นผลรวมเชิงเส้นของตัวแปรอิสระในแต่ละกลุ่ม โดยจำนวนฟังก์ชันจำแนกกลุ่มจะมีจำนวนเท่ากับจำนวนกลุ่มที่ต้องการจำแนก ซึ่งมีฟังก์ชันการจำแนกกลุ่มดังนี้

$$\hat{Y} = b_1X_1 + b_2X_2 + ... + b_pX_p$$

หรือ $$\hat{Y} = \underline{b}'\underline{X}$$

โดยที่ X เป็นเวกเตอร์ของตัวแปรอิสระ p ตัว

$$\underline{X}' = (X_1, X_2, ..., X_p)$$

และ $\underline{b}$ เป็นเวกเตอร์ของสัมประสิทธิ์จำแนกประเภท

$$\underline{b}' = (b_1, b_2, ..., b_p)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) ในการจำแนกกลุ่มให้พิจารณาจากค่าฟังก์ชันจำแนกกลุ่ม (Classification Function) โดยจัดให้สมาชิกใด ๆ เข้าเป็นสมาชิกกลุ่มที่มีค่าฟังก์ชันจำแนกประเภทสูงที่สุด

2.2.5.1 กรณีการวิเคราะห์จำแนกประเภทมากกว่า 2 กลุ่ม

เป็นกรณีที่แบ่งข้อมูลออกเป็นกลุ่มย่อยมากกว่า 2 กลุ่ม หรือแบ่งออกเป็น k กลุ่ม โดยมีข้อสมมติเบื้องต้นเช่นเดียวกับการวิเคราะห์จำแนกประเภท 2 กลุ่ม ซึ่งสมการในการจำแนกกลุ่มมีดังนี้

$$\hat{Y} = b_{i1}X_1 + b_{i2}X_2 + ... + b_{ip}X_p \; ; i = 1, 2,...,k-1$$

หรือ $\hat{Y}_i = \underline{b}' \underline{X}_p$

และค่าเฉลี่ย $\overline{Y}_i = \underline{b}' \overline{\underline{X}}_p$

เมื่อ $\underline{X}$ คือ เวกเตอร์ของตัวแปรอิสระ p ตัว

$\underline{b}$ คือ เวกเตอร์สัมประสิทธิ์จำแนกประเภท

p คือ จำนวนกลุ่มข้อมูล

ซึ่งในงานวิจัยนี้ ผู้วิจัยได้เลือกใช้การวิเคราะห์จำแนกประเภท 2 กลุ่ม ในการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์

2.2.6 สถิติสำคัญของการวิเคราะห์จำแนกกลุ่ม

2.2.6.1 ค่าลักษณะเฉพาะ (Eigenvalue) เป็นค่าที่แสดงอัตราส่วนการผันแปรระหว่างกลุ่มต่อการผันแปรภายในกลุ่ม ถ้าค่าลักษณะเฉพาะมีค่าสูง แสดงว่าสมการดีหรือมีค่าการจำแนกสูง หรือกล่าวอีกนัยหนึ่งว่าค่าลักษณะเฉพาะ คือ ความแปรปรวนของคะแนนแปลงในรูป Y ที่แปลงมาจากค่า X_1, X_2,...,X_p นั่นเอง

2.2.6.2 สหสัมพันธ์คานอนิคัล (Canonical Correlation) เป็นสถิติซึ่งสามารถใช้ในการตัดสินความสำคัญของสมการจำแนก เป็นมาตรวัดความสัมพันธ์ของสมการกับกลุ่มของตัวแปรซึ่งระบุการเป็นสมาชิกของกลุ่มนั้น ๆ ของตัวแปรตาม โดยชี้ให้เห็นว่าการเป็นสมาชิกกลุ่มมีความสัมพันธ์กับสมการที่หามาได้มากน้อยเพียงใด ดังนั้นถ้า Canonical Correlation มีค่าสูง แสดงว่าการเป็นสมาชิกกลุ่มสามารถอธิบายความผันแปรของตัวแปรกับสมการจำแนกได้มาก

2.2.6.3 แลมด้าของวิลด์ (Wilk's Lambda) เป็นสถิติที่ใช้ทดสอบความแตกต่างระหว่างกลุ่มและเป็นมาตรวัดอำนาจในการจำแนกกลุ่มของตัวแปรด้วย ถ้าแลมด้าของวิลด์มีค่ามาก แสดงว่าตัวแปรจะสามารถอธิบายการเป็นสมาชิกของกลุ่มได้น้อย และถ้าแลมด้าของวิลด์มีค่าน้อย แสดงว่าตัวแปรจะสามารถอธิบายการเป็นสมาชิกของกลุ่มได้มาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.6.4 ร้อยละสัมพัทธ์ (Relative Percentage) หรือร้อยละสะสม (Cumulative Percentage) เป็นค่าที่แสดงความสัมพันธ์ของค่าลักษณะเฉพาะกับสมการที่ได้ว่ามีอำนาจในการแบ่งแยกคิดเป็นร้อยละเท่าใดของอำนาจในการจำแนกรวม

2.2.7 วิธีการสร้างสมการวิเคราะห์จำแนกประเภท มี 2 วิธี คือ

2.2.7.1 วิธีตรง (Enter Method) เป็นวิธีการนำตัวแปรอิสระทุกตัวตามที่ระบุไว้เข้าสมการ ด้วยเหตุผลทางทฤษฎี

2.2.7.2 วิธีแบบขั้นตอน (Stepwise Method) เป็นวิธีการเลือกตัวแปรอิสระทีละตัวเข้าสมการ โดยหาตัวแปรที่ดีที่สุดในการจำแนกเข้าสมการ จากนั้นก็หาตัวแปรที่ดีที่สุดในการจำแนกกลุ่มข้อมูลลำดับที่ 2, 3 เข้าสมการตามลำดับ

2.2.8 ตัวอย่างการวิเคราะห์จำแนกประเภท

จากตัวอย่างข้อมูลความสนใจต่อเครื่องตัดหญ้าที่มีการทำงานแบบต้องมีผู้ควบคุมและอัตโนมัติดังที่แสดงในตารางที่ 1 ซึ่งมีตัวแปร 2 ตัวคือ

X_1 คือ รายได้ครอบครัว (หน่วยพันดอลลาร์)

X_2 คือ ขนาดพื้นที่ (หน่วยพันตารางฟุต)

ซึ่งจากข้อมูลมีผลการดำเนินการดังนี้

ตารางที่ 2.1 ตัวอย่างข้อมูลรายได้ครอบครัวและขนาดพื้นที่

กลุ่มที่ 1 : แบบมีผู้ควบคุม		**กลุ่มที่ 2 : แบบอัตโนมัติ**	
X_1	X_2	X_1	X_2
60.0	18.4	75.0	19.6
85.5	16.8	52.8	20.8
64.8	21.6	64.8	17.2
61.5	20.8	43.2	20.4
87.0	23.6	84.0	17.6
110.1	19.2	49.2	17.6
108.0	17.6	59.4	16.0
82.8	22.4	66.0	18.4
69.0	20.0	47.4	16.4
93.0	20.8	33.0	18.8
51.0	22.0	51.0	14.0
81.0	20.0	63.0	14.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ $X_1 = \begin{bmatrix} 79.475 \\ 20.267 \end{bmatrix}$ $X_2 = \begin{bmatrix} 57.400 \\ 17.633 \end{bmatrix}$

$S_p = \begin{bmatrix} 276.675 & -7.204 \\ -7.204 & 4.273 \end{bmatrix}$ $S_p^{-1} = \begin{bmatrix} 0.004 & 0.006 \\ 0.006 & 0.244 \end{bmatrix}$

ดังนั้นฟังก์ชันในการจำแนกกลุ่มคือ

$$\hat{Y} = b'\underline{X}$$
$$= b_1X_1 + b_2X_2$$
$$Y_1 = 0.42959X_1 + 5.46675X_2$$
$$Y_2 = 0.32936X_1 + 4.68157X_2$$

จากกนั้นตรวจสอบสมมติฐานของการวิเคราะห์จำแนกกลุ่ม โดยใช้ตัวสถิติทดสอบ Hotelling's T^2

สมมติฐานการทดสอบ

$H_0: \underline{\mu}_1 = \underline{\mu}_2$

$H_1: \underline{\mu}_1 \neq \underline{\mu}_2$

หรือ H_0 : ค่าเฉลี่ยของตัวแปรอิสระทุกตัวในกลุ่มที่ 1 เท่ากับกลุ่มที่ 2

H_1 : ค่าเฉลี่ยของตัวแปรอิสระทุกตัวในกลุ่มที่ 1 ไม่เท่ากับกลุ่มที่ 2

ซึ่ง df = $p, (n_1 + n_2 - p - 1)$

= $2, (12+ 12 - 2 - 1) = 2, 21$

เมื่อ $\overline{Y}_1 = \underline{b}'\overline{\underline{X}}_1 = 124.1769$

$\overline{Y}_2 = \underline{b}'\overline{\underline{X}}_2 = 104.02986$

ดังนั้น $D^2 = \overline{Y}_1 - \overline{Y}_2$

= 124.1769 – 104.02986

= 20.14704

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2$$

$$= \frac{(12)(12)}{12+12}(20.14704)$$

$$= 120.88224$$

โดย $$F = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

$$= \frac{12+12-2-1}{2(12+12-2)}(120.88224)$$

$$= 57.6938$$

เปิดตาราง F ที่องศาเสรี 2 และ 21 ตามลำดับ ณ ระดับนัยสำคัญ 0.05 จะได้ค่าตัวสถิติ $F_{(2,21)} = 3.47$ ดังนั้นจึงปฏิเสธ H_0 ซึ่งแสดงว่าค่าเฉลี่ยของตัวแปรอิสระทุกตัวในกลุ่มที่ 1 ไม่เท่ากับกลุ่มที่ 2 อย่างมีนัยสำคัญทางสถิติ หรือสมการที่สร้างสามารถนำมาใช้ในการจำแนกกลุ่มได้

จากตัวอย่างข้างต้น หากต้องการพยากรณ์กลุ่มตัวอย่างของหน่วยใหม่โดยใช้สมการจำแนกกลุ่มที่สร้างไว้ได้ดังนี้

เมื่อ $X_1 = 60$ และ $X_2 = 18$ นำมาแทนค่าลงในสมการ

จะได้ $$Y_1 = 0.42959X_1 + 5.46675X_2$$

$$= 0.42959(60) + 5.46675(18)$$

$$= 124.1769$$

$$Y_2 = 0.32936X_1 + 4.68157X_2$$

$$= 0.32936(60) + 4.68157(18)$$

$$= 104.02986$$

จากเกณฑ์การจำแนกกลุ่มแบบ Simple Classification Function พบว่า Y_1 มีค่ามากกว่า Y_2 แสดงว่าควรจัดหน่วยตัวอย่างใหม่นี้ให้อยู่ในกลุ่มที่ 1 หรือกลุ่มผู้ที่ความสนใจในเครื่องตัดหญ้าแบบมีผู้ควบคุม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การวิเคราะห์การถดถอยลอจิสติก (Logistic Regression Analysis)

การวิเคราะห์การถดถอยลอจิสติกเป็นวิธีการทางสถิติอีกวิธีหนึ่งที่ถูกใช้ในการพยากรณ์โอกาสที่จะเกิดขึ้นของเหตุการณ์ โดยตัวแปรอิสระอาจจะเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงกลุ่ม หรืออาจจะมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงกลุ่มก็ได้ ส่วนตัวแปรตามเป็นตัวแปรเชิงกลุ่ม ซึ่งการวิเคราะห์การถดถอยลอจิสติกต่างจากการวิเคราะห์จำแนกประเภทที่การวิเคราะห์การถดถอยลอจิสติกไม่มีข้อสมมติเบื้องต้นเกี่ยวกับการแจกแจงของตัวแปรอิสระและเมทริกซ์ค่าความแปรปรวนร่วมของแต่ละกลุ่ม ในขณะที่การวิเคราะห์การถดถอยลอจิสติกมีข้อสมมติเบื้องต้นดังกล่าว (กัลยา วานิชย์บัญชา, 2552)

2.3.1 วัตถุประสงค์ของการวิเคราะห์การถดถอยลอจิสติก

2.3.1.1 เพื่อศึกษาความสัมพันธ์ของตัวแปรอิสระที่มีโอกาสที่จะเกิดเหตุการณ์ พร้อมทั้งศึกษาระดับความสัมพันธ์ของตัวแปรอิสระแต่ละตัว

2.3.1.2 เพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจจากสมการที่เหมาะสม หรือใช้สมการโดยการเลือกตัวแปรอิสระที่เหมาะสม เพื่อทำให้เปอร์เซ็นต์ของความถูกต้องในการพยากรณ์มีค่าสูงสุด

2.3.2 ประเภทของการวิเคราะห์การถดถอยลอจิสติก

การวิเคราะห์การถดถอยลอจิสติกแบ่งออกเป็น 2 ประเภท คือ

2.3.2.1 การวิเคราะห์การถดถอยลอจิสติกทวิภาค (Binary Logistic Regression Analysis)

การวิเคราะห์การถดถอยลอจิสติกทวิภาคจะใช้เมื่อตัวแปรตาม Y เป็นตัวแปรเชิงกลุ่มที่มีค่าได้พียง 2 ค่า (Dichotomous Variable) คือ 0 กับ 1 ดั้งนั้นตัวแปรตาม Y จะมีการแจกแจงแบบแบร์นูลลี (Bernoulli Distribution) ซึ่งเป็นวิธีการหนึ่งในการพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจจากสมการที่เหมาะสม โดยการเลือกตัวแปรอิสระที่เหมาะสมเพื่อทำให้เปอร์เซ็นต์ของความถูกต้องในการพยากรณ์มีค่าสูงสุด ซึ่งการวิเคราะห์การถดถอยลอจิสติก จะมีข้อสมมติเบื้องต้นน้อยกว่าการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ข้อสมมติเบื้องต้นของการวิเคราะห์การถดถอยลอจิสติกเป็นดังนี้

1) ตัวแปรอิสระอาจจะเป็นข้อมูลที่มีได้เพียง 2 ค่า (Dichotomous Variable) คือ 0 กับ 1 เท่านั้น หรือเป็นสเกลอันตรภาค (Interval Scale) หรือสเกลอัตราส่วน (Ratio Scale)

2) ค่าคาดหวังของค่าคลาดเคลื่อนเป็นศูนย์ หรือ $E(e) = 0$

3) e_i และ e_j เป็นอิสระกัน ; $i , j = 1, 2, 3,...,n$ (เมื่อ n คือ จำนวนข้อมูลทั้งหมด)

4) e_i และ X_i เป็นอิสระกัน ; $i = 1, 2, 3,...,n$ (เมื่อ n คือ จำนวนข้อมูลทั้งหมด)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) ตัวแปรอิสระไม่มีความสัมพันธ์กันหรือไม่มีความสัมพันธ์เชิงเส้นแบบพหุ (Multicollinearity) ซึ่งการตรวจสอบว่าตัวแปรอิสระมีความสัมพันธ์กันหรือไม่ สามารถพิจารณาได้จากค่าความคลาดเคลื่อนยินยอม (Tolerance) หรือ Variance Inflation Factor (VIF) โดยมีหลักเกณฑ์ในการพิจารณาคือ ถ้าค่า ความคลาดเคลื่อนยินยอมเข้าใกล้ 0 แสดงว่าตัวแปรอิสระนั้น ๆ มีความสัมพันธ์กัน และถ้าค่า VIF มากกว่า 10 แสดงว่าตัวแปรอิสระนั้น ๆ มีความสัมพันธ์กัน

$$\text{โดยที่ } (\text{VIP})_j = \frac{1}{1\text{-}R_j^2} \; ; j = 1,2,3,...,p$$

เมื่อ R_j^2 คือ ค่าสัมประสิทธิ์การกำหนดรูปแบบการถดถอยที่ j

หากข้อมูลตัวแปรอิสระที่นำมาศึกษามีความสัมพันธ์กัน ซึ่งมีวิธีการแก้ไขปัญหากรณีที่ตัวแปรอิสระที่จะนำมาศึกษามีความสัมพันธ์กันอยู่หลายวิธีได้แก่ ตัดตัวแปรใดตัวแปรหนึ่งที่สัมพันธ์กันมากออกไป สร้างตัวแปรใหม่จากตัวแปรเดิมที่มีความสัมพันธ์กัน และการขจัดผลของตัวแปรหนึ่งออกจากอีกตัวแปรหนึ่ง ซึ่งในงานวิจัยนี้เลือกใช้วิธีการตัดตัวแปรใดตัวแปรหนึ่งที่สัมพันธ์กันมากออกไป

สำหรับการวิเคราะห์การถดถอยลอจิสติกทวิภาค จะแบ่งออกเป็น 2 กรณี คือ

1) กรณีที่มีตัวแปรอิสระ 1 ตัว จะพบว่า สมการการถดถอยเชิงเส้นอย่างง่ายหรือสมการที่แสดงความสัมพันธ์ระหว่าง X กับ Y จะอยู่ในรูปแบบเชิงเส้นดังนี้

$$Y = \beta_0 + \beta_1 X_1 + e$$

$$\text{หรือ } E(Y) = \beta_0 + \beta_1 X_1 + e \quad \text{โดยที่ } -\infty < E(Y) < \infty$$

สำหรับการวิเคราะห์การถดถอยลอจิสติกนั้น เมื่อ Y มีได้เพียง 2 ค่า จะพบว่าความสัมพันธ์ระหว่าง X และ Y ไม่ได้อยู่ในรูปเชิงเส้น แต่จะอยู่ในรูปดังนี้

$$E(Y) = \frac{e^{\beta_0+\beta_1 X_i}}{1+e^{\beta_0+\beta_1 X_i}} \quad \text{โดยที่ } 0 \le Y \le 1$$

หรือ $E(Y) = P(\text{เกิดเหตุการณ์ที่น่าสนใจ})$

2) กรณีที่มีตัวแปรอิสระมากกว่า 1 ตัว เมื่อมีตัวแปรอิสระมากกว่า 1 ตัว หรือมีตัวแปรอิสระหลายตัว จะได้ว่า

$$P(\text{เกิดเหตุการณ์}) = \frac{e^{\beta_0+\beta_1 X_1+....+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+....+\beta_p X_p}}$$

P(ไม่เกิดเหตุการณ์) = 1 – P(เกิดเหตุการณ์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่งการวิเคราะห์การถดถอยลอจิสติกนั้น ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระไม่ได้อยู่ในรูปเชิงเส้น จึงมีการปรับให้ความสัมพันธ์อยู่ในรูปเชิงเส้นโดยให้

$$\text{Odds Ratio} = \frac{P(Y=1)}{1 - P(Y=1)}$$

$$= \frac{e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}}{\frac{1+e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}}{1}}$$

$$= e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}$$

เมื่อออดส์ (Odds) หรือ อัตราส่วนออดส์ (Odd Ratio) แสดงถึงโอกาสที่จะเกิดเหตุการณ์เป็นกี่เท่าของโอกาสที่จะเกิดเหตุการณ์มากกว่าการไม่เกิดเหตุการณ์
เมื่อนำสมการอัตราส่วนออดส์มาหาค่าล็อก (log) จะได้

$$\log \text{Odds} = \log\left(e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}\right)$$

$$= \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$$

ซึ่งสมการล็อกออดส์อยู่ในรูปความสัมพันธ์เชิงเส้น เรียกว่าฟังก์ชันตอบสนองลอจิต (Logit Response Function)

ถ้าอัตราส่วนออดส์เท่ากับ 0.4 แสดงว่า ค่าโอกาสที่จะเกิดเหตุการณ์ที่สนใจเป็น 0.4 เท่าของโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

ถ้าอัตราส่วนออดส์มากกว่า 1 แสดงว่า ค่าโอกาสที่จะเกิดเหตุการณ์ที่สนใจมากกว่าโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

ถ้าอัตราส่วนออดส์น้อยกว่า 1 แสดงว่า ค่าโอกาสที่จะเกิดเหตุการณ์ที่สนใจน้อยกว่าโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

2.3.2.2 การวิเคราะห์การถดถอยลอจิสติกมัลติโนเมียล (Multinomial Logistic Regression Analysis)

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตามที่เป็นตัวแปรเชิงกลุ่มที่มีค่ามากกว่า 2 กลุ่ม โดยมีข้อสมมติเบื้องต้นดังนี้

1. ตัวแปรอิสระบางตัวอาจเป็นตัวแปรเชิงปริมาณ หรือบางตัวอาจเป็นตัวแปรเชิงกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ตัวแปรตามต้องเป็นตัวแปรเชิงกลุ่มที่มีค่ามากกว่า 2 ค่า

3. อัตราส่วนออดส์ของตัวแปรเชิงกลุ่ม 2 ค่า จะต้องเป็นอิสระกับค่าอื่น ๆ

ซึ่งคุณสมบัติที่สำคัญของการวิเคราะห์การถดถอยลอจิสติกมัลติโนเมียลคือ เมื่อตัวแปรตามมี 3 กลุ่มย่อยแล้ว ตัวแบบที่จะใช้จะมี 2 ตัวแบบ ในทำนองเดียวกันถ้าตัวแปรตามมี 4 กลุ่มย่อยแล้ว ตัวแบบที่จะใช้จะมี 3 ตัวแบบ เนื่องจากอีก 1 กลุ่ม จะใช้สำหรับกลุ่มอ้างอิง (Reference Group)

สำหรับการวิจัยในครั้งนี้จะใช้เทคนิคการวิเคราะห์การถดถอยลอจิสติกทวิมาใช้ในการวิเคราะห์ข้อมูลที่มีตัวแปรอิสระมากกว่า 1 ตัว มาจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตของผู้ป่วยเบาหวาน ชนิดที่ 2 กรณีศึกษา : ผู้ป่วยโรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์

2.3.3 ขั้นตอนการวิเคราะห์การถดถอยลอจิสติก

2.3.3.1 เลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตาม โดยที่ตัวแปรอิสระอาจมีได้มากกว่า 1 ตัว

2.3.3.2 ตรวจสอบหาค่าที่ผิดปกติของตัวแปรอิสระแต่ละตัว

2.3.3.3 สร้างสมการ Logistic Response Function แล้วตรวจสอบความเหมาะสมของสมการโดยพิจารณาจากค่า Pseudo R^2, ค่า Wald Statistic และค่า Hosmer Lemeshow Goodness of Fit Test

2.3.3.4 ตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์การถดถอยลอจิสติก

2.3.3.5 กรณีที่ต้องการจำแนกกลุ่มรายใหม่ ว่าจะเกิดเหตุการณ์ที่สนใจหรือไม่ เช่น ผู้ป่วยโรคเบาหวาน ชนิดที่ 2 มีโอกาสจะเกิดภาวะแทรกซ้อนทางไตหรือไม่ โดยใช้สมการ $\hat{P}$ ในการจำแนกกลุ่มหรือประมาณค่าความน่าจะเป็นของการเกิดเหตุการณ์ โดยอาศัยตัวแปรอิสระ

ถ้า $\hat{P}$(เหตุการณ์) < 0.5 จะได้ Y = 0 หรือไม่เกิดเหตุการณ์

$\hat{P}$(เหตุการณ์) ≥ 0.5 จะได้ Y = 1 หรือเกิดเหตุการณ์

ค่า 0.5 เป็นค่าความน่าจะเป็นที่ใช้เป็นคะแนนการตัด (Cutting Score) ที่ผู้วิจัยกำหนดขึ้นมาเอง ซึ่งอาจจะกำหนดเป็น 0.6 หรือ 0.4 หรือค่าอื่นๆ ก็ได้ แล้วแต่ผู้วิจัยเห็นสมควร แต่โดยทั่วไปแล้วมักนิยมกำหนดเป็น 0.5

2.3.4 การประมาณค่าสัมประสิทธิ์การถดถอยลอจิสติก

การประมาณค่าสัมประสิทธิ์การถดถอยลอจิสติก ในงานวิจัยนี้ ผู้วิจัยเลือกใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method) โดยเริ่มจากการสร้างฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) ก่อน หลังจากนั้นจึงทำการสร้างสมการภาวะน่าจะเป็นต่อไป ซึ่งมีขั้นตอนวิธีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$L = \prod_{i=1}^{n} p^{Y_i}(1-p)^{1-Y_i}$$

$$= \prod_{i=1}^{n}\left\{\frac{e^{(\beta_0+\beta_1X_1+\ldots+\beta_pX_p)}}{1+e^{(\beta_0+\beta_1X_1+\ldots+\beta_pX_p)}}\right\}^{Y_i}\left\{1-\frac{e^{(\beta_0+\beta_1X_1+\ldots+\beta_pX_p)}}{1+e^{(\beta_0+\beta_1X_1+\ldots+\beta_pX_p)}}\right\}^{1-Y_i}$$

P {ไม่เกิดเหตุการณ์} = P{Y = 0} = 1 - p

$$= 1-\frac{e^{(\beta_0+\beta_1X_1+\ldots+\beta_pX_p)}}{1+e^{(\beta_0+\beta_1X_1+\ldots+\beta_pX_p)}}$$

$$= \frac{1+e^{Z}-e^{Z}}{1+e^{Z}}$$

$$= \frac{1}{1+e^{Z}}$$

โดยที่ $Z = e^{(\beta_0+\beta_1X_1+\ldots+\beta_pX_p)}$

จะได้ $L = \prod_{i=1}^{n}\left\{\frac{e^{Z}}{1+e^{Z}}\right\}^{Y_i}\left\{\frac{1}{1+e^{Z}}\right\}^{1-Y_i}$

หาค่าล็อกของสมการ L จะได้ $\log_e(L)$ ซึ่งเรียกว่า ฟังก์ชันภาวะน่าจะเป็นแบบล็อก (log - likelihood function) ดังนี้

$$\log_e(L) = \ln L = \sum_{i=1}^{n}\left(Y_i\ln\left[p(Y_i)\right]+\left(1-Y_i\right)\ln\left[1-p\left(Y_i\right)\right]\right)$$

สำหรับการประมาณ β_0 และ β_1 จะใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method) โดยประมาณค่า β_0 และ β_1 ที่ทำให้ ln L มีค่ามากที่สุด ใช้การหาอนุพันธ์บางส่วน (Partial Derivative) เทียบกับ β_0 และ β_1 แล้วให้เท่ากับศูนย์ อย่างไรก็ตาม ไม่สามารถหาค่า β_0 และ β_1 ได้โดยตรง เนื่องจากสมการไม่ได้อยู่ในรูปเชิงเส้น จึงใช้เทคนิคการทำซ้ำ (Iterative techniques) ดังนั้นจึงใช้โปรแกรมสำเร็จรูปทางสถิติในการประมาณค่า β_0 และ β_1 โดยที่ค่าประมาณของ β_0 คือ b_0 และ β_1 คือ b_1 และเรียก b_0 และ b_1 ว่า ตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.5 การทดสอบสมมติฐานของค่าสัมประสิทธิ์การถดถอยลอจิสติก

เป็นการทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระแต่ละตัว โดยใช้ตัวสถิติทดสอบของ Wald ซึ่งเป็นการเปรียบเทียบ β_i กับค่าความคลาดเคลื่อนมาตรฐานของ β_i โดยมีวิธีดังนี้

สมมติฐานการทดสอบ

H_0 : $\beta_i = 0$; i = 1, 2,...,p

H_1 : $\beta_i \neq 0$

หรือ H_0 : ตัวแปรอิสระแต่ละตัวในสมการไม่มีความสัมพันธ์กับตัวแปรตาม

H_1 : ตัวแปรอิสระแต่ละตัวในสมการมีความสัมพันธ์กับตัวแปรตาม

ตัวสถิติทดสอบ Wald

$$W = \left[\frac{b_i}{SE(b_i)}\right]^2$$

เมื่อ b_i คือ ค่าสัมประสิทธิ์ของตัวแปรอิสระ

$SE(b_i)$ คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าสัมประสิทธิ์ของตัวแปรอิสระ

ตัวสถิติทดสอบ Wald จะมีการแจกแจงไคกำลังสอง (χ^2) โดยมีองศาเสรี (Degree of Freedom : df) เท่ากับ 1

จากนั้นนำค่าตัวสถิติทดสอบ Wald ที่ได้จากการคำนวณมาเปรียบเทียบกับค่าไคกำลังสอง (χ^2) ที่ระดับนัยสำคัญ(α) 0.05 และองศาเสรี เท่ากับ 1 ถ้าค่าตัวสถิติทดสอบ Wald มีค่ามากกว่าค่าไคกำลังสองจะปฏิเสธ H_0 (ยอมรับ H_1) หรือสรุปได้ว่าตัวแปรอิสระแต่ละตัวในตัวแบบมีความสัมพันธ์กับตัวแปรตาม

2.3.6 การตรวจความเหมาะสมของสมการการวิเคราะห์การถดถอยลอจิสติก

ตัวแบบการวิเคราะห์การถดถอยลอจิสติกที่ได้จำเป็นต้องมีการตรวจความเหมาะสมก่อนที่จะนำไปใช้งานจริง โดยการตรวจสอบความเหมาะสมของตัวแบบมีหลายวิธี ได้แก่ Pearson Chi – Square, Hosmer and Lemeshow และ Likelihood Ratio ซึ่งในงานวิจัยนี้ ผู้วิจัยเลือกใช้วิธี Hosmer and Lemeshow โดยมีวิธีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมมติฐานการทดสอบ

H_0 : สมการการวิเคราะห์การถดถอยลอจิสติกเหมาะสมกับข้อมูล

H_1 : สมการการวิเคราะห์การถดถอยลอจิสติกไม่เหมาะสมกับข้อมูล

ตัวสถิติทดสอบ Hosmer and Lemeshow

$$H\text{–}L = \sum_{i=1}^{10} \frac{\left(\sum_j y_{ij} - \sum_j \hat{p}_{ij}\right)^2}{\left(\sum_j \hat{p}_{ij}\right)\left(1 - \frac{\sum_j \hat{p}_{ij}}{n_i}\right)}$$

เมื่อ Y_{ij} คือ ค่าของตัวแปรตามของหน่วยตัวอย่างที่ j ในกลุ่มที่ i ; i = 1, 2,...,p และ j = 1, 2,...,n_i

n_i คือ ขนาดตัวอย่างของกลุ่มที่ i

$\hat{p}_{ij}$ คือ ค่าความน่าจะเป็นของหน่วยตัวอย่างที่ j ในกลุ่มที่ i

โดยตัวสถิติทดสอบ Hosmer and Lemeshow มีการแจกแจงโดยประมาณไคกำลังสองที่องศาเสรี p - 2 ซึ่ง Hosmer and Lemeshow ได้แบ่งข้อมูลออกเป็น 10 กลุ่มย่อย ดังนั้น p = 10

จากนั้นนำค่าสถิติทดสอบ Hosmer and Lemeshow ที่ได้จากการคำนวณมาเปรียบเทียบกับค่าไคกำลังสอง (χ^2) ถ้าค่าตัวสถิติทดสอบ Hosmer and Lemeshow มีค่ามากกว่าค่าไคกำลังสองจะปฏิเสธ H_0 หรือสรุปได้ว่าสมการการวิเคราะห์การถดถอยลอจิสติกไม่เหมาะสมกับข้อมูล

2.3.7 ตัวอย่างการวิเคราะห์การถดถอยลอจิสติก

จากข้อมูลการศึกษาการซื้อเครื่องสำอางยี่ห้อ A กับรายได้ของลูกค้า (กัลยา วานิชย์บัญชา, 2552) โดยกำหนดตัวแปร

X_1 แทนรายได้ (หน่วยบาท)

X_2 แทนอายุ (หน่วยปี)

Y แทนตัวแปรตามที่มีค่า 1 (ลูกค้าซื้อเครื่องสำอาง A) และ 0 (ลูกค้าไม่ซื้อเครื่องสำอาง A)

โดยเก็บข้อมูลของลูกค้าเพศหญิง 60 ราย ซึ่งจากข้อมูลมีผลการคำนวณดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.2 ตัวอย่างข้อมูลรายได้และอายุ

การซื้อ	อายุ (ปี)	รายได้ (บาท)	การซื้อ	อายุ (ปี)	รายได้ (บาท)	การซื้อ	อายุ (ปี)	รายได้ (บาท)
1	40	15,000	0	30	19,000	1	29	32,000
1	45	27,000	0	35	22,000	1	32	12,000
0	20	8,000	1	40	12,000	1	51	51,000
1	30	17,000	0	20	10,000	1	60	20,000
1	55	45,000	0	19	12,000	0	18	10,000
0	15	3,000	0	22	20,000	0	19	12,000
0	18	5,000	0	19	15,000	1	25	9,000
1	20	9,000	1	45	40,000	0	30	9,000
0	25	15,000	1	22	20,000	1	41	30,000
0	40	27,000	1	29	25,000	0	30	20,000
1	30	32,000	1	37	37,000	0	21	18,000
0	27	19,000	1	43	32,000	0	26	31,000
1	35	30,000	1	45	28,000	1	47	43,000
1	42	32,000	1	40	19,000	1	38	32,000
1	53	17,000	0	32	25,000	1	39	29,000
0	35	20,000	1	18	9,000	0	30	35,000
0	18	5,000	1	25	17,000	0	40	41,000
1	27	11,000	1	48	27,000	0	18	17,000
0	19	4,000	1	51	31,000	0	27	20,000
0	30	7,000	1	45	29,000	0	31	30,000

จากการประมาณผลด้วยโปรแกรม SAS ที่ใช้ตัวแบบลอจิตมีผลลัพธ์ดังนี้

ตารางที่ 2.3 ผลการประมาณการศึกษาการซื้อเครื่องสำอางยี่ห้อ A กับรายได้ของลูกค้า

ตัวแปร	df	Estimate	Standard Error	Wald Chi - Square	Pr > ChiSq
Intercept	1	-4.5680	1.2298	13.7956	0.0002
อายุ	1	0.1626	0.052	9.7833	0.0018
รายได้	1	-0.00002	0.000042	0.1811	0.6704

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ P(เกิดเหตุการณ์) $= \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2}}$

$= \frac{e^{-4.568+0.1626X_1-0.0002X_2}}{1+e^{-4.568+0.1626X_1-0.0002X_2}}$

ซึ่ง Odds Ratio $= e^{\beta_0+\beta_1X_1+\beta_2X_2}$

$= e^{-4.568+0.1626X_1-0.0002X_2}$

และ log odds $= \log(e^{\beta_0+\beta_1X_1+\beta_2X_2})$

$= \beta_0+\beta_1x_1+\beta_2x_2$

$= -4.5680 + 0.1626X_1 - 0.00002X_2$

ถ้าต้องการพยากรณ์ลูกค้าเพศหญิงที่มีอายุ 32 ปี และมีรายได้ 12,000 บาท จะได้ว่า

log odds $= -4.5680 + 0.1626(32) - 0.00002(12,000) = 0.3952$

จะได้ P(เกิดเหตุการณ์) $= \frac{e^{0.3952}}{1+e^{0.3952}} = 0.5975$

สรุปได้ว่า ลูกค้าเพศหญิงที่มีอายุ 32 ปี และมีรายได้ 12,000 บาท มีโอกาสที่จะซื้อเครื่องสำอาง A อยู่ 0.5975 หรือ 59.75%

2.4 ต้นไม้ตัดสินใจ (Decision Tree)

เทคนิคต้นไม้ตัดสินใจ คือ แบบจำลองทางคณิตศาสตร์ เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) สามารถสร้างแบบจำลองการจัดหมวดหมู่ (Clustering) ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้า (Training set) ได้โดยอัตโนมัติและสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย

โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้ว ผลลัพธ์” เช่น (พยูน พาณิชย์กุล, 2548)

“If Income = High and Married = No THEN Risk = Poor”

“ถ้าหากรายได้สูงและยังไม่ได้แต่งงาน แล้วจะมีฐานะยากจน”

“If Income = High and Married = Yes THEN Risk = Good”

“ถ้าหากรายได้สูงและแต่งงาน แล้วจะมีฐานะร่ำรวย”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.1 ส่วนประกอบของต้นไม้ตัดสินใจ

2.4.1.1 โหนด (Node) คือ คุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก (Root Node)

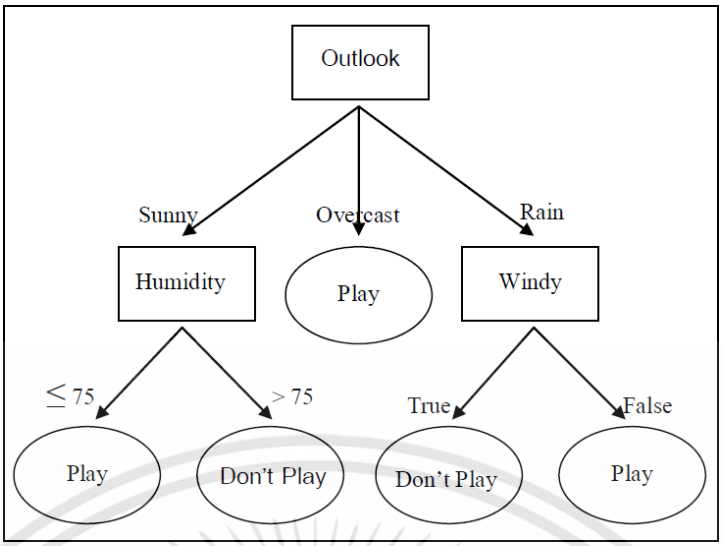
2.4.1.2 กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด

2.4.1.3 ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงตัวอย่างของต้นไม้ตัดสินใจได้ดังตารางและรูปดังต่อไปนี้

ตารางที่ 2.4 ตัวอย่างชุดข้อมูลการเล่นเทนนิส

Outlook	Temperature	Humidity	Windy	Play / Don’t Play
Sunny	85	85	False	Don't Play
Sunny	80	90	True	Don't Play
Overcast	83	78	False	Play
Rain	70	96	False	Play
Rain	68	80	False	Play
Rain	65	70	True	Don't Play
Overcast	64	65	True	Play
Sunny	72	95	False	Don't Play
Sunny	69	70	False	Play
Rain	75	80	False	Play
Sunny	75	70	True	Play
Overcast	72	90	True	Play
Overcast	81	75	False	Play
Rain	71	80	True	Don't Play

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 ตัวอย่างการตัดสินใจจากชุดข้อมูลการเล่นเทนนิส

2.4.2 การสร้างต้นไม้ตัดสินใจ

หลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจ เป็นการสร้างในลักษณะจากบนลงล่าง (Top-Down) คือเริ่มจากการสร้างรากของต้นไม้ก่อนแล้วจึงแตกกิ่งไปจนถึงใบ โดยแสดงขั้นตอนการสร้างต้นไม้ตัดสินใจได้ดังนี้

2.4.2.1 ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึก (Training Set)

2.4.2.2 ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น

2.4.2.3 ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่ จะต้องวัดค่าเกน (Gain) ของแต่ละคุณลักษณะ (Attribute) เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกคุณลักษณะที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่างๆ ได้ดีที่สุด โดยคุณลักษณะที่มีค่าเกนมากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือคุณลักษณะที่ใช้ในการตัดสินใจ โดยแสดงในรูปของโหนดบนต้นไม้

2.4.2.4 กิ่งของต้นไม้ ถูกสร้างขึ้นจากค่าต่างๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่างๆ ที่สร้างขึ้น

2.4.2.5 ทำการวนซ้ำเพื่อหาคุณลักษณะที่มีค่าเกนมากที่สุด สำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำคุณลักษณะนี้มาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่คุณลักษณะที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีก สำหรับโหนดในระดับต่อๆ ไป

2.4.2.6 ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อย ๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้เป็นจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.3 การคำนวณค่า Information Gain (ขจรศักดิ์ ศรีอ่อน, 2552)

ต้นไม้ตัดสินใจเป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดยต้นไม้ตัดสินใจจะมีลักษณะคล้ายโครงสร้างต้นไม้ ที่แต่ละโหนดแสดงคุณลักษณะในการสร้างต้นไม้ตัดสินใจ ปัญหาสำคัญที่ต้องพิจารณาคือ ควรจะตัดสินใจเลือกคุณลักษณะใดมาทำหน้าที่เป็นโหนดราก ในแต่ละขั้นตอนของการสร้างต้นไม้และต้นไม้ย่อย (Subtree) ของต้นไม้ตัดสินใจ เกณฑ์ที่ใช้ช่วยประกอบการเลือกคุณลักษณะคือการคำนวณค่ามาตรฐานเกน (Gain Criterion) ซึ่งเป็นค่าที่บ่งบอกว่าคุณลักษณะนั้นสามารถจำแนกกลุ่มของข้อมูลได้ดีเพียงใด โดยทดลองเลือกแต่ละคุณลักษณะที่เป็นไปได้จากชุดข้อมูลมาทำหน้าที่เป็นโหนดราก ถ้าคุณลักษณะใดให้ค่าเกนสูงที่สุด แสดงว่าคุณลักษณะนั้นสามารถจำแนกกลุ่มของข้อมูลได้ดีที่สุด การใช้ค่า Information Gain จะช่วยลดจำนวนครั้งของการทดสอบในการแยกแยะข้อมูล อีกทั้งยังรับประกันว่าต้นไม้ตัดสินใจที่ได้ไม่มีความซับซ้อนมากจนเกินไป

ซึ่งค่า Information Gain นั้นสามารถคำนวณได้จากสมการดังสมการดังนี้

$$I(S_1,S_2,....,S_n) = -\sum_{i=1}^{n}\frac{S_i}{S}\log_2\frac{S_i}{S}$$

เมื่อ S เป็นเซตของข้อมูลซึ่งประกอบด้วยข้อมูล S ระเบียน

n เป็นจำนวนกลุ่มทั้งหมดที่ต่างกันของข้อมูลชุดนั้น

C_i แทนกลุ่มในลำดับ ที่ i โดย ที่ i มีค่าระหว่าง 1 ถึง n

S_i แทนจำนวนข้อมูลที่เป็นสมาชิกของ S และอยู่ในกลุ่ม C_i

ค่าคาดคะเนของข้อมูล (Entropy) ของ A ซึ่งมีค่าของคุณลักษณะเป็น $(a_1,a_2,a_3,...,a_V)$ หาได้ในสมการดังนี้

$$E(A)=\sum_{j=1}^{v}\frac{(S_{1j}+...+S_{nj})}{S}I(S_{1j},S_{2j},...,S_{nj})$$

S_{ij} แทน จำนวนข้อมูลที่เป็นสมาชิกของ S ในกลุ่ม C_i จากการแบ่งข้อมูลด้วยค่าที่เป็นไปได้ของคุณลักษณะ A

ดังนั้นจะสามารถพิจารณาค่ามาตรฐานเกนได้ดังสมการดังนี้

$$Gain(A) = I(S_1,S_2,...,S_n) - E(A)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสร้างต้นไม้ตัดสินใจนั้นมีวิธีหลายวิธีด้วยกัน โดยที่วิธีที่นิยมอย่างมาก คือ ID3 และ C4.5 เนื่องจากใช้งานและทำความเข้าใจง่าย และได้ผลลัพธ์ที่ดีในการนำไปใช้งาน

2.4.4 ขั้นตอนวิธี ID3 (ID3 Algorithm)

ID3 (Iterative Dichotomiser 3) เป็นขั้นตอนวิธีพื้นฐานที่ใช้ในการสร้างการตัดสินใจแบบโครงสร้างต้นไม้ที่ใช้หลักการของการใช้ทฤษฎีข่าวสาร (Information Theory) และค่าที่วัดได้จะนามาใช้ตัดสินใจว่าจะใช้ตัวแปรใดใช้ในการทำนาย หรือแบ่งประเภทของข้อมูล โดยที่ชุดตัวอย่าง (Sample) คือชุดของข้อมูลที่ใช้ในการเรียนรู้ (Training Sample) ตัวแปรเป้าหมาย (Target Attribute) คือตัวแปรที่นำค่าไปใช้ในการทำนายผลในโครงสร้างต้นไม้และคุณลักษณะ คือตัวแปรอื่นๆ ที่ใช้ในการสร้างโหนดในต้นไม้และไม่ใช่ตัวแปรเป้าหมาย (Target Attribute)

2.4.5 ขั้นตอนวิธี C4.5 (C4.5 Algorithm)

C4.5 หรือ J48 เป็นส่วนขยายของขั้นตอนวิธี ID3 ถูกพัฒนาโดย Ross Quinlan ใช้ในการสร้างแผนภาพต้นไม้เพื่อประกอบการตัดสินใจสำหรับใช้ในการจำแนกประเภทข้อมูล มีการใช้ค่าเกน (Gain) และค่าคาดคะเนของข้อมูลเช่นเดียวกับ ID3 แต่จะมีส่วนที่เพิ่มเติมจาก ID3 ที่สำคัญดังนี้

2.4.5.1 สามารถหลีกเลี่ยงการสร้างโครงสร้างต้นไม้ที่ใหญ่เกินไป เนื่องจากมีข้อมูลจำนวนมาก อย่างไรก็ตามขึ้นอยู่กับการกำหนดความลึก เมื่อมีการเจริญเติบของต้นไม้ตัดสินใจ

2.4.5.2 ความผิดพลาดลดลง เพราะมีการตัดทอนความผิดพลาดออกไป (Pruning node)

2.4.5.3 มีการสร้างกฎหลังการตัดทอนข้อมูลที่ผิดพลาดออก

2.4.5.4 สามารถใช้งานได้ทั้งข้อมูลแบบต่อเนื่อง (Continuous Data) ที่เป็นตัวเลขได้ เช่น ค่าอุณหภูมิ จำนวนเงิน เป็นต้น และแบบไม่ต่อเนื่อง (Discrete Data) โดยในส่วนของข้อมูลแบบต่อเนื่องนั้น C4.5 จะสร้างจุดแบ่งแยก (Threshold) คุณลักษณะนั้นออกเป็น 2 ส่วน คือส่วนที่มีค่ามากกว่ากับน้อยกว่าเท่ากับค่าที่ใช้ในการสร้างจุดเริ่ม

2.4.5.5 การเลือกคุณลักษณะที่วัดการเลือกให้เหมาะสม

2.4.5.6 สามารถใช้กับชุดข้อมูลทดสอบ (Training Data) ที่มีค่าข้อมูลขาดหายได้ (Missing Data) โดยจะแทนค่าด้วย “?” และไม่นำค่านั้นมาคำนวณในกฎของความรู้จากทฤษฎีสารสนเทศ

2.4.5.7 สามารถใช้สำหรับคุณลักษณะกับ Costs ที่แตกต่างกันได้ มีการยอมรับประสิทธิภาพกระบวนการคำนวณที่ปรับปรุง

ขั้นตอนวิธี C4.5 ใช้หลักการเลือกข้อมูลตามลำดับของค่า Information Gain สูงที่สุดเป็นโหนดเริ่มต้น และโหนดถัดไปจะมีค่าเป็น Information Gain ลดหลั่นกันไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม

2.5.1 การวัดค่าความถูกต้อง (Accuracy)

ในการประมวลผลข้อมูลนั้นจำเป็นต้องมีการวัดประสิทธิภาพของการจำแนกข้อมูล โดยใช้การวัดค่าความถูกต้อง ซึ่งสามารถคำนวณได้จากตารางเมทริกซ์ความสับสน (Confusion Matrix)

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{(TP+FP+FN+TN)}} \times 100\%$$

ตารางที่ 2.5 เมทริกซ์ความสับสน

	ค่าที่ทำนาย	
ค่าที่แท้จริง	กลุ่ม 1	กลุ่ม 2
กลุ่ม 1	True Positive (TP)	False Positive (FP)
กลุ่ม 2	False Negative (FN)	True Negative (TN)

โดยที่ True Positive (TP) คือ จำนวนข้อมูลที่จำแนกถูกว่าเป็นกลุ่มที่ 1

True Negative (TN) คือ จำนวนข้อมูลที่จำแนกถูกว่าเป็นกลุ่มที่ 2

False Positive (FP) คือ จำนวนข้อมูลที่จำแนกผิดว่าเป็นกลุ่มที่ 2 ซึ่งค่าที่แท้จริงเป็นกลุ่มที่ 1

False Negative (FN) คือ จำนวนข้อมูลที่จำแนกผิดว่าเป็นกลุ่มที่ 1 ซึ่งค่าที่แท้จริงเป็นกลุ่มที่ 2

2.5.2 ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) หากยิ่งมีค่าน้อยหมายถึงจะมีความแม่นยำมากยิ่งขึ้น

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{\left(y_i - \hat{y}_i\right)^2}{n}}$$

โดยที่ y_i แทน ค่าจริง

$\hat{y}_i$ แทน ค่าพยากรณ์

n แทน ขนาดข้อมูลตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6 งานวิจัยที่เกี่ยวข้อง

พัชรี เพ็งประโคน (2539) ได้ศึกษาการเปรียบเทียบผลการวิเคราะห์ข้อมูลระหว่างวิธีวิเคราะห์จำแนกประเภทกับการวิเคราะห์การถดถอยลอจิสติก โดยใช้ตัวแปรจำแนกกลุ่มวิถีชีวิตแบบประชาธิปไตยของนิสิตนักศึกษา กลุ่มตัวอย่างเป็นนิสิตนักศึกษาชั้นปีที่ 2 - 4 ระดับปริญญาบัณฑิต จำนวน 1,056 คน จาก มหาวิทยาลัยในเขตกรุงเทพมหานคร 7 สถาบัน เครื่องมือที่ใช้ในการศึกษาเป็นแบบสอบถามวัดตัวแปรด้านภูมิหลัง ด้านสถาบันการศึกษา และด้านคุณลักษณะของนิสิตนักศึกษา วิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์จำแนกประเภทและการวิเคราะห์การถดถอยลอจิสติก ผลการวิจัยพบว่า การวิเคราะห์ข้อมูลด้วยวิธีวิเคราะห์จำแนกประเภทได้ตัวแปรที่มีผลต่อการมีวิถีชีวิตแบบประชาธิปไตยของนิสิตนักศึกษาตามลำดับดังนี้ การเข้าร่วมกิจกรรมเสริมหลักสูตร การเปิดรับสื่อสารมวลชน บรรยากาศในชั้นเรียน ส่วนการวิเคราะห์การถดถอยลอจิสติกได้ตัวแปรที่มีผลต่อการมีวิถีชีวิตแบบประชาธิปไตยของนิสิตนักศึกษาตามลำดับดังนี้ การเข้าร่วมกิจกรรมเสริมหลักสูตร การเปิดรับสื่อสารมวลชน บรรยากาศในชั้นเรียน นิสิตนักศึกษาสาขาวิศวกรรมศาสตร์ บุคลิกภาพ บิดาทำงานในบริษัทเอกชน ซึ่งสมการการวิเคราะห์จำแนกประเภทสามารถอธิบายความแปรปรวนของกลุ่มได้ร้อยละ 43.33859 มีประสิทธิภาพในการจัดกลุ่มได้ร้อยละ 82.92 ส่วนสมการการวิเคราะห์การถดถอยลอจิสติก สามารถอธิบายความแปรปรวนของการมีวิถีชีวิตแบบประชาธิปไตยได้ร้อยละ 42.6265 และมีประสิทธิภาพในการจัดกลุ่มร้อยละ 82.21 จะเห็นได้ว่าสมการการวิเคราะห์จำแนกประเภท สามารถอธิบายความแปรปรวนได้ดีกว่าสมการการวิเคราะห์การถดถอยลอจิสติกเล็กน้อย ส่วนประสิทธิภาพในการจัดกลุ่มไม่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติ

ศยามล ลำลองรัตน์ (2550) ศึกษาเปรียบเทียบการวินิจฉัยโรคตับด้วยการวิเคราะห์จำแนกกลุ่ม และการวิเคราะห์ถดถอยโลจิสติก กรณีศึกษาผู้ป่วยโรงพยาบาลฉะเชิงเทรา มีวัตถุประสงค์เพื่อศึกษาวินิจฉัยการเป็นหรือไม่เป็นโรคตับของผู้ป่วยโรงพยาบาลเมืองฉะเชิงเทราที่ได้รับการตรวจการทำงานของตับ ซึ่งแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มผู้ป่วยที่มีแนวโน้มเป็น และไม่เป็นโรคตับ โดยศึกษาเปรียบเทียบเทคนิคการวิเคราะห์ทางสถิติ 2 วิธี คือ การวิเคราะห์จำแนกกลุ่มและการวิเคราะห์ถดถอยโลจิสติก เกณฑ์ในการเปรียบเทียบคือ ค่าอัตราการพยากรณ์ที่ผิดพลาด (Apparent Error Rate : APER) โดยวิธีที่มีประสิทธิภาพสูงสุดคือ วิธีที่มีค่าอัตราการพยากรณ์ผิดพลาดน้อยที่สุด ซึ่งผลการศึกษาพบว่า การวิเคราะห์จำแนกกลุ่ม มีความถูกต้องแม่นยำในการพยากรณ์ 97.67% หรือมีความคลาดเคลื่อนในการพยากรณ์ประมาณ 2.3% และการวิเคราะห์ถดถอยโลจิสติก มีความถูกต้องแม่นยำในการพยากรณ์ 94.33% หรือมีความผิดพลาดในการพยากรณ์ประมาณ 5.7% แต่เมื่อทดสอบความเหมาะสมของตัวแบบแล้วพบว่า ตัวแบบไม่เหมาะสมกับข้อมูล จึงสรุปได้ว่า การวิเคราะห์จำแนกกลุ่มมีประสิทธิภาพสูงกว่าการวิเคราะห์ถดถอยโลจิสติก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชลธิศา พลทองมาก (2553) งานวิจัยนี้นำเสนอการวิเคราะห์ความเสี่ยงการเป็นโรคไวรัสตับอักเสบซี (Hepatitis C : HCV) โดยใช้เทคนิคการจัดแบ่งประเภทข้อมูล (Data Classification) ด้วยหลักการต้นไม้ประกอบการตัดสินใจ (Decision Tree) และทฤษฎีเบย์เซี่ยน (Bayesian Theorem) ในการศึกษาได้รวบรวมปัจจัยเสี่ยงในการเกิดโรคไวรัสตับอักเสบซี จากงานวิจัยทางการแพทย์ และแพทย์ผู้เชี่ยวชาญ เพื่อจัดทำแบบสอบถามในการจัดเก็บข้อมูลจากกลุ่มตัวอย่าง ในเขตอำเภอเมืองจังหวัดขอนแก่น จำนวน 661 ราย โดยเป็นผู้ป่วยติดเชื้อ HCV 76 ราย และเป็นผู้ป่วยที่ไม่ติดเชื้อจำนวน 585 ราย และทำการวิเคราะห์ปัจจัยเสี่ยงโดยใช้ โปรแกรม Weka และสร้างตัวแบบเพื่อพยากรณ์ความเสี่ยงโรคไวรัสตับอักเสบซี โดยใช้โมเดลต้นไม้ตัดสินใจและทฤษฎีเบย์เซี่ยน ผลการทดลองพบว่าทฤษฎีเบย์เซี่ยน ให้ความแม่นยำ (Accuracy) มากที่สุดคือ 93.86% และวิธีการต้นไม้ตัดสินใจ ให้ความแม่นยำ 92.99% ตามลำดับ

สิริมา เยาวสิทธิ์ (2553) ศึกษาการเปรียบเทียบวิธีการทางสถิติและการโปรแกรมเชิงเป้าหมาย สำหรับจำแนกกลุ่มกำไรขาดทุนของสหกรณ์การเกษตร งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาผลการดำเนินงานของสหกรณ์การเกษตรโดยการจำแนกกลุ่มสหกรณ์ว่า สหกรณ์ใดจัดอยู่ในกลุ่มสหกรณ์ที่มีผลกาไรหรือขาดทุน โดยเปรียบเทียบวิธีการจำแนกกลุ่ม 3 วิธี คือ การวิเคราะห์จำแนกประเภท การวิเคราะห์การถดถอยโลจิสติก และการโปรแกรมเชิงเป้าหมายด้วยตัวแบบ GPMEAN และ GPMED ข้อมูลที่ใช้เป็นอัตราส่วนทางการเงินของสหกรณ์การเกษตร 12 อัตราส่วนจากกรมตรวจบัญชีสหกรณ์ตั้งแต่ปี พ.ศ. 2551-2553 จากผลการวิจัยพบว่า วิธีการวิเคราะห์จำแนกประเภทเป็นวิธีที่ให้ค่าร้อยละความถูกต้องในการจำแนกมากที่สุดคือร้อยละ 100 รองลงมาคือวิธีการวิเคราะห์การถดถอยโลจิสติก การโปรแกรมเชิงเป้าหมายด้วยตัวแบบ GPMEAN และ GPMED ซึ่งสามารถจำแนกกลุ่มสหกรณ์การเกษตรได้ถูกต้องร้อยละ 95.83, 89.81 และ 89.35 ตามลำดับ

สุริยะ ตรีราภี (2553) การศึกษาเปรียบเทียบการทำนายความเสี่ยงต่อโรคฝันผุโดยใช้การเรียนรู้เบย์อย่างง่าย ต้นไม้การตัดสินใจและโครงข่ายประสาทเทียม สำหรับข้อมูลที่ใช้ประกอบด้วยกลุ่มข้อมูลผู้ป่วยจังหวัดระยอง จำนวน 500 คน และจังหวัดเชียงใหม่ จำนวน 500 คน ผลการศึกษาพบว่าแบบจำลอง Decision Tree, Naïve Bayes และ Artificial Neural Network โดยให้ความถูกต้องในการทำนายความเสี่ยงต่อโรคฟันผุ (Mean ± Standard Error (SE)) เท่ากับ 67± 3.80, 69.75 ± 4.32 และ 76.25 ± 3.16 ตามลำดับ โดยมีความแตกต่างอย่างมีนัยสำคัญทางสถิติ และพบว่า จำนวนฟันผุเดิมที่มีอยู่เดิมและใหม่ (Secondary Caries) เป็นตัวแปรสำคัญในการทำนายความเสี่ยงต่อการเกิดโรคฟันผุได้ถูกต้องมากกว่า ฟันผุในระยะแรก (White Sport), ฟันผุแบบแข็ง (Brown Sport), ฟันผุที่ไม่มีการลุกลาม (Arrested Caries) ฟันผุที่ได้รับการถอนแล้วและฟันที่ถูกถอนออกไปเพราะฟันผุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รุจิรา ธรรมสมบัติ (2554) งานวิจัยนี้ มีวัตถุประสงค์เพื่อพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือ โดยใช้ต้นไม้ตัดสินใจ และเป็นแนวทางในการตัดสินใจเลือกใช้แพคเกจอินเทอร์เน็ตมือถือจากพฤติกรรมของผู้ใช้บริการแต่ละคน โดยการเปรียบเทียบโมเดลที่ถูกสร้างขึ้นจากอัลกอริธึมต้นไม้ตัดสินใจ คือ ID3 และ C4.5 (J48) เพื่อหาโมเดลที่มีค่าความถูกต้องมากที่สุด ก่อนนำมาพัฒนาระบบ ในการพัฒนาระบบใช้ภาษา ASP.NET ด้วย C# และฐานข้อมูล SQL Server 2008 โดยพัฒนาขึ้นในลักษณะของเว็บแอพพลิเคชั่น (Web-Based Application) และใช้โปรแกรม Weka 3.6.2 เพื่อตรวจสอบความถูกต้องของโมเดลที่ระบบได้สร้างขึ้นมา ผลที่ได้ คือ โมเดลที่ถูกสร้างจาก อัลกอริธึมต้นไม้ตัดสินใจ ID3 มีค่าความถูกต้องมากกว่า C4.5 (J48) โดยมีค่าความถูกต้อง (Correctly Classified Instances) เมื่อทดสอบกับกลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) จำนวน 1,000 ชุดเท่ากับ 92.3% และเมื่อนำอัลกอริธึมต้นไม้ตัดสินใจ ID3 ไปทดสอบกับชุดข้อมูลทดสอบ (Testing Data) จำนวน 500 ชุดให้ผลการทดสอบโดยมีค่าความถูกต้องเท่ากับ 92.2 % และเมื่อพิจารณาค่า Confusion Matrix พบว่าผลของการทำนายจากโมเดลมีจำนวนข้อมูลค่าจริง กับจำนวนข้อมูลจากการทำนายของโมเดลมีผลลัพธ์ตรงกัน ได้ค่าเฉลี่ยรวมเท่ากับ 83.06% ซึ่งเป็นค่าเฉลี่ยที่อยู่ในระดับค่อนข้างสูง สามารถนาโมเดลที่ได้ไปพัฒนาระบบต่อไป

อภิวัฒน์ อ่อนสุระทุม (2555) รายงานการศึกษาเป็นการนำเสนอการวินิจฉัยโรคไข้เลือดออกโดยใช้เทคนิคการจัดแบ่งประเภทข้อมูลด้วยหลักการต้นไม้ตัดสินใจ และเครือข่ายความเชื่อเบย์ ในการศึกษานี้ได้รวบรวมปัจจัยเสี่ยงในการเกิดโรคไข้เลือดออกจากข้อมูลการเกิดโรคที่แสดงทางคลินิก ผลตรวจจากห้องปฏิบัติการ งานวิจัยทางการแพทย์และแพทย์ผู้เชี่ยวชาญ โดยทำการเก็บข้อมูลประชากรจากผู้ป่วยที่เข้ารับการรักษาที่โรงพยาบาลสมเด็จพระยุพราชสว่างแดนดิน จังหวัดสกลนคร ย้อนหลัง 2 ปี ตั้งแต่ปี พ.ศ.2553 – 2554 และทำการทดสอบผลบนพื้นฐานวิธี K – Fold Cross Validation โดยใช้โปรแกรม WEKA และสร้างตัวแบบเพื่อวินิจฉัยโรคไข้เลือดออก โดยใช้ตัวแบบต้นไม้ตัดสินใจ และเครือข่ายความเชื่อเบย์ ผลการทดลองพบว่า ตัวแบบการตัดสินใจที่การแบ่งข้อมูลจำนวน 5-ชุด ให้ความแม่นยำมากที่สุด 98.5324% และเครือข่ายความเชื่อเบย์ที่การแบ่งข้อมูลจำนวน 5-ชุด ให้ความแม่นยำมากที่สุด 98.0867% ตามลำดับ ผลการศึกษาพบว่าตัวแบบต้นไม้ตัดสินใจมีประสิทธิภาพสูงกว่าเครือข่ายความเชื่อเบย์

กนกพัฒน์ ลักษณานุรักษ์ (2556) งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลผู้ป่วยโรคความดันโลหิตสูง โดยอธิบายถึงกระบวนการแบบปกติและกระบวนการใหม่ ซึ่งกระบวนการใหม่จะมีการเพิ่มการจัดการข้อมูลก่อนการนำข้อมูลไปประมวลผล โดยวิธีการจัดการกับข้อมูลจะมีการทำความสะอาดข้อมูลและการคัดเลือกคุณลักษณะ ซึ่งการคัดเลือกคุณลักษณะจะใช้วิธีการหาอัตราขยายร่วมกับฟังก์ชันแยกแยะ จากนั้นจะแสดงการจัดลำดับความสำคัญของแต่ละแอตตริบิวส์ และเมื่อเสร็จสิ้นกระบวนการคัดเลือกคุณลักษณะก็จะเข้าสู่ขั้นตอนการจำแนกข้อมูล โดยใช้เทคนิคการถดถอยโลจิสติก ซัพพอร์ตเวกเตอร์แมชชีน และนาอีฟเบย์ โดยเทคนิคที่ให้ค่าความถูกต้องสูงที่สุดคือการถดถอยโลจิสติก ซึ่งจะได้ค่าความถูกต้องเท่ากับ 74.75%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น และห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารนี้ทุกครั้งที่มีการนำไปใช้

และเทคนิคที่ให้ค่าความถูกต้องสูงลำดับถัดมาคือนาอีฟเบย์ เท่ากับ 74.42% ส่วนเทคนิคที่ได้ค่าความถูกต้องน้อยที่สุดคือซัพพอร์ตเวกเตอร์แมชชีน 73.43% ซึ่งจากการจำแนกข้อมูลทั้งหมดนี้ สิ่งที่ขาดไม่ได้ก็คือชุดข้อมูลที่ผ่านกระบวนการคัดเลือกคุณลักษณะแล้ว โดยแหล่งที่มาของชุดข้อมูลนั้นทางผู้วิจัยได้ขอความอนุเคราะห์ไปยังสาธารณสุขจังหวัดยโสธร เพื่อขอชุดข้อมูลผู้ป่วยโรคความดันโลหิตสูง โดยได้รับข้อมูลทั้งหมด จำนวน 606 เรคคอร์ด และจากผลการทดลองสรุปได้ว่ากระบวนการใหม่ที่มีการเพิ่มเทคนิคการคัดเลือกคุณลักษณะก่อนการจำแนกข้อมูล สามารถเพิ่มประสิทธิภาพการจำแนกข้อมูลได้

ปภากร อินทร์โท่โล่ (2557) การศึกษาการพยากรณ์การพบหรือไม่พบของเชื้อซาลโมเนลล่าเอนเทอริกา และการดื้อหรือไม่ดื้อยาต้านจุลชีพของเชื้อซาลโมเนลล่าเอนเทอริกา ข้อมูลจากโครงการเขียงสะอาดร้านค้าหมูตลาดสด จังหวัดน่าน ซึ่งแบ่งออกเป็น 2 ส่วน คือ ส่วนที่ 1 คือ ด้านการพบหรือไม่พบเชื้อซาลโมเนลล่าเอนเทอริกา สำหรับข้อมูล ระหว่างเดือนมกราคม - กันยายน พ.ศ.2554 ตัวแปรอิสระคือจำนวนแบคทีเรียทั้งหมด (TotalBac) หน่วยเป็น log cfu/cm^2 และตำแหน่งที่ตรวจหาเชื้อซาลโมเนลล่าเอนเทอริกา 6 ตำแหน่ง คือ เนื้อสุกร(Site1) เขียง (Site2) ตราชั่ง(Site3) มีด(Site4) ถาด(Site5) และมือผู้ขาย(Site6) และส่วนที่ 2 คือ การดื้อหรือไม่ดื้อยาต้านจุลชีพของเชื้อซาลโนเนลล่าเอนเทอริกา ระห่างเดือนมกราคม – มีนาคม พ.ศ.2554 ตัวแปรอิสระคือ ชนิดของยาจำนวน 7 ชนิด คือ Trimethoprim(D_1), Ampicillin(D_2), Streptomycin(D_3), Sulfonamides(D_4), Tetracycline(D_5), Trimethoprim + Sulfamethoxazole(D_6) และChloramphenicol(D_7) ผู้วิจัยได้รับการอนุเคราะห์ข้อมูลดังกล่าวจาก รศ.น.สพ.ดร.ธงชัย เฉลิมชัยกิจ และ รศ.สพ.ญ.ดร.รุ่งทิพย์ ชวนชื่น โดยการใช้วิธีการวิเคราะห์ถดถอยโลจิสติก ผลสรุปได้ดังนี้ ส่วนที่ 1 สมการถดถอยโลจิสติกด้านการพบ/ไม่พบเชื้อซาลโมเนลล่าเอนเทอริกา

log odds = - 2.9015 + 0.4099 ($Site_1$) + 0.1357 ($Site_2$) + 0.2913 ($Site_3$) – 0.2430 ($Site_4$) + 0.5227 ($Site_5$) + 0.6602 (TotalBac)

โดยที่ มือผู้ขาย (Site 1) เป็นตัวแปรอ้างอิง

ส่วนที่ 2 สมการถดถอยโลจิสติกด้านการดื้อหรือไม่ดื้อยาต้านจุลชีพของเชื้อซาลโมเนลล่าเอนเทอริกา

log odds = - 2.5648 + 2.8525(D_1) + 2.2771(D_2) + 4.6850(D_3) + 2.2771(D_4) + 3.8641 (D_5) + 1.0387(D_6)

โดยที่ Chloramphenicol (D_7) เป็นตัวแปรอ้างอิง

โดยด้านการพบหรือไม่พบเชื้อซาลโนเนลล่าเอนเทอริกา พบว่า อัตราการพยากรณ์ผิดพลาดของความน่าจะเป็น 30.81% เมื่อกำหนดจุดตัดเป็น 0.4 และการดื้อหรือไม่ดื้อยาต้านจุลชีพของเชื้อซาลโมเนลล่าเอนเทอริกา พบว่า อัตราการพยากรณ์ผิดพลาดของความน่าจะเป็น 26.53% เมื่อกำหนดจุดตัดเป็น 0.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชณัฐดาภรณ์ เย็นประเสริฐ (2557) งานวิจัยนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำการพยากรณ์สถานะการชำระหนี้ของลูกหนี้ โดยใช้เทคนิคการถดถอยโลจิสติก เทคนิคเนอีฟเบย์ และเทคนิคต้นไม้ตัดสินใจ ซึ่งมีการคัดเลือกคุณลักษณะของชุดข้อมูล ด้วยเทคนิคอัตราขยาย (Information Gain) ส่วนเกณฑ์เปรียบเทียบความแม่นยำในการพยากรณ์ใช้ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ ค่าความแตกต่างระหว่างค่าจริงและค่าที่ประมาณได้จากแบบจำลอง ข้อมูลที่ใช้ในการวิจัย คือ ข้อมูลประวัติสถานะการชำระหนี้ของลูกหนี้จากสถาบันการเงินแห่งหนึ่ง ซึ่งเป็นข้อมูลในช่วงปี พ.ศ. 2556 จำนวน 392 ข้อมูล ผลลัพธ์จากวิธีการพยากรณ์ความแม่นยำของสถานะการชำระหนี้ของลูกค้า พบว่า เทคนิคต้นไม้ตัดสินใจ มีความแม่นยำเท่ากับ 89.80% มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ 10.20% รากที่สองของค่าเฉลี่ยคลาดเคลื่อนกำลังสอง เท่ากับ 0.30 รองลงมา คือ เทคนิคการถดถอยโลจิสติก มีความแม่นยำ เท่ากับ 89.80% ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ เท่ากับ 10.20% รากที่สองของค่าเฉลี่ยคลาดเคลื่อนกำลังสอง เท่ากับ 0.31 และเทคนิคนาอีฟเบย์ มีความแม่นยำ เท่ากับ 88.78% ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ เท่ากับ 11.22% รากที่สองของค่าเฉลี่ยคลาดเคลื่อนกำลังสอง เท่ากับ 0.33 จากผลการทดลองกรณีศึกษา พบว่า เทคนิคต้นไม้ตัดสินใจ และเทคนิคการถดถอยโลจิสติก มีค่าความแม่นยำ มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ และรากที่สองของค่าเฉลี่ยคลาดเคลื่อนกำลังสอง ใกล้เคียงกัน ในขณะที่เทคนิคนาอีฟเบย์ มีค่าความแม่นยำต่ำ มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ และรากที่สองของค่าเฉลี่ยคลาดเคลื่อนกำลังสองสูงกว่า

สุรวัชร ศรีเปารยะ (2559) งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม การเป็นโรคไตเรื้อรังโรงพยาบาลอพอลโล ประเทศอินเดีย โดยเลือกใช้ 7 วิธี ได้แก่ วิธีความใกล้เคียงกันมากที่สุด วิธีต้นไม้ตัดสินใจ วิธีโครงข่ายประสาทเทียม วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีฐานกฎ วิธีการถดถอยลอจิสติก และวิธีนาอีฟ เบย์ โดยแบ่งข้อมูลอย่างสุ่มเป็น 2 ชุด คือ ชุดสร้างตัวแบบและชุดทดสอบตัวแบบ ในอัตราส่วน 70 และ 30 ตามลำดับ จากการเปรียบเทียบประสิทธิภาพโดยพิจารณาจากค่าความถูกต้องและค่าความคลาดเคลื่อนกำลังสองเฉลี่ย พบว่าวิธีการจำแนกกลุ่มที่มีประสิทธิภาพการจำแนกดีที่สุดสำหรับข้อมูลดังกล่าวนี้คือ วิธีต้นไม้ตัดสินใจ ซึ่งให้ค่าความถูกต้องคือ 100% และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยคือ 0.0059

Kritiyaporn Kunsook (2012) การวิจัยนี้แสดงการใช้ต้นไม้ตัดสินใจ กฎริปเปอร์ และการวิเคราะห์การถดอยลอจิสติกในการจำแนกประเภทของระบบปลูกพืชไร้ดิน 2 รูปแบบ คือปลูกในรูปแบบให้สารละลายธาตุอาหารในถาดปลูกระดับลึก และการปลูกแบบให้สารละลายธาตุอาหารไหลผ่านเป็นแผ่นบางตลอดเวลา ในการจำแนกประเภทของระบบปลูกพืชไร้ดิน นำตัวอย่างทดสอบมาจากพืชที่ปลูกในรูปแบบให้สารละลายธาตุอาหารในถาดปลูกระดับลึก และการปลูกแบบให้สารละลายธาตุอาหารไหลผ่านเป็นแผ่นบางตลอดเวลา ข้อมูลที่ใช้ในการทดสอบประกอบด้วย อายุการเก็บเกี่ยว ความสูง ความกว้าง อุณหภมิที่เหมาะสม คุณลักษณะความต้องการแสง และลักษณะสีของพืช สำหรับใช้การจำแนกประเภทของระบบปลูกพืชไร้ดิน ต้นไม้ตัดสินใจ กฎริปเปอร์ และการถดถอยลอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จิสติกถูกนำมาใช้เป็นแบบจำลองในการจำแนกความสัมพันธ์ของข้อมูล มาใช้ในการเรียนรู้ของข้อมูลพืช จากการทดลอง พบว่า ผลลัพธ์ของต้นไม้ตัดสินใจ สำหรับการจำแนกมีประสิทธิภาพใกล้เคียงกับกฎริปเปอร์และการถดถอยลอจิสติก ว่าร้อยละความถูกต้องเท่ากับ 97.73, 98.73 และ 98 ตามลำดับ

Pimpa Cheewaprakobkit (2013) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาอัตราความถูกต้องในการทำนายระหว่างเทคนิคการทำเหมืองข้อมูล 2 วิธี ได้แก่ ต้นไม้ตัดสินใจและโครงข่ายประสาทเทียมในการจำแนกกลุ่มผลสัมฤทธิ์ทางการเรียนของนักศึกษา และวิเคราะห์ปัจจัยที่มีผลต่อผลสัมฤทธิ์ทางการเรียน ในการศึกษาครั้งนี้ผู้วิจัยใช้เครื่องมือการทำเหมืองข้อมูล WEKA เพื่อวิเคราะห์คุณลักษณะสำหรับคาดการณ์ผลการเรียนในระดับปริญญาตรีของนักศึกษาระดับปริญญาตรีในโครงการระดับนานาชาติ ชุดข้อมูลประกอบด้วยนักศึกษา 1,600 คน โดยมีนักเรียน 22 คนที่ลงทะเบียนระหว่างปี 2544 ถึง 2554 ในมหาวิทยาลัยในประเทศไทย ผลการศึกษาพบว่าการจำแนกกลุ่มด้วยเทคนิคต้นไม้ตัดสินใจมีความถูกต้องเท่ากับ 85.188% ซึ่งสูงกว่าการจำแนกกลุ่มด้วยเทคนิคโครงข่ายประสาทเทียมเท่ากับ 1.313%

ซึ่งจากการศึกษางานวิจัยที่เกี่ยวข้องดังกล่าว จะเห็นได้ว่าจากการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มในแต่ละวิธีมีประสิทธิภาพที่แตกต่างกันไป เช่น เมื่อเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มระหว่างการวิเคราะห์จำแนกประเภทและการวิเคราะห์การถดถอยลอจิสติก พบว่าการวิเคราะห์จำแนกประเภทมีประสิทธิภาพสูงกว่าหรือเท่ากัน ส่วนการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มระหว่างการวิเคราะห์การถดถอยลอจิสติกและต้นไม้ตัดสินใจ พบว่า ต้นไม้ตัดสินใจมีประสิทธิภาพในการจำแนกกลุ่มสูงกว่า ทั้งนี้อาจจะขึ้นอยู่กับลักษณะของข้อมูลแต่ละประเภทว่าข้อมูลแต่ละลักษณะเหมาะสมกับวิธีการจำแนกกลุ่มวิธีใด ถึงจะสามารถจำแนกกลุ่มได้ถูกต้องหรือแม่นยำมากที่สุด ดังนั้นผู้วิจัยจึงนำความรู้จากการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องมาเพื่อเป็นแนวทางในการทำวิจัยในครั้งนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินการวิจัย

สำหรับการศึกษางานวิจัยในครั้งนี้ ผู้วิจัยได้ศึกษาการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ ซึ่งผู้วิจัยได้กำหนดวิธีการดำเนินการวิจัยไว้ดังนี้

3.1 การเก็บรวบรวมข้อมูล

3.2 เครื่องมือที่ใช้ในการวิจัย

3.3 ขั้นตอนในการดำเนินการวิจัย

3.4 วิธีการวิเคราะห์ข้อมูล

3.1 การเก็บรวบรวมข้อมูล

การศึกษางานวิจัยในครั้งนี้ ผู้วิจัยใช้ข้อมูลทุติยภูมิของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ที่มารับการรักษาในโรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ ตั้งแต่เดือนตุลาคม พ.ศ.2559 – กันยายน พ.ศ. 2560 เป็นระยะเวลา 12 เดือน จำนวนทั้งหมด 507 ราย เนื่องจากข้อมูลที่เก็บรวบไว้ มีข้อมูลบางตัวแปรไม่สมบูรณ์ คือ ข้อมูลมีค่าสูญหายหรือผิดพลาด ซึ่งอาจจะมาจากกระบวนการหรือขั้นตอนในการเก็บรวบรวมข้อมูล เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลหรืออาจจะมีสาเหตุมาจากตัวผู้ป่วยเอง ดังนั้นผู้วิจัยจึงนำเฉพาะข้อมูลที่มีความถูกต้องและครบถ้วนสมบูรณ์มากที่สุดมาใช้ในการวิเคราะห์ในงานวิจัยครั้งนี้ และตัดข้อมูลที่มีค่าสูญหายหรือผิดพลาดออกไป ซึ่งเหลือข้อมูลที่มีความถูกต้องและครบถ้วนสมบูรณ์มากที่สุดเพียง จำนวน 320 ราย ซึ่งประกอบด้วยผู้ป่วย 2 กลุ่ม คือ กลุ่มที่ไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 200 ราย และกลุ่มที่เกิดภาวะแทรกซ้อนทางไต จำนวน 120 ราย ซึ่งประกอบด้วยตัวแปรอิสระและตัวแปรตามดังนี้

3.1.1 ตัวแปรอิสระ (X_i) คือ ปัจจัยที่อาจจะส่งผลต่อการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จำนวน 23 ตัว ดังนี้

X_1 คือ เพศ (ชาย, หญิง)

X_2 คือ อายุ (ปี)

X_3 คือ น้ำหนัก (กิโลกรัม)

X_4 คือ ส่วนสูง (เซนติเมตร)

X_5 คือ ดัชนีมวลกาย (กิโลกรัมต่อเมตร^2)

X_6 คือ รอบเอว (นิ้ว)

X_7 คือ ความดันโลหิตซีสโตลิค (มิลลิเมตรปรอท)

X_8 คือ ความดันโลหิตไดแอสโตลิค (มิลลิเมตรปรอท)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X_9 คือ Creatinine (มิลลิกรัมต่อเดซิลิตร)
X_{10} คือ HbA1c (มิลลิกรัมต่อเปอร์เซ็นต์)
X_{11} คือ Triglyceride (มิลลิกรัมต่อเดซิลิตร)
X_{12} คือ LDL Cholesterol (มิลลิกรัมต่อเดซิลิตร)
X_{13} คือ HDL Cholesterol (มิลลิกรัมต่อเดซิลิตร)
X_{14} คือ Total Cholesterol (มิลลิกรัมต่อเดซิลิตร)
X_{15} คือ Urine protein (มิลลิกรัมต่อเดซิลิตร)
X_{16} คือ Uric Acid (มิลลิกรัมต่อเปอร์เซ็นต์)
X_{17} คือ Albuminuria (ไมโครกรัมต่อนาที)
X_{18} คือ FBS (มิลลิกรัมต่อเดซิลิตร)
X_{19} คือ BUN (มิลลิกรัมต่อเปอร์เซ็นต์)
X_{20} คือ ระยะเวลาที่ป่วยเป็นโรคเบาหวาน ชนิดที่ 2 (ปี)
X_{21} คือ การดื่มสุรา (ไม่ดื่ม, ดื่ม)
X_{22} คือ การสูบบุหรี่ (ไม่สูบ, สูบ)
X_{23} คือ ภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 (ไม่มี, มี)

3.1.2 ตัวแปรตาม (Y) คือ ภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 (เกิด, ไม่เกิด)

3.2 เครื่องมือที่ใช้ในการวิจัย

สำหรับการวิจัยในครั้งนี้ ผู้วิจัยได้ทำการวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติ SPSS (Statistical Package for the Social Science for Windows) เวอร์ชั่น 20 และโปรแกรม WEKA (Waikato Environment for Knowledge Analysis) เวอร์ชั่น 3.8.1

3.3 ขั้นตอนในการดำเนินการวิจัย

3.3.1 ศึกษาและทบทวนวรรณกรรมหรือผลงานวิจัยที่เกี่ยวข้อง (Review Literature) เพื่อเป็นแนวทางในการดำเนินงานวิจัย

3.3.2 ศึกษาขั้นตอนการวิเคราะห์ข้อมูล โดยใช้เทคนิคการพยากรณ์ 3 วิธี คือ การวิเคราะห์จำแนกประเภท (Discriminant Analysis) การวิเคราะห์ถดถอยลอจิสติก (Logistic Regression) และต้นไม้ตัดสินใจ (Decision Tree) และวิธีการเปรียบเทียบประสิทธิภาพของการจำแนกกลุ่ม

3.3.3 เก็บรวบรวมข้อมูล และจัดเตรียมข้อมูลเพื่อใช้ในการวิเคราะห์

3.3.4 นำข้อมูลมาทำการวิเคราะห์ด้วยการวิเคราะห์จำแนกประเภท และการวิเคราะห์ถดถอยลอจิสติก โดยใช้โปรแกรมสำเร็จรูปทางสถิติ SPSS และต้นไม้ตัดสินใจทำการวิเคราะห์ข้อมูลโดยใช้โปรแกรม WEKA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.5 นำผลที่ได้จากการวิเคราะห์ข้อมูลด้วยวิธีการจำแนกกลุ่มทั้ง 3 วิธี มาเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่ม โดยอาศัยค่าความถูกต้อง (Accuracy) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เพื่อหาวิธีการจำแนกกลุ่มที่มีประสิทธิภาพสูงที่สุด

3.3.6 สรุปผลงานวิจัย อภิปรายผล และข้อเสนอแนะ

3.4 วิธีการวิเคราะห์ข้อมูล

3.4.1 การวิเคราะห์จำแนกประเภท

3.4.1.1 ตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์จำแนกประเภท

3.4.1.1.1 ตรวจสอบตัวแปรอิสระทั้ง 23 ตัว ว่ามีการแจกแจงปรกติหลายตัวแปรหรือไม่

3.3.2.1.2 ตรวจสอบความแปรปรวนร่วมของตัวแปรอิสระของแต่ละกลุ่มว่าเท่ากันหรือไม่

3.4.1.2 สร้างสมการการวิเคราะห์จำแนกประเภท โดยมีฟังก์ชันการวิเคราะห์จำแนกประเภทดังนี้

$$\hat{Y}_i = b_1X_1 + b_2X_2 + ... + b_PX_P$$

โดยที่ $\hat{Y}_i$ คือ ค่าพยากรณ์

b_i คือ ค่าสัมประสิทธิ์ของการวิเคราะห์จำแนกประเภท

X_i คือ ตัวแปรอิสระ เมื่อ i = 1, 2,..., n

3.4.1.3 นำสมการการวิเคราะห์จำแนกประเภทไปจำแนกกลุ่มผู้ป่วยโรคเบาหวานชนิดที่ 2 รายใหม่ ว่าควรจะอยู่ในกลุ่มเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไต

3.4.2 การวิเคราะห์การถดถอยลอจิสติก

3.4.2.1 ตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์การถดถอยลอจิสติก

3.4.2.1.1 ตรวจสอบชนิดของข้อมูลของตัวแปรอิสระว่าเป็นข้อมูลที่มีค่าได้ 2 ค่า คือ 0 กับ 1 เท่านั้น หรือเป็นสเกลอันตรภาคหรือสเกลอัตราส่วนหรือไม่

3.4.2.1.2 ค่าคาดหวังของค่าความคาดเคลื่อนเป็นศูนย์ หรือ E(e) = 0

3.4.2.1.3 e_i และ e_j เป็นอิสระกัน ; i , j = 1, 2, 3,...,n (เมื่อ n คือ จำนวนข้อมูลทั้งหมด)

3.4.2.1.4 e_i และ X_i เป็นอิสระกัน ; i = 1, 2, 3,..., n (เมื่อ n คือ จำนวนข้อมูลทั้งหมด)

3.4.2.1.5 ตรวจสอบตัวแปรอิสระไม่มีความสัมพันธ์กันหรือไม่มีความสัมพันธ์เชิงเส้นแบบพหุ (Multicollinearity)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.2.2 สร้างสมการการวิเคราะห์การถดถอยลอจิสติก โดยมีฟังก์ชันการวิเคราะห์การถดถอยลอจิติกดังนี้

$$P(\text{เกิดเหตุการณ์}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + ... + \beta_p X_p)}}$$

$$P(\text{ไม่เกิดเหตุการณ์}) = 1 - P(\text{เกิดเหตุการณ์})$$

3.4.2.3 นำสมการการวิเคราะห์การถดถอยลอจิสติกไปจำแนกกลุ่มกลุ่มผู้ป่วยโรคเบาหวาน ชนิดที่ 2 รายใหม่ ว่าควรจะอยู่ในกลุ่มเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไต

3.4.3 ต้นไม้ตัดสินใจ

3.4.3.1 สร้างแผนภาพต้นไม้เพื่อประกอบการตัดสินใจสำหรับใช้ในการจำแนกประเภทข้อมูล โดยใช้ขั้นตอนวิธีชนิด C4.5 ซึ่งถูกพัฒนามาจากขั้นตอนวิธีชนิด ID3 ต่างจากขั้นตอนวิธีชนิด ID3 ใช้ได้เพียงข้อมูลไม่ต่อเนื่องเท่านั้น แต่ขั้นตอนวิธีชนิด C4.5 ใช้ได้ทั้งข้อมูลแบบต่อเนื่องและไม่ต่อเนื่อง

3.4.3.2 นำแผนภาพต้นไม้ตัดสินใจไปใช้เพื่อประกอบการตัดสินใจสำหรับจำแนกกลุ่มกลุ่มผู้ป่วยโรคเบาหวาน ชนิดที่ 2 รายใหม่ ว่าควรจะอยู่ในกลุ่มเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไต

3.4.4 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม

3.4.4.1 ค่าความถูกต้อง (Accuracy) ซึ่งสามารถคำนวณได้จากตารางเมทริกซ์ความสับสน (Confusion Matrix) ดังตารางที่ 3.1

ตารางที่ 3.1 เมทริกซ์ความสับสน

	ค่าที่ทำนาย	
ค่าที่แท้จริง	กลุ่ม 1	กลุ่ม 2
กลุ่ม 1	True Positive (TP)	False Positive (FP)
กลุ่ม 2	False Negative (FN)	True Negative (TN)

โดยที่ True Positive (TP) คือ จำนวนข้อมูลที่จำแนกถูกว่าเป็นกลุ่ม 1

True Negative (TN) คือ จำนวนข้อมูลที่จำแนกถูกว่าเป็นกลุ่ม 2

False Positive (FP) คือ จำนวนข้อมูลที่จำแนกผิดว่าเป็นกลุ่ม 2 ซึ่งค่าที่แท้จริงเป็นกลุ่ม 1

False Negative (FN) คือ จำนวนข้อมูลที่จำแนกผิดว่าเป็นกลุ่ม 1 ซึ่งค่าที่แท้จริงเป็นกลุ่ม 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าความถูกต้อง (Accuracy) คือ อัตราส่วนของข้อมูลที่พยากรณ์ถูกต้องและข้อมูลทั้งหมด สามารถคำนวณได้จากกสูตรดังนี้

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{(TP+FP+FN+TN)}} \times 100\%$$

3.4.4.2 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root mean Squared Error : RMSE) หากยิ่งมีค่าน้อยหมายถึงจะมีความแม่นยำมากยิ่งขึ้น

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i\right)^2}{n}}$$

โดยที่ y_i แทน ค่าจริง

$\hat{y}_i$ แทน ค่าพยากรณ์

n แทน ขนาดข้อมูลตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิเคราะห์ข้อมูล

ในบทนี้จะแสดงรายละเอียดของผลการวิเคราะห์ข้อมูล ซึ่งเป็นการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ จำนวน 320 ราย ซึ่งประกอบด้วยผู้ป่วย 2 กลุ่ม คือ กลุ่มผู้ป่วยที่ไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 200 และกลุ่มผู้ป่วยที่เกิดภาวะแทรกซ้อนทางไต จำนวน 120 ราย โดยใช้วิธีการวิเคราะห์จำแนกกลุ่ม 3 วิธี คือ การวิเคราะห์จำแนกประเภท การวิเคราะห์การถดถอยลอจิสติก และต้นไม้ตัดสินใจ เพื่อเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มทั้ง 3 วิธี และเพื่อสามารถนำวิธีการจำแนกกลุ่มที่มีประสิทธิภาพสูงที่สุดไปใช้ในการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ได้ โดยอาศัยค่าความถูกต้องและค่ารากที่สองของความเคลื่อนกำลังสองเฉลี่ย โดยพิจารณาจากตัวแปรอิสระทั้งหมด 23 ตัว ซึ่งมีรายละเอียดดังต่อไปนี้

4.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีการวิเคราะห์จำแนกประเภท

4.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก

4.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ

4.4 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม

4.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีการวิเคราะห์จำแนกประเภท (Discriminant Analysis)

4.1.1 การตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์จำแนกประเภท

4.1.1.1 การตรวจสอบว่าตัวแปรอิสระของแต่ละกลุ่มมีการแจกแจงปรกติหลายตัวแปรหรือไม่ ด้วยการตรวจสอบการแจกแจงปรกติตัวแปรเดียว โดยทำการแยกตรวจสอบตัวแปรอิสระทีละตัว โดยใช้ตัวสถิติทดสอบ Kolmogorov-Smirnov ปรากฏผลดังตารางที่ 4.1

สมมติฐานการทดสอบ

H_0 : ตัวแปรอิสระแต่ละตัวมีการแจกแจงปรกติ

H_1 : ตัวแปรอิสระแต่ละตัวไม่มีการแจกแจงปรกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ผลการตรวจสอบการแจกแจงปรกติของตัวแปรอิสระหลายตัวแปร โดยใช้ตัวสถิติทดสอบ Kolmogorov-Smirnov

ตัวแปรอิสระ(X_i)	ภาวะแทรกซ้อนทางไต (Y)	Kolmogorov-Smirnov[a]		
		Statistic	df	p-value
X_1	ไม่เกิด	.392	200	.000
	เกิด	.400	120	.000
X_2	ไม่เกิด	.047	200	**.200***
	เกิด	.073	120	**.168**
X_3	ไม่เกิด	.093	200	.000
	เกิด	.103	120	.003
X_4	ไม่เกิด	.071	200	.017
	เกิด	.098	120	.007
X_5	ไม่เกิด	.107	200	.000
	เกิด	.072	120	**.191**
X_6	ไม่เกิด	.155	200	.000
	เกิด	.106	120	.002
X_7	ไม่เกิด	.153	200	.000
	เกิด	.140	120	.000
X_8	ไม่เกิด	.090	200	.000
	เกิด	.105	120	.002
X_9	ไม่เกิด	.163	200	.000
	เกิด	.218	120	.000
X_{10}	ไม่เกิด	.140	200	.000
	เกิด	.161	120	.000
X_{11}	ไม่เกิด	.182	200	.000
	เกิด	.138	120	.000
X_{12}	ไม่เกิด	.055	200	**.200***
	เกิด	.077	120	**.074**
X_{13}	ไม่เกิด	.075	200	.008
	เกิด	.102	120	.004
X_{14}	ไม่เกิด	.050	200	**.200***
	เกิด	.088	120	.023
X_{15}	ไม่เกิด	.514	200	.000
	เกิด	.384	120	.000
X_{16}	ไม่เกิด	.088	200	.001
	เกิด	.063	120	**.200***
X_{17}	ไม่เกิด	.298	200	.000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวแปรอิสระ(X_i)	ภาวะแทรกซ้อนทางไต (Y)	Kolmogorov-Smirnov[a]		
		Statistic	df	p-value
X_{18}	เกิด	.271	120	.000
	ไม่เกิด	.120	200	.000
X_{19}	เกิด	.194	120	.000
	ไม่เกิด	.105	200	.000
X_{20}	เกิด	.118	120	.000
	ไม่เกิด	.177	200	.000
X_{21}	เกิด	.103	120	.003
	ไม่เกิด	.540	200	.000
X_{22}	เกิด	.535	120	.000
	ไม่เกิด	.541	200	.000
X_{23}	เกิด	.528	120	.000
	ไม่เกิด	.537	200	.000
	เกิด	.540	120	.000

จากตารางที่ 4.1 ผลการทดสอบการแจกแจงปรกติของตัวแปรอิสระแต่ละตัว โดยใช้ตัวสถิติทดสอบ Kolmogorov-Smirnov พบว่า p-value ของตัวแปรอิสระ จำนวน 18 ตัว ได้แก่ X_1, X_3, X_4, X_5(Y= ไม่เกิด), X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}(Y= เกิด), X_{15}, X_{16}(Y= ไม่เกิด), X_{17}, X_{18}, X_{19}, X_{20}, X_{21}, X_{22}, และ X_{23} มีค่าน้อยกว่าระดับนัยสำคัญ .05 จึงปฏิเสธ H_0 แสดงว่าตัวแปรอิสระแต่ละตัวไม่มีการแจกแจงปรกติ และมีตัวแปรอิสระเพียง 5 ตัว ได้แก่ X_2, X_5(Y= เกิด), X_{12}, X_{14}(Y= ไม่เกิด) และ X_{16}(Y= เกิด) ที่มีค่า p-value มากกว่าระดับนัยสำคัญ .05 หรือตัวแปรอิสระแต่ละตัวมีการแจกแจงปรกติ แสดงว่าตัวแปรอิสระทุกตัวไม่ได้มีการแจกแจงปรกติ ดังนั้นจึงสรุปได้ว่าตัวแปรอิสระไม่มีการแจกแจงปรกติหลายตัวแปร หรือไม่เป็นไปตามข้อสมมติเบื้องต้นดังกล่าว

Barbara and Linda (2001) กล่าวว่า เมื่อตัวแปรอิสระที่นำมาศึกษามีทั้งตัวแปรต่อเนื่องและตัวแปรเชิงกลุ่ม จึงเป็นเรื่องยากที่ข้อมูลจะมีการแจกแจงแบบปรกติหลายตัวแปร แต่ถ้าจำนวนตัวอย่างที่ใช้ในงานวิจัยมีขนาดใหญ่ และวัตถุประสงค์ของงานวิจัยคือ ต้องการสร้างสมการจำแนกประเภทที่มีประสิทธิภาพในการจำแนกกลุ่มข้อมูลให้ถูกต้องมากที่สุด ดังนั้นข้อสมมติดังกล่าวจึงไม่ใช่สิ่งที่จำเป็น

4.1.1.2 การตรวจสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม โดยใช้ตัวสถิติ Box's M ปรากฏผลดังตารางที่ 4.2

สมมติฐานการทดสอบ

$$H_0 : \Sigma_1 = \Sigma_2$$

$$H_1 : \Sigma_1 \neq \Sigma_2$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หรือ H_0 : เมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระทั้ง 2 กลุ่ม เท่ากัน
H_1 : เมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระทั้ง 2 กลุ่ม ไม่เท่ากัน

ตารางที่ 4.2 ผลการทดสอบด้วยวิธี Box's M

Box's M	F	df1	df2	p-value
926.439	3.082	276	196457.109	< .001

จากตารางที่ 4.2 ผลของการทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนร่วม โดยใช้ตัวสถิติทดสอบ Box's M พบว่า p-value มีค่า < .000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญที่ .05 จึงปฏิเสธ H_0 ดังนั้นจึงสรุปได้ว่าเมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระทั้ง 2 กลุ่ม ไม่เท่ากัน หรือไม่เป็นไปตามข้อสมมติเบื้องต้นดังกล่าว

เนื่องข้อมูลที่นำมาศึกษาในงานวิจัยครั้งนี้มีขนาดใหญ่ จึงทำให้ค่า p-value มีค่าน้อย ซึ่งจะทำให้ปฏิเสธ H_0 เสมอ จึงทำให้การทดสอบความเท่ากันของเมทริกซ์ความแปรปรวนไม่เป็นไปตามข้อสมมติเบื้องต้นของการวิเคราะห์จำแนกประเภทข้อดังกล่าว และ Barbara and Linda (2001) กล่าวว่า หากข้อสมมติเบื้องต้นเกี่ยวกับความเท่ากันของเมทริกซ์ความแปรปรวนร่วมของกลุ่มที่จะจำแนกนั้น และหากข้อมูลไม่สอดคล้องกับข้อสมมติเบื้องต้นยังสามารถใช้การปรับเกี่ยวกับเมทริกซ์ความแปรปรวนร่วมที่นำมาใช้โดยการเปลี่ยนมาพิจารณาเมทริกซ์ความแปรปรวนแบบแยก (Separate Variance – Covariance Matrices) แทนเมทริกซ์ความแปรปรวนร่วมแบบรวม (Pooled Variance – Covariance Matrices) ในการสร้างตัวแบบจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ได้

ซึ่งในงานวิจัยครั้งนี้ผู้วิจัยแก้ปัญญาโดยเลือกใช้วิธีการปรับค่าเมทริกซ์ความแปรปรวนร่วมโดยใช้เมทริกซ์ความแปรปรวนร่วมแบบแยก (Separate Variance – Covariance Matrices) แทนเมทริกซ์ความแปรปรวนร่วมแบบรวม (Pool Variance – Covariance Matrices) ในการสร้างตัวแบบพยากรณ์การเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2

4.1.2 ผลการวิเคราะห์จำแนกประเภท

จากข้อมูลที่ใช้ในการวิเคราะห์การเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ จำนวน 320 ราย โดยอาศัยตัวแปรอิสระที่คาดว่าจะส่งผลต่อการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จำนวน 23 ตัว พบว่า ปรากฏผลดังตารางที่ 4.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 ค่าสัมประสิทธิ์การวิเคราะห์จำแจกประเภทของตัวแปรอิสระโดยใช้วิธีของ Fisher's (Fisher's Linear Discrimination Function)

ตัวแปรอิสระ (X_i)	b_j	
	ไม่เกิดภาวะแทรกซ้อนไต	เกิดภาวะแทรกซ้อนทางไต
Constant	-7741.442	-7739.378
X_1	56.364	58.916
X_2	2.067	2.261
X_3	-112.086	-111.956
X_4	96.141	96.011
X_5	284.006	283.508
X_6	-1.281	-1.291
X_7	1.557	1.561
X_8	0.001	-0.008
X_9	1.391	5.424
X_{10}	-2.544	-2.806
X_{11}	-0.086	-0.091
X_{12}	0.248	0.257
X_{13}	-0.297	-0.320
X_{14}	-0.129	-0.139
X_{15}	11.142	11.593
X_{16}	-3.118	-1.918
X_{17}	-0.060	-0.053
X_{18}	0.086	0.090
X_{19}	2.161	2.211
X_{20}	-1.202	-1.180
X_{21}	-8.561	-7.040
X_{22}	21.289	20.072
X_{23}	26.555	28.337

จากกตารางที่ 4.3 พบว่า ค่าสัมประสิทธิ์ของการวิเคราะห์จำแนกประเภท โดยใช้วิธี Fisher's จากการวิเคราะห์ข้อมูลจะได้สมการดังนี้

$$\hat{Y}_1 = -7741.442 + 56.364(X_1) + 2.067(X_2) - 112.086(X_3) + 96.141(X_4) + 284.006(X_5) - 1.281(X_6) + 1.557(X_7) + 0.001(X_8) + 1.391(X_9) - 2.544(X_{10}) - 0.086(X_{11}) + 0.248(X_{12}) - 0.297(X_{13}) - 0.129(X_{14}) + 11.142(X_{15}) - 3.118(X_{16}) - 0.060(X_{17}) + 0.086(X_{18}) + 2.161(X_{19}) - 1.202(X_{20}) - 8.561(X_{21}) + 21.289(X_{22}) + 26.555(X_{23})$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้ง

$$\hat{Y}_2 = -7739.378 + 58.916(X_1) + 2.261(X_2) - 111.956(X_3) + 96.011(X_4) + 283.508(X_5) - 1.291(X_6) + 1.561(X_7) - 0.008(X_8) + 5.424(X_9) - 2.806(X_{10}) - 0.091(X_{11}) + 0.257(X_{12}) - 0.320(X_{13}) - 0.139(X_{14}) + 11.593(X_{15}) - 1.918(X_{16}) - 0.053(X_{17}) + 0.090(X_{18}) + 2.211(X_{19}) - 1.180(X_{20}) - 7.040(X_{21}) + 20.072(X_{22}) + 28.337(X_{23})$$

เมื่อ $\hat{Y}_1$ แทน สมการจำแนกกลุ่มไม่เกิดภาวะแทรกซ้อนทางไต

$\hat{Y}_2$ แทน สมการจำแนกกลุ่มเกิดภาวะแทรกซ้อนทางไต

สำหรับการจำแนกกลุ่มโดยใช้สมการจำแนกประเภทของหน่วยตัวอย่างใหม่นั้นทำได้โดยแทนค่าตัวแปรลงไปในสมการทั้ง 2 สมการ โดยมีเกณฑ์พิจารณาเพื่อจัดกลุ่มผู้ป่วยดังนี้

ถ้า $\hat{Y}_1 > \hat{Y}_2$ จะจำแนกกลุ่มให้ผู้ป่วยโรคเบาหวาน ชนิดที่ 2 อยู่ในกลุ่มไม่เกิดภาวะแทรกซ้อนทางไต

ถ้า $\hat{Y}_1 < \hat{Y}_2$ จะจำแนกกลุ่มให้ผู้ป่วยโรคเบาหวาน ชนิดที่ 2 อยู่ในกลุ่มเกิดภาวะแทรกซ้อนทางไต

4.1.3 ผลการทดสอบสมมติฐานของการวิเคราะห์จำแนกประเภท

ในการวิเคราะห์จำแนกประเภท เบื้องต้นผู้วิจัยต้องทราบจำนวนกลุ่มของข้อมูลก่อนว่าข้อมูลแบ่งออกเป็นกี่กลุ่ม ซึ่งในงานวิจัยนี้ข้อมูลการเกิดภาวะแทรกซ้อนทางในผู้ป่วยโรคเบาหวานชนิดที่ 2 ได้แบ่งกลุ่มออกเป็น 2 กลุ่ม คือ กลุ่มไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 200 รายและกลุ่มเกิดภาวะแทรกซ้อนทางไต จำนวน 120 ราย โดยพิจารณาตัวแปรอิสระทั้ง 23 ตัว ซึ่งการแบ่งกลุ่มนี้ควรทำการตรวจสอบก่อนว่าทั้ง 2 กลุ่ม แตกต่างกันหรือไม่ ก่อนที่จะทำการวิเคราะห์ในขั้นตอนต่อไป ซึ่งในงานวิจัยนี้ใช้วิธี Hotelling's T^2 ในการทดสอบความแตกต่างของกลุ่มข้อมูลปรากฏผลดังตารางที่ 4.4

สมมติฐานการทดสอบ

$H_0: \underline{\mu}_1 = \underline{\mu}_2$

$H_1: \underline{\mu}_1 \neq \underline{\mu}_2$

หรือ H_0 : ค่าเฉลี่ยของตัวแปรอิสระทุกตัวทั้ง 2 กลุ่ม ไม่แตกต่างกัน

H_1 : ค่าเฉลี่ยของตัวแปรอิสระทุกตัวทั้ง 2 กลุ่ม แตกต่างกัน

ตารางที่ 4.4 ผลการทดสอบสมการจำแนกกลุ่มข้อมูล โดยใช้วิธี Hotelling's T^2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Hotelling's T^2	F	p-value
1.803	23.203	< .001

จากตารางที่ 4.4 ผลการทดสอบสมการจำแนกกลุ่มข้อมูลการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยใช้วิธี Hotelling's T^2 พบว่า p-value มีค่า < .001 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญทางสถิติ .05 จึงปฏิเสธ H_0 ดังนั้นจึงสรุปได้ว่า ค่าเฉลี่ยของตัวแปรอิสระทุกตัวของข้อมูลทั้ง 2 กลุ่มแตกต่างกันอย่างมีนัยสำคัญทางสถิติ หรือสมการการจำแนกกลุ่มที่วิเคราะห์ได้สามารถจำแนกข้อมูลการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ออกเป็น 2 กลุ่มได้

4.1.4 ผลการประเมินความเหมาะสมของสมการการวิเคราะห์จำแนกประเภท

ซึ่งผู้วิจัยได้ประเมินความเหมาะสมของสมการการวิเคราะห์จำแนกประเภทโดยใช้ค่าความถูกต้องและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏผลดังตารางที่ 4.5

ตารางที่ 4.5 ผลการประเมินความเหมาะสมของสมการการวิเคราะห์จำแนกประเภท

กลุ่ม	ขนาดตัวอย่าง (n_i)	ผลการจำแนกกลุ่ม		ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย
		ไม่เกิดภาวะแทรกซ้อนทางไต	เกิดภาวะแทรกซ้อนทางไต	
ไม่เกิดภาวะแทรกซ้อนทางไต	200	195(97.50%)	5(2.50%)	0.1768
เกิดภาวะแทรกซ้อนทางไต	120	5(4.17%)	115(95.83%)	

$$\text{ค่าความถูกต้อง} = \frac{(195 + 115) \times 100}{320} = 96.875\%$$

$$\text{หรือ} \quad \text{ค่าความผิดพลาด} = \frac{(5 + 5) \times 100}{320} = 3.125\%$$

จากตารางที่ 4.5 พบว่า จากข้อมูลจริงของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จำนวน 320 ราย ซึ่งประกอบด้วยผู้ป่วย 2 กลุ่ม คือ กลุ่มที่ไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 200 ราย เมื่อทำการจำแนกกลุ่มโดยใช้วิธีการวิเคราะห์จำแนกประเภท พบว่า จำแนกกลุ่มถูกต้องว่าไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 195 ราย และจำแนกกลุ่มผิดพลาดว่าเกิดภาวะแทรกซ้อนทางไต จำนวน 5 ราย นั่นคือ สมการการวิเคราะห์จำแนกประเภทสามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อยละ 97.5 และกลุ่มที่เกิดภาวะแทรกซ้อนทางไต จำนวน 120 ราย พบว่า จำแนกกลุ่มถูกต้องว่าเกิดภาวะแทรกซ้อนทางไต จำนวน 115 ราย และจำแนกกลุ่มผิดพลาดว่าไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 5 ราย นั่นคือ สมการการวิเคราะห์การจำแนกประเภทสามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อยละ 95.83 ดังนั้นจึงสรุปได้ว่า สมการการวิเคราะห์จำแนกประเภทสามารถจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยเบาหวาน ชนิดที่ 2 ได้ถูกต้องร้อยละ 96.875 และมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 0.1768

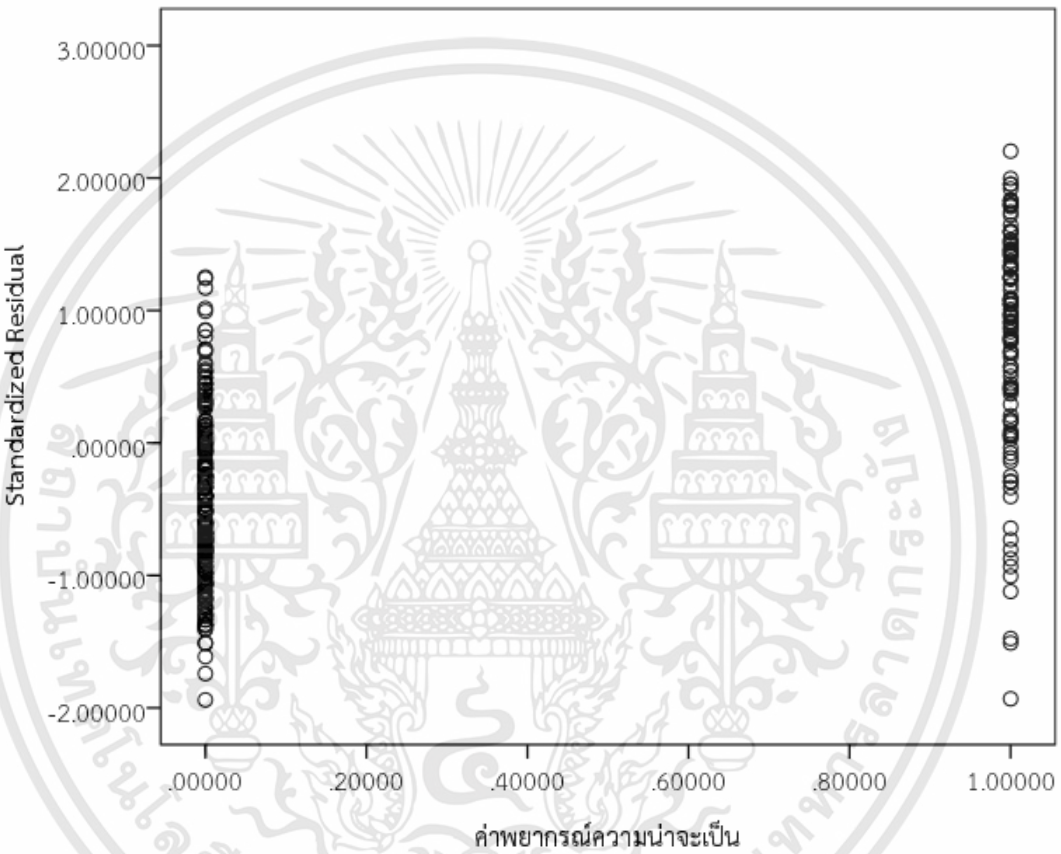
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา

4.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก

4.2.1 ตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์การถดถอยลอจิสติก

4.2.1.1 ตรวจสอบว่าตัวแปรอิสระอาจจะเป็นข้อมูลที่มีค่าได้ 2 ค่า (Dichotomous Variable) คือ 0 กับ 1 เท่านั้น หรือเป็นสเกลอันตรภาค (Interval Scale) หรือสเกลอัตราส่วน (Ratio Scale) ซึ่งในงานวิจัยนี้ข้อมูลที่ใช้ในการวิเคราะห์เป็นไปตามข้อสมมติเบื้องต้นดังกล่าว

4.2.1.2 ตรวจสอบค่าคาดหวังของค่าความคาดเคลื่อนเป็นศูนย์ หรือ E(e) = 0



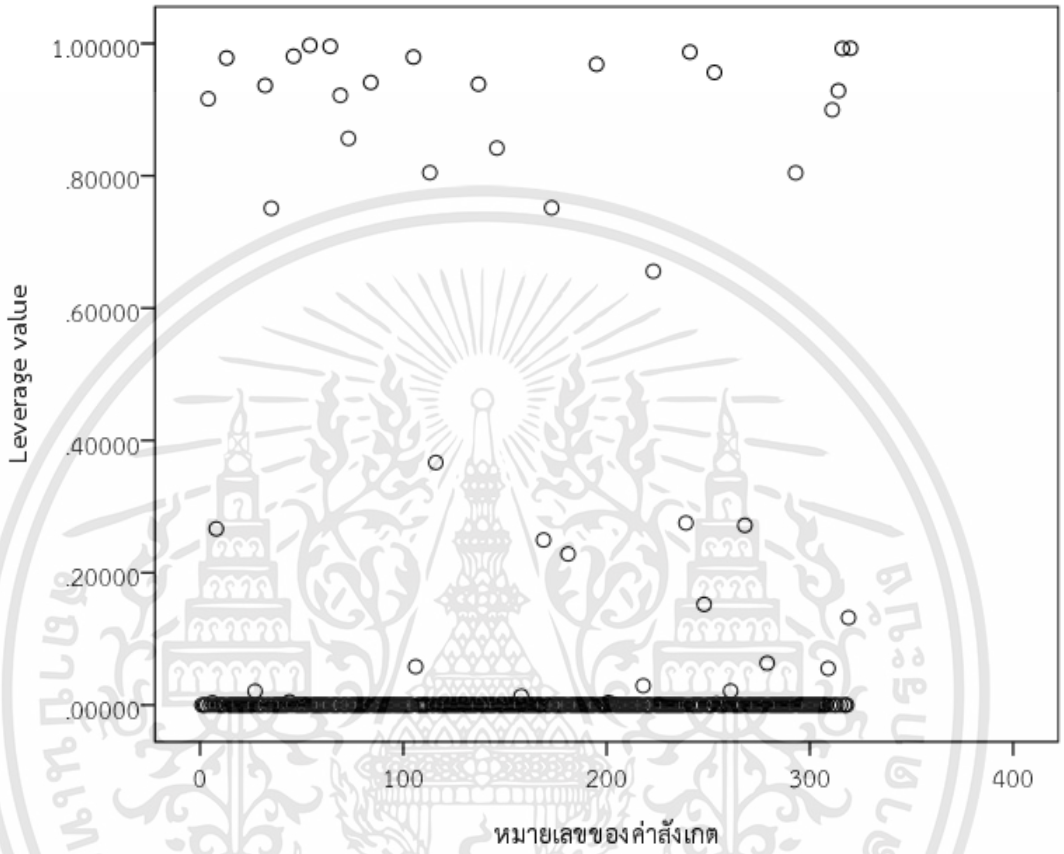
รูปที่ 4.1 การกระจายของค่าส่วนเหลือมาตรฐาน (Standardized Residual)

จากรูปที่ 4.1 จากการพล็อตกระจายของค่าส่วนเหลือมาตรฐานกับค่าพยากรณ์ความน่าจะเป็น เพื่อตรวจสอบว่าสมการที่นำมาใช้กับข้อมูลสอดคล้องกับข้อสมมติเบื้องต้นของการวิเคราะห์การถดถอยลอจิสติกที่ว่าค่าคาดหวังของค่าความคลาดเคลื่อนเป็นศูนย์ หรือ E(e) = 0 หรือไม่ ซึ่งจากการพิจารณารูปแบบการกระจายของค่าส่วนเหลือมาตรฐาน พบว่า ส่วนใหญ่มีค่าอยู่รอบ ๆ ศูนย์ ดังนั้นจึงสรุปได้ว่า ข้อสมมติเบื้องต้นดังกล่าวเป็นจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

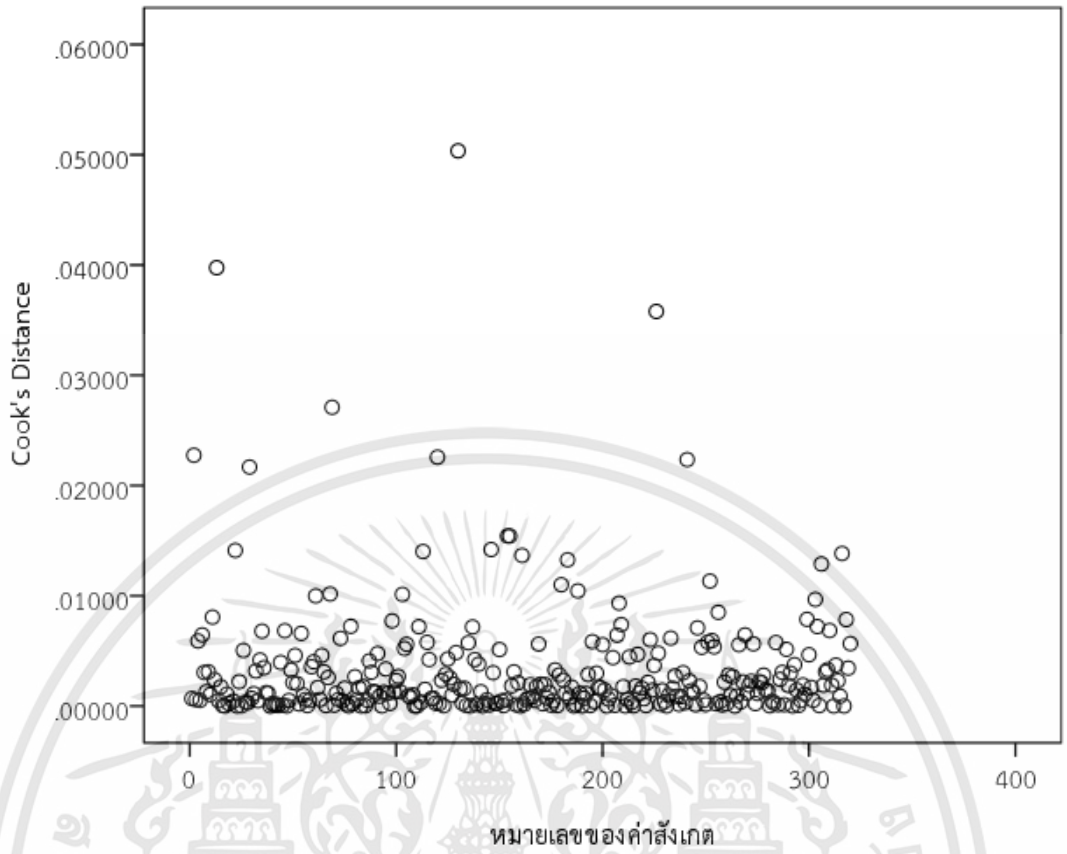
4.2.1.3 ตรวจสอบ e_i และ e_j เป็นอิสระกัน ; i , j = 1, 2, 3,..., n (เมื่อ n คือ จำนวนข้อมูลทั้งหมด) และ

4.2.1.4 ตรวจสอบ e_i และ X_i เป็นอิสระกัน ; i = 1, 2, 3,..., n (เมื่อ n คือ จำนวนข้อมูลทั้งหมด) ปรากฏผลดังรูปที่ 4.2 และ 4.3



รูปที่ 4.2 การกระจายของค่า Leverage

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 การกระจายของค่า Cook’s Distance

จากรูปที่ 4.2 – 4.3 จากการพล็อตกระจายของค่า Leverage และ Cook’s Distance กับหมายเลขของค่าสังเกต เพื่อตรวจสอบว่าสมการที่นำมาใช้กับข้อมูลตัวอย่างสอดคล้องกับข้อสมมติเบื้องต้นของการวิเคราะห์การถดถอยลอจิสติกที่ว่า e_i และ e_j เป็นอิสระกันหรือไม่ และข้อสมมติเบื้องต้นที่ว่า e_i และ X_i เป็นอิสระกันหรือไม่ เมื่อพิจารณารูปแบบการกระจายของ Leverage และ Cook’s Distance พบว่า ส่วนใหญ่มีค่าอยู่รอบ ๆ ศูนย์ ดังนั้นจึงสรุปได้ว่า ข้อสมมติเบื้องต้นดังกล่าวเป็นจริง

4.1.1.3 ตรวจสอบความสัมพันธ์ของตัวแปรอิสระ

การตรวจสอบว่าตัวแปรอิสสระแต่ละตัวมีความสัมพันธ์กันหรือไม่ โดยพิจารณาจากค่าความคลาดเคลื่อนยินยอม (Tolerance) และ VIF โดยมีเกณฑ์ในการพิจารณา คือ ถ้า Tolerance มีค่าเข้าใกล้ 0 หรือ VIF มีค่ามากกว่า 10 แสดงว่าตัวแปรอิสระมีความสัมพันธ์กัน ปรากฏผลดังตารางที่ 4.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 ค่า Tolerance และ VIF ของตัวแปรอิสระทั้ง 23 ตัว

ตัวแปรอิสระ (X_i)	Tolerance	VIF
X_1	.720	2.070
X_2	.007	1.389
X_3	.027	**153.285**
X_4	.008	**36.856**
X_5	.400	**121.109**
X_6	.352	2.502
X_7	.354	2.841
X_8	.471	2.829
X_9	.564	2.122
X_{10}	.355	1.774
X_{11}	.075	2.818
X_{12}	.432	**13.263**
X_{13}	.061	2.317
X_{14}	.699	**16.488**
X_{15}	.626	1.430
X_{16}	.894	1.596
X_{17}	.537	1.119
X_{18}	.487	1.862
X_{19}	.761	2.052
X_{20}	.483	1.314
X_{21}	.699	1.430
X_{22}	.696	1.437
X_{23}	.875	1.143

จากตารางที่ 4.6 เมื่อพิจารณาจากค่า Tolerance และ VIF ของตัวแปรอิสระทั้ง 23 ตัว พบว่า มีค่า Tolerance เข้าใกล้ 0 และ VIF มีค่ามากกว่า 10 จำนวน 5 ตัว คือ X_3, X_4, X_5 ,X_{12} และ X_{14} จึงสรุปได้ว่าตัวแปรอิสระมีความสัมพันธ์กัน ซึ่งในงานวิจัยนี้แก้ปัญหาโดยการใช้วิธีการตัดตัวแปรตัวใดตัวหนึ่งที่มีความสัมพันธ์กันมากออกไป ปรากฏผลดังตารางที่ 4.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ค่าความสัมพันธ์ของตัวแปรอิสระทั้ง 5 ตัว โดยใช้วิธีของเพียร์สัน (Pearson Correlation)

ตัวแปรอิสระ	X_3	X_4	X_5	X_{12}	X_{14}
X_3	1.000	**.473****	**.874****	.089	.057
X_4	**.473****	1.000	-.006	-.006	.008
X_5	**.874****	-.006.	1.000	105	.065
X_{12}	.089	-.006	.105	1.000	**.929****
X_{14}	.057	.008	.065	**.929****	1.000

จากตารางที่ 4.7 เมื่อพิจารณาค่าความสัมพันธ์ของตัวแปรอิสระทั้ง 5 ตัว คือ X_3, X_4 ,X_5 ,X_{12} และ X_{14} พบว่า มีตัวแปรอิสระที่มีความสัมพันธ์กันมากอยู่จำนวน 3 คู่ คือ คู่ที่ 1 ระหว่าง ตัวแปรอิสระ X_3 กับ X_4 คู่ที่ 2 คือ ระหว่างตัวแปรอิสระ X_3 กับ X_5 และคู่ที่ 3 ระหว่างตัวแปรอิสระ X_{12} กับ X_{14} ที่ระดับนัยสำคัญ .01 แล้วนำตัวแปรอิสระทั้ง 5 ตัว มาตรวจสอบความสัมพันธ์และหาค่าความสัมพันธ์กับตัวแปรตามอีกครั้งหนึ่ง เพื่อพิจารณาว่าจะตัดตัวแปรอิสระตัวใดตัวหนึ่งออกไป ปรากฏผลดังตารางที่ 4.8

ตารางที่ 4.8 ค่าความสัมพันธ์ของตัวแปรอิสระที่มีความสัมพันธ์กันมากทั้ง 5 ตัว กับตัวแปรตามโดยการใช้การวิเคราะห์ความแปรปรวนแบบทางเดียว (One Way ANOVA)

ตัวแปรอิสระ (X)	ค่าความสัมพันธ์กับตัวแปรตาม (Y)	p-value
X_3	.460	**.000***
X_4	.340	.133
X_5	.872	**.000***
X_{12}	.590	.443
X_{14}	.662	.202

จากตารางที่ 4.8 พบว่า เมื่อพิจารณาค่าความสัมพันธ์ของตัวแปรอิสระทั้ง 5 ตัว พบว่า มีตัวแปรอิสระเพียง 2 ตัว ที่มีความสัมพันธ์กับตัวแปรตาม คือ X_3 และ X_5 ที่ระดับนัยสำคัญ .05 และตัวแปรอิสระที่เหลืออีก 3 ตัว ไม่มีความสัมพันธ์กับตัวแปรตาม คือ X_4, X_{12} และ X_{14} เมื่อนำตัวแปรอิสระทั้ง 5 ตัว มาเปรียบเทียบความสัมพันธ์กับตัวแปรตามคู่ที่ 1 ระหว่างตัวแปรอิสระ X_3 กับ X_4 พบว่า X_3 มีความสัมพันธ์กับตัวแปรตาม แต่ X_4 ไม่มีความสัมพันธ์กับตัวแปรตาม จึงตัดตัวแปรอิสระ X_4 ออก คู่ที่ 2 ระหว่างตัวแปรอิสระ X_3 กับ X_5 พบว่า ตัวแปรทั้ง 2 มีความสัมพันธ์กับตัวแปรตาม แต่ X_5 มีความสัมพันธ์กับตัวแปรตามมากกว่าตัวแปรอิสระ X_3 จึงตัดตัวแปรอิสระ X_3 ออก คู่ที่ 3 ระหว่างตัวแปรอิสระ X_{12} กับ X_{14} พบว่า ตัวแปรทั้ง 2 ไม่มีความสัมพันธ์กับตัวแปรตาม แต่ X_{12} ไม่มีความสัมพันธ์กับตัวแปรตามมากกว่าตัวแปรอิสระ X_{14} จึงตัดตัวแปรอิสระ X_{14} ออก สรุปได้ว่าจะนำตัวแปรอิสระเพียง 2 ตัว คือ X_5 และ X_{12} และทำการตัดตัวแปรอิสระ X_3, X_4 และ X_{14} ออก แล้วนำตัวแปรอิสระที่เหลือ 20 ตัว มาตรวจสอบความสัมพันธ์อีกครั้งหนึ่ง ปรากฏผลดังตารางที่ 4.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.9 ค่า Tolerance และ VIF ของตัวแปรอิสระทั้ง 20 ตัว

ตัวแปรอิสระ (X_i)	Tolerance	VIF
X_1	.756	1.323
X_2	.784	1.275
X_5	.417	2.396
X_6	.453	2.209
X_7	.355	2.819
X_8	.364	2.747
X_9	.475	2.107
X_{10}	.575	1.740
X_{11}	.665	1.503
X_{12}	.728	1.373
X_{14}	.746	1.341
X_{15}	.629	1.591
X_{16}	.901	1.110
X_{17}	.538	1.860
X_{18}	.501	1.994
X_{19}	.766	1.305
X_{20}	.711	1.406
X_{21}	.699	1.430
X_{22}	.887	1.128
X_{23}	.865	1.156

จากตารางที่ 4.7 เมื่อพิจารณาค่า Tolerance และ VIF ของตัวแปรอิสระทั้ง 20 ตัว พบว่า Tolerance มีค่าไม่เข้าใกล้ 0 และ VIF มีค่าต่ำกว่า 10 ทุกตัว จึงสรุปได้ว่าตัวแปรอิสระไม่มีความสัมพันธ์กัน และสามารถนำตัวแปรอิสระทั้งหมดมาทำการวิเคราะห์การถดถอยลอจิสติก เพื่อสร้างสมการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ต่อไปได้

4.2.2 ผลการวิเคราะห์การถดถอยลอจิสติก

จากการวิเคราะห์ข้อมูลการเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไตของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ จำนวน 320 ราย ซึ่งประกอบด้วยตัวแปรอิสระ 20 ตัว โดยใช้การวิเคราะห์การถดถอยลอจิสติก และทำการประมาณค่าพารามิเตอร์โดยใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) ปรากฏผลดังตารางที่ 4.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 ค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระทั้ง 20 ตัว

ตัวแปรอิสระ (X_i)	b_j	SE(b_i)	Exp(b_i)
Constant	-338.578	32438.119	0.000
X_1	-66.549	3354.111	0.000
X_2	3.126	198.709	22.778
X_5	-0.899	394.049	0.407
X_6	0.255	67.663	1.291
X_7	-0.017	179.110	0.983
X_8	-0.337	117.337	0.714
X_9	230.767	10399.597	1.663E+100
X_{10}	-3.569	1143.595	0.028
X_{11}	-0.083	17.495	0.920
X_{12}	-0.148	73.000	0.862
X_{14}	-0.290	110.766	0.748
X_{15}	5.987	1182.654	398.368
X_{16}	10.102	1321.177	24393.158
X_{17}	0.040	15.412	1.040
X_{18}	0.075	30.515	1.078
X_{19}	0.198	213.915	1.218
X_{20}	1.120	212.466	3.066
X_{21}	-42.014	25620.633	0.000
X_{22}	0.526	16047.825	1.692
X_{23}	-11.282	15613.997	0.000

จะได้สมการตอบสนองลอจิตเป็นดังนี้

$$\text{Log odds} = -338.578 - 66.549(X_1) + 3.126(X_2) - 0.899(X_5) + 0.255(X_6) - 0.017(X_7) - 0.337(X_8) + 230.767(X_9) - 3.569(X_{10}) - 0.083(X_{11}) - 0.148(X_{12}) - 0.290(X_{13}) + 5.987(X_{15}) + 10.102(X_{16}) + 0.040(X_{17}) + 0.075(X_{18}) + 0.198(X_{19}) + 1.120(X_{20}) - 42.014(X_{21}) + 0.526(X_{22}) - 11.282(X_{23})$$

โดยมีตัวแปรอิสระที่คาดว่าจะส่งผลต่อการเป็นภาวะแทรกซ้อนทางไตของผู้ป่วยโรคเบาหวานชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ทั้ง 20 ตัว ดังต่อไปนี้

X_1 คือ เพศ (ชาย, หญิง)

X_2 คือ อายุ (ปี)

X_5 คือ ดัชนีมวลกาย (กิโลกรัมต่อเมตร^2)

X_6 คือ รอบเอว (นิ้ว)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X_7 คือ ความดันโลหิตซีสโตลิค (มิลลิเมตรปรอท)
X_8 คือ ความดันโลหิตไดแอสโตลิค (มิลลิเมตรปรอท)
X_9 คือ Creatinine (มิลลิกรัมต่อเดซิลิตร)
X_{10} คือ HbA1c (มิลลิกรัมต่อเปอร์เซ็นต์)
X_{11} คือ Triglyceride (มิลลิกรัมต่อเดซิลิตร)
X_{12} คือ LDL Cholesterol (มิลลิกรัมต่อเดซิลิตร)
X_{13} คือ HDL Cholesterol (มิลลิกรัมต่อเดซิลิตร)
X_{15} คือ Urine protein (มิลลิกรัมต่อเดซิลิตร)
X_{16} คือ Uric Acid (มิลลิกรัมต่อเปอร์เซ็นต์)
X_{17} คือ Albuminuria (ไมโครกรัมนาที)
X_{18} คือ FBS (มิลลิกรัมต่อเดซิลิตร)
X_{19} คือ BUN (มิลลิกรัมต่อเปอร์เซ็นต์)
X_{20} คือ ระยะเวลาที่ป่วยเป็นโรคเบาหวาน ชนิดที่ 2 (ปี)
X_{21} คือ การดื่มสุรา (ไม่ดื่ม, ดื่ม)
X_{22} คือ การสูบบุหรี่ (ไม่สูบ, สูบ)
X_{23} คือ ภาวะแทรกซ้อนทางตาในผู้ป่วยโรคเบาหวาน (ไม่มี, มี)

สำหรับการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยใช้สมการการวิเคราะห์การถดถอยลอจิสติก สามารถทำได้โดยแทนค่าของตัวแปรอิสระทั้ง 20 ตัว ลงในสมการได้ดังนี้

$$P(\text{เกิดภาวะแทรกซ้อนทางไต}) = \frac{e^{Z}}{1+e^{Z}}$$

โดยที่ $Z = -338.578 - 66.549(X_1) + 3.126(X_2) - 0.899(X_5) + 0.255(X_6) - 0.017(X_7) - 0.337(X_8) + 230.767(X_9) - 3.569(X_{10}) - 0.083(X_{11}) - 0.148(X_{12}) - 0.290(X_{13}) + 5.987(X_{15}) + 10.102(X_{16}) + 0.040(X_{17}) + 0.075(X_{18}) + 0.198(X_{19}) + 1.120(X_{20}) - 42.014(X_{21}) + 0.526(X_{22}) - 11.282(X_{23})$

ค่าที่คำนวณได้จะเป็นค่าจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ว่าจะเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไต

จากค่า $Exp(b_j)$ ของตัวแปรอิสระแต่ละตัวสามารถอธิบายได้ดังนี้ เช่น X_1 มีค่าเท่ากับ .000 และมีค่า b_j เท่ากับ - 66.549 หมายความว่า ผู้ป่วยโรคเบาหวาน ชนิดที่ 2 มีโอกาสเกิดภาวะแทรกซ้อนทางไตลดลงเป็น .000 เท่าของการไม่เกิดภาวะแทรกซ้อนทางไต หรือ X_2 มีค่า $Exp(b_j)$ เท่ากับ 22.778 และมีค่า b_j เท่ากับ 3.126 หมายความว่า ผู้ป่วยโรคเบาหวาน ชนิดที่ 2 มีโอกาสเกิดภาวะแทรกซ้อนทางไตเพิ่มขึ้นเป็น 22.778 เท่าของการไม่เกิดภาวะแทรกซ้อนทางไต และค่าของตัวอิสระตัวอื่น ๆ ก็สามารถอธิบายได้เช่นเดียวกัน

เอกสารนี้เป็น ไม่ว่ากรณีใดๆ

4.2.2.1 การทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระทีละตัวโดยใช้การทดสอบวอดล์ (Wald) ปรากฏผลดังตารางที่ 4.11

สมมติฐานการทดสอบ

$H_0 : \beta_j = 0$
$H_1 : \beta_j \neq 0$

โดยที่ j เท่ากับ 1, 2, 3, ..., P

หรือ H_0 : ตัวแปรอิสระแต่ละตัวในสมการไม่มีความสัมพันธ์กับตัวแปรตาม
H_1 : ตัวแปรอิสระแต่ละตัวในสมการมีความสัมพันธ์กับตัวแปรตาม

ตารางที่ 4.11 ผลทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระโดยใช้การทดสอบวอดล์

ตัวแปรอิสระ (x_i)	b_j	Wald	p-value
Constant	-338.578	.000	.984
X_1	-66.549	.000	.987
X_2	3.126	.000	.998
X_5	-.899	.000	.997
X_6	.255	.000	1.000
X_7	-.017	.000	.998
X_8	-.337	.000	.982
X_9	230.767	.000	.998
X_{10}	-3.569	.000	.996
X_{11}	-.083	.000	.998
X_{12}	-.148	.000	.998
X_{14}	-.290	.000	.996
X_{15}	5.987	.000	.994
X_{16}	10.102	.000	.998
X_{17}	.040	.000	.998
X_{18}	.075	.000	.999
X_{19}	.198	.000	.996
X_{20}	1.120	.000	.999
X_{21}	-42.014	.000	1.000
X_{22}	.526	.000	.999
X_{23}	-11.282	.000	.992

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.11 ผลการทดสอบโดยใช้ตัวสถิติทดสอบของวอดล์ พบว่า มีค่า p–value มากกว่า .05 จึงสรุปได้ว่า ตัวแปรอิสระแต่ละตัวในสมการไม่มีความสัมพันธ์กับตัวแปรตาม แต่เนื่องจากตัวแปรอิสระทุกตัวที่นำมาวิเคราะห์นั้นมีความสำคัญต่อการเกิดภาวะแทรกซ้อนทางไตของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ดังนั้นจึงยังคงตัวแปรอิสระทั้ง 20 ตัว ไว้ เพื่อนำมาวิเคราะห์จำแนกประเภทต่อไป

4.2.2.2 การทดสอบค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระทั้งหมด โดยใช้วิธีอัตราส่วนภาวะน่าจะเป็น ปรากฏผลดังตารางที่ 4.12

สมมติฐานการทดสอบ

$H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0$

โดยที่ i = 1, 2,...,p

หรือ H_0 : ตัวแปรอิสระแต่ละตัวในสมการไม่มีอิทธิพลต่อตัวแปรตาม

H_1 : ตัวแปรอิสระแต่ละตัวในสมการมีอิทธิพลต่อตัวแปรตาม

ตารางที่ 4.12 ผลการทดสอบค่าสัมประสิทธิ์การถดถอยลอจิสติก

Chi - Square	df	p-value
423.400	20	< .001

จากตารางที่ 4.12 ผลการทดสอบค่าสัมประสิทธิ์การถดถอยลอจิสติกด้วยวิธีอัตราส่วนภาวะน่าจะเป็น พบว่า ไคกำลังสองมีค่าเท่ากับ 423.400 และ p-value มีค่า < .001 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ .05 จึงสรุปได้ว่า ตัวแปรอิสระแต่ละตัวในสมการมีอิทธิพลต่อการเกิดภาวะแทรกซ้อนทางไตของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 อย่างมีนัยสำคัญทางสถิติ

4.2.1.3 การตรวจสอบความเหมาะสมของสมการการวิเคราะห์การถดถอยลอจิสติก ซึ่งในงานวิจัยครั้งนี้ ผู้วิจัยตรวจสอบความเหมาะสมของสมการโดยใช้วิธีการทดสอบของ Hosmer and Lemeshow ที่มีการแบ่งข้อมูลออกเป็น p กลุ่มย่อย ปรากฏผลดังตารางที่ 4.13

สมมติฐานการทดสอบ

H_0 : สมการมีความเหมาะสมกับข้อมูล

H_1 : สมการไม่มีความเหมาะสมกับข้อมูล

เอกสารนี้เป็น... ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.13 ผลการตรวจสอบความเหมาะสมของสมการการวิเคราะห์การถดถอยลอจิสติก

Chi - Square	df	P - Value
.000	7	1.000

จากตารางที่ 4.13 ผลการทดสอบความเหมาะสมของสมการวิธีของ Hosmer and Lemeshow พบว่ามีค่าไคกำลังสองมีค่าเท่ากับ .000 และ p-value มีค่าเท่ากับ 1.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ .05 จึงปฏิเสธ H_0 สรุปได้ว่า สมการมีความเหมาะสมกับข้อมูลอย่างมีนัยสำคัญทางสถิติ

4.2.2 ผลการประเมินความเหมาะสมของสมการการวิเคราะห์การถดถอยลอจิสติก

การประเมินความเหมาะสมของสมการการวิเคราะห์การถดถอยลอจิสติกที่ใช้ในการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยพิจารณาจากตัวแปรอิสระ 20 ตัว เมื่อกำหนดจุดตัดในการแบ่งกลุ่มเป็น 0.5 โดยใช้ค่าความถูกต้องและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏผลดังตารางที่ 4.14

ตารางที่ 4.14 ผลการประเมินความเหมาะสมของตัวแบบการวิเคราะห์การถดถอยลอจิสติก

กลุ่ม	ขนาดตัวอย่าง (n_i)	ผลการพยากรณ์		ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย
		ไม่เกิดภาวะแทรกซ้อนทางไต	เกิดภาวะแทรกซ้อนทางไต	
ไม่เกิดภาวะแทรกซ้อนทางไต	200	200(100%)	0(0%)	0.000
เกิดภาวะแทรกซ้อนทางไต	120	0(0%)	120(100%)	

$$\text{ค่าความถูกต้อง} = \frac{(200 + 120) \times 100}{320} = 100\%$$

หรือ $$\text{ค่าความผิดพลาด} = \frac{(0 + 0) \times 100}{320} = 0\%$$

จากตารางที่ 4.14 จากข้อมูลจริงของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จำนวน 320 ราย ซึ่งประกอบด้วยผู้ป่วย 2 กลุ่ม คือ กลุ่มที่ไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 200 ราย เมื่อทำการจำแนกกลุ่มโดยใช้วิธีการวิเคราะห์การถดถอยลอจิสติก พบว่า จำแนกกลุ่มถูกต้องว่าไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 200 ราย และจำแนกกลุ่มผิดพลาดว่าเกิดภาวะแทรกซ้อนทางไต จำนวน 0 ราย นั่นคือ สมการการวิเคราะห์การถดถอยลอจิสติกสามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อยละ 100 และกลุ่มที่เกิดภาวะแทรกซ้อนทางไต จำนวน 120 ราย พบว่า จำแนกกลุ่มถูกต้องว่าเกิดภาวะแทรกซ้อนทางไต จำนวน 120 ราย และจำแนกกลุ่มผิดพลาดว่าไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 0 ราย นั่นคือ สมการการวิเคราะห์การถดถอยลอจิสติกสามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อย 100 ดังนั้นจึงสรุปได้ว่า สมการการวิเคราะห์การถดถอยลอจิสติกสามารถจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยเบาหวาน ชนิดที่ 2 ได้ถูกต้องคิดเป็นร้อยละ 100 และมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น

4.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision tree)

จากการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยใช้วิธีต้นไม้ตัดสินใจ ปรากฏผลดังตารางที่ 4.15

ตารางที่ 4.15 ผลการประเมินความเหมาะสมของตัวแบบต้นไม้ตัดสินใจ

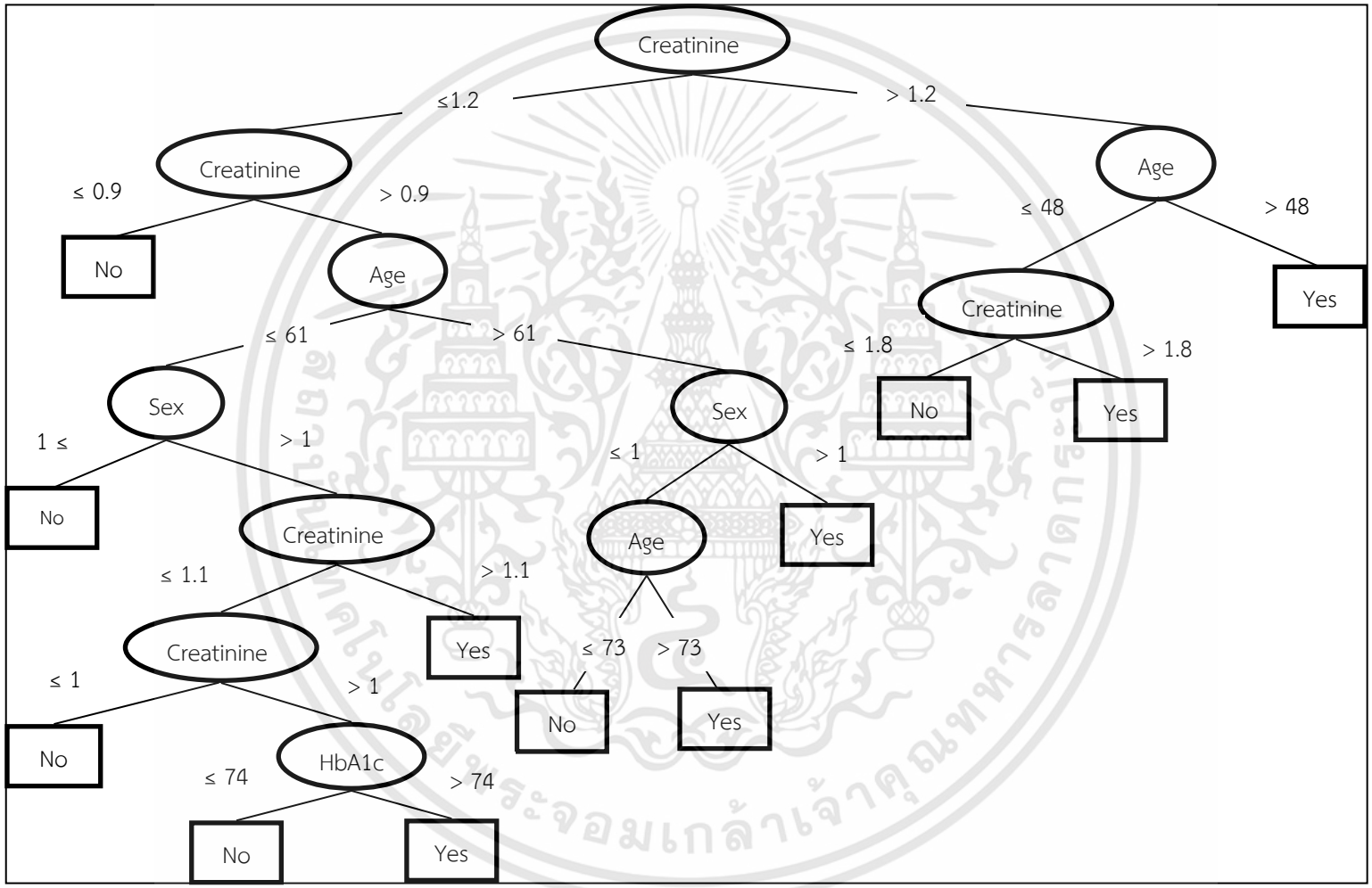
กลุ่ม	ขนาดตัวอย่าง (n_i)	ผลการพยากรณ์		ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย
		ไม่เกิดภาวะแทรกซ้อนทางไต	เกิดภาวะแทรกซ้อนทางไต	
ไม่เกิดภาวะแทรกซ้อนทางไต	200	196 (98%)	4(2%)	0.1579
เกิดภาวะแทรกซ้อนทางไต	120	4(3.33%)	116(96.7%)	

$$\text{ค่าความถูกต้อง} = \frac{(196 + 116) \times 100}{320} = 97.5\%$$

$$\text{หรือ ค่าความผิดพลาด} = \frac{(4 + 4) \times 100}{320} = 2.5\%$$

จากตารางที่ 4.15 จากข้อมูลจริงของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จำนวน 320 ราย ซึ่งประกอบด้วยผู้ป่วย 2 กลุ่ม คือ กลุ่มที่ไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 200 ราย เมื่อทำการจำแนกกลุ่มโดยใช้วิธีต้นไม้ตัดสินใจ พบว่า จำแนกกลุ่มถูกต้องว่าไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 196 ราย และจำแนกกลุ่มผิดพลาดว่าเกิดภาวะแทรกซ้อนทางไต จำนวน 4 ราย นั่นคือ ตัวแบบต้นไม้ตัดสินใจสามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อยละ 98 และกลุ่มที่เกิดภาวะแทรกซ้อนทางไต จำนวน 120 ราย พบว่า จำแนกกลุ่มถูกต้องว่าเกิดภาวะแทรกซ้อนทางไต จำนวน 116 ราย และจำแนกกลุ่มผิดพลาดว่าไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 4 ราย นั่นคือ ตัวแบบต้นไม้ตัดสินใจสามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อย 96.7 ดังนั้นจึงสรุปได้ว่า ตัวแบบต้นไม้ตัดสินใจสามารถจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยเบาหวาน ชนิดที่ 2 ได้ถูกต้องคิดเป็นร้อยละ 97.5 และมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.1579

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.4 ต้นไม้ตัดสินใจเพื่อใช้ในการจำแนกกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

จากรูปที่ 4.4 ต้นไม้ตัดสินใจเพื่อใช้ในการจำแนกกลุ่มสามารถอธิบายได้ดังนี้

กฎข้อที่ 1 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มีปริมาณต่ำกว่าหรือเท่ากับ 0.9 มิลลิกรัมต่อเดซิลิตร แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะไม่เป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 2 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร เป็นเพศชาย และอายุต่ำกว่าหรือเท่ากับ 61 ปี แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะไม่เป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 3 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร และอายุมากกว่า 61 ปี และเป็นเพศชาย แต่อายุมากกว่า 73 ปี แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะเป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 4 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร และอายุมากกว่า 61 ปี และเป็นเพศชาย แต่อายุต่ำกว่าหรือเท่ากับ 73 ปี แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะไม่เป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 5 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร และอายุมากกว่า 61 ปี และเป็นเพศหญิง แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะเป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 6 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร และอายุต่ำกว่าหรือเท่ากับ 61 ปี และเป็นเพศชาย แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะไม่เป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 7 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร และอายุต่ำกว่าหรือเท่ากับ 61 ปี และเป็นเพศหญิง แต่ครีเอตินินมีปริมาณมากกว่า 1.1 มิลลิกรัมต่อเดซิลิตร แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะเป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 8 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร และอายุต่ำกว่าหรือเท่ากับ 61 ปี เพศหญิง แต่ครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.1 มิลลิกรัมต่อเดซิลิตร และน้อยกว่าหรือเท่ากับ 1 มิลลิกรัมต่อเดซิลิตร แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะไม่เป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 9 : ถ้าครีเอตินิน มีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร และอายุต่ำกว่าหรือเท่ากับ 61 ปี เพศหญิง แต่ครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.1 มิลลิกรัมต่อเดซิลิตร และมากกว่า 1 มิลลิกรัมต่อเดซิลิตร และ HbA1c มีปริมาณน้อยกว่าหรือเท่ากับ 74 มิลลิกรัมต่อเปอร์เซ็นต์ แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะไม่เป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 10 : ถ้าครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.2 มิลลิกรัมต่อเดซิลิตร แต่มากกว่า 0.9 มิลลิกรัมต่อเดซิลิตร และอายุต่ำกว่าหรือเท่ากับ 61 ปี เพศหญิง แต่ครีเอตินินมีปริมาณน้อยกว่าหรือเท่ากับ 1.1 มิลลิกรัมต่อเดซิลิตร และมากกว่า 1 มิลลิกรัมต่อเดซิลิตร และ HbA1c มีปริมาณมากกว่า 74 มิลลิกรัมต่อเปอร์เซ็นต์ แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะเป็นภาวะแทรกซ้อนทางไต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กฎข้อที่ 11 : ถ้าครีเอตินินมีปริมาณมากกว่า 1.2 มิลลิกรัมต่อเดซิลิตร และอายุมากกว่า 48 ปี แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะเป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 12 : ถ้าครีเอตินินมีปริมาณมากกว่า 1.2 มิลลิกรัมต่อเดซิลิตร และอายุต่ำกว่าหรือเท่ากับ 48 ปี และครีเอตินินมีปริมาณมากกว่า 1.8 มิลลิกรัมต่อเดซิลิตร แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะเป็นภาวะแทรกซ้อนทางไต

กฎข้อที่ 13 : ถ้าครีเอตินินมีปริมาณมากกว่า 1.2 มิลลิกรัมต่อเดซิลิตร และอายุต่ำกว่าหรือเท่ากับ 48 ปี และครีเอตินินมีปริมาณต่ำกว่าหรือเท่ากับ 1.8 มิลลิกรัมต่อเดซิลิตร แล้วผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จะไม่เป็นภาวะแทรกซ้อนทางไต

4.5 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม

จากการศึกษาการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ โดยใช้การวิเคราะห์จำแนกประเภท การวิเคราะห์การถดถอยลอจิสติก และต้นไม้ตัดสินใจ ผลการเปรียบเทียบปริทธิภาพของวิธีการจำแนกกลุ่ม ปรากฏผลดังตารางที่ 4.16

ตารางที่ 4.16 ผลการเปรียบเทียบประสิทธิภาพการพยากรณ์โดยใช้ค่าความถูกต้องและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย

วิธีการจำแนกกลุ่ม	ค่าความถูกต้อง	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย
1) การวิเคราะห์จำแนกประเภท	96.875%	0.1768
2) การวิเคราะห์การถดถอยลอจิสติก	100.00%	0.0000
3) ต้นไม้ตัดสินใจ	97.50%	0.1579

จากตารางที่ 4.16 ผลการเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มเมื่อพิจารณาจากค่าความถูกต้อง พบว่า วิธีการวิเคราะห์ถดถอยลอจิสติก มีค่าความถูกต้องมากที่สุดคือ สามารถจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ได้ถูกต้องคิดเป็นร้อยละ 100 และมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.0000 รองลงมาคือ วิธีต้นไม้ตัดสินใจสามารถจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ได้ถูกต้องคิดเป็นร้อยละ 97.5 และมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.1579 และวิธีการวิเคราะห์จำแนกประเภทสามารถจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ได้ถูกต้องคิดเป็นร้อยละ 96.875 และมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.1768 ตามลำดับ จึงสรุปว่าวิธีการจำแนกกลุ่มที่มีประสิทธิภาพสูงที่สุดและเหมาะสมสำหรับจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 คือ การวิเคราะห์ถดถอยลอจิสติก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 โดยข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลทุติยภูมิของผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ที่มารับการรักษาในโรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ ตั้งแต่เดือนตุลาคม พ.ศ.2559 – กันยายน พ.ศ.2560 เป็นระยะเวลา 12 เดือน จำนวน 320 ราย ซึ่งประกอบด้วยผู้ป่วย 2 กลุ่ม คือ กลุ่มที่ไม่เกิดภาวะแทรกซ้อนทางไต จำนวน 200 ราย และกลุ่มที่เกิดภาวะแทรกซ้อนทางไต จำนวน 120 ราย โดยใช้วิธีการจำแนกกลุ่ม 3 วิธี คือ การวิเคราะห์จำแนกประเภท การวิเคราะห์ถดถอยลอจิสติกและต้นไม้ตัดสินใจในการจำแนกกกลุ่มผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ว่าจะเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไต โดยพิจารณาจากค่าความถูกต้องและค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งวิธีการจำแนกกลุ่มที่มีค่าความถูกต้องสูงที่สุดและมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด จะสรุปผลได้ว่าวิธีการจำแนกกลุ่มนั้นมีประสิทธิภาพสูงที่สุด ซึ่งสรุปผลการวิจัยได้ดังนี้

จากผลการเปรียบเทียบประสิทธิภาพการจำแนกกลุ่ม เมื่อพิจารณาจากค่าความถูกต้องและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย พบว่า วิธีการจำแนกกลุ่มที่มีค่าความถูกต้องมากที่สุดและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด คือ การวิเคราะห์ถดถอยลอจิสติก ซึ่งสามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อยละ 100 และมีค่าค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.000 รองลงมาคือ ต้นไม้ตัดสินใจ สามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อยละ 97.5 และมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.1579 และการวิเคราะห์จำแนกประเภท สามารถจำแนกกลุ่มได้ถูกต้องคิดเป็นร้อยละ 96.875 และมีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 0.1768 ตามลำดับ ดังนั้นจึงสรุปได้ว่า วิธีการจำแนกกลุ่มที่มีประสิทธิภาพสูงที่สุดและเหมาะสมสำหรับการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 คือ การวิเคราะห์ถดถอยลอจิสติก ซึ่งเป็นไปตามวัตถุประสงค์ของงานวิจัยในครั้งนี้ คือ ได้ทราบผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มทั้ง 3 วิธี ว่าวิธีการจำแนกกลุ่มวิธีใดที่มีประสิทธิภาพสูงที่สุด และสามารถนำผลการวิจัยดังกล่าวไปใช้เพื่อจำแนกกลุ่มผู้ป่วยโรคเบาหวาน ชนิดที่ 2 รายใหม่ ในกรณีที่ผู้ป่วยยังไม่ได้รับการตรวจรักษาภาวะแทรกซ้อนทางไตได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 อภิปรายผล

จากสรุปผลการวิจัยในครั้งนี้ พบว่า วิธีการวิเคราะห์จำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ที่มีประสิทธิภาพสูงที่สุดคือ การวิเคราะห์การถดถอยลอจิสติก รองลงมาคือ ต้นไม้ตัดสินใจ และการจำแนกประเภท ตามลำดับ ซึ่งจากสรุปผลการวิจัยดังกล่าว สามารถนำวิธีการจำแนกกลุ่มที่มีประสิทธิภาพสูงที่สุดมาเพื่อเป็นแนวทางสำหรับผู้ที่สนใจศึกษาหรือหน่วยงาน/องค์กรที่เกี่ยวข้อง เช่น โรงพยาบาล กระทรวงสาธารณสุข เป็นต้น โดยสามารถนำวิธีการจำแนกกลุ่มที่มีประสิทธิภาพสูงที่สุดไปใช้ในการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ได้ และสามารถนำผลลัพธ์ที่ได้ไปใช้ในการสนับสนุนการตัดสินใจในทางการแพทย์เพื่อวินิจฉัยการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 ในกรณีที่ผู้ป่วยยังไม่ได้รับการตรวจวินิจฉัยภาวะแทรกซ้อนทางไตได้ ว่าจะเกิดหรือไม่เกิดภาวะแทรกซ้อนทางไต เพื่อให้ได้ผลการวินิจฉัยที่รวดเร็วยิ่งขึ้น สามารถรักษาได้ทันท่วงที และเพื่อเป็นแนวทางในการป้องกันไม่ให้ผู้ป่วยโรคเบาหวาน ชนิดที่ 2 เกิดภาวะแทรกซ้อนทางไตเพิ่มขึ้น

ซึ่งจากวัตถุประสงค์ของการวิจัยในครั้งนี้ และผลการวิจัยพบว่า การวิเคราะห์การถดถอยลอจิสติก มีประสิทธิภาพในการจำแนกกลุ่มสูงกว่าการวิเคราะห์จำแนกประเภท และต้นไม้ตัดสินใจ ซึ่งจากผลการวิจัยดังกล่าว จะเห็นได้ว่าผลการวิจัยที่ได้ไม่สอดคล้องกับงานวิจัยที่ผู้วิจัยได้ศึกษาที่ผ่านมาเกี่ยวกับการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่ม โดยใช้วิธีการจำแนกกลุ่มทั้ง 3 วิธี เช่น การศึกษางานวิจัยเกี่ยวกับการเปรียบเทียบการวินิจฉัยโรคตับด้วยการวิเคราะห์จำแนกประเภทและการวิเคราะห์การถดถอยลอจิสติก กรณีศึกษาผู้ป่วยโรงพยาบาลฉะเชิงเทรา ผลการวิจัยพบว่าการวิเคราะห์จำแนกประเภทมีประสิทธิภาพสูงกว่าการวิเคราะห์การถดถอยลอจิสติก (ศยามล ลำลองรัตน์, 2550) และศึกษาการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรังโรงพยาบาลอพอลโล ประเทศอินเดีย โดยใช้วิธีความใกล้เคียงกันมากที่สุด วิธีต้นไม้ตัดสินใจ วิธีโครงข่ายประสาทเทียม วิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีฐานกฎ วิธีการถดถอยลอจิสติก และวิธีนาอีฟเบย์ ผลการวิจัยพบว่า วิธีการจำแนกกลุ่มที่มีประสิทธิภาพการจำแนกดีที่สุดคือ วิธีต้นไม้ตัดสินใจ (สุรวัชร ศรีเปารยะ, 2559) ซึ่งอาจจะมีสาเหตุมาจากข้อมูลที่ใช้ในการศึกษาครั้งนี้มีจำนวนน้อยเกินไป หรือข้อมูลมีค่าความผิดพลาด หรือวิธีการจำแนกกลุ่มที่เหมาะสมสำหรับจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 คือ การวิเคราะห์การถดถอยลอจิสติกมากกว่าการวิเคราะห์จำแนกประเภทและต้นไม้ตัดสินใจ จึงทำให้ผลการวิจัยที่ได้ไม่สอดคล้องกับงานวิจัยดังกล่าว ดังนั้นในการศึกษาครั้งต่อไป ผู้วิจัยสามรถนำการอภิปรายผลไปใช้เป็นแนวทางในการปรับปรุงงานวิจัยให้มีประสิทธิภาพยิ่งขึ้น เช่น อาจจะเพิ่มจำนวนข้อมูล หรือตรวจสอบความถูกต้องของข้อมูลกับทางหน่วยงานที่ทำการเก็บรวบรวมข้อมูล ว่ามีความถูกต้อง ครบถ้วนสมบูรณ์หรือข้อมูลมีความผิดพลาดหรือไม่ เพื่อให้ผู้วิจัยสามารถวิเคราะห์ถึงปัญหาที่เกิดขึ้นได้และสามารถแก้ไขปัญหาได้อย่างถูกต้อง

เอกสารนี้เป็น... ไม่ว่ากรณี...

5.3 ข้อเสนอแนะ

5.3.1 จากการตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์จำแนกประเภทเกี่ยวกับตัวแปรอิสระของแต่ละกลุ่มมีการแจกแจงปรกติหลายตัวแปร และเมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระของแต่ละกลุ่มเท่ากัน พบว่า ข้อสมมติเบื้องต้นดังกล่าวไม่เป็นจริง เนื่องจากข้อมูลที่นำมาใช้ในวิเคราะห์อาจจะมีจำนวนน้อยเกินไป หรือข้อมูลเป็นข้อมูลทุติยภูมิ ผู้วิจัยไม่ได้เก็บรวบรวมเอง ซึ่งอาจจะทำให้ข้อมูลเกิดความผิดพลาด ซึ่งผู้วิจัยได้ทำการแก้ปัญหาโดยการปรับข้อมูลแล้ว แต่ข้อมูลก็ไม่เป็นไปตามข้อสมมติเบื้องต้นดังกล่าว ดังนั้นเพื่อให้ผลการวิเคราะห์มีประสิทธิภาพ ในการศึกษาครั้งต่อไป ผู้วิจัยควรเลือกใช้วิธีการจำแนกกลุ่มวิธีอื่นที่ไม่มีข้อสมมติเบื้องต้นดังกล่าว เช่น การวิเคราะห์การถดถอยลอจิสติก เป็นต้น

5.3.2 ควรนำข้อมูลการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเหวาน ชนิดที่ 2 ของโรงพยาบาลอื่น ๆ มาทำการวิเคราะห์ด้วย เพื่อเปรียบเทียบผลการวิเคราะห์ว่ามีความสอดคล้องกันหรือไม่

5.3.3 เนื่องจากงานวิจัยในครั้งนี้ใช้ข้อมูลทุติยภูมิ ซึ่งมีการเก็บข้อมูลตัวแปรอิสระที่คาดว่าจะส่งผลต่อการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 เพียง 23 ตัว ในการศึกษาครั้งต่อไปผู้วิจัยอาจจะเพิ่มตัวแปรอิสระที่เกี่ยวข้องกับการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 จากการสัมภาษณ์เพิ่ม เช่น มีประวัติครอบครัวเป็นโรคไตจากเบาหวาน โรคความดันโลหิตสูง การออกกำลังกาย การรับประทานอาหาร เป็นต้น เพื่อเพิ่มประสิทธิภาพในการจำแนกกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

กัลยา วานิชย์บัญชา. 2552. **การวิเคราะห์ข้อมูลตัวแปรหลายตัว**. พิมพ์ครั้งที่ 4. กรุงเทพมหานคร : ธรรมสาร.

สายชล สินสมบูรณ์ทอง. 2559. **การวิเคราะห์ตัวแปรหลายตัว**. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร : จามจุรีโปรดักส์.

ยุทธ ไกยวรรณ์. 2557. **การวิเคราะห์สถิติหลายตัวแปรสำหรับงานวิจัย**. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร : วี.พริ้นท์ (1991).

กัลยา วานิชย์บัญชา. 2558. **การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows**. พิมพ์ครั้งที่ 10. กรุงเทพมหานคร : หจก.สามลดา.

สายชล สินสมบูรณ์ทอง. 2558. **การทำเหมืองข้อมูล**. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร : จามจุรีโปรดักส์.

สายชล สินสมบูรณ์ทอง. 2559. **การทำเหมืองข้อมูล เล่ม 2 วิธีการและตัวแบบ**. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร : จามจุรีโปรดักส์.

ทรงศักดิ์ ภูสีอ่อน. 2551. การประยุกต์ใช้ SPSS วิเคราะห์ข้อมูลงานวิจัย. พิมพ์ครั้งที่ 2. กาฬสินธุ์ : ประสานการพิมพ์.

พัชรี เพ็งประโคน. 2539. การเปรียบเทียบผลการวิเคราะห์ข้อมูลระหว่างวิธีวิเคราะห์จำแนกกับวิธีวิเคราะห์ถดถอยลอจิสติก : กรณีศึกษาตัวแปรจำแนกวิถีชีวิตแบบประชาธิปไตยของนิสิตนักศึกษา. สาขาวิชาวิจัยการศึกษา ภาควิจัยการศึกษา จุฬาลงกรณ์มหาวิทยาลัย.

ศยามล ลำลองรัตน์. 2550. ศึกษาเปรียบเทียบการวินิจฉัยโรคตับด้วยการวิเคราะห์จำแนกกลุ่มและการวิเคราะห์ถดถอยโลจิสติก กรณีศึกษาผู้ป่วยโรงพยาบาลฉะเชิงเทรา. สาขาสถิติ ภาควิชาสถิติ มหาวิทยาลัยเกษตรศาสตร์.

ชลธิศา พลทองมาก. 2553. การวิเคราะห์ความเสี่ยงการเป็นโรคไวรัสตับอักเสบซี โดยต้นไม้การตัดสินใจ. สาขาวิชาเทคโนโลยีสารสนเทศ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

สิริมา เยาวสิทธิ์. 2553. การเปรียบเทียบวิธีการทางสถิติและการโปรแกรมเชิงเป้าหมายสำหรับจำแนกกลุ่มกำไรขาดทุนของสหกรณ์การเกษตร. ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

สุริยะ ตรีราภี. 2553. การศึกษาเปรียบเทียบการทานายความเสี่ยงต่อโรคฝันผุโดยใช้การเรียนรู้เบย์อย่างง่าย ต้นไม้การตัดสินใจและโครงข่ายประสาทเทียม. สาขาวิชาวิทยาการคอมพิวเตอร์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต.

รุจิรา ธรรมสมบัติ. 2554. ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือโดยใช้ต้นไม้ตัดสินใจ. สาขาคอมพิวเตอร์ธุรกิจ คณะบริหารธุรกิจ มหาวิทยาลัยราชพฤกษ์.

อภิวัฒน์ อ่อนสุระทุม. 2555. การวินิจฉัยโรคไข้เลือดออกด้วยเทคนิคต้นไม้ตัดสินใจและเครือข่ายความเชื่อเบย์. สาขาวิชาเทคโนโลยีสารสนเทศ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กนกพัฒน์ ลักษณานุรักษ์. 2556. การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลผู้ป่วยโรคความดันโลหิตสูงโดยใช้เทคนิคการถดถอยโลจิสติก ซัพพอร์ตเวกเตอร์แมชชีน และนาอีฟเบย์. สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ปภากร อินทร์โท่โล่. 2557. การวิเคราะห์ถดถอยโลจิสติกเพื่อการพยากรณ์การพบเชื้อซาลโมเนลล่าเอนเทอริกาในสุกร. ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

ชณัฐดาภรณ์ เย็นประเสริฐ. 2557. การเปรียบเทียบความแม่นยาการพยากรณ์สถานการณ์ชาระหนี้ของลูกหนี้ โดยใช้เทคนิคการถดถอยลอจิสติก นาอีฟย์เบย์ และต้นไม้การตัดสินใจ. สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

สุรวัชร ศรีเปารยะ. 2559. การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย. สาขาวิชาสถิติประยุกต์ ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

พยูน พาณิชย์กุล. 2548. การพัฒนาระบบดาต้าไมน์นิ่งโดยใช้ Decision Tree. โครงการพัฒนาระบบงานปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนงวิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

ขจรศักดิ์ ศรีอ่อน. 2552. การทำนายสาเหตุของเหตุการณ์กระแสไฟฟ้าขัดข้องโดยใช้เทคนิคการทำเหมืองข้อมูลในระบบจำหน่ายของการไฟฟ้าส่วนภูมิภาค เขต 1 ภาคกลาง. มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพมหานคร.

สมาคมโรคเบาหวานแห่งประเทศไทยฯ. 2558. **ประเด็นสารรณรงค์วันเบาหวานโลก ปี 2558.** เข้าถึงได้จาก http://www.dmthai.org/

สำนักโรคไม่ติดต่อ. 2557. **จำนวนและอัตราตายด้วยโรคไม่ติดต่อและการบาดเจ็บประจำปีปฏิทิน พ.ศ.2557**.เข้าถึงได้จาก http://thaincd.com/information-statistic/non-communicable-disease-data.php

กลุ่มงานควบคุมโรค สำนักงานสาธารณสุขจังหวัดบุรีรัมย์. 2559 **สถานการณ์โรคเบาหวาน ความดันโลหิตสูง ปี 2559.** 17 มีนาคม 2560**.**

อัจจิมา กาญจนาภา. 2558. โรคไตจากเบาหวานกับวิธีการตรวจ. **R&D NEWSLETTER**, 22(4), 22-24.

เทพ หิมะทองคำ. 2547. **ความรู้เรื่องโรคเบาหวานฉบับสมบูรณ์.** พิมพ์ครั้งที่ 4. กรุงเทพมหานคร : บริษัทวิทยาพัฒน์ จำกัด

พานิช แกมนิล. 2553. พฤติกรรมการบริโภคอาหารของผู้ป่วยเบาหวาน อำเภอสามชัย จังหวัดกาฬสินธุ์. คณะสาธารณสุขศาสตร์ มหาวิทยาลัยบูรพา.

บุญทิพย์ สิริธรังศรี. 2539. **ผู้ป่วยเบาหวาน : การดูแลตนเองแบบองค์รวม.** พิมพ์ครั้งที่ 3. ฝ่ายการพิมพ์ ศูนย์อาเยน มหาวิทยาลัยมหิดล.

เบญจวรรณ เกิดแพร. 2549. การพยาบาลผู้ป่วยโรคเบาหวาน (กรณีศึกษา). ศูนย์อนามัยที่ 9 พิษณุโลก กรมอนามัย กระทรวงสาธารณสุข.

เกตุแก้ว จันทร์จำรัส. 2559. การสำรวจความชุกของผู้ป่วยเบาหวานชนิดที่ 2 ที่มีภาวะแทรกซ้อนที่ไตในโรงพยาบาลสมเด็จพระปิ่นเกล้า. **วารสารเทคนิคการแพทย์.** 44(1), 5508-5521

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Kritiyaporn Kunsook 2012. Classification of hydroponics system case study : deep flow technique (DFT) and nutrient film tecnique (NFT). Master of Science Technology of Information System Management. Mahidol University

Pimpa Cheewaprakobkit. 2015. Predicting Student Academic Achievement by Using the Decision Tree and Neural Network Techniques. **Catalyst**. 12(2) : 34–43.

Barbara, G. and S. Linda. 2007. **Using Multivariate Statistics**. 5th ed. Inc., Boston, Pearson Education.

Johnson, A. 2002. **Applied Multivariate Statistical Analysis**. 5th ed. Inc., New Jersey, Prentice-Hall

Johnson, Richard A. and Wichern Dean W. 1988. Applied Multivariate Statistical Analysis. 2nd ed. Englewood Cliffs, New Jersey : Prentice Hall.

Hosmer, D.W. & Lemeshow, S. 1989. Applied Logistic Regression. New York: John Wiley & Sons.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก
รายละเอียดและคุณลักษณะของข้อมูลและตัวอย่างข้อมูลที่นำมาวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก-1 รายละเอียดและคุณลักษณะของข้อมูลของผู้ป่วยโรคเบาหวานชนิดที่ 2

ตัวแปร	ความหมาย	หน่วย	มาตราวัด
X_1	เพศ	ชาย, หญิง	นามบัญญัติ
X_2	อายุ	ปี	อัตราส่วน
X_3	น้ำหนัก	กิโลกรัม	อัตราส่วน
X_4	ส่วนสูง	เซนติเมตร	อัตราส่วน
X_5	ดัชนีมวลกาย	กิโลกรัมต่อเมตร^2	อัตราส่วน
X_6	รอบเอว	นิ้ว	อัตราส่วน
X_7	ความดันโลหิตซีสโตลิค	มิลลิเมตรปรอท	อัตราส่วน
X_8	ความดันโลหิตไดแอสโตลิค	มิลลิเมตรปรอท	อัตราส่วน
X_9	Creatinine	มิลลิกรัมต่อเดซิลิตร	อัตราส่วน
X_{10}	HbA1c	มิลลิกรัมต่อเปอร์เซ็นต์	อัตราส่วน
X_{11}	Triglyceride	มิลลิกรัมต่อเดซิลิตร	อัตราส่วน
X_{12}	LDL Cholesterol	มิลลิกรัมต่อเดซิลิตร	อัตราส่วน
X_{13}	HDL Cholesterol	มิลลิกรัมต่อเดซิลิตร	อัตราส่วน
X_{14}	Total Cholesterol	มิลลิกรัมต่อเดซิลิตร	อัตราส่วน
X_{15}	Urine protein	มิลลิกรัมต่อเดซิลิตร	อัตราส่วน
X_{16}	Uric Acid	มิลลิกรัมต่อเปอร์เซ็นต์	อัตราส่วน
X_{17}	Albuminuria	ไมโครกรัมต่อนาที	อัตราส่วน
X_{18}	FBS	มิลลิกรัมต่อเดซิลิตร	อัตราส่วน
X_{19}	BUN	มิลลิกรัมต่อเปอร์เซ็นต์	อัตราส่วน
X_{20}	ระยะเวลาที่ป่วยโรคเบาหวาน ชนิดที่ 2	ปี	อัตราส่วน
X_{21}	การดื่มสุรา	ดื่ม, ไม่ดื่ม	นามบัญญัติ
X_{22}	การสูบบุหรี่	สูบ, ไม่สูบ	นามบัญญัติ
X_{23}	ภาวะแทรกซ้อนทางตา	มี, ไม่มี	นามบัญญัติ
Y	ภาวะแทรกซ้อนทางไต	เกิด, ไม่เกิด	นามบัญญัติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก-2 ตัวอย่างข้อมูลของผู้ป่วยโรคเบาหวาน ชนิดที่ 2

ID	Sex	Age	Weight	Height	BMI	รอบเอว	SBP	DBP	HbA1c	FBS	Cholesterol	LDLCholesterol
1	2	69	68	160	25	80	130	80	9	119	196	211
2	1	70	62	165	22.41	84	106	60	8.2	142	185	171
3	2	61	70	160	23.83	89	143	60	6.5	100	234	225
4	2	66	51	155	21.64	84	134	69	7.7	106	138	84
5	2	82	48	158	23.72	73	120	80	13.5	143	208	230
6	1	62	54	157	25.56	83	130	80	6.4	142	162	180
7	2	46	55	148	27.85	87	109	60	8.8	128	179	80
8	1	71	61	155	25.81	90	130	80	9.2	197	225	107
9	1	66	74	163	27.85	92	155	90	4.2	106	216	133
10	2	64	67	153	24.97	100	130	80	8.1	64	255	454
11	2	66	60	155	24.97	96	140	75	10.9	134	217	120
12	1	71	51	170	17.65	77	130	80	5.7	154	137	92
13	2	63	50	165	18.36	80	130	80	10.1	148	271	179
14	2	73	52	155	21.36	77	130	80	11.3	251	184	244
15	1	60	55	158	20.7	78	127	69	5.8	112	167	101
16	2	64	55	150	25.78	93	125	72	9.5	195	177	191
17	2	59	68	160	26.23	95	125	67	5.2	155	225	153
18	2	67	64	165	23.14	88	120	70	6.8	128	252	85
19	2	70	40	145	21.4	69	110	70	6.9	124	256	111

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	Sex	Age	Weight	Height	BMI	รอบเอว	SBP	DBP	HbA1c	FBS	Cholesterol	LDLCholesterol
20	1	62	63	170	21.11	82	134	71	7.2	237	146	98
21	1	57	64	165	24.98	92	129	76	6.2	237	129	250
22	2	62	40	158	18.98	64	120	70	7.1	138	200	112
23	2	60	40	150	20.89	72	130	80	7.3	89	144	314
24	2	75	54	150	24.44	84	110	80	4.4	112	146	85
25	2	76	80	165	27.55	92	120	80	7.3	161	168	162
26	2	59	54	140	25.21	92	128	79	8.8	133	170	276
27	2	68	40	155	17.78	62	90	60	9.2	247	253	150
28	2	39	58	165	21.3	80	132	86	10	149	181	100
29	1	65	65	170	25.86	92	120	70	14	178	217	108
30	2	58	66	150	30.67	104	130	80	5.8	219	190	422
31	2	67	48	148	22.35	82	130	80	9.5	208	124	72
32	2	51	52	155	21.64	72	130	77	8.6	142	214	92
33	2	69	60	165	22.66	90	131	69	7.7	140	180	139
34	2	67	54	146	23.56	80	130	80	8.3	116	230	242
35	2	49	67	156	31.63	87	126	80	6	157	144	145
36	2	51	70	150	30.22	91	116	69	9.5	184	210	162
37	2	63	56	145	29.01	93	130	80	5.6	107	209	100
38	1	81	61	172	20.96	82	110	70	11	279	309	180
39	1	65	71	165	23.14	97	130	80	11.5	180	163	183
40	1	75	77	170	25.8	100	142	80	5.3	144	166	202

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	Sex	Age	Weight	Height	BMI	รอบเอว	SBP	DBP	HbA1c	FBS	Cholesterol	LDLCholesterol
41	2	53	50	150	21.78	71	107	70	7.8	160	225	146
42	2	70	81	156	34.11	104	130	80	6.9	88	213	116
43	2	59	52	160	21.48	77	130	80	10.4	144	239	143
44	2	52	66	159	27.24	88	135	80	7	137	173	139
45	2	69	60	155	24.97	92	127	87	7.8	140	230	240
46	1	75	0	162	20.7	81	110	70	5.1	123	211	305
47	1	71	63	150	23.05	84	130	70	10.6	139	273	97
48	1	50	66	172	22.31	82	103	60	9.6	274	147	185
49	1	70	70	167	24.74	86	120	80	7.7	148	236	576
50	1	66	74	165	26.81	92	130	80	7.1	180	187	59
51	2	68	62	155	25.81	93	130	80	10.2	121	224	283
52	2	65	65	155	23.51	90	150	90	7.5	150	222	149
53	2	58	60	157	27.06	80	116	71	9.8	241	206	155
54	2	47	53	154	23.61	67	130	80	7.5	177	182	55
55	1	70	69	157	26.04	88	130	80	12.5	120	211	292
56	2	68	56	147	28.23	92	135	71	5.2	90	224	399
57	2	69	53	159	22.55	82	140	70	5.4	197	201	196
58	2	67	60	160	23.05	87	130	60	6.5	122	213	266
59	2	68	55	152	24.78	82	126	76	11.1	144	176	107
60	1	60	54	160	19.14	80	116	63	5.9	134	97	118
61	2	72	44	156	18.9	77	160	90	8.1	108	192	120

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	Sex	Age	Weight	Height	BMI	รอบเอว	SBP	DBP	HbA1c	FBS	Cholesterol	LDLCholesterol
62	2	50	48	145	25.56	82	130	90	7	186	179	81
63	1	67	70	171	23.45	90	130	80	6.1	154	211	452
64	1	66	79	170	27.68	99	120	80	8.7	177	158	198
65	1	71	69	168	23.46	83	110	70	5.3	103	222	142
66	1	71	85	160	33.2	105	150	90	6.5	135	199	100
67	1	55	88	160	33.59	102	165	95	11.7	230	168	213
68	2	58	58	160	22.66	86	130	70	12.9	140	201	102
69	1	56	73	160	28.12	100	136	75	4	97	164	433
70	2	69	61	155	20.57	87	130	80	6.8	154	179	349
71	1	63	57	163	20.32	82	132	82	6.4	137	133	156
72	1	62	65	167	23.31	85	108	64	5.3	129	226	209
73	2	77	65	156	28.76	100	100	60	6.6	102	145	165
74	2	68	58	149	29.55	92	130	80	7	116	150	90
75	2	73	50	150	25.43	78	146	64	6.1	109	241	190
76	1	55	58	173	24.91	96	131	91	10.7	214	176	114
77	2	66	57	155	24.97	80	128	76	10.2	110	204	212
78	2	57	75	160	29.69	96	130	80	9.1	179	171	163
79	2	42	49	144	22.35	78	114	70	10.9	234	256	94
80	2	67	51	155	21.23	76	129	72	5.3	135	114	181
81	2	71	57	156	23.83	83	98	60	5.6	120	184	87
82	1	62	70	164	25.65	92	120	80	6.5	134	276	387

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	Sex	Age	Weight	Height	BMI	รอบเอว	SBP	DBP	HbA1c	FBS	Cholesterol	LDLCholesterol
83	1	59	47	159	18.59	74	130	80	6.2	114	194	86
84	2	69	72	160	27.73	107	160	90	8.3	138	240	192
85	1	75	90	168	34.26	102	130	70	9	130	209	142
86	2	67	80	155	32.88	97	130	70	7.7	143	194	176
87	2	81	44	155	17.9	80	160	76	10.5	193	227	564
88	1	77	61	150	27.33	91	130	70	14	143	230	151
89	2	57	82	162	31.24	91	148	82	9.2	104	180	190
90	1	58	57	154	22.77	83	130	70	10.2	274	208	178
91	2	53	78	158	30.44	105	130	75	6	134	220	158
92	2	63	53	140	21.93	83	139	81	7.5	134	144	206
93	2	66	60	150	26.67	93	100	60	6.8	135	190	127
94	2	63	67	150	29.78	95	123	81	8	165	139	151
95	1	68	67	165	26.08	90	130	80	6.2	113	145	207
96	2	59	60	154	24.14	80	118	54	9.6	184	149	167
97	2	58	52	145	25.68	81	120	80	5.8	265	186	91
98	2	58	56	160	22.27	81	130	80	8.1	128	217	142
99	2	67	67	157	25.39	86	90	60	7	89	201	68
100	1	47	78	169	26.26	93	120	80	9.5	159	230	314
101	2	44	62	160	24.22	84	148	78	7	191	168	147
102	2	57	57	156	23.42	82	112	70	9.3	118	199	120
103	1	52	82	172	27.68	102	120	80	5.6	168	218	343

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	Sex	Age	Weight	Height	BMI	รอบเอว	SBP	DBP	HbA1c	FBS	Cholesterol	LDLCholesterol
104	2	69	61	156	24.13	82	135	61	6.9	181	173	277
105	1	47	72	162	28.96	80	130	80	6.7	182	182	287
106	2	70	48	150	23.28	85	120	80	8.3	114	176	90
107	2	62	63	162	24.39	82	140	80	7.8	141	235	138
108	2	58	67	160	26.17	82	130	80	5.9	142	182	443
109	1	58	87	165	30.49	102	153	74	7.5	141	147	157
110	1	78	68	156	29.14	100	132	80	6.5	145	278	340
111	2	66	62	159	24.92	87	130	80	6.1	114	176	248
112	2	77	60	160	24.22	77	120	66	6.4	137	227	91
113	2	57	60	157	24.34	90	123	62	6.8	164	208	189
114	2	59	45	153	20.5	72	130	70	8	138	212	284
115	2	80	35	146	16.89	70	120	70	8.5	83	209	183
116	2	71	56	150	26.67	80	164	85	6.3	133	154	121
117	2	39	80	160	29.69	96	120	80	6.3	108	187	115
118	2	52	57	150	25.99	82	135	81	6	153	140	85
119	2	73	43	150	20.44	78	126	70	6.3	108	162	159
120	1	78	50	160	19.53	74	150	87	7.3	131	147	75
121	2	63	36	145	18.55	69	90	60	6.6	139	141	63
122	2	65	61	150	25.78	87	130	80	7.4	149	171	182
123	1	63	60	160	23.44	87	130	80	10.9	134	196	178
124	2	53	61	145	29.01	93	165	90	5.8	189	191	474

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ตารางที่ ก-2 ตัวอย่างข้อมูลของผู้ป่วยโรคเบาหวานชนิดที่ 2 (ต่อ)

ID	HDL Cholesterol	Triglyceride	Microalbumin	Creatine	Urineprotein	Uric Acid	Albuminuria	BUN	ระยะเวลาที่ป่วยเป็นโรคเบาหวานชนิดที่ 2	Alcohol	Smoking	DKA
1	31	106	10	0.9	0	4.2	10	13	7	0	0	no
2	40	111	10	1.1	0	6.9	80	20	14	0	0	no
3	35	111	10	2.1	0	10.5	80	33	24	0	0	yes
4	40	63	300	1.2	1	3.7	300	13	20	0	0	no
5	34	105		1.6	2	4.1	1	30	37	0	0	yes
6	66	64	10	1.8	0	8.1	10	24	12	0	0	no
7	52	116	10	0.8	0	3.4	30	5	13	0	0	no
8	49	146	1000	1.2	2	6.4	300	15	4	0	0	no
9	50	138	80	1.3	1	6	300	20	11	0	0	yes
10	47	132	10	0.9	0	4.4	30	11	4	0	0	no
11	45	142	150	1	0	4	30	24	11	0	0	no
12	44	86	80	1	0	2.7	80	7	6	0	0	no
13	37	148	80	1	0	2.9	30	12	21	0	0	no
14	36	89	150	1.4	0	5.5	150	21	17	0	0	yes
15	47	100	10	1.1	0	4.7	80	16	5	0	0	no
16	32	114	80	1.2	0	6.4	150	21	17	0	0	yes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	HDL Cholesterol	Triglyceride	Microalbumin	Creatine	Urineprotein	Uric Acid	Albuminuria	BUN	ระยะเวลาที่ป่วยเป็นโรคเบาหวานชนิดที่ 2	Alcohol	Smoking	DKA
17	63	120	1000	1.6	2	6	300	18	11	0	0	no
18	65	115	80	1	0	5.3	10	17	11	0	0	no
19	64	162	300	0.8	0	3.1	30	25	11	0	0	no
20	48	88	30	1.3	0	3.3	10	15	23	0	0	no
21	28	58	30	1	0	6.9	30	15	19	0	0	no
22	73	78	10	0.9	0	2.4	10	13	18	0	0	no
23	23	72	30	1.1	0	3.1	80	21	12	0	0	no
24	25	88	10	1.3	0	4.9	10	12	11	0	0	yes
25	41	76	10	1	0	6.4	10	9	7	0	0	yes
26	39	75.8	10	0.9	0	3.3	10	13	6	0	0	no
27	55	167	150	1.2	0	5.9	80	10.6	14	0	0	no
28	65	103	10	0.7	0	3.8	10	11	2	0	0	no
29	46	103	80	1.3	0	3.9	30	21	18	0	0	no
30	45	99	10	0.9	0	7.3	80	14	4	0	0	no
31	82	48	10	0.9	0	0.5	10	8	7	0	0	yes
32	36	115	300	1	0	3.4	150	26	3	0	0	no
33	60	108	10	1.1	0	6.3	80	17	9	0	0	no
34	46	109	10	0.9	0	2.4	10	15	14	0	0	no

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	HDL Cholesterol	Triglyceride	Microalbumin	Creatine	Urineprotein	Uric Acid	Albuminuria	BUN	ระยะเวลาที่ป่วยเป็นโรคเบาหวานชนิดที่ 2	Alcohol	Smoking	DKA
35	62	68	80	0.7	0	4.4	30	8	3	0	0	no
36	31	127		0.8	0	3.1	10	10	6	0	0	no
37	37	117	300	1.1	1	4.6	300	17	13	0	0	yes
38	55	199	30	1.1	0	5.3	150	15	31	0	1	no
39	45	92	300	1.1	0	5.1	300	7	13	0	0	yes
40	46	96	150	1.4	0	7	10	15	14	0	0	yes
41	42	118	10	0.9	0	2.5	80	13	5	0	0	no
42	82	89	10	1.4	0	5.7	10	23	12	0	0	no
43	43	124	30	0.9	0	4.7	30	17	16	0	0	yes
44	43	78	30	0.6	0	3.5	10	8	5	0	0	no
45	36	129	10	0.8	0	4.4	10	13	10	0	0	no
46	22	125	10	1.4	1	5.1	300	13	8	0	0	no
47	56	142	300	1.1	0	4.8	80	11	4	0	0	yes
48	23	73	30	1.1	0	4.4	80	20	4	0	1	no
49	19	113	10	1.2	0	5.5	30	19	6	0	0	no
50	59	116	30	1.2	0	7.3	10	16	4	0	0	yes
51	33	134		1.3	0	6.2	30	19	9	0	0	no
52	57	142	10	0.8	0	3	10	14	7	0	0	yes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	HDL Cholesterol	Triglyceride	Microalbumin	Creatine	Urineprotein	Uric Acid	Albuminuria	BUN	ระยะเวลาที่ป่วยเป็นโรคเบาหวานชนิดที่ 2	Alcohol	Smoking	DKA
53	59	129	10	1.6	0	5	10	18	13	0	0	no
54	50	94	80	0.9	0	2.9	10	12	11	0	0	yes
55	47	120	10	1.1	0	2.6	80	11	1	0	0	no
56	34	128	80	1.3	0	2.6	30	19	20	0	0	no
57	43	101	10	1.2	1	3.3	300	17	34	0	0	yes
58	48	131	10	1.2	0	5.2	30	14	3	0	0	yes
59	44	79	150	0.9	0	2.6	150	12	11	0	0	no
60	33	48	10	1	0	4.8	30	12	14	0	0	yes
61	32	109	10	1.2	0	3.5	10	14	4	0	0	no
62	54	86	300	0.8	0	2.7	10	8	10	0	0	yes
63	45	110	10	1.2	0	2.8	10	14	4	0	0	no
64	35	79	30	1.2	0	5.8	80	12	8	0	0	no
65	62	135	10	1.2	0	4	30	10	9	0	0	no
66	54	128	30	1.4	0	5.3	10	21	3	0	0	no
67	56	68	150	1.1	1	4.3	1	12	12	0	1	no
68	34	107	150	1	0	2.1	150	12.4	11	0	0	no
69	54	75		0.8	0	3.2	80	10	2	1	0	no
70	33	115	30	1.1	0	3.5	10	13	11	0	0	no

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	HDL Cholesterol	Triglyceride	Microalbumin	Creatine	Urineprotein	Uric Acid	Albuminuria	BUN	ระยะเวลาที่ป่วยเป็นโรคเบาหวานชนิดที่ 2	Alcohol	Smoking	DKA
71	36	64	80	1	1	3.9	300	9	3	1	1	yes
72	57	149	10	0.9	0	4.9	10	11	3	0	0	no
73	29	98	150	0.9	1	4.6	300	17	9	0	0	no
74	63	61	10	0.7	0	5.3	10	8	3	0	0	no
75	37	123	1000	1.3	3	5.8	1	28	4	0	0	yes
76	33	99	1000	1.1	3	5.5	5	23	20	0	0	yes
77	38	127	30	0.9	0	2.5	10	17	17	0	0	yes
78	26	93	10	0.9	0	4	30	11	22	0	0	no
79	55	130	30	0.9	0	2.5	30	11	6	0	0	no
80	33	53	300	2.1	1	4.8	300	32	16	0	0	no
81	43	105	10	1	0	4.6	10	14	3	0	0	yes
82	44	167	30	1.2	0	6.6	10	17	2	0	0	no
83	64	109	300	1.1	1	4.8	300	15	8	0	1	no
84	61	147	10	0.9	0	4.4	80	22	4	0	0	no
85	61	120	80	1.3	0	4	150	18	12	0	0	yes
86	55	111	10	1.1	0	4.9	10	12	3	0	0	yes
87	89	132	150	1.3	0	5.9	30	18	12	0	0	no
88	33	134	80	1.1	3	3.5	1	12	3	0	0	yes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	HDL Cholesterol	Triglyceride	Microalbumin	Creatine	Urineprotein	Uric Acid	Albuminuria	BUN	ระยะเวลาที่ป่วยเป็นโรคเบาหวานชนิดที่ 2	Alcohol	Smoking	DKA
89	53	81	10	0.8	0	3.1	150	12	13	0	0	yes
90	72	84	30	1.2	0	5.1	30	12	2	0	0	no
91	47	143	30	0.8	0	2.9	30	10	2	0	0	no
92	34	67	30	1.1	0	6.3	30	16	5	0	0	no
93	41	91	10	1.1	0	7.1	10	14	13	0	0	yes
94	31	64	30	1	0	4.8	150	13	3	0	0	yes
95	41	69	30	1.3	0	4.7	10	12	3	0	0	no
96	28	69	10	0.9	0	4.9	30	17	22	0	0	yes
97	72	104	10	0.8	0	2.6	10	9	4	0	0	no
98	56	104	30	1.2	0	6.3	30	12	8	0	0	no
99	53	97	10	1.2	0	5.2	150	20	33	0	0	no
100	32	123	10	0.9	0	5.8	10	14	7	0	0	yes
101	37	109	10	1.1	0	3.9	30	12	13	0	0	no
102	45	113	30	1	0	1.6	30	12	5	0	0	no
103	28	88	80	1	0	5.6	10	15.1	1	1	1	no
104	41	94	150	1.3	0	3.9	150	19	10	0	0	no
105	31	92	10	1.3	0	7.9	10	20	4	0	0	no
106	64	93	10	0.8	0	3.1	30	11	8	0	0	no

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ID	HDL Cholesterol	Triglyceride	Microalbumin	Creatine	Urineprotein	Uric Acid	Albuminuria	BUN	ระยะเวลาที่ป่วยเป็นโรคเบาหวานชนิดที่ 2	Alcohol	Smoking	DKA
107	63	115	10	1	0	3.9	10	11	4	0	0	no
108	49	89	10	1	0	4.9	80	15	17	0	0	no
109	33	132	300	1.2	0	8.3	80	10	6	0	0	yes
110	39	139	300	1.2	0	5.9	150	20	9	0	0	no
111	29	115	10	1	0	3.2	10	10	9	0	0	yes
112	66	139	10	0.9	0	3.9	10	12	2	0	0	yes
113	42	110	10	0.8	0	3	10	15	22	0	0	no
114	38	120	300	1.5	2	5	1	13	14	0	0	no
115	46	126		1.2	3	5.1	1	28	4	0	0	yes
116	53	69	80	1	0	5.3	80	16	10	0	0	no
117	43	96	10	0.8	0	4.2	10	14	5	0	0	yes
118	53	83	10	0.9	0	4.8	10	12	6	0	0	no
119	41	80	10	1.3	0	5.4	10	13	19	0	0	no
120	29	74	1000	2.2	2	6.7	300	15	5	0	0	yes
121	49	74	10	1	0	4.6	300	27	2	0	0	yes
122	36	82	10	0.8	0	3.8	30	11	2	0	0	no
123	37	125	300	1.6	1	5.4	300	18.3	7	0	0	no
124	45	86	150	0.9	2	4.1	1	8	15	0	0	no

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ภาคผนวก ข
ผลการวิเคราะห์ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์จำแนกประเภท

วิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติ SPSS เวอรชั่น 20 ซึ่งได้ผลการวิเคราะห์ดังนี้

1.1 การตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์จำแนกประเภท

1.1.1 การตรวจสอบว่าตัวแปรอิสระของแต่ละกลุ่มมีการแจกแจงปกติหลายตัวแปรหรือไม่ โดยใช้ตัวสถิติทดสอบ Kolmogorov-Smirnov ปรากฏผลการวิเคราะห์ดังตารางที่ ข-1

ตารางที่ ข-1 ผลการทดสอบการแจกแจงปรกติของตัวแปรอิสระ

Tests of Normality

ภาวะแทรกซ้อนทางไต		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
อายุ	ไม่เกิด	.047	200	.200*	.991	200	.292
	เกิด	.073	120	.168	.985	120	.203
น้ำหนัก	ไม่เกิด	.093	200	.000	.909	200	.000
	เกิด	.103	120	.003	.935	120	.000
ส่วนสูง	ไม่เกิด	.071	200	.017	.987	200	.058
	เกิด	.098	120	.007	.984	120	.158
ดัชนีมวลกาย	ไม่เกิด	.107	200	.000	.855	200	.000
	เกิด	.072	120	.191	.928	120	.000
รอเอว	ไม่เกิด	.155	200	.000	.762	200	.000
	เกิด	.106	120	.002	.983	120	.140
ความดันโลหิตซีสโตลิค	ไม่เกิด	.153	200	.000	.941	200	.000
	เกิด	.140	120	.000	.855	120	.000
ความดันโลหิตไดแอสโตลิค	ไม่เกิด	.090	200	.000	.976	200	.002
	เกิด	.105	120	.002	.882	120	.000
Creatinine	ไม่เกิด	.163	200	.000	.953	200	.000
	เกิด	.218	120	.000	.543	120	.000
HbA1c	ไม่เกิด	.140	200	.000	.927	200	.000
	เกิด	.161	120	.000	.910	120	.000
Triglyceride	ไม่เกิด	.182	200	.000	.770	200	.000
	เกิด	.138	120	.000	.830	120	.000
LDLCholesterol	ไม่เกิด	.055	200	.200*	.993	200	.429
	เกิด	.077	120	.074	.973	120	.016
HDLCholesterol	ไม่เกิด	.075	200	.008	.979	200	.004
	เกิด	.102	120	.004	.974	120	.019
TotalCholesterol	ไม่เกิด	.050	200	.200*	.975	200	.001
	เกิด	.088	120	.023	.979	120	.052
Urineprotein	ไม่เกิด	.514	200	.000	.380	200	.000
	เกิด	.384	120	.000	.672	120	.000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tests of Normality

ภาวะแทรกซ้อนทางไต		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
UricAcid	ไม่เกิด	.088	200	.001	.921	200	.000
	เกิด	.063	120	.200*	.967	120	.005
Albuminuria	ไม่เกิด	.298	200	.000	.700	200	.000
	เกิด	.271	120	.000	.748	120	.000
FBS	ไม่เกิด	.120	200	.000	.913	200	.000
	เกิด	.194	120	.000	.829	120	.000
BUN	ไม่เกิด	.105	200	.000	.949	200	.000
	เกิด	.118	120	.000	.927	120	.000
ระยะเวลาที่ป่วยโรคเบาหวาน ชนิดที่ 2	ไม่เกิด	.177	200	.000	.885	200	.000
	เกิด	.103	120	.003	.934	120	.000
Sex_Male	ไม่เกิด	.392	200	.000	.622	200	.000
	เกิด	.400	120	.000	.616	120	.000
Alcohol_Yes	ไม่เกิด	.540	200	.000	.177	200	.000
	เกิด	.535	120	.000	.108	120	.000
Smoking_Yes	ไม่เกิด	.541	200	.000	.209	200	.000
	เกิด	.528	120	.000	.065	120	.000
DR_Yes	ไม่เกิด	.537	200	.000	.121	200	.000
	เกิด	.540	120	.000	.248	120	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

1.1.2 การตรวจสอบความเท่ากันของเมตริกซ์ความแปรปรวนร่วม โดยใช้ตัวสถิติทดสอบ Box's M ปรากฏผลการวิเคราะห์ดังตารางที่ ข-2

ตารางที่ ข-2 การตรวจสอบความเท่ากันของเมตริกซ์ความแปรปรวนร่วม โดยใช้ตัวสถิติทดสอบ Box's M

Test Results

Box's M		926.439
F	Approx.	3.082
	df1	276
	df2	196457.109
	Sig.	.000

Tests null hypothesis of equal population covariance matrices.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 ผลการวิเคราะห์จำแนกประเภท

จากการวิเคราะห์จำแนกประเภท โดยใช้วิธีของ Fisher's ค่าสัมประสิทธิ์จำแนกประเภทของตัวแปรอิสระ (X_i) ปรากฏผลการวิเคราะห์ดังตารางที่ ข-3

ตารางที่ ข-3 ค่าสัมประสิทธิ์จำแจกประเภท โดยใช้วิธีของ Fisher's

Classification Function Coefficients

	ภาวะแทรกซ้อนทางไต	
	ไม่เกิด	เกิด
เพศ	56.364	58.916
อายุ	2.067	2.261
น้ำหนัก	-112.086	-111.956
ส่วนสูง	96.141	96.011
ดัชนีมวลกาย	284.006	283.508
รอเอว	-1.281	-1.291
ความดันโลหิตซิสโตลิค	1.557	1.561
ความดันโลหิตไดแอสโตลิค	.001	-.008
Creatinine	1.391	5.424
HbA1c	-2.544	-2.806
Triglyceride	-.086	-.091
LDLCholesterol	.248	.257
HDLCholesterol	-.297	-.320
TotalCholesterol	-.129	-.139
Urineprotein	11.142	11.593
UricAcid	-3.118	-1.918
Albuminuria	-.060	-.053
FBS	.086	.090
BUN	2.161	2.211
ระยะเวลาที่ป่วยโรคเบาหวาน ชนิดที่ 2	-1.202	-1.180
การดื่มสุรา	-8.561	-7.040
การสูบบุหรี่	21.289	20.072
ภาวะแทรกซ้อนทางตา	26.555	28.337
(Constant)	-7741.442	-7739.378

Fisher's linear discriminant functions

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ผลการทดสอบสมมติฐานของการวิเคราะห์จำแนกประเภท

ในการวิเคราะห์จำแนกประเภท ซึ่งในงานวิจัยนี้ใช้วิธี Hotelling's T^2 ในการทดสอบความแตกต่างของกลุ่มข้อมูล ปรากฏผลการวิเคราะห์ดังตารางที่ ข-4

ตารางที่ ข-4 ผลการทดสอบสมมติฐานของการวิเคราะห์จำแนกประเภท โดยใช้วิธี Hotelling's T^2

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	1.000	187914.147[b]	23.000	296.000	.000
	Wilks' Lambda	.000	187914.147[b]	23.000	296.000	.000
	Hotelling's Trace	14601.437	187914.147[b]	23.000	296.000	.000
	Roy's Largest Root	14601.437	187914.147[b]	23.000	296.000	.000
DN	Pillai's Trace	.643	23.203[b]	23.000	296.000	.000
	Wilks' Lambda	.357	23.203[b]	23.000	296.000	.000
	Hotelling's Trace	1.803	23.203[b]	23.000	296.000	.000
	Roy's Largest Root	1.803	23.203[b]	23.000	296.000	.000

a. Design: Intercept + DN

b. Exact statistic

1.4 ผลการประเมินความเหมาะสมของสมการการวิเคราะห์จำแนกประเภท

ในงานวิจัยนี้ ได้ประเมินความเหมาะสมของสมการการวิเคราะห์จำแนกประเภทโดยเลือกใช้ค่าความถูกต้อง (Accuracy) และค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error : RMSE) ปรากฏผลการวิเคราะห์ดังตารางที่ ข-5

ตารางที่ ข-5 ผลการประเมินความเหมาะสมของสมการการวิเคราะห์จำแนกประเภท

Classification Results[a]

			Predicted Group Membership		
		ภาวะแทรกซ้อนทางไต	ไม่เกิด	เกิด	Total
Original	Count	ไม่เกิด	195	5	200
		เกิด	5	115	120
	%	ไม่เกิด	97.5	2.5	100.0
		เกิด	4.2	95.8	100.0

a. 96.9% of original grouped cases correctly classified.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

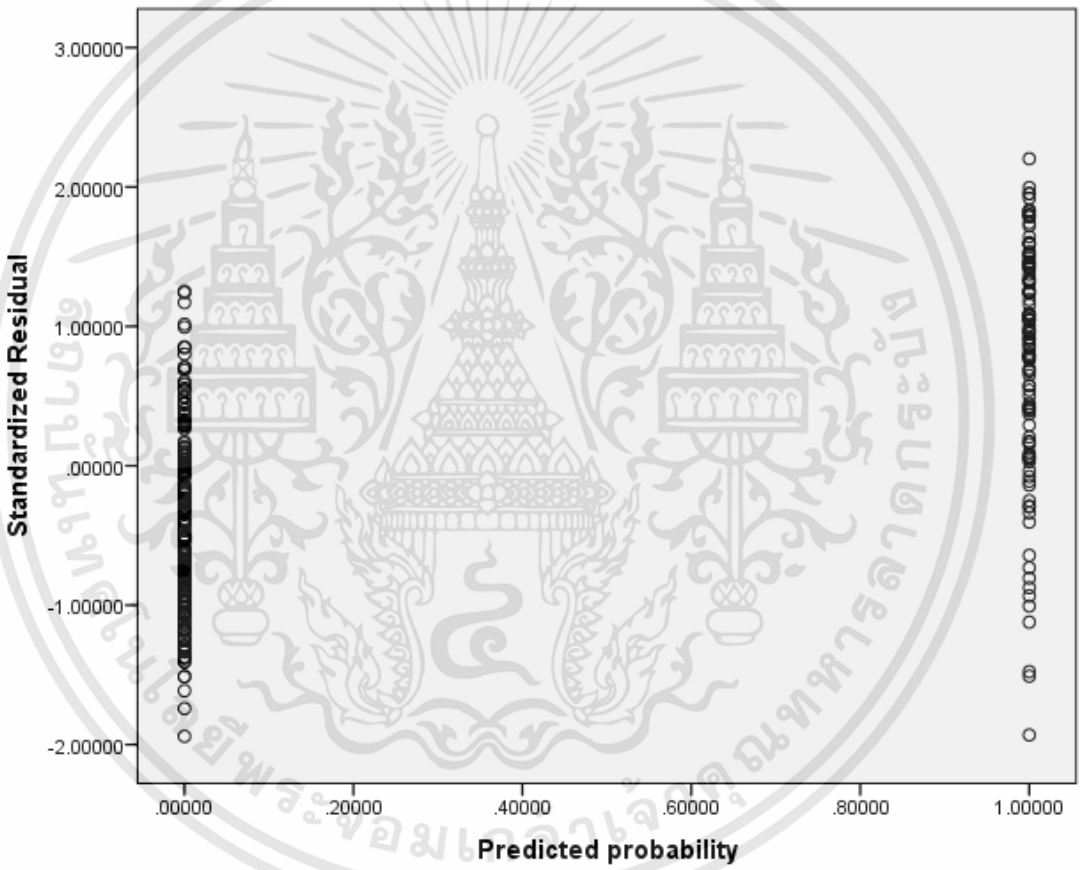
2. ผลการวิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์การถดถอยลอจิสติก

วิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติ SPSS เวอรชั่น 20 ซึ่งได้ผลการวิเคราะห์ดังนี้

2.1 ตรวจสอบเงื่อนไขเบื้องต้นของการวิเคราะห์การถดถอยลอจิสติก

2.1.1 ตัวแปรอิสระอาจจะเป็นข้อมูลที่มีค่าได้ 2 ค่า (Dichotomous Variable) คือ 0 กับ 1 เท่านั้น หรือเป็นสเกลอันตรภาค (Interval Scale) หรือสเกลอัตราส่วน (Ratio Scale) ซึ่งในงานวิจัยนี้ข้อมูลที่ใช้ในการวิเคราะห์เป็นไปตามเงื่อนไขเบื้องต้นดังกล่าว

2.1.2 ค่าคาดหวังของค่าความคาดเคลื่อนเป็นศูนย์ หรือ E(e) = 0 ปรากฏผลการวิเคราะห์ดังรูปที่ ข-1

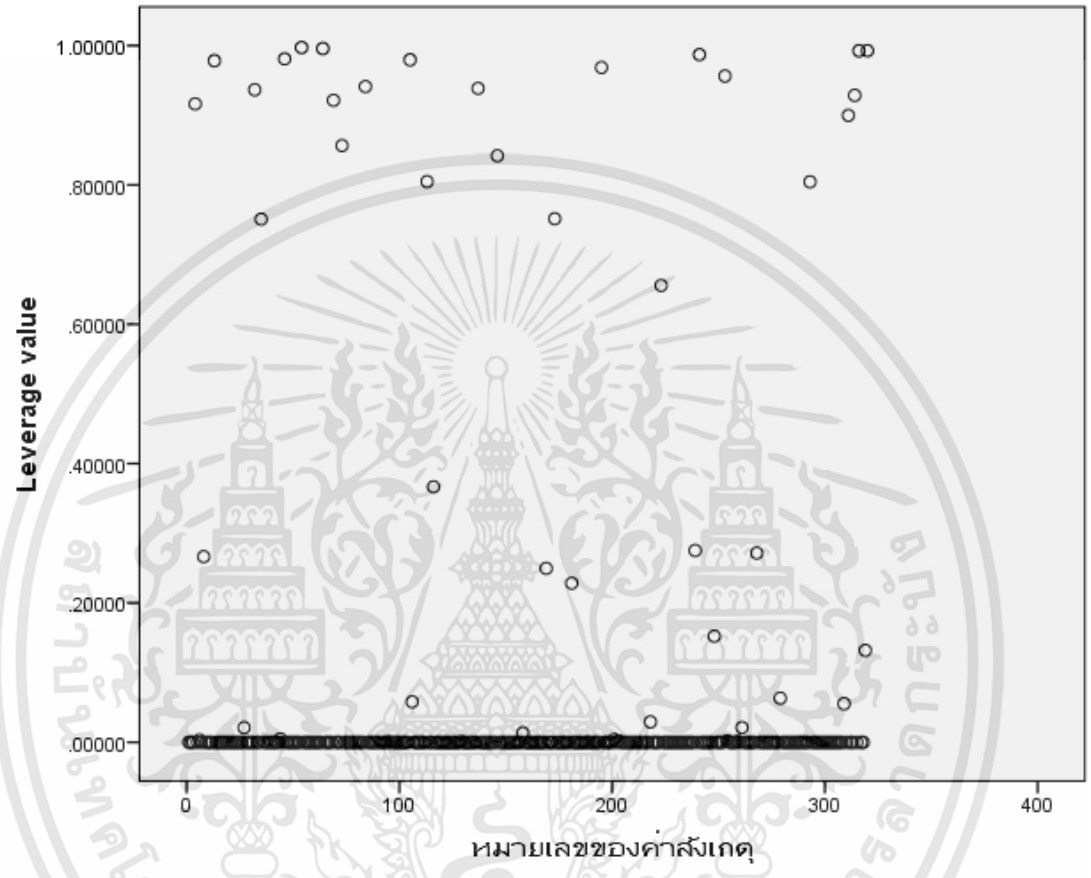


รูปที่ ข-1 การกระจายของค่าส่วนเหลือมาตรฐาน (Standardized Residual)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

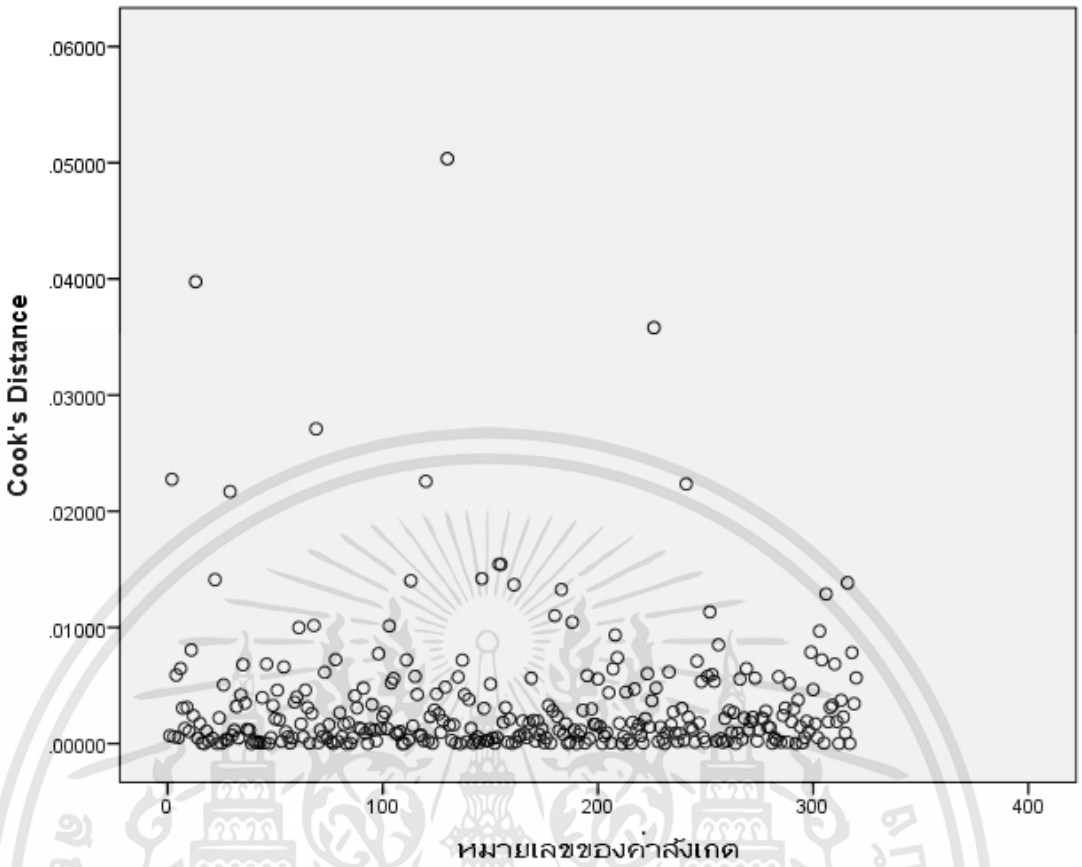
2.1.3 e_i และ e_j เป็นอิสระกัน ; I , j = 1, 2, 3,..., n (เมื่อ n คือ จำนวนข้อมูลทั้งหมด) ปรากฏผลการวิเคราะห์ดังรูปที่ ข-2 และ ข-3

2.1.4 e_i และ x_i เป็นอิสระกัน ; I = 1, 2, 3,..., n (เมื่อ n คือ จำนวนข้อมูลทั้งหมด) ปรากฏผลการวิเคราะห์ดังรูปที่ ข-2 และ ข-3



รูปที่ ข-2 การกระจายของค่า Leverage

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข-3 การกระจายของค่า Cook’s Distance

2.1.5 การตรวจสอบตัวแปรอิสระไม่มีความสัมพันธ์กัน

โดยพิจารณาจากค่า Tolerance และ Variance Inflation Factor (VIF) ของตัวแปรอิสระทั้ง 23 ตัว โดยมีเกณฑ์ในการพิจารณา คือ ถ้า Tolerance มีค่าเข้าใกล้ 0 หรือ VIF มีค่ามากกว่า 10 แสดงว่าตัวแปรอิสระมีความสัมพันธ์กัน ปรากฏผลการวิเคราะห์ดังตารางที่ ข-6 และ ข-7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข-6 ค่า Tolerance และ VIF ของตัวแปรอิสระทั้ง 23 ตัว

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	เพศ	.483	2.070
	อายุ	.720	1.389
	น้ำหนัก	.007	153.285
	ส่วนสูง	.027	36.856
	ดัชนีมวลกาย	.008	121.109
	รอบเอว	.400	2.502
	ความดันโลหิตซิสโตลิค	.352	2.841
	ความดันโลหิตไดแอสโตลิค	.354	2.829
	Creatinine	.471	2.122
	HbA1c	.564	1.774
	Triglyceride	.355	2.818
	LDLCholesterol	.075	13.263
	HDLCholesterol	.432	2.317
	TotalCholesterol	.061	16.488
	Urineprotein	.699	1.430
	UricAcid	.626	1.596
	Albuminuria	.894	1.119
	FBS	.537	1.862
	BUN	.487	2.052
	ระยะเวลาที่ป่วยโรคเบาหวาน ชนิดที่ 2	.761	1.314
	การดื่มสุรา	.699	1.430
	การสูบบุหรี่	.696	1.437
	ภาวะแทรกซ้อนทางตา	.875	1.143

a. Dependent Variable: ภาวะแทรกซ้อนทางไต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข-8 ค่าความสัมพันธ์ของตัวแปรอิสระทั้ง 5 ตัว โดยใช้วิธีของเพียร์สัน (Pearson Correlation)

Correlations

		น้ำหนัก	ส่วนสูง	ดัชนีมวลกาย	LDLCholesterol	TotalCholesterol
น้ำหนัก	Pearson Correlation	1	.473**	.874**	.089	.057
	Sig. (2-tailed)		.000	.000	.113	.306
	Sum of Squares and Cross-products	55625.374	15513.411	17049.359	10599.533	10025.956
	Covariance	174.374	48.631	53.446	33.227	31.429
	N	320	320	320	320	320
ส่วนสูง	Pearson Correlation	.473**	1	-.006	-.006	.008
	Sig. (2-tailed)	.000		.916	.916	.891
	Sum of Squares and Cross-products	15513.411	19352.347	-67.826	-417.488	794.012
	Covariance	48.631	60.666	-.213	-1.309	2.489
	N	320	320	320	320	320
ดัชนีมวลกาย	Pearson Correlation	.874**	-.006	1	.105	.065
	Sig. (2-tailed)	.000	.916		.061	.247
	Sum of Squares and Cross-products	17049.359	-67.826	6835.264	4397.611	3977.383
	Covariance	53.446	-.213	21.427	13.786	12.468
	N	320	320	320	320	320
LDLCholesterol	Pearson Correlation	.089	-.006	.105	1	.929**
	Sig. (2-tailed)	.113	.916	.061		.000
	Sum of Squares and Cross-products	10599.533	-417.488	4397.611	256660.829	348685.672
	Covariance	33.227	-1.309	13.786	804.579	1093.059
	N	320	320	320	320	320
TotalCholesterol	Pearson Correlation	.057	.008	.065	.929**	1
	Sig. (2-tailed)	.306	.891	.247	.000	
	Sum of Squares and Cross-products	10025.956	794.012	3977.383	348685.672	548789.550
	Covariance	31.429	2.489	12.468	1093.059	1720.343
	N	320	320	320	320	320

**. Correlation is significant at the 0.01 level (2-tailed).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข-9 ตรวจสอบความสัมพันธ์ของตัวแปรอิสระทั้ง 5 ตัว กับตัวแปรตามโดยใช้การวิเคราะห์ความแปรปรวนแบบทางเดียว (One Way ANOVA)

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
น้ำหนัก	Between Groups	2829.773	1	2829.773	17.044	.000
	Within Groups	52795.601	318	166.024		
	Total	55625.374	319			
ส่วนสูง	Between Groups	137.025	1	137.025	2.268	.133
	Within Groups	19215.322	318	60.426		
	Total	19352.347	319			
ดัชนีมวลกาย	Between Groups	335.382	1	335.382	16.408	.000
	Within Groups	6499.882	318	20.440		
	Total	6835.264	319			
LDLCholesterol	Between Groups	475.147	1	475.147	.590	.443
	Within Groups	256185.682	318	805.615		
	Total	256660.829	319			
TotalCholesterol	Between Groups	2802.963	1	2802.963	1.633	.202
	Within Groups	545986.587	318	1716.939		
	Total	548789.550	319			

ตารางที่ ข-10 ค่าความสัมพันธ์ของตัวแปรอิสระ (X_3) กับตัวแปรตาม (Y)

Directional Measures

			Value
Nominal by Interval	Eta	น้ำหนัก Dependent	.226
		ภาวะแทรกซ้อนทางไต Dependent	.460

ตารางที่ ข-11 ค่าความสัมพันธ์ของตัวแปรอิสระ (X_4) กับตัวแปรตาม (Y)

Directional Measures

			Value
Nominal by Interval	Eta	ส่วนสูง Dependent	.084
		ภาวะแทรกซ้อนทางไต Dependent	.340

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข-12 ค่าความสัมพันธ์ของตัวแปรอิสระ (X_5) กับตัวแปรตาม (Y)

Directional Measures

			Value
Nominal by Interval	Eta	ดัชนีมวลกาย Dependent	.222
		ภาวะแทรกซ้อนทางไต Dependent	.872

ตารางที่ ข-13 ค่าความสัมพันธ์ของตัวแปรอิสระ (X_{12}) กับตัวแปรตาม (Y)

Directional Measures

			Value
Nominal by Interval	Eta	LDLCholesterol Dependent	.043
		ภาวะแทรกซ้อนทางไต Dependent	.590

ตารางที่ ข-14 ค่าความสัมพันธ์ของตัวแปรอิสระ (X_5) กับตัวแปรตาม (Y)

Directional Measures

			Value
Nominal by Interval	Eta	TotalCholesterol Dependent	.071
		ภาวะแทรกซ้อนทางไต Dependent	.662

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข-15 ค่า Tolerance และ VIF ของตัวแปรอิสระทั้ง 20 ตัว

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	เพศ	.756	1.323
	อายุ	.784	1.275
	ดัชนีมวลกาย	.417	2.396
	รอเอว	.453	2.209
	ความดันโลหิตซิสโตลิค	.355	2.819
	ความดันโลหิตไดแอสโตลิค	.364	2.747
	Creatinine	.475	2.107
	HbA1c	.575	1.740
	Triglyceride	.665	1.503
	LDLCholesterol	.865	1.156
	HDLCholesterol	.728	1.373
	Urineprotein	.746	1.341
	UricAcid	.629	1.591
	Albuminuria	.901	1.110
	FBS	.538	1.860
	BUN	.501	1.994
	ระยะเวลาที่ป่วยโรคเบาหวาน ชนิดที่ 2	.766	1.305
	การดื่มสุรา	.711	1.406
	การสูบบุหรี่	.699	1.430
	ภาวะแทรกซ้อนทางตา	.887	1.128

a. Dependent Variable: ภาวะแทรกซ้อนทางไต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 ผลการวิเคราะห์การถดถอยลอจิสติก

จากการวิเคราะห์ข้อมูล โดยใช้การวิเคราะห์การถดถอยลอจิสติก และทำการประมาณค่าพารามิเตอร์โดยใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) ปรากฏผลการวิเคราะห์ดังตารางที่ ข-16

ตารางที่ ข-16 แสดงค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระทั้ง 20 ตัว

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Sex(1)	-66.549	3354.111	.000	1	.984	.000
	Age	3.126	198.709	.000	1	.987	22.778
	BMI	-.899	394.049	.000	1	.998	.407
	Waistline	.255	67.663	.000	1	.997	1.291
	Lbloodpress	-.017	179.110	.000	1	1.000	.983
	Hbloodpress	-.337	117.337	.000	1	.998	.714
	Creatinine	230.767	10399.597	.000	1	.982	1.663E+100
	HbA1c	-3.569	1143.595	.000	1	.998	.028
	Triglyceride	-.083	17.495	.000	1	.996	.920
	LDLCholesterol	-.148	73.000	.000	1	.998	.862
	HDLCholesterol	-.290	110.766	.000	1	.998	.748
	Urineprotein	5.987	1182.654	.000	1	.996	398.368
	UricAcid	10.102	1321.177	.000	1	.994	24393.158
	Albuminuria	.040	15.412	.000	1	.998	1.040
	FBS	.075	30.515	.000	1	.998	1.078
	BUN	.198	213.915	.000	1	.999	1.218
	Year_DM	1.120	212.466	.000	1	.996	3.066
	Alcohol(1)	-42.014	25620.633	.000	1	.999	.000
	Smoking(1)	.526	16047.825	.000	1	1.000	1.692
	DR(1)	-11.282	15613.997	.000	1	.999	.000
	Constant	-338.578	32438.119	.000	1	.992	.000

a. Variable(s) entered on step 1: Sex, Age, BMI, Waistline, Lbloodpress, Hbloodpress, Creatinine, HbA1c, Triglyceride, LDLCholesterol, HDLCholesterol, Urineprotein, UricAcid, Albuminuria, FBS, BUN, Year_DM, Alcohol, Smoking, DR.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.1 การทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระทีละตัวโดยใช้วิธีการทดสอบวอดล์ ปรากฏผลการวิเคราะห์ดังตารางที่ ข-17

ตารางที่ ข-17 การทดสอบโดยใช้วิธีของวอดล์ (Wald)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Sex(1)	-66.549	3354.111	.000	1	.984	.000
	Age	3.126	198.709	.000	1	.987	22.778
	BMI	-.899	394.049	.000	1	.998	.407
	Waistline	.255	67.663	.000	1	.997	1.291
	Lbloodpress	-.017	179.110	.000	1	1.000	.983
	Hbloodpress	-.337	117.337	.000	1	.998	.714
	Creatinine	230.767	10399.597	.000	1	.982	1.663E+100
	HbA1c	-3.569	1143.595	.000	1	.998	.028
	Triglyceride	-.083	17.495	.000	1	.996	.920
	LDLCholesterol	-.148	73.000	.000	1	.998	.862
	HDLCholesterol	-.290	110.766	.000	1	.998	.748
	Urineprotein	5.987	1182.654	.000	1	.996	398.368
	UricAcid	10.102	1321.177	.000	1	.994	24393.158
	Albuminuria	.040	15.412	.000	1	.998	1.040
	FBS	.075	30.515	.000	1	.998	1.078
	BUN	.198	213.915	.000	1	.999	1.218
	Year_DM	1.120	212.466	.000	1	.996	3.066
	Alcohol(1)	-42.014	25620.633	.000	1	.999	.000
	Smoking(1)	.526	16047.825	.000	1	1.000	1.692
	DR(1)	-11.282	15613.997	.000	1	.999	.000
	Constant	-338.578	32438.119	.000	1	.992	.000

a. Variable(s) entered on step 1: Sex, Age, BMI, Waistline, Lbloodpress, Hbloodpress, Creatinine, HbA1c, Triglyceride, LDLCholesterol, HDLCholesterol, Urineprotein, UricAcid, Albuminuria, FBS, BUN, Year_DM, Alcohol, Smoking, DR.

2.1.2 การทดสอบค่าสัมประสิทธิ์การถดถอยลอจิสติกของตัวแปรอิสระ โดยใช้วิธีอัตราส่วนภาวะน่าจะเป็น ปรากฏผลการวิเคราะห์ดังตารางที่ ข-18

ตารางที่ ข-18 ผลการทดสอบค่าสัมประสิทธิ์การถดถอยลอจิสติก

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	423.400	23	.000
	Block	423.400	23	.000
	Model	423.400	23	.000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3 การตรวจสอบความเหมาะสมของตัวแบบการวิเคราะห์การถดถอยลอจิสติก ในงานวิจัยนี้ผู้วิจัยได้ใช้วิธีการทดสอบของ Hosmer and Lemeshow ที่มีการแบ่งข้อมูลออกเป็น p กลุ่มย่อย ให้ผลการวิเคราะห์ดังตารางที่ ข-19

ตารางที่ ข-19 ผลการตรวจสอบความเหมาะสมของตัวแบบการวิเคราะห์การถดถอยลอจิสติก

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	.000	7	1.000

2.3 ผลการประเมินความเหมาะสมของตัวแบบการวิเคราะห์การถดถอยลอจิสติก

การประเมินความเหมาะสมของตัวแบบการวิเคราะห์การถดถอยลอจิสติกที่ใช้ในการจำแนกกลุ่ม โดยพิจารณาจากตัวแปรอิสระทั้ง 20 ตัวแปร เมื่อกำหนดจุดตัดในการแบ่งกลุ่มเป็น 0.5 โดยใช้ค่าความถูกต้องและค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง ปรากฏผลการวิเคราะห์ดังตารางที่ ข-20

ตารางที่ ข-20 ผลการประเมินความเหมาะสมของตัวแบบการวิเคราะห์การถดถอยลอจิสติก

Classification Table[a]

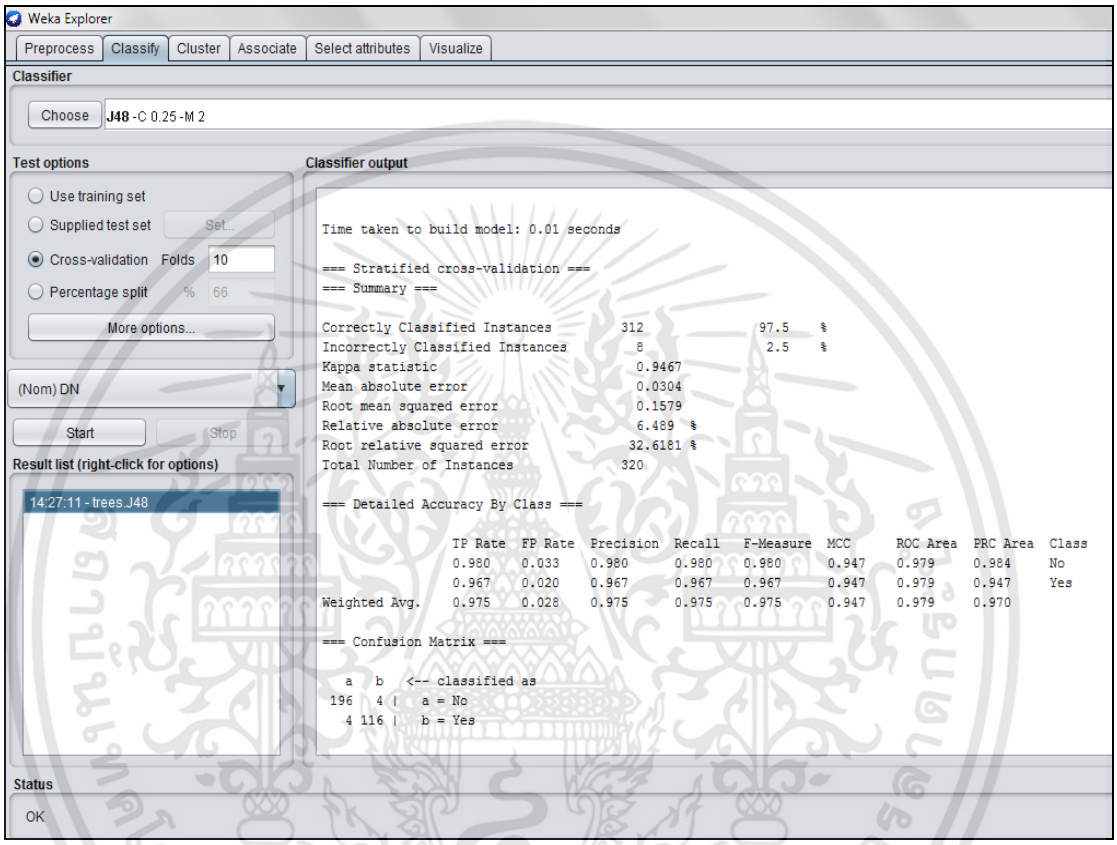
Observed			Predicted		
			ภาวะแทรกซ้อนทางไต		Percentage Correct
			ไม่เกิด	เกิด	
Step 1	ภาวะแทรกซ้อนทางไต	ไม่เกิด	200	0	100.0
		เกิด	0	120	100.0
	Overall Percentage				100.0

a. The cut value is .500

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

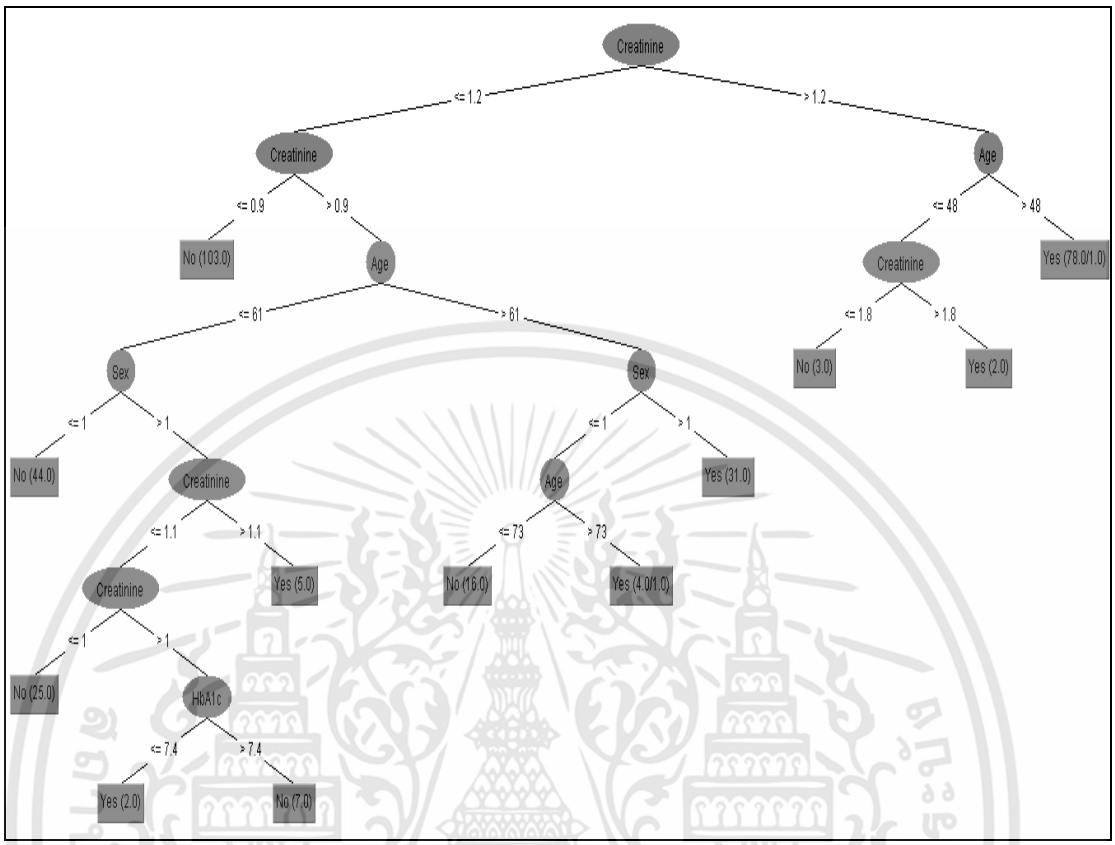
3. ผลการวิเคราะห์ข้อมูลโดยใช้วิธีต้นไม้ตัดสินใจ

วิเคราะห์ข้อมูลโดยใช้โปรแกรม WEKA เวอรชั่น 3.8.1 ซึ่งได้ผลการวิเคราะห์ดังนี้ ปรากฏผลการวิเคราะห์ดังรูปที่ ข-4 และ ข-5



รูปที่ ข-4 ผลการวิเคราะห์ข้อมูลโดยใช้วิธีต้นไม้ตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข-5 ต้นไม้ตัดสินใจเพื่อใช้ในการทำนาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นางสาวทองสา บุตรงาม
วัน เดือน ปีเกิด	16 พฤษภาคม 2531
ที่อยู่ปัจจุบัน	89 หมู่ 10 ตำบลจรเข้มาก อำเภอประโคนชัย จังหวัดบุรีรัมย์ 31140
ประวัติการศึกษา	(2553) วิทยาศาสตรบัณฑิต สาขาวิชาสถิติประยุกต์ เกรดเฉลี่ย 3.38 (มหาวิทยาลัยราชภัฏบุรีรัมย์)
ผลงานทางวิชาการ	การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเกิดภาวะแทรกซ้อนทางไตในผู้ป่วยโรคเบาหวาน ชนิดที่ 2 กรณีศึกษา : โรงพยาบาลแห่งหนึ่งในจังหวัดบุรีรัมย์ (การประชุมวิชาการระดับชาติ ครั้งที่ 11 มหาวิทยาลัยราชภัฏนครปฐม “วิจัยสร้างนวัตกรรม เพื่อพัฒนาท้องถิ่นและสังคมไทย สู่ Disruptive Society” ในวันที่ 11-12 กรกฎาคม 2562)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้