

การคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท  
โดยใช้ฟังก์ชันโพโรบิตเป็นฟังก์ชันเชื่อมโยง

นางสาวจารุตา ฤทธิเดชะ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต  
สาขาวิชาสถิติ ภาควิชาสถิติ  
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2554

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ทั้งหมดที่กองกลางมหาวิทยาลัย  
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)  
are the thesis authors' files submitted through the Graduate School.

THE STRICTLY NON-NESTED MODEL OF BINARY LOGISTIC REGRESSION MODEL  
USING PROBIT FUNCTION AS A LINK FUNCTION

Miss Jaruta Ritdacha

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2011

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพโรบิตเป็นฟังก์ชันเชื่อมโยง
โดย	นางสาวจารุตา ฤทธิเดชะ
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา

---

คณะแพทยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....คณบดีคณะแพทยศาสตร์และการบัญชี  
(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

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

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

..... กรรมการ  
(อาจารย์ ดร.อนุภาพ สมบูรณ์สวัสดิ์)

..... กรรมการภายนอกมหาวิทยาลัย  
(รองศาสตราจารย์ ดร. มนต์ทิพย์ เทียนสุวรรณ)

จารุตา ฤทธิเดช : คัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอย  
โลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง. (THE STRICTLY  
NON-NESTED MODEL OF BINARY LOGISTIC REGRESSION MODEL USING  
PROBIT FUNCTION AS A LINK FUNCTION) อ. ที่ปรึกษาวิทยานิพนธ์หลัก:  
รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา, 85 หน้า.

การศึกษาค้นคว้าครั้งนี้มีวัตถุประสงค์เพื่อคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการ  
ถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง ปัจจัยที่ส่งผลต่อตัวแบบที่  
ถูกคัดเลือกคือ คู่ลำดับของ จำนวนตัวแปรในตัวแบบแรก ( $p_1$ ) และจำนวนตัวแปรในตัวแบบที่สอง  
( $p_2$ ); ( $p_1, p_2$ ); (2, 2), (2, 3), (2, 4), (2, 5), (3, 2), (3, 3), (3, 4), (3, 5), (4, 2), (4, 3), (4, 4), (4, 5),  
(5, 2), (5, 3), (5, 4), (5, 5) ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระเท่ากับ 0.33, 0.66 และ  
0.99 และขนาดตัวอย่าง( $n$ ) คือ 50, 100, 150, 200 และ 250 ซึ่งข้อมูลทั้งหมดนี้ใช้การจำลองโดย  
เทคนิคมอนติคาร์โล ด้วยโปรแกรม R โดยใช้พื้นที่ใต้โค้ง ROC สูงสุดเป็นเกณฑ์ในการคัดเลือก จาก  
การวิจัยสรุปผลได้ดังนี้

ปัจจัยที่เกี่ยวข้องกับการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ ของตัวแบบการถดถอย  
โลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง ที่ทำให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC  
เปลี่ยนแปลงไปในแต่ละตัวแบบ กรณีจำนวนตัวแปรอิสระในตัวแบบแรก( $p_1$ ) และจำนวนตัวแปรอิสระ  
ในตัวแบบที่สอง( $p_2$ ); ( $p_1, p_2$ ) เปลี่ยนแปลงไป ภายใต้ขอบเขตขนาดตัวอย่างและระดับความสัมพันธ์  
ในแต่ละคู่ของตัวแปรอิสระคงที่ พบว่าเมื่อจำนวนตัวแปรอิสระในตัวแบบแรก( $p_1$ ) มีค่ามากกว่าจำนวน  
ตัวแปรอิสระในตัวแบบที่สอง( $p_2$ ) ตัวแบบที่ถูกเลือกคือตัวแบบแรก( $p_1$ ) และเมื่อจำนวนตัวแปรอิสระใน  
ตัวแบบที่สอง( $p_2$ ) มีค่ามากกว่าจำนวนตัวแปรอิสระในตัวแบบแรก( $p_1$ ) ตัวแบบที่ถูกเลือกคือตัวแบบที่  
สอง( $p_2$ ) กรณีระดับความสัมพันธ์ในแต่ละคู่ ของตัวแปรอิสระเปลี่ยนแปลงไป แต่ ขนาดตัวอย่างและ  
จำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่ พบว่าเมื่อระดับความสัมพันธ์ในแต่ละคู่ ของตัวแปรอิสระมี  
ระดับต่ำ ตัวแบบจะมีค่าความน่าเชื่อถือสูง แต่เมื่อระดับความสัมพันธ์ในแต่ละคู่ ของตัวแปรอิสระอยู่  
ในระดับปานกลางและสูง ตัวแบบจะมีค่าความน่าเชื่อถือลดลง กรณีขนาดตัวอย่างเปลี่ยนแปลงไป แต่  
จำนวนตัวแปรอิสระของแต่ละตัวแบบและ ความสัมพันธ์ในแต่ละคู่ ของตัวแปรอิสระคงที่ พบว่าเมื่อ  
ขนาดตัวอย่างมีค่าเพิ่มมากขึ้น ตัวแบบจะมีค่าความน่าเชื่อถือสูงขึ้น

ภาควิชา ..... สถิติ ..... ลายมือชื่อ.....  
สาขาวิชา..... สถิติ..... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....  
ปีการศึกษา.....2554.....

## 5281774026: MAJOR STATISTICS

KEYWORDS : BINARY LOGISTIC REGRESSION / STRICTLY NON-NESTED MODEL /  
ROC CURVE / PROBIT FUNCTION

JARUTA RITDACHA : THE STRICTLY NON-NESTED MODEL OF BINARY  
LOGISTIC REGRESSION MODEL USING PROBIT FUNCTION AS A LINK  
FUNCTION. ADVISOR : ASSOC.PROF.SUPOL DURONGKAWATANA, Ph.D.,  
85 pp.

The purpose of this study is to select the strictly non-nested model of binary logistic regression model using probit function as a link function. The factors affecting selected model are ordered pairs of number independent variables in the first model ( $p_1$ ) and the second model ( $p_2$ ); ( $p_1, p_2$ ); (5, 2), (5, 3), (5, 4), (5, 5), (2, 5), (3, 5), (4, 5), (4, 2), (4, 3), (4, 4), (2, 4), (3, 4), (3, 2), (3, 3), (2, 3), (2, 2), the degree of pair-wise correlation independent variables of the level low medium and high and the sample size ( $n$ ); 50, 100, 150, 200 and 250 The data in all situations are generated using Monte Carlo technique through R-program. The selection criterion is the maximum of the area under ROC curve. The results can be summarized as follow:

Factors involved in selecting the strictly non-nested model of binary logistic regression model using probit function as a link function. Area under the ROC curve of the average change for each model. As ordered pairs of number independent variables in the first model ( $p_1$ ) and the second model ( $p_2$ ); ( $p_1, p_2$ ) changed but the other factors are kept constant, with the number independent variables in the first model ( $p_1$ ) is greater than the second model ( $p_2$ ) the model chosen is the first model ( $p_1$ ) and the number independent variables in the second model ( $p_2$ ) is greater than the first model ( $p_1$ ) the model chosen is the first model ( $p_2$ ). As the degree of pair-wise correlation independent variables of the values changed but the other factors are kept constant, that is, if the degree of pair-wise correlation is at low level, the reliability of the model increases. But if degree of pair-wise correlation is at medium or higher, the reliability of the model decreases. As the sample size changed but the other factors are kept constant, that is, if the size of sample increases, the reliability of the model increases.

Department : .....Statistics..... Student's Signature .....

Field of Study : .....Statistics..... Advisor's Signature .....

Academic Year : .....2011.....

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งจากอาจารย์ที่ปรึกษา วิทยานิพนธ์ รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา ที่กรุณาให้คำแนะนำปรึกษา ตลอดจนช่วยเหลือตรวจสอบ แก้ไขข้อบกพร่องต่างๆ จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณและสำนึกในพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. ชีระพร วีระถาวร ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร. อนุภาพ สมบูรณ์สวัสดิ์ กรรมการสอบวิทยานิพนธ์ และ รองศาสตราจารย์ ดร. มนต์ทิพย์ เทียนสุวรรณ กรรมการภายนอกมหาวิทยาลัย ที่กรุณาตรวจแก้ วิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น และขอกราบขอบพระคุณคณาจารย์ทุกท่านที่ให้โอกาสทางการศึกษา และประสิทธิประสาทความรู้ให้แก่ผู้วิจัยกระทั่งสำเร็จการศึกษา

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา พี่สาว พี่ชาย ซึ่งสนับสนุนในด้านทุนการศึกษา และให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา ขอขอบคุณผู้ที่ให้คำปรึกษา ด้านโปรแกรม นาย กรกฎ วัฒนวีร์ รวมถึงเพื่อนๆปริญญาโท รุ่น 52 ทุกคน ที่ช่วยกันส่งเสริมและให้กำลังใจกันและกันตลอดมา

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ฎ
บทที่	1
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ขอบเขตการวิจัย.....	3
1.4 ข้อตกลงเบื้องต้น.....	4
1.5 คำจำกัดความที่ใช้ในการวิจัย.....	5
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	6
1.7 วิธีการดำเนินการวิจัย.....	6
2 แนวคิด ทฤษฎีและสถิติที่เกี่ยวข้อง.....	8
2.1 ส่วนประกอบของ GLM.....	8
2.2 ตัวแบบพหุคูณแบบ 2 ประเภท.....	10
2.3 ฟังก์ชันความน่าจะเป็นของข้อมูลการถดถอยพหุคูณ.....	12
2.4 การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด.....	13
2.5 สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient).....	14
2.6 ตัวแบบไม่ติดกลุ่ม (Non-Nested Model).....	15
2.7 การตรวจสอบความเหมาะสมของตัวแบบ.....	15
2.8 Receiver Operating Characteristic curve หรือ ROC curve.....	16
2.9 พื้นที่ใต้โค้ง ROC (Area under the Curve หรือ AUC).....	18

บทที่	ช หน้า
3 วิธีดำเนินการวิจัย.....	21
3.1 เทคนิคมอนติคาร์โล.....	21
3.2 แผนการดำเนินการวิจัย.....	22
3.3 การจำลองข้อมูลที่ใช้ในการวิจัย.....	23
3.4 ขั้นตอนในการดำเนินการวิจัย.....	24
3.5 การคำนวณพื้นที่ใต้โค้ง ROC.....	25
3.6 ทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุด.....	26
3.7 สรุปผลการวิจัยในแต่ละสถานการณ์.....	26
3.8 ขั้นตอนการทำงานของโปรแกรม.....	27
4 ผลการวิเคราะห์ข้อมูล.....	29
4.1 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ.....	31
5 สรุปผลการวิจัย และข้อเสนอแนะ.....	80
5.1 สรุปผลการคัดเลือกตัวแบบที่เหมาะสมสำหรับตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง ภายใต้สถานการณ์ต่างๆ.....	81
5.2 ข้อเสนอแนะ.....	83
รายการอ้างอิง.....	84
ประวัติผู้เขียนวิทยานิพนธ์.....	85



สารบัญตาราง

ตารางที่	หน้า
4.1 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=2, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	32
4.2 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=3, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	35
4.3 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=4, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	38
4.4 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	41
4.5 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	44
4.6 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	47
4.7 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	50
4.8 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	53

4.9	แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	56
4.10	แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	59
4.11	แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	62
4.12	แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	65
4.13	แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	68
4.14	แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	71
4.15	แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	74
4.16	แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ.....	77

## สารบัญภาพ

ภาพที่		หน้า
4.1	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=2, p_2=2, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	33
4.2	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=2, p_2=2, Max\{r_{ij}\} = Low, Medium, High$ และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	34
4.3	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=2, p_2=3, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	36
4.4	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=2, p_2=3, Max\{r_{ij}\} = Low, Medium, High$ และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	37
4.5	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=2, p_2=4, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	39
4.6	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=2, p_2=4, Max\{r_{ij}\} = Low, Medium, High$ และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	40
4.7	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=2, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	42
4.8	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=2, p_2=5, Max\{r_{ij}\} = Low, Medium, High$ และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	43
4.9	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=3, p_2=2, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	45
4.10	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=3, p_2=2, Max\{r_{ij}\} = Low, Medium, High$ และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	46
4.11	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=3, p_2=3, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	48
4.12	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=3, p_2=3, Max\{r_{ij}\} = Low, Medium, High$ และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	49
4.13	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=3, p_2=4, n=50, n=100, n=150, n=200, n=250$ และ $Max\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	51

ภาพที่	หน้า	
4.14	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=3, p_2=4, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	52
4.15	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=3, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	54
4.16	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=3, p_2=5, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	55
4.17	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=4, p_2=2, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	57
4.18	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=4, p_2=2, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	58
4.19	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=4, p_2=3, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	60
4.20	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=4, p_2=3, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	61
4.21	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=4, p_2=4, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	63
4.22	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=4, p_2=4, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	64
4.23	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=4, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	66
4.24	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=4, p_2=5, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	67
4.25	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=5, p_2=2, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	69
4.26	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=5, p_2=2, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	70
4.27	แสดงค่าของ $\overline{AUC}$ เมื่อ $p_1=5, p_2=3, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	72

ภาพที่	หน้า	
4.28	แสดงค่าของ $\overline{A\hat{U}C}$ เมื่อ $p_1=5, p_2=3, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	73
4.29	แสดงค่าของ $\overline{A\hat{U}C}$ เมื่อ $p_1=5, p_2=4, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	75
4.30	แสดงค่าของ $\overline{A\hat{U}C}$ เมื่อ $p_1=5, p_2=4, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	76
4.31	แสดงค่าของ $\overline{A\hat{U}C}$ เมื่อ $p_1=5, p_2=5, n=50, n=100, n=150, n=200, n=250$ และ $\text{Max}\{r_{ij}\}$ เปลี่ยนแปลงไปเป็น Low, Medium และ High.....	78
4.32	แสดงค่าของ $\overline{A\hat{U}C}$ เมื่อ $p_1=5, p_2=5, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$ และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น $n= 50, 100, 150, 200, 250$ .....	79

## บทที่ 1

### บทนำ

#### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันงานวิจัยด้านต่างๆ หลายแขนงทั้งทางด้านเศรษฐศาสตร์ สังคมศาสตร์ วิทยาศาสตร์ การแพทย์ ธุรกิจ การเงิน ล้วนแล้วแต่จะต้องอาศัยวิธีการทางสถิติมาเป็นเครื่องมือในการวิเคราะห์ข้อมูลเพื่อประกอบการตัดสินใจทั้งสิ้น โดยการเลือกใช้วิธีการทางสถิตินั้น จะต้องพิจารณาจากชนิดของข้อมูล วัตถุประสงค์ที่ต้องการศึกษาและข้อสมมติเบื้องต้น เพื่อผลการวิเคราะห์ทางสถิติที่ถูกต้องและแม่นยำ

การวิเคราะห์ข้อมูลด้วยการถดถอยโลจิสติก (Logistic Regression Analysis) เป็นการวิเคราะห์ทางสถิติที่ได้รับความนิยม ตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท (Binary Logistic Regression Model) ซึ่งใช้ในการพยากรณ์ตัวแปรตามเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า (Dichotomy or Binary Variable) คือค่า 0 และ 1 ว่าจะอยู่ในกลุ่มใดกลุ่มหนึ่งใน 2 กลุ่มโดยใช้ตัวแปรอิสระเป็นตัวพยากรณ์ จึงถูกนำมาใช้ประโยชน์อย่างแพร่หลาย และได้มีการประยุกต์ใช้กับตัวแบบเชิงเส้นที่วางนัยทั่วไป (Generalized Linear Model: GLM) ซึ่งจะประกอบด้วย 3 ส่วน คือ ส่วนประกอบเชิงสุ่ม (Random Component) เป็นส่วนแสดงการแจกแจงความน่าจะเป็นของตัวแปรตาม อยู่ในกลุ่ม Exponential ส่วนประกอบแบบมีระบบ (Systematic Component) แสดงฟังก์ชันเชิงเส้นของตัวแปรอิสระ (Linear Predictor) และส่วนประกอบการเชื่อมโยงหรือเรียกอีกอย่างว่าฟังก์ชันการเชื่อมโยง (Link Function) สำหรับเชื่อมส่วนประกอบเชิงสุ่ม กับส่วนประกอบแบบมีระบบเข้าด้วยกัน ซึ่งในฟังก์ชันการเชื่อมโยง (Link Function) แบ่งเป็น ฟังก์ชันการเชื่อมโยงโลจิส (Logit Link Function) ฟังก์ชันการเชื่อมโยง โพรบิต (Probit Link Function) และ ฟังก์ชันการเชื่อมโยงคอมพลีเมนทารี ล็อก-ล็อก (Log-Log Link Function) การพิจารณาใช้ Link Function ควรพิจารณาเพื่อสามารถสร้างตัวแบบความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามได้อย่างเหมาะสม

กุลพัชร หมื่นมา ได้ทำการศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบโลจิส ตัวแบบโพรบิต และตัวแบบคอมพลีเมนทารี ล็อก-ล็อก เมื่อตัวแปรตอบสนองมีสองกลุ่ม ซึ่งการเปรียบเทียบกระทำภายใต้ข้อมูลทางด้านการแพทย์ ข้อมูลทางด้านวิทยาศาสตร์ (ชีววิทยา) ข้อมูลทางด้านวิศวกรรมศาสตร์ และข้อมูลทางด้านสังคมศาสตร์ วิธีวิเคราะห์คือ การประมาณค่าของตัวแบบโลจิส ตัวแบบโพรบิต และตัวแบบคอมพลีเมนทารี ล็อก-ล็อก ซึ่งใช้ตัวสถิติ Deviance เป็นเกณฑ์ในการเปรียบเทียบ ผลการวิเคราะห์พบว่า ตัวแบบโลจิสเป็นตัวแบบที่เหมาะสมกับ

ข้อมูลทางด้านวิทยาศาสตร์ (ชีววิทยา) ตัวแบบโพรบิตเป็นตัวแบบที่เหมาะสมกับข้อมูลทางด้าน การแพทย์ ข้อมูลทางด้านวิทยาศาสตร์ (ชีววิทยา) และข้อมูลทางด้านสังคมศาสตร์ ตัวแบบคอม พลิเมนทารี ล็อก-ล็อก เป็นตัวแบบที่เหมาะสมกับข้อมูลทางด้านวิศวกรรมศาสตร์ ซึ่งในงานวิจัยนี้ ผู้วิจัยสนใจที่จะศึกษาตัวแบบโพรบิตเนื่องจากเป็นตัวแบบที่เหมาะสมกับข้อมูลหลายด้าน

ในการพยากรณ์เพื่อมุ่งหวังให้เกิดการจำแนกประเภท การแบ่งกลุ่ม เช่น การเป็น โรคเบาหวานหรือการไม่เป็นโรคเบาหวาน การเลือกซื้อสินค้าหรือไม่เลือกซื้อสินค้าและลูกค้าเป็น ลูกหนี้ที่มีปัญหาหรือลูกค้าเป็นลูกหนี้ที่ดี เป็นต้น ดังนั้น ตัวแบบโพรบิตแบบ 2 ประเภท ( Binary Probit Model) จึงเป็นอีกตัวแบบที่นิยมกันแพร่หลายสำหรับตัวแบบที่มีข้อจำกัดเกี่ยวกับตัวแปร ตามที่เป็นตัวแปรเชิงคุณภาพมีค่าได้เพียง 2 ค่าเท่านั้น คือค่า 0 และ 1 โดยกำหนดให้ตัวแปรตาม มีค่าเท่ากับ 1 เมื่อเกิดเหตุการณ์ที่สนใจ (Success) และตัวแปรตามมีค่าเท่ากับ 0 เมื่อไม่เกิด เหตุการณ์ที่สนใจ(Failure) ของสิ่งที่สนใจศึกษา ซึ่งเป็นการวิเคราะห์ข้อมูลโดยการหาตัวแบบ ความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ เพื่อนำตัวแบบที่ได้นำไปใช้ในการพยากรณ์

นอกจากนี้การใช้ตัวแบบโพรบิตยังเป็นเทคนิคอย่างหนึ่งในการแก้ปัญหาของตัวแบบ ความน่าจะเป็นเชิงเส้นที่มีค่าประมาณความน่าจะเป็นอยู่นอกช่วง 0 ถึง 1 การวิเคราะห์ตัวแบบ โพรบิตได้ใช้ฟังก์ชันการแจกแจงสะสมปกติมาตรฐาน (standard normal cumulative distribution function) แปลงค่าเฉลี่ยของตัวแปรตาม ให้เป็นค่าความน่าจะเป็นของการเกิด เหตุการณ์ที่สนใจ ตัวแบบโพรบิตมีลักษณะไม่เป็นเชิงเส้น (Nonlinear) กล่าวคือมีลักษณะเป็นโค้ง รูปตัวเอส (S-Shaped หรือ Sigmoid Curve) ซึ่งจะทำให้ค่าความน่าจะเป็นของการเกิดเหตุการณ์ ที่สนใจ มีค่าอยู่ในช่วง 0 ถึง 1 ตามคุณสมบัติเบื้องต้นของความน่าจะเป็น และมีความสัมพันธ์ ในทางเดียวกัน (monotonic) ระหว่างตัวแปรตาม และความน่าจะเป็นของการเกิดเหตุการณ์ที่ สนใจ

การคัดเลือกตัวแบบโพรบิตแบบ 2 ประเภท เป็นวิธีการทางสถิติที่เกี่ยวข้องกับการเลือกตัว แปรอิสระที่ดีที่สุดซึ่งใช้เป็นตัวพยากรณ์ โดยใช้พื้นที่ใต้โค้ง ROC (Receiver Operating Characteristic curve) สูงสุดเป็นเกณฑ์หรือเรียกอีกอย่างว่า พื้นที่ใต้โค้ง (Area Under the Curve: AUC) โดยพิจารณาพื้นที่ใต้โค้งมีค่ามาก จะหมายความว่า ตัวแบบนั้นมีอัตราความ ถูกต้องมาก ซึ่งในหลายๆงานวิจัยจะเปรียบเทียบกลุ่มของตัวแปรอิสระที่จะเลือกเป็นกลุ่มของตัว พยากรณ์ที่ดีที่สุดในตัวแบบทางสถิติแบบติดกลุ่ม (Nested) และแบบไม่ติดกลุ่ม(Non-Nested) ซึ่ง ตัวแบบไม่ติดกลุ่มสามารถแบ่งออกเป็น 2 ประเภทหลักๆคือ ตัวแบบไม่ติดกลุ่มบางส่วน (Partially non-nested) และตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ (Strictly non-nested) นั่นคือ 2 ตัวแบบไม่ติด กกลุ่มจะถูกเรียกว่า ตัวแบบไม่ติดกลุ่มบางส่วน (Partially non-nested) ก็ต่อเมื่อตัวแบบที่หนึ่ง

ประกอบด้วยตัวแปรอิสระบางตัวของตัวแบบที่สอง และ 2 ตัวแบบไม่ติดกลุ่มจะถูกเรียกว่า ตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ (Strictly non-nested) ก็ต่อเมื่อตัวแบบที่หนึ่งไม่มีตัวแปรอิสระใดๆ ของตัวแบบที่สองอยู่เลย

ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาว่าตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยงแบบใดที่จะทำให้ ค่าพื้นที่ใต้เส้นโค้ง ROC สูงสุด โดยพิจารณาจากขนาดตัวอย่าง , ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ และจำนวนตัวแปรอิสระ ถ้าขนาดตัวอย่างเท่ากันและระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระเท่ากัน เซตของจำนวนอิสระเท่าใดจึงเหมาะสมที่ถูกเลือกเข้ามาอยู่ในตัวแบบ หรือในกรณีที่ว่าจำนวนตัวแปรอิสระหรือขนาดตัวอย่างแตกต่างกันแล้ว สาเหตุใดที่ทำให้เราเลือกตัวแบบนั้นเข้ามาอยู่ในตัวแบบ

## 1.2 วัตถุประสงค์ของการวิจัย

- 1.2.1 เพื่อศึกษาระดับความสัมพันธ์ในแต่ละคู่ (degree of pair-wise correlation) ของตัวแปรอิสระที่ส่งผลให้การเลือกตัวแบบโพรบิตแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างสมบูรณ์ดีที่สุด เมื่อขนาดตัวอย่างเท่ากัน จำนวนของตัวแปรอิสระเท่ากัน โดยใช้พื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือก
- 1.2.2 เพื่อศึกษาหาจำนวนตัวแปรอิสระที่ส่งผลให้การเลือกตัวแบบโพรบิตแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างสมบูรณ์ดีที่สุด เมื่อขนาดตัวอย่างเท่ากัน และมีระดับความสัมพันธ์ในแต่ละคู่ (degree of pair-wise correlation) อยู่ในระดับเดียวกัน โดยใช้พื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือก

## 1.3 ขอบเขตของการวิจัย

การวิจัยครั้งนี้มีขอบเขตของการวิจัยสำหรับการดำเนินการวิจัย ดังนี้

- 1.3.1 ทำการศึกษาตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกส์แบบ 2 ประเภทที่มีฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง เพื่อหาตัวแบบที่เหมาะสมที่สุด
- 1.3.2 ตัวแปรตาม (Y) เป็นข้อมูลเชิงคุณภาพที่มี 2 ค่า คือ 0 และ 1
- 1.3.3 จำนวนตัวแปรอิสระ ( $p_1, p_2$ ) ถูกกำหนดเป็น
 
$$(p_1, p_2) = (2, 2), (2, 3), (2, 4), (2, 5),$$

$$(p_1, p_2) = (3, 2), (3, 3), (3, 4), (3, 5)$$



$$(p_1, p_2) = (4, 2), (4, 3), (4, 4), (4, 5)$$

$$(p_1, p_2) = (5, 2), (5, 3), (5, 4), (5, 5)$$

1.3.4 ขนาดตัวอย่าง ( $n$ ) ในการศึกษาครั้งนี้เป็น  $n = 50, 100, 150, 200, 250$

1.3.5 คัดเลือกคู่ของตัวแปรอิสระเข้าในตัวอย่างแบบโพรบิตแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างสมบูรณ์ โดยใช้ค่า  $\text{Max}\{r_{jj}\}$  เป็นเกณฑ์ในการคัดเลือก นั่นคือเลือกคู่ของตัวแปรอิสระที่มีระดับความสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) มากที่สุดเข้าในตัวอย่าง

$$0 \leq \text{Max}\{r_{jj}\} \leq 0.99$$

โดยที่  $r_{jj}$  คือ สหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่  $j$  และตัวแปรอิสระตัวที่  $j'$  เกณฑ์สำหรับการแบ่งระดับความสัมพันธ์ระหว่างตัวแปรอิสระ แสดงดังนี้

$$\text{ตัวแปรอิสระมีความสัมพันธ์กันในระดับอย่างต่ำ: } 0 \leq \text{Max}\{r_{jj}\} \leq 0.33$$

$$\text{ตัวแปรอิสระมีความสัมพันธ์กันในระดับปานกลาง: } 0.33 < \text{Max}\{r_{jj}\} \leq 0.66$$

$$\text{ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูง: } 0.66 < \text{Max}\{r_{jj}\} \leq 0.99$$

1.3.6 ตัวแปรอิสระ ( $X$ ) มีการแจกแจงเริ่มต้น เป็นการแจกแจงแบบยูนิฟอร์ม

1.3.7 กำหนดค่าพารามิเตอร์เริ่มต้นของสมการถดถอยเป็นค่าใดๆในการศึกษาครั้งนี้ คือ

$$\beta_i = 0.1 ; i = 0, 1, 2, \dots, p_1 + p_2 \text{ และ } \varepsilon_i \sim N(0, 1); i = 1, 2, \dots, n$$

1.3.8 กำหนดระดับนัยสำคัญ ( $\alpha$ ) ในการศึกษาครั้งนี้ที่ระดับ 0.05 ( $\alpha = 0.05$ )

1.3.9 ในการศึกษาครั้งนี้ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation) โดยการจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 500 รอบ ( $N=500$ )

#### 1.4 ข้อตกลงเบื้องต้น

สำหรับการศึกษาวิจัยครั้งนี้มาจาก ตัวอย่างที่จำลองขึ้นมาจากตัวแบบ และการแจกแจงดังนี้

1.4.1 ศึกษาตัวแบบโพรบิตแบบ 2 ประเภท โดยมีรูปแบบคือ

$$\begin{aligned}
 \pi_i &= P(Y_i = 1) \\
 &= P(Y_i^* > 0) \\
 \text{จาก} \quad &= P(\varepsilon_i < \beta' X_i) ; i=1,2,\dots,n \\
 &= \int_{-\infty}^{\beta' X_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2}\right) d\varepsilon
 \end{aligned}$$

โดยที่  $\pi_i$  คือ ค่าความน่าจะเป็นเมื่อเกิดเหตุการณ์ที่สนใจในหน่วยที่  $i$   
 $Y_i$  คือ ตัวแปรตามเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า คือ 0 และ 1  
 $Y_i^*$  คือ ตัวแปรแฝง (Latent Variable) เป็นค่าที่วัดไม่ได้ จึงไม่ทราบค่าที่แท้จริง ทราบเพียงแต่ผลที่เกิดขึ้น

$\varepsilon_i$  คือ ตัวแปรสุ่มปกติมาตรฐาน โดย  $\varepsilon_i \sim N(0,1)$  <sup>iid</sup>

$\beta'$  คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยจำนวน  $k+1$  ตัว ขนาด  $1 \times (k+1)$

$X_i$  คือ เวกเตอร์ของตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณจำนวน  $k$  ตัว  
 ขนาด  $(k+1) \times 1$

4.2.1 ตัวแปรอิสระเริ่มต้นมีการแจกแจงแบบยูนิฟอร์ม ที่มีฟังก์ชันความน่าจะเป็นของ  $x$  อยู่ในรูปของ

$$f(x) = \frac{1}{b-a} ; a < x < b \quad \text{เมื่อ } a < b \text{ และ } a, b \text{ เป็นค่าคงที่}$$

โดยมี

$$\text{ค่าเฉลี่ย } E(X) = \mu = \frac{a+b}{2}$$

$$\text{ค่าความแปรปรวน } \text{Var}(X) = \sigma^2 = \frac{(b-a)^2}{12}$$

## 1.5 คำจำกัดความที่ใช้ในการวิจัย

1.5.1 พารามิเตอร์ (Parameter) หมายถึง ค่าคงที่ที่แสดงคุณลักษณะบางประการของประชากร

1.5.2 ค่าเฉลี่ย (Mean) หมายถึง ค่าที่แสดงว่าข้อมูลส่วนใหญ่อยู่ที่ไหน

1.5.3 ฟังก์ชันการเชื่อมโยง (Link function) หมายถึง ส่วนที่ใช้อธิบายความสัมพันธ์ระหว่างองค์ประกอบเกี่ยวกับตัวแปรสุ่มและองค์ประกอบเชิงระบบ

1.5.4 ตัวแบบโพรบิตแบบ 2 ประเภท (Binary Probit Model) หมายถึง ตัวแบบที่ศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ เมื่อตัวแปรตามเป็นตัวแปรเชิงคุณภาพที่มี 2 ลักษณะ คือมีค่าเท่ากับ 1 หรือ 0 และตัวแปรอิสระเป็นตัวแปรเชิงปริมาณ

หรือตัวแปรหุ่น เมื่อได้รูปแบบความสัมพันธ์ระหว่างตัวแปรแล้ว จะนำไปใช้ในการประมาณค่าตัวแปรตาม

1.5.5 ความสัมพันธ์ระหว่างตัวแปรอิสระ(multicollinearity) หมายถึง สถานการณ์ที่ตัวแปรอิสระมีความสัมพันธ์กัน

1.5.6 การแจกแจงแบบเบอร์นูลลี (Bernoulli Distribution) หมายถึงตัวแปรสุ่ม  $Y$  เรียกว่าตัวแปรสุ่มเบอร์นูลลี กล่าวคือ

ถ้า  $Y = 0$  เมื่อ การทดลองไม่เกิดเหตุการณ์ที่สนใจ

และถ้า  $Y = 1$  เมื่อ การทดลองเกิดเหตุการณ์ที่สนใจ

โดยที่  $P(Y = 1) = p$ ,  $0 < p < 1$  และ  $P(Y = 0) = 1 - p$

เราอาจเขียนแทนด้วย  $Y \sim \text{Ber}(p)$  ซึ่งฟังก์ชันความน่าจะเป็นอยู่ในรูปของ

$$P(Y=y) = p^y (1-p)^{1-y}$$

1.5.7 ค่าประมาณพื้นที่ใต้โค้ง ROC ( $AUC$ ) หมายถึงดัชนีในการบ่งชี้ความสามารถในการจำแนกกลุ่มหรือความเชื่อถือได้ของตัวแบบ

1.5.8 ตัวแบบไม่ติดกลุ่ม (Non-nested Model) หมายถึง ตัวแบบ 2 ตัวแบบจะไม่ติดกลุ่มกัน ถ้าตัวแบบหนึ่งไม่สามารถลดรูปไปเป็นอีกตัวแบบหนึ่งได้ ด้วยการสมมติให้พารามิเตอร์บางตัวมีค่าเป็นศูนย์

## 1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 เพื่อเป็นแนวทางสำหรับนักวิจัยสาขาใดๆ ที่วางแผนจะใช้ตัวแบบโพรบิตแบบ 2 ประเภท เพื่อคัดเลือกตัวแบบที่ดีที่สุดแบบไม่ติดกลุ่มอย่างสมบูรณ์ ซึ่งจะสามารถประเมินและพยากรณ์ตัวแปรอิสระที่ถูกเลือก รวมไปถึงตัวแบบโดยดูจากจำนวนตัวแปรอิสระในแต่ละชุด, ขนาดตัวอย่าง และความสัมพันธ์ของแต่ละคู่ระหว่างตัวแปรอิสระ ซึ่งจะทำให้นักวิจัยได้ตัวแบบที่ดีที่สุดเพื่อนำไปใช้ในการพยากรณ์หรือการวิจัยต่อไป

1.6.2 เพื่อใช้เป็นแนวทางในการศึกษาวิจัยต่อไป

## 1.7 วิธีดำเนินการวิจัย

- 1.7.1 ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับงานวิจัย
- 1.7.2 จำลองข้อมูลตามขอบเขตที่ต้องการศึกษา
- 1.7.3 สร้างตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกแบบ 2 ประเภทที่มีฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง
- 1.7.4 หาระดับความสัมพันธ์ในแต่ละคู่ (degree of pair-wise colleration) ระหว่างแต่ละคู่ของตัวแปรอิสระจากตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกแบบ 2 ประเภทที่มีฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง
- 1.7.5 นำตัวแบบที่ได้ไปพล็อต ROC curve และคำนวณหาค่าประมาณพื้นที่ใต้โค้ง ROC
- 1.7.6 เปรียบเทียบพื้นที่ใต้โค้ง ROC เพื่อหาตัวแบบที่ดีที่สุด
- 1.7.7 สรุปผลที่ได้จากการวิจัย

## บทที่ 2

### แนวคิด ทฤษฎีและสถิติที่เกี่ยวข้อง

การคัดเลือกตัวแบบที่เหมาะสมที่สุดสำหรับการวิจัยครั้งนี้จะคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ โดยใช้ฟังก์ชันเชื่อมโยงโพรบิตแบบ 2 ประเภท ซึ่งตัวแบบที่เหมาะสมที่สุดจะให้ค่าความน่าเชื่อถือของตัวแบบสูงสุด ในบทนี้จะกล่าวถึงรายละเอียดเกี่ยวกับ ฟังก์ชันเชื่อมโยง ตัวแบบโพรบิตแบบ 2 ประเภท การประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบโพรบิตแบบ 2 ประเภทด้วยวิธีภาวะความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) ตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ (Stricly Non-nested Model) และการหาความน่าเชื่อถือของตัวแบบโดยใช้พื้นที่ใต้โค้ง ROC

#### 2.1 ส่วนประกอบของ GLM

GLM เป็นตัวแบบที่ขยายจากตัวแบบเชิงเส้นแบบคลาสสิก (classical linear model) โดยส่วนประกอบแรกที่เกี่ยวข้องกับตัวแปรเชิงสุ่มนั้น นอกจากมีการแจกแจงแบบปกติแล้วยังสามารถขยายไปสู่การแจกแจงในกลุ่มเอกซ์โปเนนเชียล (exponential family) ได้และส่วนประกอบของ link function นอกจากจะใช้ identity link แล้วก็ยังสามารถขยายให้ใช้กับฟังก์ชันตัวเชื่อมอื่นๆอีกหลายแบบที่เป็นฟังก์ชันแบบ monotonic differentiable function ใตๆก็ได้ส่วนประกอบทั้ง 3 ส่วนมีรายละเอียดของแต่ละส่วนประกอบดังนี้

ส่วนประกอบที่ 1 ของ GLM คือส่วนประกอบที่เกี่ยวข้องกับสมมติฐานของการแจกแจงของตัวแปรเชิงสุ่ม (Y) ที่เป็นตัวแปรตอบสนอง สมมติว่าค่าสังเกตจาก Y มีขนาด n หน่วยที่เป็นอิสระต่อกัน นั่นคือ  $Y = (y_1, \dots, y_n)$  แต่ละส่วนประกอบของ Y คือ  $y_i, i=1, \dots, n$  มีการแจกแจงในกลุ่มเอกซ์โปเนนเชียล ซึ่งอยู่ในรูปของ

$$f_y(y; \theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right] \dots \dots (1)$$

โดยที่ a(.) , b(.) และ c(.) แทนฟังก์ชันต่างๆเมื่อทราบ  $\phi$  แล้ว (1) คือ ตัวแบบหนึ่งในกลุ่มเอกซ์โปเนนเชียลที่มีพารามิเตอร์  $\theta$  แต่ถ้าไม่ทราบ  $\phi$  (1) อาจเป็นหรือไม่เป็นตัวแบบหนึ่งในกลุ่มเอกซ์โปเนนเชียลที่มีพารามิเตอร์ 2 ตัว  $(\theta, \phi)$  สำหรับพารามิเตอร์  $\theta$  ที่เรียกว่า natural parameter ส่วน  $\phi$  มักเรียกว่า dispersion parameter และฟังก์ชัน  $a(\phi)$  มักจะมีรูปแบบเป็น

$$a(\phi) = \phi / w_i$$

โดยที่  $w_i$  แทนน้ำหนักที่ทราบค่า เช่นเมื่อ  $\bar{y}_i$  แทนค่าเฉลี่ยของ  $n_i$  หน่วยที่เป็นอิสระต่อกัน จะใช้โดยทั่วไปว่า  $w_i = n_i$  และเพื่อให้เกิดความเข้าใจในตัวแบบ (1) ได้ชัดเจนขึ้นจะยกตัวอย่าง การแจกแจงแบบปกติ ซึ่งสามารถเขียนให้อยู่ในรูปแบบของ (1) ได้ดังนี้

$$\begin{aligned} f_y(y; \theta, \phi) &= \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{\frac{(y\mu - \mu^2/2)}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2)\right)\right\} \end{aligned}$$

โดย  $\theta = \mu$ ,  $b(\theta) = \theta^2 / 2$ ,  $a(\phi) = \phi = \sigma^2$  และ  $c(y, \phi) = -1/2(y^2 / \sigma^2 + \log(2\pi\sigma^2))$

ทำนองเดียวกันกับการแจกแจงแบบปกติ การแจกแจงแบบปัวซอง (Poisson) การแจกแจงแบบทวินาม (binomial) และการแจกแจงในกลุ่มเอกซ์โปเนนเชียลอื่นๆ ก็สามารถเขียนให้อยู่ในรูปแบบของ (1) ได้เช่นกัน และในกรณีที่  $\phi$  เป็นค่าคงที่ที่ทราบค่า (1) จะอยู่ในรูปของ (1) คือ

$$f(y_i; \theta_i) = a(\theta_i)b(y_i) \exp\{y_i Q(\theta)\} \dots \dots \dots (2)$$

โดยที่  $Q(\theta)$  ใน (2) คือ  $\theta / a(\phi)$  ใน (2)

$a(\theta)$  ใน (2) คือ  $\exp\{-b(\theta)\} / a(\phi)$  ใน (2)

$b(y)$  ใน (2) คือ  $\exp\{c(y, \phi)\}$  ใน (2)

จะเห็นว่าตัวแบบ (1) มีรูปแบบทั่วไปที่สามารถนำไปใช้ประโยชน์กับการแจกแจงหลายรูปแบบ โดยเฉพาะสำหรับกลุ่มพารามิเตอร์ 2 ตัว (two-parameter families) เช่นการแจกแจงแบบปกติและการแจกแจงแบบแกมมา (gamma) ซึ่ง  $\phi$  จะเป็นพารามิเตอร์ของความคลาดเคลื่อน (nuisance parameter) ส่วนการแจกแจงสำหรับกลุ่มที่มีพารามิเตอร์ตัวเดียว (one-parameter families) เช่น การแจกแจงแบบปัวซองของแบบทวินาม ไม่จำเป็นต้องใช้เทอม  $\phi$

ส่วนประกอบที่ 2 ของ GLM คือ ส่วนประกอบแบบมีระบบ ทำหน้าที่เชื่อมเวกเตอร์  $\eta$  โดยที่  $\eta = (\eta_1, \dots, \eta_N)'$  กับเซตของตัวแปรอธิบาย ให้มีรูปแบบเชิงเส้นดังนี้

$$\eta = X\beta \quad ; \quad \eta_j = \sum_i \beta_i x_{ij} \quad ; \quad i = 1, \dots, p \quad j = 1, \dots, N$$

โดยที่  $X$  แทนเมทริกซ์ของตัวแปรอธิบายที่ประกอบด้วยค่าสังเกตขนาด  $N$  อาจเรียก  $X$  ว่า

design matrix ที่มีขนาด  $(N \times p)$

$\beta$  เชนเวกเตอร์ของพารามิเตอร์  $(\beta_1, \dots, \beta_p)'$

$\eta$  เชนตัวพยากรณ์เชิงเส้น (linear predictor)

ส่วนประกอบที่ 3 ของ GLM คือ link function ต่างๆสำหรับเชื่อมส่วนประกอบเชิงสุ่มและส่วนประกอบมีระบบเข้าด้วยกัน เช่น

ให้  $\mu_j = E(Y_j) \quad ; i=1, \dots, p \quad j=1, \dots, N$

$\mu_j$  จะเกี่ยวข้องกับ  $\eta_j$  ในรูปฟังก์ชันของ  $\eta_j = g(\mu_j)$

โดยที่  $g$  เชนฟังก์ชันแบบ monotonic differentiable function ดังนั้นตัวแบบที่จะต้องการเชื่อมระหว่างค่าเฉลี่ยของค่าสังเกตของ  $Y$  กับตัวแปรอธิบาย คือ

$$g(\mu_j) = \sum_i \beta_i x_{ij}, \quad ; i=1, \dots, p \quad j=1, \dots, N$$

โดยที่  $p$  เชนจำนวนตัวแปรอธิบาย

ถ้า  $g(\mu) = \mu$  จะได้ว่า  $\eta_j = \mu_j$  คือ identity link หรือเรียกอีกอย่างหนึ่งว่า canonical link โดยมีการเปลี่ยนแปลงค่าเฉลี่ยให้อยู่ในเทอมของพารามิเตอร์นั้นคือ

$$g(\mu_j) = Q(\theta_j)$$

และ  $Q(\theta_j) = \sum_i \beta_i x_{ij} \quad ; i=1, \dots, p \quad j=1, \dots, N$

สรุปว่า GLM เป็นตัวแบบเชิงเส้นสำหรับค่าเฉลี่ยที่แปลงแล้วของตัวแปรซึ่งมีการแจกแจงอยู่ในกลุ่มเอกซ์โปเนนเชียล (กุลพัชร หมิ่นมา, 2551)

ซึ่งในฟังก์ชันการเชื่อมโยง (Link Function) แบ่งเป็น ฟังก์ชันการเชื่อมโยงโลจิท (Logit Link Function) ฟังก์ชันการเชื่อมโยงโพรบิต (Probit Link Function) และฟังก์ชันการเชื่อมโยงคอมพลีเมนทารี ล็อก-ล็อก (Log-Log Link Function) โดยการเลือกใช้ฟังก์ชันเชื่อมโยงขึ้นอยู่กับลักษณะของการกระจายของข้อมูล

## 2.2 ตัวแบบโพรบิตแบบ 2 ประเภท (Binary Probit Model)

ตัวแบบที่ศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระเมื่อตัวแปรตามเป็นตัวแปรเชิงคุณภาพที่มี 2 ลักษณะ และตัวแปรอิสระเป็นตัวแปรเชิงปริมาณหรือตัวแปรหุ่น เมื่อได้รูปแบบความสัมพันธ์ระหว่างตัวแปรแล้ว จะนำไปใช้ในการประมาณค่าตัวแปรตาม และพยากรณ์โอกาสที่แต่ละหน่วยจะอยู่ในกลุ่มใดกลุ่มหนึ่งได้

จากตัวแปรตาม ( $Y_i$ ) ซึ่งเป็นตัวแปรเชิงคุณภาพที่เป็นได้ 2 ค่า คือ 0 กับ 1 และ  $X_{i1}, X_{i2}, \dots, X_{ik}$  เป็นตัวแปรอิสระของค่าสังเกตที่ได้จากหน่วยตัวอย่างที่  $i$  โดยที่  $i = 1, 2, 3, \dots, n$  โดยมี ตัวแปรแฝง คือ  $Y_i^*$  ซึ่งเป็นค่าที่วัดไม่ได้ จึงไม่ทราบค่าที่แท้จริง ทราบเพียงแต่ผลที่เกิดขึ้น โดย  $Y_i^*$  ของหน่วยตัวอย่างที่  $i$  เป็นฟังก์ชันเชิงเส้นของตัวแปรอิสระ  $X_{i1}, X_{i2}, \dots, X_{ik}$  และนั่นคือ

$$Y_i^* = \beta'X_i + \varepsilon_i$$

เมื่อ  $\varepsilon_i$  คือค่าความคลาดเคลื่อนของหน่วยตัวอย่าง และ  $\varepsilon_i \sim N(0,1)$  <sup>iid</sup>

จะได้

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* > 0 \text{ or } \varepsilon_i < \beta'X_i \\ 0 & \text{if } Y_i^* \leq 0 \text{ or } \varepsilon_i \geq \beta'X_i \end{cases}$$

จาก

$$\begin{aligned} \pi_i &= P(Y_i = 1) \\ &= P(Y_i^* > 0) \\ &= P(\varepsilon_i < \beta'X_i) \\ &= \Phi(\beta'X_i) \\ &= \int_{-\infty}^{\beta'X_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2}\right) d\varepsilon \end{aligned}$$

โดยที่  $\pi_i$  คือ ค่าความน่าจะเป็นเมื่อเกิดเหตุการณ์ที่สนใจในหน่วยที่  $i$

$Y_i$  คือ ตัวแปรตามเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า คือ 0 และ 1

$Y_i^*$  คือ ตัวแปรแฝง (Latent Variable) เป็นค่าที่วัดไม่ได้ จึงไม่ทราบค่าที่แท้จริง ทราบเพียงแต่ผลที่เกิดขึ้น

$\Phi(\cdot)$  คือ ฟังก์ชันการแจกแจงสะสมปกติมาตรฐาน

$\varepsilon_i$  คือ ตัวแปรสุ่มปกติมาตรฐาน โดย  $\varepsilon_i \sim N(0,1)$  <sup>iid</sup>



$\beta'$  คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยจำนวน  $k+1$  ตัว ขนาด  $1 \times (k+1)$   
 $X_i$  คือ เวกเตอร์ของตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณจำนวน  $k$  ตัว  
 ขนาด  $(k+1) \times 1$

ดังนั้น Probit Model สามารถเขียนแปลงให้อยู่ในรูป

$$\begin{aligned}\Phi^{-1}(\pi_i) &= \beta' X_i \\ &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} \dots\dots\dots(1)\end{aligned}$$

จากสมการที่(1) พบว่า การแปลงดังกล่าวจะทำให้เราได้ความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปร

ฟังก์ชันโพรบิต คือ

$$\text{Probit}(E[Y_i | \tilde{X}_i^T = \tilde{x}_i^T]) = \Phi^{-1}(E[Y_i | \tilde{X}_i^T = \tilde{x}_i^T]) = \Phi^{-1}(\pi_i)$$

ซึ่งฟังก์ชันโพรบิตจะทำการแปลงค่า  $\pi_i$  จากช่วง  $(0, 1)$  เป็นค่าที่อยู่ในช่วง  $(-\infty, \infty)$

## 2.3 ฟังก์ชันความน่าจะเป็น (Likelihood function) ของข้อมูลการถดถอยโพรบิต

จากประชากรที่ทำการศึกษาเป็นตัวแปรตาม ( $Y_i$ ) ที่มีเพียง 2 ค่า คือ 0 กับ 1 จึงใช้ฟังก์ชันการแจกแจงแบบเบอร์นูลลี

$$P(Y_i = y_i) = p^{y_i} (1-p)^{1-y_i} \quad ; y_i = 0, 1$$

สร้างฟังก์ชันของการแจกแจงความน่าจะเป็นร่วม (Joint Probability Density Function) ของหน่วยตัวอย่างอิสระ  $n$  ค่า โดยการคูณฟังก์ชันการแจกแจงความน่าจะเป็นของทุกหน่วยตัวอย่าง ( $g(Y_i)$ )

$$\begin{aligned}g(Y_1, Y_2, \dots, Y_n) &= \prod_{i=1}^n g(Y_i) \\ &= \prod_{i=1}^n p_i^{y_i} (1-p_i)^{1-y_i} \\ &= \prod_{i=1}^n [\Phi(\beta X_i')]^{y_i} [1-\Phi(\beta X_i')]^{1-y_i} \\ L(\beta) &= \prod_{i=1}^n [\Phi(\beta X_i')]^{y_i} [1-\Phi(\beta X_i')]^{1-y_i} \dots\dots\dots(2)\end{aligned}$$

## 2.4 การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation)

การหาค่าตัวประมาณค่าสัมประสิทธิ์การถดถอย ( $\hat{\beta}$ ) ด้วยวิธีภาวะความน่าจะเป็นสูงสุดคือต้องทำให้ L มีค่ามากที่สุดโดยทำการหาอนุพันธ์เทียบกับ  $\beta$  ต่างๆ โดยเราสามารถเขียนให้อยู่ในรูปของลอการิทึม (Logarithm) หรือความควรจะเป็นลอการิทึม (Log-Likelihood) ได้ดังนี้

$$\ln L(\beta) = \sum_{i=1}^n \{ y_i \ln [\Phi(X_i' \beta)] + (1 - y_i) \ln [1 - \Phi(X_i' \beta)] \} \dots (3)$$

$$= \sum_{y_i=0} \ln [1 - \Phi(X_i' \beta)] + \sum_{y_i=1} \ln \Phi(X_i' \beta) \dots (4)$$

และเงื่อนไขอันดับแรก (First Order) สำหรับการให้สมการที่ (4) มีค่าสูงสุด (Maximization) และสามารถแก้สมการหาค่า First Order Condition ของแบบจำลองโพรบิต ได้ดังนี้

$$\begin{aligned} \frac{\partial \ln L(\beta)}{\partial \beta} &= \sum_{i=1}^n \left\{ \frac{y_i \phi(\cdot)}{\Phi(\cdot)} + (1 - y_i) \left[ \frac{-\phi(\cdot)}{1 - \Phi(\cdot)} \right] \right\} X_i = 0 \\ &= \sum_{y_i=0} \left[ \frac{-\phi(X_i' \beta)}{1 - \Phi(X_i' \beta)} \right] X_i + \sum_{y_i=1} \left[ \frac{\phi(X_i' \beta)}{\Phi(X_i' \beta)} \right] X_i = 0 \dots (5) \end{aligned}$$

เมื่อ  $\phi_i$  คือ ฟังก์ชันความหนาแน่นของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ

$\Phi_i$  คือ ฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ

เมื่อประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุดในตัวแบบโพรบิตแบบ 2 ประเภทแล้ว จะสามารถนำไปใช้ในการพยากรณ์การจำแนกกลุ่มของตัวแบบ ดังนี้

- หน่วยที่  $i$  จะถูกจัดให้อยู่ในกลุ่มที่เกิดที่สนใจ ( $Y=1$ ) ถ้า

$$\hat{\pi}_i = \Phi(\beta' X_i) > c \quad ; 0 \leq c \leq 1$$

- หน่วยที่  $i$  จะถูกจัดให้อยู่ในกลุ่มที่ไม่เกิดเหตุการณ์ที่สนใจ ( $Y=0$ ) ถ้า

$$\hat{\pi}_i = \Phi(\beta' X_i) \leq c \quad ; 0 \leq c \leq 1$$

เมื่อ  $c$  คือ จุดแบ่ง หรือระดับของความน่าจะเป็นที่ใช้ในการพิจารณาการจำแนกกลุ่มว่าแต่ละหน่วยจะอยู่ในกลุ่มใดระหว่างกลุ่มที่เกิดเหตุการณ์ที่สนใจ และกลุ่มไม่เกิดเหตุการณ์ที่สนใจ

## 2.5 สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

การวิเคราะห์สหสัมพันธ์อย่างง่าย เป็นการวัดค่าความสัมพันธ์ของตัวแปร 2 ตัวแปร เพื่อแสดงว่า 2 ตัวนั้นมีความสัมพันธ์กันมากน้อยเพียงใด จะหาได้จากค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Coefficient of Simple Correlation:  $r$ ) ดังสูตรต่อไปนี้

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

โดยที่  $r$  คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร  $x, y$

$x$  คือ ตัวแปรที่ 1

$y$  คือ ตัวแปรที่ 2

$n$  คือ จำนวนตัวอย่าง

งานวิจัยนี้ผู้วิจัยคัดเลือกคู่ของตัวแปรอิสระเข้าในตัวแบบโดยใช้ค่า  $Max\{|r_{ij}|\}$  เป็นเกณฑ์ในการคัดเลือกระดับความสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) ซึ่งมีสูตรดังนี้

$$0 \leq Max\{|r_{ij}|\} = Max\left\{\frac{n \sum x_j x_j - (\sum x_j)(\sum x_j)}{\sqrt{n \sum x_j^2 - (\sum x_j)^2} \sqrt{n \sum x_j^2 - (\sum x_j)^2}}\right\} \leq 0.99$$

โดยที่  $r_{ij}$  คือ สหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่  $j$  และตัวแปรอิสระตัวที่  $j'$

$x_i$  คือ ตัวแปรที่ 1

$x_j$  คือ ตัวแปรที่ 2

$n$  คือ จำนวนตัวอย่าง

เกณฑ์สำหรับการแบ่งระดับความสัมพันธ์ระหว่างตัวแปรอิสระ แสดงดังนี้

ตัวแปรอิสระมีความสัมพันธ์กันในระดับอย่างต่ำ:  $0 \leq Max\{|r_{ij}|\} \leq 0.33$

ตัวแปรอิสระมีความสัมพันธ์กันในระดับปานกลาง:  $0.33 < Max\{|r_{ij}|\} \leq 0.66$

ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูง:  $0.66 < Max\{|r_{ij}|\} \leq 0.99$

## 2.6 ตัวแบบไม่ติดกลุ่ม (Non-Nested Model)

ตัวแบบพหุคูณแบบ 2 ประเภท เป็นตัวแบบไม่ติดกลุ่มไม่ว่าจะเป็นเพียงบางส่วนหรืออย่างสมบูรณ์ ถ้าหนึ่งตัวแบบไม่สามารถลดตัวแบบอื่น ๆ ได้โดยกำหนดการเชิงเส้นซึ่งมีข้อจำกัดบนเวกเตอร์พารามิเตอร์ เช่น

$$M_1 : \Phi^{-1}(E[Y_i | \tilde{X}_i^T = \tilde{x}_i^T]) = \Phi^{-1}(\pi_i) = \tilde{x}_i^T \tilde{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}; i = 1, 2, \dots, n$$

$$M_2 : \Phi^{-1}(E[Y_i | \tilde{X}_i^T = \tilde{x}_i^T]) = \Phi^{-1}(\pi_i) = \tilde{x}_i^T \tilde{\beta} = \beta_0 + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i}; i = 1, 2, \dots, n$$

ตัวแบบทั้งสองเป็นตัวแบบไม่ติดกลุ่มเพราะแม้ว่าเราจะกำหนดข้อจำกัดที่  $\beta_4 = 0, \beta_5 = 0$   $M_2$  จะไม่กลายเป็น  $M_1$  แต่จะเป็น  $M_2$  เป็นส่วนหนึ่งของตัวแบบ  $M_1$  ซึ่งจะเรียกตัวแบบดังกล่าวว่า “ตัวแบบไม่ติดกลุ่มบางส่วน (partially Non-Nested Model)” เพราะยังมีตัวแปรอิสระใน  $X_3$  ร่วมกัน

ถ้าในกรณีที่  $M_1$  และ  $M_2$  ไม่มีตัวแปร  $X_3$  ร่วมกัน จะได้ดังนี้

$$M_1 : \Phi^{-1}(E[Y_i | \tilde{X}_i^T = \tilde{x}_i^T]) = \Phi^{-1}(\pi_i) = \tilde{x}_i^T \tilde{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}; i = 1, 2, \dots, n$$

$$M_2 : \Phi^{-1}(E[Y_i | \tilde{X}_i^T = \tilde{x}_i^T]) = \Phi^{-1}(\pi_i) = \tilde{x}_i^T \tilde{\beta} = \beta_0 + \beta_4 x_{4i} + \beta_5 x_{5i}; i = 1, 2, \dots, n$$

ถ้าให้พารามิเตอร์บางตัวมีค่าเป็นศูนย์ จะเห็นว่าตัวแบบ  $M_2$  ไม่สามารถลดรูปไปเป็นตัวแบบ  $M_1$  ได้ ซึ่งจะเรียกตัวแบบนี้ว่าเป็นตัวแบบ “ไม่ติดกลุ่มอย่างสมบูรณ์ (Strictly Non-nested Model)”

## 2.7 การตรวจสอบความเหมาะสมของตัวแบบ

ตัวแบบที่เป็นไปได้มีหลายตัวแบบ ซึ่งมีวิธีการวัดทางสถิติที่หลากหลาย สำหรับใช้ในการวัดว่าตัวแบบพหุคูณแบบ 2 ประเภทมีความสามารถในการจำแนกกลุ่มให้แต่ละหน่วยอยู่ในกลุ่มใดกลุ่มหนึ่งระหว่างกลุ่มที่เกิดเหตุการณ์การที่สนใจ และกลุ่มที่ไม่เกิดเหตุการณ์ที่สนใจว่าเหมาะสมมากน้อยเพียงใด ซึ่ง มีวิธีต่างๆ ดังนี้

- Chi-square Goodness of Fit Test และสถิติ Deviance
- HosMer-LoMeshow Test
- Classification Table
- Receiver Operating Characteristic Curve (ROC)
- สัมประสิทธิ์การตัดสินใจ สำหรับการถดถอยพหุคูณ ( $R^2$ )
- ตรวจสอบความถูกต้องของตัวแบบ (Model validation) ด้วยวิธีการใช้ชุด

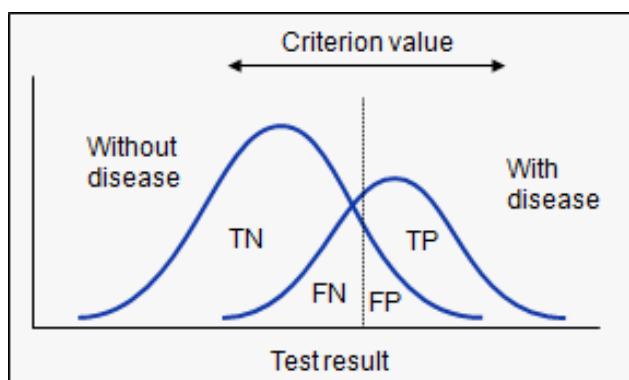
ข้อมูลภายนอกหรือโดยแบ่งชุดข้อมูลออกเป็น 2 ส่วน

Receiver Operating Characteristic Curve (ROC) จะสอดคล้องกับการคัดเลือกตัวแบบที่ถูกคัดเลือก ซึ่งจะถูกใช้เป็นเกณฑ์สำหรับการคัดเลือกตัวแบบโพรบิตแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างสมบูรณ์ดีที่สุด

## 2.8 Receiver Operating Characteristic curve หรือ ROC curve

Receiver Operating Characteristic curve หรือ ROC curve ถูกนำมาใช้ครั้งแรกในสมัยสงครามโลกครั้งที่สองสำหรับการตรวจจับและวิเคราะห์สัญญาณ ก่อนหน้าที่จะถูกนำมาใช้อย่างแพร่หลายในทฤษฎีการตรวจจับสัญญาณ (signal detection theory) และหลังจากในปี ค.ศ. 1950 เป็นต้นมา ROC curve ก็เริ่มเป็นที่รู้จักและได้รับการยอมรับอย่างกว้างขวาง โดยถูกนำมาประยุกต์ใช้กับศาสตร์ด้านต่างๆมากมาย เช่น ด้านจิตวิทยา ด้านการแพทย์ ด้านรังสีเอ็กซเรย์ รวมถึงด้านวิศวกรรม

การพยากรณ์ที่มีค่าพยากรณ์เพียง 2 ค่า ถือว่ามีความสำคัญอีกอย่างหนึ่งสำหรับการพยากรณ์และเป็นส่วนสำคัญในการสร้าง Receiver Operating Characteristic curve หรือ ROC Curve โดยตัวแปรตามตามคุณภาพแบ่งออกเป็น 2 กรณีคือ ตัวแปรตามมีค่าเท่ากับ 1 หรือผลการทดสอบเป็นบวก (Positive) เมื่อเกิดเหตุการณ์ที่สนใจในหน่วยที่  $i$  และตัวแปรตามมีค่าเท่ากับ 0 หรือผลการทดสอบเป็นลบ (Negative) เมื่อเกิดเหตุการณ์ที่ไม่สนใจในหน่วยที่  $i$  การพยากรณ์สามารถจำแนกประชากรออกเป็น 2 กลุ่ม โดยกลุ่มหนึ่งเป็นเหตุการณ์ที่สนใจ และอีกกลุ่มหนึ่งก็จะเป็นกลุ่มของเหตุการณ์ที่ไม่สนใจ



ภาพที่ 4.1 แสดงผลการพยากรณ์จำแนกประชากรออกเป็นกลุ่มเหตุการณ์ที่สนใจและกลุ่มเหตุการณ์ที่ไม่สนใจ

หากใช้จุดตัด ซึ่งเป็นตำแหน่งตรงเส้นตรงเป็นเกณฑ์ในการจำแนกเหตุการณ์ออกเป็นกลุ่มของเหตุการณ์ที่ไม่สนใจ และกลุ่มของเหตุการณ์ที่สนใจ พบว่า

- True positive (TP) คือ เหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นบวก ผลดังกล่าวเป็นผลบวกจริง
- False positive (FP) คือ เหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นบวก ผลดังกล่าวเป็นผลบวกลวง
- False negative (FN) คือ เหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นลบ ผลดังกล่าวเป็นผลลบลวง
- True negative (TN) คือ เหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นลบ ผลดังกล่าวเป็นผลลบจริง

ค่าสังเกต	ค่าพยากรณ์		
	จำนวนเหตุการณ์ที่ให้ผลบวก ( $y_i = 1$ )	จำนวนเหตุการณ์ที่ให้ผลลบ ( $y_i = 0$ )	รวม
จำนวนเหตุการณ์ที่สนใจ ( $\hat{y}_i = 1$ )	True Positive (TP)	False Positive (FP)	TP+FP
จำนวนเหตุการณ์ที่ไม่สนใจ ( $\hat{y}_i = 0$ )	False Negative (FN)	True Negative (TN)	FN+TN

ROC curve เป็นกราฟที่พล็อตระหว่างค่าของ Sensitivity หรือ True-Positive Rate ซึ่งเป็นอัตราส่วนของการพยากรณ์ถูกต้องของการเกิดเหตุการณ์ที่สนใจ ( $Y=1$ ) และค่า  $1-\text{Specificity}$  หรือ False-Positive Rate ซึ่งเป็นอัตราส่วนของการพยากรณ์ผิดของการไม่เกิดเหตุการณ์ที่สนใจ ( $Y=0$ ) ซึ่งการพล็อตกราฟจะได้จากการกำหนดจุดตัด ที่ระดับต่างๆ เพื่อแบ่งผลลัพธ์ของการพยากรณ์ออกเป็น 2 กลุ่ม คือกลุ่มที่เกิดเหตุการณ์  $P(Y = 1) \geq$  จุดตัด และกลุ่มที่ไม่เกิดเหตุการณ์  $P(Y = 1) <$  จุดตัด

- Sensitivity (SN) คือสัดส่วนของเหตุการณ์ที่ให้ผลการพยากรณ์เป็นผลบวกจริง (TP) ต่อจำนวนเหตุการณ์ที่สนใจทั้งหมด นั่นคือ

$$\text{Sensitivity or TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- Specificity (SP) คือสัดส่วนของเหตุการณ์ที่ให้ผลการพยากรณ์เป็นผลลบจริง (TN) ต่อจำนวนเหตุการณ์ที่ไม่สนใจทั้งหมด นั่นคือ

$$\text{Specificity or TNR} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

จะเห็นได้ว่า ค่า Sensitivity จะขึ้นอยู่กับจำนวนเหตุการณ์ที่สนใจทั้งหมด ส่วนค่า Specificity ก็ขึ้นอยู่กับจำนวนเหตุการณ์ที่ไม่สนใจทั้งหมด

ข้อได้เปรียบของการใช้ ROC curve ในการตรวจสอบความถูกต้องของตัวแบบอีกประการหนึ่ง นั่นคือ เป็นการง่ายมากที่เราสามารถมองเห็นการเปลี่ยนแปลงระหว่างค่า sensitivity และ Specificity สำหรับทุกๆจุดตัด ซึ่งส่งผลให้เห็นภาพรวมทั้งหมดของความถูกต้องหรือความเชื่อถือได้ของตัวแบบ รวมถึงสามารถคำนวณหาร้อยละของการพยากรณ์ที่ถูกต้องในแต่ละจุดตัดได้

## 2.9 พื้นที่ใต้โค้ง ROC (Area under the Curve หรือ AUC)

เนื่องจาก ROC curve มักถูกนำไปใช้ในการตรวจสอบความถูกต้องของตัวแบบในการพยากรณ์ โดยค่าที่ใช้เป็นดัชนีในการบ่งชี้ความถูกต้องหรือความเชื่อถือได้ของตัวแบบ ในการพยากรณ์ โดยค่าที่ใช้เป็นดัชนีในการบ่งชี้ความถูกต้องหรือความเชื่อถือได้ของตัวแบบ คือ ค่าประมาณพื้นที่ใต้โค้ง ROC (Area under the Curve หรือ AUC) ซึ่งในการใช้ค่าประมาณพื้นที่ใต้โค้ง ROC สำหรับบ่งชี้ความถูกต้องของตัวแบบนี้ อาศัยความคิดในการตีความได้หลายรูปแบบ

1. ค่าประมาณพื้นที่ใต้โค้ง ROC คือ ความน่าจะเป็นที่ตัวแบบสามารถสร้างค่าพยากรณ์ในกลุ่มเหตุการณ์ที่สนใจให้มีความมากกว่าค่าพยากรณ์ในกลุ่มเหตุการณ์ที่ไม่สนใจ ซึ่งนั่นส่งผลให้สามารถประเมินได้ว่าตัวแบบดังกล่าวมีความสามารถในการจำแนกระหว่างกลุ่มเหตุการณ์ที่สนใจกับเหตุการณ์ที่ไม่สนใจออกจากกันได้ดีเพียงใด

2. ค่าประมาณพื้นที่ใต้โค้ง ROC เป็นการรวมเครื่องมือพื้นฐานที่ใช้วัดความถูกต้องของการพยากรณ์ที่มีค่าพยากรณ์เพียง 2 ค่า ระหว่าง Sensitivity และ Specificity สำหรับทุกๆจุดตัดให้เป็นดัชนีบ่งชี้แค่เพียงค่าเดียว คือ ค่าเฉลี่ย Sensitivity สำหรับทุก Specificity หรือ ค่าเฉลี่ย Specificity สำหรับทุก Sensitivity

โดยในการคำนวณค่าพื้นที่ใต้โค้ง สามารถคำนวณได้หลายวิธี สำหรับในวานวิจัยนี้ขอกล่าวอ้างอิงระเบียบวิธีที่ไม่ใช้พารามิเตอร์ (Nonparametric method) ดังรายละเอียดต่อไปนี้

### ระเบียบวิธีที่ไม่ใช้พารามิเตอร์ (Nonparametric method)

ในการหาค่าประมาณพื้นที่ใต้โค้ง ROC โดยใช้ค่าที่ได้จริงจากการพยากรณ์ เป็นการสร้างโดยการเชื่อมต่อแต่ละจุดของข้อมูลซึ่งมีค่า อยู่ในแนวแกน และ อยู่ในแนวแกน เป็นเส้นตรง หลังจากนั้นทำการประมาณค่าพื้นที่ใต้โค้ง โดยอาศัยหลักเกณฑ์ดังนี้

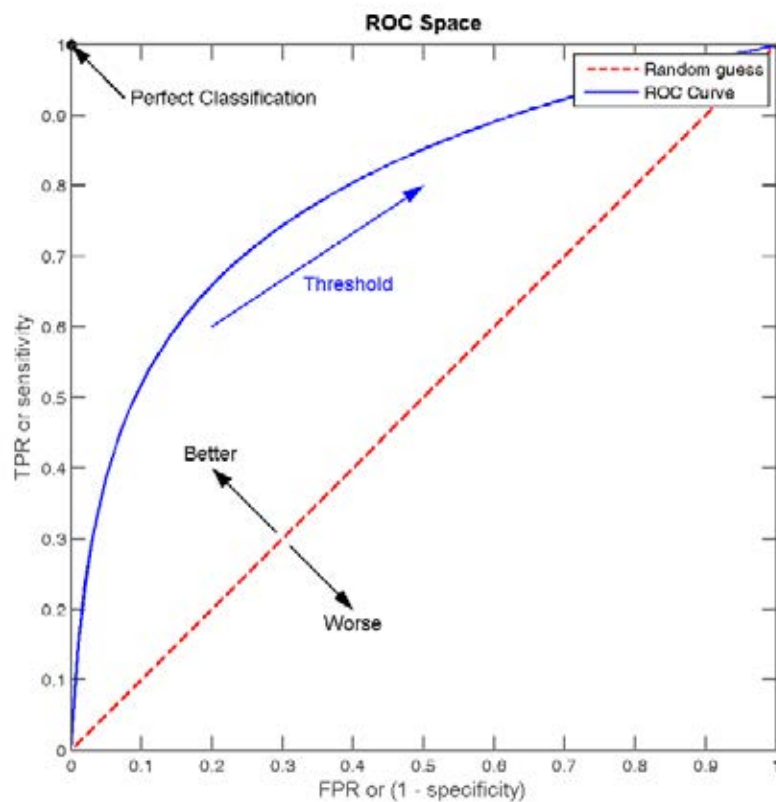
1. หาค่าพื้นที่ใต้โค้ง ROC โดยอาศัยหลักเกณฑ์เชิงสี่เหลี่ยมคางหมู(trapezoidal rule)
2. หาค่าพื้นที่ใต้โค้ง ROC โดยอาศัยหลักเกณฑ์ การทดสอบของแมนวิทนี่(Mann-Whitney U Test) ซึ่งเป็นการทดสอบโดยการคำนวณค่าจากจำนวนคู่ที่มีความเป็นไปได้ของเหตุการณ์ที่สนใจกับเหตุการณ์ที่ไม่สนใจ
3. หาค่าพื้นที่ใต้โค้ง ROC โดยอาศัยหลักเกณฑ์ การทดสอบผลรวมอันดับของวิลคอกสัน (Wilcoxon-rank-sum)

ทั้งสามวิธีข้างต้นล้วนเป็นระเบียบวิธีที่ไม่ใช้พารามิเตอร์ ซึ่งเป็นระเบียบวิธีที่ไม่จำเป็นต้องมีข้อสมมติเกี่ยวกับข้อมูล ในการคำนวณค่าสถิติที่ใช้ในการวิเคราะห์ก็ทำได้ง่าย ไม่ยุ่งยากจึงส่งผลให้ระเบียบวิธีเหล่านี้เป็นที่นิยมและถูกนำไปใช้ในงานด้านต่างๆอย่างกว้างขวาง ซึ่งในงานวิจัยนี้หาค่าพื้นที่ใต้โค้ง ROC โดยอาศัยหลักเกณฑ์เชิงสี่เหลี่ยมคางหมู (trapezoidal rule) เนื่องจากเป็นวิธีที่ง่ายและไม่ซับซ้อน

พื้นที่ใต้โค้ง ROC เป็นค่าที่บ่งบอกถึงความสามารถในการพยากรณ์ได้ถูกต้องหรือแสดงถึงความเชื่อถือได้ของตัวแบบ มีพิสัยอยู่ระหว่าง 0 ถึง 1 คือ ค่าประมาณพื้นที่ใต้โค้ง ROC (Area under the Curve หรือ  $A\hat{U}C$  ) ซึ่ง Hosmer, David W., and Stanley Lemeshow.(200) ได้กำหนดเกณฑ์ทั่วไปไว้ดังนี้

- ถ้า  $0.5 \leq A\hat{U}C < 0.6$  แล้ว ตัวแบบเป็นตัวแบบที่เชื่อถือไม่ได้
- ถ้า  $0.6 \leq A\hat{U}C < 0.7$  แล้ว ตัวแบบเป็นตัวแบบที่เชื่อถือได้น้อย
- ถ้า  $0.7 \leq A\hat{U}C < 0.8$  แล้ว ตัวแบบเป็นตัวแบบที่สามารถยอมรับได้
- ถ้า  $0.8 \leq A\hat{U}C < 0.9$  แล้ว ตัวแบบเป็นตัวแบบที่เชื่อถือได้ในระดับดี
- และถ้า  $A\hat{U}C \geq 0.9$  แล้ว ตัวแบบเป็นตัวแบบที่เชื่อถือได้ในระดับดีมาก





ภาพที่ 4.3 แสดงตัวอย่างเส้นโค้งและพื้นที่ใต้โค้ง

การทดสอบการพยากรณ์กับข้อมูลชุดเดียวกัน ตัวแบบใดมีเส้นโค้ง ROC อยู่ข้างบนหรือมีพื้นที่ใต้โค้ง ROC มากกว่า แสดงว่าเป็นตัวแบบที่มีประสิทธิภาพหรือให้ความถูกต้องในการพยากรณ์ได้ดีกว่า

## บทที่ 3

### วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อต้องการหาตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ที่เหมาะสมที่สุดสำหรับการพยากรณ์โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยงสำหรับแต่ละสถานการณ์ที่ต้องการศึกษา โดยตัวแบบที่เหมาะสมที่สุดจะให้ค่า ความน่าเชื่อถือของตัวแบบสูงสุด จากนั้นจะนำค่าความน่าเชื่อถือของตัวแบบมาทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุด

การจำลองข้อมูลในแต่ละสถานการณ์จะจำลองขึ้นด้วยการทำงานของเครื่องคอมพิวเตอร์ โดยใช้เทคนิคมอนติคาร์โล ด้วยโปรแกรม R เนื่องจากวิธีมอนติคาร์โลเป็นเทคนิคที่ใช้ในการวิจัยครั้งนี้ ดังนั้น ในตอนแรกของบทนี้จะกล่าวถึงวิธีมอนติคาร์โลก่อน แล้วจึงแสดงรายละเอียดของแผนการดำเนินการวิจัย ขั้นตอนในแผนการดำเนินการวิจัย ตลอดจนโปรแกรมที่ใช้ในการวิจัย ซึ่งมีรายละเอียดต่างๆดังนี้

#### 3.1 เทคนิคมอนติคาร์โล

เทคนิคมอนติคาร์โลเป็นการจำลองระบบที่ไม่เปลี่ยนแปลงตามเวลา ซึ่งตัวแบบของการจำลองจะมีลักษณะเป็นตัวแทนทางคณิตศาสตร์ โดยการนำตัวเลขสุ่ม มาประยุกต์ใช้ในการแก้ปัญหาหรือหาคำตอบให้กับระบบที่ยังไม่แน่ใจในผลที่จะเกิดขึ้น ซึ่งมีขั้นตอนที่สำคัญ 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การสร้างเลขสุ่ม (Generate Random Number) การสร้างเลขสุ่มจะกำหนดให้มีการแจกแจงแบบยูนิฟอร์มในช่วง  $[0, 1]$  และเป็นอิสระซึ่งกันและกัน จากนั้นนำเลขสุ่มนี้ไปสร้างตัวแปรตามลักษณะการแจกแจงที่ต้องการศึกษา เพื่อเป็น ข้อมูลของปัญหานั้น ๆ

ขั้นตอนที่ 2 การประยุกต์ใช้เลขสุ่มในการแก้ปัญหา ขั้นตอนนี้เป็นการนำตัวแปรที่ได้จากขั้นตอนแรกมาใช้ในการหาค่าต่างๆ ตามปัญหาที่ต้องการศึกษา

ขั้นตอนที่ 3 การทดลองซ้ำๆ ขั้นตอนนี้เป็นการทำวิธีนั้น ซ้ำๆ กัน (Replication) จำนวนหลายครั้งโดยถือว่าการทำซ้ำๆ กันนั้นเป็นวิธีการเก็บรวบรวมข้อมูลให้มีจำนวนมาก เพื่อลดความไม่แน่นอนของคำตอบในการวิเคราะห์หาค่าต่างๆ ได้

จากหลักการของเทคนิคมอนติคาร์โล จะเห็นว่าการใช้เลขสุ่มเพื่อเป็นพื้นฐานในการหาคำตอบของปัญหาเป็นวิธีที่จะนำไปสู่แนวคิดในทางทฤษฎีที่เกี่ยวข้องกับการคำนวณโดยเฉพาะ ทฤษฎีความน่าจะเป็นที่จะนำไปสู่การอ้างอิงผลสรุปในสถานการณ์ของข้อมูลจริงเพราะไม่มีผลกระทบจากเรื่องอื่นๆ เข้ามาเกี่ยวข้องในการทดลอง เมื่อทำซ้ำๆ กันเป็นจำนวนมากแล้ว ความคลาดเคลื่อนอย่างสุ่มที่เกิดขึ้นในการวิเคราะห์หาค่าต่างๆ ในแต่ละครั้งให้หมดไป

### 3.2 แผนการดำเนินการวิจัย

ในการวิจัยครั้งนี้ได้จำลองข้อมูลขึ้นโดยกำหนดสถานการณ์จำลองต่างๆ ดังนี้

1. กำหนดให้ข้อมูลของตัวแปรอิสระ (X) เริ่มต้นมีการแจกแจงแบบยูนิฟอร์ม
2. กำหนดให้ข้อมูลตัวแปรตาม (Y) ให้มีความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระและความผิดพลาด กำหนดค่าพารามิเตอร์เริ่มต้นของสมการถดถอยเป็นค่าใดๆในการศึกษาครั้งนี้ คือ  $\beta_i = 0.1 ; i = 0, 1, 2, \dots, p_1 + p_2$  และ  $\varepsilon_i \sim N(0, 1); i = 1, 2, \dots, n$

3. ตัวแปรตาม (Y) เป็นข้อมูลเชิงคุณภาพที่มี 2 ค่า คือ 0 และ 1

4. กำหนดให้จำนวนคู่ของตัวแปรอิสระ ( $p_1, p_2$ ) ในการวิจัยครั้งนี้ คือ

$$(p_1, p_2) = (2, 2), (2, 3), (2, 4), (2, 5),$$

$$(p_1, p_2) = (3, 2), (3, 3), (3, 4), (3, 5)$$

$$(p_1, p_2) = (4, 2), (4, 3), (4, 4), (4, 5)$$

$$(p_1, p_2) = (5, 2), (5, 3), (5, 4), (5, 5)$$

5. กำหนดให้ขนาดตัวอย่าง (n) ในการวิจัยครั้งนี้เป็น  $n = 50, 100, 150, 200, 250$

6. กำหนดให้ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) ในการวิจัยครั้งนี้โดยใช้ค่า  $Max\{|r_{jj}|\}$  นั่นคือ

$$0 \leq Max\{|r_{jj}|\} \leq 0.99$$

โดยที่  $r_{jj}$  คือ สหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่  $j$  และตัวแปรอิสระตัวที่  $j'$

เกณฑ์สำหรับการแบ่งระดับความสัมพันธ์ระหว่างตัวแปรอิสระ แสดงดังนี้

$$\text{ตัวแปรอิสระมีความสัมพันธ์กันในระดับอย่างต่ำ: } 0 \leq Max\{|r_{jj}|\} \leq 0.33$$

$$\text{ตัวแปรอิสระมีความสัมพันธ์กันในระดับปานกลาง: } 0.33 < Max\{|r_{jj}|\} \leq 0.66$$

$$\text{ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูง: } 0.66 < Max\{|r_{jj}|\} \leq 0.99$$

7. กำหนดระดับนัยสำคัญ ( $\alpha$ ) ในการวิจัยครั้งนี้ที่ระดับ 0.05

8. ในการวิจัยครั้งนี้ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โลโดยการจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 500 รอบ

9. ทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุด

### 3.3 การจำลองข้อมูลที่ใช้ในการวิจัย

1. สร้างตัวแปรอิสระ  $z_i \sim N(0,1); i = 1, \dots, p_1 + p_2$

โดยที่  $p_1$  คือ จำนวนตัวแปรอิสระของตัวแบบที่ 1

$p_2$  คือ จำนวนตัวแปรอิสระของตัวแบบที่ 2

2. ทำให้ตัวแปรอิสระในแต่ละตัวแบบมีการแจกแจงแบบ Multivariate normal โดยที่

$$Z_{\sim 1} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{p_1} \end{bmatrix} \sim N_{p_1} \left( \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}, \Sigma_1 \right) \quad \text{และ} \quad Z_{\sim 2} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{p_2} \end{bmatrix} \sim N_{p_2} \left( \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}, \Sigma_2 \right)$$

ซึ่งจะได้ตัวแปรอิสระที่มีความสัมพันธ์กันของแต่ละตัวแบบดังนี้

ตัวแบบ 1:  $z_1, \dots, z_{p_1}$

ตัวแบบ 2:  $z_{p_1+1}, \dots, z_{p_1+p_2}$

3. Transform ตัวแปรอิสระจากการแจกแจงแบบ  $z_i \sim N(0,1)$  เป็น  $U_i \sim U(0,1)$

4. แปลง  $U_i \sim U(0,1)$  เป็น  $x_i \sim U(-10,10)$

5. ตรวจสอบความสัมพันธ์ระหว่างตัวแปรอิสระที่แท้จริงโดยใช้ค่า  $\text{Max}\{|r_{jj'}|\}$  ดังนี้

$$0 \leq \text{Max}\{|r_{jj'}|\} \leq 0.99$$

โดยที่  $r_{jj'}$  คือ สหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่  $j$  และตัวแปรอิสระตัวที่  $j'$

เกณฑ์สำหรับการแบ่งระดับความสัมพันธ์ระหว่างตัวแปรอิสระ แสดงดังนี้

ตัวแปรอิสระมีความสัมพันธ์กันในระดับอย่างต่ำ:  $0 \leq \text{Max}\{|r_{jj'}|\} \leq 0.33$

ตัวแปรอิสระมีความสัมพันธ์กันในระดับปานกลาง:  $0.33 < \text{Max}\{|r_{jj'}|\} \leq 0.66$

ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูง:  $0.66 < \text{Max}\{|r_{jj'}|\} \leq 0.99$

6. จะได้ตัวแปรอิสระที่มีการแจกแจงแบบ  $x_i \sim U(-10,10)$  มีระดับความสัมพันธ์ระหว่าง

ตัวแปรอิสระของแต่ละแบบตามที่ได้ตรวจสอบระดับความสัมพันธ์ที่แท้จริงในข้อ 5 ซึ่ง

ตัวแปรอิสระของทั้งสองตัวแบบที่ได้เป็นดังนี้

ตัวแบบ 1:  $x_1, \dots, x_{p_1}$

ตัวแบบ 2:  $x_{p_1+1}, \dots, x_{p_1+p_2}$

7. สร้างตัวแปรตามโดยกำหนดให้ข้อมูลตัวแปรตาม ( $Y^*$ ) ให้มีความสัมพันธ์เชิงเส้นตรง

กับตัวแปรอิสระและความผิดพลาด กำหนดค่าพารามิเตอร์เริ่มต้นของสมการถดถอย

เป็นค่าใดๆในการศึกษาครั้งนี้ คือ  $\beta_i = 0.1; i = 0, 1, 2, \dots, p_1 + p_2$  และ

$\varepsilon_i \sim N(0, 1); i = 1, 2, \dots, n$  ซึ่งสมการตัวแปรตาม ( $Y^*$ ) เป็นดังนี้

$$Y_i^* = \tilde{X}_i^T \tilde{\beta} + \varepsilon_i; i = 1, 2, \dots, n$$

ทำการปรับ  $y_i^*$  ให้เป็นตัวแปรหุ่น (Dummy Variable) แทนด้วยค่า  $y_i$  ดังนี้

$$y_i = 1 \text{ ถ้า } y_i^* > 0$$

$$y_i = 0 \text{ ถ้า } y_i^* \leq 0$$

8. นำข้อมูลจากตัวแปรตามในข้อที่ 7 และตัวแปรอิสระมาทำการประมาณค่าพารามิเตอร์ ด้วยการวิเคราะห์ความถอยแบบโลจิสติกโดยใช้ฟังก์ชันโพรบิตเป็นการเชื่อมโยงเพื่อนำมาหาค่าพารามิเตอร์และสร้างตัวแบบที่ใช้ในการพยากรณ์

$$M_1 : \Phi^{-1}(E[Y_i | \tilde{X}_i^T = \tilde{x}_i^T]) = \Phi^{-1}(\pi_i)$$

$$= \tilde{x}_i^T \tilde{\beta}$$

$$= \beta_0 + \beta_1 x_{1i} + \dots + \beta_{p_1} x_{p_1 i}; i = 1, 2, \dots, n$$

$$M_2 : \Phi^{-1}(E[Y_i | \tilde{X}_i^T = \tilde{x}_i^T]) = \Phi^{-1}(\pi_i)$$

$$= \tilde{x}_i^T \tilde{\beta}$$

$$= \beta_0 + \beta_{p_1+1} x_{(p_1+1)i} + \dots + \beta_{p_1+p_2} x_{(p_1+p_2)i}; i = 1, 2, \dots, n$$

สร้างค่าพยากรณ์  $\hat{y}_i^*$  จากตัวแบบที่ได้ ดังนี้

$$\hat{y}_{1i}^* = b_0 + b_1 x_1 + \dots + b_{p_1} x_{p_1 i}$$

$$\hat{y}_{2i}^* = b_0 + b_{p_1+1} x_{p_1+1} + \dots + b_{p_1+p_2} x_{p_1+p_2}$$

### 3.4 ขั้นตอนในการดำเนินการวิจัย

การดำเนินการวิจัยมีขั้นตอนดังนี้

1. ศึกษาตัวแบบโพรบิตและวิธีการหาค่าพื้นที่ใต้โค้ง ROC
2. จำลองข้อมูลตามลักษณะที่กำหนดเพื่อใช้ในการวิจัย
3. หาค่า  $y^*$  จากการสร้างความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระ
4. ทำการแปลงค่าตัวแปรตาม  $y^*$  ที่ได้เป็น  $y$  ที่มีค่าเป็น 0 หรือ 1 ตามสัดส่วนของการไม่เกิดเหตุการณ์ที่สนใจและขนาดตัวอย่างตามที่กำหนดไว้
5. ประมาณค่าพารามิเตอร์ของตัวแบบโพรบิตแบบ 2 ประเภทโดยใช้วิธีภาวะน่าจะเป็นสูงสุดในกรณีที่ข้อมูลมีลักษณะที่ต้องการศึกษา

6. คำนวณหาพื้นที่ใต้โค้ง ROC ที่สูงที่สุดในแต่ละสถานการณ์สำหรับตัวแบบโพรบิตแบบ 2 ประเภท ซึ่งทำให้ค่าความน่าเชื่อถือของตัวแบบมีค่าสูงสุด
7. ทำการทดลองซ้ำ 500 รอบในแต่ละสถานการณ์
9. ทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุด
10. สรุปผลการวิจัยในแต่ละสถานการณ์

### 3.5 การคำนวณหาพื้นที่ใต้โค้ง ROC

พื้นที่ใต้โค้ง ROC เป็นค่าที่บ่งบอกถึงความสามารถในการพยากรณ์ได้ถูกต้องหรือแสดงถึงความเชื่อถือได้ของตัวแบบ

ตาราง : Binary Probit Regression Classification Table

ค่าสังเกต	ค่าพยากรณ์		
	จำนวนเหตุการณ์ที่ให้ผลบวก ( $y_i = 1$ )	จำนวนเหตุการณ์ที่ให้ผลลบ ( $y_i = 0$ )	รวม
จำนวนเหตุการณ์ที่สนใจ ( $\hat{y}_i = 1$ )	True Positive (TP) (1,1) นับเป็น 1	False Positive (FP) (1,0) นับเป็น 1	TP+FP
จำนวนเหตุการณ์ที่ไม่สนใจ ( $\hat{y}_i = 0$ )	False Negative (FN) (0,1) นับเป็น 1	True Negative (TN) (0,0) นับเป็น 1	FN+TN

จะได้ค่า  $y$  ที่เป็นเฉพาะค่า 0 กับ 1 ซึ่งส่งผลให้ได้ค่าอื่นๆดังนี้

TP = ผลรวมของจำนวนที่นับเป็น 1 ของช่อง (1, 1)

FP = ผลรวมของจำนวนที่นับเป็น 1 ของช่อง (1, 0)

FN = ผลรวมของจำนวนที่นับเป็น 1 ของช่อง (0, 1)

TN = ผลรวมของจำนวนที่นับเป็น 1 ของช่อง (0, 0)

$$SN = \frac{TP}{TP+FN}$$

$$SP = \frac{TN}{FP+TN}$$

นำ SN กับ 1-SP ไป plot graph ROC curve

หาพื้นที่ใต้โค้ง ROC curve หรือ ค่าความน่าเชื่อถือของตัวแบบ

### 3.6 ทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุด

ทดสอบสมมติฐานเพื่อหาตัวแบบที่เหมาะสมที่สุด โดยการเปรียบเทียบพื้นที่ใต้โค้ง ROC ในแต่ละตัวแบบ

สมมติฐานที่ใช้ในการทดสอบ

$$H_0 : \mu_{\text{model1}} \geq \mu_{\text{model2}}$$

$$H_a : \mu_{\text{model1}} < \mu_{\text{model2}}$$

ตัวสถิติทดสอบ คือ

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

เกณฑ์ในการตัดสินใจ

ถ้า  $t < t_{0.05, df=498}$  ( $t_{0.05, df=498} = -1.645$ ) จะปฏิเสธ  $H_0$  นั่นคือ ค่าเฉลี่ยของพื้นที่โดยใช้พื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของพื้นที่โดยใช้พื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2

หรือ

สมมติฐานที่ใช้ในการทดสอบ

$$H_0 : \mu_{\text{model1}} \leq \mu_{\text{model2}}$$

$$H_a : \mu_{\text{model1}} > \mu_{\text{model2}}$$

เกณฑ์ในการตัดสินใจ

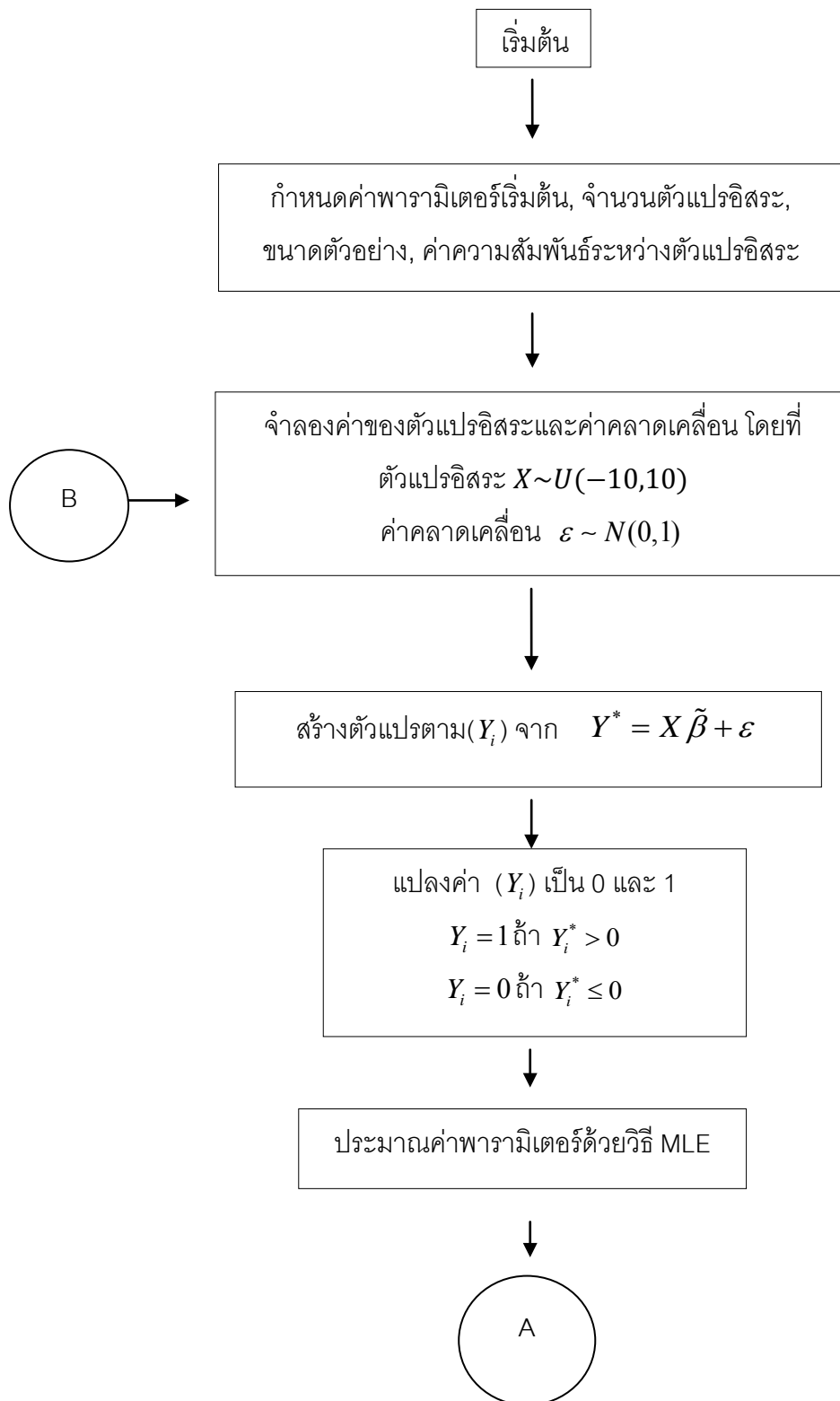
ถ้า  $t > t_{0.95, df=498}$  ( $t_{0.95, df=498} = 1.645$ ) จะปฏิเสธ  $H_0$  นั่นคือ ค่าเฉลี่ยของพื้นที่โดยใช้พื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของพื้นที่โดยใช้พื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2

หมายเหตุ : งานวิจัยนี้ใช้สถิติทดสอบ t-test เนื่องจาก  $n_1 = n_2 = 500$  ซึ่งทำให้  $t\text{-distribution} \approx z$

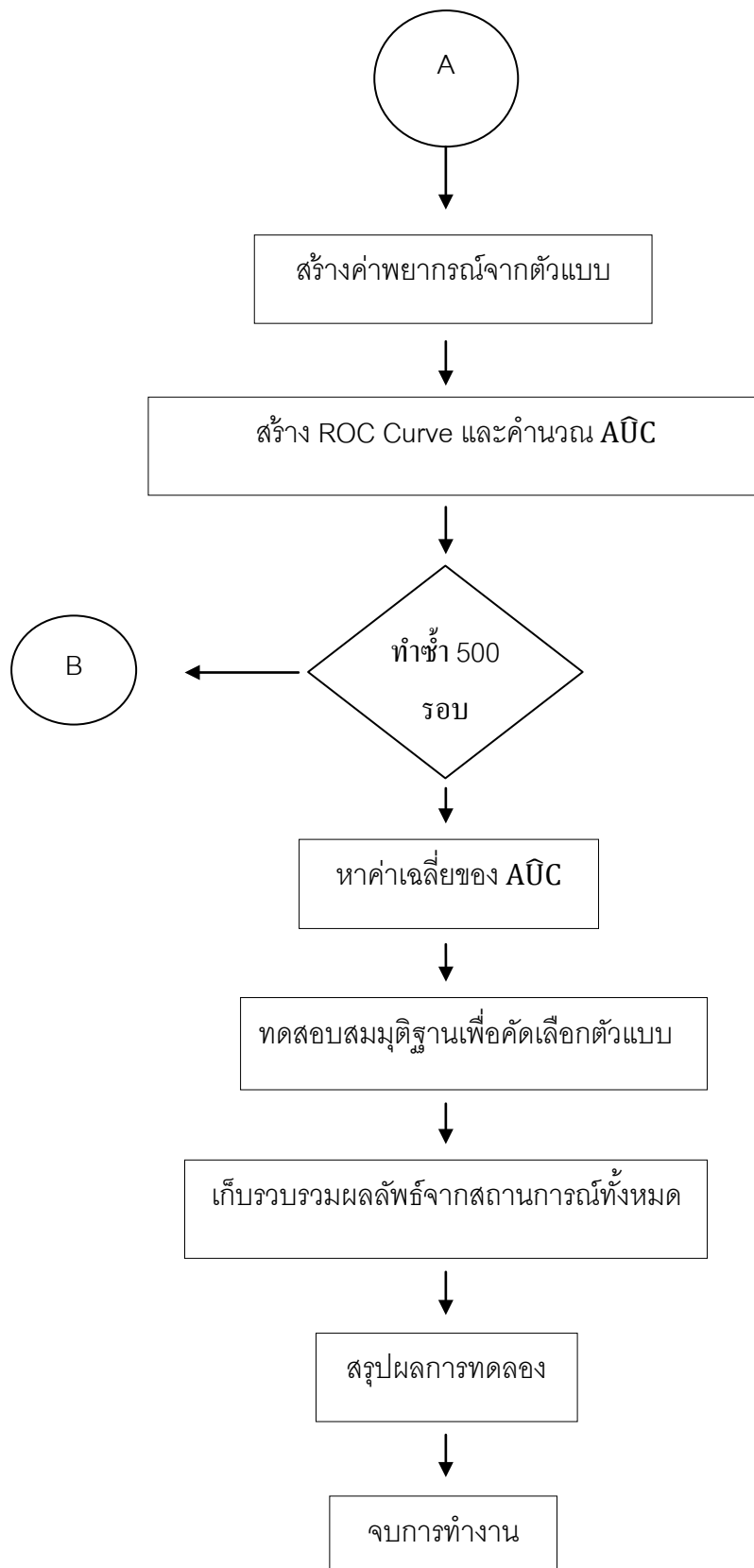
### 3.7 สรุปผลการวิจัยในแต่ละสถานการณ์

เมื่อทำการหาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ ครบทุกสถานการณ์ที่ต้องการศึกษาแล้ว นำผลการทดลองมาสรุปในภาพตาราง เพื่อดูแนวโน้มว่าปัจจัยที่ต้องการศึกษามีผลต่อค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ อย่างไรในแต่ละสถานการณ์

## 3.8 ขั้นตอนการทำงานของโปรแกรม







ภาพที่ 3.1 แสดงขั้นตอนการทำงานของโปรแกรม

## บทที่ 4

### ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ที่เหมาะสมที่สุดสำหรับการพยากรณ์โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง เมื่อตัวแปรอิสระเริ่มต้นมีการแจกแจงแบบยูนิฟอร์ม, จำนวนคู่ของตัวแปรอิสระเปลี่ยนแปลงไป, ระดับความสัมพันธ์ของตัวแปรอิสระเปลี่ยนแปลงไป และขนาดตัวอย่างเปลี่ยนแปลงไป เมื่อนำปัจจัยต่างๆเหล่านี้มาพิจารณารวมกัน ซึ่งแปรเปลี่ยนไปพร้อมกัน โดยศึกษาภายใต้สถานการณ์ต่อไปนี้

1. ข้อมูลที่กำหนดให้คู่ลำดับของจำนวนตัวแปรในตัวแบบแรก( $p_1$ ) และจำนวนตัวแปรในตัวแบบที่สอง( $p_2$ ) ; ( $p_1, p_2$ ) เปลี่ยนแปลงไปเป็น (2, 2), (2, 3), (2, 4), (2, 5), (3, 2), (3, 3), (3, 4), (3, 5), (4, 2), (4, 3), (4, 4), (4, 5), (5, 2), (5, 3), (5, 4), (5, 5) ตามลำดับ แต่ ระดับความสัมพันธ์ของตัวแปรอิสระ และ ขนาดตัวอย่างคงที่
2. ข้อมูลที่กำหนดให้ระดับความสัมพันธ์ของตัวแปรอิสระเปลี่ยนแปลงไปเป็น ความสัมพันธ์ระดับต่ำ, ระดับกลาง และ ระดับสูง ดังนี้  
ตัวแปรอิสระมีความสัมพันธ์กันในระดับอย่างต่ำ:  $0 \leq \text{Max}\{|r_{ij}|\} \leq 0.33$   
ตัวแปรอิสระมีความสัมพันธ์กันในระดับปานกลาง:  $0.33 < \text{Max}\{|r_{ij}|\} \leq 0.66$   
ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูง:  $0.66 < \text{Max}\{|r_{ij}|\} \leq 0.99$
3. ข้อมูลที่กำหนดให้ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น 50, 100, 150, 200 และ 250 ตามลำดับ แต่จำนวนตัวแปรอิสระ และ ระดับความสัมพันธ์ของตัวแปรอิสระคงที่
4. กำหนดระดับนัยสำคัญ ( $\alpha$ ) ในการวิจัยครั้งนี้ที่ระดับ 0.05
5. ในการวิจัยครั้งนี้ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล โดยการจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 500 รอบ

โดยการหาค่าความน่าเชื่อถือของตัวแบบในตัวแบบโพรบิต 2 ประเภทในแต่ละสถานการณ์ จะพิจารณาจากค่าพื้นที่ใต้โค้ง ROC สูงที่สุดในแต่ละรอบที่ต้องการศึกษา และทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบ ในแต่ละสถานการณ์ เพื่อให้ได้ตัวแบบที่เหมาะสมของสถานการณ์นั้นๆ

การวิจัยครั้งนี้จะนำเสนอผลการวิเคราะห์ในแต่ละสถานการณ์ที่ต้องการศึกษา ตามวัตถุประสงค์หลักของการวิจัย ดังนี้

- 4.1 จำนวนตัวแปรอิสระที่ส่งผลให้การเลือกตัวแบบโพรบิตแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างสมบูรณ์ดีที่สุด
- 4.2 ระดับความสัมพันธ์ในแต่ละคู่ (degree of pair-wise correlation) ของตัวแปรอิสระจากทั้งสองตัวแบบที่ส่งผลให้การเลือกตัวแบบโพรบิตแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างสมบูรณ์ดีที่สุด

โดยจะนำเสนอผลการวิจัยในภาพแบบตาราง เพื่อความสะดวกในการอธิบาย ผู้วิจัยขอใช้สัญลักษณ์ต่อไปนี้แทนความหมายต่างๆ ดังนี้

$p_1$  หมายถึง จำนวนตัวแปรในตัวแบบแรก

$p_2$  หมายถึง จำนวนตัวแปรในตัวแบบที่สอง

$n$  หมายถึง ขนาดตัวอย่าง

$\text{Max}\{|r_{ij}|\}$  หมายถึง ระดับความสัมพันธ์ของตัวแปรอิสระ

$\overline{AUC}_1$  หมายถึง ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ในตัวแบบแรก

$\overline{AUC}_2$  หมายถึง ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ในตัวแบบที่สอง

Selected Model หมายถึง ตัวแบบที่ถูกเลือก

$\mu_{model1}$  หมายถึง ค่าเฉลี่ย  $AUC$  ในตัวแบบแรก

$\mu_{model2}$  หมายถึง ค่าเฉลี่ย  $AUC$  ในตัวแบบที่สอง

4.1 จำนวนตัวแปรอิสระที่ส่งผลให้การเลือกตัวแบบโพรบิตแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างสมบูรณ์ดีที่สุด

คัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ที่เหมาะสมที่สุด โดยใช้พื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือก เมื่อนำปัจจัยมาพิจารณา คือ คู่ลำดับของจำนวนตัวแปรอิสระ ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ และ ขนาดตัวอย่าง ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จำลอง

ขึ้นมาหลายสถานการณ์ด้วยกันโดยการจำลองสถานการณ์ทั้งหมดใช้เทคนิคมอนติคาร์โล ด้วยโปรแกรม R โดยให้ตัวแปรอิสระเริ่มต้นมีการแจกแจงแบบยูนิฟอร์ม คู่ลำดับของจำนวนตัวแปรในตัวแบบแรก( $p_1$ ) และจำนวนตัวแปรในตัวแบบที่สอง ( $p_2$ ) ; ( $p_1, p_2$ ) มีทั้งหมด 16 คู่ ระดับความสัมพันธ์ของตัวแปรอิสระแบ่งเป็น 3 ระดับ คือ ระดับต่ำ ระดับกลาง และระดับสูง ขนาดตัวอย่าง( $n$ ) เป็น 50, 100, 150, 200 และ 250 การจำลองข้อมูลจะกระทำซ้ำ 500 ครั้งในแต่ละสถานการณ์

ซึ่งผลการวิจัยได้นำเสนอในตารางที่ 4.1 - 4.16 ดังต่อไปนี้

#### 4.1.1 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ ดังนี้

เมื่อคู่ลำดับของจำนวนตัวแปรในตัวแบบแรก( $p_1$ ) และจำนวนตัวแปรในตัวแบบที่สอง( $p_2$ ) ; ( $p_1, p_2$ ) เปลี่ยนแปลงไปเป็น (2, 2), (2, 3), (2, 4), (2, 5), (3, 2), (3, 3), (3, 4), (3, 5), (4, 2), (4, 3), (4, 4), (4, 5), (5, 2), (5, 3), (5, 4), (5, 5) ตามลำดับ แต่ ระดับความสัมพันธ์ของตัวแปรอิสระ และ ขนาดตัวอย่างคงที่

เมื่อระดับความสัมพันธ์ของตัวแปรอิสระเปลี่ยนแปลงไปเป็น ความสัมพันธ์ระดับต่ำ, ระดับกลาง และ ระดับสูง แต่ คู่ลำดับของจำนวนตัวแปรในตัวแบบแรก( $p_1$ ) แต่ จำนวนตัวแปรในตัวแบบที่สอง ( $p_2$ ); ( $p_1, p_2$ ) และ ขนาดตัวอย่างคงที่

เมื่อขนาดตัวอย่างเปลี่ยนแปลงไปเป็น 50, 100, 150, 200 และ 250 แต่คู่ลำดับของจำนวนตัวแปรในตัวแบบแรก ( $p_1$ ) แต่จำนวนตัวแปรในตัวแบบที่สอง ( $p_2$ ); ( $p_1, p_2$ ) และ ระดับความสัมพันธ์ของตัวแปรอิสระคงที่

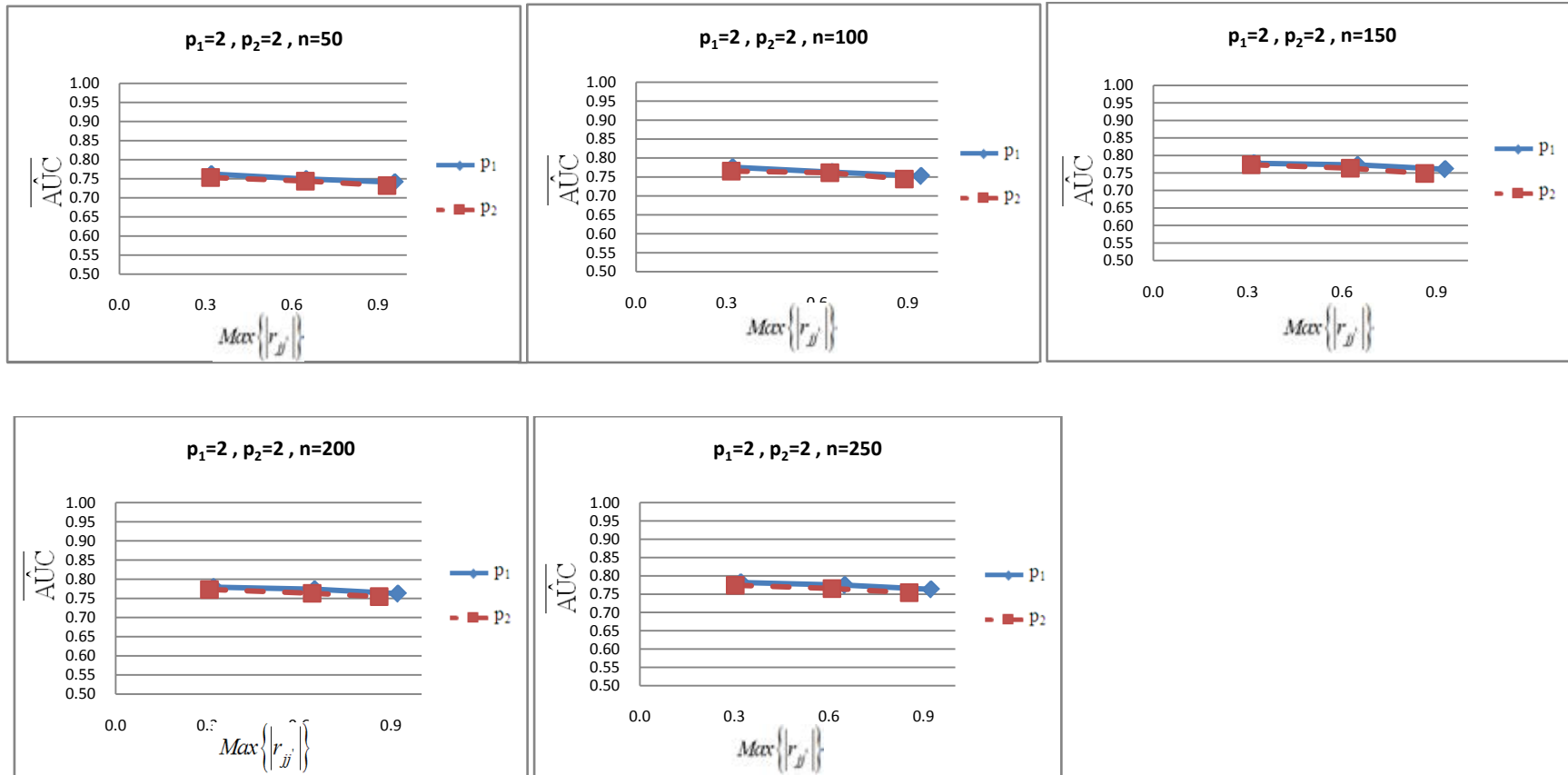
4.1.1 เมื่อจำนวนขนาดตัวอย่างและระดับความสัมพันธ์ของตัวแปรอิสระเปลี่ยนแปลงไป นำเสนอในภาพตารางที่ 4.1-4.16 ดังนี้

**ตารางที่ 4.1** แสดง ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=2$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

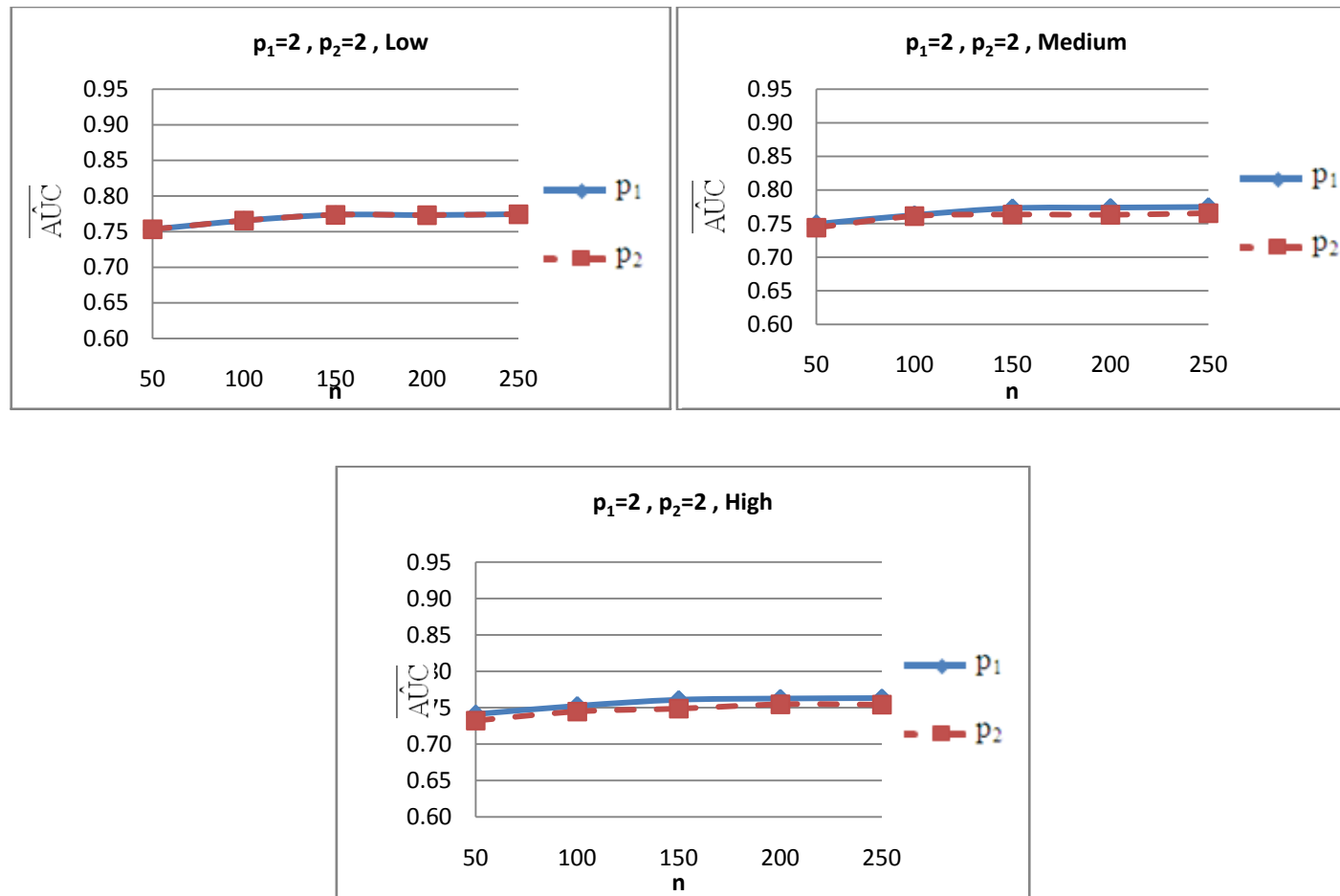
$p_1$	$p_2$	n	$Max\{r_{ij}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
2	2	50	Low	0.7623	0.7533	Model 1
			Medium	0.7494	0.7440	Model 1
			High	0.7410	0.7322	Model 1
		100	Low	0.7756	0.7655	Model 1
			Medium	0.7628	0.7612	Model 1 หรือ Model 2
			High	0.7525	0.7446	Model 1
		150	Low	0.7776	0.7736	Model 1 หรือ Model 2
			Medium	0.7730	0.7636	Model 1
			High	0.7607	0.7487	Model 1
		200	Low	0.7794	0.7731	Model 1
			Medium	0.7739	0.7633	Model 1
			High	0.7624	0.7546	Model 1
		250	Low	0.7820	0.7744	Model 1
			Medium	0.7748	0.7655	Model 1
			High	0.7631	0.7541	Model 1

จากตารางที่ 4.1 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=2$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0 : \mu_{model1} \leq \mu_{model2}$  แย้งกับ  $H_0 : \mu_{model1} > \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ส่วนใหญ่ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกกรณี ซึ่งตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ยกเว้นกรณี  $n=100$ ,  $150$  ค่า  $Max\{r_{ij}\}$  อยู่ในช่วง Medium, Low ตามลำดับ ซึ่งตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 หรือ ตัวแบบที่ 2 (p-value<0.05)

ภาพที่ 4.1 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=2, p_2=2, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.2 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=2, p_2=2, \text{Max}\{r_{ij}\}=\text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n=50, 100, 150, 200, 250$



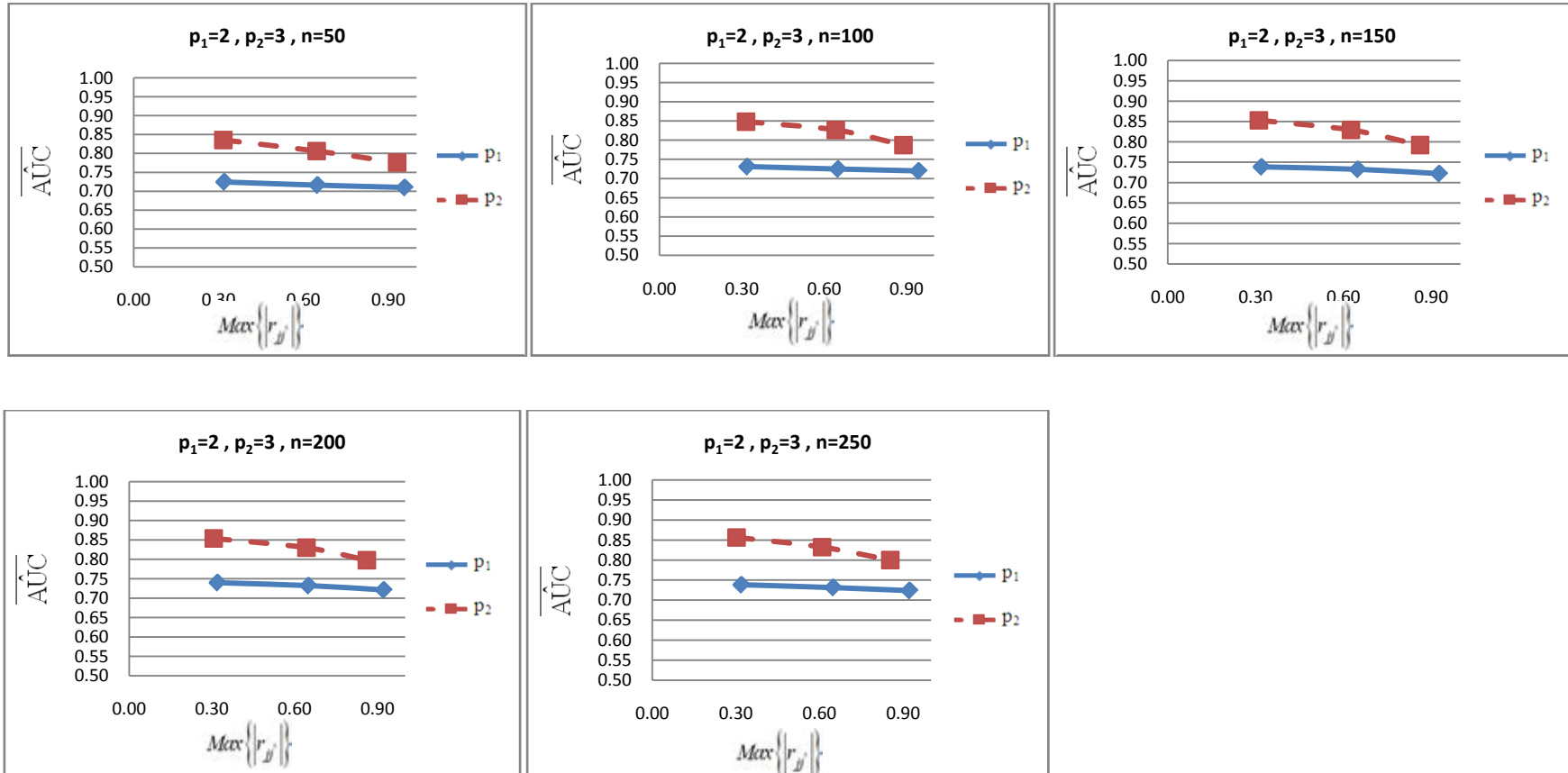
ตารางที่ 4.2 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=2$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

$p_1$	$p_2$	$n$	$Max\{r_{ij}\}$	$\overline{A\hat{U}C}_1$	$\overline{A\hat{U}C}_2$	Selected Model
2	3	50	Low	0.7245	0.8355	Model 2
			Medium	0.7162	0.8063	Model 2
			High	0.7099	0.7759	Model 2
		100	Low	0.7311	0.8480	Model 2
			Medium	0.7247	0.8265	Model 2
			High	0.7197	0.7870	Model 2
		150	Low	0.7387	0.8532	Model 2
			Medium	0.7326	0.8297	Model 2
			High	0.7221	0.7924	Model 2
		200	Low	0.7399	0.8539	Model 2
			Medium	0.7323	0.8302	Model 2
			High	0.7211	0.7979	Model 2
		250	Low	0.7384	0.8565	Model 2
			Medium	0.7312	0.8324	Model 2
			High	0.7238	0.8002	Model 2

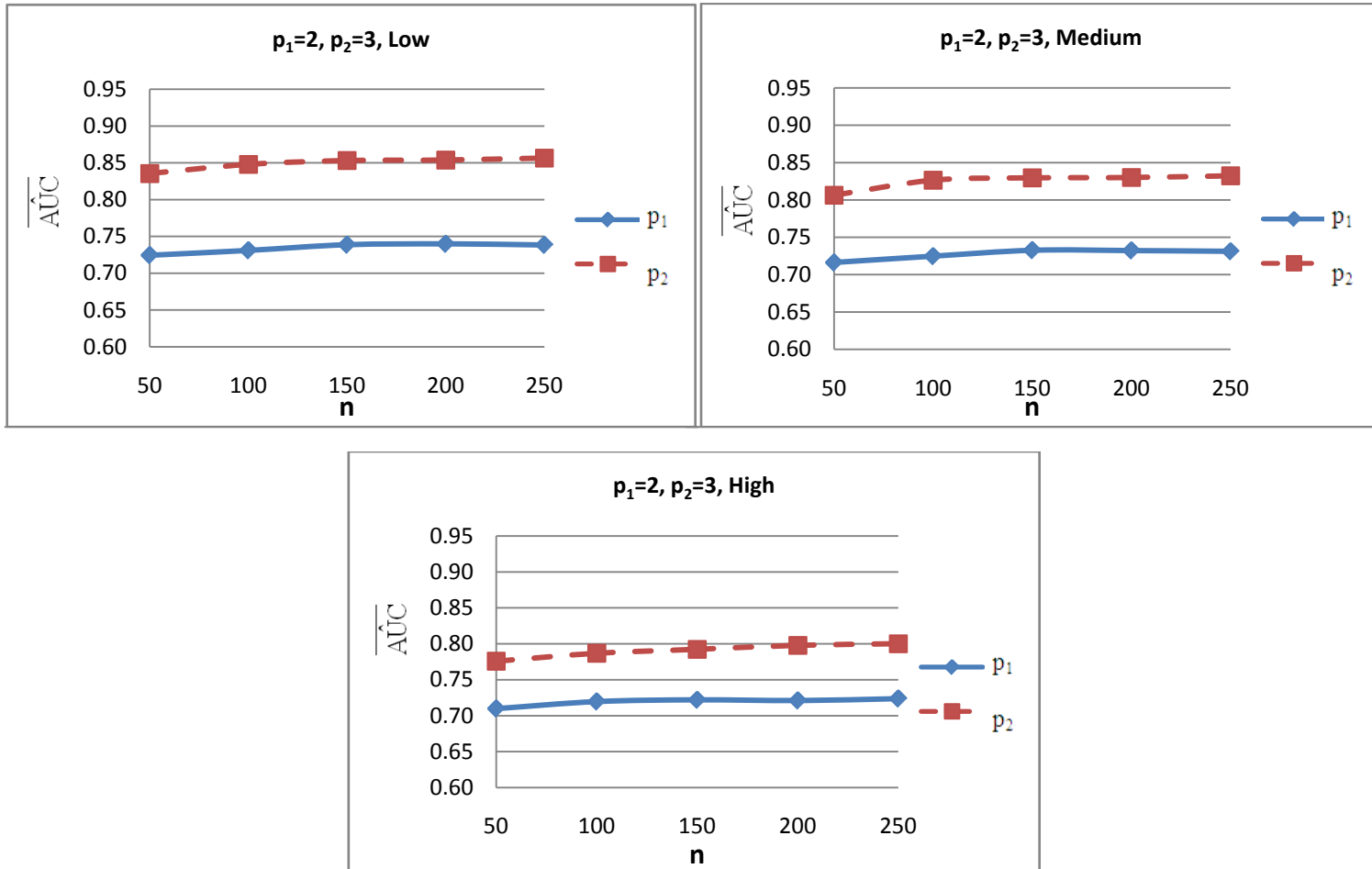
จากตารางที่ 4.2 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=2$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0 : \mu_{model1} \geq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} < \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 มากกว่าตัวแบบที่ 1 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2 ( $p\text{-value} < 0.05$ )



ภาพที่ 4.3 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=2, p_2=3, n=50, n=100, n=150, n=200, n=250$  และ  $\text{Max}\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.4 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=2, p_2=3, \text{Max}\{r_{jj}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n = 50, 100, 150, 200, 250$

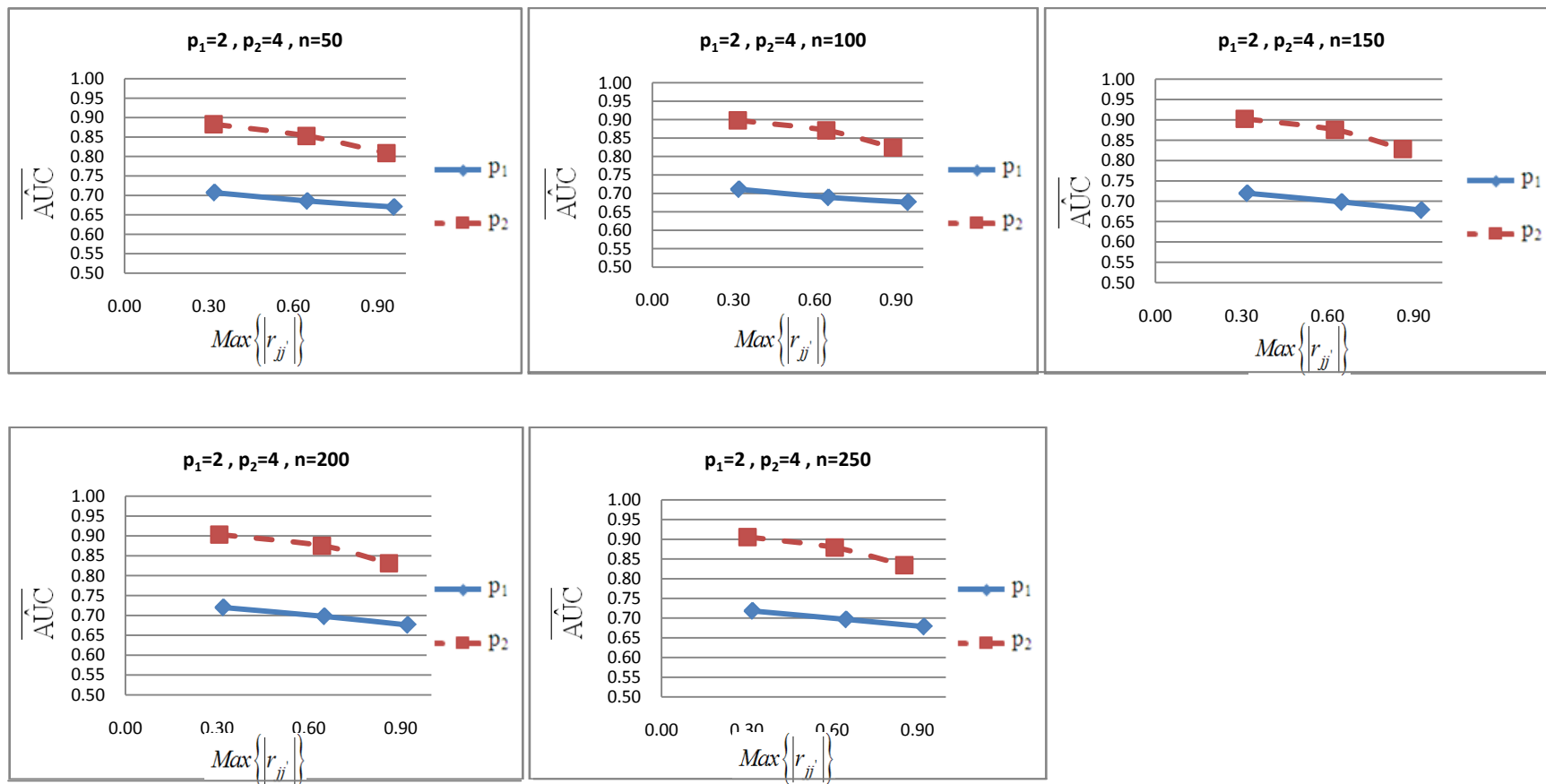


ตารางที่ 4.3 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=2$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

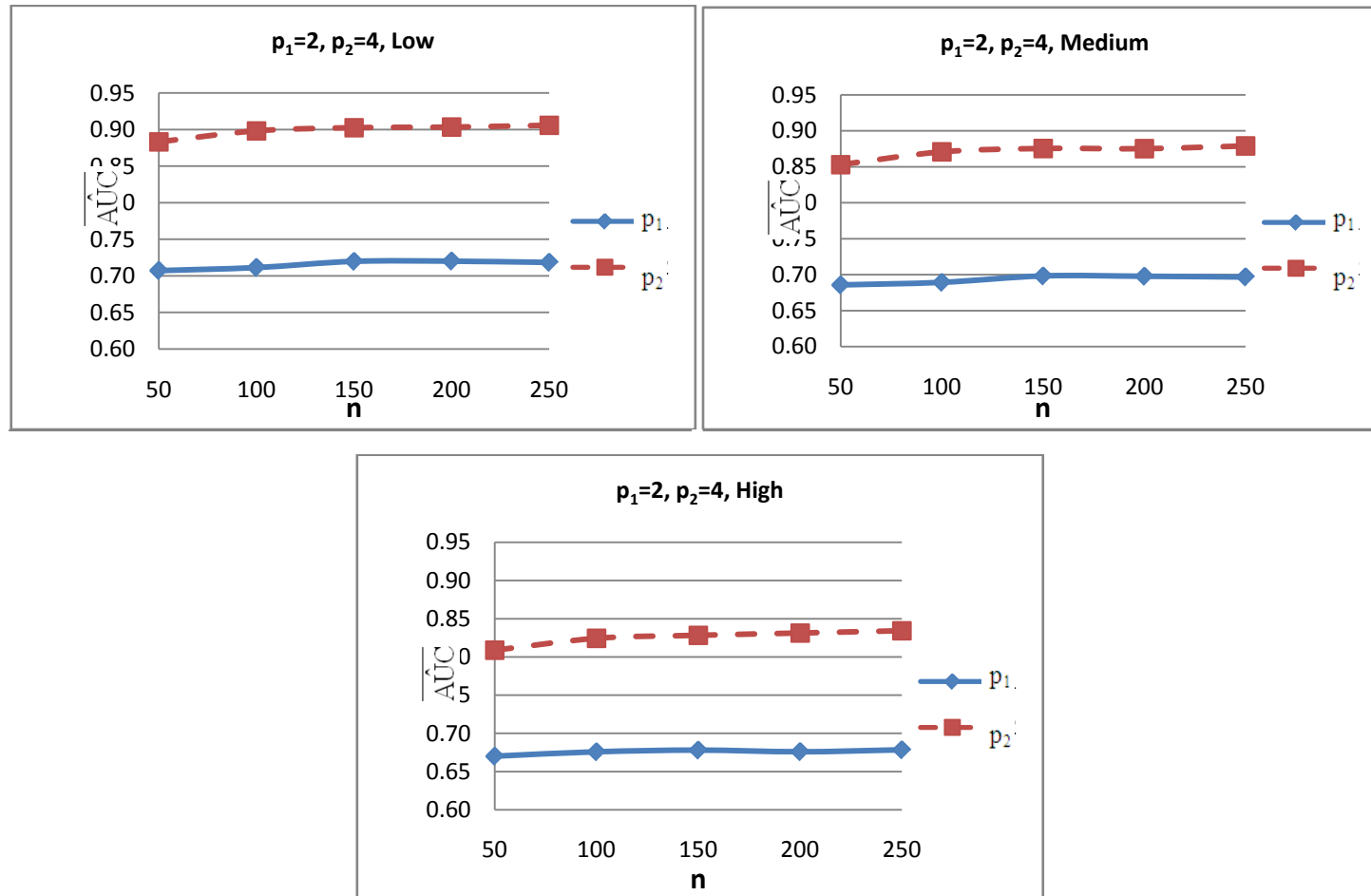
$p_1$	$p_2$	n	$Max\{r_{jj}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
2	4	50	Low	0.7072	0.8832	Model 2
			Medium	0.6857	0.8530	Model 2
			High	0.6702	0.8089	Model 2
		100	Low	0.7114	0.8982	Model 2
			Medium	0.6893	0.8709	Model 2
			High	0.6759	0.8245	Model 2
		150	Low	0.7198	0.9025	Model 2
			Medium	0.6982	0.8756	Model 2
			High	0.6782	0.8283	Model 2
		200	Low	0.7200	0.9035	Model 2
			Medium	0.6977	0.8753	Model 2
			High	0.6760	0.8313	Model 2
		250	Low	0.7184	0.9059	Model 2
			Medium	0.6968	0.8792	Model 2
			High	0.6785	0.8343	Model 2

จากตารางที่ 4.3 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=2$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0 : \mu_{model1} \leq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} < \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 มากกว่าตัวแบบที่ 1 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.5 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=2, p_2=4, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.6 แสดงค่าของ  $\overline{\widehat{AUC}}$  เมื่อ  $p_1=2, p_2=4, \text{Max}\{r_{jj}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n=50, 100, 150, 200, 250$

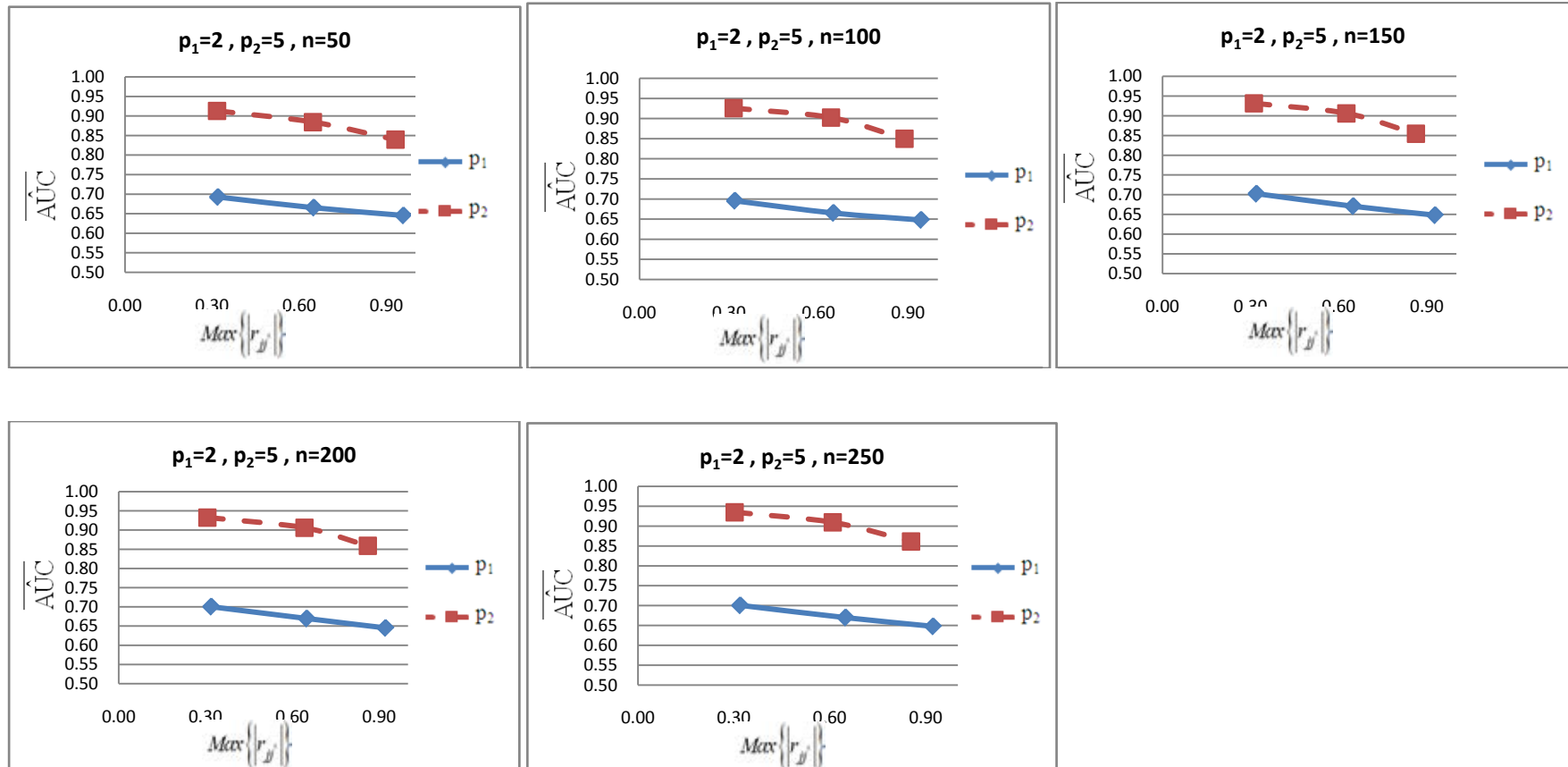


ตารางที่ 4.4 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=2$ ,  $p_2=5$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

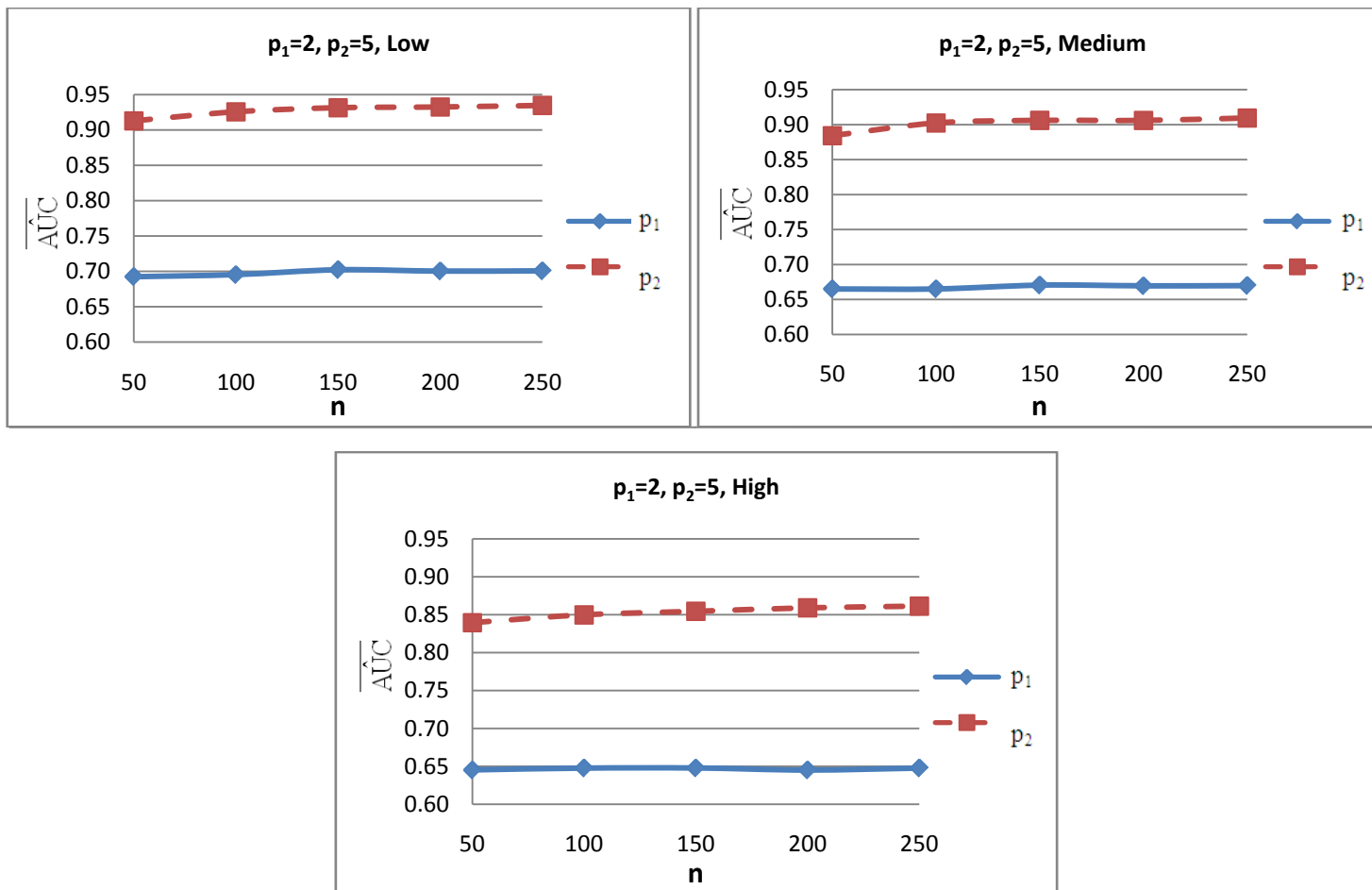
$p_1$	$p_2$	$n$	$Max\{r_{ij}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
2	5	50	Low	0.6925	0.9129	Model 2
			Medium	0.6652	0.8842	Model 2
			High	0.6453	0.8395	Model 2
		100	Low	0.6954	0.9258	Model 2
			Medium	0.6651	0.9024	Model 2
			High	0.6477	0.8499	Model 2
		150	Low	0.7023	0.9316	Model 2
			Medium	0.6706	0.9060	Model 2
			High	0.6478	0.8545	Model 2
		200	Low	0.7004	0.9326	Model 2
			Medium	0.6696	0.9060	Model 2
			High	0.6451	0.8591	Model 2
		250	Low	0.7007	0.9349	Model 2
			Medium	0.6699	0.9095	Model 2
			High	0.6477	0.8613	Model 2

จากตารางที่ 4.4 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=2$ ,  $p_2=5$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0 : \mu_{model1} \geq \mu_{model2}$  แย้งกับ  $H_0 : \mu_{model1} < \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 มากกว่าตัวแบบที่ 1 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.7 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $P_1=2, P_2=5, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.8 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=2$ ,  $p_2=5$ ,  $\text{Max}\{r_{ij}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n=50, 100, 150, 200, 250$



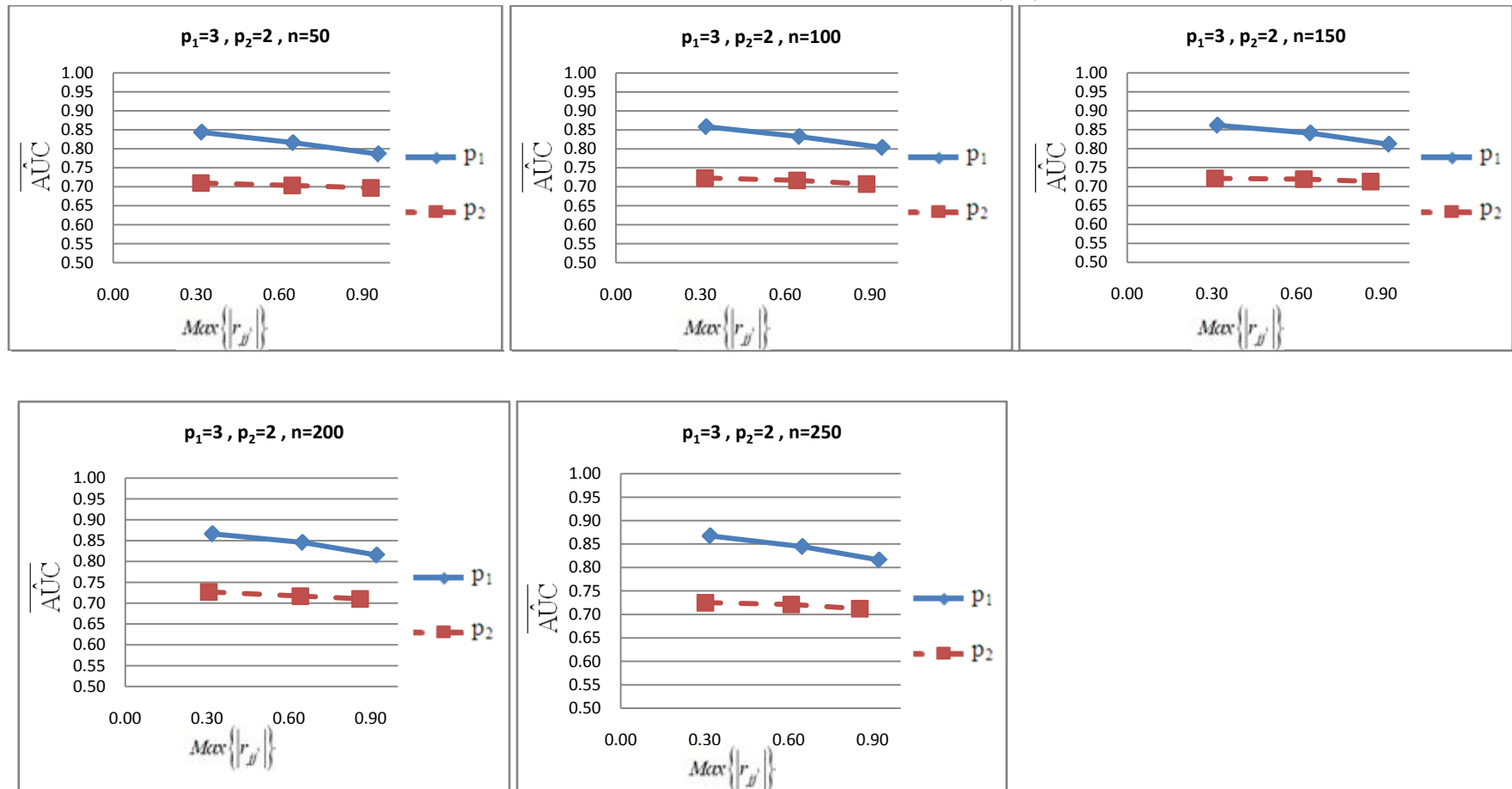


ตารางที่ 4.5 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=3$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

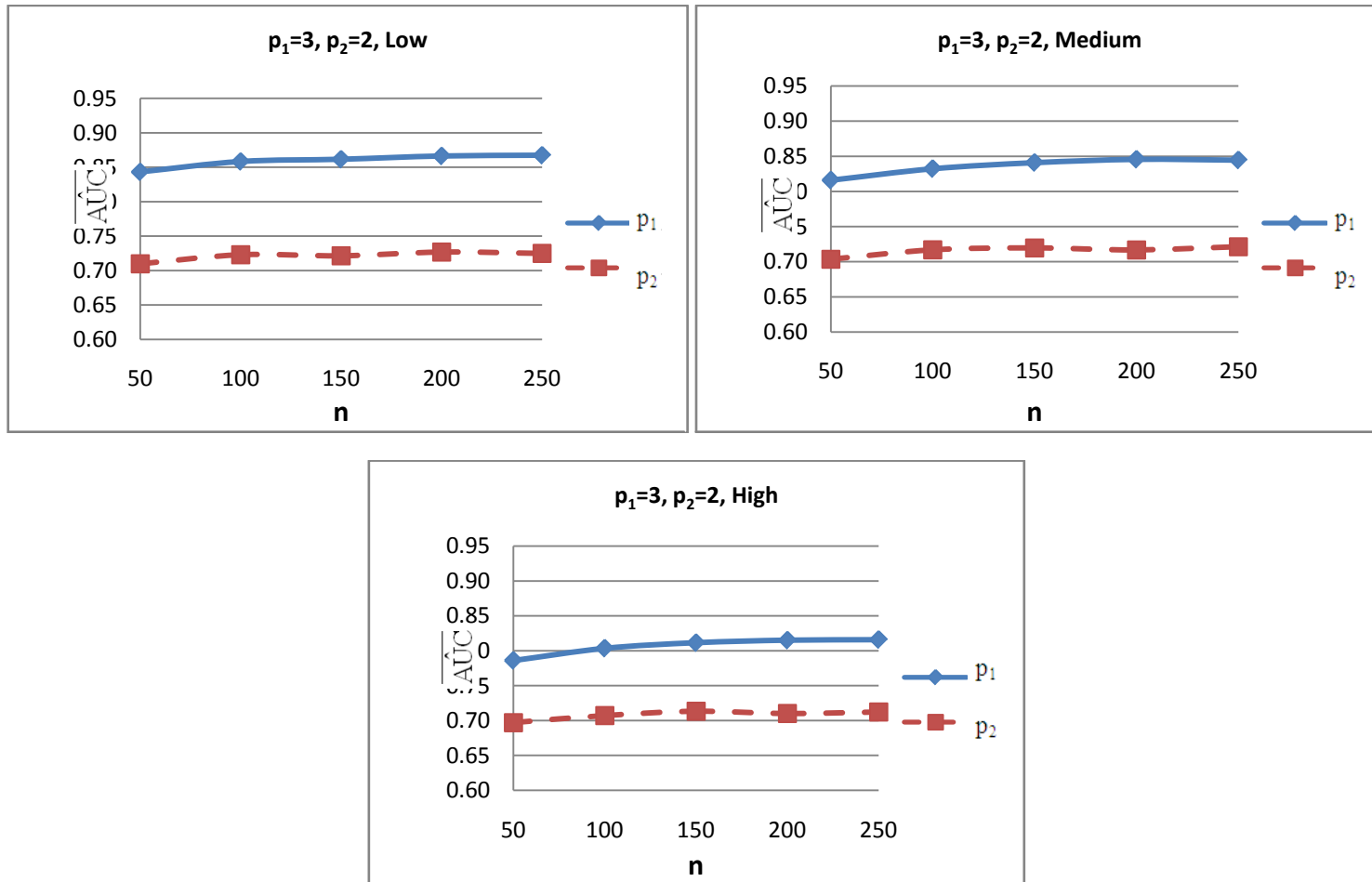
$p_1$	$p_2$	$n$	$Max\{r_{jj}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
3	2	50	Low	0.8432	0.7099	Model 1
			Medium	0.8160	0.7035	Model 1
			High	0.7861	0.6970	Model 1
		100	Low	0.8583	0.7230	Model 1
			Medium	0.8322	0.7169	Model 1
			High	0.8035	0.7072	Model 1
		150	Low	0.8616	0.7215	Model 1
			Medium	0.8409	0.7196	Model 1
			High	0.8115	0.7133	Model 1
		200	Low	0.8663	0.7269	Model 1
			Medium	0.8456	0.7168	Model 1
			High	0.8151	0.7101	Model 1
		250	Low	0.8674	0.7250	Model 1
			Medium	0.8445	0.7213	Model 1
			High	0.8160	0.7123	Model 1

จากตารางที่ 4.5 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=3$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0 : \mu_{model1} \leq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} > \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.9 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=3, p_2=2, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.10 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=3, p_2=2, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n= 50, 100, 150, 200, 250$

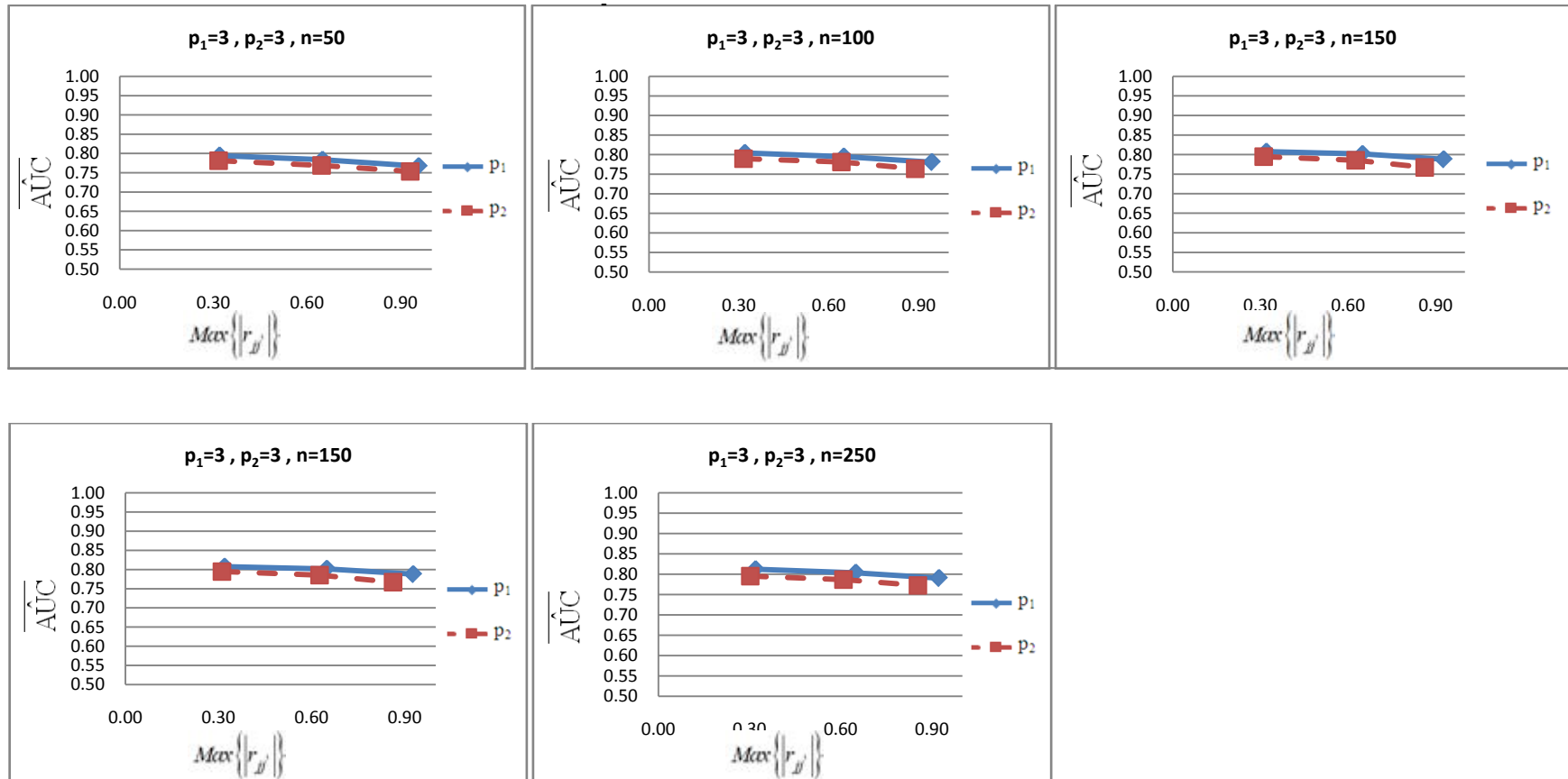


ตารางที่ 4.6 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=3$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

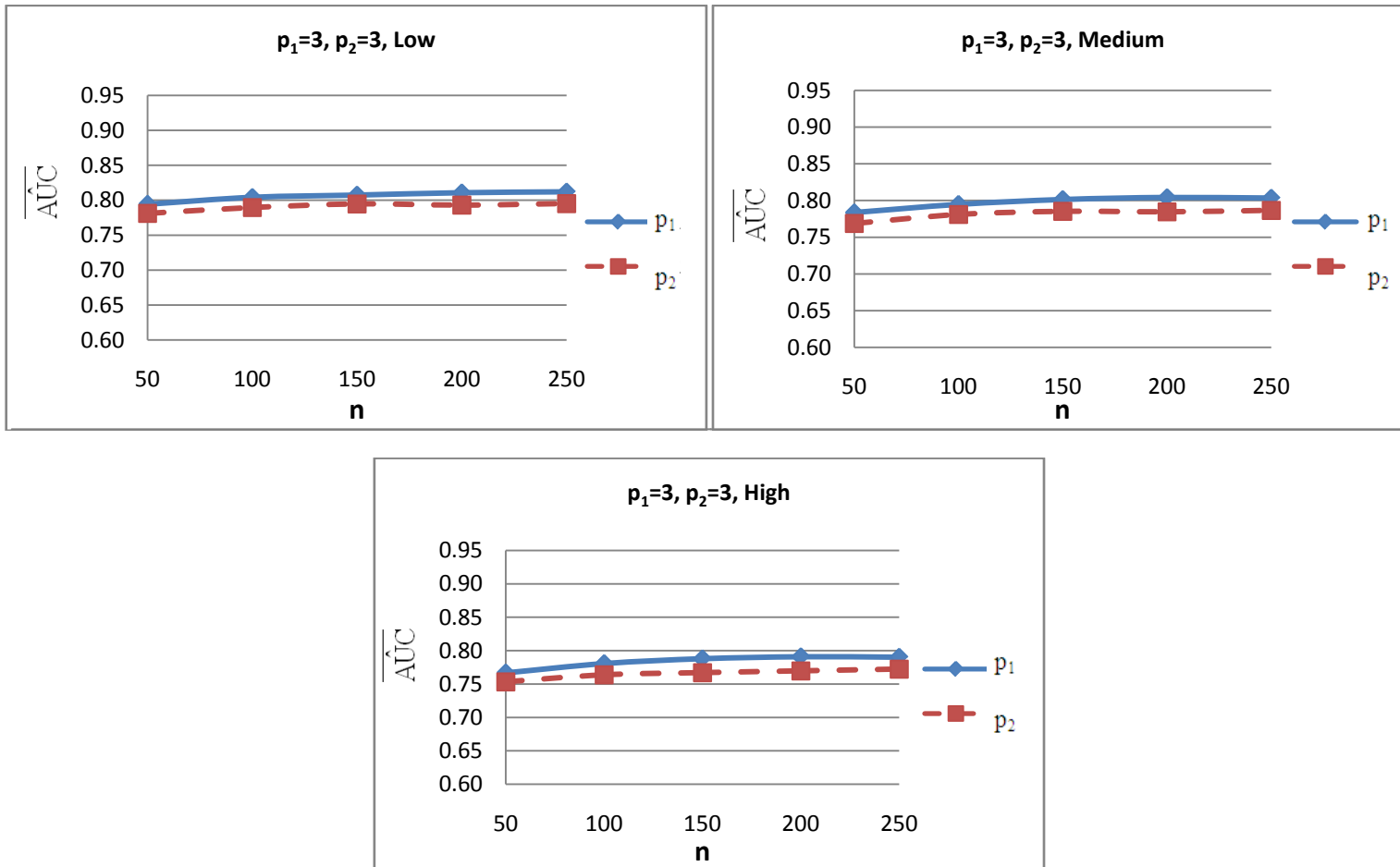
$p_1$	$p_2$	$n$	$Max\{r_{jj}\}$	$\overline{A\hat{U}C}_1$	$\overline{A\hat{U}C}_2$	Selected Model
3	3	50	Low	0.7943	0.7811	Model 1
			Medium	0.7837	0.7689	Model 1
			High	0.7667	0.7533	Model 1
		100	Low	0.8041	0.7895	Model 1
			Medium	0.7948	0.7811	Model 1
			High	0.7807	0.7639	Model 1
		150	Low	0.8074	0.7947	Model 1
			Medium	0.8015	0.7854	Model 1
			High	0.7879	0.7670	Model 1
		200	Low	0.8107	0.7931	Model 1
			Medium	0.8043	0.7847	Model 1
			High	0.7909	0.7697	Model 1
		250	Low	0.8120	0.7952	Model 1
			Medium	0.8034	0.7867	Model 1
			High	0.7902	0.7724	Model 1

จากตารางที่ 4.6 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=3$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0: \mu_{model1} \leq \mu_{model2}$  แยังกับ  $H_0: \mu_{model1} > \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.11 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=3, p_2=3, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.12 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=3, p_2=3, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n = 50, 100, 150, 200, 250$

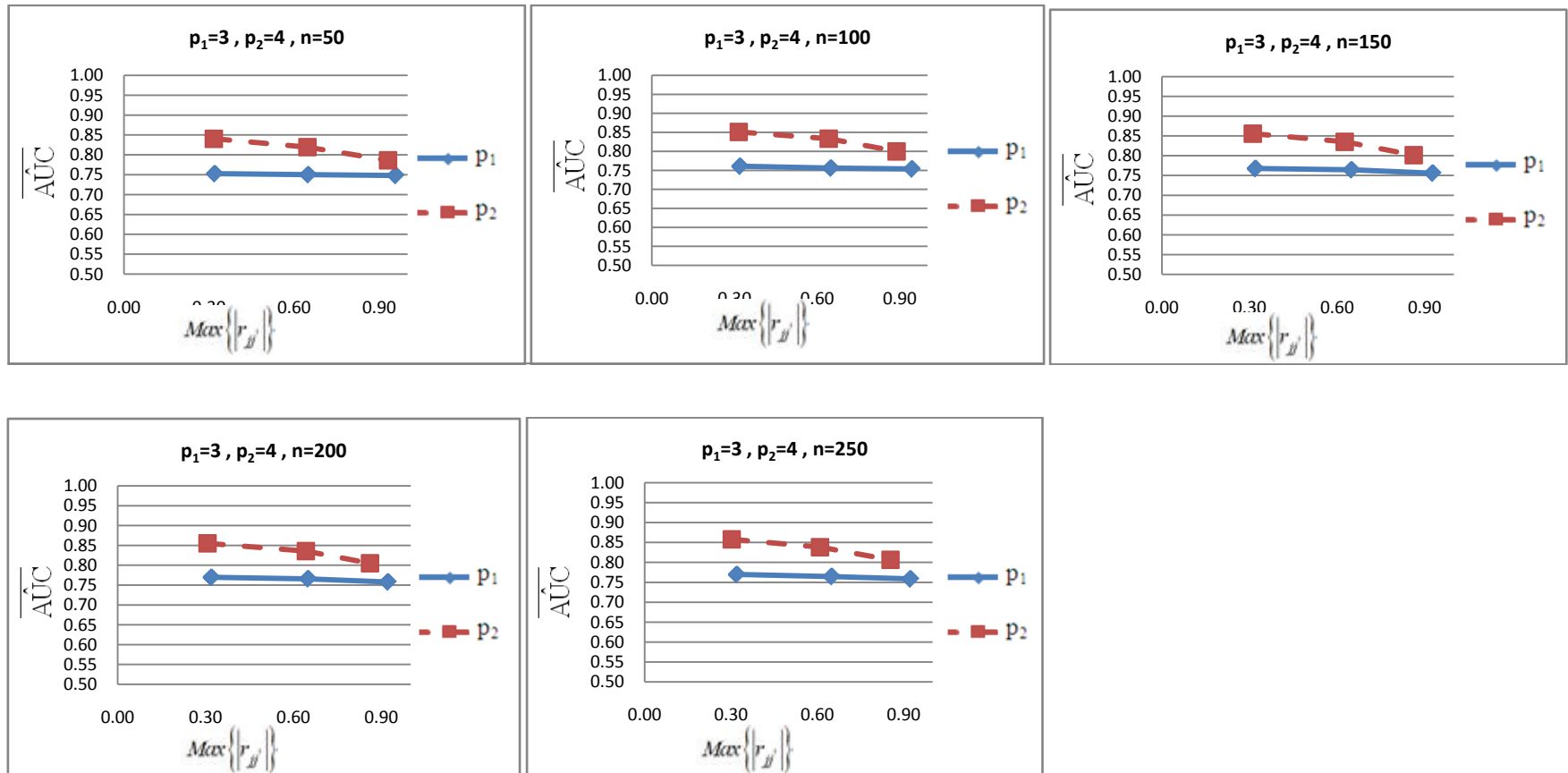


ตารางที่ 4.7 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=3$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

$p_1$	$p_2$	$n$	$Max\{r_{ij}\}$	$\overline{A\hat{U}C}_1$	$\overline{A\hat{U}C}_2$	Selected Model
3	4	50	Low	0.7526	0.8403	Model 2
			Medium	0.7500	0.8194	Model 2
			High	0.7481	0.7861	Model 2
		100	Low	0.7609	0.8510	Model 2
			Medium	0.7559	0.8333	Model 2
			High	0.7536	0.8000	Model 2
		150	Low	0.7675	0.8558	Model 2
			Medium	0.7644	0.8353	Model 2
			High	0.7558	0.8015	Model 2
		200	Low	0.7696	0.8553	Model 2
			Medium	0.7657	0.8358	Model 2
			High	0.7578	0.8051	Model 2
		250	Low	0.7698	0.8579	Model 2
			Medium	0.7644	0.8384	Model 2
			High	0.7585	0.8066	Model 2

จากตารางที่ 4.7 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=3$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0 : \mu_{model1} \geq \mu_{model2}$  แอ้งกับ  $H_0 : \mu_{model1} < \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 มากกว่าตัวแบบที่ 1 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2 ( $p\text{-value} < 0.05$ )

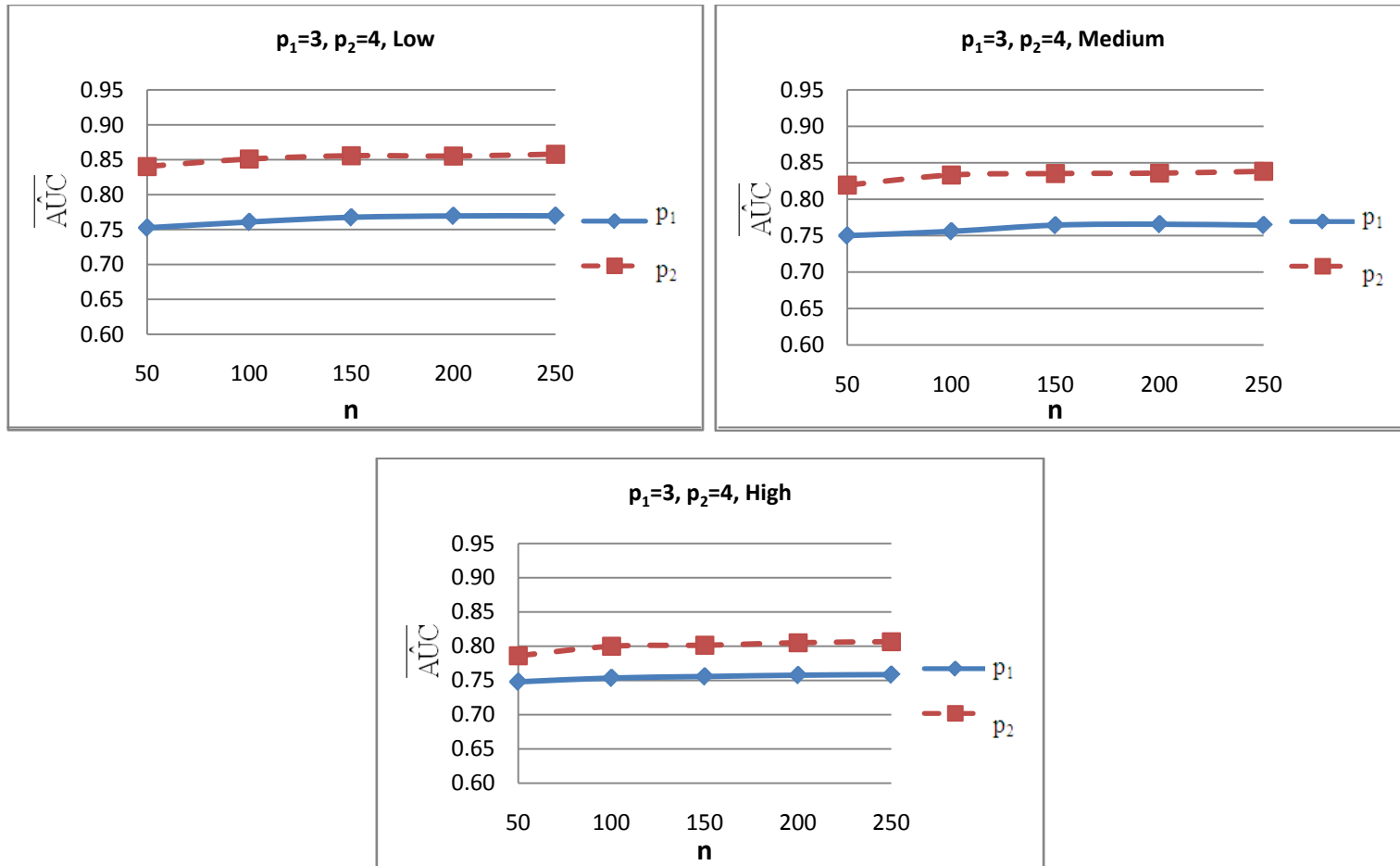
ภาพที่ 4.13 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=3, p_2=4, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High





ภาพที่ 4.14 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=3, p_2=4, \text{Max}\{r_{jj}\} = \text{Low, Medium, High}$  และ ขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n= 50, 100, 150, 200,$

250

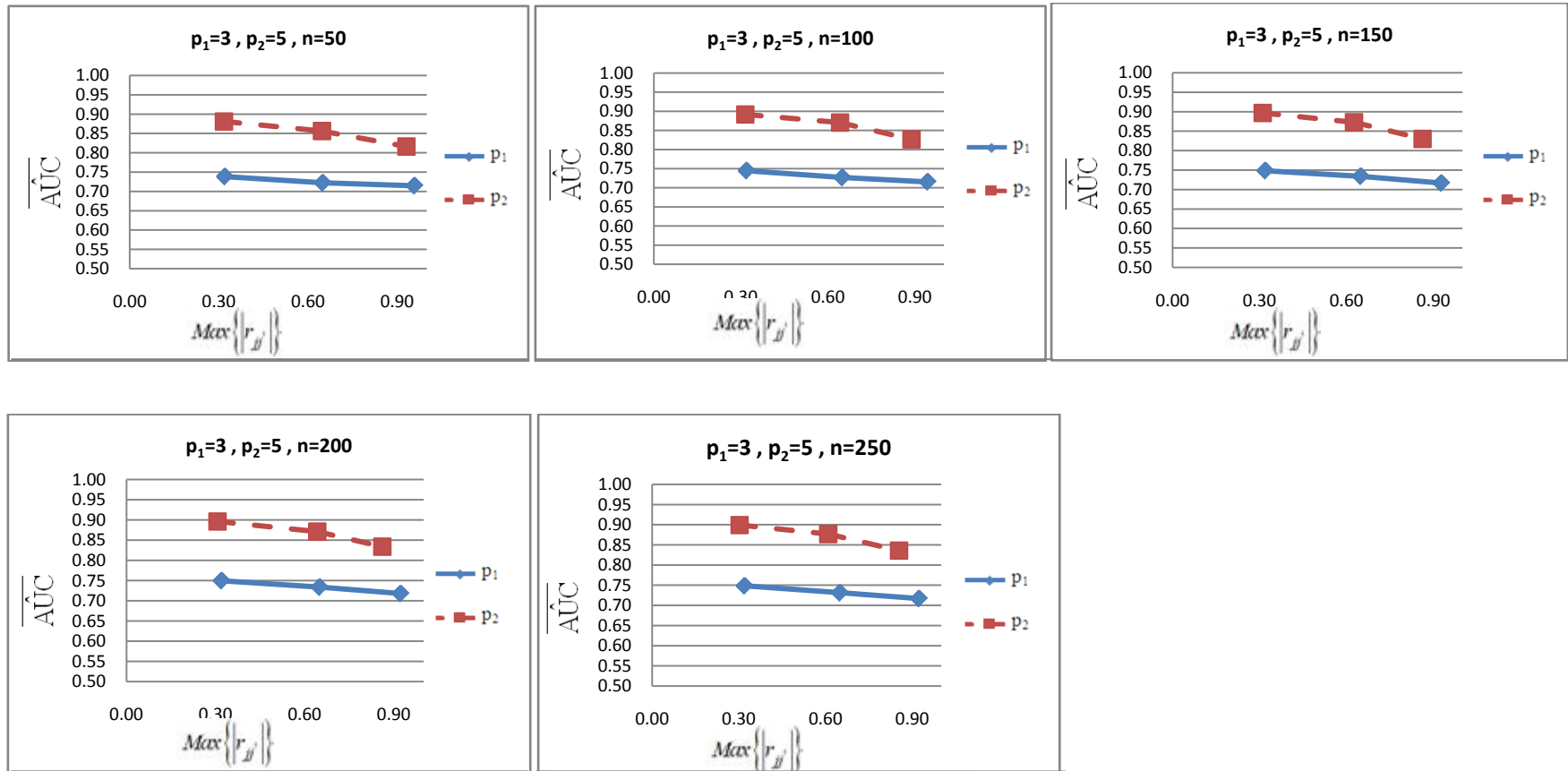


ตารางที่ 4.8 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=3$ ,  $p_2=5$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

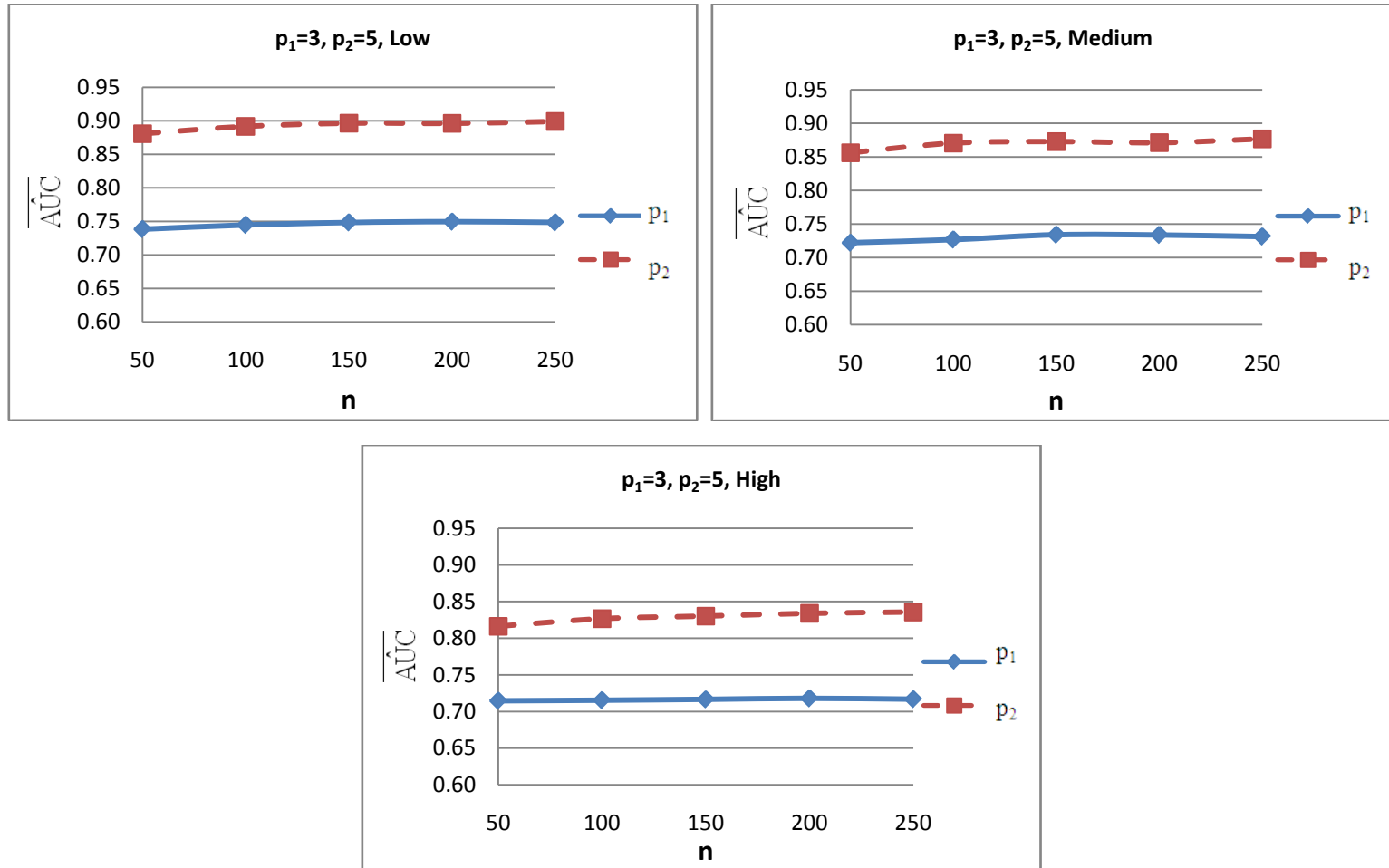
$p_1$	$p_2$	$n$	$Max\{r_{ij}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
3	5	50	Low	0.7384	0.8809	Model 2
			Medium	0.7222	0.8563	Model 2
			High	0.7147	0.8163	Model 2
		100	Low	0.7447	0.8918	Model 2
			Medium	0.7268	0.8706	Model 2
			High	0.7155	0.8267	Model 2
		150	Low	0.7484	0.8966	Model 2
			Medium	0.7340	0.8729	Model 2
			High	0.7166	0.8302	Model 2
		200	Low	0.7496	0.8964	Model 2
			Medium	0.7338	0.8713	Model 2
			High	0.7179	0.8340	Model 2
		250	Low	0.7485	0.8991	Model 2
			Medium	0.7313	0.8769	Model 2
			High	0.7168	0.8358	Model 2

จากตารางที่ 4.8 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=3$ ,  $p_2=5$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0 : \mu_{model1} \geq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} < \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 มากกว่าตัวแบบที่ 1 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.15 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=3, p_2=5, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.16 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=3, p_2=5, \text{Max}\{r_{jj}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n= 50, 100, 150, 200, 250$

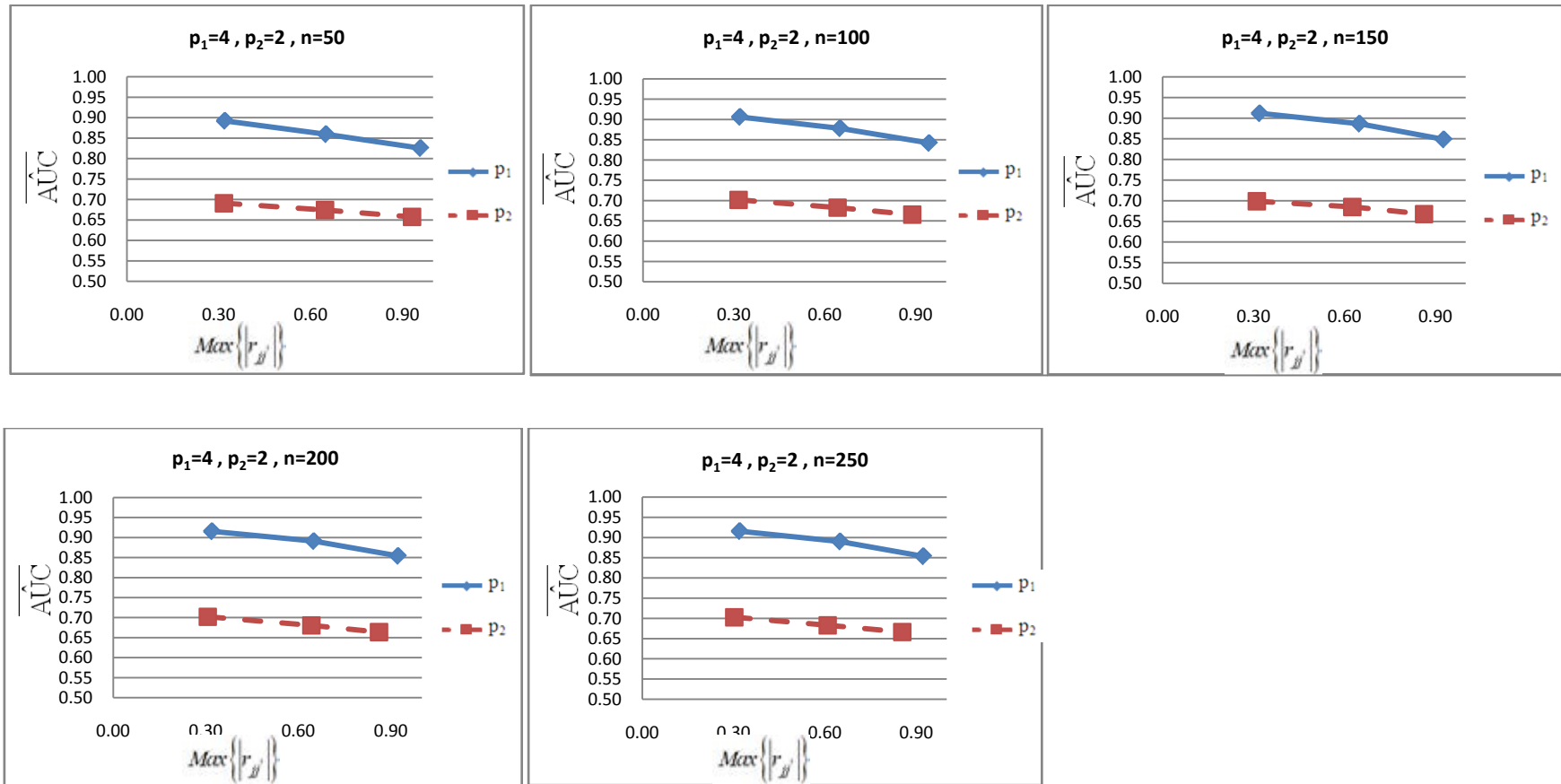


ตารางที่ 4.9 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=4$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $\text{Max}\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

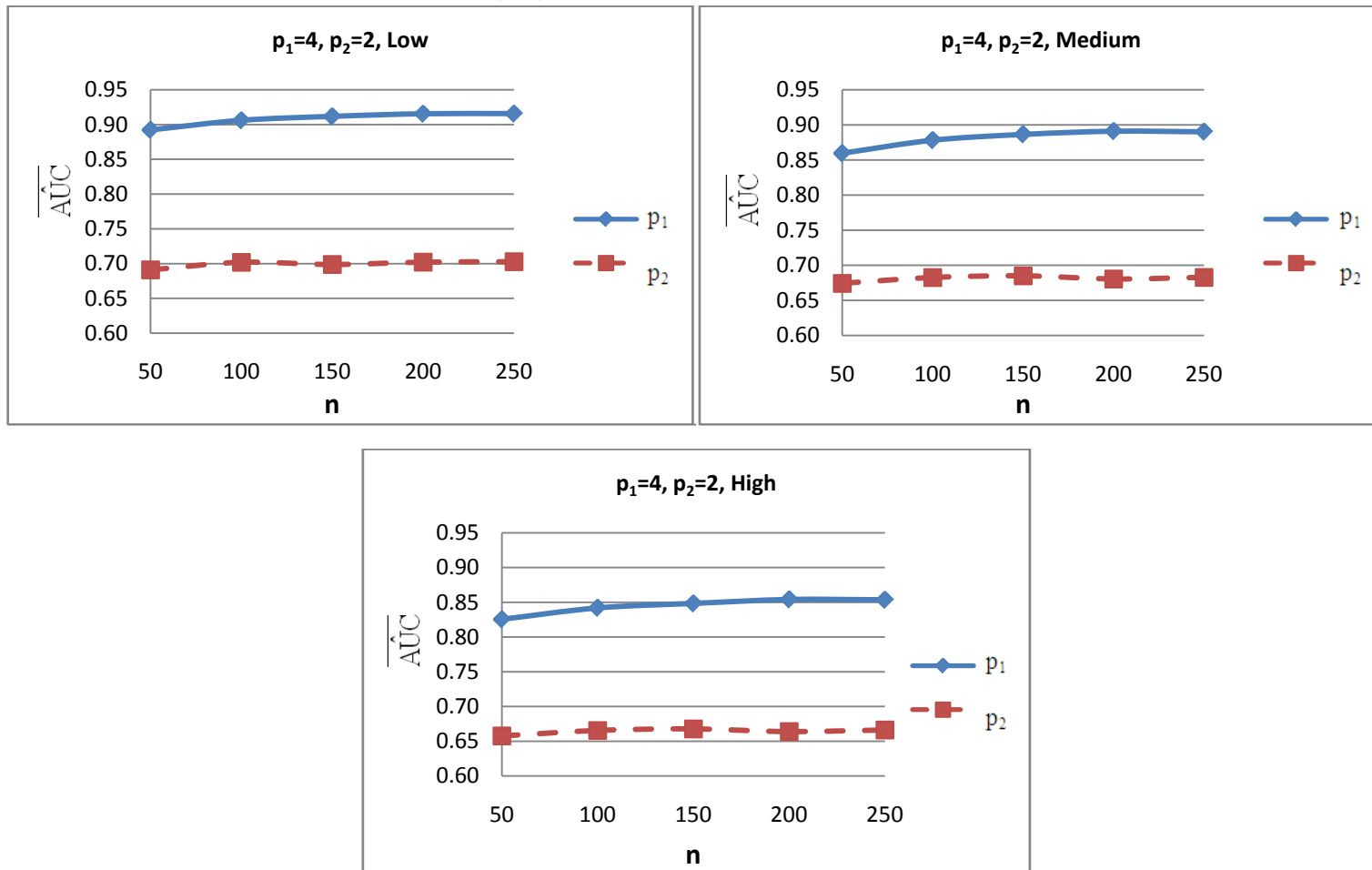
$p_1$	$p_2$	$n$	$\text{Max}\{r_{jj}\}$	$\overline{\text{AUC}}_1$	$\overline{\text{AUC}}_2$	Selected Model
4	2	50	Low	0.8923	0.6910	Model 1
			Medium	0.8596	0.6744	Model 1
			High	0.8256	0.6576	Model 1
		100	Low	0.9062	0.7018	Model 1
			Medium	0.8782	0.6827	Model 1
			High	0.8420	0.6655	Model 1
		150	Low	0.9118	0.6988	Model 1
			Medium	0.8865	0.6850	Model 1
			High	0.8484	0.6677	Model 1
		200	Low	0.9156	0.7021	Model 1
			Medium	0.8911	0.6806	Model 1
			High	0.8542	0.6639	Model 1
		250	Low	0.9158	0.7028	Model 1
			Medium	0.8903	0.6829	Model 1
			High	0.8536	0.6663	Model 1

จากตารางที่ 4.9 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=4$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $\text{Max}\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0: \mu_{\text{model1}} \leq \mu_{\text{model2}}$  แยังกับ  $H_0: \mu_{\text{model1}} > \mu_{\text{model2}}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.17 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=4, p_2=2, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{j\cdot}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.18 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=4, p_2=2, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n=50, 100, 150, 200, 250$



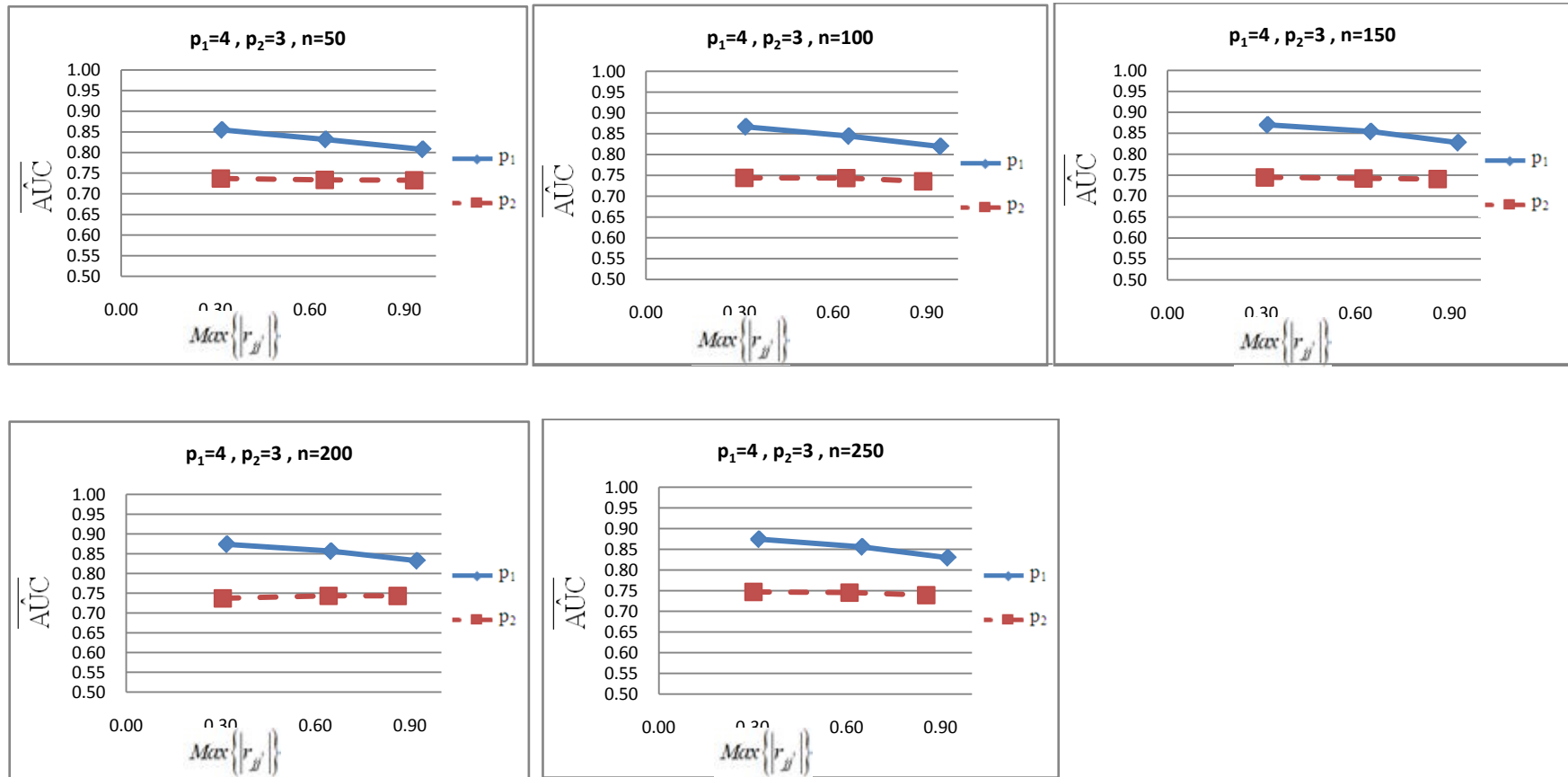
ตารางที่ 4.10 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=4$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

$P_1$	$P_2$	n	$Max\{r_{jj}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
4	3	50	Low	0.8548	0.7371	Model 1
			Medium	0.8316	0.7337	Model 1
			High	0.8075	0.7329	Model 1
		100	Low	0.8667	0.7442	Model 1
			Medium	0.8445	0.7436	Model 1
			High	0.8194	0.7358	Model 1
		150	Low	0.8699	0.7446	Model 1
			Medium	0.8542	0.7422	Model 1
			High	0.8273	0.7408	Model 1
		200	Low	0.8738	0.7372	Model 1
			Medium	0.8566	0.7435	Model 1
			High	0.8322	0.7434	Model 1
		250	Low	0.8744	0.7470	Model 1
			Medium	0.8558	0.7454	Model 1
			High	0.8295	0.7393	Model 1

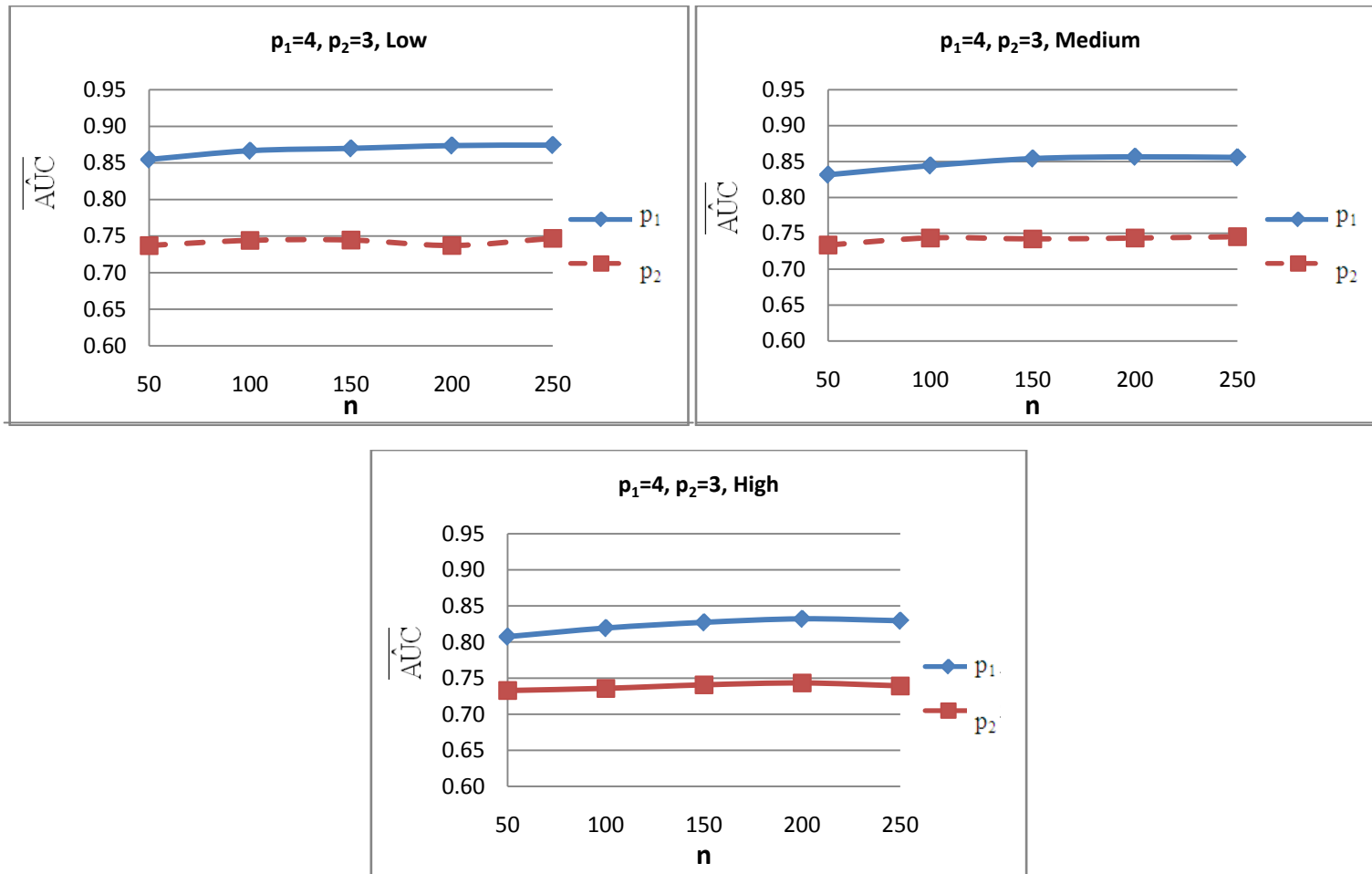
จากตารางที่ 4.10 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=4$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0 : \mu_{model1} \leq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} > \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )



ภาพที่ 4.19 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=4, p_2=3, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.20 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=4, p_2=3, \text{Max}\{r_{jj}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n= 50, 100, 150, 200, 250$

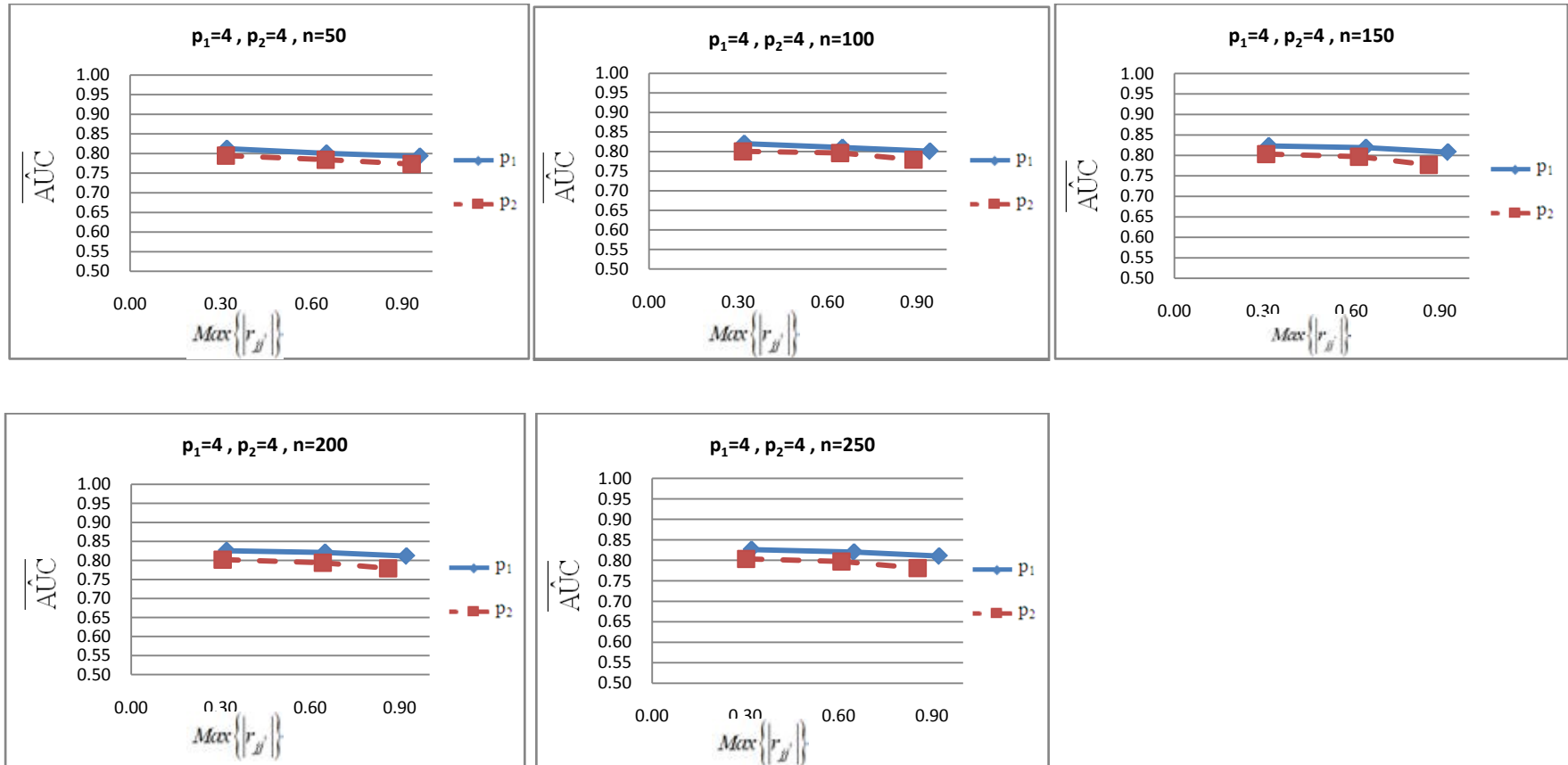


ตารางที่ 4.11 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=4$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $\text{Max}\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

$p_1$	$p_2$	$n$	$\text{Max}\{r_{ij}\}$	$\overline{\text{AUC}}_1$	$\overline{\text{AUC}}_2$	Selected Model
4	4	50	Low	0.8126	0.7948	Model 1
			Medium	0.8006	0.7843	Model 1
			High	0.7925	0.7734	Model 1
		100	Low	0.8204	0.8006	Model 1
			Medium	0.8103	0.7965	Model 1
			High	0.8009	0.7794	Model 1
		150	Low	0.8233	0.8032	Model 1
			Medium	0.8190	0.7969	Model 1
			High	0.8079	0.7769	Model 1
		200	Low	0.8256	0.8022	Model 1
			Medium	0.8210	0.7940	Model 1
			High	0.8115	0.7794	Model 1
		250	Low	0.8262	0.8033	Model 1
			Medium	0.8200	0.7970	Model 1
			High	0.8103	0.7811	Model 1

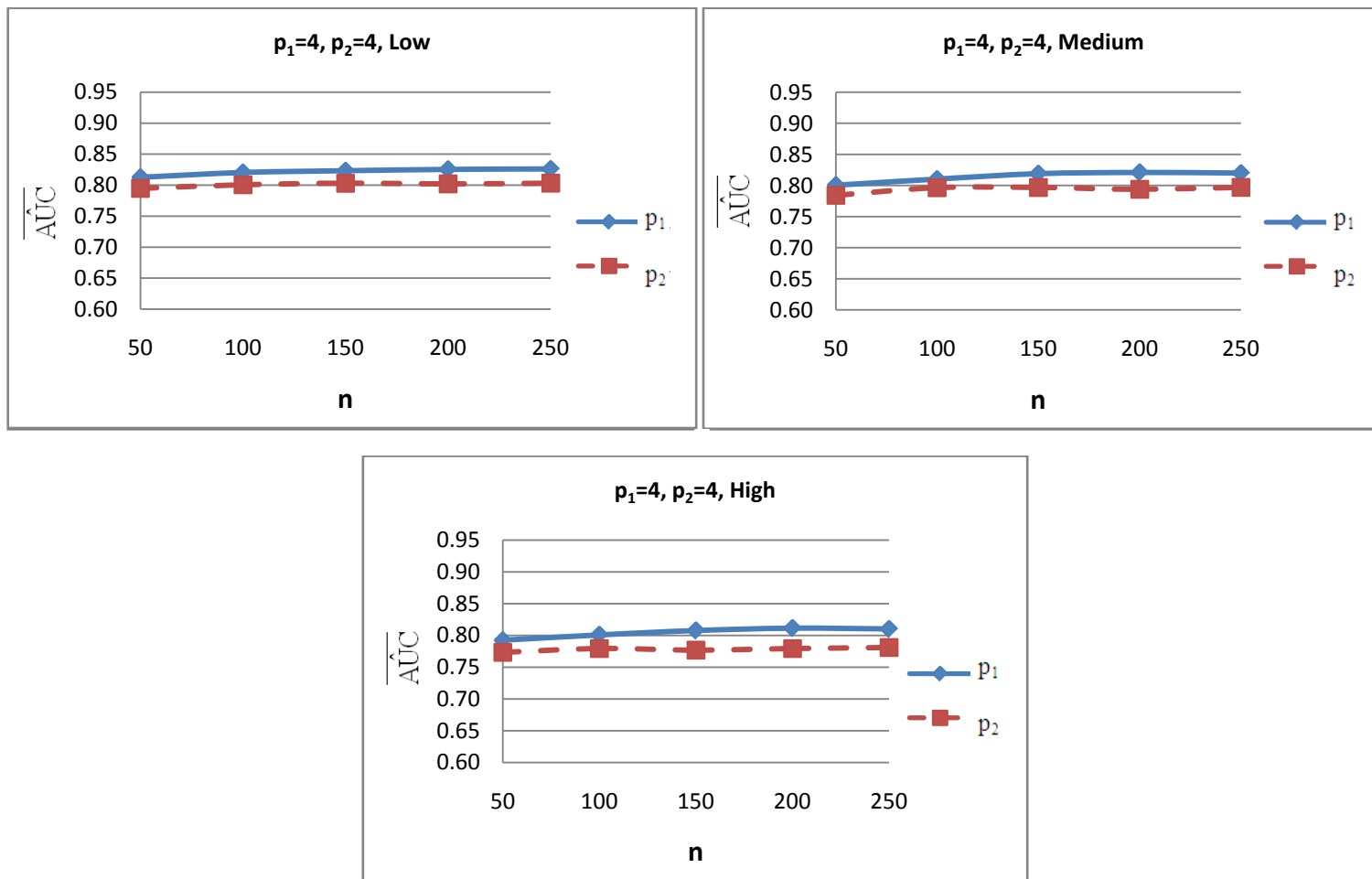
จากตารางที่ 4.11 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=4$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $\text{Max}\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้องมากที่สุดคือ  $H_0: \mu_{\text{model1}} \leq \mu_{\text{model2}}$  แยังกับ  $H_0: \mu_{\text{model1}} > \mu_{\text{model2}}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.21 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=4, p_2=4, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.22 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=4, p_2=4, \text{Max}\{r_{jj}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n=50, 100, 150, 200, 250$

250

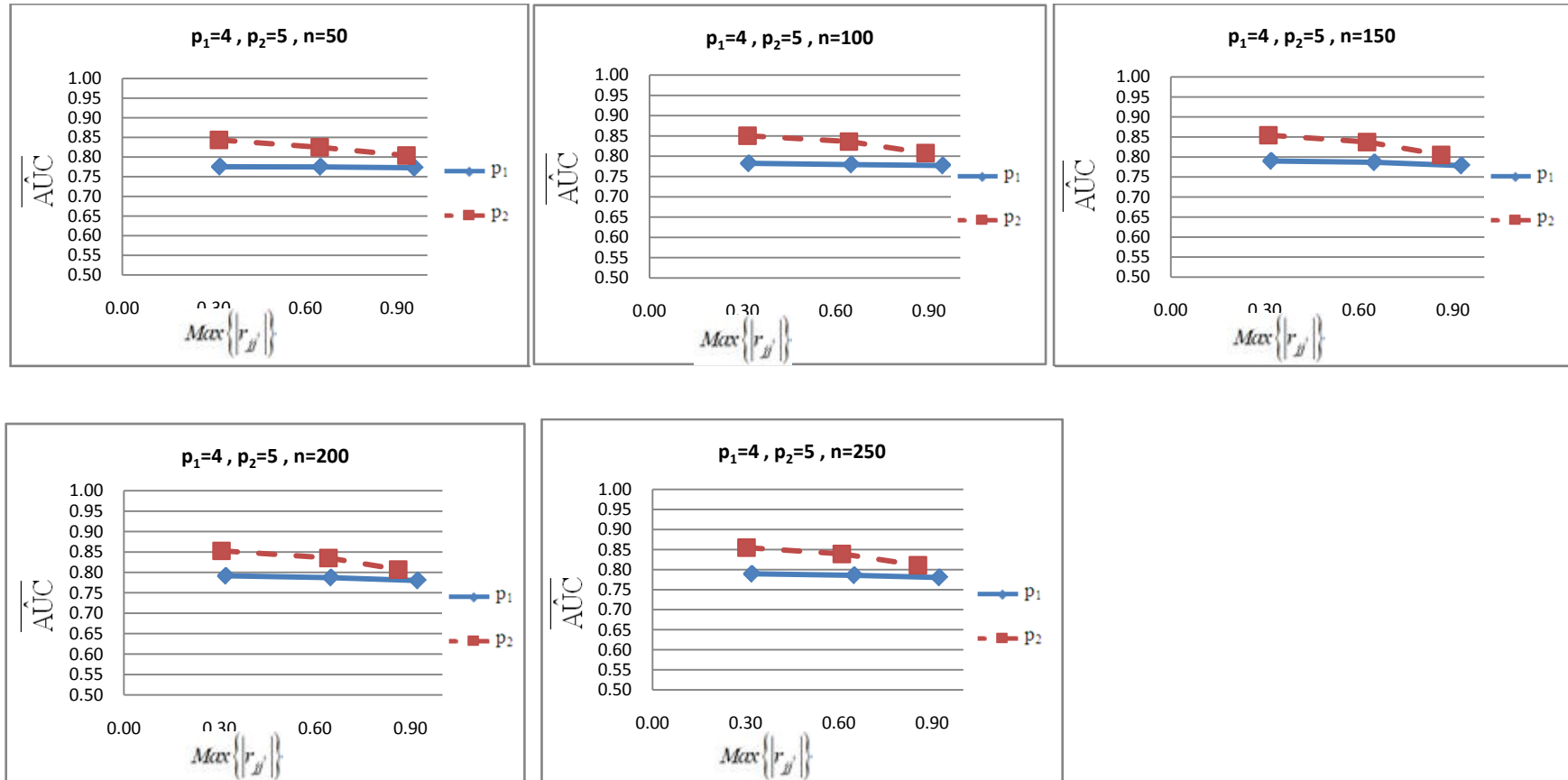


ตารางที่ 4.12 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=4$ ,  $p_2=5$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

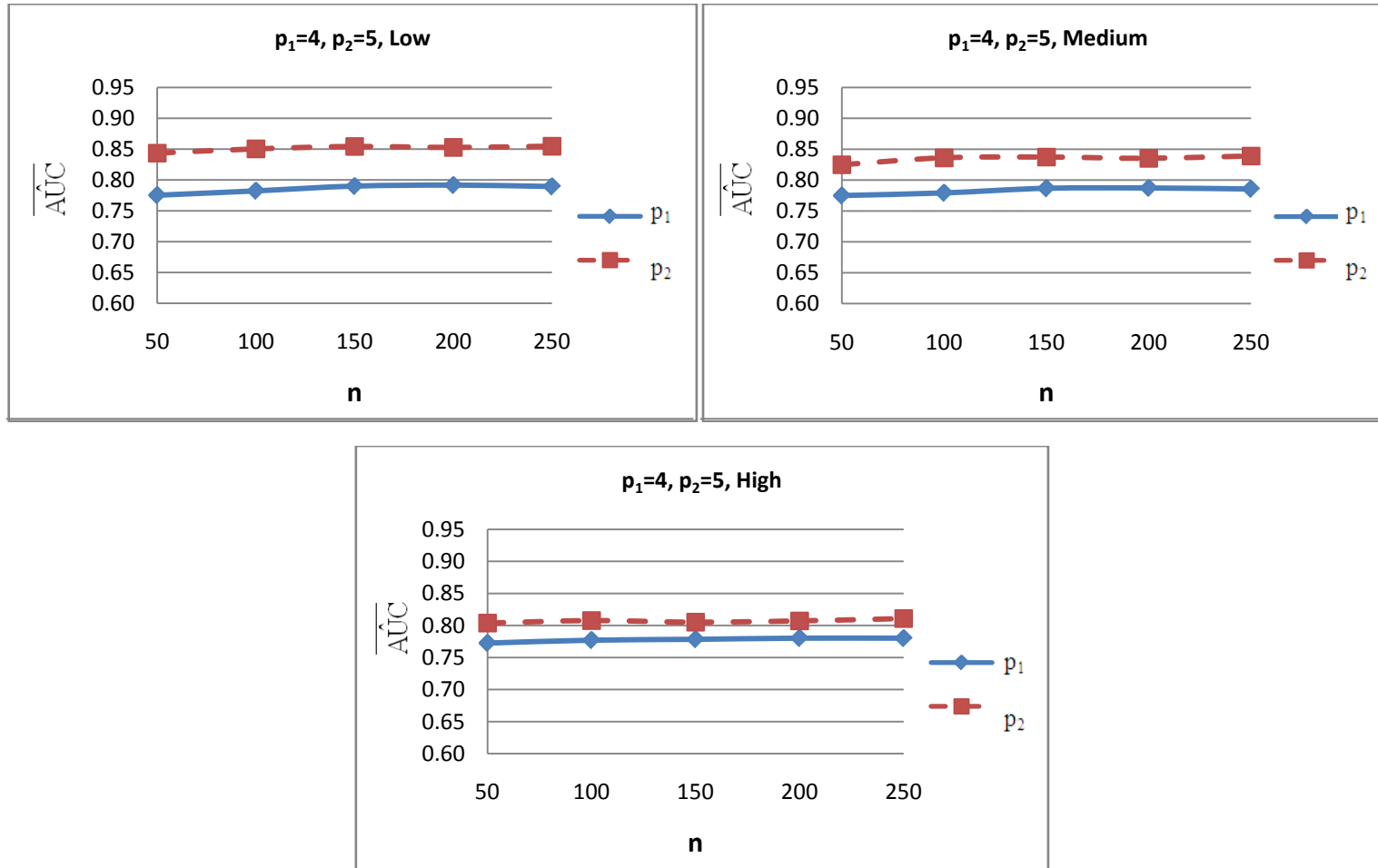
$p_1$	$p_2$	n	$Max\{r_{jj}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
4	5	50	Low	0.7753	0.8436	Model 2
			Medium	0.7750	0.8250	Model 2
			High	0.7726	0.8039	Model 2
		100	Low	0.7824	0.8505	Model 2
			Medium	0.7794	0.8362	Model 2
			High	0.7771	0.8078	Model 2
		150	Low	0.7901	0.8544	Model 2
			Medium	0.7868	0.8374	Model 2
			High	0.7785	0.8053	Model 2
		200	Low	0.7917	0.8528	Model 2
			Medium	0.7873	0.8355	Model 2
			High	0.7803	0.8072	Model 2
		250	Low	0.7895	0.8548	Model 2
			Medium	0.7857	0.8391	Model 2
			High	0.7802	0.8110	Model 2

จากตารางที่ 4.12 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=4$ ,  $p_2=5$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้อง มากที่สุดคือ  $H_0 : \mu_{model1} \geq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} < \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 มากกว่าตัวแบบที่ 1 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.23 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=4, p_2=5, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.24 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=4, p_2=5, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n=50, 100, 150, 200, 250$



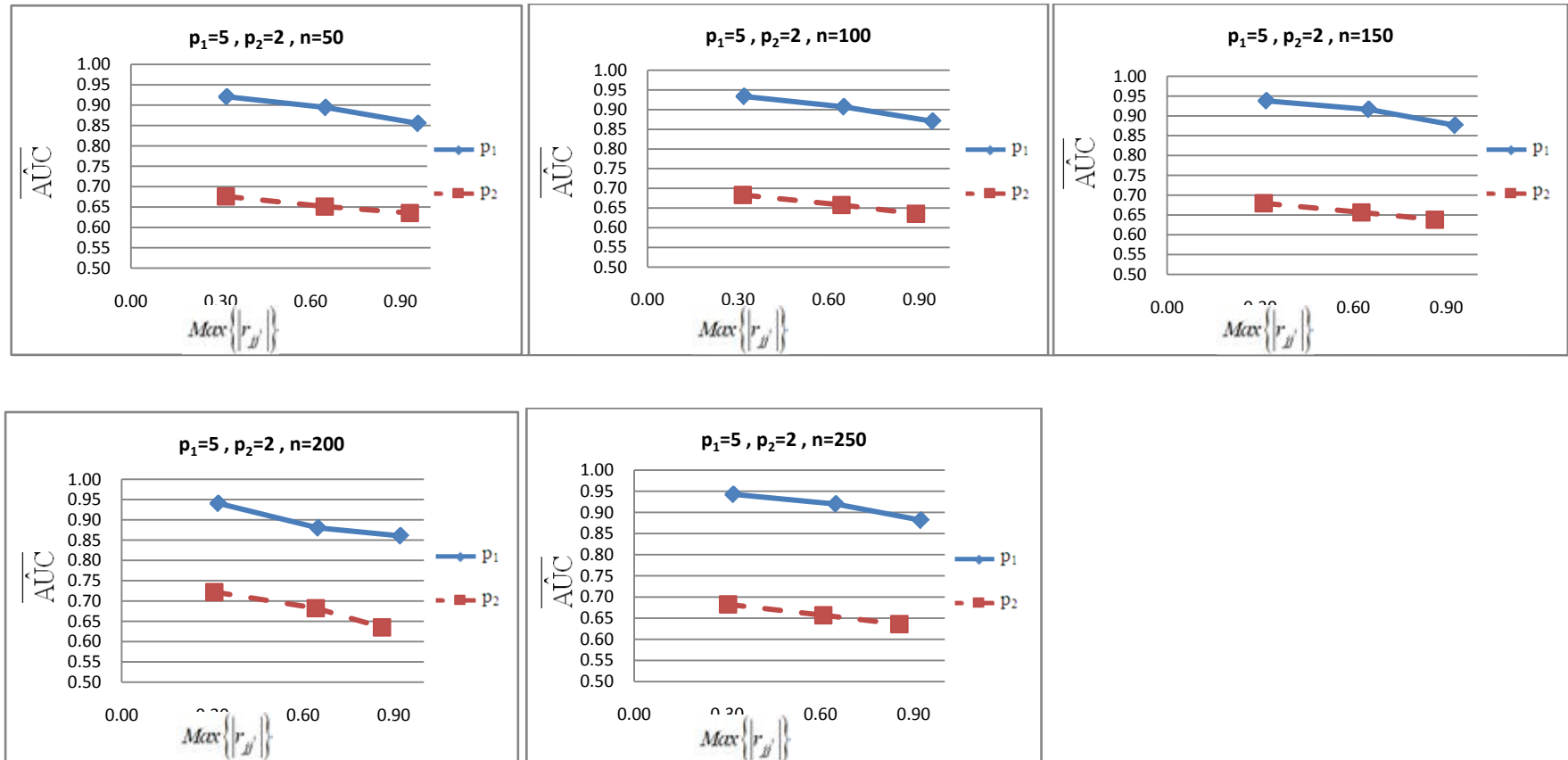


ตารางที่ 4.13 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=5$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

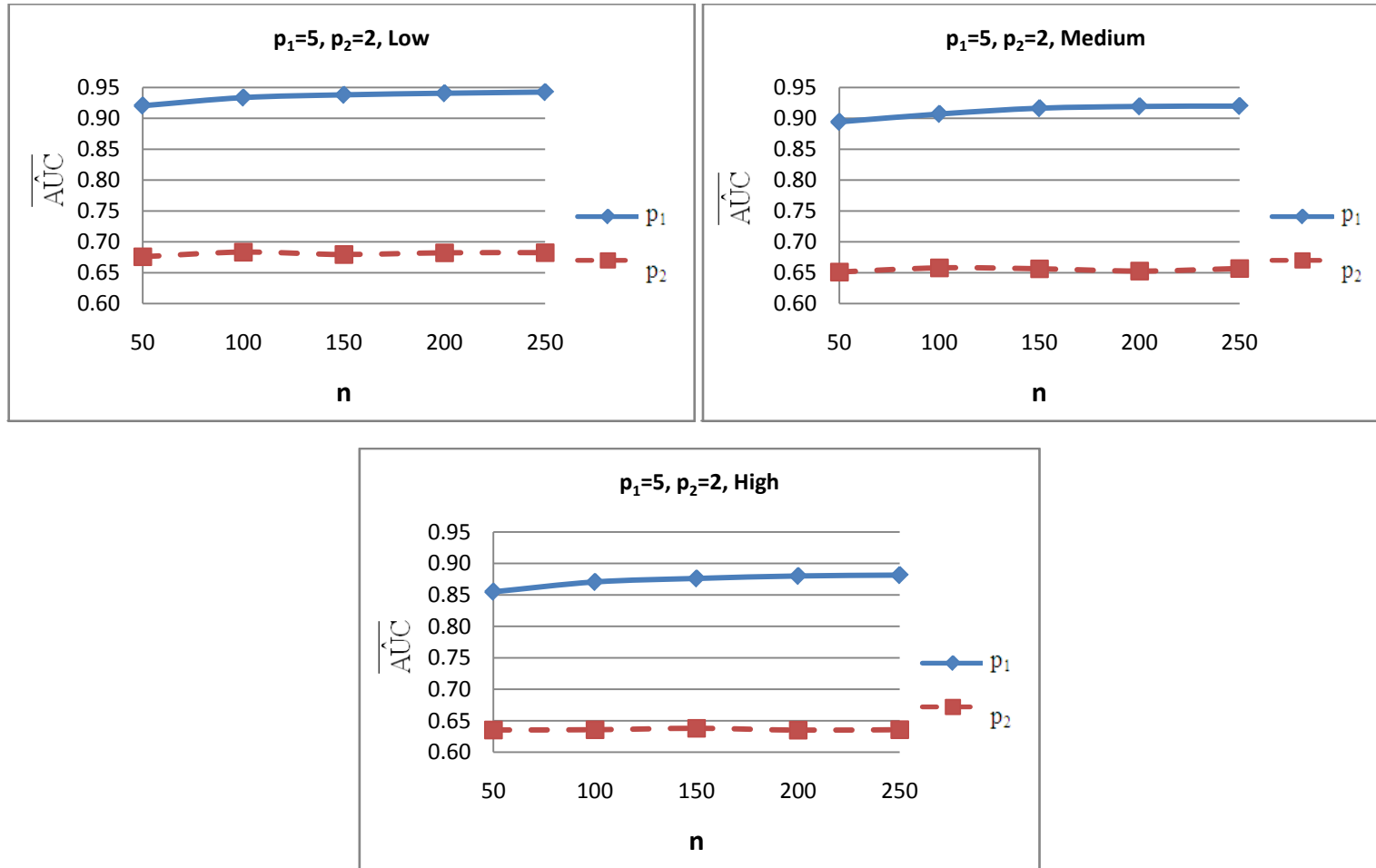
$p_1$	$p_2$	n	$Max\{r_{ij}\}$	$\overline{A\hat{U}C}_1$	$\overline{A\hat{U}C}_2$	Selected Model
5	2	50	Low	0.9206	0.6760	Model 1
			Medium	0.8943	0.6512	Model 1
			High	0.8551	0.6353	Model 1
		100	Low	0.9337	0.6834	Model 1
			Medium	0.9071	0.6579	Model 1
			High	0.8706	0.6357	Model 1
		150	Low	0.9381	0.6796	Model 1
			Medium	0.9165	0.6564	Model 1
			High	0.8762	0.6380	Model 1
		200	Low	0.9407	0.7221	Model 1
			Medium	0.8801	0.6823	Model 1
			High	0.8606	0.6351	Model 1
		250	Low	0.9426	0.6826	Model 1
			Medium	0.9199	0.6569	Model 1
			High	0.8815	0.6359	Model 1

จากตารางที่ 4.13 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=5$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้อง มากที่สุดคือ  $H_0 : \mu_{model1} \leq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} > \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.25 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=5$ ,  $p_2=2$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.26 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=5, p_2=2, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n=50, 100, 150, 200, 250$

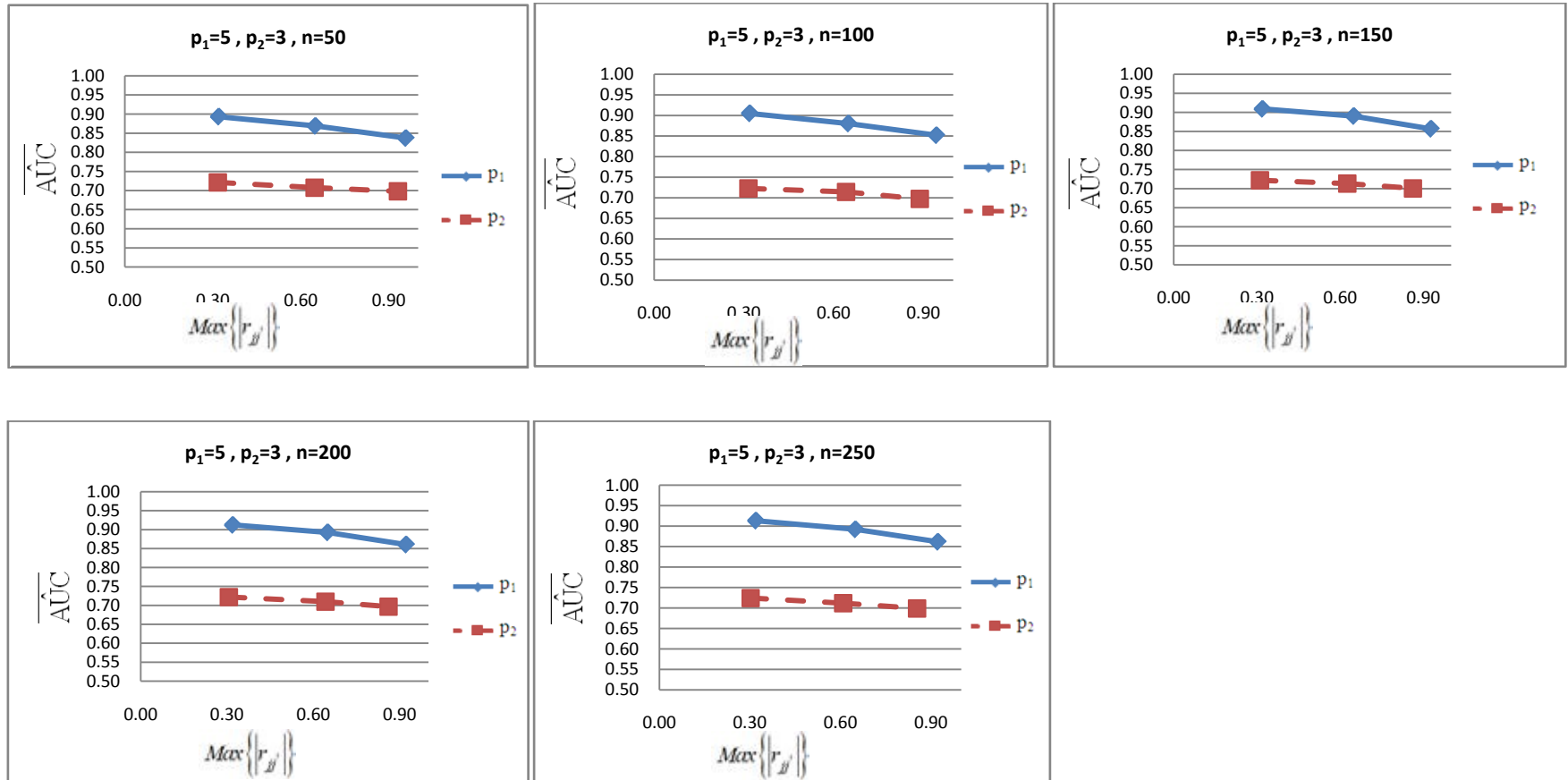


ตารางที่ 4.14 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=5$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

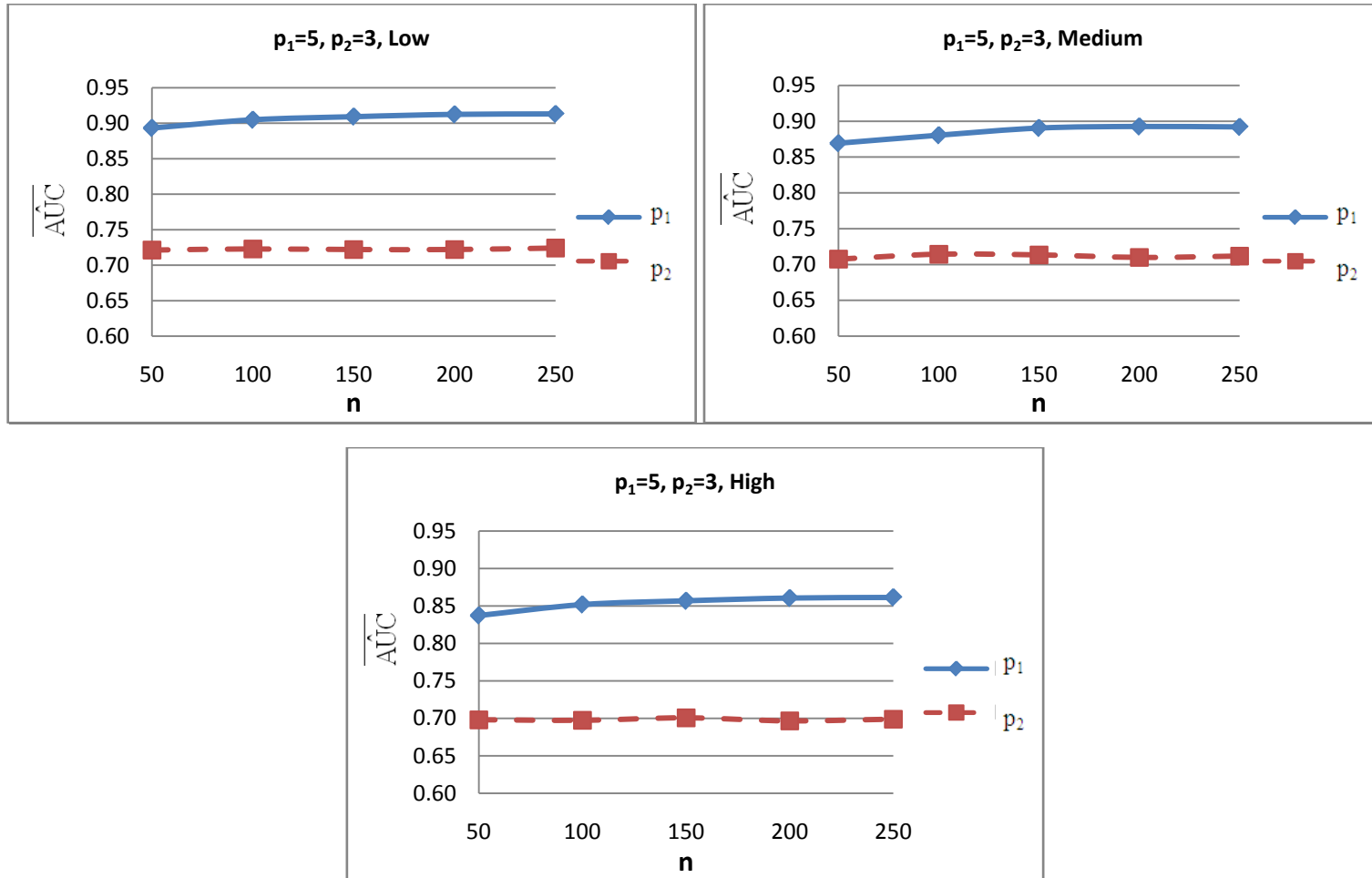
$p_1$	$p_2$	n	$Max\{r_{ij}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
5	3	50	Low	0.8933	0.7213	Model 1
			Medium	0.8692	0.7076	Model 1
			High	0.8373	0.6981	Model 1
		100	Low	0.9049	0.7229	Model 1
			Medium	0.8804	0.7144	Model 1
			High	0.8519	0.6975	Model 1
		150	Low	0.9094	0.7221	Model 1
			Medium	0.8904	0.7134	Model 1
			High	0.8569	0.7008	Model 1
		200	Low	0.9125	0.7221	Model 1
			Medium	0.8926	0.7099	Model 1
			High	0.8606	0.6967	Model 1
		250	Low	0.9131	0.7242	Model 1
			Medium	0.8920	0.7118	Model 1
			High	0.8615	0.6989	Model 1

จากตารางที่ 4.14 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=5$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้อง มากที่สุดคือ  $H_0 : \mu_{model1} \leq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} > \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.27 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=5$ ,  $p_2=3$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{j'}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.28 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=5, p_2=3, \text{Max}\{|r_{ij}|\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n= 50, 100, 150, 200, 250$

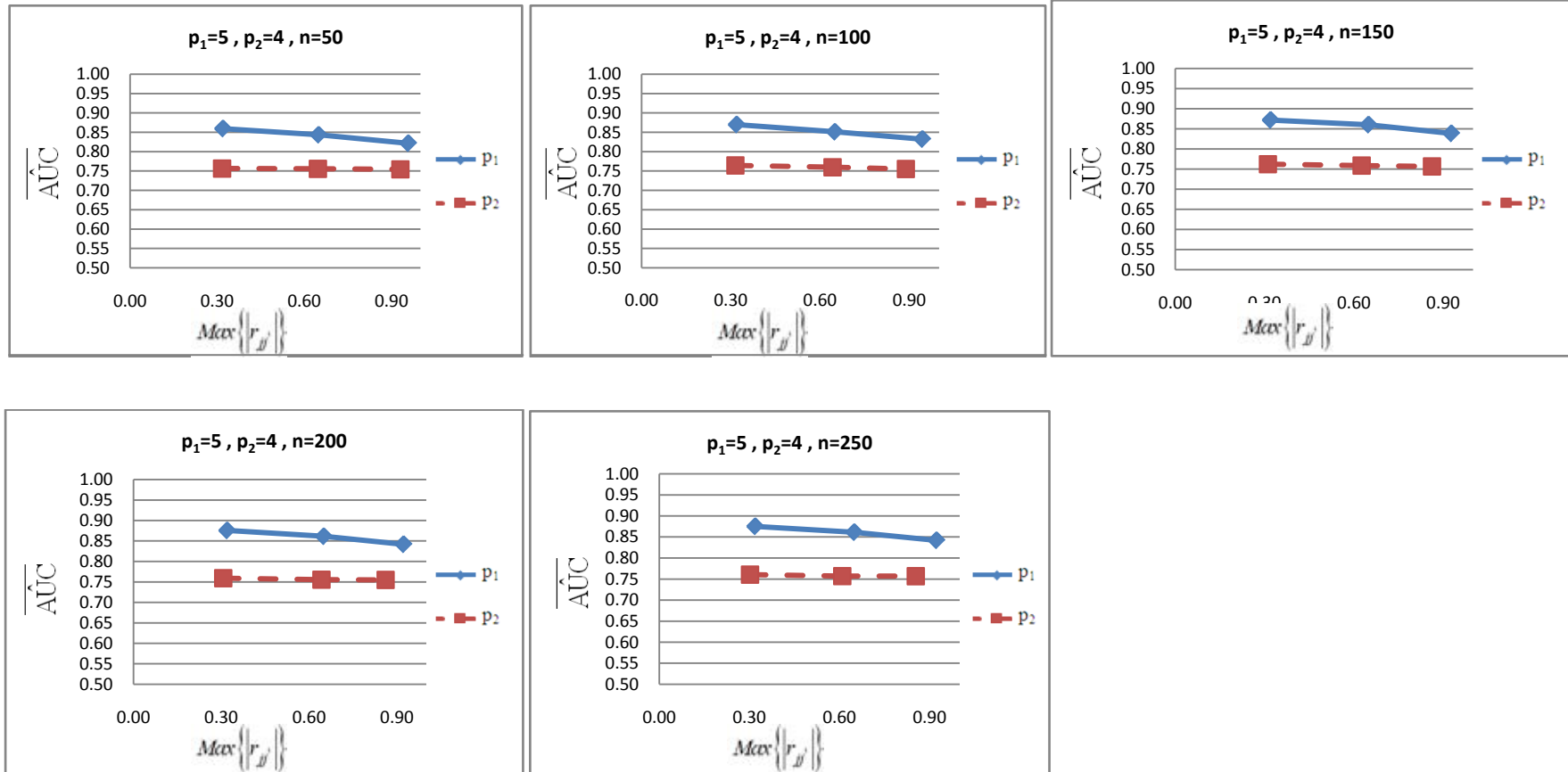


ตารางที่ 4.27 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=5$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

$p_1$	$p_2$	n	$Max\{r_{ij}\}$	$\overline{AUC}_1$	$\overline{AUC}_2$	Selected Model
5	4	50	Low	0.8595	0.7565	Model 1
			Medium	0.8435	0.7557	Model 1
			High	0.8216	0.7544	Model 1
		100	Low	0.8699	0.7644	Model 1
			Medium	0.8512	0.7598	Model 1
			High	0.8323	0.7551	Model 1
		150	Low	0.8718	0.7621	Model 1
			Medium	0.8601	0.7588	Model 1
			High	0.8385	0.7566	Model 1
		200	Low	0.8756	0.7591	Model 1
			Medium	0.8614	0.7555	Model 1
			High	0.8418	0.7548	Model 1
		250	Low	0.8751	0.7606	Model 1
			Medium	0.8613	0.7571	Model 1
			High	0.8419	0.7570	Model 1

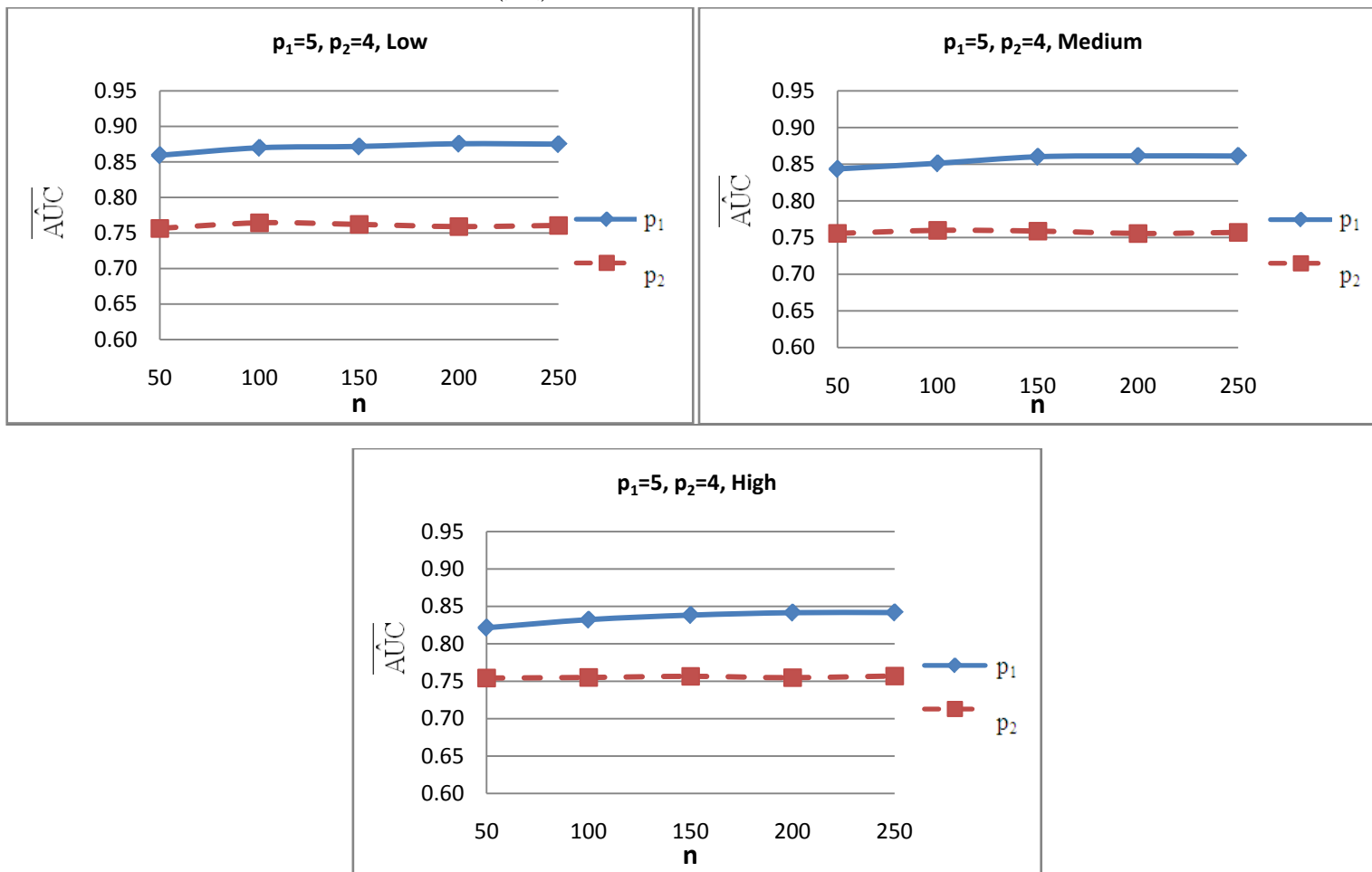
จากตารางที่ 4.15 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=5$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้อง มากที่สุดคือ  $H_0 : \mu_{model1} \leq \mu_{model2}$  แยังกับ  $H_0 : \mu_{model1} > \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.29 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=5, p_2=4, n=50, n=100, n=150, n=200, n=250$  และ  $\text{Max}\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High





ภาพที่ 4.30 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=5, p_2=4, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n = 50, 100, 150, 200, 250$

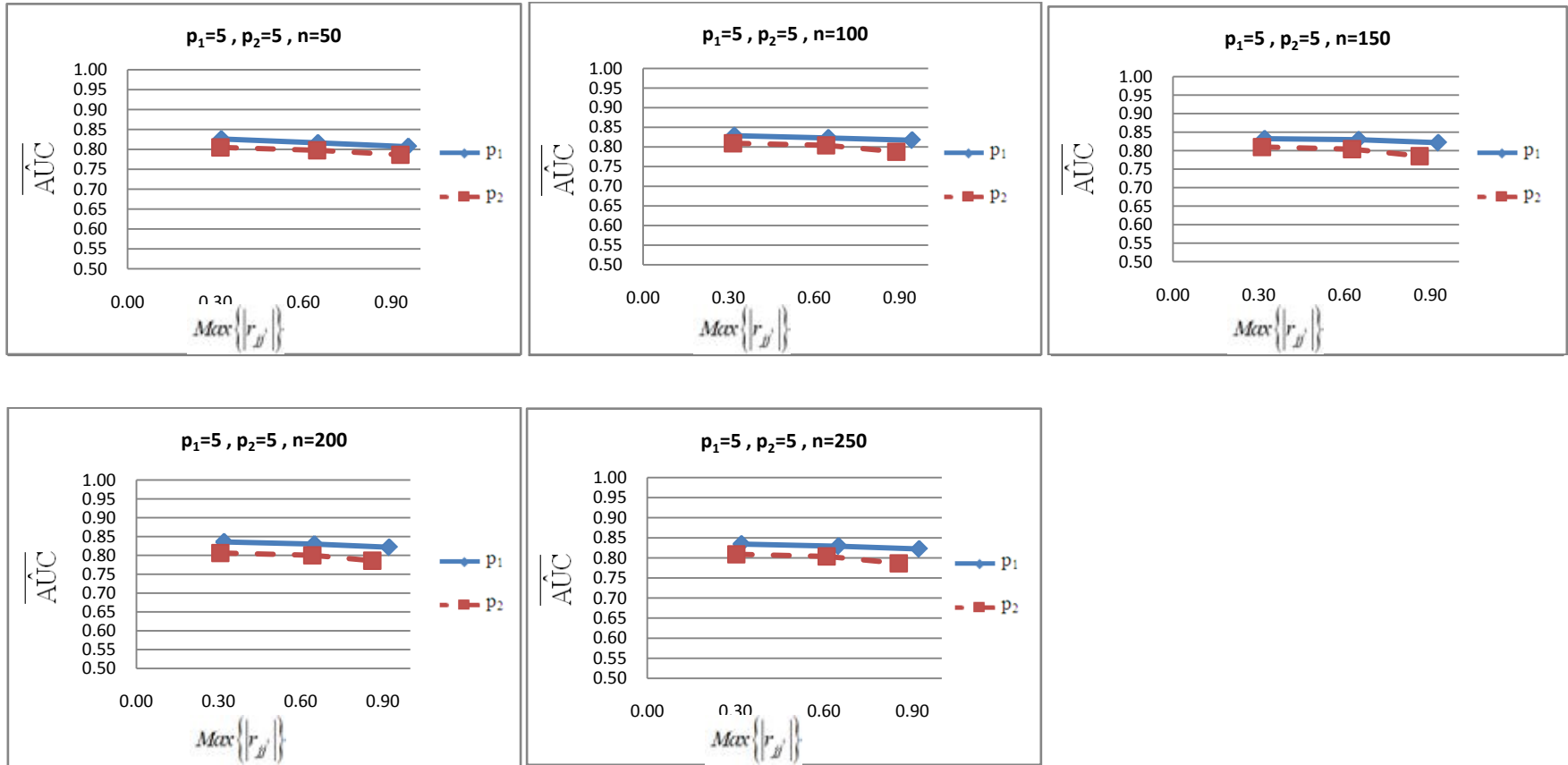


ตารางที่ 4.16 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ และการคัดเลือกตัวแบบที่เหมาะสม เมื่อ  $p_1=5$ ,  $p_2=4$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  และ  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High ตามลำดับ

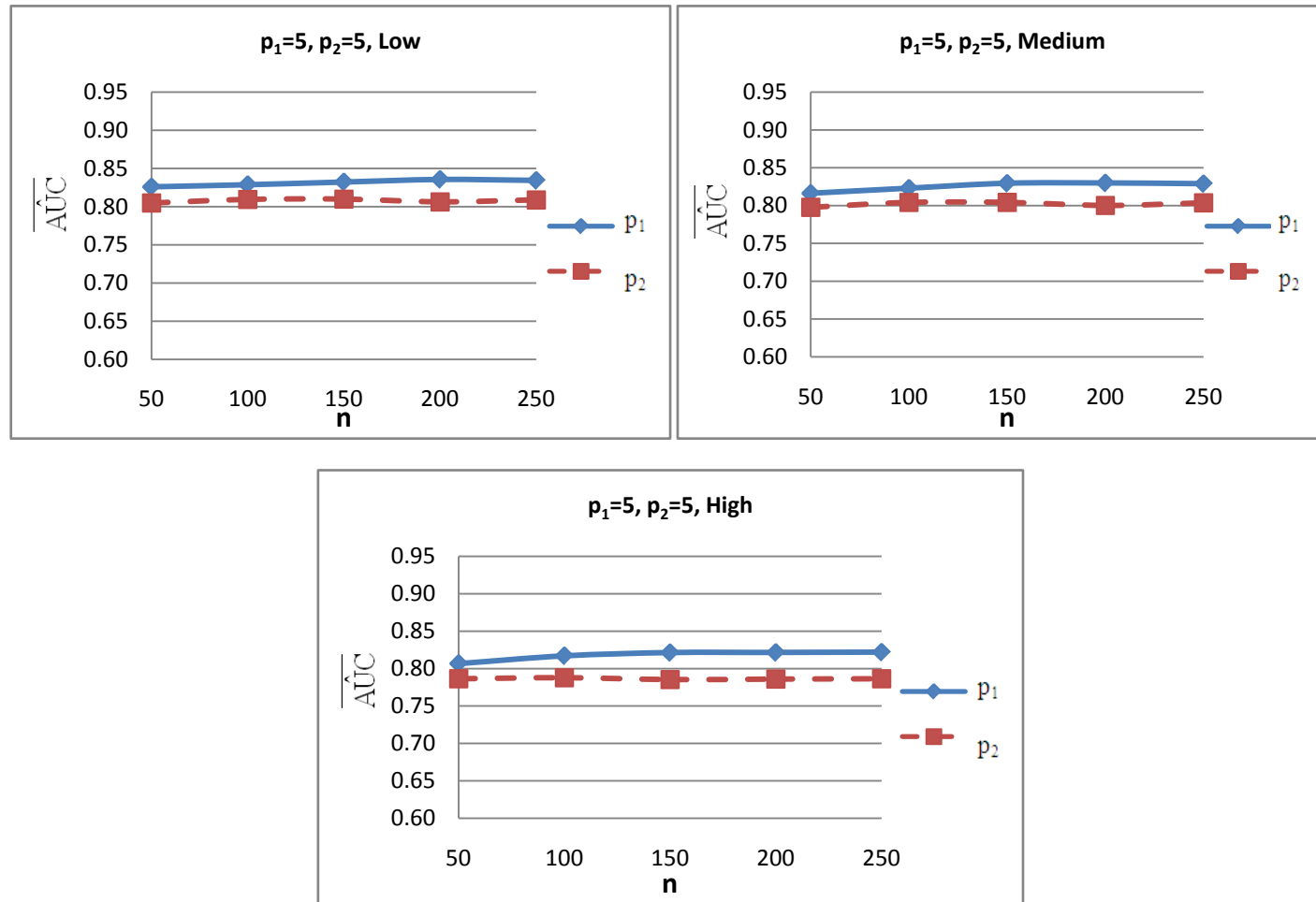
$p_1$	$p_2$	$n$	$Max\{r_{jj}\}$	$\overline{A\hat{U}C}_1$	$\overline{A\hat{U}C}_2$	Selected Model
5	5	50	Low	0.8260	0.8049	Model 1
			Medium	0.8162	0.7977	Model 1
			High	0.8067	0.7864	Model 1
		100	Low	0.8288	0.8095	Model 1
			Medium	0.8230	0.8042	Model 1
			High	0.8171	0.7878	Model 1
		150	Low	0.8323	0.8099	Model 1
			Medium	0.8296	0.8043	Model 1
			High	0.8214	0.7854	Model 1
		200	Low	0.8356	0.8063	Model 1
			Medium	0.8299	0.8002	Model 1
			High	0.8215	0.7860	Model 1
		250	Low	0.8344	0.8089	Model 1
			Medium	0.8290	0.8037	Model 1
			High	0.8220	0.7865	Model 1

จากตารางที่ 4.15 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่  $p_1=5$ ,  $p_2=5$ ,  $n=50$ ,  $n=100$ ,  $n=150$ ,  $n=200$ ,  $n=250$  แต่  $Max\{r_{jj}\}$  เปลี่ยนแปลงไปเป็น Low, Medium, High พบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มลดลง เมื่อระดับความสัมพันธ์มีค่าเพิ่มมากขึ้น และจากการทดสอบสมมติฐานเพื่อคัดเลือกตัวแบบที่มีค่าความถูกต้อง มากที่สุดคือ  $H_0 : \mu_{model1} \leq \mu_{model2}$  แยกกับ  $H_0 : \mu_{model1} > \mu_{model2}$  ผลการทดสอบพบว่าค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าตัวแบบที่ 2 ในทุกๆกรณี ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1 ( $p\text{-value} < 0.05$ )

ภาพที่ 4.31 แสดงค่าของ  $\overline{AUC}$  เมื่อ  $p_1=5, p_2=5, n=50, n=100, n=150, n=200, n=250$  และ  $Max\{r_{ij}\}$  เปลี่ยนแปลงไปเป็น Low, Medium และ High



ภาพที่ 4.32 แสดงค่าของ  $\overline{\hat{AUC}}$  เมื่อ  $p_1=5, p_2=5, \text{Max}\{r_{ij}\} = \text{Low, Medium, High}$  และขนาดตัวอย่างเปลี่ยนแปลงไปเป็น  $n = 50, 100, 150, 200, 250$



## บทที่ 5

### สรุปผลการวิจัย และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อต้องการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ที่เหมาะสมที่สุดสำหรับการพยากรณ์โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยงสำหรับแต่ละสถานการณ์ที่ต้องการศึกษา โดยใช้ พื้นที่ใต้โค้ง ROC (Receiver Operating Characteristic curve) เป็นเกณฑ์ในการคัดเลือก การจำลองข้อมูลในแต่ละสถานการณ์จะจำลองขึ้นโดยใช้เทคนิคมอนติคาร์โล ด้วยโปรแกรม R

สำหรับงานวิจัยนี้ ผู้วิจัยได้ทำการศึกษาการคัดเลือกตัวแบบภายใต้สถานการณ์ต่าง ๆ ที่สนใจศึกษาดังนี้

1. คู่ลำดับของจำนวนตัวแปรในตัวแบบแรก( $p_1$ ) และจำนวนตัวแปรในตัวแบบที่สอง ( $p_2$ ); ( $p_1, p_2$ ) ถูกกำหนดเป็น  
 $(p_1, p_2) = (2, 2), (2, 3), (2, 4), (2, 5),$   
 $(p_1, p_2) = (3, 2), (3, 3), (3, 4), (3, 5)$   
 $(p_1, p_2) = (4, 2), (4, 3), (4, 4), (4, 5)$   
 $(p_1, p_2) = (5, 2), (5, 3), (5, 4), (5, 5)$
2. กำหนดระดับความสัมพันธ์เริ่มต้นของตัวแปรอิสระเป็นดังนี้  
ตัวแปรอิสระมีความสัมพันธ์กันในระดับอย่างต่ำ:  $0 \leq \text{Max}\{|r_{jj'}|\} \leq 0.33$   
ตัวแปรอิสระมีความสัมพันธ์กันในระดับปานกลาง:  $0.33 \leq \text{Max}\{|r_{jj'}|\} \leq 0.66$   
ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูง:  $0.66 \leq \text{Max}\{|r_{jj'}|\} \leq 0.99$
3. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 50, 100, 150, 200 และ 250

การคัดเลือกตัวแบบข้างต้นจะพิจารณาจากค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ โดยใช้ค่าเฉลี่ย  $AUC$  (Area under the Curve) และการทดสอบสมมติฐานเพื่อเปรียบเทียบค่าเฉลี่ย ซึ่งเกณฑ์การคัดเลือกตัวแบบที่ให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ สูงจะเป็นเกณฑ์การคัดเลือกตัวแบบที่สามารถคัดเลือกตัวแบบได้ถูกต้องที่สุด

## 5.1 สรุปผลการวิจัย

5.1.1 ปัจจัยที่เกี่ยวข้องกับการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง ในแต่ละตัวแบบ

ปัจจัยที่เกี่ยวข้องกับการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง ที่ทำให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เปลี่ยนแปลงไปในแต่ละตัวแบบ โดยมีปัจจัยที่เกี่ยวข้องดังนี้

### ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ

ระดับความสัมพันธ์ของตัวแปรอิสระ ( $Max\{|r_{ij}|\}$ ) ในการศึกษาคั้งนี้คือ ระดับความสัมพันธ์ระดับต่ำ ระดับกลางและระดับสูง ซึ่งระดับความสัมพันธ์ระดับต่ำจะอยู่ในช่วง 0 ถึง 0.33 ระดับความสัมพันธ์ระดับกลางจะอยู่ในช่วง 0.33 ถึง 0.66 และระดับความสัมพันธ์ระดับสูงจะอยู่ในช่วง 0.66 ถึง 0.99 เกณฑ์การคัดเลือกตัวแบบพิจารณาจากค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีค่าสูงสุด ผลการศึกษาพบว่าเมื่อระดับความสัมพันธ์ในแต่ละคู่มีระดับต่ำ ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC จะมีค่าสูง เมื่อระดับความสัมพันธ์ในแต่ละคู่มีระดับปานกลางและสูง ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC จะมีค่าลดลง ซึ่งหมายความว่า เมื่อระดับความสัมพันธ์ในแต่ละคู่มีระดับต่ำ ตัวแบบจะมีความน่าเชื่อถือสูง แต่เมื่อระดับความสัมพันธ์อยู่ในระดับปานกลางและสูง ตัวแบบจะมีความน่าเชื่อถือลดลง ภายใต้ขนาดตัวอย่างและจำนวนตัวแปรอิสระอิสระของแต่ละตัวแบบคงที่

### จำนวนตัวแปรอิสระของแต่ละตัวแบบ

จำนวนตัวแปรอิสระของแต่ละตัวแบบ ในการศึกษาคั้งนี้คือ (2, 2), (2, 3), (2, 4), (2, 5), (3, 2), (3, 3), (3, 4), (3, 5), (4, 2), (4, 3), (4, 4), (4, 5), (5, 2), (5, 3), (5, 4), (5, 5) ผลการศึกษาพบว่า เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เพิ่มสูงขึ้น ซึ่งหมายความว่า ตัวแบบจะมีความน่าเชื่อถือมาก ภายใต้ขนาดตัวอย่างและระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระคงที่

### ขนาดตัวอย่าง

ขนาดตัวอย่างในการศึกษาคั้งนี้เท่ากับ 50, 100, 150, 200 และ 250 เกณฑ์การคัดเลือกตัวแบบพิจารณาจากพื้นที่ใต้โค้ง ROC มีค่าสูงสุด ผลการศึกษาพบว่าเมื่อขนาดตัวอย่างมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มเพิ่มสูงขึ้น ซึ่งหมายความว่า เมื่อขนาดตัวอย่างมีค่า

เพิ่มมากขึ้น ตัวแบบจะมีความน่าเชื่อถือเพิ่มมากขึ้น ภายใต้ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ และจำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่

ดังนั้นจึงสามารถสรุปได้ว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ ขนาดตัวอย่าง และจำนวนตัวแปรอิสระ เป็นปัจจัยที่มีผลต่อการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง

### 5.1.2 ผลการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง

ผลการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระเปลี่ยนแปลงไป สามารถแยกได้เป็น 3 กรณี ดังนี้

จำนวนตัวแปรอิสระในตัวแบบแรกน้อยกว่าจำนวนตัวแปรอิสระในตัวแบบที่สอง ( $p_1 < p_2$ )

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 เท่ากับ ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 2 เนื่องจากตัวแบบที่ 2 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 มากกว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 2 เนื่องจากตัวแบบที่ 2 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

จำนวนตัวแปรอิสระในตัวแบบแรกเท่ากับจำนวนตัวแปรอิสระในตัวแบบที่สอง ( $p_1 = p_2$ )

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 มากกว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

จำนวนตัวแปรอิสระในตัวแบบแรกเท่ากับจำนวนตัวแปรอิสระในตัวแบบที่สอง ( $p_1 \geq p_2$ )

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 เท่ากับ ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

## 5.2 ข้อเสนอแนะ

ผลการวิจัยในครั้งนี้มีข้อเสนอแนะ 2 ด้าน คือ

### 5.2.1 ด้านการนำไปใช้ประโยชน์

1. เพื่อเป็นแนวทางสำหรับนักวิจัยสาขาใดๆ ที่วางแผนจะใช้ตัวแบบโพรบิตแบบ
2. ประเภทเพื่อคัดเลือกตัวแบบที่ดีที่สุดแบบไม่ติดกลุ่มอย่างสมบูรณ์ ซึ่งจะสามารถประเมินและพยากรณ์ตัวแปรอิสระที่ถูกเลือก รวมไปถึงตัวแบบโดยดูจากจำนวนตัวแปรอิสระในแต่ละชุด, ขนาดตัวอย่าง และความสัมพันธ์ของแต่ละคู่ระหว่างตัวแปรอิสระ ซึ่งจะทำให้ นักวิจัยได้ตัวแบบที่ดีที่สุดเพื่อนำไปใช้ในการพยากรณ์หรือการวิจัยต่อไป

2. เมื่อต้องการหาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ ที่เหมาะสมสำหรับการพยากรณ์โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยง สามารถนำค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ นี้ไปใช้ได้ ตามแต่ละสถานการณ์ที่ต้องการศึกษา

### 5.2.2 ด้านการศึกษาวิจัย

1. ในการวิจัยครั้งนี้ได้ทำการศึกษาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ สำหรับการพยากรณ์โดยใช้ฟังก์ชันโพรบิตเป็นฟังก์ชันเชื่อมโยงเท่านั้น ในการวิจัยครั้งต่อไปอาจทำการศึกษาค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ โพรบิตแบบหลายกลุ่ม (Multinomial Probit Model)

2. ในการวิจัยครั้งนี้ได้ทำการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ ( Strictly non-nested) ในการวิจัยครั้งต่อไปอาจทำการศึกษาคัดเลือกตัวแบบติดกลุ่ม (Nested) หรือ ตัวแบบไม่ติดกลุ่มบางส่วน (Partially Non-Nested Model)

3. ในงานวิจัยนี้ใช้เกณฑ์การคัดเลือกตัวแบบ คือ พื้นที่ใต้โค้ง ROC ในการวิจัยครั้งต่อไปอาจใช้เกณฑ์การคัดเลือกตัวแบบอื่นๆ เช่น AIC และ BIC เป็นต้น



## รายการอ้างอิง

### ภาษาไทย

สุพล ดุรงค์วัฒนา. ตัวแบบและการวิเคราะห์ความถดถอยสำหรับการวิจัยขั้นสูง. กรุงเทพมหานคร : โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2549.

กัลยา วานิชย์บัญชา. ตัวแบบและการวิเคราะห์ความถดถอยสำหรับการวิจัยขั้นสูง. พิมพ์ครั้งที่ 1 โรงพิมพ์บริษัท ธรรมสาร จำกัด กรุงเทพมหานคร : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2548.

กุลพัชร เหมือนมา. การเปรียบเทียบการประมาณค่าพารามิเตอร์ของตัวแบบโลจิสต์ ตัวแบบโพรบิต และตัวแบบคอมพลีเมนต์ารีล็อก-ล็อก เมื่อตัวแปรตอบสนองมี 2 วิทยานิพนธ์ปริญญา มหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2551

บุญจิรา มาก้อน. การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยแบบไม่ติดกลุ่ม. วิทยานิพนธ์ปริญญา มหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2545

ปฐมาภรณ์ สาณกุล. อำนาจการทดสอบของตัวสถิติสำหรับพื้นที่ใต้โค้ง ROC บนตัวแบบโพรบิต. วิทยานิพนธ์ปริญญา มหาบัณฑิต, ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2552.

### ภาษาอังกฤษ

Alvin C. Rencher. 1995. Methods of Multivariate Analysis. New York :John Wiley

Richard A. Johnson and Dean W.Wichern. 2002. Applied Multivariate Statistical Analysis. 5thed.,United States of America: Prentice-Hal

D.W. Hosmer and S. Lemeshow, Applied Logistic Ression.New York, Second Edition,2000.

J.A.Hanley, and B.J.McNeil, "The meaning and Use of the Area under a Receiver Operating Characteristic(ROC) Curve."Radiologo,Vol.143,April 1982, pp.29-36

Lutz Hamel. Model Assessment with ROC Curves. Encyclopedia of Data Warehousing and Mining 2(2009) : 1316-1323.

Walter Zucchini.An Introduction to Model Selection. Journal of Mathematical Psychology 44(2000):41-61

## ประวัติผู้เขียนวิทยานิพนธ์

นางสาวจรรุตา ฤทธิเดชะ เกิดเมื่อวันที่ 11 พฤษภาคม พุทธศักราช 2529 สำเร็จการศึกษา  
ระดับปริญญาวิทยาศาสตรบัณฑิต สาขาสถิติ จากภาควิชาสถิติ คณะวิทยาศาสตร์และเทคโนโลยี  
มหาวิทยาลัยธรรมศาสตร์ ในปีการศึกษา 2551 และเข้าศึกษาต่อในหลักสูตร สถิติศาสตร  
มหาบัณฑิต สาขาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์  
มหาวิทยาลัย ในปีการศึกษา 2552