การคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง

นางสาวศริญณา มาปลูก

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

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

Miss Sarinna Maplook

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

หัวข้อวิทยานิพนธ์ การคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภทโดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง

โดย นางสาวศิริญญา มาปลูก

สาขาวิชา สถิติ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…………………………………………..คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

…………………………………………… ประธานกรรมการ

(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)

…………………………………………... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

…………………………………………… กรรมการ

(อาจารย์ ดร. อนันตณัฐ กันต์ธัญญรัตน์)

…………………………………………… กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร.บุญอ้อม โฉมที)

ศริญญา มาปลูก: การคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง (THE STRICTLY NON-NESTED MODEL SELECTION OF BINARY LOGISTIC REGRESSION MODEL USING LOGIT FUNCTION AS A LINK FUNCTION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก:รศ.ดร.สุพล ดุรงค์วัฒนา, 94 หน้า.

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง ปัจจัยที่ส่งผลต่อตัวแบบที่ถูกคัดเลือกคือ คู่ลำดับของตัวแปรในตัวแบบแรก (p_1) และจำนวนตัวแปรในตัวแบบที่สอง(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) ขนาดตัวอย่าง(n) คือ 50,100,150, 200 และ 250 และระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีความสัมพันธ์กันในระดับน้อย ปานกลาง และมาก ซึ่งข้อมูลทั้งหมดนี้ใช้การจำลองโดยเทคนิคมอนติคาร์โล ด้วยโปรแกรม R โดยใช้พื้นที่ใต้โค้ง ROC สูงสุดเป็นเกณฑ์ในการคัดเลือก ผลการวิจัยสรุปได้ดังนี้

กรณีพิจารณาในแต่ละตัวแบบ กรณีขนาดตัวอย่าง เมื่อขนาดตัวอย่างมีขนาดเพิ่มขึ้น ภายใต้ขอบเขตระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับเดียวกันและจำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่ ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC จะมีค่าเพิ่มขึ้นด้วย ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีค่าเพิ่มขึ้น ภายใต้ขอบเขตขนาดตัวอย่างและจำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่ ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC จะมีค่าลดลง กรณีพิจารณาจำนวนตัวแปรอิสระ เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น ภายใต้ขอบเขตระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับเดียวกันและขนาดตัวอย่างของแต่ละตัวแบบคงที่ จะทำให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เพิ่มสูงขึ้นด้วย ดังนั้นระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ ขนาดตัวอย่างและจำนวนตัวแปรอิสระ เป็นปัจจัยที่มีผลต่อการคัดเลือกตัวแบบ ผลการคัดเลือกตัวแบบ เมื่อจำนวนตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่าจำนวนตัวแปรอิสระในตัวแบบที่ 2 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 มากกว่าหรือเท่ากับระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 2 เมื่อจำนวนตัวแปรอิสระในตัวแบบที่ 1 มากกว่าจำนวนตัวแปรอิสระในตัวแบบที่ 2เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่าหรือเท่ากับระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1

ภาควิชา สถิติ ลายมือชื่อนิสิต...

สาขาวิชา สถิติ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก..

ปีการศึกษา 2554

5281904226 : MAJOR STATISTICS
KEYWORDS: BINARY LOGISTIC REGRESSION / STRICTLY NON-NESTED MODEL / ROC CURVE/ AREA UNDER THE CURVE / LOGIT

SARINNA MAPLOOK : THE STRICTLY NON-NESTED MODEL SELECTION OF BINARY LOGISTIC REGRESSION MODEL USING LOGIT FUNCTION AS A LINK FUNCTION. ADVISOR: ASSOC. PROF. SUPOL DURONGWATANA, Ph.D., 93 pp.

The purpose of this study is to select the strictly non-nested model of binary logistic regression model using logit 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 sample size (n); 50, 100, 150, 200 and 250 and the degree of pair-wise correlation independent variables of the levels low, medium and high. The data in all situations are generated using Monte Cario technique through R-program. The selection criterion is the maximum of the area under ROC curve. The results are summarized as follow:

As sample size changed but the other factors are kept constant, when the sample size increases, area under the curve increased. As the degree of pair-wise correlation of variables but the other factors are kept constant, when the degree of pair-wise correlation of variables increases, area under the curve decreased. As the number of independent variables changed but the other factors are kept constant, when the number of independent variables increases, area under the curve increased. Therefore, sample size, the degree of pair-wise correlation of variables and the number of independent variables affect the selection of the strictly non-nested binary logistic regression model.

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

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

Academic Year : 2011

กิตติกรรมประกาศ

งานวิจัยครั้งนี้สำเร็จลุล่วงไปด้วยความช่วยเหลืออย่างดียิ่งจาก รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำ ปรึกษา ตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่างๆ จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณและสำนึกในพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ธีระพร วีระถาวร ในฐานะประธานกรรมการ อาจารย์ ดร.อนันตณัฐ กันต์ธัญญรัตน์ กรรมการสอบวิทยานิพนธ์ ที่ท่านช่วยเหลือ รวมถึงคำแนะนำในการทำวิจัยครั้งนี้ ซึ่งทำให้วิทยานิพนธ์มีความสมบูรณ์

ผู้วิจัยขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.บุญอ้อม โฉมที ที่ท่านได้เสียสละเวลาอันมีค่ามาเป็นกรรมการภายนอกมหาวิทยาลัย ซึ่งทำให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น และขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ให้โอกาสทางการศึกษาและประสิทธิประสาทความรู้ให้แก่ผู้วิจัยกระทั่งสำเร็จการศึกษา

ท้ายนี้ผู้วิจัยใคร่ขอกราบขอบคุณ บิดา มารดา ที่ให้การส่งเสริม สนับสนุนด้านทุนการศึกษา ให้ความรักและกำลังใจเสมอมาจนสำเร็จการศึกษา ตลอดจน พี่ๆ เพื่อนๆทุกคนที่ส่งเสริมและให้กำลังใจแก่ผู้วิจัยมาตลอด

สารบัญ

หน้า

หน้า

สารบัญตาราง

ตารางที่ หน้า

ตารางที่ หน้า

ตารางที่ หน้า

สารบัญภาพ

ภาพที่ หน้า

ภาพที่		หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันงานวิจัยด้านต่างๆเกือบทุกด้านทั้งทางด้านวิทยาศาสตร์ เศรษฐศาสตร์ สังคมศาสตร์ การแพทย์ การเงิน ได้นำเทคนิคการพยากรณ์มาใช้ด้วยกันหลายวิธี ตามลักษณะของข้อมูลและจุดประสงค์ของการพยากรณ์ วิธีการหนึ่งที่เป็นที่นิยมในปัจจุบัน คือ การพยากรณ์ด้วยตัวแบบการถดถอยโลจิสติก จะมีทั้งตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท (Binary Logistic Regression Model) และตัวแบบการถดถอยโลจิสติกแบบเชิงพหุ (Multiple Logistic Regression Model) ซึ่งเป็นการวิเคราะห์ข้อมูลโดยการหาตัวแบบความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ เพื่อนำตัวแบบที่ได้นำไปใช้ในการพยากรณ์ โดยที่ตัวแปรตามเป็นข้อมูลเชิงคุณภาพ ส่วนตัวแปรอิสระอาจเป็นข้อมูลเชิงปริมาณหรือเชิงคุณภาพอย่างใดอย่างหนึ่งหรือทั้งสองอย่าง บางครั้งการวิจัยในด้านต่างๆ เช่น ทางด้านการแพทย์ หรือทางด้านเศรษฐศาสตร์ จำเป็นต้องพิจารณาหาตัวแบบที่เหมาะสมที่สุดสำหรับการจำแนกกลุ่ม

ตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท คือตัวแปรตามสามารถแบ่งได้เป็น 2 กลุ่ม คือ กลุ่มที่เราสนใจ และกลุ่มที่เราไม่สนใจ ซึ่งเป็นตัวแบบที่นิยมใช้กันมาก เช่น ในทางด้านธุรกิจ คือ ความพึงพอใจในสินค้า(พอใจ ไม่พอใจ) การตัดสินใจในการซื้อสินค้า (ซื้อ ไม่ซื้อ) ในการวิเคราะห์การถดถอยโลจิสติกแบบ 2 ประเภทนั้นจะมีการใช้ฟังก์ชันการเชื่อมโยง(Link function) เพื่อเปลี่ยนตัวแบบให้ง่ายขึ้นสำหรับการวิเคราะห์ เช่นฟังก์ชันโลจิท (Logit Function) ฟังก์ชันโพรบิต (Probit Function) และฟังก์ชันคอมพลีเมนทารี ล็อก-ล็อก (Complementary log-log Function) เป็นต้น การที่เราจะเลือกใช้ฟังก์ชันใดเป็นฟังก์ชันเชื่อมโยงนั้น จะพิจารณาจากลักษณะของข้อมูล ในวิจัยงานนี้ เราสนใจตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง เพราะว่าฟังก์ชันโลจิทนั้นจะใช้กับข้อมูลที่มีการกระจายเท่ากัน และมีค่าคลาดเคลื่อน (error terms) มีการกระจายแบบโลจิสติก ซึ่งคนส่วนใหญ่นิยมใช้ฟังก์ชันโลจิท เพราะว่ามีวิธีการคำนวณที่ซับซ้อนน้อยกว่าฟังก์ชันการเชื่อมโยงอื่นๆ การแปลความหมายของฟังก์ชันง่ายกว่า

เพราะฟังก์ชันโลจิทเป็นโอกาสของการเกิดเหตุการณ์ที่สนใจต่อเหตุการณ์ที่ไม่สนใจ ซึ่งทำให้ง่ายต่อการแปลความหมาย

ขั้นตอนที่สำคัญที่สุดอีกขั้นตอนหนึ่งในการพยากรณ์คือการเลือกตัวแบบที่มีประสิทธิภาพ การวัดประสิทธิภาพของตัวแบบสามารถทำได้โดยการวัดอัตราความถูกต้องของการพยากรณ์ ซึ่งมีหลายวิธีที่นิยมใช้ วิธีหนึ่งที่ได้รับความนิยมคือ การใช้ Receiver Operating Characteristic curve หรือ ROC curve โดยค่าที่ถูกใช้เป็นตัวแบบในการบอกความถูกต้องของการพยากรณ์คือพื้นที่ใต้โค้ง ROC โดยที่ถ้าพื้นที่ใต้โค้ง ROC มีค่ามาก จะหมายถึงว่า ตัวแบบนั้นมีอัตราความถูกต้องมากเช่นกัน

Lutz Hamel (2009) ได้ทำการศึกษาถึงการคัดเลือกตัวแบบโดยใช้ ROC curve ผลการศึกษากล่าวว่า การใช้ ROC curve เป็นวิธีการที่มีประสิทธิภาพ มีความถูกต้องแม่นยำมากกว่าทำให้มีความเชื่อถือได้สูง และเป็นวิธีการที่ไม่ซับซ้อน

ผู้วิจัยสนใจที่จะทำการศึกษาว่าตัวแบบทางสถิติแบบไม่ติดกลุ่มแบบใดที่จะทำให้ ค่าพื้นที่ใต้เส้นโค้ง ROC สูงสุด โดยพิจารณาจากขนาดตัวอย่างและจำนวนตัวแปรอิสระ ถ้าขนาดตัวอย่างและจำนวนตัวแปรอิสระเท่ากัน เซตของจำนวนอิสระเท่าใดจึงเหมาะสมที่ถูกเลือกเข้ามาอยู่ในตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันการเชื่อมโยงโลจิท หรือในกรณีที่ถ้าจำนวนตัวแปรอิสระหรือขนาดตัวอย่างแตกต่างกันแล้ว สาเหตุใดที่ทำให้เราเลือกตัวแบบนั้นเข้ามาอยู่ในตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อศึกษาระดับความสัมพันธ์ในแต่ละคู่ (degree of pair -wise correlation) ของตัวแปรอิสระจากทั้งสองตัวแบบที่ส่งผลให้การเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภทโดยใช้ฟังก์ชันการเชื่อมโยงโลจิท เมื่อขนาดตัวอย่างเท่ากัน จำนวนของตัวแปรอิสระเท่ากัน โดยใช้พื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือก

1.2.2 เพื่อศึกษาหาจำนวนตัวแปรอิสระที่ส่งผลให้การเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภทโดยใช้ฟังก์ชันการเชื่อมโยงโลจิท เมื่อขนาดตัวอย่างเท่ากัน และมีระดับความสัมพันธ์ในแต่ละคู่ (degree of pair -wise correlation)อยู่ในระดับเดียวกัน โดยใช้พื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือก

1.3 ขอบเขตของการวิจัย

การวิจัยครั้งนี้มีขอบเขตของการวิจัยสำหรับการดำเนินวิจัย ดังนี้

1.3.1 ทำการศึกษาตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภทโดยใช้ฟังก์ชันการเชื่อมโยงโลจิท เพื่อหาตัวแบบที่เหมาะสมที่สุดสำหรับการจำแนกกลุ่มโดยการเปรียบเทียบทีละ 2 ตัวแบบ

1.3.2 ตัวแปรตาม (Y) เป็นข้อมูลเชิงคุณภาพที่มี 2 ค่า คือ 0 และ 1

1.3.3 ตัวแบบทั้งสองที่นำมาเปรียบเทียบกัน จำนวนตัวแปรอิสระ 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) ในการศึกษาครั้งนี้ คือ 50,100,150,200,250

1.3.5 ตัวแปรอิสระ (X) มีการแจกแจงเริ่มต้น เป็นการแจกแจงแบบยูนิฟอร์ม

1.3.6 กำหนดค่าพารามิเตอร์เริ่มต้นของสมการถดถอยเป็นค่าใดๆในการศึกษาครั้งนี้ คือ $\beta_i = 0.1; i = 0,1,2,..., p_1 + p_2$

1.3.7 กำหนดให้ความคลาดเคลื่อน(ε) มีการแจกแจงแบบโลจิท

1.3.8 ใช้ค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ j และตัวแปรอิสระตัวที่ j’ ในการวัดระดับความสัมพันธ์ระหว่างตัวแปรอิสระ

$$Max\{|r_{jj'}|\}$$

โดยที่ $r_{jj'}$ คือ สหสัมพันธ์ระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ระดับความสัมพันธ์ระหว่างตัวแปรอิสระแบ่งเป็น 3 ระดับ ดังนี้

- ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับต่ำ เมื่อ 0 < $Max\{|r_{jj'}|\} \leq$ 0.33
- ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับปานกลางเมื่อ 0.33< $Max\{|r_{jj'}|\} \leq$ 0.66
- ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับสูง เมื่อ 0.66 < $Max\{|r_{jj'}|\} \leq$ 0.99

1.3.9 กำหนดระดับนัยสำคัญ (α) ในการศึกษาครั้งนี้ที่ระดับ 0.05 $(\alpha = 0.05)$

1.3.10 ในการศึกษาครั้งนี้ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation) โดยการจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 500 รอบ (N=500)

1.4 วิธีดำเนินการวิจัย

1.4.1 ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับงานวิจัย

1.4.2 จำลองข้อมูลตามขอบเขตที่ต้องการศึกษา

1.4.3 หาระดับความสัมพันธ์ในแต่ละคู่ (degree of pair -wise correlation) ระหว่างแต่ละคู่ของตัวแปรอิสระจากตัวแบบโลจิสติกแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างแท้จริง

1.4.4 นำตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันการเชื่อมโยงโลจิท ที่ได้ไปพล็อตโค้ง ROC และคำนวณหาค่าประมาณพื้นที่ใต้โค้ง ROC

1.4.5 เปรียบเทียบพื้นที่ใต้โค้ง ROC เพื่อหาตัวแบบที่ดีที่สุด

1.4.6 สรุปผลที่ได้จากการวิจัย

1.5 เกณฑ์ที่ใช้ในการตัดสินใจ

ตัวแบบโลจิสติกแบบ 2 ประเภทแบบไม่ติดกลุ่มอย่างแท้จริงที่นำไปพล็อต ROC curve แล้วได้พื้นที่ใต้โค้ง ROC มากที่สุด ถือว่าเป็นตัวแบบที่ดีที่สุด

ถ้า $A\hat{U}C = 0.50$ ถือเป็นตัวแบบที่มีความเชื่อถือได้น้อย ไม่สามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้

ถ้า $0.70 \leq A\hat{U}C < 0.80$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้พอใช้

ถ้า $0.80 \le A\hat{U}C < 0.90$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ดี โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้ดี

ถ้า $A\hat{U}C \ge 0.90$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ในระดับดีมาก โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้อย่างชัดเจน

1.6 คำจำกัดความที่ใช้ในการวิจัย

1.6.1 ตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท (Binary Logistic Regression Model) หมายถึงตัวแบบที่ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหลายตัวกับตัวแปรตามซึ่งเป็นตัวแปรคุณภาพที่มีค่า 2 ค่า ส่วนตัวแปรอิสระอาจเป็นข้อมูลเชิงปริมาณหรือเชิงคุณภาพอย่างหนึ่งหรือทั้งสองอย่างก็ได้ ซึ่งจะนำไปใช้ในการจำแนกกลุ่มตามโอกาสที่แต่ละหน่วยจะอยู่ในกลุ่มใดกลุ่มหนึ่ง

1.6.2 ฟังก์ชันโลจิท (logit function) หมายถึง ฟังก์ชันที่แปลงฟังก์ชันโลจิสติก ซึ่งแปลงค่าความน่าจะเป็นจากช่วง (0,1) เป็นสมการ logit(π_i) ในช่วง $(-\infty, \infty)$

1.6.3 ความสัมพันธ์ระหว่างตัวแปรอิสระ (multicollinearity) หมายถึง สถานการณ์ที่ตัวแปรอิสระมีความสัมพันธ์กัน

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

1.6.5 ค่าประมาณพื้นที่ใต้โค้ง ROC $(A\hat{U}C)$ คือ ดัชนีในการบ่งชี้ความสามารถในการจำแนกกลุ่มหรือความเชื่อถือได้ของตัวแบบ

1.6.6 การแจกแจงแบบเบอร์นูลลี (Bernoulli Distribution) หมายถึง ตัวแปรสุ่มY เป็นตัวแปรสุ่มเบอร์นูลลีคือ $Y \sim Ber(p)$ มีฟังก์ชันการแจกแจงความน่าจะเป็นจะอยู่ในรูป

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

Y=1 ถ้าการทดลองเกิดเหตุการณ์ที่สนใจ (ความสำเร็จ) โดยที่ $P(Y=1) = p$

Y=0 ถ้าการทดลองไม่เกิดเหตุการณ์ที่สนใจ (ความล้มเหลว)

โดยที่ $P(Y=0) = 1-p$

1.6.7 ฟังก์ชันเชื่อมโยง (link function) คือ ส่วนที่ใช้อธิบายความสัมพันธ์ระหว่างองค์ประกอบเกี่ยวกับตัวแปรสุ่มและองค์ประกอบเชิงระบบเป็นการเชื่อมโยงระหว่างส่วนตัวแปรสุ่มและส่วนเชิงระบบ

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 เพื่อเป็นแนวทางในการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภทโดยใช้ฟังก์ชันการเชื่อมโยงโลจิท สำหรับไปใช้ในการจำแนกกลุ่มซึ่งจะสามารถหามาจากปัจจัยตัวแปรอิสระที่ถูกคัดเลือก ขนาดตัวอย่างและความสัมพันธ์ของแต่ละคู่ระหว่างตัวแปรอิสระ

1.7.2 เพื่อใช้เป็นแนวทางในการศึกษาวิจัยต่อไป

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 ตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภทและการประมาณค่า

ตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภทสำหรับการพยากรณ์ในการจัดประเภท ดังนี้

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

$$\tilde{X}_i^T = (1 \quad X_{1i} \quad X_{2i} \quad \ldots \quad X_{pi}), \tilde{\beta}^T = (\beta_0 \quad \beta_1 \quad \beta_2 \quad \ldots \quad \beta_p)$$

เมื่อ

$$\Pr\left(Y_i = 1 \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = \pi_i , \Pr\left(Y_i = 0 \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = 1 - \pi_i ; 0 < \pi_i < 1$$

เมื่อ Y_i มีการแจกแจงแบบเบอร์นูลลีที่เป็นอิสระกัน มีค่าเฉลี่ยและความแปรปรวน ดังนี้

$$E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = \tilde{x}_i^T \tilde{\beta} , Var\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = \pi_i (1 - \pi_i); i = 1,2,\ldots,n$$

ถ้าตัวแบบการถดถอยเป็น

$$Y_i = \tilde{X}_i^T \tilde{\beta} + \varepsilon_i , E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = \tilde{x}_i^T \tilde{\beta}; i = 1,2,\ldots,n$$

เมื่อ $E\left(\varepsilon_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = 0; i = 1,2,\ldots,n$

ฟังก์ชั่นโลจิทคือ

$$Logit\left(E\left[Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right]\right) = \ln\left(\frac{E\left[Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right]}{1 - E\left[Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right]}\right) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right)$$

ซึ่งฟังก์ชันโลจิทจะทำการแปลงค่า π_i จากช่วง (0,1) เป็นค่าที่อยู่ในช่วง$(-\infty, \infty)$ โดยฟังก์ชันผกผันของฟังก์ชันโลจิท จะแปลงค่า $\tilde{x}_i^T \tilde{\beta}$ เป็นค่าที่อยู่ในช่วง [0,1] ซึ่งเรียกว่า ฟังก์ชันโลจิสติก (logistic function) ซึ่งจะได้ตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท ดังนี้

$$\left(E\left[Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right]\right) = \frac{\exp\left(\tilde{x}_i^T \tilde{\beta}\right)}{1 + \exp\left(\tilde{x}_i^T \tilde{\beta}\right)}; i = 1,2,\ldots,n$$

ตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภทแปลงเป็นเชิงเส้นได้ดังนี้

$$\ln\left(\frac{E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right)}{1-E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right)}\right) = \tilde{x}_i^T \tilde{\beta}; i = 1,2,\ldots,n$$

$$Y_i^* = \tilde{X}_i^T \tilde{\beta} + \varepsilon_i$$

$$Y_i^* = \ln\left(\frac{E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right)}{1-E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right)}\right) + \varepsilon_i; \varepsilon_i \sim N\left(0,\sigma^2\right); i = 1,2,\ldots,n$$

ฟังก์ชั่นภาวะน่าจะเป็น (Likelihood function)

เมื่อแต่ละค่าสังเกต Y_i คือตัวแปรสุ่มแบบเบอร์นูลลี ซึ่ง

$$P\left(Y_i = 1 \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = \pi_i, \; P\left(Y_i = 0 \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = 1-\pi_i; i = 1,2,\ldots,n$$

ซึ่งมีการแจกแจงความน่าจะเป็นคือ

$$f_i\left(Y_i = y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = \pi_i^{y_i}\left(1-\pi_i\right)^{1-y_i}, y_i = 0,1; i = 1,2,\ldots,n$$

ฟังก์ชั่นความน่าจะเป็นร่วมคือ

$$g\left(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n\right) = \prod_{i=1}^{n} f_i(Y_i = y_i) = \prod_{i=1}^{n} \pi_i^{y_i}\left(1-\pi_i\right)^{1-y_i}$$

ลอการิทึมธรรมชาติของฟังก์ชั่นความน่าจะเป็นร่วมคือ

$$\begin{aligned}\ln g\left(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n\right) &= \ln\left\{\prod_{i=1}^{n} \pi_i^{y_i}\left(1-\pi_i\right)^{1-y_i}\right\} \\ &= \sum_{i=1}^{n}\left[y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right)\right] + \sum_{i=1}^{n}\ln\left(1-\pi_i\right)\end{aligned}$$

เมื่อ

$$E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right) = \pi_i, 1-\pi_i = \frac{1}{1+\exp\left(\tilde{x}_i^T \tilde{\beta}\right)}$$

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \tilde{x}_i^T \tilde{\beta}; i = 1,2,\ldots,n$$

ดังนั้น ลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น (log-likelihood function) คือ

$$\ln L\left(\tilde{\beta}^T\right)=\sum_{i=1}^{n} y_i\left(\tilde{x}_i^T\tilde{\beta}\right)-\sum_{i=1}^{n}\ln\left[1+\exp\left(\tilde{x}_i^T\tilde{\beta}\right)\right]$$

ถัดมา ฟังก์ชั่นภาวะน่าจะเป็น คือ

$$l\left(\tilde{\beta}^T\middle|Y_i=y_i,\tilde{X}_i^T=\tilde{x}_i^T;i=1,2,...,n\right)=\prod_{i=1}^{n}\left[\pi_i\left(\tilde{x}_i^T\right)\right]^{y_i}\left[1-\pi_i\left(\tilde{x}_i^T\right)\right]^{1-y_i}$$

ดังนั้นลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น คือ

$$\ln\left\{l\left(\tilde{\beta}^T\middle|Y_i=y_i,\tilde{X}_i^T=\tilde{x}_i^T;i=1,2,...,n\right)\right\}=L\left(\tilde{\beta}^T\right)=\ln\left\{\prod_{i=1}^{n}\left[\pi_i\left(\tilde{x}_i^T\right)\right]^{y_i}\left[1-\pi_i\left(\tilde{x}_i^T\right)\right]^{1-y_i}\right\}$$

$$=\sum_{i=1}^{n}\left[y_i\ln\left(\frac{\pi_i\left(\tilde{x}_i^T\right)}{1-\pi_i\left(\tilde{x}_i^T\right)}\right)\right]+\sum_{i=1}^{n}\ln\left[1-\pi_i\left(\tilde{x}_i^T\right)\right]$$

เมื่อ

$$E\left(Y_i\middle|\tilde{X}_i^T=\tilde{x}_i^T\right)=\pi_i,1-\pi_i=\frac{1}{1+\exp\left(\tilde{x}_i^T\tilde{\beta}\right)}$$

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right)=\tilde{x}_i^T\tilde{\beta};i=1,2,...,n$$

ดังนั้นฟังก์ชั่นภาวะน่าจะป็นคือ

$$L\left(\tilde{\beta}^T\right)=\sum_{i=1}^{n} y_i\left(\tilde{x}_i^T\tilde{\beta}\right)-\sum_{i=1}^{n}\ln\left[1+\exp\left(\tilde{x}_i^T\tilde{\beta}\right)\right];i=1,2,...,n$$

เมื่อเราทราบการแจกแจงของ Y เราจะสามารถสร้างฟังก์ชันภาวะน่าจะเป็นได้ เนื่องจากการประมาณค่าไม่ได้เป็นไปตามรูปแบบ เราจึงต้องใช้วิธีการประมาณเชิงตัวเลข (จำเป็นต้องทำซ้ำเพื่อให้ได้มาซึ่งตัวประมาณภาวะน่าจะเป็นสูงสุด (MLE)) โดยในการทำซ้ำ 1 ครั้งจะได้ตัวประมาณภาวะน่าจะเป็นสูงสุด (b_0 b_1 b_2 … b_p) แล้วทำการคำนวณค่าประมาณของ π_i=1,2,…,n ดังนี้

$$\hat{\pi}_i=\frac{\exp\left(\tilde{x}_i^T\tilde{b}\right)}{1+\exp\left(\tilde{x}_i^T\tilde{b}\right)};i=1,2,...,n$$

เมื่อประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดในตัวแบบการถดถอยโลจิสติก แบบ 2 ประเภทแล้วจะสามารถนำไปใช้ในการพยากรณ์การจำแนกกลุ่มของตัวแบบ ดังนี้

- หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มของความสำเร็จของลักษณะที่สนใจศึกษา (Y=1) ถ้า

$$\hat{\pi}_i = \frac{\exp\left(\tilde{x}_i^T \tilde{b}\right)}{1+\exp\left(\tilde{x}_i^T \tilde{b}\right)} \geq c \text{ where } 0 \leq c \leq 1$$

- หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มของความล้มเหลวของลักษณะที่สนใจศึกษา (Y=0) ถ้า

$$\hat{\pi}_i = \frac{\exp\left(\tilde{x}_i^T \tilde{b}\right)}{1+\exp\left(\tilde{x}_i^T \tilde{b}\right)} < c \text{ where } 0 \leq c \leq 1$$

เมื่อ c คือ จุดแบ่งหรือระดับของความน่าจะเป็นที่ใช้ในการพิจารณาการจำแนกกลุ่มว่าแต่ละหน่วยจะอยู่ในกลุ่มใดระหว่างกลุ่มของความสำเร็จและกลุ่มของความล้มเหลวของลักษณะที่สนใจศึกษา

2.2 **ตัวแบบไม่ติดกลุ่ม** (Non-Nested Model)

ตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท เป็นตัวแบบไม่ติดกลุ่มไม่ว่าจะเป็นเพียงบางส่วนหรืออย่างแท้จริง ถ้าหนึ่งตัวแบบไม่สามารถลดตัวแบบอื่น ๆได้โดยกำหนดการเชิงเส้นซึ่งมีข้อจำกัดบนเวกเตอร์พารามิเตอร์ เช่น

$$M_1: ln\left(\frac{E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right)}{1-E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right)}\right) = \tilde{x}_i^T \tilde{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}; i = 1,2,\ldots,n$$

$$M_2: ln\left(\frac{E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right)}{1-E\left(Y_i \middle| \tilde{X}_i^T = \tilde{x}_i^T\right)}\right) = \tilde{x}_i^T \tilde{\beta} = \beta_0 + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i}; i = 1,2,\ldots,n$$

ตัวแบบทั้งสองเป็นตัวแบบไม่ติดกลุ่ม เพราะแม้ว่าเราจะกำหนดข้อจำกัดที่ $\beta_4 = 0$, $\beta_5 = 0$ M_2จะไม่กลายเป็น 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

$$M_1: ln\left(\frac{E\left(Y_i\middle|\tilde{X}_i^T = \tilde{x}_i^T\right)}{1 - E\left(Y_i\middle|\tilde{X}_i^T = \tilde{x}_i^T\right)}\right) = \tilde{x}_i^T\tilde{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}; i = 1,2,\dots,n$$

$$M_2: ln\left(\frac{E\left(Y_i\middle|\tilde{X}_i^T = \tilde{x}_i^T\right)}{1 - E\left(Y_i\middle|\tilde{X}_i^T = \tilde{x}_i^T\right)}\right) = \tilde{x}_i^T\tilde{\beta} = \beta_0 + \beta_4 x_{4i} + \beta_5 x_{5i}; i = 1,2,\dots,n$$

ถ้าให้พารามิเตอร์บางตัวมีค่าเป็นศูนย์ จะเห็นว่าตัวแบบ M_2 ไม่สามารถลดรูปไปเป็นตัวแบบ M_1ได้ ซึ่งจะเรียกตัวแบบนี้ว่าเป็นตัวแบบ "ไม่ติดกลุ่มอย่างแท้จริง(Strictly Non-nested Model)"

2.3 โค้ง ROC (Receiver Operating Characteristic curve หรือ ROC curve)

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

หากใช้จุดตัด ซึ่งเป็นตำแหน่งตรงเส้นตรงเป็นเกณฑ์ในการจำแนกเหตุการณ์ออกเป็นกลุ่มของเหตุการณ์ที่ไม่สนใจ และกลุ่มของเหตุการณ์ที่สนใจ พบว่า

- ในเหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นบวก ผลดังกล่าวเป็นผลบวกจริง หรือเรียกว่า true positive (TP)
- ในเหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นบวก ผลดังกล่าวเป็นผลบวกลวง หรือเรียกว่า false positive (FP)

- ในเหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นลบ ผลดังกล่าวเป็นผลลบจริง หรือเรียกว่า true negative (TN)
- ในเหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นลบ ผลดังกล่าวเป็นผลลบลวง หรือเรียกว่า false negative (FN)

ค่าพยากรณ์ / ค่าสังเกต	จำนวนเหตุการณ์ที่ให้ผลบวก $\hat{y}_i = 1$	จำนวนเหตุการณ์ที่ให้ผลลบ $\hat{y}_i = 0$	รวม
จำนวนเหตุการณ์ที่สนใจ $y_i = 1$	TP	FN	TP+FN
จำนวนเหตุการณ์ที่ไม่สนใจ $y_i = 0$	FP	TN	FP+TN

เครื่องมือที่ใช้วัดความถูกต้องของการพยากรณ์ซึ่งผลที่เกิดจากการพยากรณ์มีเพียง 2 ค่า คือ

-Sensitivity (SN) คือสัดส่วนของเหตุการณ์ที่ให้ผลการพยากรณ์เป็นผลบวกจริง (TP) ต่อจำนวนเหตุการณ์ที่สนใจทั้งหมด นั่นคือ

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

-Specificity (SP) คือสัดส่วนของเหตุการณ์ที่ให้ผลการพยากรณ์เป็นผลลบจริง (TN) ต่อจำนวนเหตุการณ์ที่ไม่สนใจทั้งหมด นั่นคือ

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

จะเห็นได้ว่า ค่า Sensitivity จะขึ้นอยู่กับจำนวนเหตุการณ์ที่สนใจทั้งหมด ส่วนค่า Specificity ก็จะขึ้นอยู่กับจำนวนเหตุการณ์ที่ไม่สนใจทั้งหมด

ROC curve เป็นกราฟที่พล็อตระหว่างค่าของ Sensitivity และ 1-Specificity ซึ่งการพล็อตกราฟจะได้จากการกำหนดจุดตัด ที่ระดับต่างๆ เพื่อแบ่งผลลัพธ์ของการพยากรณ์ออกเป็น 2 กลุ่ม คือกลุ่มที่เกิดเหตุการณ์ $P(Y=1) \geq$ จุดตัด และกลุ่มที่ไม่เกิดเหตุการณ์ $P(Y=1) <$ จุดตัด

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

2.3.1 **พื้นที่ใต้โค้ง** ROC (Area Under the Curve **หรือ** $A\hat{U}C$)

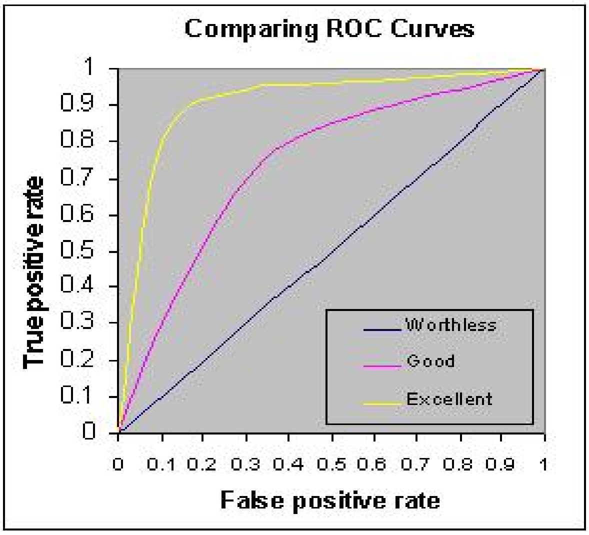
เนื่องจาก ROC curve มักถูกนำไปใช้ในการตรวจสอบความถูกต้องของตัวแบบในการพยากรณ์ โดยค่าที่ใช้เป็นดัชนีในการบ่งชี้ความถูกต้องหรือความเชื่อถือได้ของตัวแบบ คือ ค่าประมาณพื้นที่ใต้โค้ง ROC (Area Under the Curve หรือ $A\hat{U}C$)

การคำนวณค่า $A\hat{U}C$ สามารถทำได้ 2 วิธี ซึ่งวิธีแรกคือวิธีที่ไม่ใช้พารามิเตอร์จะขึ้นอยู่กับการสร้างการประมาณรูปสี่เหลี่ยมที่มีสองด้านขนานกันใต้เส้นโค้ง ซึ่งเป็นการประมาณพื้นที่ วิธีที่สองคือวิธีการประมาณค่าพารามิเตอร์โดยใช้ตัวประมาณค่าความควรจะเป็นสูงสุด (maximum likelihood estimator) เพื่อให้เส้นโค้งเรียบพอดีกับจุดข้อมูล โดยทั้งสองวิธีใช้ได้กับโปรแกรมคอมพิวเตอร์และสามารถหาค่าประมาณของพื้นที่และความคลาดเคลื่อนมาตรฐานได้ การประมาณค่าโดยใช้รูปสี่เหลี่ยมที่มีสองด้านขนานกัน หรือวิธีการที่เรียกว่า กฎของรูปสี่เหลี่ยมที่มีสองด้านขนานกัน หรือกฎรูปสี่เหลี่ยมคางหมู ซึ่งในการศึกษาครั้งนี้จะใช้วิธีนี้ในการคำนวณ $A\hat{U}C$ วิธีนี้เป็นเทคนิคการประมาณค่าเกี่ยวกับการคำนวณอินทริกรัลจำกัดเขต ซึ่งแสดงดังนี้

$$\int_{a}^{\mathrm{b}} \mathrm{f}\,(x)dx\,;a \leq x \leq b$$

ฟังก์ชัน $\mathrm{f}\,(x)\,;a \leq x \leq b$ ซึ่งจะคำนวณโดยอินทริกรัลจำกัดเขต

ค่าประมาณพื้นที่ใต้โค้ง ROC เป็นค่าที่บ่งบอกถึงความสามารถในการแบ่งกลุ่มได้ถูกต้อง หรือความเชื่อถือได้ของตัวแบบ ซึ่งมีพิสัยอยู่ระหว่าง 0 ถึง 1 โดยที่เกณฑ์ทั่วไปของการวัดความสามารถหรือความเชื่อถือได้ของตัวแบบ สามารถตีความจากค่าดังกล่าวดังนี้



ภาพที่ 2.1 แสดงความน่าเชื่อถือของตัวแบบภายใต้พื้นที่ใต้โค้ง ROC

ถ้า $A\hat{U}C = 0.50$ ถือเป็นตัวแบบที่มีความเชื่อถือได้น้อย ไม่สามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้

ถ้า $0.70 \le A\hat{U}C < 0.80$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้พอใช้

ถ้า $0.80 \le A\hat{U}C < 0.90$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ดี โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้ดี

ถ้า $A\hat{U}C \ge 0.90$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ในระดับดีมาก โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้อย่างชัดเจน

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ต้องการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยงสำหรับแต่ละสถานการณ์ที่ต้องการศึกษา โดยตัวแบบที่ดีที่สุดคือตัวแบบที่ให้ค่าพื้นที่ใต้โค้ง ROC สูงสุด

การจำลองข้อมูลในแต่ละสถานการณ์จะจำลองขึ้นด้วยการทำงานของเครื่องคอมพิวเตอร์ โดยใช้เทคนิคมอนติคาร์โล ด้วยโปรแกรม R รายละเอียดต่างๆของแผนการวิจัยมีดังนี้

3.1 ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้มีขอบเขตการศึกษาวิจัยดังนี้

3.1.1 ทำการศึกษาตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันการเชื่อมโยงโลจิท เพื่อหาตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์ โดยการเปรียบเทียบทีละ 2 ตัวแบบ

3.1.2 ตัวแปรอิสระ (X) มีการแจกแจงเริ่มต้นเป็นการแจกแจงแบบยูนิฟอร์ม

3.1.3 ตัวแบบทั้งสองที่นำมาเปรียบเทียบกัน จะจำนวนตัวแปรอิสระ (p_1,p_2) ถูกกำหนดเป็น

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

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

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

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

3.1.4 ขนาดตัวอย่าง (n) ในการศึกษาครั้งนี้ คือ 50,100,150,200,250

3.1.5 ใช้ค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ j และตัวแปรอิสระตัวที่ j’ ในการวัดระดับความสัมพันธ์ระหว่างตัวแปรอิสระ

$$Max\{|r_{jj'}|\}$$

โดยที่ $r_{jj'}$ คือ สหสัมพันธ์ระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ระดับความสัมพันธ์ระหว่างตัวแปรอิสระแบ่งเป็น 3 ระดับ ดังนี้

- ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับต่ำ เมื่อ 0 < $Max\{|r_{jj'}|\} \leq$ 0.33
- ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับปานกลาง เมื่อ 0.33 < $Max\{|r_{jj'}|\} \leq$ 0.66
- ความสัมพันธ์ระหว่างตัวแปรอิสระอยู่ในระดับสูง เมื่อ 0.66 < $Max\{|r_{jj'}|\} \leq$ 0.99

3.1.6 ตัวแปรตาม (Y) เป็นตัวแปรเชิงคุณภาพมีค่าได้เพียง 2 ค่า คือ 0 และ 1

3.1.7 กำหนดค่าพารามิเตอร์เริ่มต้นของสมการถดถอยเป็นค่าใดๆ คือ

$\beta_i = 0.1; i = 0,1,2,..., p_1 + p_2$

3.1.8 กำหนดให้มีความคลาดเคลื่อน (ε) มีการแจกแจงแบบโลจิท

3.1.9 กำหนดระดับนัยสำคัญ (α) ที่ระดับ 0.05 $(\alpha = 0.05)$

3.1.10 การจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 500 รอบ

3.2 ขั้นตอนในการดำเนินการวิจัย

3.2.1 กำหนดขนาดตัวอย่าง n =50,100,150,200,250

3.2.2 สร้างตัวแปรอิสระ (X) ให้มีการแจกแจงแบบยูนิฟอร์ม ซึ่งมีขั้นตอนในการจำลองดังนี้

3.2.2.1 สร้างตัวแปรอิสระให้มีการอจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 1 $Z_i \sim N(0,1); i = 1,..., p_1 + p_2$

โดยที่ p_1 คือ จำนวนตัวแปรอิสระของตัวแบบที่ 1

p_2 คือ จำนวนตัวแปรอิสระของตัวแบบที่ 2

3.2.2.2 ทำให้ตัวแปรอิสระในแต่ละตัวแบบมีการแจกแจงแบบ Multivariate normal โดยที่

$$\tilde{Z}_1 = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{p_1} \end{bmatrix} \sim N_{p_1}(\tilde{0}_{p_1\times 1}, \Sigma_1) \quad \text{และ} \quad \tilde{Z}_2 = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{p_2} \end{bmatrix} \sim N_{p_2}(\tilde{0}_{p_2\times 1}, \Sigma_2)$$

ซึ่งจะได้ตัวแปรอิสระที่มีความสัมพันธ์กันของแต่ละตัวแบบดังนี้

ตัวแบบ 1: $Z_1, ..., Z_{p_1}$

ตัวแบบ 2: $Z_{p_{1+1}}, ..., Z_{p_1+p_2}$

3.2.2.3. แปลงตัวแปรอิสระจากการแจกแจงแบบ $Z_i \sim N(0,1)$ เป็น $U_i \sim U\,(0,1)$

3.2.2.4 แปลง $U_i \sim U\,(0,1)$ เป็น $X_i \sim U\,(-10,10)$

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

$$0 \le Max\{|r_{jj'}|\} \le 0.99$$

โดยที่ r_{ij} คือ สหสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ j และตัวแปรอิสระตัวที่ j’

เกณฑ์สำหรับการแบ่งระดับความสัมพันธ์ระหว่างตัวแปรอิสระ แสดงดังนี้

- ตัวแปรอิสระมีความสัมพันธ์กันในระดับอย่างต่ำ: $0 < Max\{|r_{jj'}|\} \le 0.33$
- ตัวแปรอิสระมีความสัมพันธ์กันในระดับปานกลาง: $0.33 < Max\{|r_{jj'}|\} \le 0.66$
- ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูง: $0.66 < Max\{|r_{jj'}|\} \le 0.99$

3.2.2.6. จะได้ตัวแปรอิสระที่มีการแจกแจงแบบ $X_i \sim U\,(-10,10)$ มีระดับความสัมพันธ์ระหว่างตัวแปรอิสระของแต่ละแบบตามที่ได้ตรวจสอบระดับความสัมพันธ์ที่แท้จริงในข้อ 5 ซึ่งตัวแปรอิสระของทั้งสองตัวแบบที่ได้เป็นดังนี้

ตัวแบบ 1: $X_1, ..., X_{p_1}$

ตัวแบบ 2: $X_{p_{1+1}}, ..., X_{p_1+p_2}$

3.2.4. สร้างค่าความคลาดเคลื่อนให้มีการแจกแจงแบบโลจิท

3.2.5. กำหนดสัมประสิทธ์การถดถอยของพารามิเตอร์ β_i เมื่อ i=0,1,2,…,10 คงที่ เท่ากับ 0.01

3.2.6. สร้างตัวแปรตาม(Y) ให้มีการแจกแจงแบบเบอร์นูลลี ซึ่งคำนวณจากข้อ 3.2.2 - 3.2.5

โดยตัวแปรตาม (Y) มีค่าได้เพียง 2 ค่า คือ 0 และ 1 ดังนี้

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

แทนค่าตามข้อกำหนดที่ 2-4 จะได้ $Y_i^* = \tilde{X}_i^T \tilde{\beta} + \varepsilon_i$

ทำการปรับ Y_i^* ให้เป็นตัวแปรหุ่น (Dummy Variable) แทนด้วยค่า Y_i ดังนี้

$Y_i = 1$ ถ้า $Y_i^* > 0$

$Y_i = 0$ ถ้า $Y_i^* \leq 0$

3.2.7. นำข้อมูลจากตัวแปรตามในข้อที่ 3.2.6.และตัวแปรอิสระมาทำการประมาณค่าพารามิเตอร์ ด้วยการวิเคราะห์ความแบบโลจิสติกโดยใช้ฟังก์ชั่นโลจิทเป็นการเชื่อมโยงเพื่อนำมา ค่าพารามิเตอร์ดังกล่าวมาสร้างตัวแบบที่ใช้ในการพยากรณ์

3.2.8. สร้างค่าพยากรณ์ ($\hat{y}_i^*$) จากตัวแบบที่ได้ในข้อที่ 3.2.7

3.2.9. คำนวณหาค่า Sensitivity (SN) และค่า SP (Specificity) โดยใช้ข้อมูลจากข้อที่ 3.2.6 และ ข้อที่ 3.2.8

3.2.10 พล็อต ROC curve จากข้อมูลในข้อ 3.2.9. โดยที่แกน x เป็นค่า 1-Specificity แกน y เป็น ค่า Sensitivity (1-SP,SN) จากนั้นคำนวณหาค่าประมาณพื้นที่ใต้โค้ง ROC การคำนวณ ค่าพื้นที่ใต้โค้งใช้หลักเกณฑ์เชิงสี่เหลี่ยมคางหมู

3.2.11 ทำซ้ำขั้นตอนที่ 3.2.1-3.2.7 จำนวน 500 รอบ

3.2.12 เปรียบเทียบพื้นที่ใต้โค้ง ROC ในแต่ละตัวแบบ โดยการทดสอบสมมติฐานเพื่อหาตัวแบบที่ เหมาะสมที่สุด

สมมติฐานที่ใช้ในการทดสอบ

H_0 : ค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 1มากกว่าหรือเท่ากับค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 2

H_a : ค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 2

$$H_0 : \mu_{\text{model}1} \geq \mu_{\text{model}2}$$
$$H_a : \mu_{\text{model}1} < \mu_{\text{model}2}$$

ตัวสถิติทดสอบ คือ

$$t = \frac{\overline{y_1} - \overline{y_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

ในงานวิจัยครั้งนี้ เมื่อเราไม่ทราบค่าความแปรปรวน (σ^2) ในงานวิจัยครั้งนี้ เราจึงใช้ s^2 แทน และใช้ t-test ในการทดสอบ แต่เนื่องจากขนาดตัวอย่างในงานวิจัยครั้งนี้มีขนาดใหญ่ เราจึงใช้ z-test แทน

เกณฑ์ในการตัดสินใจ

ถ้า $t \le t_{0.05} (t_{0.05} = -1.645)$ จะปฏิเสธ H_0 นั่นคือ ค่าเฉลี่ยของพื้นที่โดยใช้พื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยของพื้นที่โดยใช้พื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2

หรือสมมติฐานที่ใช้ในการทดสอบ

H_0 : ค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 1น้อยกว่าหรือเท่ากับค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 2

H_a : ค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 2

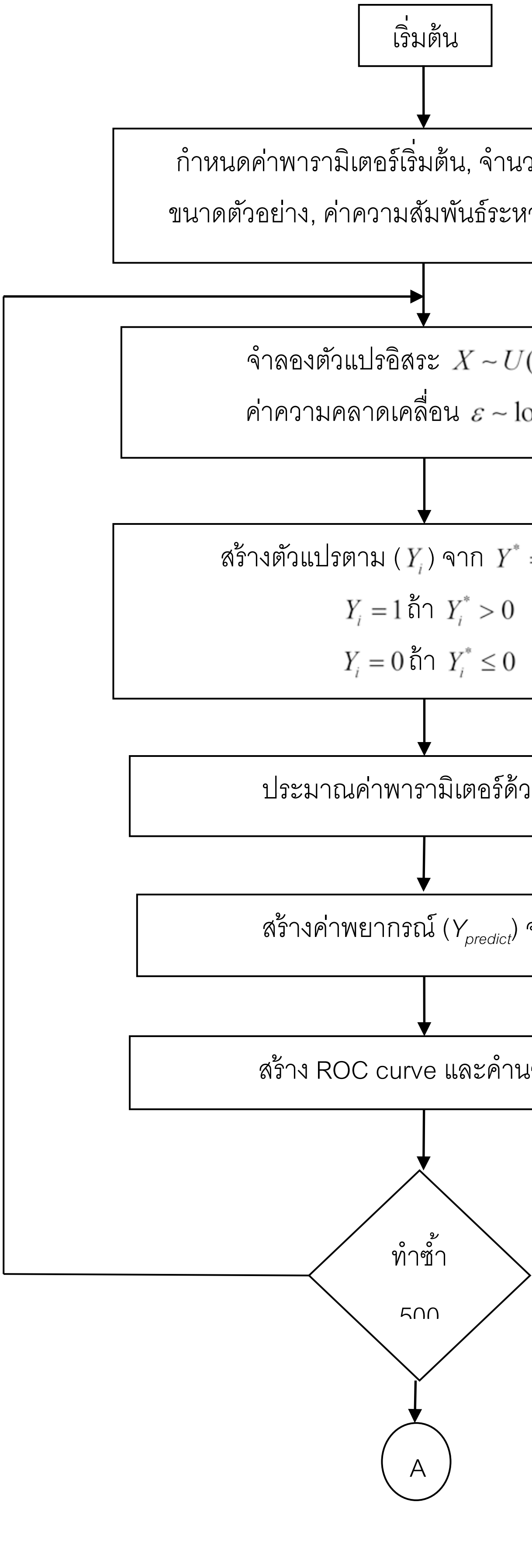
$$H_0 : \mu_{model1} \le \mu_{model2}$$
$$H_a : \mu_{model1} > \mu_{model2}$$

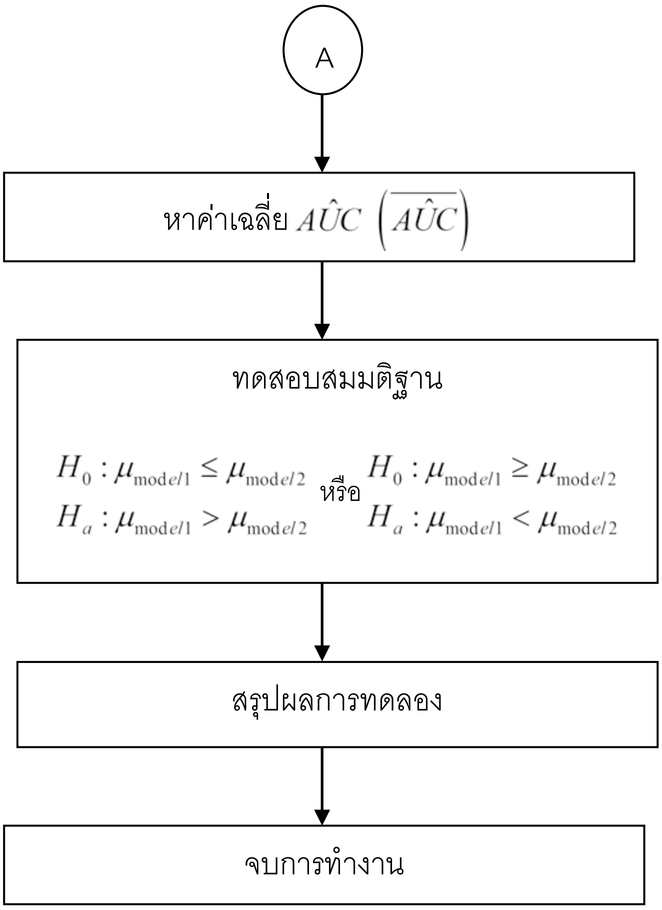
เกณฑ์ในการตัดสินใจ

ถ้า $t \ge t_{0.95} (t_{0.95} = 1.645)$ จะปฏิเสธ H_0 นั่นคือ ค่าเฉลี่ยของพื้นที่โดยใช้พื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่โดยใช้พื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2

3.2.13. สรุปผลการวิจัยในแต่ละกรณี

สำหรับขั้นตอนการจำลองในการวิจัยครั้งนี้ ได้แสดงแผนภาพแสดงขั้นตอนในการวิจัยดังนี้


เริ่มต้น
กำหนดค่าพารามิเตอร์เริ่มต้น, จำนวนตัวแปรอิสระ,
ขนาดตัวอย่าง, ค่าความสัมพันธ์ระหว่างตัวแปรอิสระ
จำลองตัวแปรอิสระ $X \sim U(-10,10)$
ค่าความคลาดเคลื่อน $\varepsilon \sim \log is(0,1)$
สร้างตัวแปรตาม (Y_i) จาก $Y^* = X\tilde{\beta} + \varepsilon$
$Y_i = 1$ ถ้า $Y_i^* > 0$
$Y_i = 0$ ถ้า $Y_i^* \leq 0$
ประมาณค่าพารามิเตอร์ด้วยวิธี MLE
สร้างค่าพยากรณ์ ($Y_{predict}$) จากตัวแบบ
สร้าง ROC curve และคำนวณ $A\hat{U}C$
ทำซ้ำ
500
A



รูปที่ 3.1 แสดงขั้นตอนการทำงานของโปรแกรม

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาหาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง ในแต่ละสถานการณ์ที่ต้องการศึกษา ซึ่งใช้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือก และทำการจำลองข้อมูลเพื่อศึกษาถึงกรณีดังกล่าว โดยการสรุปผลการวิจัยได้นำเสนอในรูปแบบของตาราง และเพื่อความสะดวกในการสรุปผล จึงกำหนดสัญลักษณ์แทนความหมายต่างๆในตาราง ดังนี้

n แทน จำนวนขนาดตัวอย่าง

p_1 แทน จำนวนตัวแปรอิสระในตัวแบบที่ 1

p_2 แทน จำนวนตัวแปรอิสระในตัวแบบที่ 2

μ_{model1} แทน ค่าพื้นที่เฉลี่ยใต้โค้ง ROC ของตัวแบบที่ 1

μ_{model2} แทน ค่าพื้นที่เฉลี่ยใต้โค้ง ROC ของตัวแบบที่ 2

$\overline{A\hat{U}C}_{model1}$ แทน ค่าประมาณพื้นที่เฉลี่ยใต้โค้ง ROC ของตัวแบบที่ 1

$\overline{A\hat{U}C}_{model2}$ แทน ค่าประมาณพื้นที่เฉลี่ยใต้โค้ง ROC ของตัวแบบที่ 2

$Max\left\{\left|r_{jj'}\right|\right\}$ แทน ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ

ผลการศึกษา

ผลการศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง เมื่อมีปัจจัยที่เกี่ยวข้อง คือ ขนาดตัวอย่าง ระดับความสัมพันธ์ในแต่ละคู่ตัวแปรอิสระ และจำนวนตัวแปรอิสระในแต่ละตัวแบบ ผลการศึกษาเป็นดังนี้

ตารางที่ 4.1 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัว และ 2 ตัว (p=(2,2)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\{\lvert r_{jj'}\rvert\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
2	2	50	Low	0.7585	0.7528	model 1 หรือ model 2
			Medium	0.7478	0.7466	model 1 หรือ model 2
			High	0.7396	0.7277	model 1
		100	Low	0.7759	0.7658	model 1
			Medium	0.7659	0.7581	model 1
			High	0.7541	0.7456	model 1
		150	Low	0.7778	0.7739	model 1 หรือ model 2
			Medium	0.7706	0.7638	model 1
			High	0.7589	0.7544	model 1 หรือ model 2
		200	Low	0.7795	0.7733	model 1 หรือ model 2
			Medium	0.7746	0.7633	model 1
			High	0.7619	0.7562	model 1 หรือ model 2
		250	Low	0.7823	0.7746	model 1
			Medium	0.7761	0.7656	model 1
			High	0.7626	0.7542	model 1

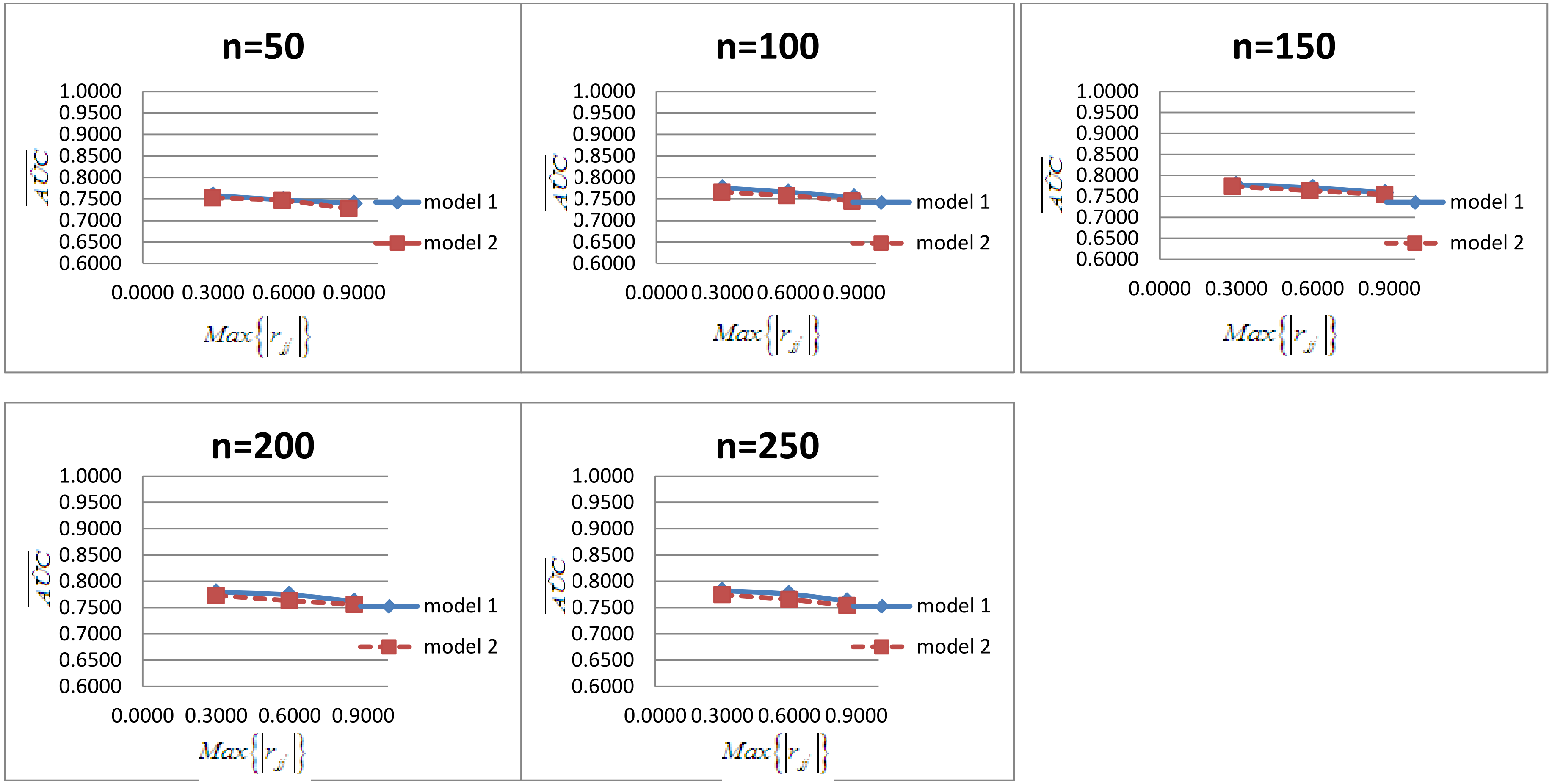
จากตารางที่ 4.1 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 2 ตัว (p=(2,2)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{\text{model1}} \leq \mu_{\text{model2}}$ และ $H_a : \mu_{\text{model1}} > \mu_{\text{model2}}$ ที่ระดับนัยสำคัญ 0.05 ผลการทดสอบพบว่า

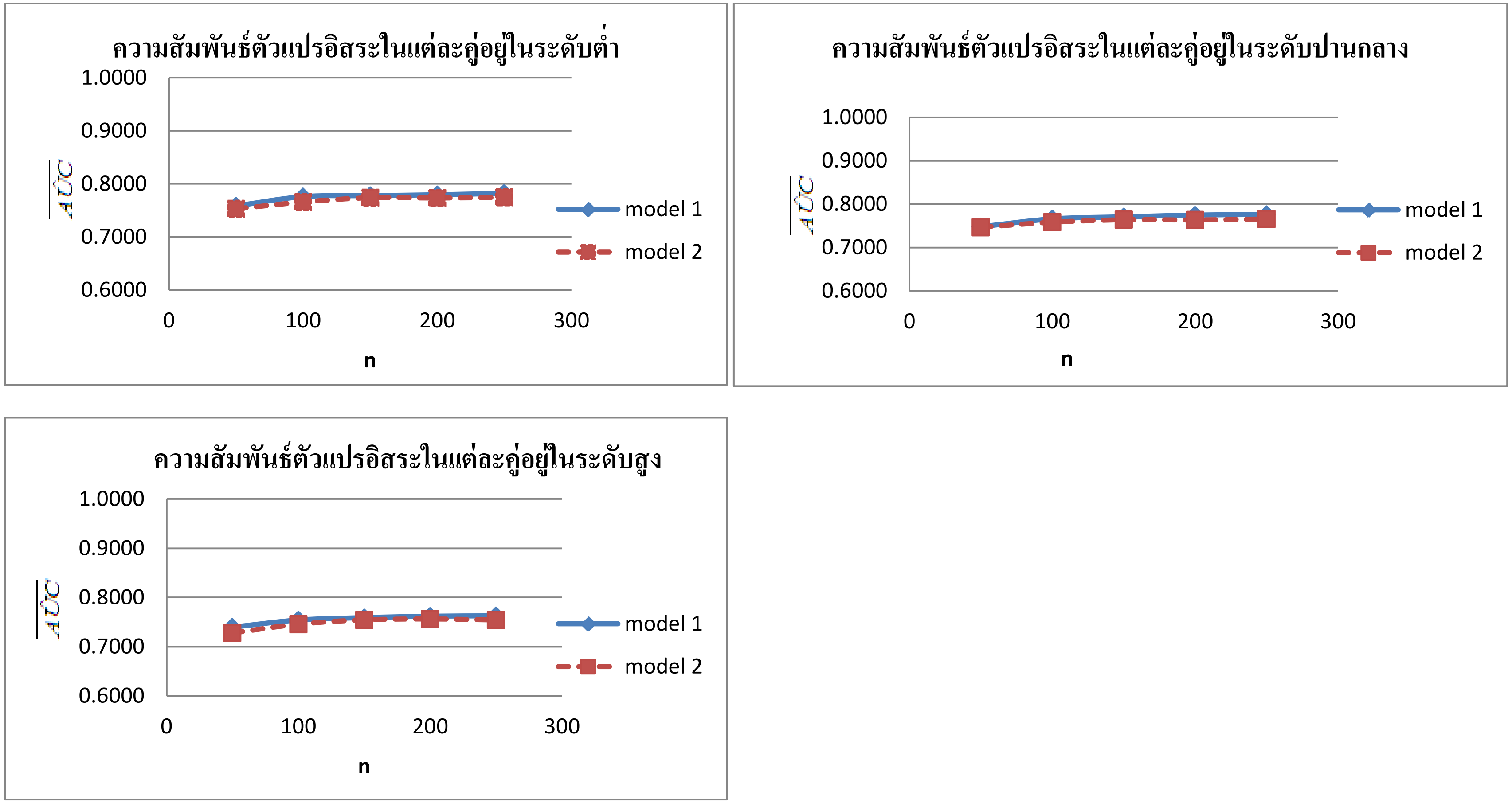
กรณีจำนวนตัวแปรอิสระในตัวแบบที่ 1 (p_1) เท่ากับ 2 ตัวแปร จำนวนตัวแปรอิสระในตัวแบบที่ 2 (p_2) เท่ากับ 2 ตัวแปร ขนาดตัวอย่าง(n) เท่ากับ 50 ในระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีความสัมพันธ์ในมาก ขนาดตัวอย่าง(n)เท่ากับ 100 และ 250 ในระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีความสัมพันธ์ในระดับน้อย ปานกลาง และมาก ขนาดตัวอย่าง(n)เท่ากับ 150 และ 200 ในระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีความสัมพันธ์ในระดับปานกลาง ผลการคัดเลือกตัวแบบโดยใช้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือกและจากการทดสอบสมมติฐานที่ระดับนัยสำคัญ 0.05 พบว่าตัวแบบที่ 1 (p_1=2)ให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่าตัวแบบที่ 1 ซึ่งหมายถึงว่า ตัวแบบที่ 1 เป็นตัวแบบที่ถูกคัดเลือก นั่นคือตัวแบบที่ 1 มีความน่าเชื่อถือของตัวแบบมากกว่าตัวแบบที่ 2(p_2=2) (P-value<0.05)

ขนาดตัวอย่าง(n)เท่ากับ 50 ในระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีความสัมพันธ์ในระดับน้อย ปานกลาง ขนาดตัวอย่าง(n)เท่ากับ 150 และ 200 ในระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีความสัมพันธ์ในระดับน้อย และมาก ผลการคัดเลือกตัวแบบโดยใช้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือก และจากการทดสอบสมมุติฐานที่ระดับนัยสำคัญ 0.05 พบว่าตัวแบบที่ 1(p_1=2) และตัวแบบที่ 2(p_2=2) ให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ไม่ต่างกัน ซึ่งหมายถึงว่า ตัวแบบที่ 1 และตัวแบบที่ 2 เป็นตัวแบบที่ใช้แทนกันได้ นั่นคือเราสามารถเลือกใช้ตัวแบบที่ 1 หรือตัวแบบที่ 2 ก็ได้ ซึ่งมีความน่าเชื่อถือของตัวแบบเท่ากัน

ภาพที่ 4.1 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2ตัวและ 2 ตัว (p=(2,2)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.2 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 2 ตัว (p=(2,2)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



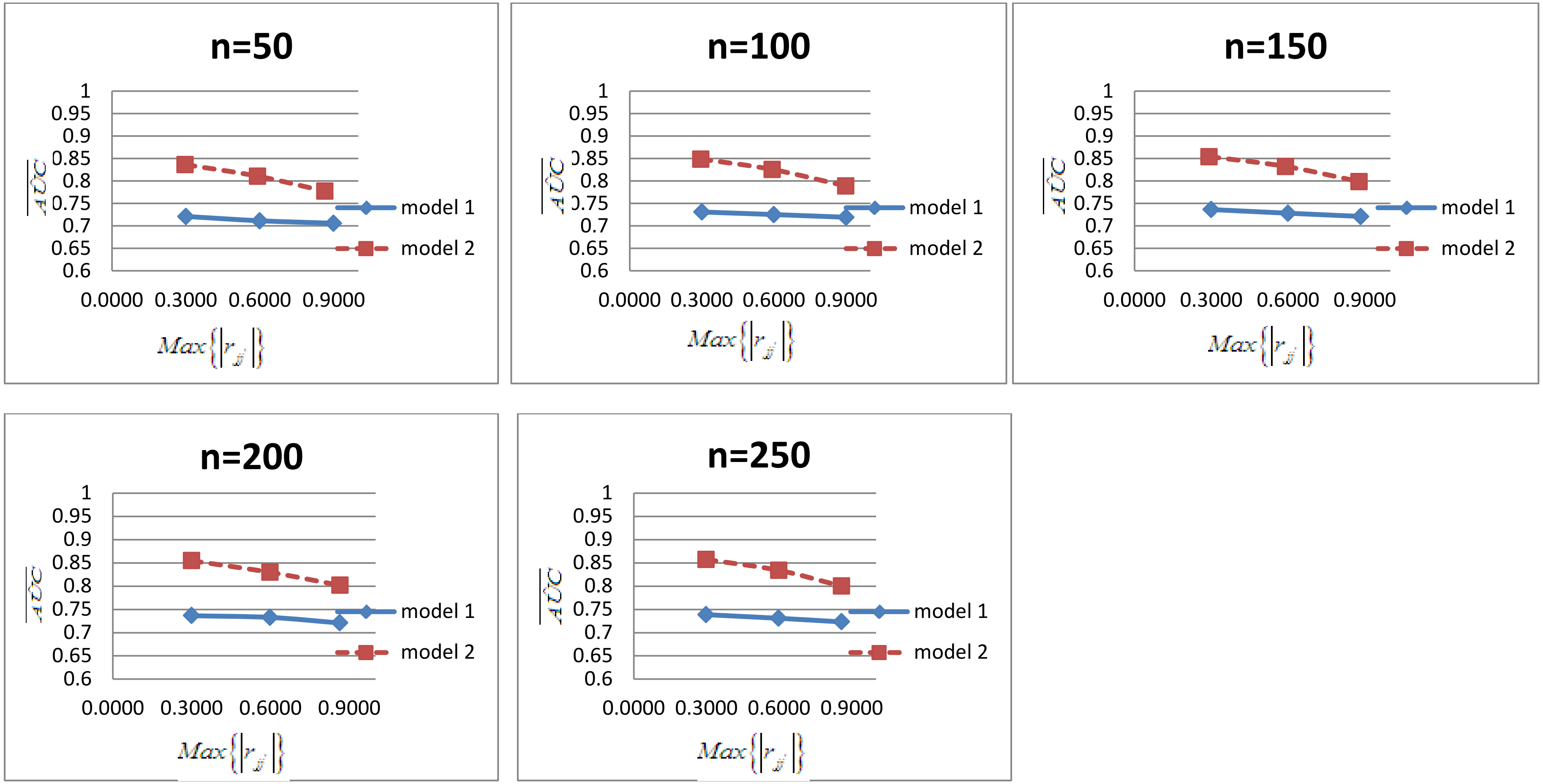
ตารางที่ 4.2 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัว และ 3 ตัว (p=(2,3)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
2	3	50	Low	0.7211	0.8364	model 2
			Medium	0.7115	0.8108	model 2
			High	0.7059	0.7774	model 2
		100	Low	0.7310	0.8488	model 2
			Medium	0.7253	0.8255	model 2
			High	0.7192	0.7888	model 2
		150	Low	0.7364	0.8542	model 2
			Medium	0.7283	0.8324	model 2
			High	0.7211	0.7988	model 2
		200	Low	0.7364	0.8546	model 2
			Medium	0.7328	0.8297	model 2
			High	0.7208	0.8017	model 2
		250	Low	0.7387	0.8573	model 2
			Medium	0.7309	0.8340	model 2
			High	0.7231	0.8000	model 2

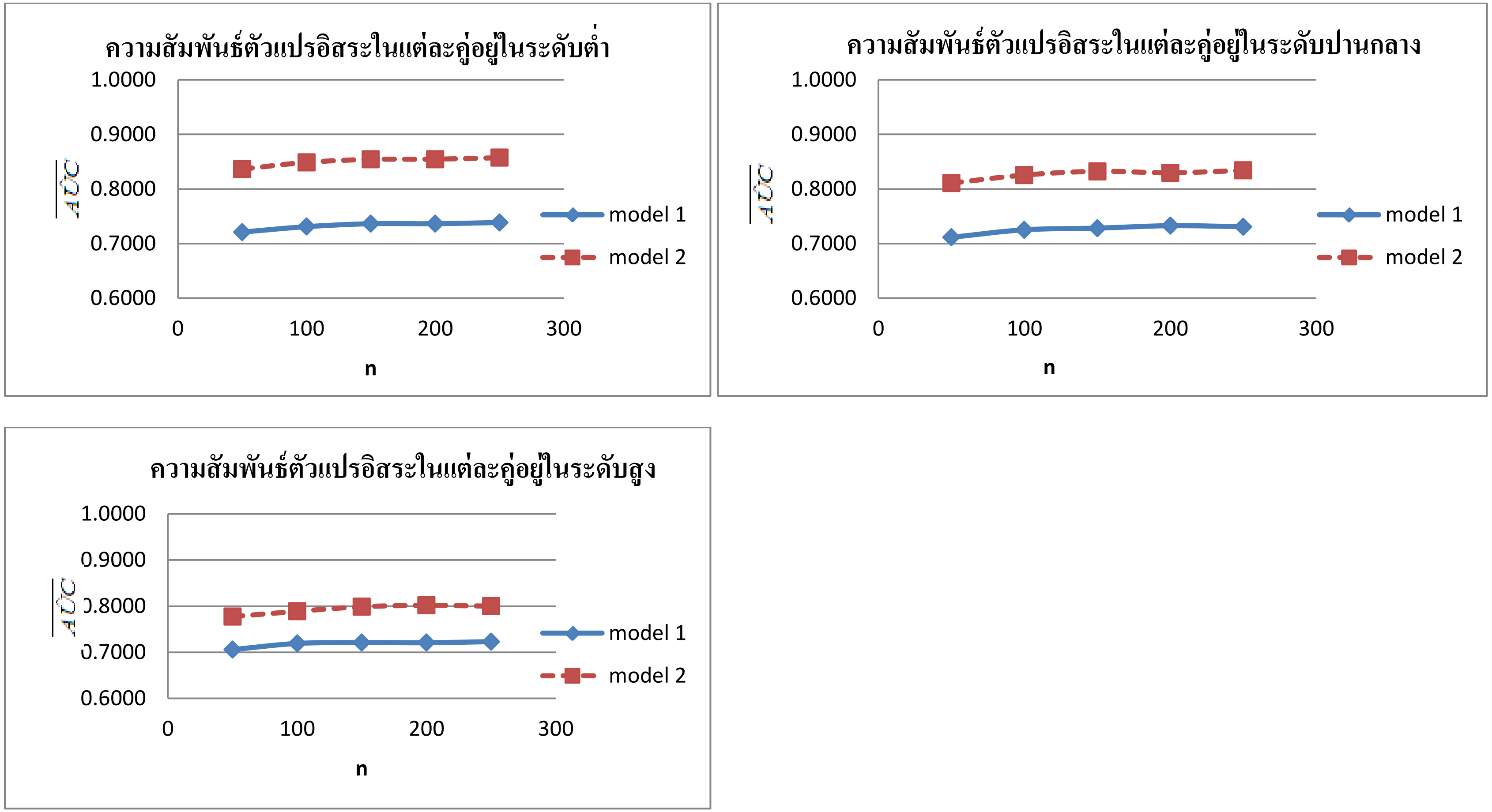
จากตารางที่ 4.2 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีจำนวนคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 3 ตัว (p=(2,3)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 2 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC น้อยกว่าตัวแปรอิสระ 3 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \geq \mu_{model2}$ และ $H_a : \mu_{model1} < \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 2 ได้ทุกกรณี นั่นคือตัวแบบที่ 2 มีความน่าเชื่อถือกว่าตัวแบบที่1(P-value<0.05)

ภาพที่ 4.3 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 3 ตัว (p=(2,3)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.4 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 3 ตัว (p=(2,3)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



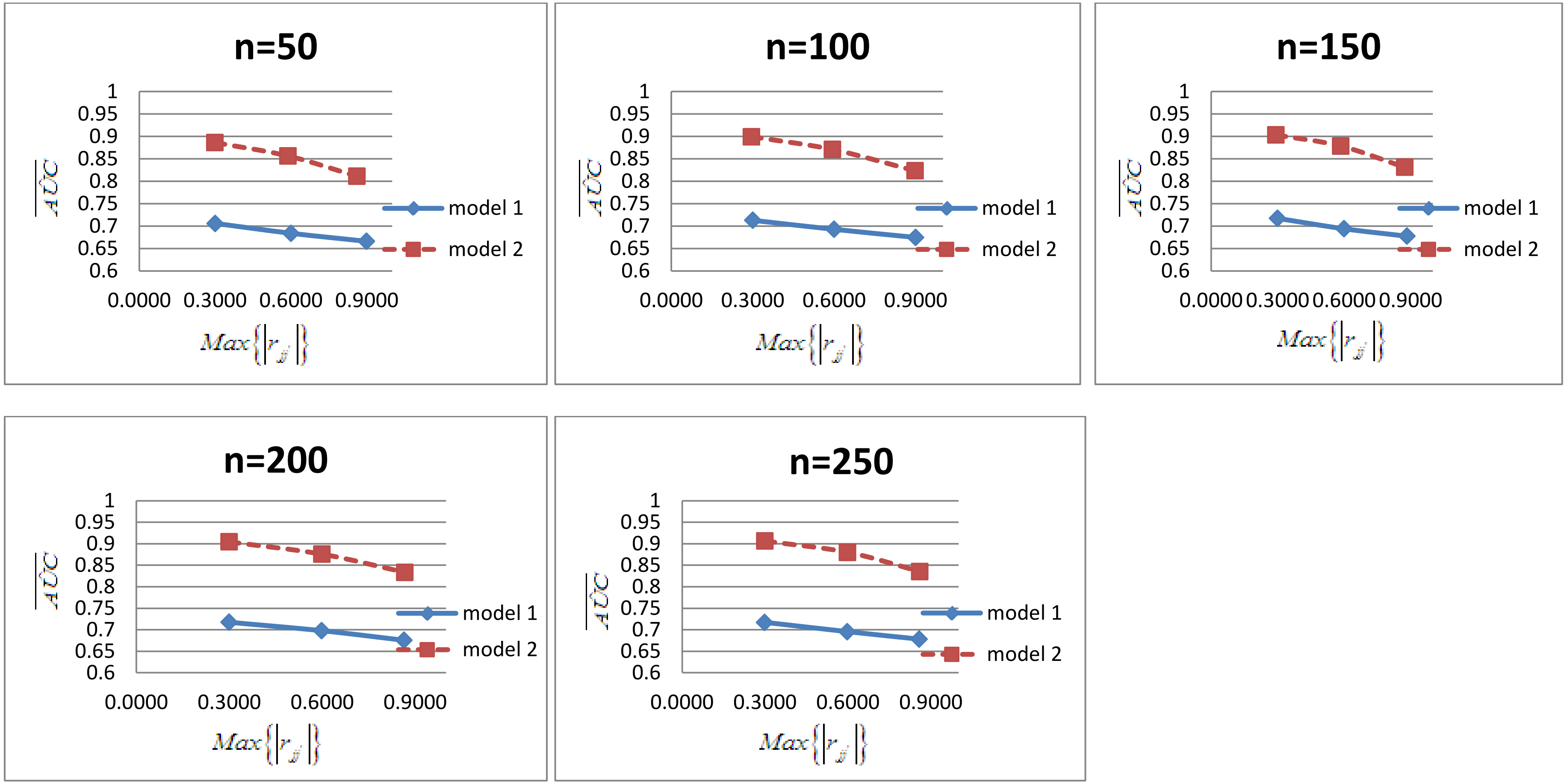
ตารางที่ 4.3 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 4 ตัว (p=(2,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลาง มาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{\text{model}1}$	$\overline{A\hat{U}C}_{\text{model}2}$	ตัวแบบที่เลือก
2	4	50	Low	0.7061	0.8864	model 2
			Medium	0.6840	0.8564	model 2
			High	0.6661	0.8111	model 2
		100	Low	0.7129	0.8995	model 2
			Medium	0.6928	0.8709	model 2
			High	0.6747	0.8233	model 2
		150	Low	0.7173	0.9034	model 2
			Medium	0.6941	0.8787	model 2
			High	0.6775	0.8313	model 2
		200	Low	0.7175	0.9044	model 2
			Medium	0.6979	0.8758	model 2
			High	0.6757	0.8335	model 2
		250	Low	0.7173	0.9063	model 2
			Medium	0.6957	0.8801	model 2
			High	0.6782	0.8356	model 2

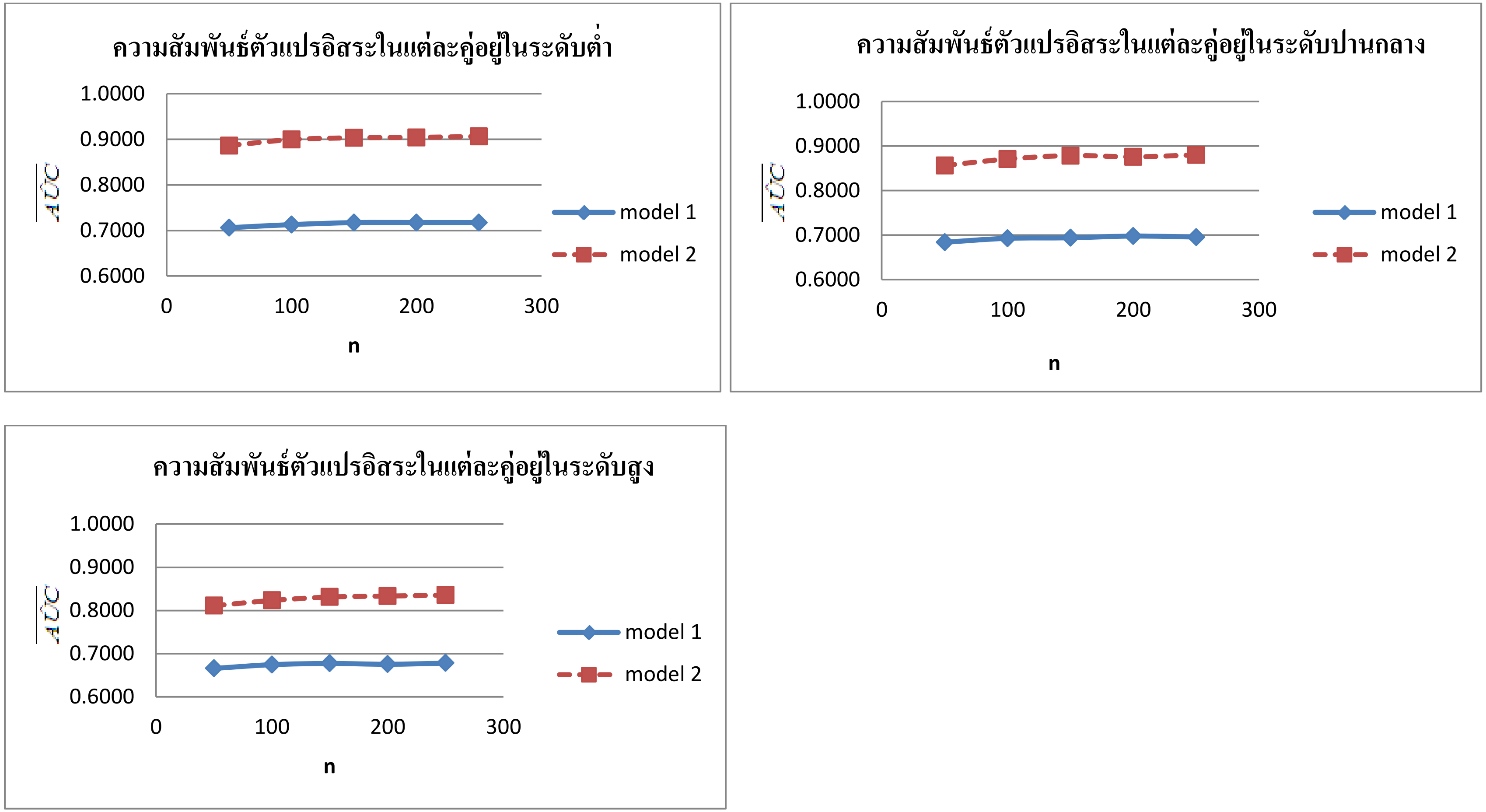
จากตารางที่ 4.3 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 4 ตัว (p=(2,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 2 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC น้อยกว่าตัวแปรอิสระ 4 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \geq \mu_{model2}$ และ $H_a : \mu_{model1} < \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 2 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.5 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 4 ตัว (p=(2,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.6 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 4 ตัว (p=(2,4)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



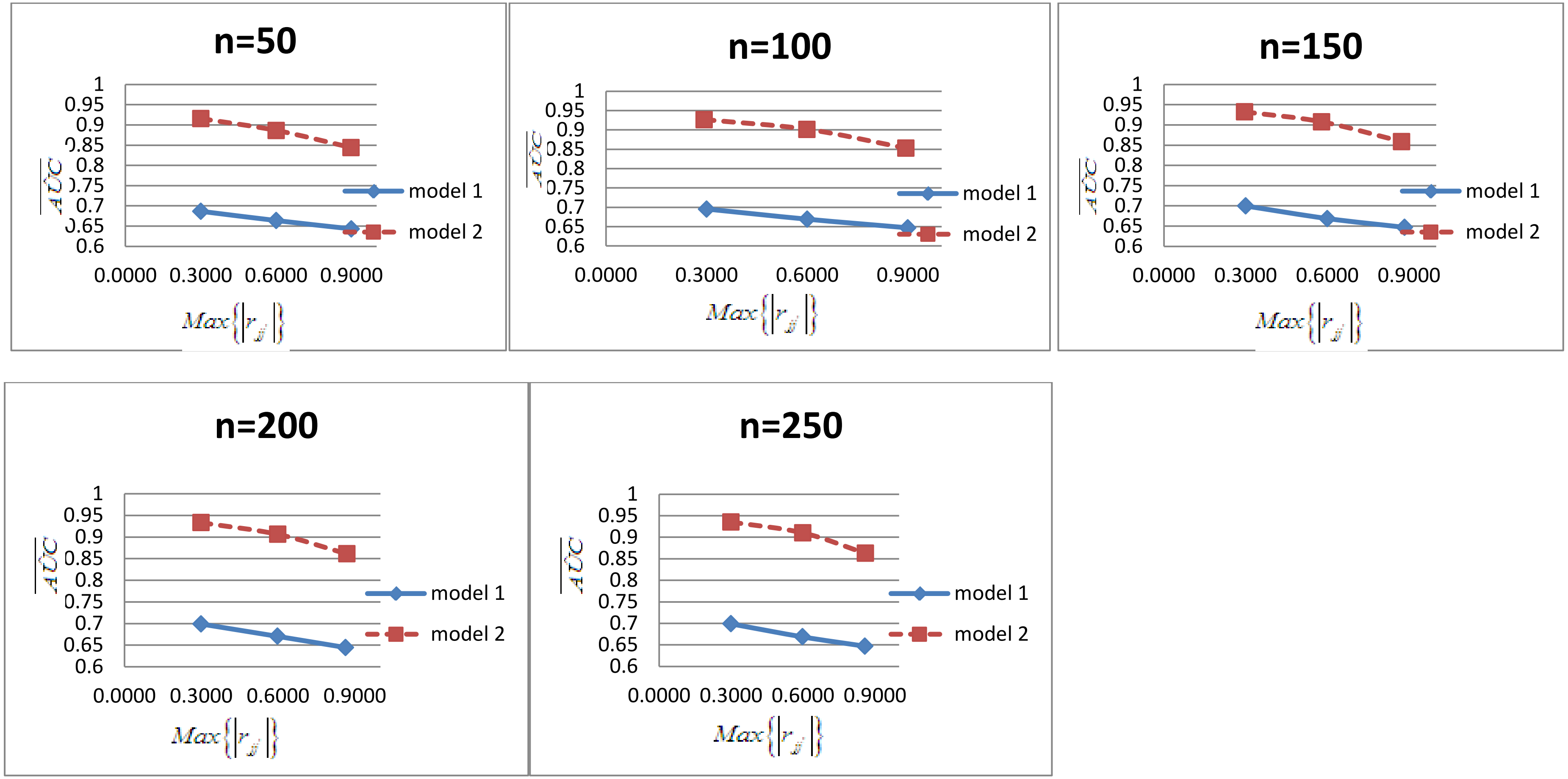
ตารางที่ 4.4 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัว และ 5 ตัว (p=(2,5)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
2	5	50	Low	0.6865	0.9157	model 2
			Medium	0.6637	0.8865	model 2
			High	0.6431	0.8442	model 2
		100	Low	0.6955	0.9267	model 2
			Medium	0.6693	0.9016	model 2
			High	0.6469	0.8527	model 2
		150	Low	0.6997	0.9323	model 2
			Medium	0.6686	0.9077	model 2
			High	0.6469	0.8588	model 2
		200	Low	0.6990	0.9332	model 2
			Medium	0.6704	0.9063	model 2
			High	0.6446	0.8611	model 2
		250	Low	0.6988	0.9353	model 2
			Medium	0.6685	0.9101	model 2
			High	0.6469	0.8626	model 2

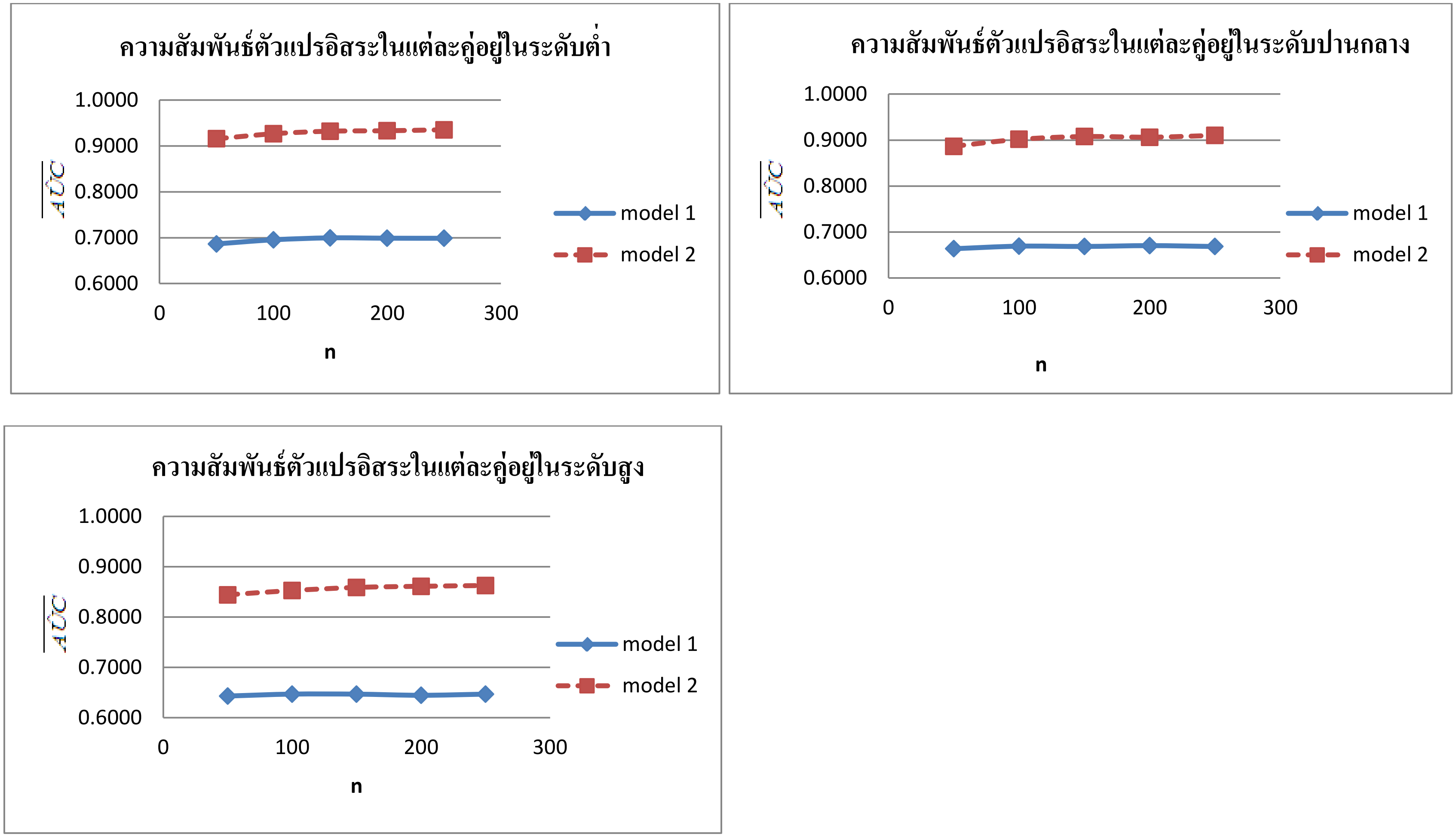
จากตารางที่ 4.4 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีจำนวนของตัวแปรอิสระเท่ากับ 2 ตัวและ 5 ตัว (p=(2,5)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 2 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC น้อยกว่าตัวแปรอิสระ 5 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \geq \mu_{model2}$ และ $H_a : \mu_{model1} < \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่2 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.7 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 5 ตัว (p=(2,5)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.8 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 2 ตัวและ 5 ตัว (p=(2,5)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



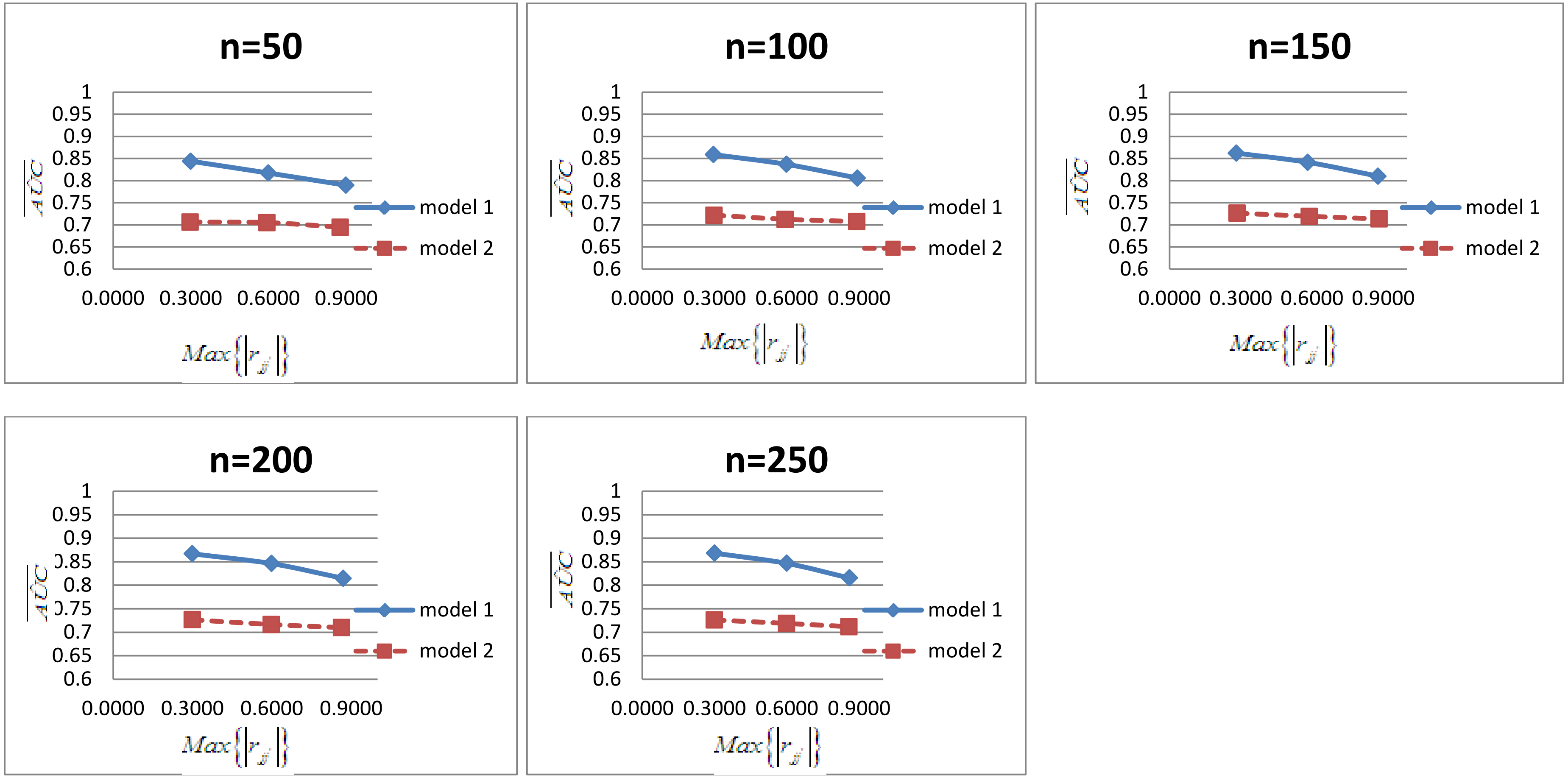
ตารางที่ 4.5 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัว และ 2 ตัว (p=(3,2)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
3	2	50	Low	0.8438	0.7064	model 1
			Medium	0.8171	0.7051	model 1
			High	0.7896	0.6947	model 1
		100	Low	0.8588	0.7218	model 1
			Medium	0.8367	0.7126	model 1
			High	0.8058	0.7075	model 1
		150	Low	0.8621	0.7264	model 1
			Medium	0.8413	0.7193	model 1
			High	0.8099	0.7133	model 1
		200	Low	0.8668	0.7267	model 1
			Medium	0.8461	0.7164	model 1
			High	0.8147	0.7099	model 1
		250	Low	0.8683	0.7262	model 1
			Medium	0.8465	0.7188	model 1
			High	0.8153	0.7119	model 1

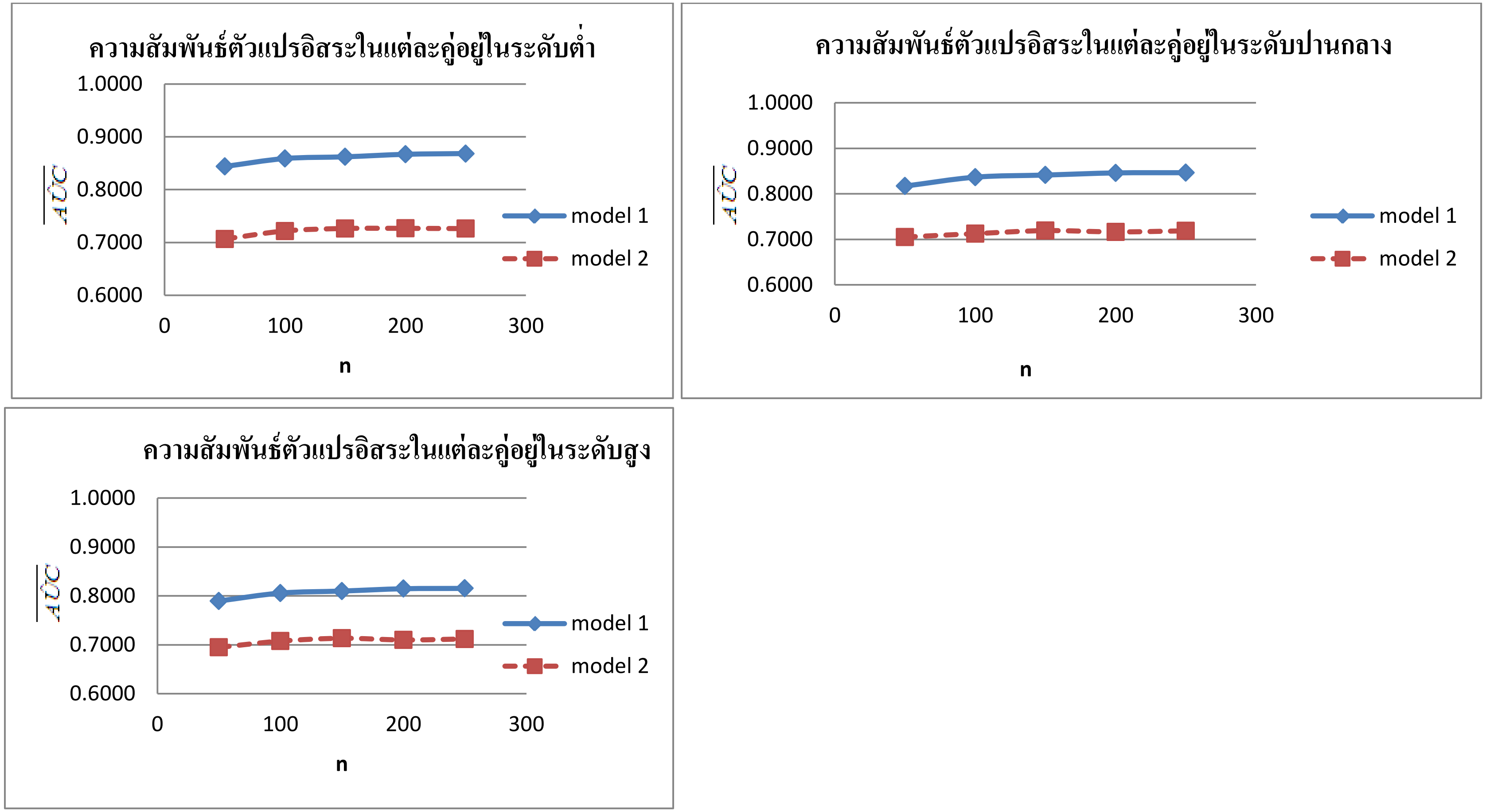
จากตารางที่ 4.5 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 2 ตัว (p=(3,2)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 3 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 2 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{\mathrm{mod}el1} \leq \mu_{\mathrm{mod}el2}$ และ $H_a : \mu_{\mathrm{mod}el1} > \mu_{\mathrm{mod}el2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.9 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 2 ตัว (p=(3,2)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.10 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 2 ตัว (p=(3,2)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



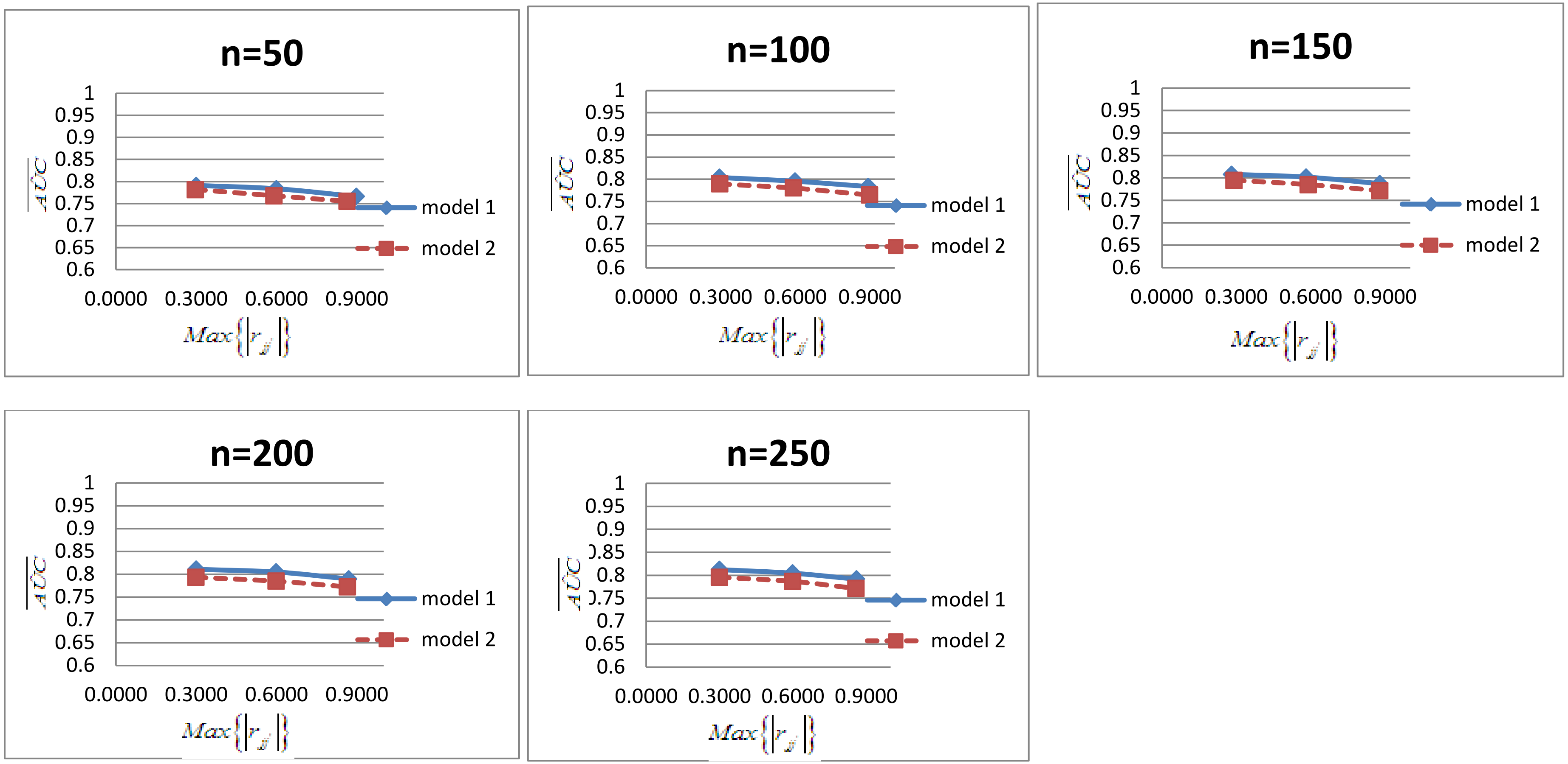
ตารางที่ 4.6 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัว และ 3 ตัว (p=(3,3)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
3	3	50	Low	0.7911	0.7818	model 1
			Medium	0.7835	0.7677	model 1
			High	0.7661	0.7551	model 1
		100	Low	0.8040	0.7895	model 1
			Medium	0.7956	0.7802	model 1
			High	0.7837	0.7645	model 1
		150	Low	0.8078	0.7947	model 1
			Medium	0.8017	0.7850	model 1
			High	0.7872	0.7715	model 1
		200	Low	0.8107	0.7932	model 1
			Medium	0.8048	0.7848	model 1
			High	0.7896	0.7722	model 1
		250	Low	0.8123	0.7955	model 1
			Medium	0.8045	0.7867	model 1
			High	0.7915	0.7713	model 1

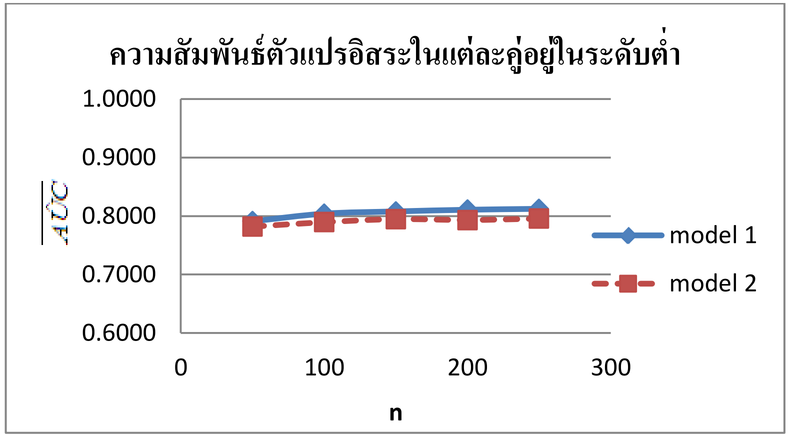
จากตารางที่ 4.6 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 3 ตัว (p=(3,3)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 3 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 3 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

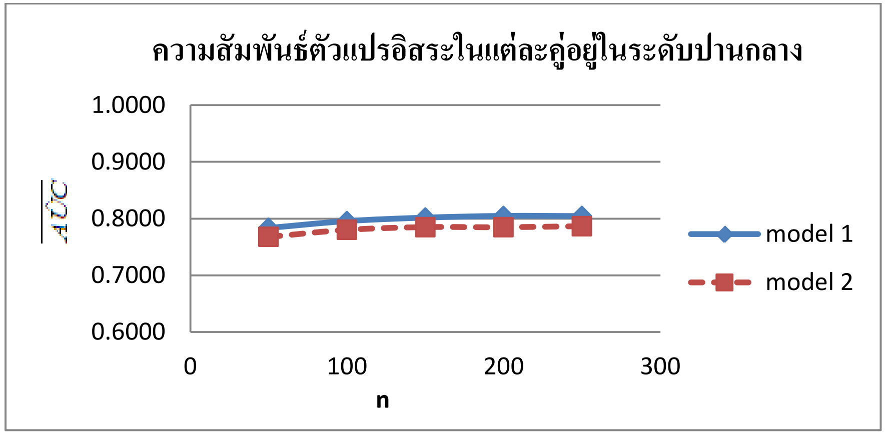
จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \le \mu_{model2}$ และ $H_a : \mu_{model1} > \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

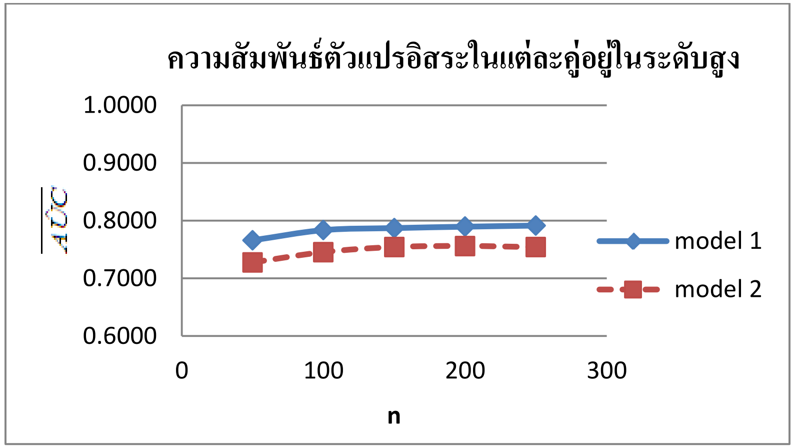
ภาพที่ 4.11 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 3 ตัว (p=(3,3)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ เปลี่ยนแปลง



ภาพที่ 4.12 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 3 ตัว (p=(3,3)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง







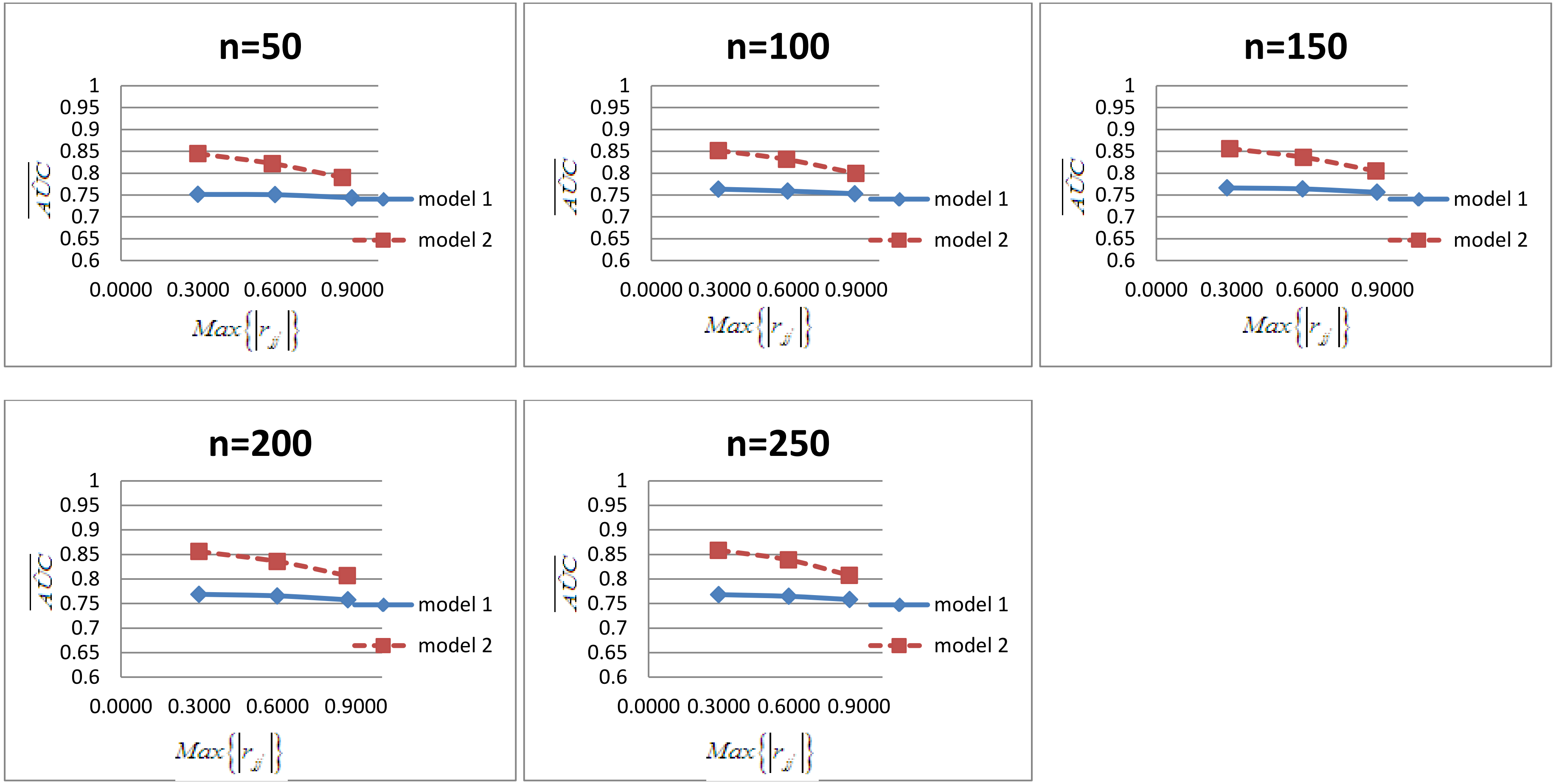
ตารางที่ 4.7 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัว และ 4 ตัว (p=(3,4)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{\text{model}1}$	$\overline{A\hat{U}C}_{\text{model}2}$	ตัวแบบที่เลือก
3	4	50	Low	0.7514	0.8446	model 2
			Medium	0.7507	0.8221	model 2
			High	0.7438	0.7903	model 2
		100	Low	0.7635	0.8517	model 2
			Medium	0.7592	0.8315	model 2
			High	0.7531	0.7995	model 2
		150	Low	0.7663	0.8562	model 2
			Medium	0.7641	0.8364	model 2
			High	0.7563	0.8049	model 2
		200	Low	0.7688	0.8559	model 2
			Medium	0.7657	0.8357	model 2
			High	0.7576	0.8066	model 2
		250	Low	0.7682	0.8586	model 2
			Medium	0.7647	0.8388	model 2
			High	0.7582	0.8074	model 2

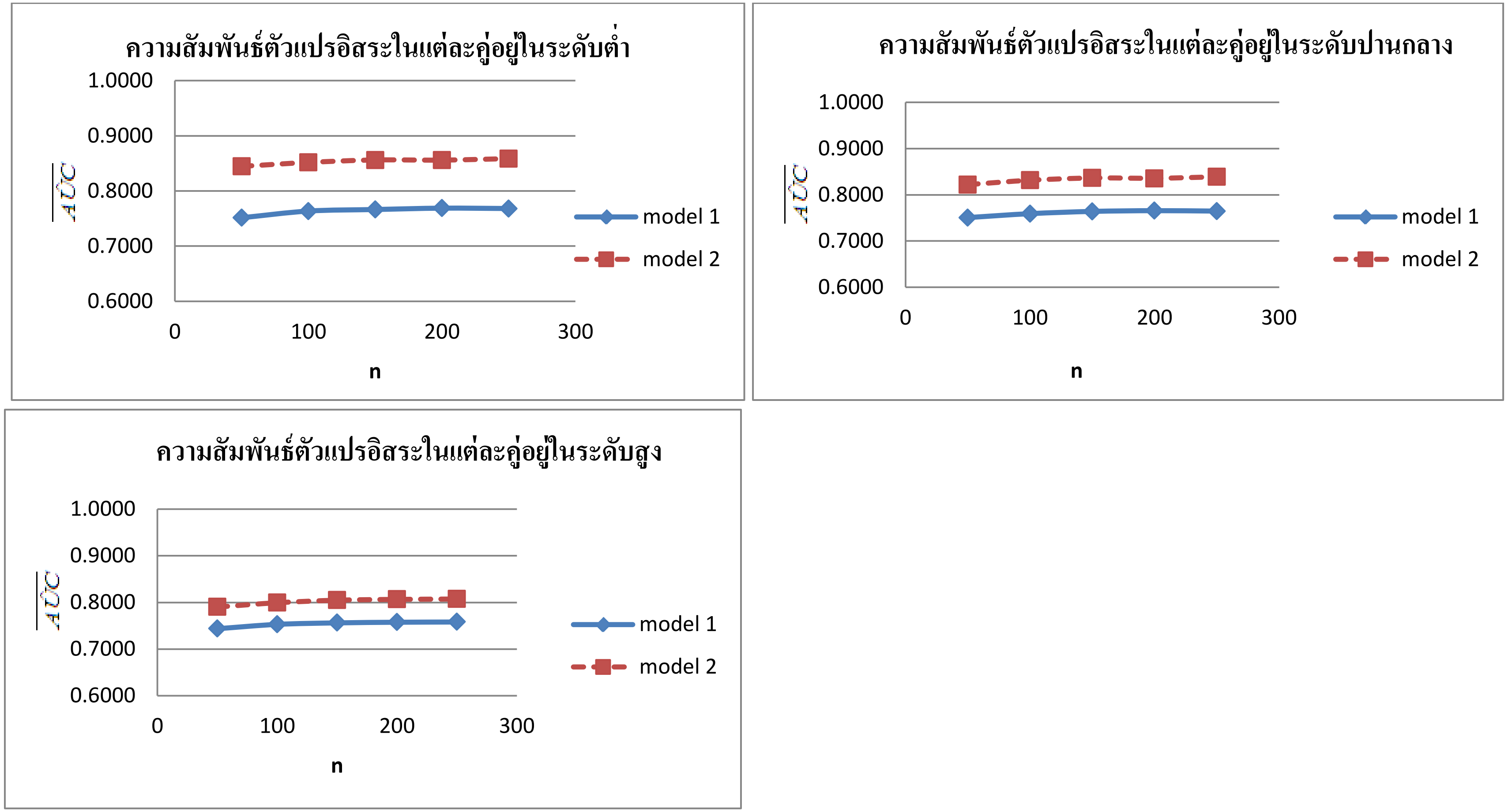
จากตารางที่ 4.7 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 4 ตัว (p=(3,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 3 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC น้อยกว่าตัวแปรอิสระ 4 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{\mathrm{mod}el1} \geq \mu_{\mathrm{mod}el2}$ และ $H_a : \mu_{\mathrm{mod}el1} < \mu_{\mathrm{mod}el2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 2 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.13 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 4 ตัว (p=(3,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ เปลี่ยนแปลง



ภาพที่ 4.14 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 4 ตัว (p=(3,4)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



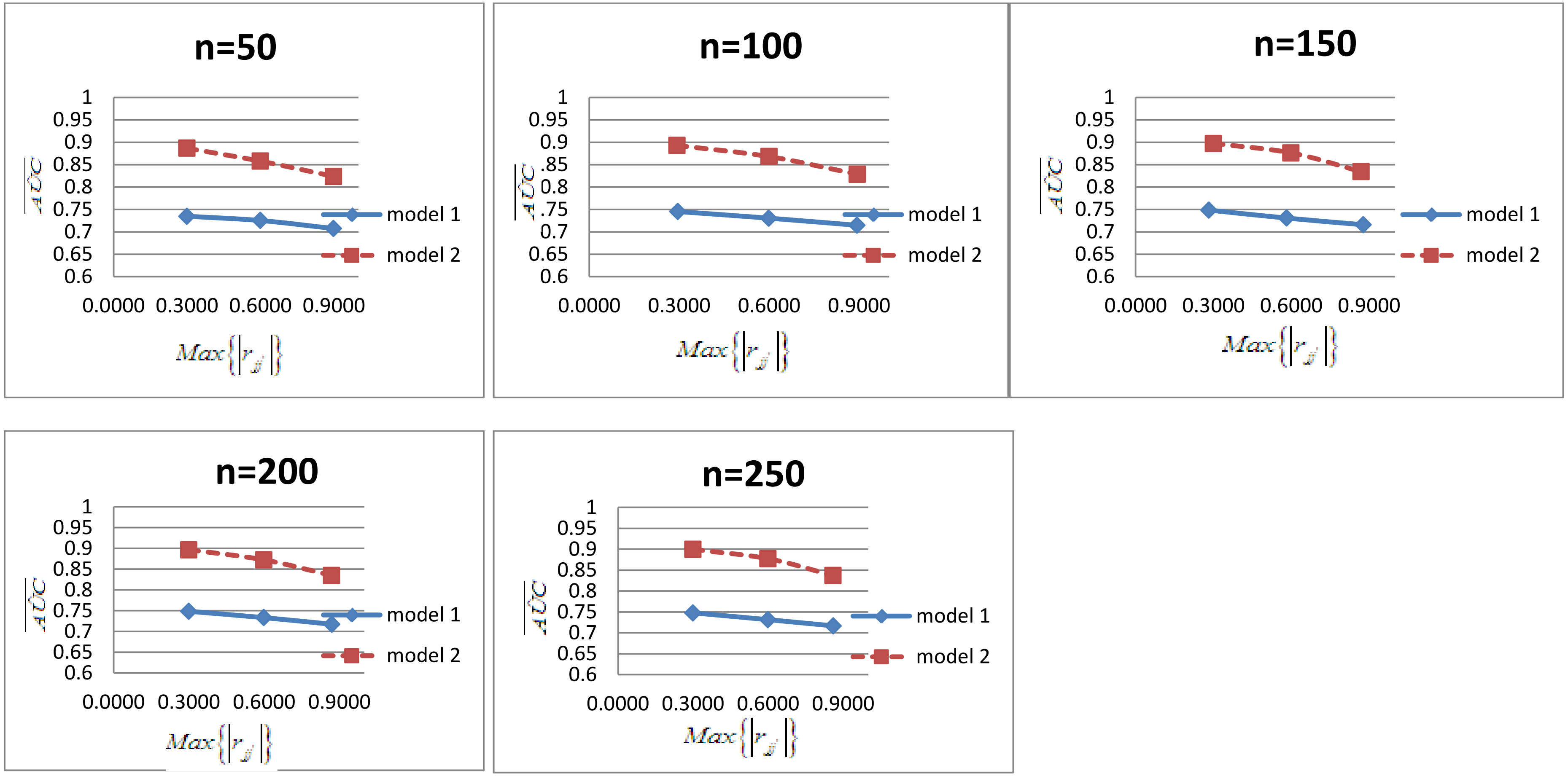
ตารางที่ 4.8 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัว และ 5 ตัว (p=(3,5)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\{\|r_{jj'}\|\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
3	5	50	Low	0.7349	0.8867	model 2
			Medium	0.7253	0.8578	model 2
			High	0.7075	0.8239	model 2
		100	Low	0.7453	0.8929	model 2
			Medium	0.7306	0.8680	model 2
			High	0.7147	0.8285	model 2
		150	Low	0.7486	0.8974	model 2
			Medium	0.7308	0.8763	model 2
			High	0.7159	0.8346	model 2
		200	Low	0.7486	0.8969	model 2
			Medium	0.7333	0.8725	model 2
			High	0.7173	0.8350	model 2
		250	Low	0.7478	0.8996	model 2
			Medium	0.7314	0.8774	model 2
			High	0.7166	0.8370	model 2

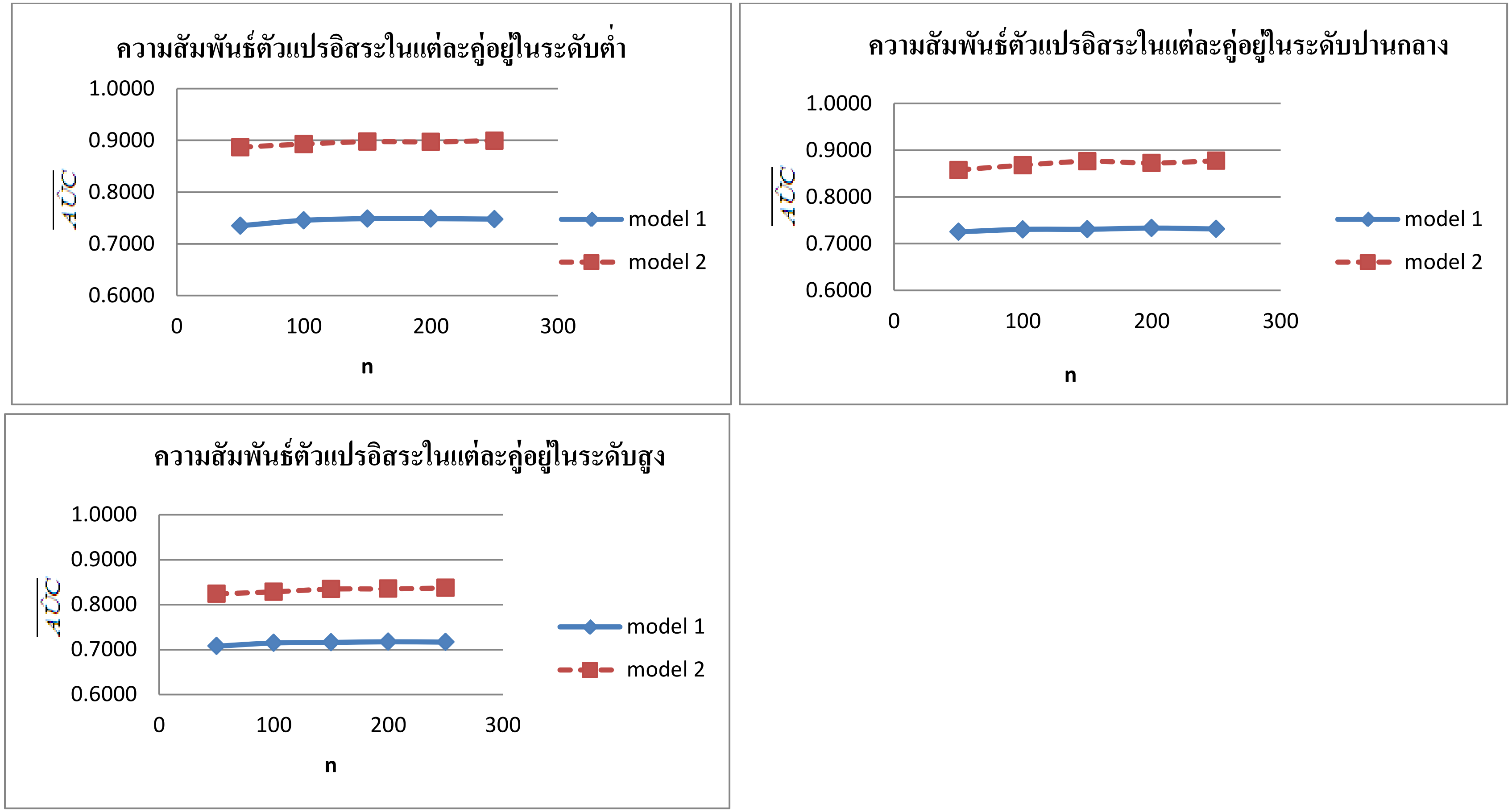
จากตารางที่ 4.8 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 5 ตัว (p=(3,5)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 3 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC น้อยกว่าตัวแปรอิสระ 5 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \geq \mu_{model2}$ และ $H_a : \mu_{model1} < \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 2 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.15 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 5 ตัว (p=(3,5)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.16 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 3 ตัวและ 5 ตัว (p=(3,5)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



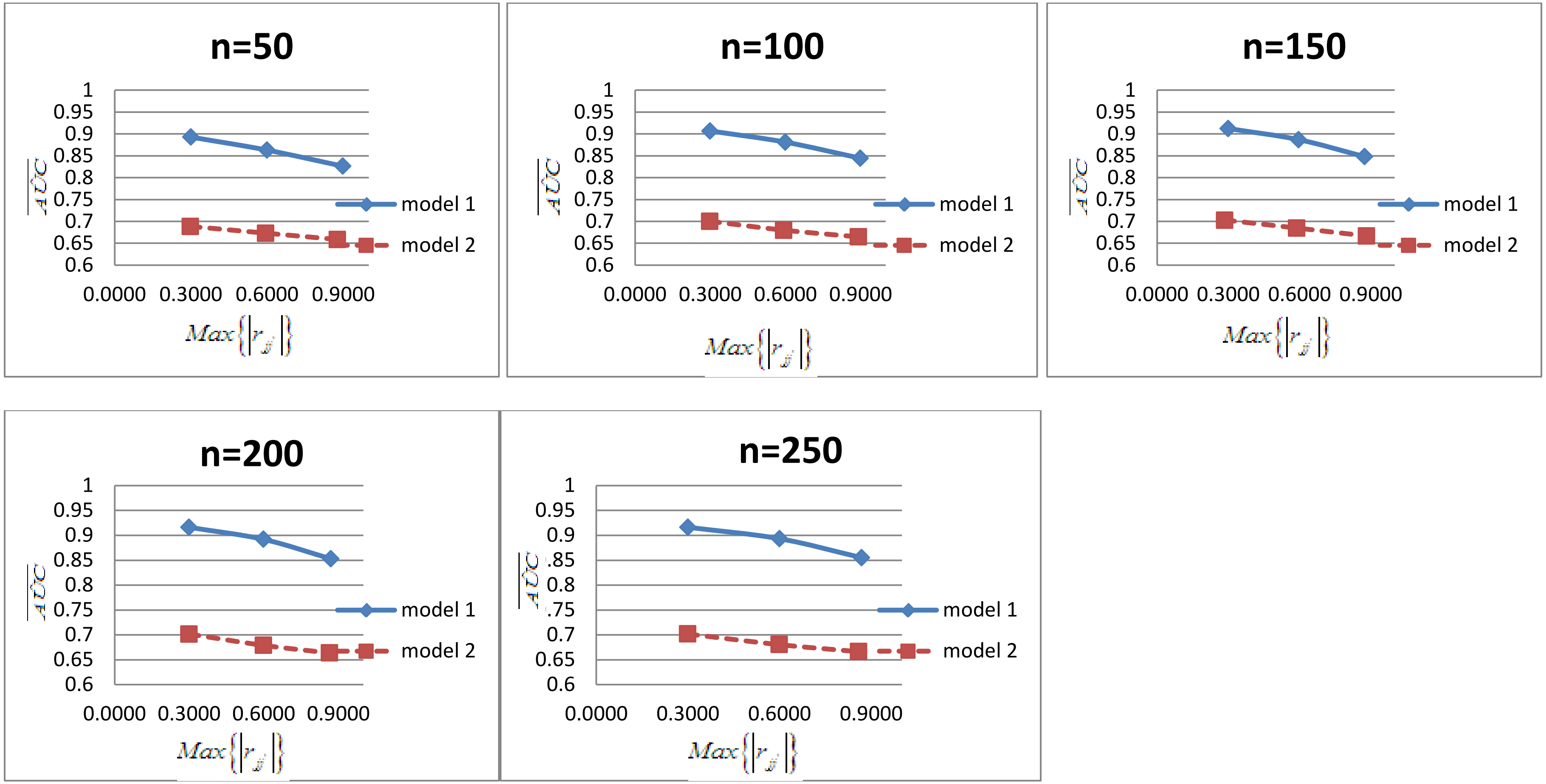
ตารางที่ 4.9 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัว และ 2 ตัว (p=(4,2)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\{\lvert r_{jj'}\rvert\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
4	2	50	Low	0.8929	0.6886	model 1
			Medium	0.8631	0.6732	model 1
			High	0.8263	0.6587	model 1
		100	Low	0.9069	0.7002	model 1
			Medium	0.8813	0.6800	model 1
			High	0.8445	0.6646	model 1
		150	Low	0.9124	0.7031	model 1
			Medium	0.8869	0.6848	model 1
			High	0.8481	0.6667	model 1
		200	Low	0.9162	0.7016	model 1
			Medium	0.8919	0.6790	model 1
			High	0.8528	0.6636	model 1
		250	Low	0.9163	0.7017	model 1
			Medium	0.8933	0.6804	model 1
			High	0.8552	0.6662	model 1

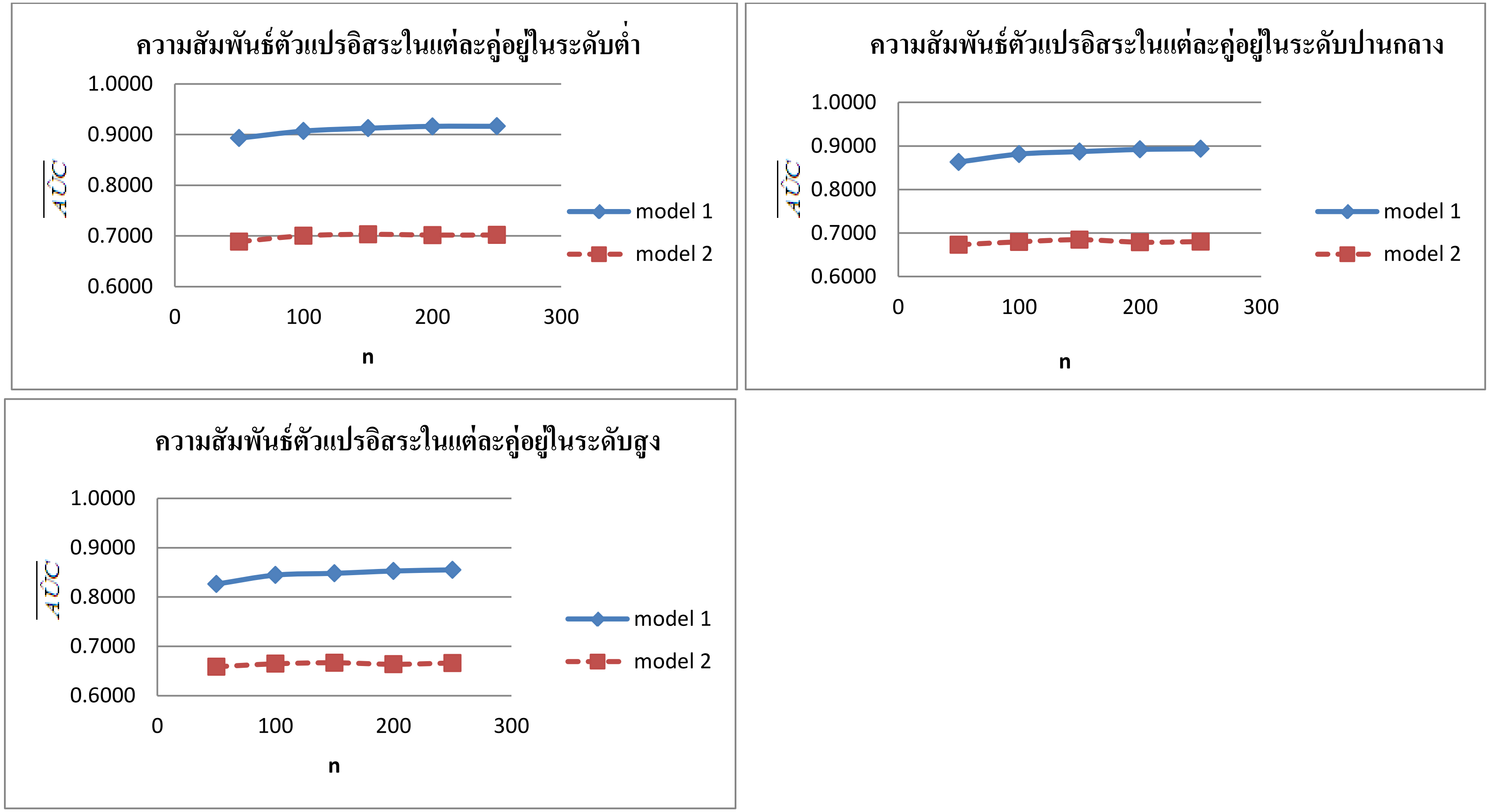
จากตารางที่ 4.9 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 2 ตัว (p=(4,2)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 4 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 2 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{\mathrm{mod}el1} \leq \mu_{\mathrm{mod}el2}$ และ $H_a : \mu_{\mathrm{mod}el1} > \mu_{\mathrm{mod}el2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบคือค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.17 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 2 ตัว (p=(4,2)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.18 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 2 ตัว (p=(4,2)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



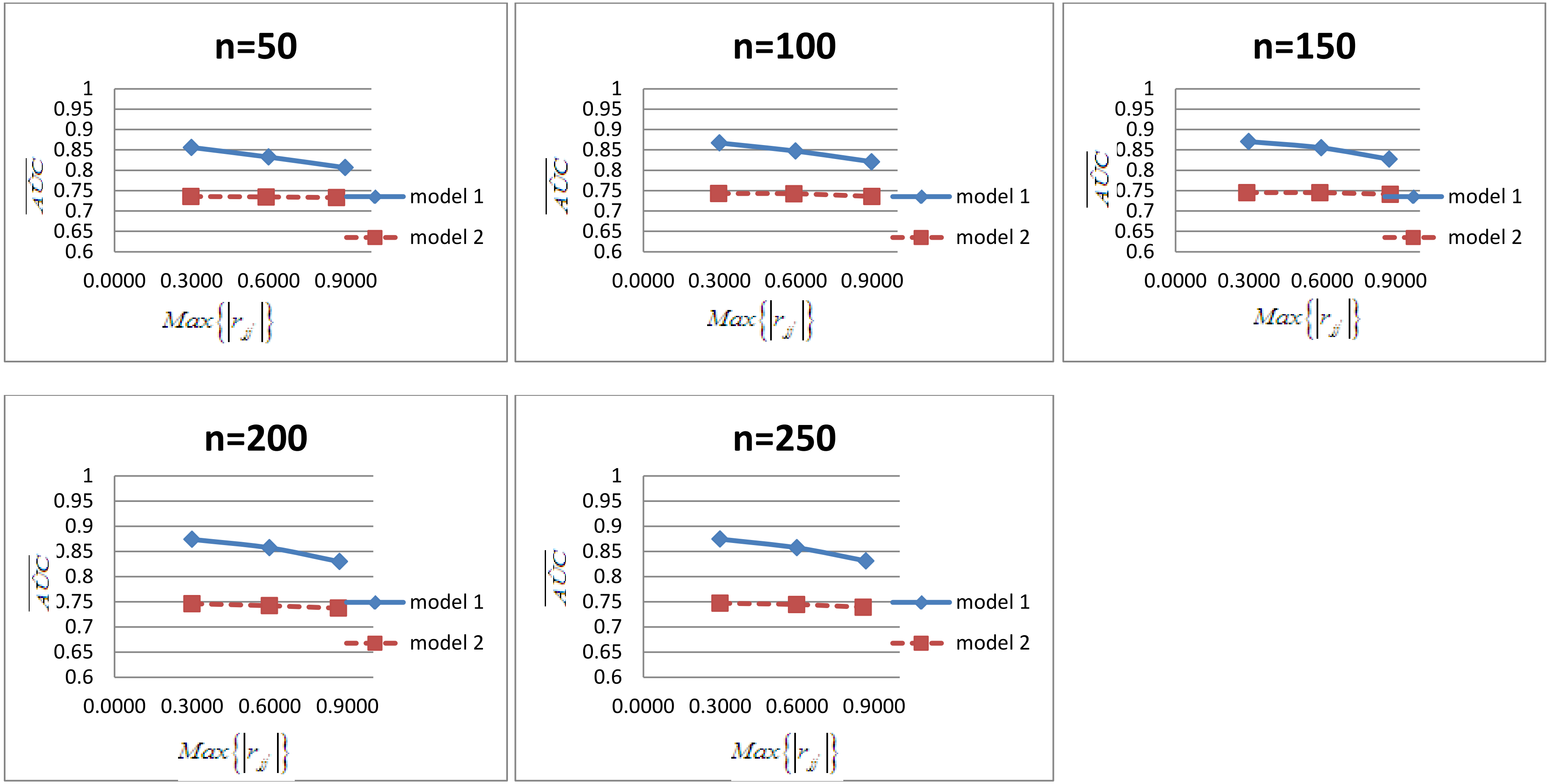
ตารางที่ 4.10 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัว และ 3 ตัว (p=(4,3)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
4	3	50	Low	0.8561	0.7356	model 1
			Medium	0.8324	0.7346	model 1
			High	0.8065	0.7329	model 1
		100	Low	0.8670	0.7429	model 1
			Medium	0.8470	0.7424	model 1
			High	0.8210	0.7358	model 1
		150	Low	0.8703	0.7451	model 1
			Medium	0.8551	0.7449	model 1
			High	0.8272	0.7410	model 1
		200	Low	0.8741	0.7463	model 1
			Medium	0.8574	0.7423	model 1
			High	0.8298	0.7373	model 1
		250	Low	0.8746	0.7472	model 1
			Medium	0.8574	0.7447	model 1
			High	0.8313	0.7393	model 1

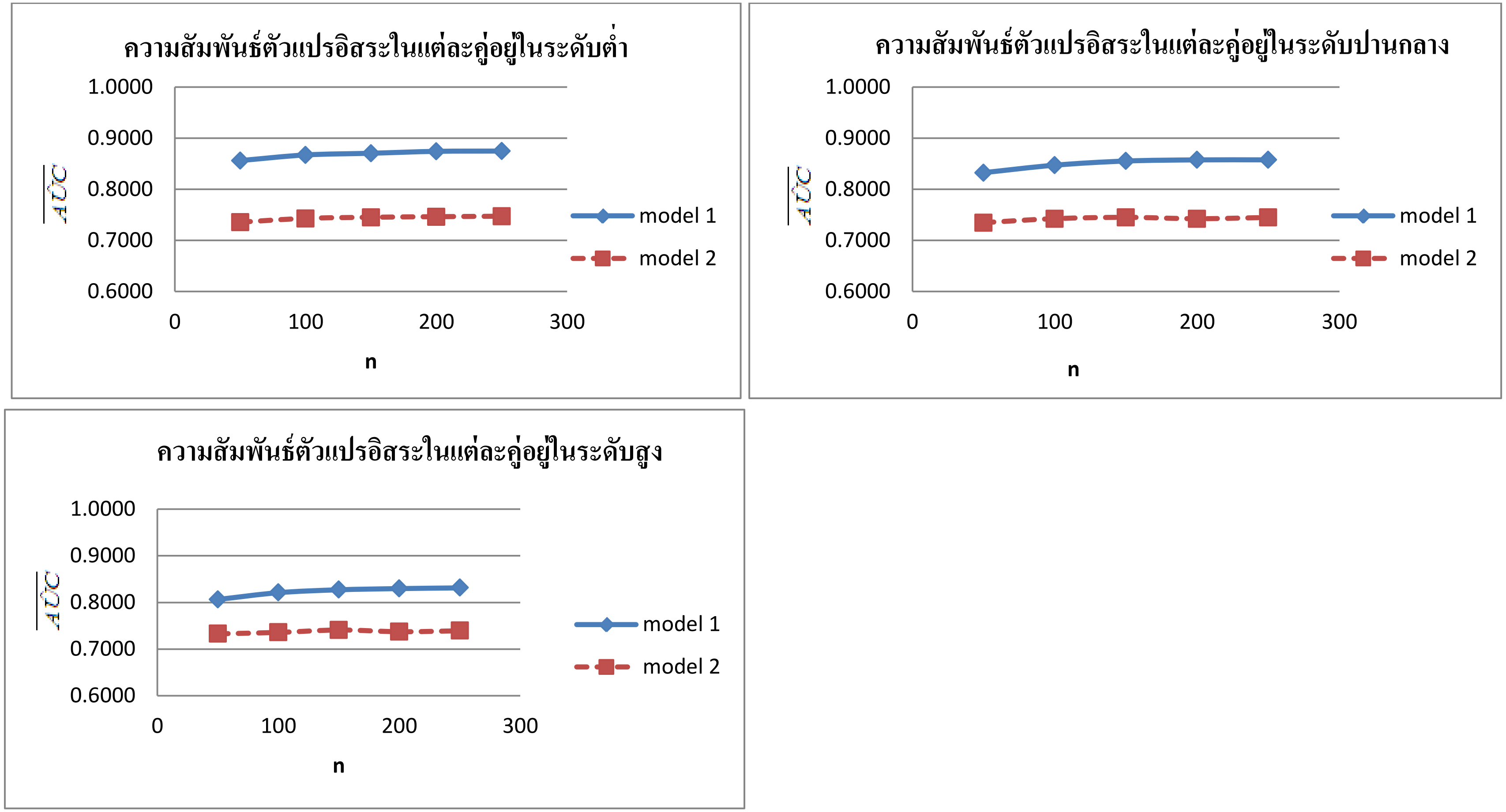
จากตารางที่ 4.10 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 3 ตัว (p=(4,3)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 4 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 3 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \le \mu_{model2}$ และ $H_a : \mu_{model1} > \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.19 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 3 ตัว (p=(4,3)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.20 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ3 ตัว (p=(4,3)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



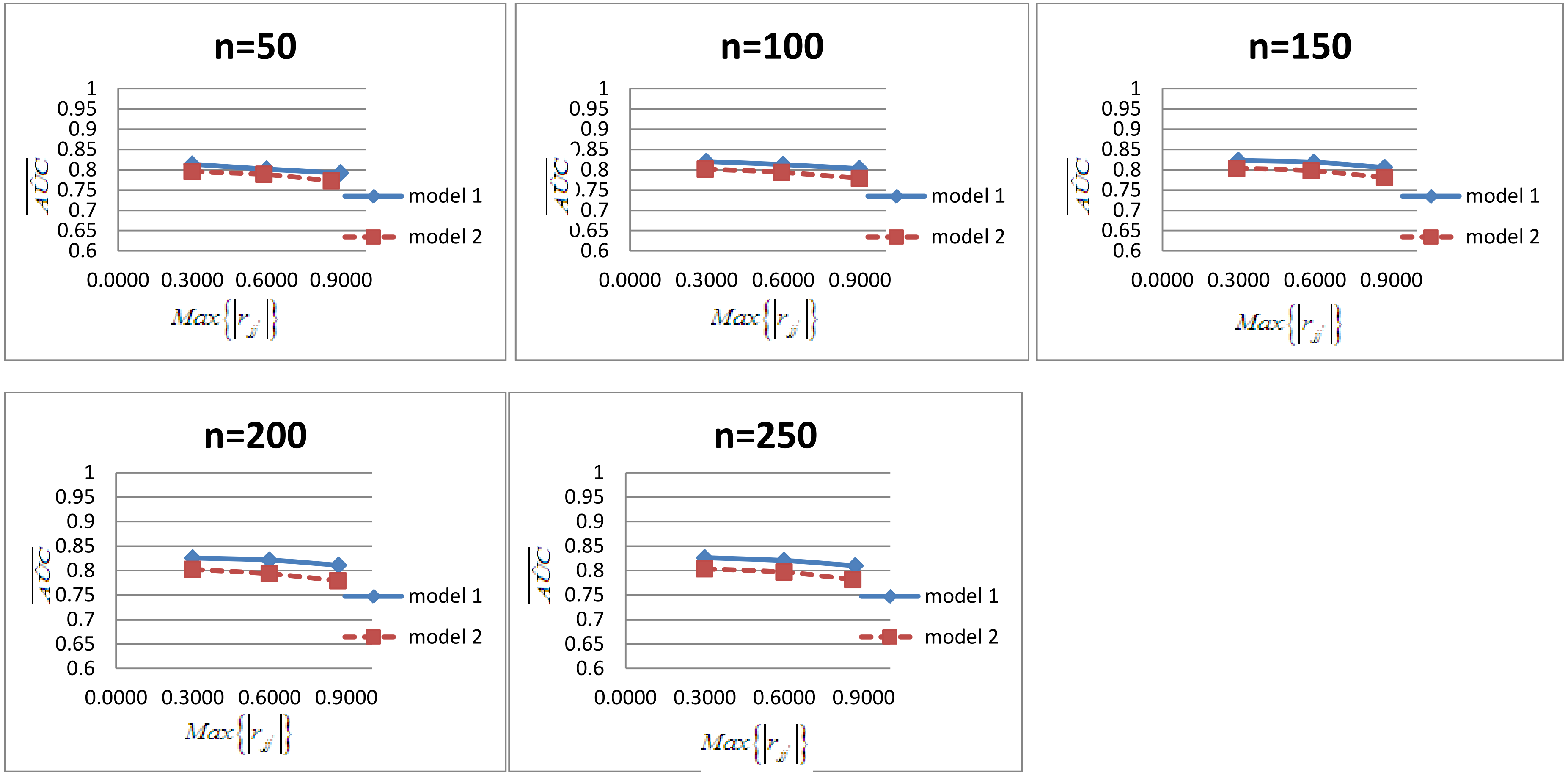
ตารางที่ 4.11 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัว และ 4 ตัว (p=(4,4)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
4	4	50	Low	0.8135	0.7954	model 1
			Medium	0.8014	0.7885	model 1 หรือ model 2
			High	0.7921	0.7728	model 1
		100	Low	0.8201	0.8018	model 1
			Medium	0.8126	0.7937	model 1
			High	0.8025	0.7790	model 1
		150	Low	0.8229	0.8035	model 1
			Medium	0.8185	0.7978	model 1
			High	0.8053	0.7810	model 1
		200	Low	0.8257	0.8024	model 1
			Medium	0.8214	0.7934	model 1
			High	0.8108	0.7792	model 1
		250	Low	0.8263	0.8035	model 1
			Medium	0.8205	0.7968	model 1
			High	0.8098	0.7816	model 1

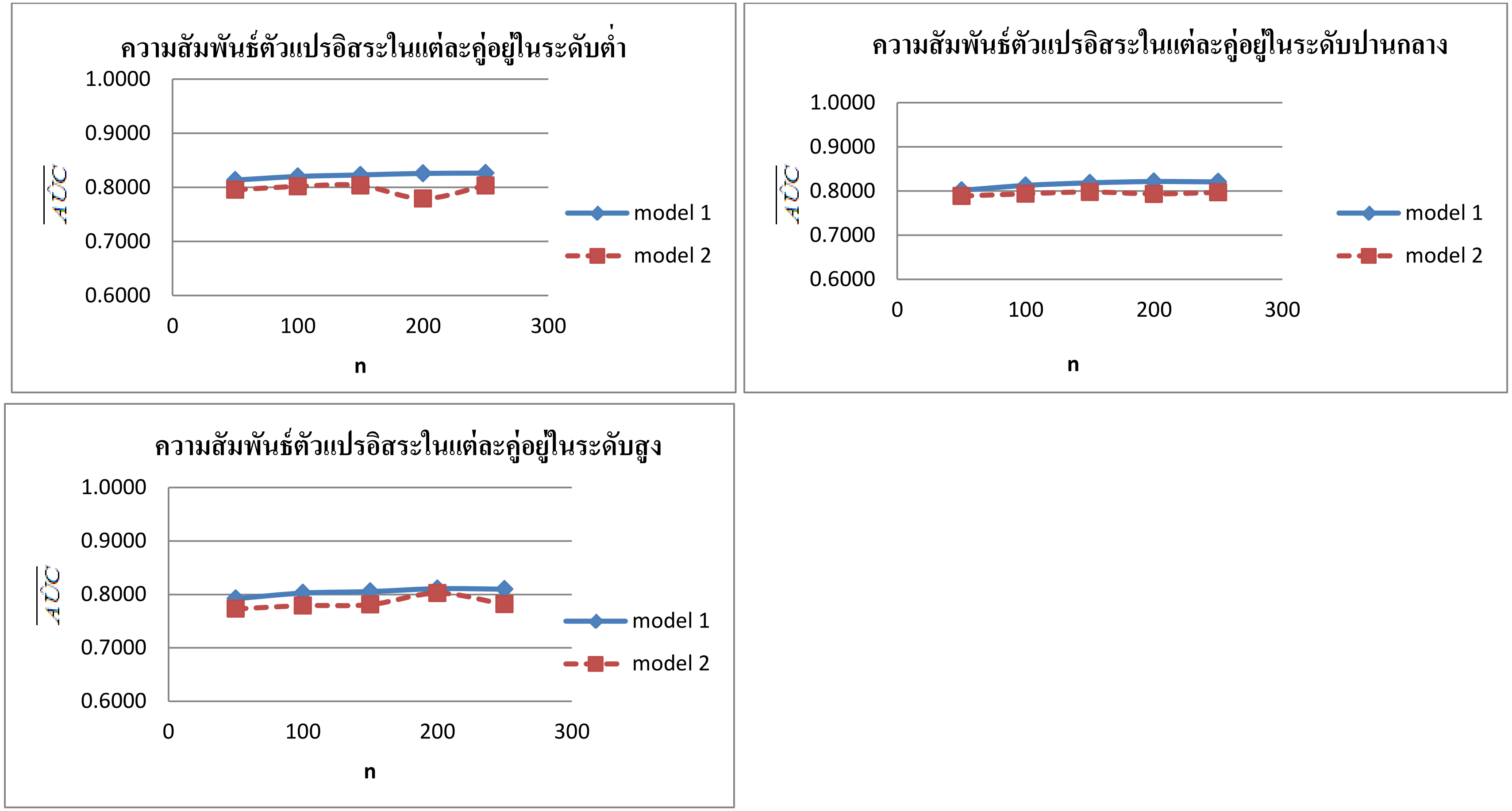
จากตารางที่ 4.11 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 4 ตัว (p=(4,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 4 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 4 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \leq \mu_{model2}$ และ $H_a : \mu_{model1} > \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบคือค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.21 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 4 ตัว (p=(4,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.22 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 4 ตัว (p=(4,4)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



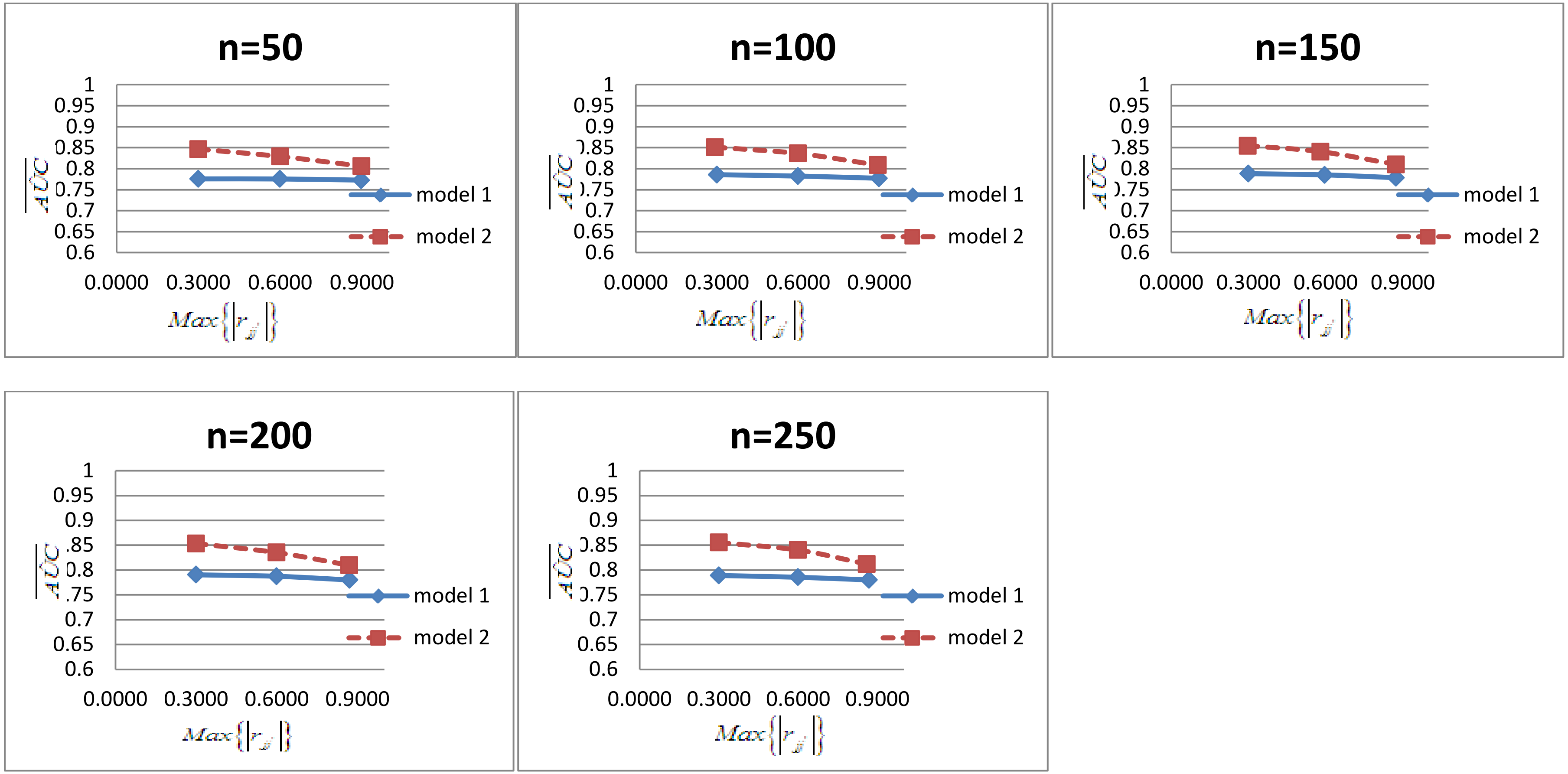
ตารางที่ 4.12 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัว และ 5 ตัว (p=(4,5)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
4	5	50	Low	0.7757	0.8467	model 2
			Medium	0.7755	0.8290	model 2
			High	0.7727	0.8060	model 2
		100	Low	0.7856	0.8511	model 2
			Medium	0.7824	0.8363	model 2
			High	0.7772	0.8086	model 2
		150	Low	0.7881	0.8548	model 2
			Medium	0.7852	0.8405	model 2
			High	0.7785	0.8102	model 2
		200	Low	0.7904	0.8534	model 2
			Medium	0.7875	0.8353	model 2
			High	0.7801	0.8098	model 2
		250	Low	0.7891	0.8554	model 2
			Medium	0.7855	0.8404	model 2
			High	0.7801	0.8121	model 2

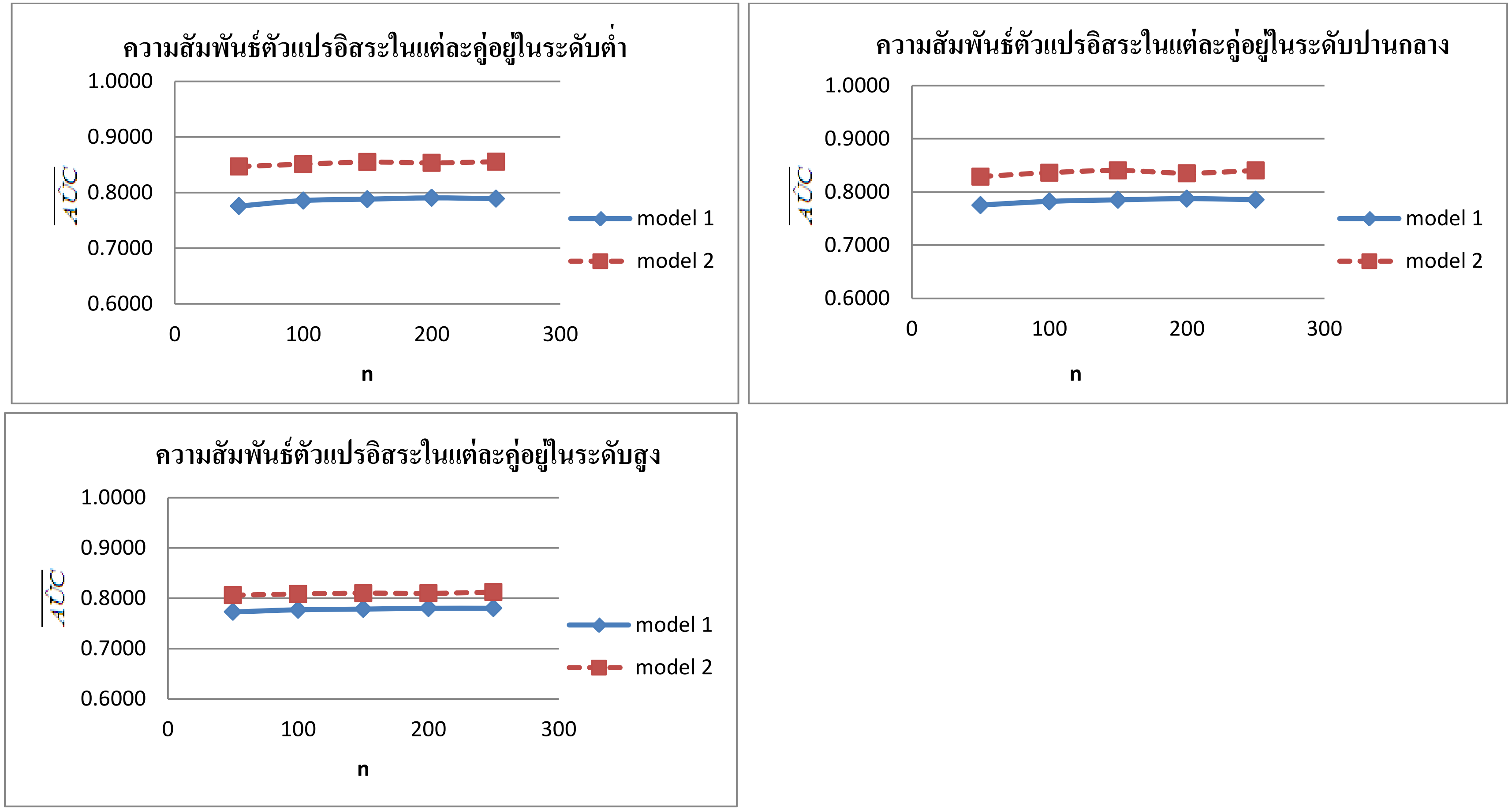
จากตารางที่ 4.12 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 5 ตัว (p=(4,5)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 4 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC น้อยกว่าตัวแปรอิสระ 5 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \geq \mu_{model2}$ และ $H_a : \mu_{model1} < \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 2 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.23 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 5 ตัว (p=(4,5)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.24 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 4 ตัวและ 5 ตัว (p=(4,5)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



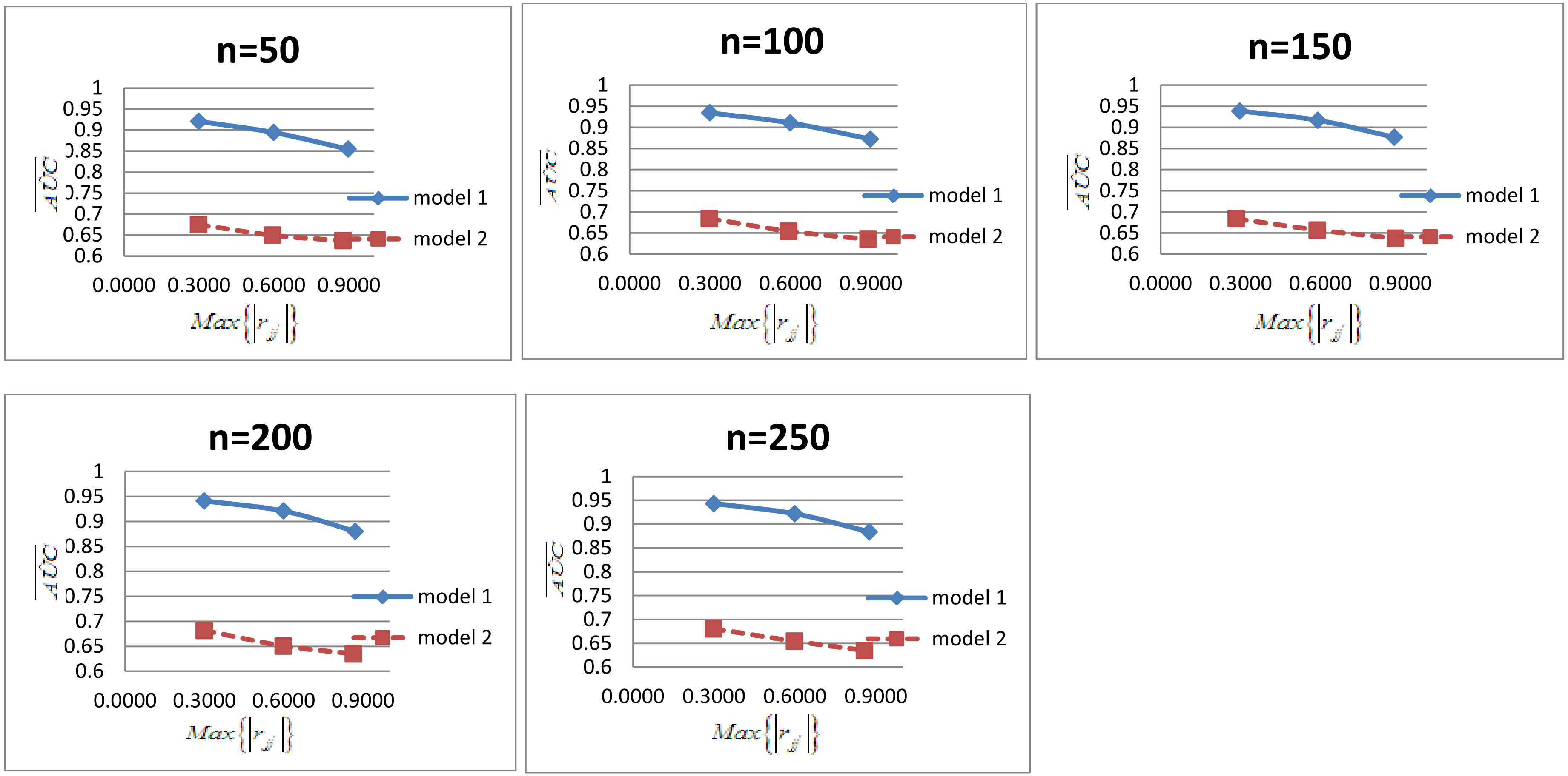
ตารางที่ 4.13 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัว และ 2 ตัว (p=(5,2)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
5	2	50	Low	0.9209	0.6749	model 1
			Medium	0.8942	0.6495	model 1
			High	0.8547	0.6368	model 1
		100	Low	0.9345	0.6842	model 1
			Medium	0.9107	0.6543	model 1
			High	0.8723	0.6352	model 1
		150	Low	0.9387	0.6837	model 1
			Medium	0.9169	0.6572	model 1
			High	0.8765	0.6375	model 1
		200	Low	0.9411	0.6818	model 1
			Medium	0.9208	0.6506	model 1
			High	0.8802	0.6350	model 1
		250	Low	0.9433	0.6805	model 1
			Medium	0.9217	0.6545	model 1
			High	0.8838	0.6348	model 1

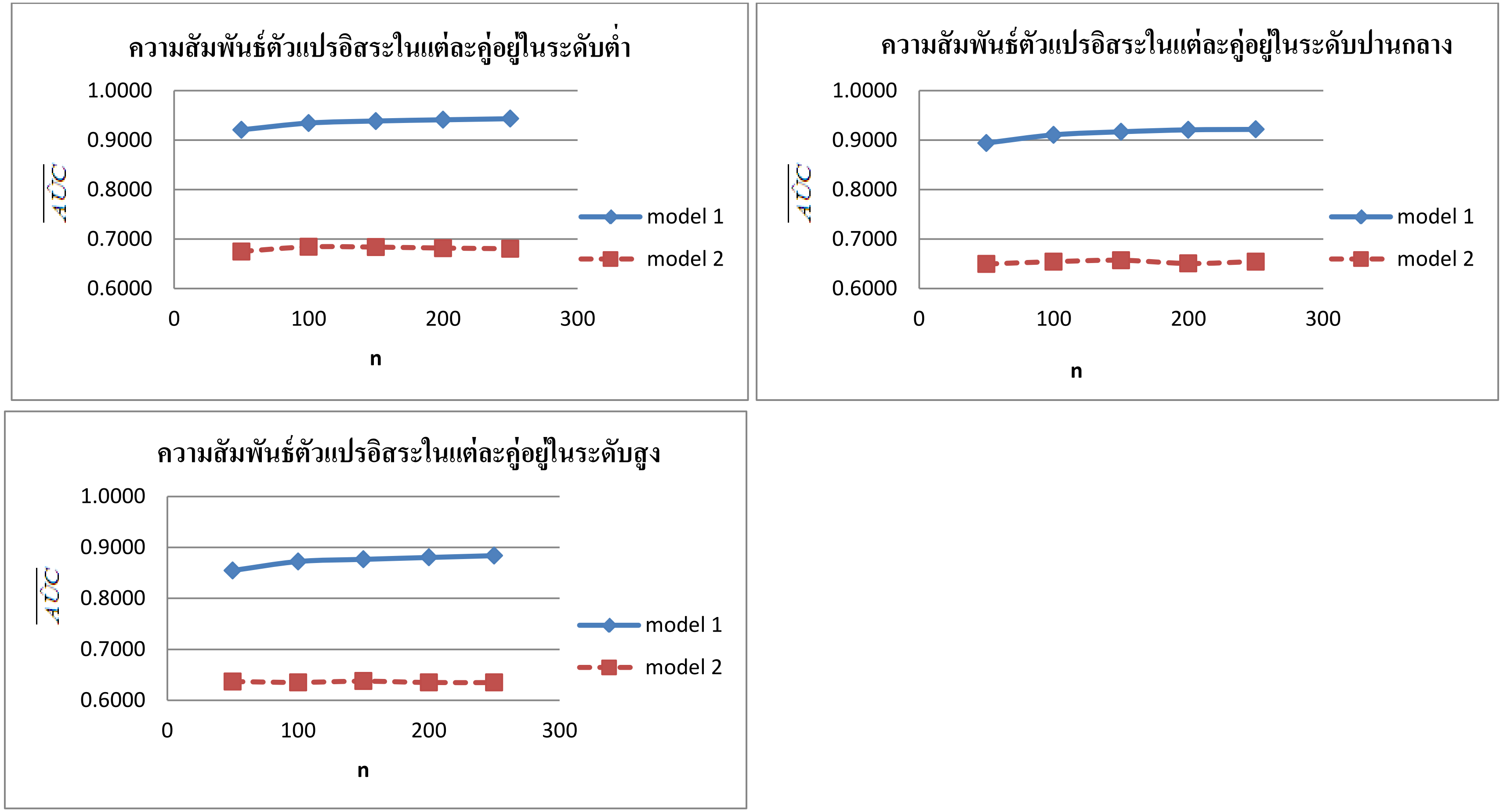
จากตารางที่ 4.13 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 2 ตัว (p=(5,2)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 5 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 2 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{\mathrm{mod}el1} \leq \mu_{\mathrm{mod}el2}$ และ $H_a : \mu_{\mathrm{mod}el1} > \mu_{\mathrm{mod}el2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.25 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 2 ตัว (p=(5,2)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.26 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 2 ตัว (p=(5,2)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



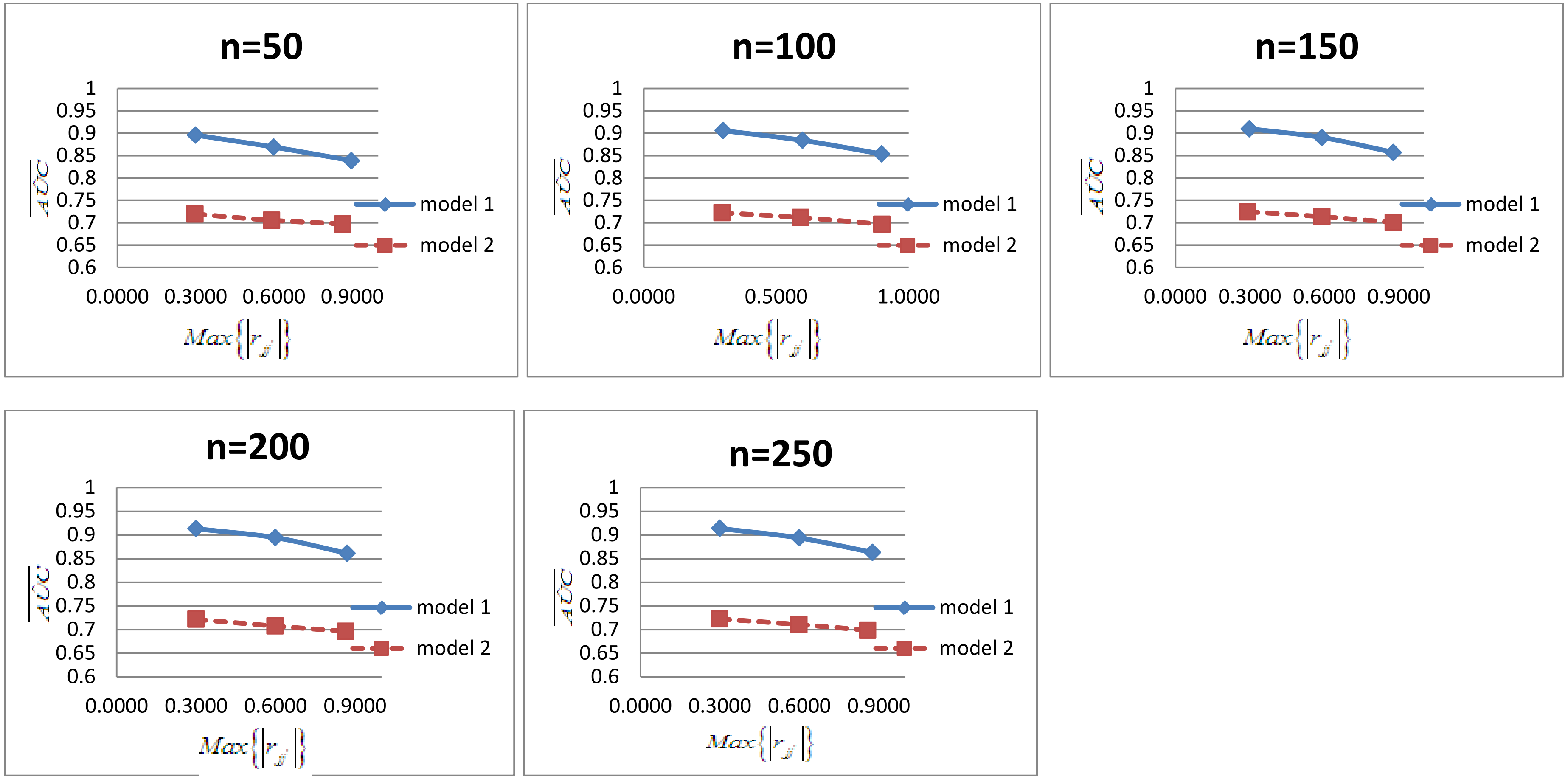
ตารางที่ 4.14 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัว และ 3 ตัว (p=(5,3)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
5	3	50	Low	0.8958	0.7194	model 1
			Medium	0.8690	0.7054	model 1
			High	0.8388	0.6972	model 1
		100	Low	0.9058	0.7224	model 1
			Medium	0.8840	0.7114	model 1
			High	0.8536	0.6966	model 1
		150	Low	0.9099	0.7247	model 1
			Medium	0.8904	0.7135	model 1
			High	0.8571	0.7004	model 1
		200	Low	0.9131	0.7217	model 1
			Medium	0.8942	0.7073	model 1
			High	0.8611	0.6963	model 1
		250	Low	0.9136	0.7225	model 1
			Medium	0.8936	0.7106	model 1
			High	0.8630	0.6985	model 1

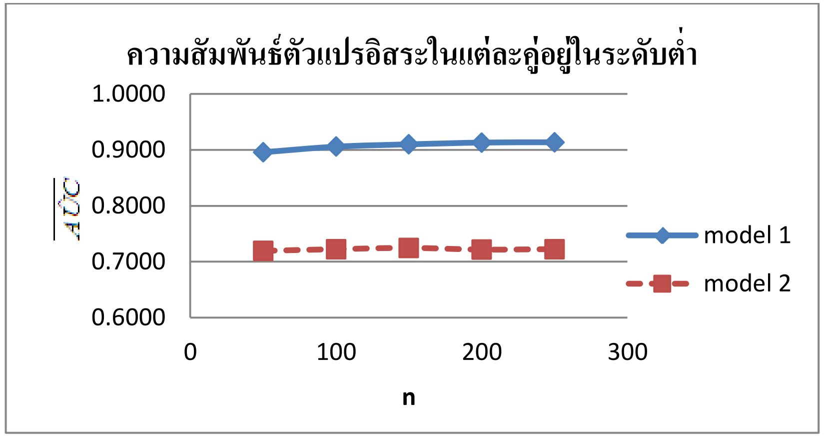
จากตารางที่ 4.14 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 3 ตัว (p=(5,3)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 5 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 3 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

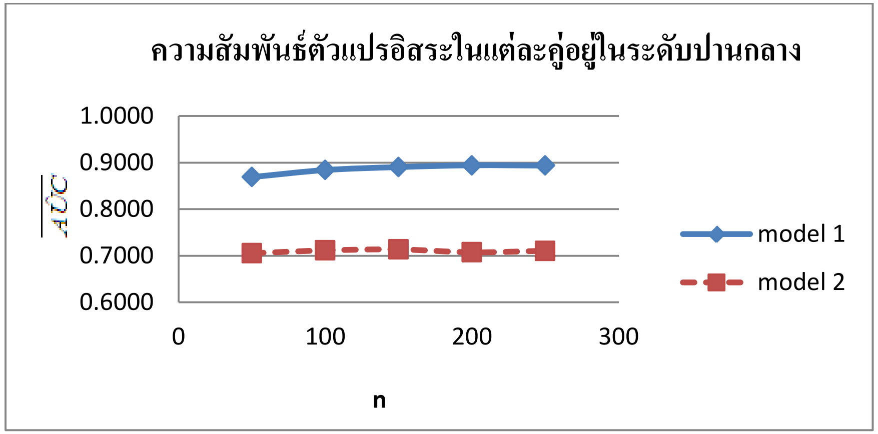
จากการทดสอบสมมติฐาน $H_0 : \mu_{\mathrm{mod}el1} \leq \mu_{\mathrm{mod}el2}$ และ $H_a : \mu_{\mathrm{mod}el1} > \mu_{\mathrm{mod}el2}$ ที่ระดับนัยสำคัญ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

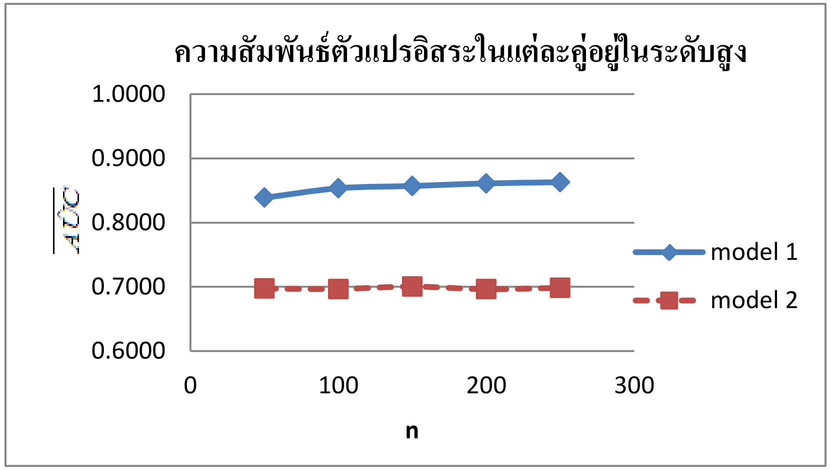
ภาพที่ 4.27 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 3 ตัว (p=(5,3)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.28 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 3 ตัว (p=(5,3)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง







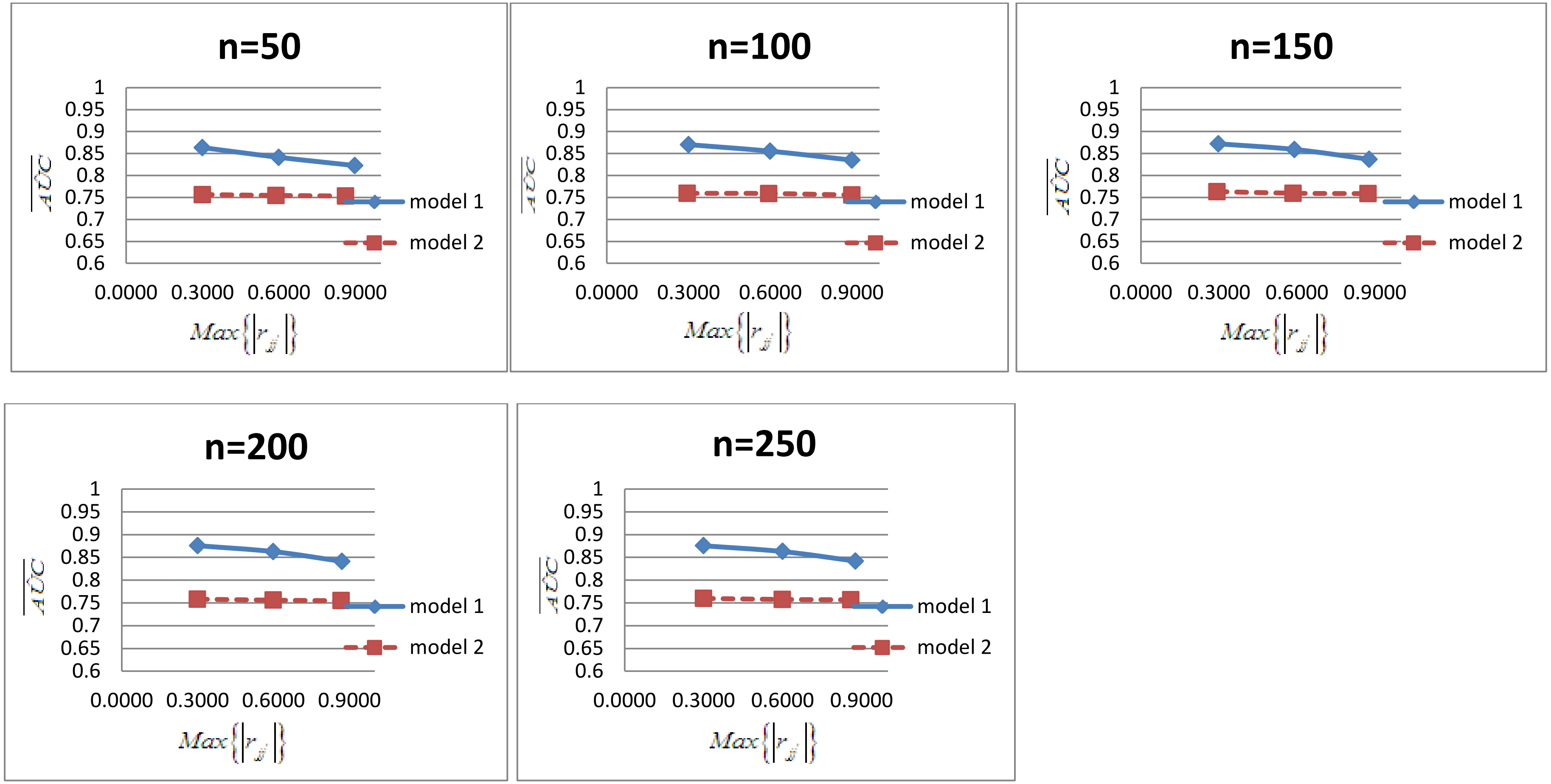
ตารางที่ 4.15 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัว และ 4 ตัว (p=(5,4)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\left\{\left\|r_{jj'}\right\|\right\}$	$\overline{A\hat{U}C}_{\text{model}1}$	$\overline{A\hat{U}C}_{\text{model}2}$	ตัวแบบที่เลือก
5	4	50	Low	0.8635	0.7566	model 1
			Medium	0.8414	0.7546	model 1
			High	0.8226	0.7533	model 1
		100	Low	0.8703	0.7595	model 1
			Medium	0.8553	0.7590	model 1
			High	0.8350	0.7556	model 1
		150	Low	0.8721	0.7635	model 1
			Medium	0.8593	0.7596	model 1
			High	0.8369	0.7587	model 1
		200	Low	0.8758	0.7583	model 1
			Medium	0.8627	0.7561	model 1
			High	0.8413	0.7549	model 1
		250	Low	0.8756	0.7600	model 1
			Medium	0.8629	0.7576	model 1
			High	0.8419	0.7569	model 1

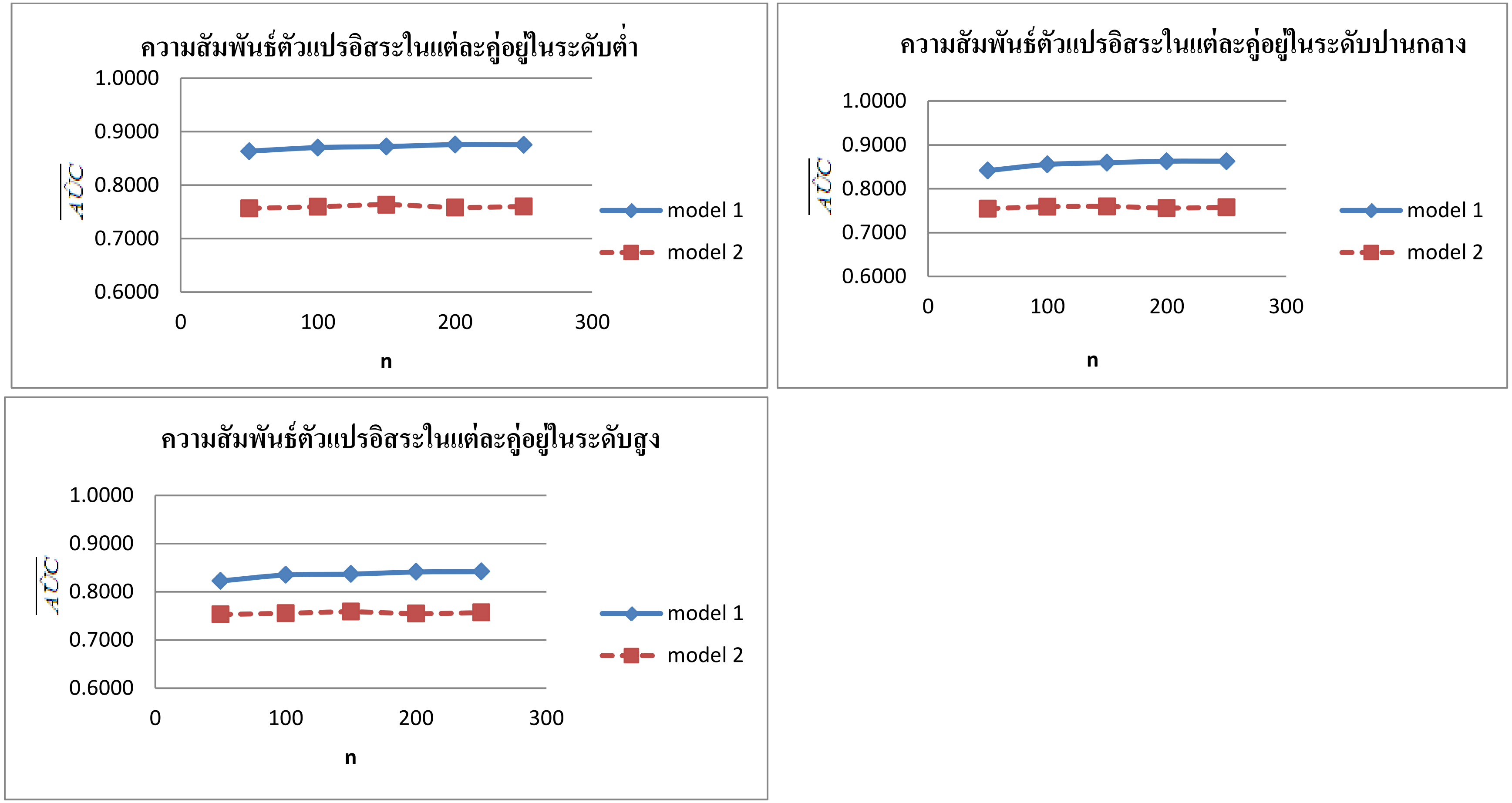
จากตารางที่ 4.15 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 4 ตัว (p=(5,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 5 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 4 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \leq \mu_{model2}$ และ $H_a : \mu_{model1} > \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.29 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 4 ตัว (p=(5,4)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.30 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 4 ตัว (p=(5,4)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



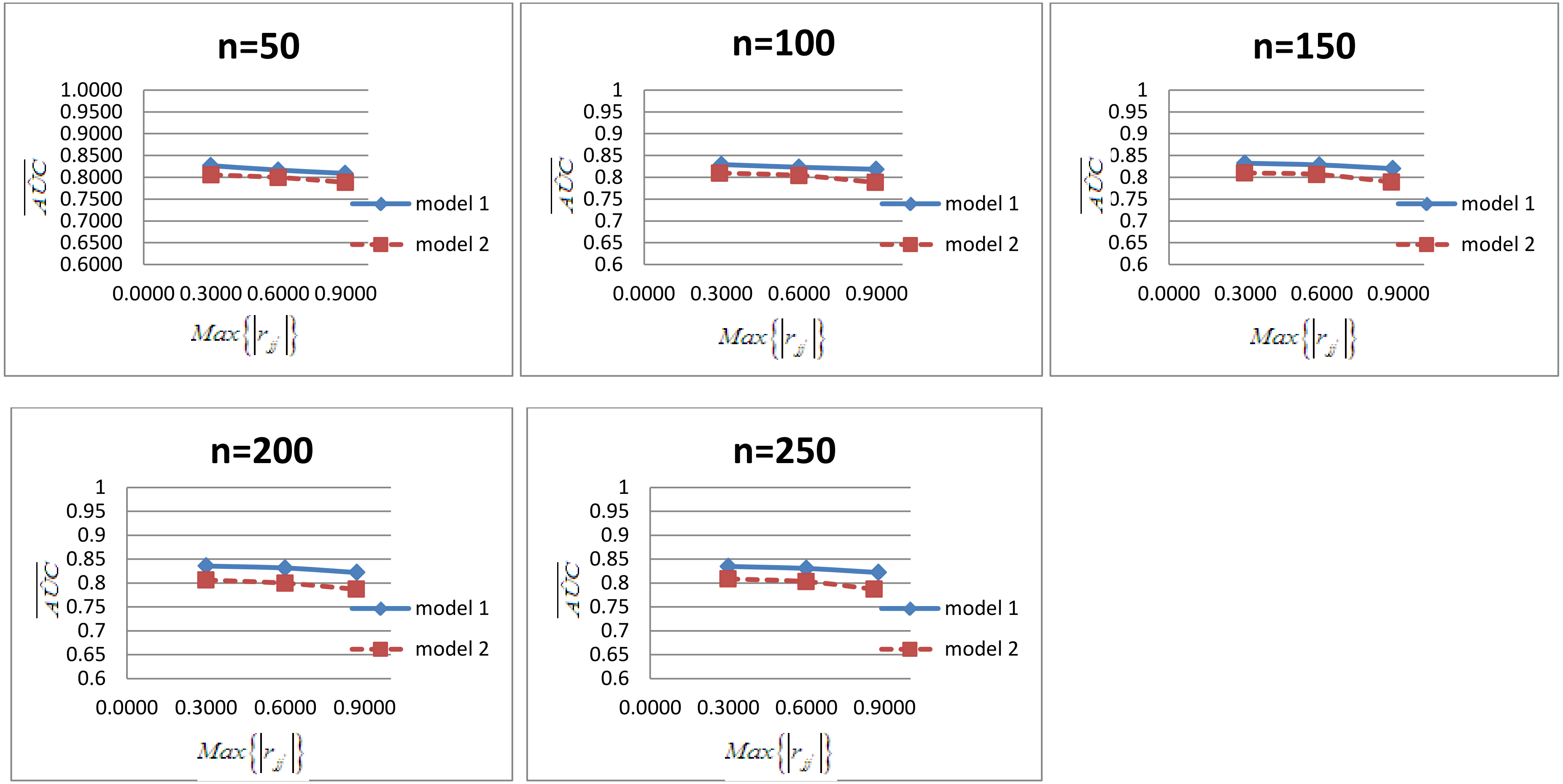
ตารางที่ 4.16 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ของตัวแบบ เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัว และ 5 ตัว (p=(5,5)) ขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับน้อย ปานกลางและมาก

p_1	p_2	n	$Max\{\lvert r_{jj'}\rvert\}$	$\overline{A\hat{U}C}_{model1}$	$\overline{A\hat{U}C}_{model2}$	ตัวแบบที่เลือก
5	5	50	Low	0.8267	0.8057	model 1
			Medium	0.8167	0.7997	model 1
			High	0.8091	0.7886	model 1
		100	Low	0.8293	0.8097	model 1
			Medium	0.8232	0.8043	model 1
			High	0.8184	0.7883	model 1
		150	Low	0.8323	0.8100	model 1
			Medium	0.8284	0.8066	model 1
			High	0.8199	0.7894	model 1
		200	Low	0.8356	0.8064	model 1
			Medium	0.8315	0.7997	model 1
			High	0.8220	0.7869	model 1
		250	Low	0.8348	0.8087	model 1
			Medium	0.8305	0.8030	model 1
			High	0.8220	0.7870	model 1

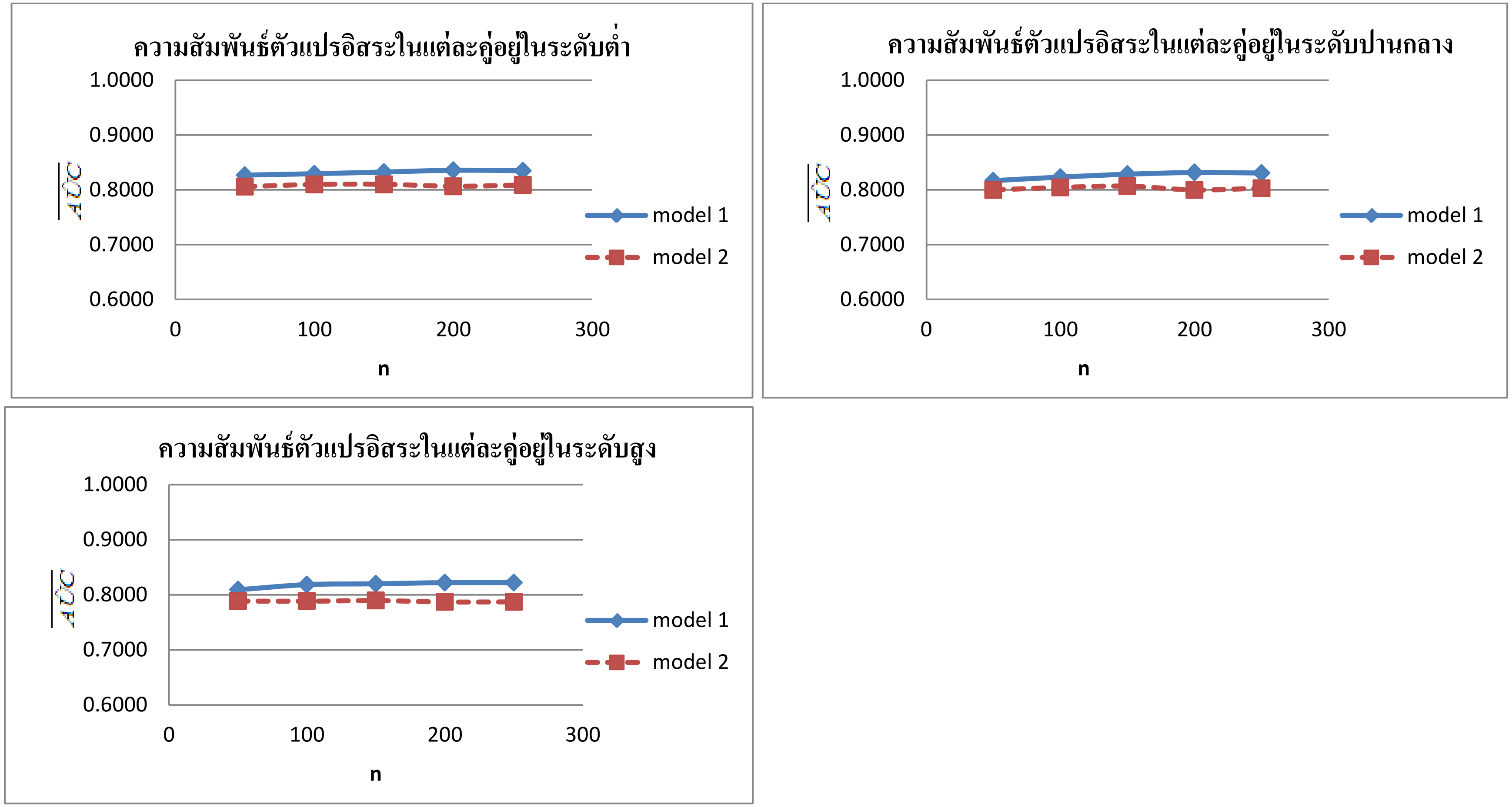
จากตารางที่ 4.16 เมื่อพิจารณาค่าเฉลี่ยพื้นที่ใต้กราฟ ROC เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 5 ตัว (p=(5,5)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 ระดับความสัมพันธ์แต่ละคู่สัมพันธ์กันในระดับน้อย ปานกลาง มาก พบว่าตัวแปรอิสระที่เท่ากับ 5 ตัวในตัวแบบที่ 1 จะให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC มากกว่าตัวแปรอิสระ 5 ตัวในตัวแบบที่ 2 และในขนาดตัวอย่างที่เท่ากัน เมื่อระดับความสัมพันธ์ในแต่ละคู่มีค่าเพิ่มมากขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้กราฟ ROC ลดลง นั่นคือมีความน่าเชื่อถือในตัวแบบลดลง

จากการทดสอบสมมติฐาน $H_0 : \mu_{model1} \leq \mu_{model2}$ และ $H_a : \mu_{model1} > \mu_{model2}$ ที่ระดับนัยสำคัญทางสถิติ 0.05 ผลการทดสอบพบว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2 สามารถสรุปได้ว่า เราสามารถตัดสินเลือกตัวแบบที่ 1 ได้ทุกกรณี(P-value<0.05)

ภาพที่ 4.31 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 5 ตัว (p=(5,5)) และขนาดตัวอย่าง (n) เท่ากับ 50 100 150 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลง



ภาพที่ 4.32 แสดงค่าเฉลี่ยพื้นที่ใต้โค้ง ROC ที่เหมาะสม เมื่อมีคู่ของตัวแปรอิสระเท่ากับ 5 ตัวและ 5 ตัว (p=(5,5)) และค่าความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $Max\left\{\left|r_{jj'}\right|\right\}$ ในระดับน้อย ปานกลางและมาก เมื่อขนาดตัวอย่าง (n) เปลี่ยนแปลง



บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีจุดประสงค์เพื่อศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง ในแต่ละสถานการณ์ที่ต้องการศึกษา ซึ่งใช้พื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการคัดเลือก ดังนี้

1. 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)

2. กำหนดระดับความสัมพันธ์เริ่มต้นของตัวแปรอิสระเป็นดังนี้

   2.1 ตัวแปรอิสระมีความสัมพันธ์กันในระดับอย่างต่ำ: $0 < Max\left\{\left|r_{jj'}\right|\right\} \le 0.33$

   2.2 ตัวแปรอิสระมีความสัมพันธ์กันในระดับปานกลาง: $0.33 < Max\left\{\left|r_{jj'}\right|\right\} \le 0.66$

   2.3 ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูง: $0.66 < Max\left\{\left|r_{jj'}\right|\right\} \le 0.99$

3. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 50, 100, 150, 200 และ 250

การคัดเลือกตัวแบบในการวิจัยครั้งนี้ใช้พื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการพิจารณาคัดเลือกตัวแบบ และใช้การทดสอบสมมติฐานในการเปรียบเทียบเพื่อคัดเลือกตัวแบบ

การจำลองข้อมูลในแต่ละสถานการณ์จะจำลองขึ้นโดยใช้เทคนิคมอนติคาร์โลด้วยโปรแกรม R ซึ่งกำหนดระดับนัยสำคัญ (α) ในการวิจัยครั้งนี้ที่ระดับ 0.05 ผลการวิจัยมีข้อสรุปดังนี้

5.1 สรุปผลการวิจัย

ผลการวิจัยเรื่องการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง แบ่งเป็น 2 ตอน ได้แก่

5.1.1 ปัจจัยที่เกี่ยวข้องกับการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยงในแต่ละตัวแบบ

ปัจจัยที่เกี่ยวข้องกับการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยงที่ทำให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เปลี่ยนแปลงไปในแต่ละตัวแบบโดยมีปัจจัยที่เกี่ยวข้อง ดังนี้

<u>ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ</u>

ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $\left(Max\left\{\left|r_{jj'}\right|\right\}\right)$ ในการศึกษาครั้งนี้คือ ระดับความสัมพันธ์ระดับต่ำ ระดับกลางและระดับสูง ซึ่งระดับความสัมพันธ์ระดับต่ำจะอยู่ในช่วง 0 ถึง 0.33 ระดับความสัมพันธ์ระดับกลางจะอยู่ในช่วง 0.33 ถึง 0.66 และระดับความสัมพันธ์ระดับสูงจะอยู่ในช่วง 0.66 ถึง 0.99 เมื่อขนาดตัวอย่างและจำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่ เกณฑ์การคัดเลือกตัวแบบพิจารณาจากค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีค่าสูงสุด ผลการศึกษาพบว่า เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีค่าความสัมพันธ์ในระดับต่ำ ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีค่ามาก เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีค่าความสัมพันธ์ในระดับปานกลาง และระดับสูง ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีค่าลดลง จึงสรุปได้ว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีค่าเพิ่มขึ้น ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC จะมีค่าลดลง ซึ่งหมายถึงตัวแบบจะมีความน่าเชื่อถือลดลงด้วย ภายใต้ขอบเขตขนาดตัวอย่างและจำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่

<u>ขนาดตัวอย่าง</u>

ขนาดตัวอย่างในการศึกษาครั้งนี้เท่ากับ 50, 100, 150, 200 และ 250 เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ และจำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่ เกณฑ์การคัดเลือกตัวแบบพิจารณาจากค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีค่าสูงสุด ผลการศึกษาพบว่า เมื่อขนาดตัวอย่างมีขนาดเล็ก (n=50) ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีค่าน้อย แต่เมื่อขนาดตัวอย่าง

มีขนาดปานกลาง (n=100,150) และมาก (n=200,250) ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มเพิ่มขึ้น จึงสรุปได้ว่า ขนาดตัวอย่างมีขนาดเพิ่มขึ้น ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC จะมีค่าเพิ่มขึ้นด้วย ซึ่งหมายถึงตัวแบบจะมีความน่าเชื่อถือมากขึ้น ภายใต้ขอบเขตระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับเดียวกันและจำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่

<u>จำนวนตัวแปรอิสระ</u>

จากผลการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง จะเห็นว่า เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น แต่ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ และขนาดตัวอย่างมีค่าคงที่ สรุปได้ว่า เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เพิ่มสูงขึ้นด้วย ซึ่งหมายถึงตัวแบบจะมีความน่าเชื่อถือมาก ภายใต้ขอบเขตระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระอยู่ในระดับเดียวกันและขนาดตัวอย่างของแต่ละตัวแบบคงที่

ดังนั้นจึงสามารถสรุปได้ว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ ขนาดตัวอย่าง และจำนวนตัวแปรอิสระ เป็นปัจจัยที่มีผลต่อการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง

5.1.2 ผลการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง

ผลการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ $\left(Max\left\{\left|r_{jj'}\right|\right\}\right)$ เปลี่ยนแปลงไป สามรถแยกได้เป็น 3 กรณี ดังนี้

<u>จำนวนตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่าจำนวนตัวแปรอิสระในตัวแบบที่ 2 (p_1<p_2)</u>

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 เท่ากับระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 2 เนื่องจากตัวแบบที่ 2 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 มากกว่าระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 2 เนื่องจากตัวแบบที่ 2 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

<u>จำนวนตัวแปรอิสระในตัวแบบที่ 1 เท่ากับจำนวนตัวแปรอิสระในตัวแบบที่ 2 (p_1=p_2)</u>

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่าระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 มากกว่าระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 2 เนื่องจากตัวแบบที่ 2 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

<u>จำนวนตัวแปรอิสระในตัวแบบที่ 1 มากกว่าจำนวนตัวแปรอิสระในตัวแบบที่ 2 (p_1>p_2)</u>

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่าระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 เท่ากับระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มากกว่า

5.2 ข้อเสนอแนะ

ผลการวิจัยในครั้งนี้มีข้อเสนอแนะ 2 ด้าน คือ

5.2.1 ด้านการนำไปใช้ประโยชน์

เพื่อเป็นแนวทางในการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยงสำหรับไปใช้ในการจำแนกกลุ่ม ซึ่งจะสามารถหามาจากปัจจัยตัวแปรอิสระที่ถูกคัดเลือก ขนาดตัวอย่าง และความสัมพันธ์ของแต่ละคู่ระหว่างตัวแปรอิสระ

5.2.2 ด้านการศึกษาวิจัย

1. ในการวิจัยครั้งนี้ได้ทำการศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง ในการวิจัยครั้งต่อไปอาจทำการศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Model)

2. ในการวิจัยครั้งนี้ได้ทำการศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง ในการวิจัยครั้งต่อไปอาจทำการศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันเชื่อมโยงแบบอื่นๆ เช่น ฟังก์ชันการเชื่อมโยงโพรบิต เป็นต้น

3. เมื่อเริ่มต้นตัวแปรอิสระมีการแจกแจงแบบยูนิฟอร์มแต่เพื่อให้ตัวแปรอิสระมีความสัมพันธ์กันตามเมทริกซ์สหสัมพันธ์ที่กำหนดอาจมีข้อจำกัดทำให้ตัวแปรอิสระมีความสัมพันธ์กันในการแจกแจงแบบยูนิฟอร์ม ดังนั้นผู้วิจัยขอเสนอแนะแนวคิดในการวิจัยครั้งต่อไปอาจกำหนดให้ตัวแปรอิสระมีการแจกแจงปกติ (ในงานวิจัยครั้งนี้ไม่ได้ทำวิธีดังกล่าวเนื่องจากเกินขอบเขตของงานวิจัย) ซึ่งจะทำให้ตัวแปรอิสระมีการแจกแจงปกติคงเดิมเมื่อถูกทำให้มีความสัมพันธ์กันตามเมทริกซ์สหสัมพันธ์ที่กำหนด

4. ในการวิจัยครั้งนี้ได้ทำการศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบ 2 ประเภท โดยใช้ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยง ได้ใช้เกณฑ์ในการคัดเลือกตัวแบบคือ พื้นที่ใต้โค้ง ROC เป็นเกณฑ์ในการตัดสินใจ ในการวิจัยครั้งต่อไปอาจใช้เกณฑ์การคัดเลือกตัวแบบอื่นๆ เช่น AIC BIC เป็นต้น

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา . <u>การวิเคราะห์ข้อมูลหลายตัวแปร</u> . พิมพ์ครั้งที่ 3. กรุงเทพฯ : สำนักพิมพ์ธรรมสาร, 2551.

กัลยา วานิชย์บัญชา . <u>การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows</u> . พิมพ์ครั้งที่ 7. กรุงเทพฯ : สำนักพิมพ์ธรรมสาร, 2552.

ปฐมาภรณ์ สานุกูล. <u>อำนาจการทดสอบของตัวสถิติสำหรับพื้นที่ใต้โค้ง ROC บนตัวแบบโพรบิท.</u> วิทยานิพนธ์ปริญญา มหาบัณฑิต, ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2552.

ภาษาอังกฤษ

Herman J. Bierens. <u>The logit model : Estimation,Testing and Interpretation</u>. 2008

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

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

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศิริญญา มาปลูก เกิดวันที่ 6 ตุลาคม พ.ศ. 2529 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ในปีการศึกษา 2551 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต (สต.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2552