

การคัดเลือกตัวแบบไม้ติคกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินาม
โดยใช้ฟังก์ชันคอมพลิเมนทารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง

นางสาวศิวรัตน์ รสจันทร์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2554

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

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 SELECTION OF BINARY LOGISTIC
REGRESSION MODEL USING THE COMPLEMENTARY LOG-LOG FUNCTION
AS A LINK FUNCTION

Miss Siwarat Rodjan

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	การคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบ ถดถอยโลจิสติกทวินาม โดยใช้ฟังก์ชันคอมพลิเมนต์ารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง
โดย	นางสาวศิวรัตน์ รสจันทร์
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. สุพล ศุภวงศ์วัฒนา

คณะพาณิชย์ศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์
ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบริหารธุรกิจ

.....คณบดีคณะพาณิชย์ศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร. ชีระพร วีระถาวร)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.สุพล ศุภวงศ์วัฒนา)

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

.....กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.บุญอ้อม โนมทิ)

ศิริรัตน์ รสจันทร์ : การคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอย
 โลจิสติก ทวินามโดยใช้ฟังก์ชันคอมพลิเมนทารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง
 (THE STRICTLY NON-NESTED MODEL SELECTION OF BINARY LOGISTIC
 REGRESSION MODEL USING THE COMPLEMENTARY LOG-LOG
 FUNCTION AS A LINK FUNCTION.) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ.ดร.สุพล
 คุรงค์วัฒนา, 79 หน้า.

การศึกษานี้มีวัตถุประสงค์เพื่อศึกษาระดับความสัมพันธ์ใน แต่ละคู่ของตัวแปรอิสระจากทั้ง
 สองตัวแบบที่จะส่งผลให้ได้เลือกตัวแบบทางสถิติแบบไม่ติดกลุ่มอย่างสมบูรณ์ ของตัวแบบถดถอยโล
 จิสติกทวินามที่มีฟังก์ชันคอมพลิเมนทารีล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยงที่ดีที่สุด เมื่อขนาดตัวอย่าง
 ของเซตตัวแปรอิสระและจำนวนของตัวแปรอิสระของเซตทั้งสองเท่ากัน และเพื่อศึกษาจำนวนของ
 ตัวแปรอิสระจากทั้งสองตัวแบบที่จะส่งผลให้ได้ตัวแบบทางสถิติแบบไม่ติดกลุ่มอย่างสมบูรณ์ของ
 ตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลิเมนทารี ล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยงที่ดีที่สุด
 เมื่อขนาดตัวอย่างของเซตตัวแปรอิสระและระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระของเซตทั้ง
 สองเท่ากัน โดยใช้พื้นที่ใต้โค้ง (Area Under the Curve: AUC) สูงสุดเป็นเกณฑ์ในการเลือกตัวแบบ

จากการวิจัยพบว่า กรณีขนาดตัวอย่างของเซตตัวแปรอิสระและจำนวนของตัวแปรอิสระของ
 เซตทั้งสองเท่ากันพบว่าเมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีค่าเพิ่มขึ้น ค่าเฉลี่ยพื้นที่ใต้
 โค้ง ROC จะมีค่าลดลง กรณีขนาดตัวอย่างของเซตตัวแปรอิสระและระดับความสัมพันธ์ในแต่ละคู่
 ของตัวแปรอิสระของเซตทั้งสองเท่ากันพบว่าเมื่อจำนวนตัวแปรอิสระในตัวแบบมีจำนวนมากขึ้น
 ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC จะมีค่าเพิ่มสูงขึ้นตาม และจากผลการวิจัยทำให้ทราบอีกว่า ในกรณีจำนวน
 ของตัวแปรอิสระของเซตทั้งสองเท่ากันและระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระของเซตทั้ง
 สองเท่ากัน พบว่าเมื่อขนาดตัวอย่างของเซตตัวแปรอิสระเพิ่มมากขึ้น ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC จะมีค่า
 เพิ่มสูงขึ้นตาม ดังนั้น ระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ จำนวนตัวแปรอิสระในตัวแบบ
 และขนาดตัวอย่างเป็นปัจจัยที่ส่งผลต่อการคัดเลือกตัวแบบ ผลการคัดเลือกตัวแบบ กรณีเมื่อจำนวน
 ตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่า จำนวนตัวแปรอิสระในตัวแบบที่ 2 และระดับสหสัมพันธ์ในแต่
 ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 มากกว่าหรือเท่ากับระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ
 ในตัวแบบที่ 2 จะเลือกตัวแบบที่ 2 กรณีเมื่อจำนวนตัวแปรอิสระในตัวแบบที่ 1 มากกว่า จำนวน
 ตัวแปรอิสระในตัวแบบที่ 2 และระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่า
 หรือเท่ากับระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะเลือกตัวแบบที่ 1

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

5281911626 : MAJOR STATISTICS

KEYWORDS : BINARY LOGISTIC REGRESSION / STRICTLY NON-NESTED / THE COMPLEMENTARY LOG-LOG FUNCTION / ROC Curve / AREA UNDER THE CURVE

SIWARAT RODJAN : THE STRICTLY NON-NESTED MODEL SELECTION OF BINARY LOGISTIC REGRESSION MODEL USING THE COMPLEMENTARY LOG-LOG FUNCTION AS A LINK FUNCTION. ADVISOR : ASSOC. PROF. SUPOL DURONGKAWATANA, Ph.D., 79 pp.

The purpose of this study is to find out how different degree of pair-wise correlations between each pair of independent variables from both models affecting the optimal model selection with binary logistic regression model using the complementary log-log function as a link function when the sample size of both sets of independent variables are the same, the number of independent variables for both sets are the same. In addition to find out how different number of independent variables affecting the optimal model selection with binary logistic regression model using the complementary log-log function as a link function when the sample size of both sets of independent variable are the same, and both sets have the same degree of pair-wise correlations. While both sets of independent variables of two purposes are strictly non-nested and the selection criterion is the maximum of the AUC.

The conclusions of this study are: when the sample size of both sets of independent variables are the same, the number of independent variables for both sets are the same. At the degree of pair-wise correlations increases the average maximum of the AUC decrease. When the sample size of both sets of independent variable are the same, and both sets have the same degree of pair-wise correlations. At the number of independent variables for both sets increases the average maximum of the AUC increases. And when the sample size of both sets of independent variables are the same, and both sets have the same degree of pair-wise correlations. At the sample size the average maximum of the AUC increases. Therefore, the degree of pair-wise correlations, the number of independent variables for both sets and the sample size affect the selection of the strictly non-nested binary logistic regression model using the complementary log-log function.

Department : .. Statistics Student's Signature

Field of Study : .. Statistics Advisor's Signature

Academic Year :2011.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือเป็นอย่างดีจาก รองศาสตราจารย์ ดร.สุพล คุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำปรึกษาตลอดจนช่วยเหลือตรวจสอบ แก้ไขข้อบกพร่องต่างๆ เป็นอย่างดีจนกระทั่ง วิทยานิพนธ์เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณด้วยความรู้สึกซาบซึ้ง เคารพและสำนึกในพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ธีระพร วีระถาวร ประธานกรรมการสอบวิทยานิพนธ์ และ อาจารย์ ดร.อนุภาพ สมบูรณ์สวัสดิ์ กรรมการสอบวิทยานิพนธ์ที่ให้คำแนะนำในการทำวิจัยครั้งนี้ และกรุณาตรวจสอบแก้ไขวิทยานิพนธ์ฉบับนี้ให้มีความสมบูรณ์ยิ่งขึ้น

ผู้วิจัยขอกราบขอบพระคุณอาจารย์ผู้ช่วยศาสตราจารย์ ดร.บุญอ้อม โฉมที่ ที่ท่านได้เสียสละเวลาอันมีค่ามาเป็นกรรมการภายนอกมหาวิทยาลัย ซึ่งทำให้วิทยานิพนธ์ฉบับนี้ ให้มีความสมบูรณ์มากยิ่งขึ้น และขอกราบขอบพระคุณคณาจารย์ประจำวิชาภาควิชาสถิติที่ให้โอกาสทางการศึกษาและประสิทธิประสาทความรู้ให้แก่ผู้วิจัยกระทั่งสำเร็จการศึกษา

สุดท้ายนี้ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา มารดา ที่ให้การส่งเสริม สนับสนุนด้านทุนการศึกษา และให้ความรัก ให้กำลังใจแก่ผู้วิจัยมาโดยตลอดจนสำเร็จการศึกษา ตลอดจนเพื่อนๆ พี่ๆ น้องๆ ทุกคนที่ให้คำปรึกษาและคอยช่วยเหลือด้วยดีมาตลอด

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ซ
สารบัญภาพ.....	ญ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ข้อตกลงเบื้องต้น.....	3
1.4 ขอบเขตการวิจัย.....	4
1.5 เกณฑ์ที่ใช้ในการตัดสินใจ.....	5
1.6 คำจำกัดความที่ใช้ในการวิจัย.....	5
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	6
2. ทฤษฎีที่เกี่ยวข้องกับงานวิจัย.....	7
2.1 แบบจำลองที่ใช้ในการวิจัย.....	7
2.2 ตัวแบบติดกลุ่มและไม่ติดกลุ่ม (Nested and Non-nested Models).....	12
2.3 เส้นโค้ง ROC (Receiver Operating Characteristic Curve).....	13
3. วิธีดำเนินการวิจัย.....	18
4. ผลการวิเคราะห์ข้อมูล.....	23
5. สรุปผลการศึกษาและข้อเสนอแนะ.....	74
รายการอ้างอิง.....	78
ประวัติผู้เขียนวิทยานิพนธ์.....	79

สารบัญตาราง

ตารางที่	หน้า
2.1 ตารางการจัดจำแนกการถดถอยโลจิสติกแบบทวินาม.....	15
4.1 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	24
4.2 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	28
4.3 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	31
4.4 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	34
4.5 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	37
4.6 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	40
4.7 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	43
4.8 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	46

ตารางที่	หน้า
4.9 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	49
4.10 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	52
4.11 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	55
4.12 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	58
4.13 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	61
4.14 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	64
4.15 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	67
4.16 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{ r_{jj} \}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	70

สารบัญญภาพ

ภาพที่	หน้า
2.1 กราฟของ π_i สำหรับตัวแบบคอมพลิเมนต์ทารีลือก-ลือก.....	8
2.2 กราฟของ π_i สำหรับตัวแบบลือก-ลือก.....	10
2.3 ตัวอย่างเส้น โคงงและพื้นที่ใต้โค้ง.....	14
2.4 ฟังก์ชัน $f(x)$ (เส้น โคงง) เป็นการประมาณโดยฟังก์ชันเชิงเส้น(เส้นตรง) ระหว่างค่าจริง a และ b.....	17
3.1 ผังงานขั้นตอนการวิจัย.....	22
4.1 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	26
4.2 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,2)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250.....	27
4.3 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	29
4.4 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,3)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250.....	30
4.5 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	32
4.6 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,4)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250.....	33
4.7 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	35

ภาพที่	หน้า
4.8 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2,5)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	36
4.9 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	38
4.10 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,2)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	39
4.11 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	41
4.12 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,3)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	42
4.13 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	44
4.14 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,4)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	45
4.15 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	47
4.16 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,5)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	48

ภาพที่	หน้า
4.17 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,2)$, $n = 50, 100, 150, 200, 250$ โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	50
4.18 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,2)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	51
4.19 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	53
4.20 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,3)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	54
4.21 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	56
4.22 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,4)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	57
4.23 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	59
4.24 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,5)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	60
4.25 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	62

ภาพที่	หน้า
4.26 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,2)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	63
4.27 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	65
4.28 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,3)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	66
4.29 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	68
4.30 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,4)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	69
4.31 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง.....	72
4.32 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5,5)$, $Max\{r_{jj}\}$ อยู่ในระดับต่ำ ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50,100,150,200 และ 250.....	73

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในงานวิจัยขององค์กรหรือแขนงสาขาวิชาต่างๆ วิธีการทางสถิติเข้ามามีบทบาทมากขึ้นเรื่อยๆ จนกระทั่งกลายเป็นวิธีการที่จำเป็นที่จะช่วยให้การวิเคราะห์ข้อมูลที่ได้มานั้นเป็นไปอย่างมีประสิทธิภาพ โดยจุดประสงค์ของงานวิจัยโดยส่วนใหญ่คือ เพื่อต้องการทราบถึงปัจจัย(ตัวแปรอิสระ)ใดบ้างที่ส่งผลโอกาสที่จะเกิดเหตุการณ์(ตัวแปรตาม) ความสัมพันธ์ระหว่างตัวแปรอิสระที่มีผลต่อโอกาสที่จะเกิดเหตุการณ์(ตัวแปรตาม) ตลอดจนเพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจซึ่งจะได้จากสมการที่เหมาะสม ปัจจุบันพบว่าลักษณะของข้อมูลและจุดประสงค์ของงานวิจัยมีหลากหลายมากขึ้น ทำให้เทคนิคการพยากรณ์มีเงื่อนไขที่หลากหลายมากขึ้นตาม ดังนั้นเพื่อให้การพยากรณ์ที่มีประสิทธิภาพการเลือกใช้เทคนิคการพยากรณ์จึงขึ้นอยู่กับลักษณะข้อมูลที่นำมาศึกษาและจุดประสงค์ของการพยากรณ์ เช่น การพยากรณ์ด้วยวิธีการวิเคราะห์ความถดถอยโลจิสติก (Logistic Regression Analysis) เป็นการหาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระโดยตัวแปรตามเป็นตัวแปรเชิงกลุ่มที่แบ่งเป็น 2 กลุ่ม(Binary Logistic Regression Model) หรือมากกว่า 2 กลุ่ม(Multinomial Logistic Regression Model) โดยทั่วไปแล้วมักเป็นการศึกษาข้อมูลเกี่ยวกับเหตุการณ์ที่เราสนใจเราจะเกิดขึ้นหรือไม่ หรือเกิดขึ้นด้วยความน่าจะเป็นเท่าไร เช่น การเป็นโรคหัวใจหรือไม่เป็นโรคหัวใจ มักเป็นเหตุการณ์ที่สาเหตุเกิดขึ้นได้ตลอดเวลาและตัวแปรตามเป็นตัวแปรคุณภาพที่มีการแจกแจงแบบเบอร์นูลลี แบ่งข้อมูลออกเป็น 2 กลุ่ม คือ มีค่าเท่ากับ 1 เมื่อเป็นเหตุการณ์ที่เราสนใจ และมีค่าเท่ากับ 0 เมื่อเป็นเหตุการณ์ที่เราไม่สนใจ ส่วนตัวแปรอิสระมีทั้งตัวแปรต่อเนื่องและตัวแปรที่ไม่ต่อเนื่องก็ได้ และสามารถเป็นได้ทั้งตัวแปรเชิงปริมาณหรือเป็นตัวแปรเชิงคุณภาพได้อีกด้วย ดังนั้นจึงไม่สามารถใช้การวิเคราะห์ความถดถอยแบบปกติได้ เนื่องจากการวิเคราะห์ความถดถอยเป็นการวิเคราะห์ข้อมูลเพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรตามและ ตัวแปรอิสระ โดยตัวแปรอิสระต้องเป็นตัวแปรเชิงปริมาณ หรือ ตัวแปรเชิงกลุ่มก็ได้ แต่ตัวแปรตามต้องเป็นตัวแปรเชิงปริมาณเพียงอย่างเดียว ถ้าทำการวิเคราะห์ด้วยความถดถอยแบบปกติ

ก็จะทำให้ผลการวิเคราะห์ไม่น่าเชื่อถือ ดังนั้นจากสถานการณ์ดังกล่าวโดยส่วนใหญ่จะใช้เทคนิคการวิเคราะห์ข้อมูลด้วยตัวแบบอื่น เช่น ตัวแบบโลจิท(Logit Model), ตัวแบบโพรบิท(Probit Model) และตัวแบบคอมพลิเมนต์ารีล็อก-ล็อก(Complementary log-log Model) เป็นต้น

ตัวแบบที่นำมาใช้แสดงสหสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระในการวิเคราะห์ของ ตัวแบบแบบโลจิท, ตัวแบบโพรบิทและตัวแบบคอมพลิเมนต์ารี ล็อก-ล็อก ซึ่งเป็นตัวแบบเชิงเส้นที่วางนัยทั่วไป(Generalized Linear Model:GLM) เรียกว่า ลิงค์ฟังก์ชัน(Link Function) โดยมี g เป็นตัวเชื่อมระหว่างตัวแปรตาม Y_i และตัวแปรอิสระ X_i ฉะนั้นการเลือกลิงค์ฟังก์ชันสามารถหาได้โดยใช้ฟังก์ชันเชื่อม โลจิทฟังก์ชัน(Logit Function) $g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right)$, โพรบิทฟังก์ชัน(Probit Function) $g(\pi) = \Phi^{-1}(\pi)$, คอมพลิเมนต์ารีล็อก-ล็อก(Complementary log-log Function) $g(\pi) = \ln[-\ln\{1 - \pi\}]$ ตามลำดับ (กุลพัชร หมั่นมา,2551:1:2) ส่วนการเลือกใช้ฟังก์ชันเชื่อมโยงขึ้นอยู่กับลักษณะของการกระจายของข้อมูล เช่น ตัวแบบคอมพลิเมนต์ารี ล็อก-ล็อก(Complementary log-log Model) เป็นฟังก์ชันเชื่อมโยง(Link Function)ที่มีลักษณะข้อมูลคือตัวแปรตามแยกเป็น 2 กลุ่ม ส่วนตัวแปรอิสระเป็นตัวแปรต่อเนื่องหรือเชิงกลุ่มก็ได้ เป็นตัวแบบที่มีเทอมหลักคล้ายกับตัวแบบโลจิท(Logit Model) เหมาะสมกับข้อมูลที่มีลักษณะฟังก์ชันความหนาแน่นสะสมที่มีการรู้เข้าและรู้ออกไม่สม่ำเสมอและไม่สมมาตรที่ความน่าจะเป็น 0.5 (สุทธิเดช ตันฑ์พาณิชย์,2543:88)

การคัดเลือกตัวแบบการถดถอยโลจิสติกแบบทวินาม(Binary logistic regression model) คือ วิธีการทางสถิติที่เกี่ยวข้องกับการเลือกเซตของตัวแปรอิสระที่ดีที่สุดซึ่งใช้เป็นตัวพยากรณ์(Predictor) แต่ละตัวแปรตามให้อยู่ในกลุ่มเหตุการณ์ที่เราสนใจหรือกลุ่มเหตุการณ์ที่เราไม่สนใจ โดยใช้พื้นที่ใต้โค้ง ROC (Receiver operating characteristic) สูงสุดเป็นเกณฑ์หรือเรียกอีกอย่างว่า “พื้นที่ใต้โค้ง”(Area Under the Curve: AUC) เพื่อให้การคัดเลือกตัวแบบที่เหมาะสมที่สุดในการพยากรณ์ เป็นการเปรียบเทียบเซตของตัวแปรอิสระสองกลุ่มที่จะถูกเลือกเป็นเซตของตัวพยากรณ์ที่ดีที่สุด ควรอยู่ภายใต้เงื่อนไขตัวแบบแบบไม่ติดกลุ่ม(Non-nested) (บุญจิรา มากอ้น, 2545:1) และ เป็นตัวแบบแบบไม่ติดกลุ่มอย่างสมบูรณ์(Strictly Non-nested) นั่นคือ ตัวแบบทางสถิติแรกจะต้องไม่มีตัวแปรอิสระใดๆ ของตัวแบบที่สองอยู่เลย

ดังนั้น ผู้วิจัยจึงสนใจที่จะศึกษาว่าตัวแบบทางสถิติแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลิเมนทารี ล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยงมีลักษณะของตัวแบบเป็นแบบใดที่จะทำให้ ค่าพื้นที่ใต้เส้นโค้ง ROC สูงสุด โดยพิจารณาลักษณะของตัวแบบจากขนาดตัวอย่าง(n), จำนวนตัวแปรอิสระ(p_1, p_2) และระดับความสัมพันธ์ในแต่ละคู่ (Degree of pair-wise correlations) ของตัวแปรอิสระ ($\rho_{jj'}$) ที่จะส่งผลให้การเลือกเซตของตัวแปรอิสระที่ดีที่สุด ซึ่งใช้เป็นตัวพยากรณ์ (Predictor) แต่ละตัวแปรตามให้อยู่ในกลุ่มเหตุการณ์ที่เราสนใจหรือกลุ่มเหตุการณ์ที่เราไม่สนใจได้อย่างมีประสิทธิภาพ

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาระดับสหสัมพันธ์ในแต่ละคู่ (Degree of pair-wise correlations) ของตัวแปรอิสระจากทั้งสองตัวแบบที่จะส่งผลให้ได้เลือกตัวแบบทางสถิติแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลิเมนทารีล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยงที่ดีที่สุดเมื่อขนาดตัวอย่างของเซตตัวแปรอิสระ (n) และจำนวนของตัวแปรอิสระ (p_1, p_2) ของเซตทั้งสองเท่ากัน โดยใช้พื้นที่ใต้โค้ง (Area Under the Curve: AUC) สูงสุดเป็นเกณฑ์ในการเลือกตัวแบบ

2. เพื่อศึกษาจำนวนของตัวแปรอิสระ (p_1, p_2) จากทั้งสองตัวแบบที่จะส่งผลให้ได้ตัวแบบทางสถิติแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลิเมนทารี ล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยงที่ดีที่สุดเมื่อขนาดตัวอย่างของเซตตัวแปรอิสระ (n) และระดับสหสัมพันธ์ในแต่ละคู่ (Degree of pair-wise correlations) ของตัวแปรอิสระของเซตทั้งสองเท่ากัน โดยใช้พื้นที่ใต้โค้ง (Area Under the Curve: AUC) สูงสุดเป็นเกณฑ์ในการเลือก

1.3 ข้อตกลงเบื้องต้น

สำหรับการวิจัยในครั้งนี้ศึกษามาจาก ตัวอย่างที่จำลองขึ้นมาจากตัวแบบและการแจกแจงดังนี้

1. ศึกษาตัวแบบคอมเมนทารีล็อก-ล็อก โดยมีรูปแบบคือ

$$\begin{aligned}\hat{\pi}_i &= 1 - \exp[-\exp(\tilde{X}_i^T \tilde{\beta})] & ; i = 1, 2, \dots, n \\ &= 1 - \exp[-\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni})]\end{aligned}$$

โดยที่ π_i คือ ค่าความน่าจะเป็นเมื่อเกิดเหตุการณ์ที่สนใจในหน่วยที่ i

Y_i คือ ตัวแปรตามเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า คือ 0 และ 1

ε_i คือ ตัวแปรสุ่มปกติมาตรฐาน โดย $\varepsilon_i \sim N(0, 1)$

β คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยจำนวน $p + 1$ ตัว

X_i^T คือ เวกเตอร์ของตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณจำนวน p ตัว

2. ตัวแปรอิสระเบื้องต้นมีการแจกแจงสม่ำเสมอ (Uniform Distribution) ที่มีฟังก์ชันความน่าจะเป็นของ x อยู่ในรูป

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

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

ค่าความแปรปรวน

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

1.4 ขอบเขตการวิจัย

1. ศึกษาตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลิเมนต์ารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง เพื่อหาตัวแบบที่ดีที่สุดโดยใช้พื้นที่ใต้โค้ง (Area Under the Curve: AUC) สูงสุดเป็นเกณฑ์ในการเลือก โดยการเปรียบเทียบที่ละ 2 ตัวแบบ

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

3. ค่าพารามิเตอร์เริ่มต้นของสมการถดถอยเป็นค่าใดๆ ในการวิจัยในครั้งนี้ คือ

$$\beta_i = 0.1 \quad ; i = 0, 1, 2, \dots, p \quad \text{และ } \varepsilon_i \sim \text{gumbel}(0, 1); i = 1, 2, \dots, n$$

4. จำนวนตัวแปรอิสระของทั้งสองตัวแบบ (p_1, p_2) ในการวิจัยในครั้งนี้ คือ

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

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

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

$$(p_1, p_2) = (2, 2)$$

5. ขนาดตัวอย่าง (n) ในการวิจัยในครั้งนี้ คือ $n = 50, 100, 150, 200, 250$

6. คัดเลือกคู่ของตัวแปรอิสระเข้าในตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลิเมนต์ารี ล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยงโดยใช้ค่า $Max\{r_{jj'}\}$ เป็นเกณฑ์ในการคัดเลือก นั่นคือ

$$0 < Max\{r_{jj'}\} \leq 0.99$$

โดยที่ $r_{jj'}$ คือ สหสัมพันธ์ระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$

สำหรับเกณฑ์ที่ใช้แบ่งระดับสหสัมพันธ์ระหว่างตัวแปรอิสระ คือ

$$- \text{ ถ้า } 0 < Max\{r_{ij}\} \leq 0.33$$

แล้วตัวแปรอิสระมีสหสัมพันธ์กันในระดับอย่างต่ำ

- ถ้า $0.33 < \text{Max}\{r_{ij}\} \leq 0.66$

แล้วตัวแปรอิสระมีสหสัมพันธ์กันในระดับปานกลาง

- ถ้า $0.66 < \text{Max}\{r_{ij}\} \leq 0.99$

แล้วตัวแปรอิสระมีสหสัมพันธ์กันในระดับสูง

7. กำหนดระดับนัยสำคัญ(α)ในครั้งนี ที่ระดับ 0.05
8. กำหนดสัดส่วนสัดส่วนความสำเร็จและล้มเหลวในศึกษา เท่ากับ 0.5
9. ตัวแปรอิสระเริ่มต้นมีการแจกแจงแบบยูนิฟอร์ม(Uniform)
10. ทำการจำลองข้อมูลโดยใช้เทคนิค Monte Carlo Simulation โดยในการจำลองในแต่ละสถานการณ์ทำซ้ำ 500 รอบ

1.5 เกณฑ์ที่ใช้ในการตัดสินใจ

คัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลี-เมนทารี ล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยงที่ดีที่สุด โดยใช้พื้นที่ใต้โค้ง ROC (Receiver Operating Characteristic curve) สูงสุดเป็นเกณฑ์ นั่นคือพิจารณาพื้นที่ใต้โค้งที่มีค่ามากที่สุดจะหมายความว่า ตัวแบบนั้นมีอัตราความถูกต้องมาก

- ถ้า $0.5 \leq AUC < 0.6$ แล้วตัวแบบเป็นตัวแบบที่เชื่อถือไม่ได้
- ถ้า $0.6 \leq AUC < 0.7$ แล้วตัวแบบเป็นตัวแบบที่เชื่อถือได้น้อย
- ถ้า $0.7 \leq AUC < 0.8$ แล้วตัวแบบเป็นตัวแบบที่สามารถยอมรับได้
- ถ้า $0.8 \leq AUC < 0.9$ แล้วตัวแบบเป็นตัวแบบที่เชื่อถือได้ในระดับดี
- และถ้า $AUC \geq 0.9$ แล้วตัวแบบเป็นตัวแบบที่เชื่อถือได้ในระดับดีมาก

1.6 คำจำกัดความที่ใช้ในการวิจัย

1. พารามิเตอร์(Parameter) คือ ค่าคงที่ที่แสดงคุณลักษณะบางประการของประชากร
2. ค่าเฉลี่ย(Mean) คือ ค่าที่แสดงว่าข้อมูลส่วนใหญ่อยู่ที่ไหน
3. ความแปรปรวน(Variance) คือ ค่าที่บอกว่าค่าของตัวแปรสุ่ม จะมีความผิดพลาดกำลังสอง โดยเฉลี่ยจากค่าของข้อมูลส่วนใหญ่ เท่ากับเท่าไร กล่าวคือ เป็นค่าที่วัดแนวโน้มที่ตัวแปรสุ่มจะห่างจากค่าเฉลี่ยของมัน
4. ตัวแบบคอมพลีเมนทารี ล็อก-ล็อก(Complementary log-log Model)เป็นฟังก์ชันเชื่อมโยง(Link function)ที่มีลักษณะข้อมูล คือ ตัวแบบที่ศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ เมื่อตัวแปรตามมีค่าอยู่ระหว่าง 0 ถึง 1 เหมาะสมกับข้อมูลที่มีลักษณะฟังก์ชัน

ความหนาแน่นสะสมที่มีการลู่เข้า และลู่ออกไม่สม่ำเสมอและไม่สมมาตรที่ความน่าจะเป็น 0.5 สำหรับค่าที่เป็นบวกของ X และ β การแปลงคอมพลิเมนต์ลอจ-ลอจ(Complementary log-log Transformation) ทำให้แน่ใจว่าค่า π จะอยู่ในช่วง $[0,1]$, $\pi \rightarrow 0$ มีแนวโน้มเข้าสู่ $-\infty$ และ $\pi \rightarrow 1$ มีแนวโน้มเข้าสู่ ∞

5. จุดแบ่ง(Cut-off Point) คือ ค่าความน่าจะเป็นที่ใช้ในการพิจารณาการจำแนกกลุ่มของข้อมูลว่าแต่ละหน่วยจะอยู่ในกลุ่มใดระหว่างกลุ่มเหตุการณ์ที่เราสนใจ(Success) หรือกลุ่มเหตุการณ์ที่เราไม่สนใจ(Failure)

6. ข้อมูลดิบ(Ungrouped Data) คือ ข้อมูลที่เพิ่งได้จากการเก็บรวบรวมข้อมูล ยังไม่ได้จัดแบ่งกลุ่ม จาแนกกลุ่ม แยกประเภทและไม่อยู่ในรูปตารางความถี่

7. ระดับสหสัมพันธ์ในแต่ละคู่(Degree of pair-wise correlations) คือ สถานการณ์ที่ตัวแปรอิสระมีความสัมพันธ์กัน

8. การแจกแจงแบบเบอร์นูลลี(Bernoulli Distribution) คือ

ตัวแปรสุ่มเบอร์นูลลี Y กล่าวคือ

ถ้า $Y_i = 0$ เมื่อ การทดลองไม่เกิดเหตุการณ์ที่สนใจ

และ ถ้า $Y_i = 1$ เมื่อ การทดลองเกิดเหตุการณ์ที่สนใจ

โดย $Pr(Y_i = 1 | \tilde{X}_i^T = \tilde{x}_i^T) = \pi_i$; $0 < \pi_i < 1$

และ $Pr(Y_i = 0 | \tilde{X}_i^T = \tilde{x}_i^T) = 1 - \pi_i$; $i = 1, 2, \dots, n$

เขียนแทนด้วย $Y_i \sim Ber(\pi_i)$

ฟังก์ชันความน่าจะเป็นอยู่ในรูปของ

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

9. การประมาณค่าแบบช่วงหรือช่วงความเชื่อมั่น(Confidence Intervals) คือ การประมาณค่าพารามิเตอร์ของประชากรในรูปแบบช่วงโดยใช้ข้อมูลตัวอย่าง การประมาณแบบช่วงนั้นจะบอกถึงค่าต่ำสุดและค่าสูงสุดของพารามิเตอร์ที่เป็นไปได้

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลิเมนต์ลอจ-ลอจเป็นฟังก์ชันเชื่อมโยงที่ดีที่สุด สำหรับไปใช้ในการพยากรณ์ ซึ่งจะสามารถประเมินจากตัวแปรอิสระที่ถูกคัดเลือก ขนาดตัวอย่าง และสหสัมพันธ์ของแต่ละคู่ระหว่างตัวแปรอิสระ

2. เพื่อใช้เป็นแนวทางในการศึกษาวิจัยต่อไป

บทที่ 2

ทฤษฎีที่เกี่ยวข้องกับงานวิจัย

ในบทนี้จะกล่าวถึงแนวคิดและทฤษฎีที่เกี่ยวข้องกับงานวิจัย ซึ่งประกอบด้วย 3 ส่วน คือ
2.1 แบบจำลองที่ใช้ในการวิจัย 2.2 ตัวแบบติดกลุ่มและไม่ติดกลุ่ม(Nested and Non-nested Models)
2.3 เส้นโค้ง ROC (Receiver Operating Characteristic Curve)

2.1 แบบจำลองที่ใช้ในการวิจัย

ตัวแบบการถดถอยโลจิสติกทวินาม(Binary Logistic Regression Model)

ตัวแบบการถดถอยโลจิสติกทวินามสำหรับการพยากรณ์ในการจัดประเภท ดังนี้

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

$$\tilde{X}_i^T = (1 \ X_{i1} \ X_{i2} \ \dots \ X_{ip}), \quad \tilde{\beta}^T = (\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_p)$$

เมื่อ

$$Pr(Y_i = 1 | \tilde{X}_i^T = \tilde{x}_i^T) = \pi_i,$$

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

เมื่อ Y_i มีการแจกแจงแบบเบอร์นูลลีที่เป็นอิสระกัน มีค่าเฉลี่ยและความแปรปรวน ดังนี้

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

ดังนั้น ตัวแบบการถดถอยคือ

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

$$\text{เมื่อ} \quad E(\varepsilon_i | \tilde{X}_i^T = \tilde{x}_i^T) = 0 \quad ; i = 1, 2, \dots, n$$

ตัวแบบคอมพลีเมนต์ลอจิสติก-ลอจิสติก(Complementary log-log Model)

ตัวแบบคอมพลีเมนต์ลอจิสติก-ลอจิสติก(Complementary log-log model) เป็นส่วนที่ขยายจากตัวแบบโลจิสติก(Logit model) และตัวแบบโพรบิท(Probit model) เมื่อค่า π_i เพิ่มขึ้นจาก 0 ก่อนข้างซ้าย แต่ค่าของ π_i จะมีค่าเข้าใกล้ 1 อย่างรวดเร็ว

สำหรับ ฟังก์ชันเชื่อมโยง(Link Function) ที่ใช้ในตัวแบบโลจิทและโพรบิทนั้นจะมีคุณสมบัติสมมาตร(Symmetric) รอบค่า 0.5 หรือ

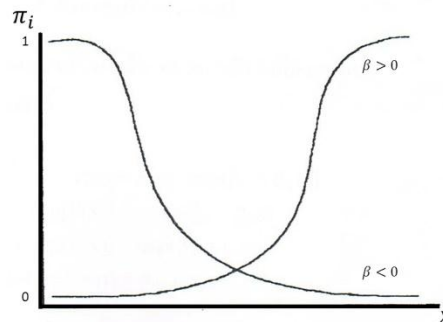
$$\text{link}[\pi_i] = -\text{link}[1 - \pi_i]$$

นั่นคือจากฟังก์ชันโลจิท

$$\begin{aligned} \text{logit}[E(Y_i|\tilde{X}_i^T = \tilde{x}_i^T)] &= \ln\left(\frac{E(Y_i|\tilde{X}_i^T = \tilde{x}_i^T)}{1 - E(Y_i|\tilde{X}_i^T = \tilde{x}_i^T)}\right) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) \\ &= \ln \pi_i - \ln[1 - \pi_i] \\ &= -\ln\left(\frac{1 - \pi_i}{\pi_i}\right) \\ &= -\text{logit}[1 - \pi_i] \end{aligned}$$

จาก $\text{link}[\pi_i] = -\text{link}[1 - \pi_i]$ หมายถึง โค้งของ π_i สำหรับตัวแบบโลจิทและตัวแบบโพรบิทจะมีรูปแบบสมมาตรรอบจุดที่ $\pi_i = 0.5$ โดยเฉพาะเมื่อ π_i จะมีค่าเข้าใกล้ 0 ด้วยอัตราที่เท่ากับ เมื่อ π_i จะมีค่าเข้าใกล้ 1 อย่างรวดเร็ว

แต่ถ้า π_i เพิ่มขึ้นจาก 0 ค่อนข้างช้า แต่มีค่าเข้าใกล้ 1 อย่างรวดเร็วดังรูปที่ 2.1 ด้วยตัวแบบโลจิทและตัวแบบโพรบิท จึงไม่เหมาะสมกับข้อมูล จึงควรใช้ตัวแบบอื่นคือตัวแบบคอมพลิเมนต์ารีลือก-ลือก และลักษณะของกราฟของ π_i สำหรับตัวแบบคอมพลิเมนต์ารีลือก-ลือกแสดงดัง ภาพที่ 2.1



ภาพที่ 2.1 กราฟของ π_i สำหรับตัวแบบคอมพลิเมนต์ารีลือก-ลือก

จากลักษณะของกราฟ π_i ข้างต้น ควรใช้เส้นโค้งฟังก์ชัน คือ

$$\pi_i = 1 - \exp[-\exp(\tilde{X}_i^T \tilde{\beta})] \quad (2)$$

$$1 - \pi_i = \exp[-\exp(\tilde{X}_i^T \tilde{\beta})]$$

ซึ่งมีรูปแบบไม่สมมาตรคือ π_i มีค่าลดลงจาก 1 รวดเร็วกว่าการเข้าใกล้ 0 โดยสอดคล้องกับรูป 1.1 ฟังก์ชัน (2) นำไปสู่ตัวแบบคอมพลิเมนต์ล็อก-ล็อกในสมการ (3) คือ

$$-\ln[1 - \pi_i] = \exp(\tilde{X}_i^T \tilde{\beta})$$

$$\ln[-\ln[1 - \pi_i]] = \tilde{X}_i^T \tilde{\beta} \quad (3)$$

การตีความหมายในตัวแบบคอมพลิเมนต์ล็อก-ล็อก สำหรับ x_1 และ x_2 ใดจะพบว่า

$$\ln\{-\ln[1 - \pi_2]\} - \ln\{-\ln[1 - \pi_1]\} = \beta_1(x_2 - x_1)$$

และ

$$\frac{\ln[1 - \pi_2]}{\ln[1 - \pi_1]} = \exp[(x_2 - x_1)]$$

ดังนั้น

$$1 - \pi_2 = \ln[1 - \pi_1]^{\exp[\beta_1(x_2 - x_1)]}$$

ซึ่งหมายความว่า ความน่าจะเป็นที่ไม่สำเร็จ (Probability of failure) ณ x_2 มีค่าเท่ากับ ความน่าจะเป็นที่ไม่สำเร็จ ณ x_1 ยกกำลัง $\exp[\beta_1(x_2 - x_1)]$

นอกจากตัวแบบ (2) และ(3)แล้ว ยังมีตัวแบบที่น่าสนใจและเกี่ยวกับตัวแบบนี้คือตัวแบบอยู่ในรูปแบบ ดังนี้

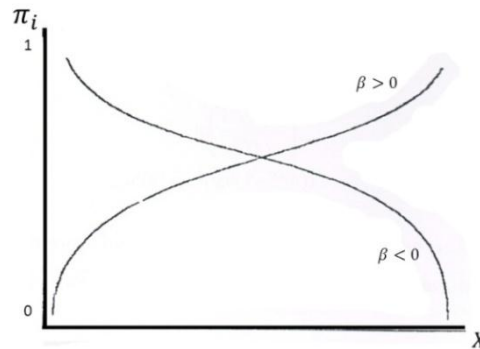
$$\pi_i = \exp[-\exp(\tilde{X}_i^T \tilde{\beta})] ; i = 1, 2, \dots, n$$

$$\ln[\pi_i] = -\exp(\tilde{X}_i^T \tilde{\beta})$$

$$-\ln[\pi_i] = \exp(\tilde{X}_i^T \tilde{\beta})$$

$$\ln[-\ln[\pi_i]] = \tilde{X}_i^T \tilde{\beta} \quad (4)$$

ฟังก์ชัน (4) มีลักษณะของกราฟของ π_i ที่มีค่าของ π_i จะลดลงจาก 1 ก่อนข้างช้า แต่จะเข้าใกล้ 0 อย่างรวดเร็ว ดังเส้นโค้งของภาพที่ 2.2 ดังนี้



ภาพที่ 2.2 กราฟของ π_i สำหรับตัวแบบล็อก-ล็อก

โดย ค่าของ π_i จะลดลงจาก 1 ก่อนข้างช้า แต่เข้าใกล้ 0 อย่างรวดเร็ว

ดังนั้น กรณีที่ใช้ฟังก์ชันเชื่อมโยง แบบล็อก-ล็อก (Log-Log Link) จะนำไปสู่ตัวแบบล็อก-ล็อก(Log-Log Model) ซึ่งมีตัวแบบดัง(4) และ ภาพที่ 1.2

สรุปว่า ถ้าเป็นตัวแบบคอมพลิเมนต์ล็อก-ล็อก จะใช้สำหรับความน่าจะเป็นที่สำเร็จ (Probability of success) ส่วนตัวแบบล็อก-ล็อกจะใช้สำหรับความน่าจะเป็นที่ไม่สำเร็จ (Probability of failure)

หมายเหตุ ตัวแบบล็อก-ล็อก $\pi_i = 1 - \exp[-\exp(\tilde{x}_i^T \tilde{\beta})]$ เป็นกรณีพิเศษของตัวแบบ $\pi_i = F(\tilde{X}_i^T \tilde{\beta})$ ด้วย cdf ของการแจกแจงค่าสุดโต่ง (Extreme value or Gumbel Distribution) เท่ากับ

$$G(x) = \exp \left[-\exp \left\{ -\frac{(x-a)}{b} \right\} \right]$$

ซึ่งมีค่าเฉลี่ยเท่ากับ $a + 0.577b$ และส่วนเบี่ยงเบนมาตรฐานเท่ากับ $\frac{\pi b}{\sqrt{6}}$ สำหรับพารามิเตอร์ $b > 0, -\infty < a < +\infty$

และฟังก์ชันผกผันของ cdf ของการแจกแจงค่าสุดโต่ง (Inverse of the cumulative Extreme value Function)

$$F(x) = 1 - \exp[-\exp(x)]$$

หรือเรียกอีกอย่างหนึ่งว่า **Gompertz Distribution**

ตัวแบบคอมพลิเมนต์ลอจิสติก-ลอจิสติกในรูปความสัมพันธ์เชิงเส้น

$$\ln[-\ln[1 - E(Y_i | \tilde{X}_i^T = \tilde{x}_i^T)]] = \ln[-\ln[1 - \pi_i]] = \tilde{X}_i^T \tilde{\beta} \quad ; i = 1, 2, \dots, n$$

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

$$Y_i^* = \ln[-\ln[1 - E(Y_i | \tilde{X}_i^T = \tilde{x}_i^T)]] + \varepsilon_i \quad ; \varepsilon_i \sim N(0, \sigma^2), i = 1, 2, \dots, n$$

โดย π_i คือ ค่าความน่าจะเป็นเมื่อเกิดเหตุการณ์ที่สนใจในหน่วยที่ i

Y_i คือ ตัวแปรตามเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า คือ 0 และ 1

Y_i^* คือ ตัวแปรแฝง(Latent variable)ที่ไม่สามารถสังเกตได้ จึงไม่ทราบค่าที่แท้จริง
ทราบเพียงแต่ผลที่เกิดขึ้น

\tilde{X}_i คือ เวกเตอร์ตัวแปรอิสระหน่วยที่ i

$\tilde{\beta}$ คือ เวกเตอร์พารามิเตอร์ของตัวแบบ

ε_i คือ ตัวแปรค่าคลาดเคลื่อนที่เป็นอิสระกัน หรือค่าคาดหมายอย่างมีเงื่อนไข

(Conditional expectation) หน่วยที่ i ตัวแปรสุ่มปกติมาตรฐานโดย

$$\varepsilon_i \sim N(0, 1)$$

เมื่อแต่ละค่าสังเกต Y_i คือตัวแปรสุ่มแบบเบอร์นูลลี ซึ่ง

$$Pr(Y_i = 1 | \tilde{X}_i^T = \tilde{x}_i^T) = \pi_i, \quad Pr(Y_i = 0 | \tilde{X}_i^T = \tilde{x}_i^T) = 1 - \pi_i \quad ; i = 1, 2, \dots, n$$

ซึ่งมีการแจกแจงความน่าจะเป็นคือ

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

ฟังก์ชันความน่าจะเป็นรวมคือ

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

ลอการิทึมธรรมชาติของฟังก์ชันความน่าจะเป็นรวมคือ

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

เมื่อ

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

$$1 - \pi_i = \exp[-\exp(\tilde{X}_i^T \tilde{\beta})]$$

$$\ln[-\ln[1 - \pi_i]] = \tilde{X}_i^T \tilde{\beta} \quad ; i = 1, 2, \dots, n$$

ดังนั้น ลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น (Log-likelihood function) คือ

$$\begin{aligned} \ln L(\tilde{\beta}^T) &= \sum_{i=1}^n y_i \ln(1 - \exp(-\exp(\tilde{X}_i^T \tilde{\beta}))) \\ &\quad + (1 - y_i) \ln(1 - \exp(-\exp(\tilde{X}_i^T \tilde{\beta}))) \end{aligned}$$

เมื่อเราทราบการแจกแจงของ Y เราสามารถสร้างฟังก์ชันภาวะน่าจะเป็นได้ เนื่องจากการประมาณค่าไม่ได้เป็นไปตามรูปแบบ เราจึงต้องใช้วิธีการประมาณเชิงตัวเลข(จำเป็นต้องทำซ้ำเพื่อให้ได้ซึ่งตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator: MLE) โดยการทำซ้ำ 1 ครั้งจะได้ตัวประมาณภาวะน่าจะเป็นสูงสุด $(b_0, b_1, b_2, \dots, b_n)$ แล้วทำการคำนวณค่าประมาณของ $\pi_i ; i = 1, 2, \dots, n$ ดังนี้

$$\hat{\pi}_i = 1 - \exp[-\exp(\tilde{X}_i^T \tilde{\beta})] \quad ; i = 1, 2, \dots, n$$

เมื่อประมาณค่าพารามิเตอร์ด้วยวิธีภาวะความน่าจะเป็นสูงสุดในตัวแบบการถดถอยโลจิสติกทวินาม แล้วจะสามารถนำไปใช้ในการพยากรณ์การจำแนกกลุ่มของตัวแบบต่อไป

2.2 ตัวแบบติดกลุ่มและไม่ติดกลุ่ม(Nested and Non-nested Models)

นิยาม 2.1: ตัวแบบติดกลุ่ม(Nested Models) ตัวแบบการถดถอยโลจิสติกทวินาม เป็นตัวแบบติดกลุ่ม ถ้าหนึ่งตัวแบบสามารถลดตัวแบบอื่นๆ ได้โดยกำหนดการเชิงเส้นซึ่งมีข้อจำกัดบนเวกเตอร์พารามิเตอร์ ตัวอย่างเช่น ให้ M_1 และ M_2 เป็นสองตัวแบบที่มีลักษณะตามรูปแบบการทำงานเหมือนกันและโครงสร้างข้อผิดพลาดเดียวกัน โดยเฉพาะข้อมูลเป็นค่าความเบี่ยงเบนจากค่าเฉลี่ยเพื่อให้ไม่มีตัว intercept ตัวแบบเหล่านี้แตกต่างเพียงในแง่ของตัวแปรอิสระ เช่น

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

$$M_2: \ln[-\ln[1 - E(Y_i | \tilde{X}_i^T = \tilde{x}_i^T)]] = \tilde{x}_i^T \tilde{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} \quad ; i = 1, 2, \dots, n$$

ตัวแบบทั้งสองเป็นตัวแบบติดกลุ่มเพราะข้อจำกัดที่ $\beta_3 = 0$ M_2 จะกลายเป็น M_1 พูดังๆ าคือ M_2 "ครอบคลุม" M_1 ความแตกต่างระหว่างตัวแบบเหล่านี้เกี่ยวข้องกับทดสอบข้อจำกัดใน β_3

นิยาม 2.2: ตัวแบบไม่ติดกลุ่ม(Non-nested Models) ตัวแบบการถดถอยโลจิสติกทวินาม เป็นตัวแบบไม่ติดกลุ่มไม่ว่าจะเป็นเพียงบางส่วนหรืออย่างสมบูรณ์ ถ้าหนึ่งตัวแบบไม่สามารถลด ตัวแบบอื่นๆ ได้โดยกำหนดการเชิงเส้นซึ่งมีข้อจำกัดบนเวกเตอร์พารามิเตอร์ เช่น

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

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

ตัวแบบทั้งสองเป็นตัวแบบติดกลุ่มเพราะแม้ว่าเราจะกำหนดข้อจำกัดที่ $\beta_4 = 0, \beta_5 = 0$ M_2 จะไม่กลายเป็น M_1 ตัวแบบดังกล่าวเป็นเพียง"บางส่วน(Partially)" ที่ไม่ติดกลุ่มเพราะมี ตัวแปรอิสระใน x_3 ร่วมกัน ซึ่งถ้า M_1 และ M_2 ไม่มี x_3 ร่วมกัน จะเป็นดังนี้

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

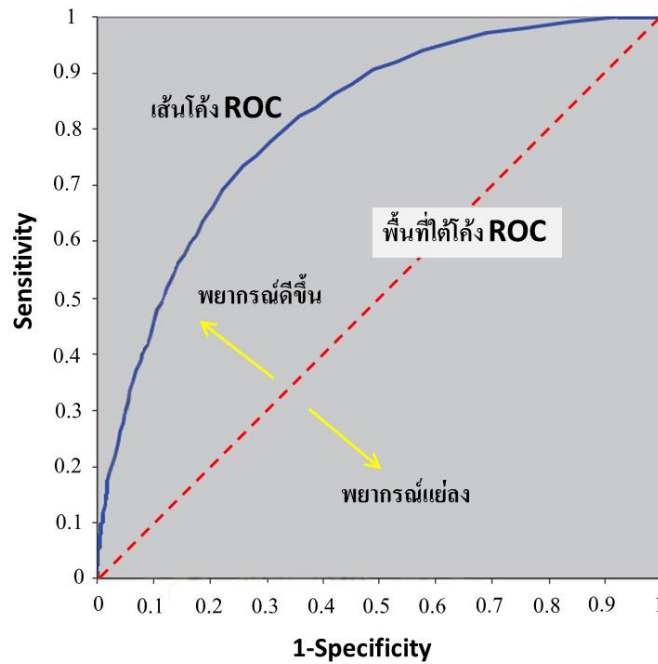
$$M_2: \ln[-\ln[1 - E(Y_i|\tilde{X}_i^T = \tilde{x}_i^T)]] = \tilde{x}_i^T \tilde{\beta} = \beta_0 + \beta_4 x_{4i} + \beta_5 x_{5i}; i = 1, 2, \dots, n$$

ตัวแบบนี้เป็นตัวแบบ “ไม่ติดกลุ่มอย่างสมบูรณ์(Strictly non-nested)”

2.3 เส้นโค้ง ROC (Receiver Operating Characteristic Curve)

Receiver Operating Characteristic Curve หรือเรียกง่าย ๆ ว่า “ROC Curve” ได้ถูกนำมาใช้ ครั้งแรกเมื่อสงครามโลกครั้งที่ 2 สำหรับการวิเคราะห์สัญญาณเรดาร์โดยทหารของสหรัฐอเมริกา และต่อจากนั้นมาก็ได้มีงานวิจัยที่นำ ROC Curve มาใช้อย่างแพร่หลายมากขึ้นทั้งด้านจิตวิทยา เกี่ยวกับมนุษย์ การวินิจฉัยโรคทางการแพทย์ และการประเมินความถูกต้องของตัวแบบของ ทางด้านวิทยาศาสตร์และสังคม

ROC Curve ถูกนำมาใช้ในการประเมินความถูกต้องของการพยากรณ์เหตุการณ์ในระบบ จำแนกกลุ่มกรณีแบ่งเป็น 2 กลุ่ม ได้แก่ กลุ่มที่เกิดเหตุการณ์ที่สนใจและกลุ่มที่ไม่เกิดเหตุการณ์ที่ สนใจ โดยอยู่ในรูปของกราฟที่พล็อตระหว่างค่า Sensitivity หรือ True-Positive Rate ซึ่งเป็น อัตราส่วนของการพยากรณ์ถูกต้องของการเกิดเหตุการณ์ที่สนใจ ($Y = 1$) และค่า 1-Specificity หรือ False-Positive Rate ซึ่งเป็นอัตราส่วนของการพยากรณ์ผิดของการไม่เกิดเหตุการณ์ที่สนใจ ($Y = 0$) โดยจะกำหนดให้ 1-Specificity อยู่บนแกน X และ Sensitivity อยู่บนแกน Y และกราฟอยู่ใน ช่วง $[0,1]$ ดังรูปที่ 2.3



ภาพที่ 2.3 ตัวอย่างเส้นโค้งและพื้นที่ใต้โค้ง

สำหรับการคำนวณค่า Sensitivity และ 1-Specificity จะทำโดยการกำหนดค่าจุดตัด (Cut-off) ที่ระดับต่างๆ ระหว่าง 0 ถึง 1 เพื่อเปรียบเทียบกับค่าความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจที่ได้จากการพยากรณ์ของแต่ละหน่วยตัวอย่างด้วยตัวแบบประมาณที่ได้จากการวิเคราะห์ ($\hat{\pi}_i$) แล้วทำการจำแนกกลุ่มของตัวแปรตามซึ่งเป็นตัวแปรที่ต้องการพยากรณ์ออกเป็น 2 กลุ่มโดยที่

ถ้า $\hat{\pi}_i = 1 - \exp[-\exp(\tilde{X}_i^T \tilde{\beta})] \leq \text{Cut-off} \rightarrow \hat{Y}_i = 0$ ตัวอย่างจะถูกพยากรณ์ให้อยู่ในกลุ่มที่ไม่เกิดเหตุการณ์

ถ้า $\hat{\pi}_i = 1 - \exp[-\exp(\tilde{X}_i^T \tilde{\beta})] > \text{Cut-off} \rightarrow \hat{Y}_i = 1$ ตัวอย่างจะถูกพยากรณ์ให้อยู่ในกลุ่มที่เกิดเหตุการณ์

จากนั้นจึงคำนวณหาสัดส่วนของการพยากรณ์เหตุการณ์ เพื่อนำค่าที่ได้ไปทำการพล็อตโค้ง ROC และคำนวณหาพื้นที่ใต้โค้ง ซึ่งสูตรการคำนวณ Sensitivity และ 1-Specificity เป็นดังนี้

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

ตารางที่ 2.1 ตารางการจัดจำแนกการถดถอยโลจิสติกแบบทวินาม

True positive (TP) คือ จำนวนตัวอย่างที่พยากรณ์ถูกต้องของการเกิดเหตุการณ์ ($\hat{Y} = 1|Y = 1$)

False positive (FP) คือ จำนวนตัวอย่างที่พยากรณ์ผิดของการเกิดเหตุการณ์ ($\hat{Y} = 1|Y = 0$)

True negative (TN) คือ จำนวนตัวอย่างที่พยากรณ์ถูกต้องของการไม่เกิดเหตุการณ์ ($\hat{Y} = 0|Y = 0$)

False negative (FN) คือ จำนวนตัวอย่างที่พยากรณ์ผิดของการไม่เกิดเหตุการณ์ ($\hat{Y} = 0|Y = 1$)

Sensitivity (True-Positive Rate) คือ ค่าความน่าจะเป็นหรืออัตราส่วนของการพยากรณ์เหตุการณ์ได้ถูกต้องของการเกิดเหตุการณ์ที่สนใจ

$$Sensitivity = P(\hat{Y} = 1|Y = 1) = \frac{TP}{TP + FN}$$

Specificity (False-Positive Rate) คือ ค่าความน่าจะเป็นหรืออัตราส่วนของการพยากรณ์เหตุการณ์ได้ถูกต้องของการไม่เกิดเหตุการณ์ที่สนใจ

$$Specificity = P(\hat{Y} = 0|Y = 0) = \frac{TN}{FP + TN}$$

1-Specificity (False-Positive Rate) คือ ค่าความน่าจะเป็นหรืออัตราส่วนของการพยากรณ์เหตุการณ์ผิดของการไม่เกิดเหตุการณ์ที่สนใจ

$$1 - Specificity = P(\hat{Y} = 1|Y = 0) = \frac{FP}{FP + TN}$$

พื้นที่ใต้โค้ง ROC เป็นค่าที่บ่งบอกถึงความสามารถในการพยากรณ์ได้ถูกต้องหรือแสดงถึงความเชื่อถือได้ของตัวแบบ มีพิสัยอยู่ระหว่าง 0 ถึง 1 คือ ค่าประมาณพื้นที่ใต้โค้ง ROC (Area under the Curve หรือ AUC) ซึ่ง Hosmer, David W., and Stanley Lemeshow.(200) ได้กำหนดเกณฑ์ทั่วไปไว้ดังนี้

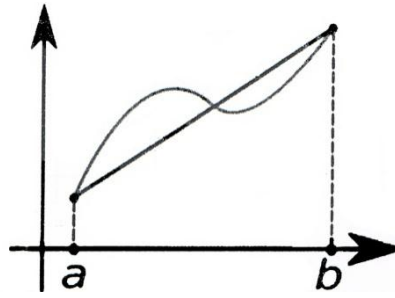
- ถ้า $0.5 \leq AUC < 0.6$ แล้ว ตัวแบบเป็นตัวแบบที่เชื่อถือไม่ได้
- ถ้า $0.6 \leq AUC < 0.7$ แล้ว ตัวแบบเป็นตัวแบบที่เชื่อถือได้น้อย
- ถ้า $0.7 \leq AUC < 0.8$ แล้ว ตัวแบบเป็นตัวแบบที่สามารถยอมรับได้
- ถ้า $0.8 \leq AUC < 0.9$ แล้ว ตัวแบบเป็นตัวแบบที่เชื่อถือได้ในระดับดี
- และถ้า $AUC \geq 0.9$ แล้ว ตัวแบบเป็นตัวแบบที่เชื่อถือได้ในระดับดีมาก

ถ้าพื้นที่ใต้โค้ง ROC เท่ากับ 0.5 แสดงว่าตัวแบบไม่สามารถจำแนกกลุ่มได้(เป็นเสมือนการเลือกอย่างสุ่มสำหรับการจำแนกกลุ่ม 2 กลุ่ม) เส้นโค้งจะมีลักษณะเป็นเส้นทแยงมุมจากจุด (0,0) ถึงจุด (1,1) ดังภาพ 3.1 และถ้าพื้นที่ใต้โค้ง ROC เท่ากับ 1 แสดงว่าตัวแบบสามารถจำแนกกลุ่มได้อย่างถูกต้องสมบูรณ์ ในกรณีที่เส้นโค้ง ROC อยู่ใต้เส้นทแยงมุมหรือมีพื้นที่ใต้ ROC < 0.5 จำเป็นจะต้องมีการปรับเปลี่ยนให้เส้นอยู่บนเส้นทแยงมุมโดยทำการกลับด้านของการพยากรณ์ของทุกหน่วยตัวอย่าง จากการสนใจการพยากรณ์ที่เป็นเกิดเหตุการณ์เปลี่ยนเป็นสนใจการพยากรณ์กลุ่มที่ไม่เกิดเหตุการณ์ นั่นคือเปลี่ยนจาก TP เป็น FN และเปลี่ยนจาก FP เป็น TN แทน(วรุณี มุริกา ,2551;4:7)

การคำนวณค่า AUC สามารถทำได้ 2 วิธี ซึ่งวิธีแรกคือวิธีที่ไม่ใช้พารามิเตอร์จะขึ้นอยู่กับ การสร้างการประมาณรูปสี่เหลี่ยมที่มีสองด้านขนานกันได้เส้น โค้ง ซึ่งเป็นการประมาณพื้นที่ วิธีที่สองคือวิธีการประมาณค่าพารามิเตอร์โดยใช้ตัวประมาณค่าความควรจะเป็นสูงสุด (Maximum Likelihood Estimator) เพื่อให้เส้นโค้งเรียบพอดีกับจุดข้อมูล โดยทั้งสองวิธีใช้ได้กับโปรแกรม คอมพิวเตอร์และสามารถหาค่าประมาณของพื้นที่และความคลาดเคลื่อนมาตรฐานได้ การประมาณค่าโดยใช้รูปสี่เหลี่ยมที่มีสองด้านขนานกัน หรือวิธีการที่เรียกว่า กฎของรูปสี่เหลี่ยมที่มีสองด้านขนานกัน หรือกฎรูปสี่เหลี่ยมคางหมู ซึ่งในการศึกษาครั้งนี้จะใช้วิธีนี้ในการคำนวณ AUC วิธีนี้เป็น เทคนิคการประมาณค่าเกี่ยวกับการคำนวณอินทิกรัลจำกัดเขต ซึ่งแสดงดังนี้

$$\int_a^b f(x)dx ; a \leq x \leq b$$

ฟังก์ชัน $f(x) ; a \leq x \leq b$ ซึ่งจะคำนวณโดยอินทิกรัลจำกัดเขต



ภาพที่ 2.4 ฟังก์ชัน $f(x)$ (เส้นโค้ง) เป็นการประมาณโดยฟังก์ชันเชิงเส้น
(เส้นตรง) ระหว่างค่าจริง a และ b

สำหรับ 2 ตัวแบบประมาณเมื่อทำการทดสอบการพยากรณ์กับข้อมูลชุดเดียวกันแล้ว
ตัวแบบใดมีเส้นโค้ง ROC อยู่ข้างบนหรือมีพื้นที่ใต้โค้ง ROC มากกว่า แสดงว่าเป็นตัวแบบที่มี
ประสิทธิภาพหรือให้ความถูกต้องในการพยากรณ์ได้ดีกว่า

บทที่ 3

วิธีดำเนินการวิจัย

ศึกษาการเลือกตัวแบบทางสถิติแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติก ทวินามที่มีฟังก์ชันคอมพลิเมนต์ทรีล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยง โดยใช้พื้นที่ใต้โค้ง (Area Under the Curve: AUC) สูงสุดเป็นเกณฑ์ในการคัดเลือก ซึ่งมีปัจจัยที่เกี่ยวข้องคือ ระดับสหสัมพันธ์ในแต่ละคู่ (Degree of pair-wise correlations) ของตัวแปรอิสระจากทั้งสองตัวแบบ ขนาดตัวอย่างของเซตตัวแปรอิสระ (n) และจำนวนของตัวแปรอิสระ (p_1, p_2) ของเซตทั้งสอง

วิธีดำเนินการวิจัย

1. ศึกษาค้นหาเอกสารและข้อมูลที่เกี่ยวข้องกับงานวิจัยและกำหนดตัวแปรที่สนใจศึกษา ดังนี้

- จำนวนตัวแปรอิสระของทั้งสองตัวแบบ คือ

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

$$(3,2), (3,3), (3,4), (3,5),$$

$$(4,2), (4,3), (4,4), (4,5),$$

$$(5,2), (5,3), (5,4), (5,5)$$

- ค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' คือ

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

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

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

- ขนาดตัวอย่าง คือ

$$n = 50, 100, 150, 200, 250$$

และทำการวางแผนงานวิจัยทำให้ได้กรณีที่ต้องการทดสอบของจุดประสงค์แรก

240 กรณี และของจุดประสงค์ที่สอง 240 กรณี

2. กำหนดค่าพารามิเตอร์เริ่มต้นของตัวแบบ (β_0, β_i) เท่ากับ 0.1

3. จำลองตัวแปร

ตัวแปรอิสระ

- จำลองตัวแปรอิสระ จำนวน $p_1 + p_2$ จากการแจกแจงแบบปกติค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 1

$$Z_i \sim N(0,1) ; i = 1, 2, \dots, p_1 + p_2$$

- แบ่งตัวแปรอิสระออกเป็นสองเซตเพื่อใช้ในตัวแบบทั้งสอง โดยในแต่ละเซตนำมาสร้างสหสัมพันธ์กัน จะได้

$$\tilde{Z}_1 \sim N_{p_1}(\mathbf{0}_{p_1 \times 1}, \Sigma_1) , \tilde{Z}_2 \sim N_{p_2}(\mathbf{0}_{p_2 \times 1}, \Sigma_2)$$

- แปลงตัวแปรอิสระที่มีการแจกแจงแบบปกติที่มีสหสัมพันธ์กันเป็นการแจกแจงแบบสม่ำเสมอช่วง 0 ถึง 1

$$Z_i \sim N(0,1) \rightarrow U_i \sim U(0,1)$$

- ปรับช่วงของการแจกแจงสม่ำเสมอ

$$U_i \sim (0,1) \rightarrow x_i \sim (-10,10)$$

- ตรวจสอบระดับสหสัมพันธ์ r_{jj} ของแต่ละคู่ของตัวแปรอิสระว่าอยู่ในช่วงของค่าสหสัมพันธ์ที่ใช้ตั้งแต่แรกแล้วหรือไม่

เกณฑ์ที่ใช้แบ่งระดับสหสัมพันธ์ระหว่างตัวแปรอิสระ คือ

ถ้า $0 < \text{Max}\{|r_{jj}|\} \leq 0.33$ แล้วตัวแปรอิสระมีค่าสหสัมพันธ์กันในระดับต่ำ

ถ้า $0.33 < \text{Max}\{|r_{jj}|\} \leq 0.66$ แล้วตัวแปรอิสระมีค่าสหสัมพันธ์กันในระดับปานกลาง

ถ้า $0.66 < \text{Max}\{|r_{jj}|\} \leq 0.99$ แล้วตัวแปรอิสระมีค่าสหสัมพันธ์กันในระดับสูง

ดังนั้นตัวแปรอิสระที่ใช้ในตัวแบบทั้งสองจะมีการแจกแจงแบบสม่ำเสมอในช่วง -10 ถึง 10 และมีระดับสหสัมพันธ์เท่ากับที่ทำการตรวจสอบ ($\text{Max}\{|r_{jj}|\}$) โดยตัวแบบที่ 1 มี p_1 และตัวแบบที่ 2 มี p_2 ตัว

ตัวแบบที่ 1 : x_1, x_2, \dots, x_{p_1} และ

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

ค่าความคลาดเคลื่อน

- จำลองค่าความคลาดเคลื่อนให้มีการแจกแจง

$$\varepsilon_i \sim \text{gumbel}(0,1); i = 1, 2, \dots, n$$

ตัวแปรตาม

- จากตัวแบบการถดถอยโลจิสติกแบบทวินาม

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

- นำตัวแปรอิสระ, ค่าพารามิเตอร์เริ่มต้นและค่าความคลาดเคลื่อนมาแทนในตัวแบบจะได้

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

- ปรับค่า Y_i^* เป็นตัวแปรหุ่น (Dummy Variable)

$$Y_i; Y_i^* > 0 \text{ และ } Y_i; Y_i^* \leq 0$$

4. ทำการประมาณค่าพารามิเตอร์โดยใช้ฟังก์ชันคอมพลีเมนต์รี ล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยง แล้วจะได้ตัวแบบที่ใช้ในการพยากรณ์ แล้วทำการสร้างค่าพยากรณ์ (Y_i^*) ขึ้นมา

5. คำนวณค่า Sensitivity และค่า 1-Specificity

6. พล็อตโค้ง ROC ด้วยค่า Sensitivity และค่า 1-Specificity

7. คำนวณหาค่าประมาณค่า AUC จาก โค้ง ROC

8. ทำซ้ำจนครบจำนวนตามที่กำหนดคือ 500 รอบ

9. หาค่าเฉลี่ยของ AUC ที่ได้ทั้ง 500 รอบจะได้ \overline{AUC} จากทั้งสองตัวแบบ

10. ทำการทดสอบสมมติฐานทางสถิติเพื่อทดสอบว่าตัวแบบไหนเป็นตัวแบบที่ดีที่สุด โดยสมมติฐานที่ใช้ในการทดสอบ

H_0 : ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 1 มากกว่าหรือเท่ากับค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 2

H_1 : ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 1 น้อยกว่าค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 2

$$H_0: \mu_1 \geq \mu_2 \text{ แยกกับ } H_1: \mu_1 < \mu_2$$

เนื่องจากงานวิจัยครั้งนี้เราไม่ทราบค่าความแปรปรวน (σ^2) จึงใช้ S^2 แทน ดังนั้นตัวสถิติที่ใช้ในการทดสอบคือ t-test

ตัวสถิติทดสอบ

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

เกณฑ์ในการตัดสินใจ

ถ้า $t \leq t_{0.05}(t_{0.05,\infty} \approx Z = -1.645)$ จะปฏิเสธ H_0 นั่นคือ ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2
หรือ

สมมติฐานที่ใช้ในการทดสอบ

H_0 : ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 1 น้อยกว่าหรือเท่ากับค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 2

H_1 : ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 1 มากกว่าค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ในตัวแบบที่ 2

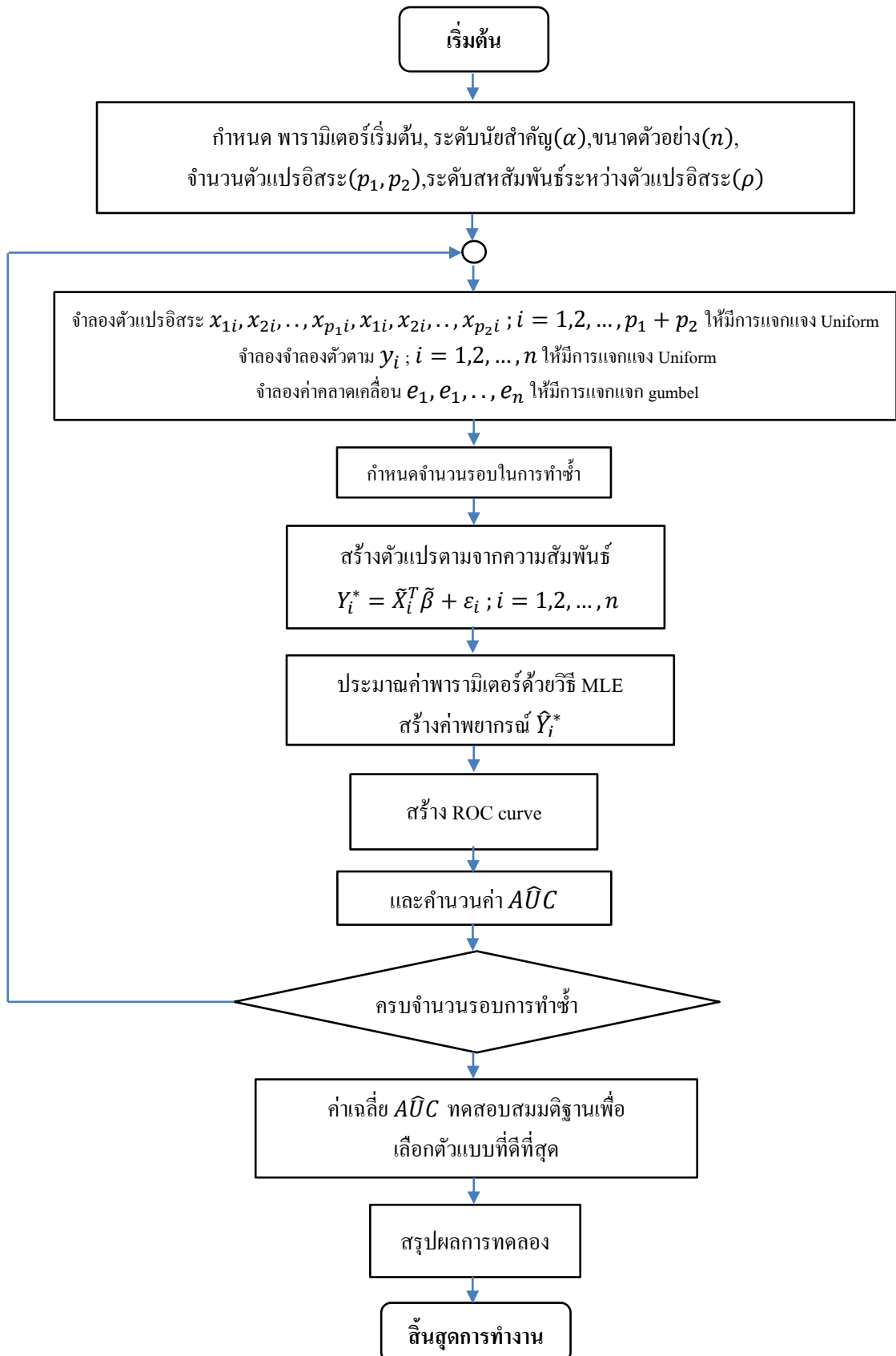
$$H_0: \mu_1 \leq \mu_2 \text{ แอ้งกับ } H_1: \mu_1 > \mu_2$$

เกณฑ์ในการตัดสินใจ

ถ้า $t \geq t_{0.95}(t_{0.95,\infty} \approx Z = 1.645)$ จะปฏิเสธ H_0 นั่นคือ ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 2

11. หรือสรุปผลที่ได้จากการวิจัย

ภาพที่ 3.1 ฟังงานขั้นตอนการวิจัย



บทที่ 4

ผลการวิเคราะห์ข้อมูล

งานวิจัยครั้งนี้เป็นการศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามโดยใช้ฟังก์ชันคอมพลิเมนต์ทารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยงเพื่อศึกษาระดับสหสัมพันธ์ในแต่ละคู่ (Degree of pair-wise correlations) ของตัวแปรอิสระจากทั้งสองตัวแบบทางสถิติที่ดีที่สุดโดยศึกษาภายใต้ตัวแปรดังต่อไปนี้

1. ข้อมูลที่กำหนดให้สหสัมพันธ์ของแต่ละคู่ (ρ_{jj}) เปลี่ยนแปลงไป 0.3, 0.6 และ 0.9 ตามลำดับ เมื่อ $i \neq j = 1, 2, \dots, 5$
2. ข้อมูลที่กำหนดให้จำนวนตัวแปรอิสระของทั้งสองตัวแบบ (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) และ (5,5)
3. ข้อมูลที่กำหนดให้ขนาดตัวอย่าง n มีค่าเท่ากับ 50, 100, 150, 200 และ 250

โดยในการศึกษานี้จะเป็นการศึกษาตัวแบบทางสถิติแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติกทวินามที่มีฟังก์ชันคอมพลิเมนต์ทารีล็อก-ล็อกเป็นฟังก์ชันเชื่อมโยงและหาพื้นที่ใต้โค้ง (Area Under the Curve: AUC) สูงสุดในแต่ละสถานการณ์

สำหรับการนำเสนอผลการวิจัยผู้วิจัยได้กำหนดสัญลักษณ์ต่างๆ เพื่อใช้ในตารางและสรุปผลดังนี้

(p_1, p_2) หมายถึง จำนวนของตัวแปรอิสระของตัวแบบที่ 1 และ ตัวแบบที่ 2 ตามลำดับ

n หมายถึง ขนาดตัวอย่างของเซตตัวแปรอิสระ

$Max\{r_{jj'}\}$ หมายถึง สหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และ

ตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$

\overline{AUC} หมายถึง ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC

M_1 หมายถึง ตัวแบบที่ 1 ประกอบไปด้วยตัวแปรอิสระ x_j ; $j = 1, 2, \dots, p_1$

M_2 หมายถึง ตัวแบบที่ 2 ประกอบไปด้วยตัวแปรอิสระ $x_{j'}$; $j' = 1, 2, \dots, p_2$

μ_1 หมายถึง ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1

μ_2 หมายถึง ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC ของตัวแบบที่ 1

โดยได้ผลการวิจัยตามจุดประสงค์ของการวิจัยดังนี้

ตารางที่ 4.1 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
2	2	50	ต่ำ	0.31977	0.70871	0.29048	0.70805	ตัวแบบที่ 1 หรือ ตัวแบบที่ 2
			ปานกลาง	0.64878	0.70058	0.64213	0.69513	ตัวแบบที่ 1 หรือ ตัวแบบที่ 2
			สูง	0.65257	0.68160	0.90113	0.67529	ตัวแบบที่ 1 หรือ ตัวแบบที่ 2
		100	ต่ำ	0.31975	0.71940	0.28400	0.71733	ตัวแบบที่ 1
			ปานกลาง	0.64995	0.70737	0.62479	0.70184	ตัวแบบที่ 1 หรือ ตัวแบบที่ 2
			สูง	0.94611	0.69246	0.86365	0.68300	ตัวแบบที่ 1 หรือ ตัวแบบที่ 2
		150	ต่ำ	0.31875	0.72616	0.31875	0.71518	ตัวแบบที่ 1
			ปานกลาง	0.64839	0.71571	0.50512	0.70521	ตัวแบบที่ 1
			สูง	0.92476	0.69862	0.85898	0.68608	ตัวแบบที่ 1
		200	ต่ำ	0.31988	0.72760	0.26054	0.71971	ตัวแบบที่ 1
			ปานกลาง	0.64943	0.71370	0.59551	0.70415	ตัวแบบที่ 1
			สูง	0.92387	0.70031	0.85168	0.68827	ตัวแบบที่ 1

ตารางที่ 4.1 (ต่อ)

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
2	2	250	ต่ำ	0.31765	0.72736	0.27772	0.71992	ตัวแบบที่ 1
			ปานกลาง	0.64448	0.71561	0.57615	0.70717	ตัวแบบที่ 1
			สูง	0.91756	0.69950	0.84157	0.68806	ตัวแบบที่ 1

จากตารางที่ 4.1 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง

และทำการคัดเลือกตัวแบบจากการสมมติฐาน

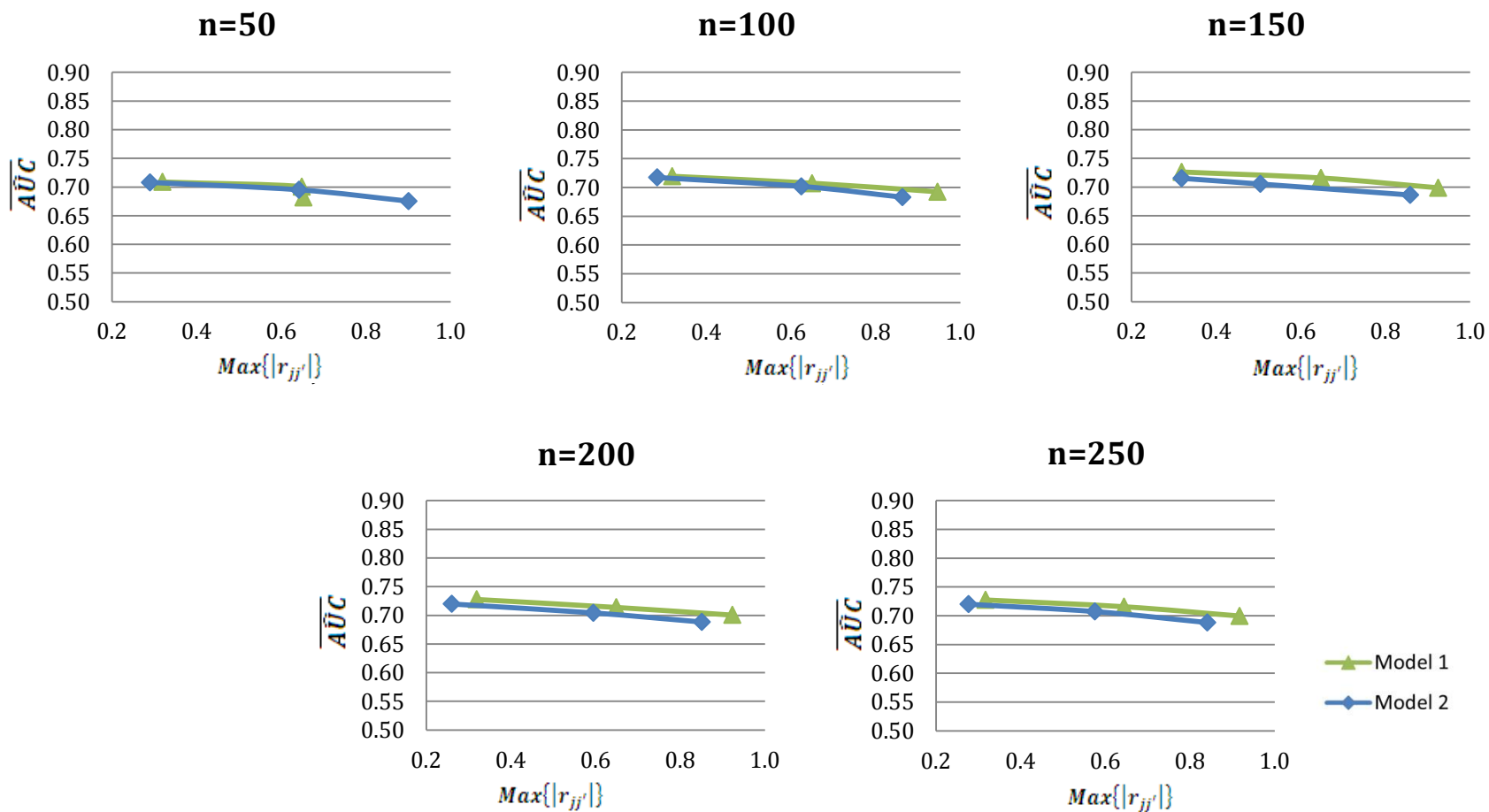
โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่

สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แอ้งกับ $H_a: \mu_1 > \mu_2$

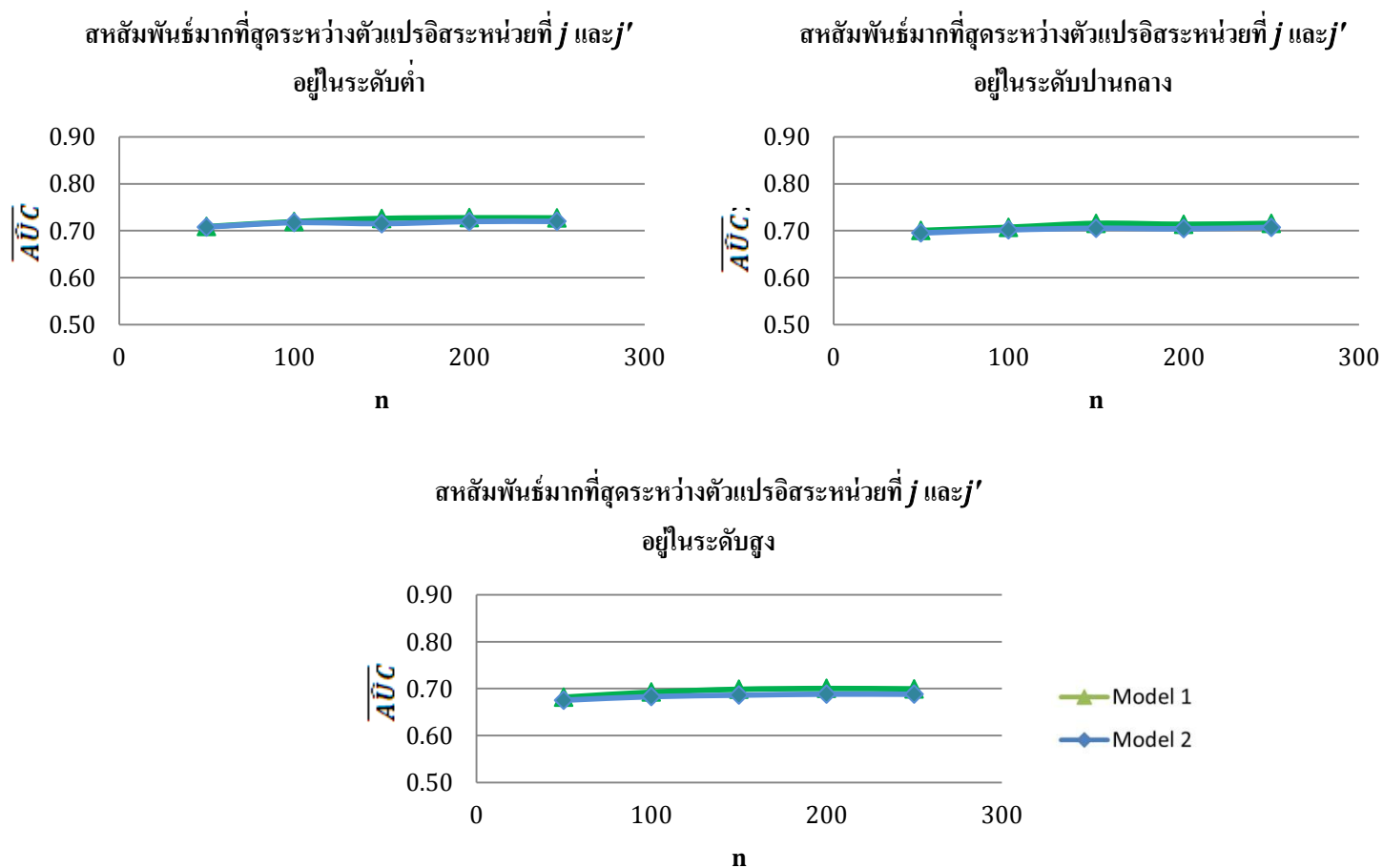
ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ($P\text{-Value} < 0.05$) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

ยกเว้นกรณีที่ ขนาดตัวอย่างของเขตตัวแปรอิสระ $n = 50$ และ $n = 100$ ที่ค่าสหสัมพันธ์มากที่สุดอยู่ในระดับปานกลางและสูง ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่าหรือเท่ากับ ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ($P\text{-Value} > 0.05$) ตัวแบบที่ถูกเลือกจะเป็นตัวแบบที่ 1 หรือตัวแบบที่ 2 ก็ได้

ภาพที่ 4.1 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.2 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 2)$, $\text{Max}\{|r_{jj'}|\}$ อยู่ใน ระดับต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



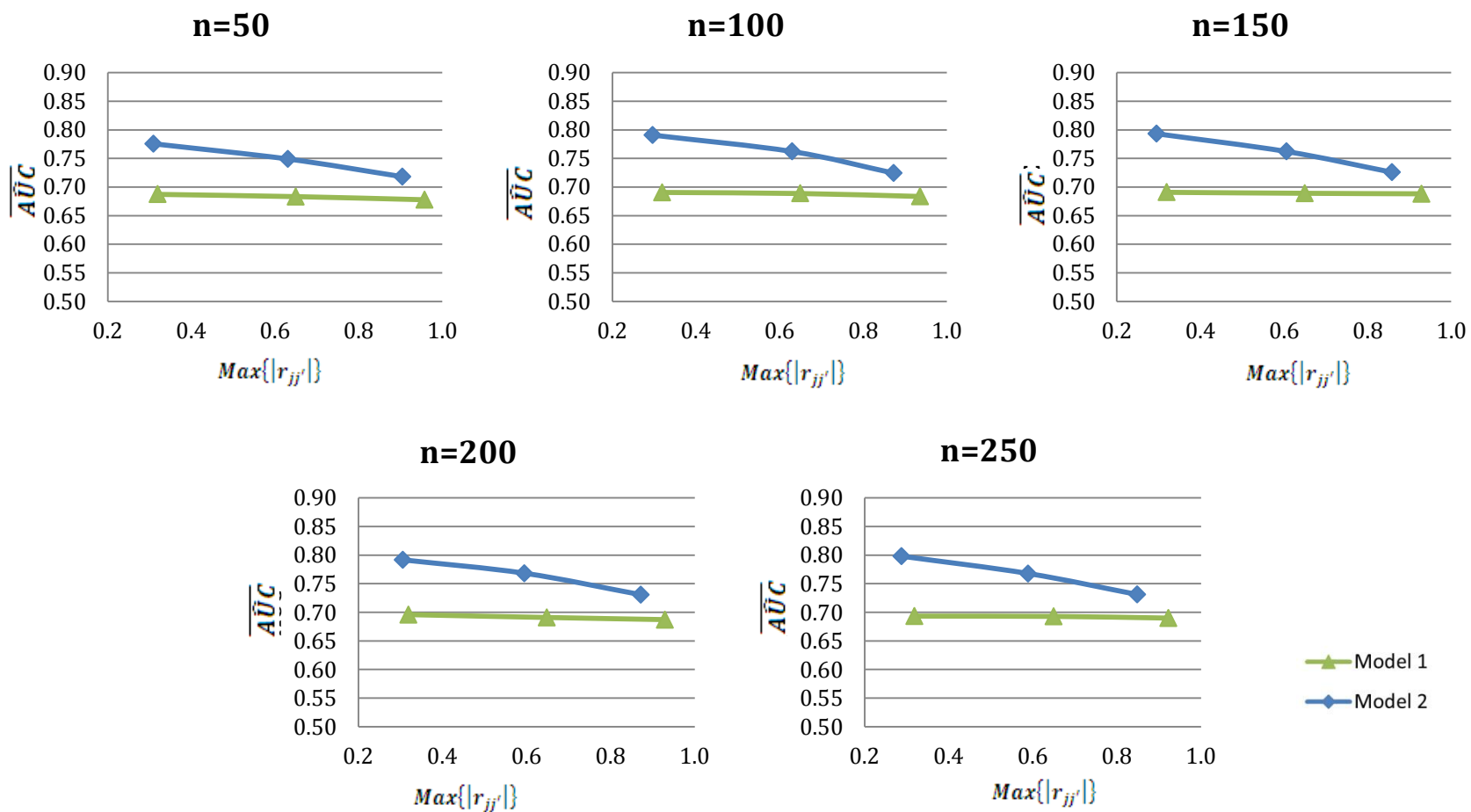
ตารางที่ 4.2 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{ r_{jj'} \}$	\overline{AUC}	$Max\{ r_{jj'} \}$	\overline{AUC}	
2	3	50	ต่ำ	0.31981	0.68721	0.30989	0.77565	ตัวแบบที่ 2
			ปานกลาง	0.64976	0.68325	0.63078	0.74911	ตัวแบบที่ 2
			สูง	0.95742	0.67779	0.90462	0.71808	ตัวแบบที่ 2
		100	ต่ำ	0.31998	0.69058	0.29674	0.79084	ตัวแบบที่ 2
			ปานกลาง	0.64956	0.68869	0.63056	0.76207	ตัวแบบที่ 2
			สูง	0.93628	0.68351	0.87313	0.72426	ตัวแบบที่ 2
		150	ต่ำ	0.31976	0.69078	0.29541	0.79325	ตัวแบบที่ 2
			ปานกลาง	0.64981	0.68907	0.60630	0.76215	ตัวแบบที่ 2
			สูง	0.92865	0.68817	0.85841	0.72584	ตัวแบบที่ 2
		200	ต่ำ	0.31993	0.69589	0.30611	0.79176	ตัวแบบที่ 2
			ปานกลาง	0.64889	0.69091	0.59551	0.76830	ตัวแบบที่ 2
			สูง	0.92992	0.68716	0.87212	0.73072	ตัวแบบที่ 2
		250	ต่ำ	0.31910	0.69335	0.28848	0.79807	ตัวแบบที่ 2
			ปานกลาง	0.64996	0.69282	0.58902	0.76785	ตัวแบบที่ 2
			สูง	0.92271	0.68987	0.84885	0.73107	ตัวแบบที่ 2

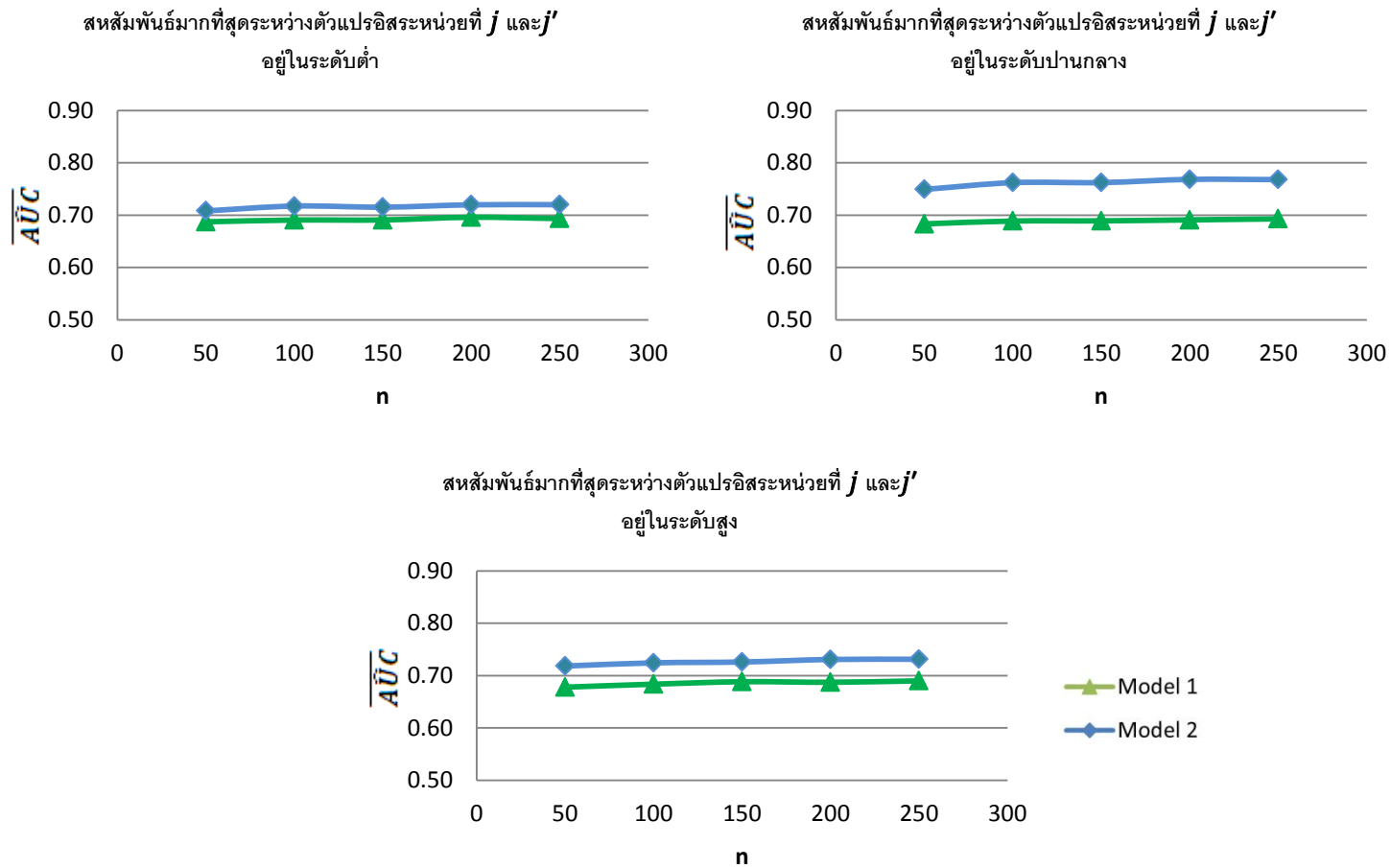
จากตารางที่ 4.2 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \geq \mu_2$ แยกกับ $H_1: \mu_1 < \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2

ภาพที่ 4.3 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.4 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 3)$, $\text{Max}\{|r_{jj'}|\}$. อยู่ใน ระดับต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



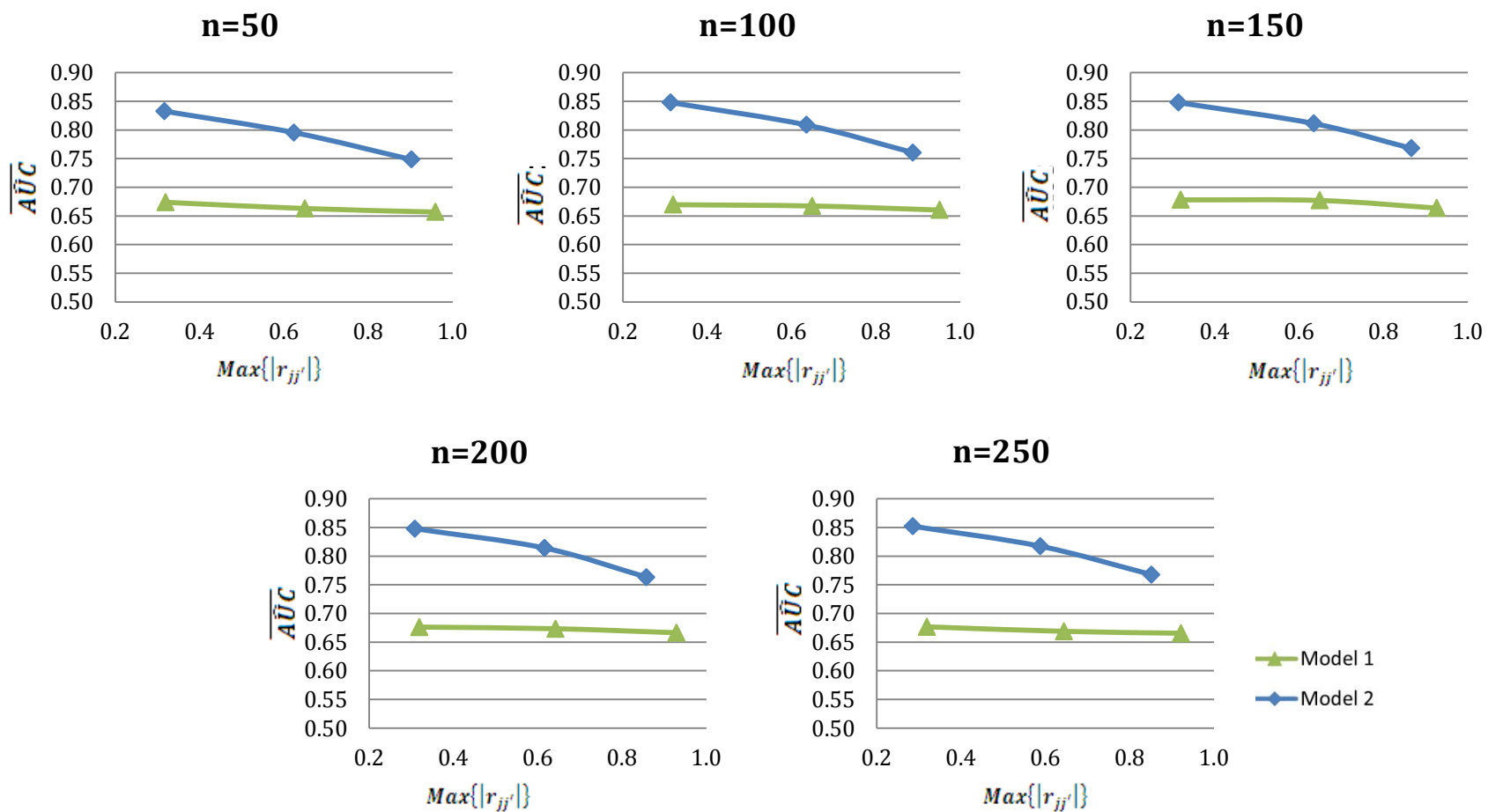
ตารางที่ 4.3 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{ r_{jj'} \}$	\overline{AUC}	$Max\{ r_{jj'} \}$	\overline{AUC}	
2	4	50	ต่ำ	0.31953	0.67375	0.31688	0.83274	ตัวแบบที่ 2
			ปานกลาง	0.64991	0.66294	0.62429	0.79543	ตัวแบบที่ 2
			สูง	0.95955	0.65677	0.90286	0.74834	ตัวแบบที่ 2
		100	ต่ำ	0.31998	0.66961	0.31383	0.84793	ตัวแบบที่ 2
			ปานกลาง	0.64968	0.66734	0.63575	0.80873	ตัวแบบที่ 2
			สูง	0.95169	0.66028	0.88840	0.76049	ตัวแบบที่ 2
		150	ต่ำ	0.31998	0.67804	0.31452	0.84775	ตัวแบบที่ 2
			ปานกลาง	0.64972	0.67698	0.63518	0.81123	ตัวแบบที่ 2
			สูง	0.92696	0.66371	0.86647	0.76796	ตัวแบบที่ 2
		200	ต่ำ	0.31992	0.67621	0.30926	0.84786	ตัวแบบที่ 2
			ปานกลาง	0.64279	0.67335	0.61678	0.81414	ตัวแบบที่ 2
			สูง	0.92992	0.66631	0.85822	0.76340	ตัวแบบที่ 2
		250	ต่ำ	0.31953	0.67664	0.28635	0.85216	ตัวแบบที่ 2
			ปานกลาง	0.64461	0.66873	0.58839	0.81733	ตัวแบบที่ 2
			สูง	0.92215	0.66546	0.85194	0.76783	ตัวแบบที่ 2

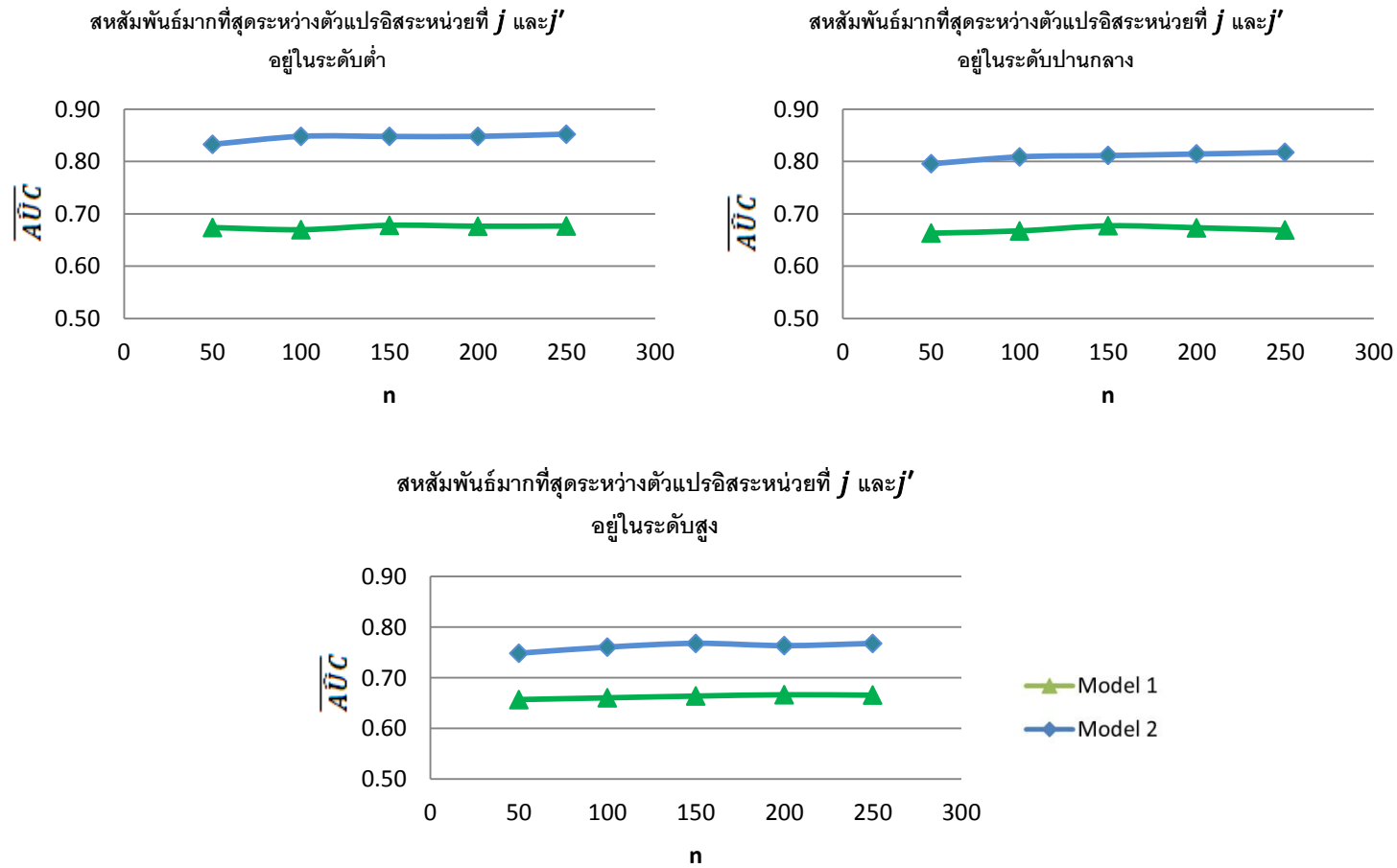
จากตารางที่ 4.3 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \geq \mu_2$ แยังกับ $H_1: \mu_1 < \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2

ภาพที่ 4.5 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.6 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 4)$, $\text{Max}\{r_{jj'}\}$, อยู่ใน ระดับต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



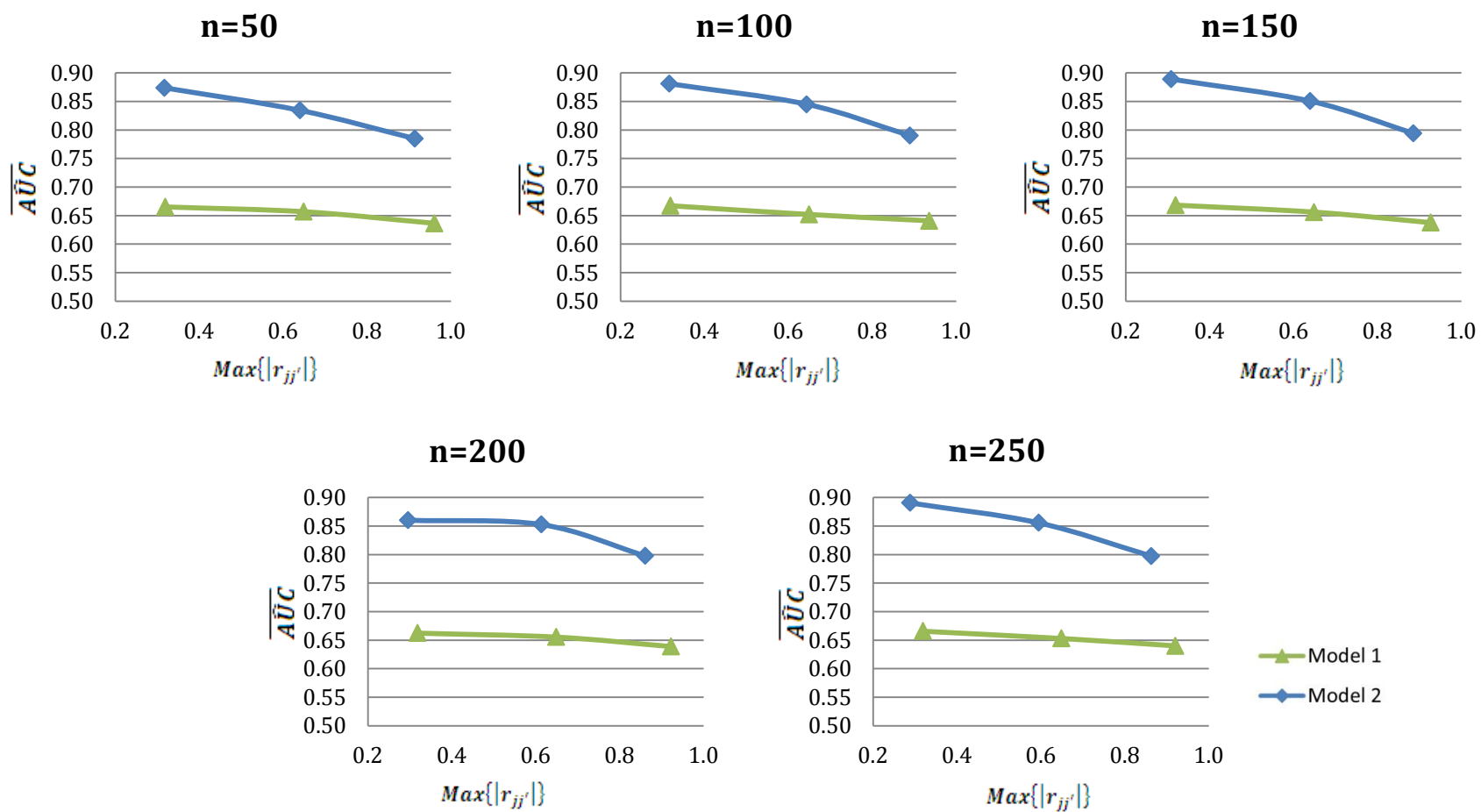
ตารางที่ 4.4 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 5)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
2	5	50	ต่ำ	0.31981	0.66510	0.31745	0.87391	ตัวแบบที่ 2
			ปานกลาง	0.64972	0.65671	0.64043	0.83436	ตัวแบบที่ 2
			สูง	0.96097	0.63641	0.91429	0.78486	ตัวแบบที่ 2
		100	ต่ำ	0.31994	0.66721	0.31714	0.88108	ตัวแบบที่ 2
			ปานกลาง	0.64995	0.65189	0.64418	0.84507	ตัวแบบที่ 2
			สูง	0.93628	0.64085	0.89004	0.79023	ตัวแบบที่ 2
		150	ต่ำ	0.31953	0.66821	0.30923	0.88912	ตัวแบบที่ 2
			ปานกลาง	0.64973	0.65599	0.63983	0.85064	ตัวแบบที่ 2
			สูง	0.92774	0.63756	0.88593	0.79424	ตัวแบบที่ 2
		200	ต่ำ	0.31878	0.66226	0.29630	0.86022	ตัวแบบที่ 2
			ปานกลาง	0.64943	0.65531	0.61394	0.85260	ตัวแบบที่ 2
			สูง	0.92292	0.63877	0.86172	0.79757	ตัวแบบที่ 2
		250	ต่ำ	0.31966	0.66558	0.28928	0.89046	ตัวแบบที่ 2
			ปานกลาง	0.64962	0.65287	0.59524	0.85549	ตัวแบบที่ 2
			สูง	0.92117	0.63997	0.86361	0.79729	ตัวแบบที่ 2

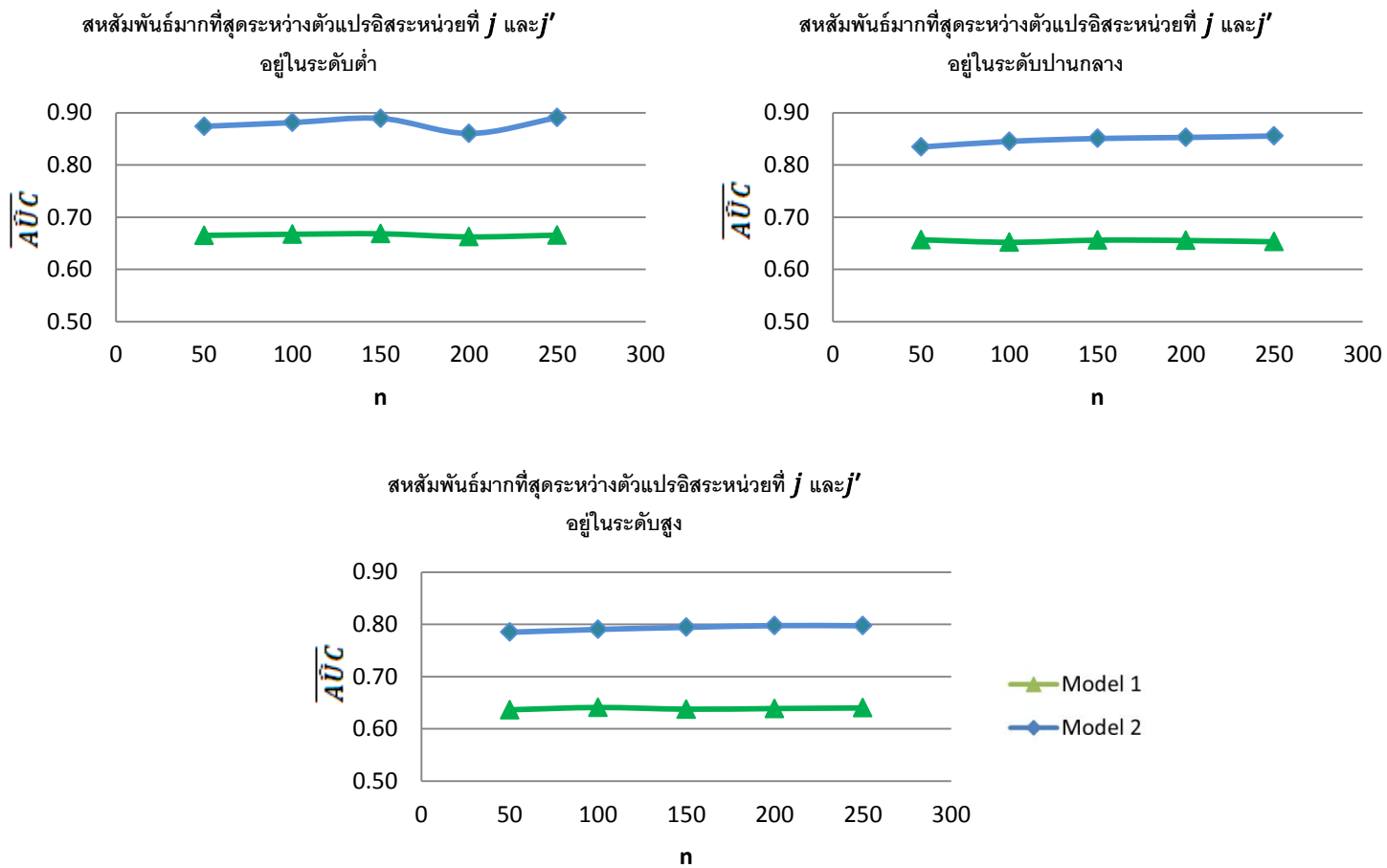
จากตารางที่ 4.4 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 5)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \geq \mu_2$ แยกกับ $H_1: \mu_1 < \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2

ภาพที่ 4.7 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.8 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (2, 5)$, $Max\{|r_{jj'}|\}$. อยู่ใน ระดับต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



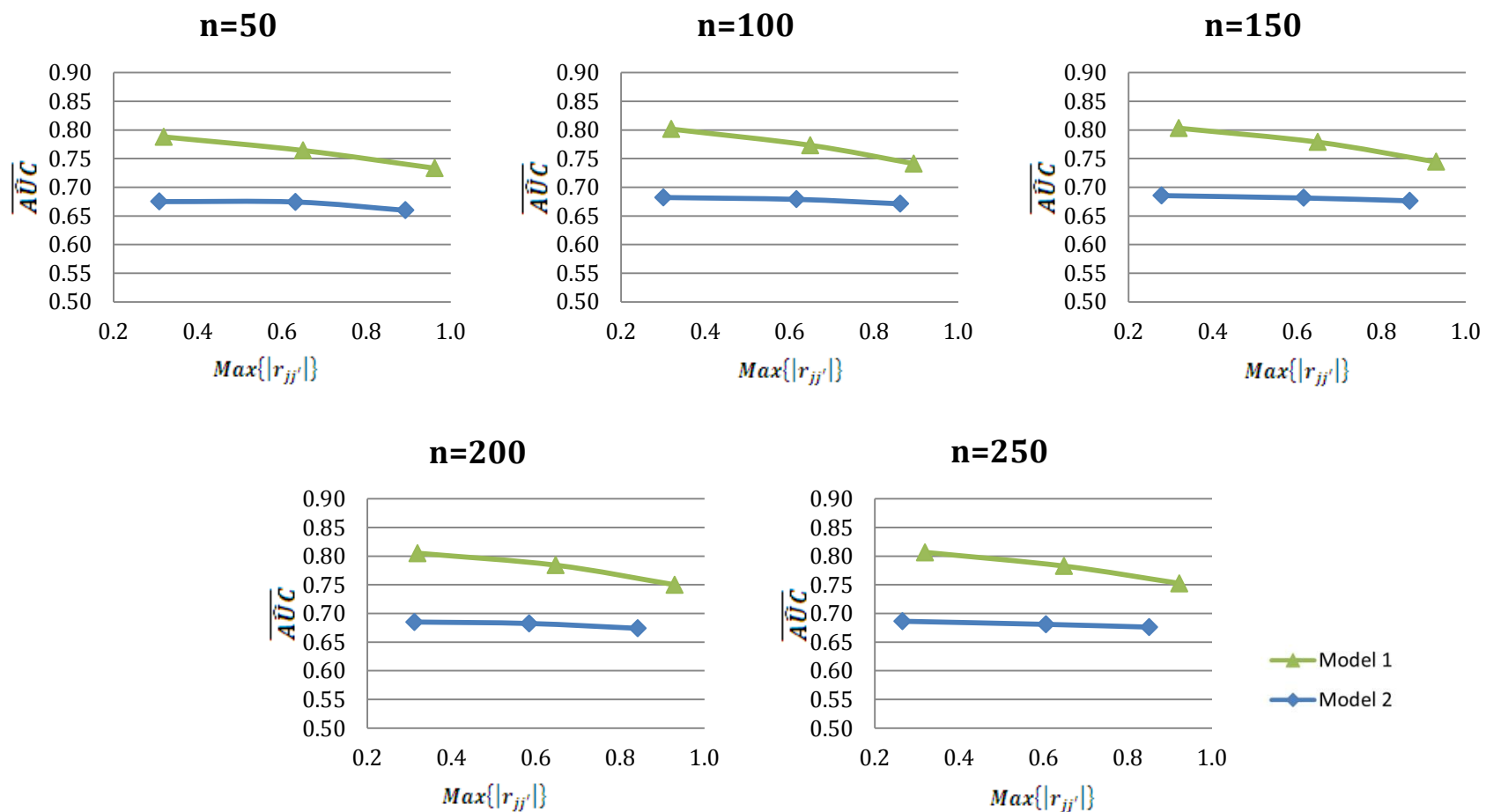
ตารางที่ 4.5 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 2)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
3	2	50	ต่ำ	0.32000	0.78791	0.30874	0.67497	ตัวแบบที่ 1
			ปานกลาง	0.64993	0.76422	0.63217	0.67428	ตัวแบบที่ 1
			สูง	0.96268	0.73326	0.89255	0.65976	ตัวแบบที่ 1
		100	ต่ำ	0.31976	0.80156	0.30067	0.68218	ตัวแบบที่ 1
			ปานกลาง	0.64968	0.77324	0.61676	0.67887	ตัวแบบที่ 1
			สูง	0.89521	0.74123	0.86274	0.67114	ตัวแบบที่ 1
		150	ต่ำ	0.31999	0.80311	0.27902	0.68547	ตัวแบบที่ 1
			ปานกลาง	0.64932	0.77876	0.61617	0.68131	ตัวแบบที่ 1
			สูง	0.93009	0.74427	0.86733	0.67636	ตัวแบบที่ 1
		200	ต่ำ	0.31983	0.80516	0.31257	0.68503	ตัวแบบที่ 1
			ปานกลาง	0.64759	0.78419	0.58477	0.68246	ตัวแบบที่ 1
			สูง	0.92970	0.75001	0.84232	0.67400	ตัวแบบที่ 1
		250	ต่ำ	0.31976	0.80664	0.26639	0.68636	ตัวแบบที่ 1
			ปานกลาง	0.64927	0.78267	0.60692	0.68091	ตัวแบบที่ 1
			สูง	0.92271	0.75255	0.85164	0.67612	ตัวแบบที่ 1

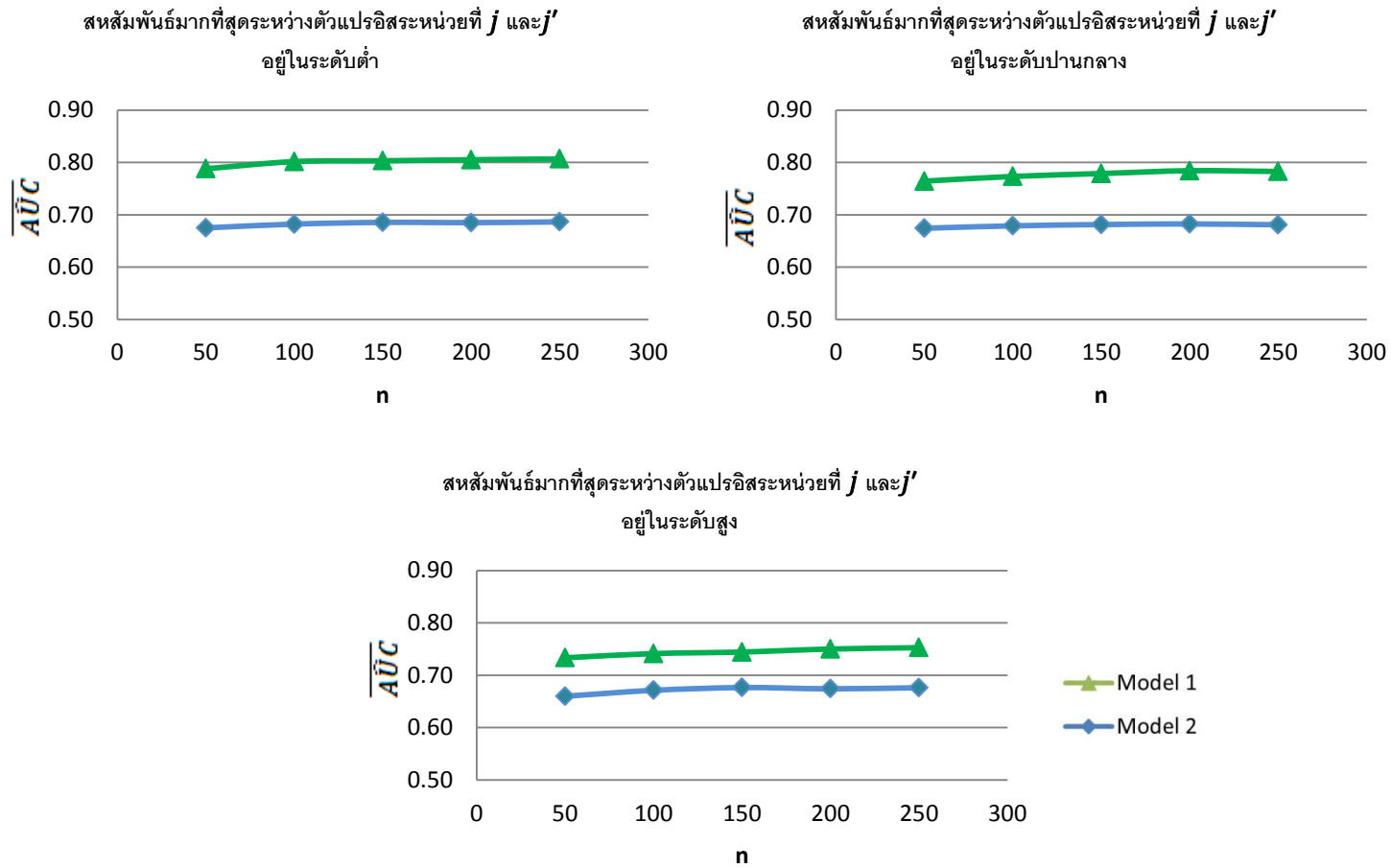
จากตารางที่ 4.5 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 2)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แยังกับ $H_1: \mu_1 > \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า เมื่อค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

ภาพที่ 4.9 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ใน ระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.10 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 2)$, $Max\{r_{jj'}\}$. อยู่ใน ระดับต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



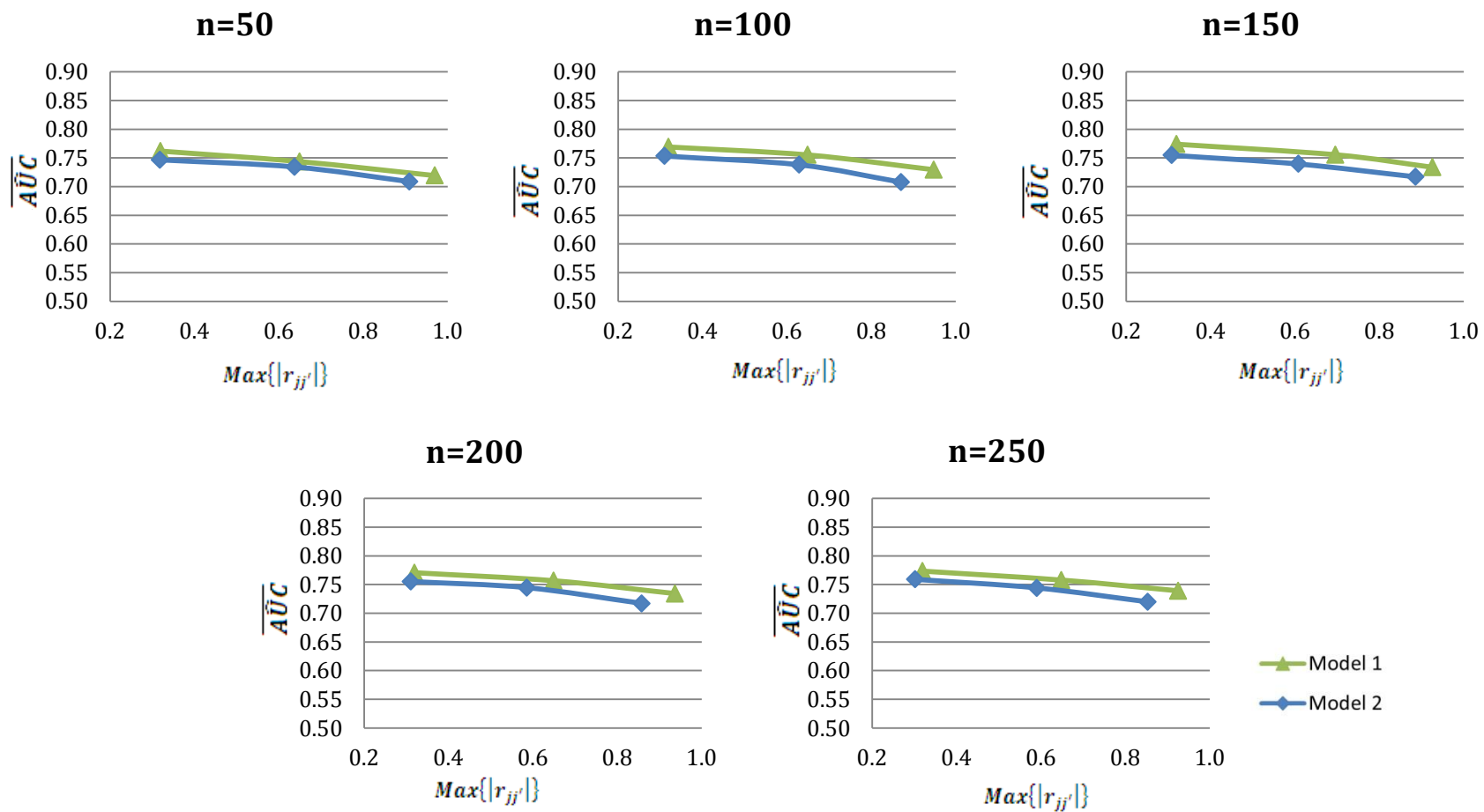
ตารางที่ 4.6 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 3)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
3	3	50	ต่ำ	0.31988	0.76194	0.31806	0.74664	ตัวแบบที่ 1
			ปานกลาง	0.64935	0.74360	0.63762	0.73427	ตัวแบบที่ 1
			สูง	0.96963	0.71923	0.90972	0.70870	ตัวแบบที่ 1
		100	ต่ำ	0.31978	0.76908	0.31051	0.75330	ตัวแบบที่ 1
			ปานกลาง	0.64964	0.75521	0.62931	0.73805	ตัวแบบที่ 1
			สูง	0.94865	0.72940	0.87106	0.70772	ตัวแบบที่ 1
		150	ต่ำ	0.31995	0.77390	0.30816	0.75465	ตัวแบบที่ 1
			ปานกลาง	0.69635	0.75553	0.60840	0.73940	ตัวแบบที่ 1
			สูง	0.92656	0.73363	0.88625	0.71700	ตัวแบบที่ 1
		200	ต่ำ	0.31993	0.77062	0.31073	0.75528	ตัวแบบที่ 1
			ปานกลาง	0.64957	0.75665	0.58573	0.74461	ตัวแบบที่ 1
			สูง	0.93691	0.73429	0.85811	0.71720	ตัวแบบที่ 1
		250	ต่ำ	0.31988	0.77342	0.30241	0.75906	ตัวแบบที่ 1
			ปานกลาง	0.64877	0.75781	0.59029	0.74417	ตัวแบบที่ 1
			สูง	0.92596	0.73905	0.85352	0.71989	ตัวแบบที่ 1

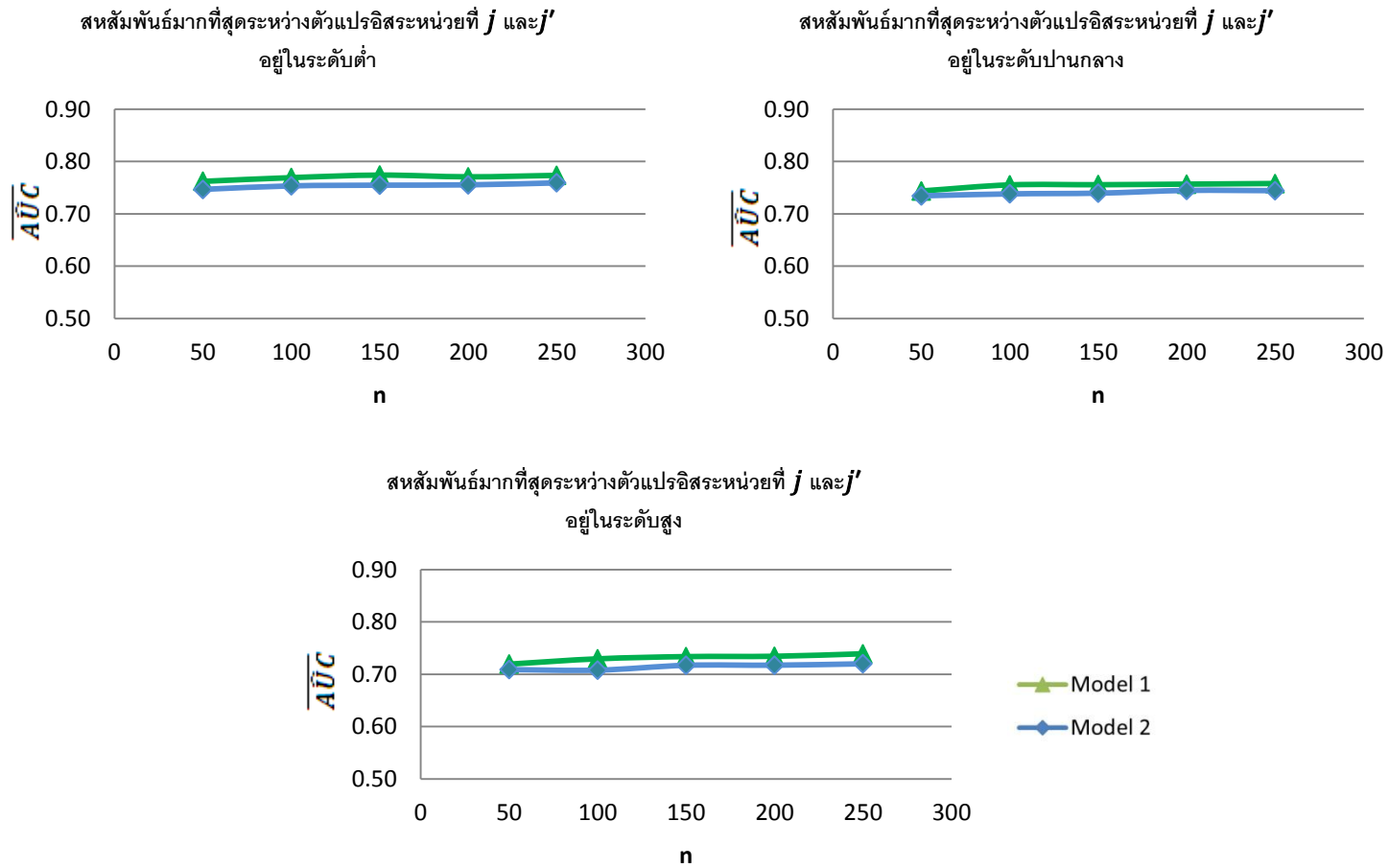
จากตารางที่ 4.6 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 3)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลงและเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 (M_1) มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แยังกับ $H_1: \mu_1 > \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

ภาพที่ 4.11 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.12 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3,3)$, $Max\{|r_{jj'}|\}$. อยู่ใน ระดับต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



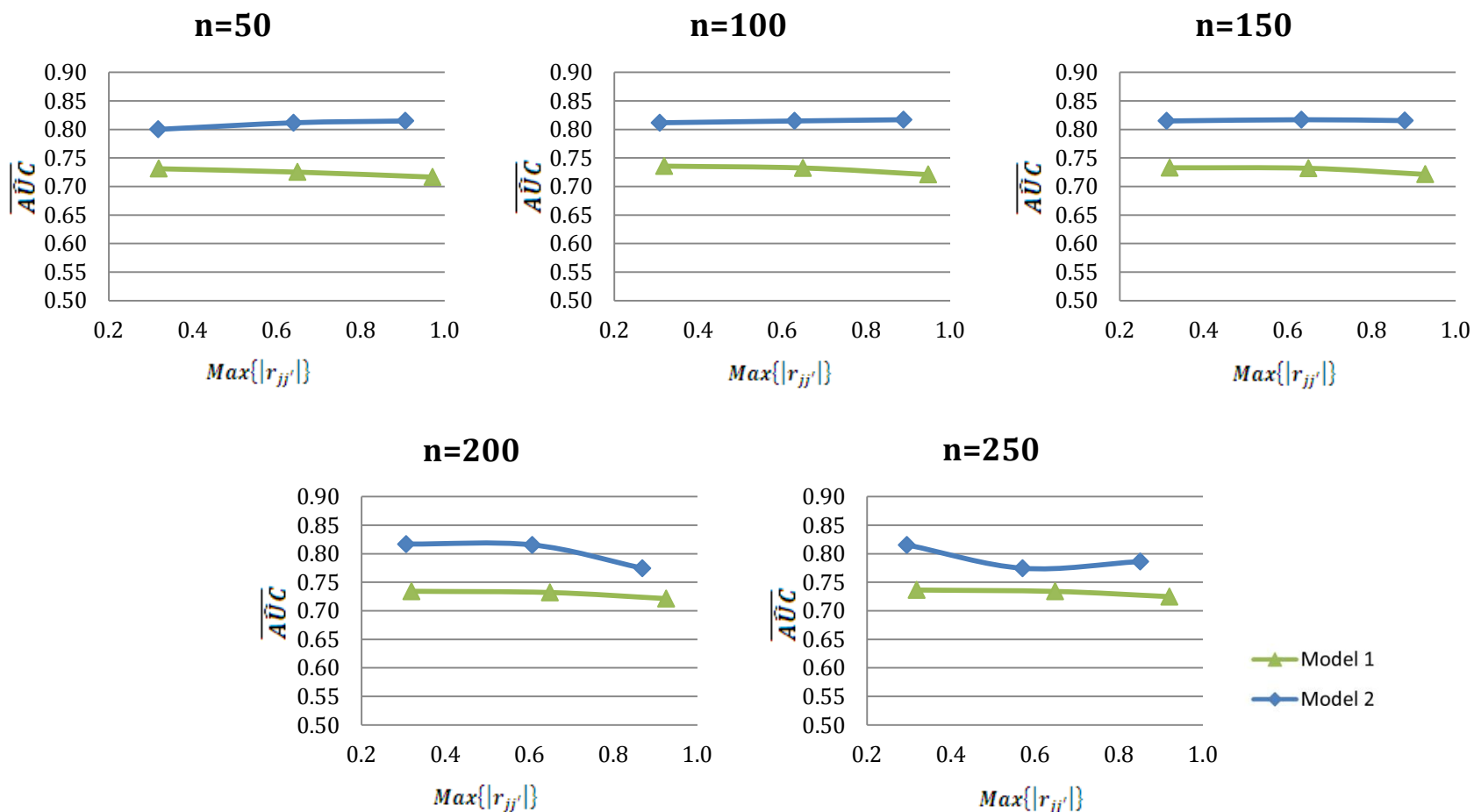
ตารางที่ 4.7 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{ r_{jj'} \}$	\overline{AUC}	$Max\{ r_{jj'} \}$	\overline{AUC}	
3	4	50	ต่ำ	0.31996	0.73107	0.31837	0.80013	ตัวแบบที่ 2
			ปานกลาง	0.64995	0.72513	0.64088	0.77461	ตัวแบบที่ 2
			สูง	0.97213	0.71640	0.90710	0.74057	ตัวแบบที่ 2
		100	ต่ำ	0.31999	0.73566	0.30882	0.81166	ตัวแบบที่ 2
			ปานกลาง	0.64995	0.73242	0.63008	0.78642	ตัวแบบที่ 2
			สูง	0.94819	0.72078	0.88933	0.74656	ตัวแบบที่ 2
		150	ต่ำ	0.31942	0.73283	0.31229	0.81483	ตัวแบบที่ 2
			ปานกลาง	0.64986	0.73189	0.63310	0.79024	ตัวแบบที่ 2
			สูง	0.92838	0.72117	0.87969	0.75012	ตัวแบบที่ 2
		200	ต่ำ	0.31945	0.73423	0.30704	0.81703	ตัวแบบที่ 2
			ปานกลาง	0.64943	0.73202	0.60743	0.78914	ตัวแบบที่ 2
			สูง	0.92652	0.72137	0.86980	0.75111	ตัวแบบที่ 2
		250	ต่ำ	0.31900	0.73641	0.29595	0.81546	ตัวแบบที่ 2
			ปานกลาง	0.64863	0.73394	0.57136	0.79358	ตัวแบบที่ 2
			สูง	0.92117	0.72481	0.85160	0.75232	ตัวแบบที่ 2

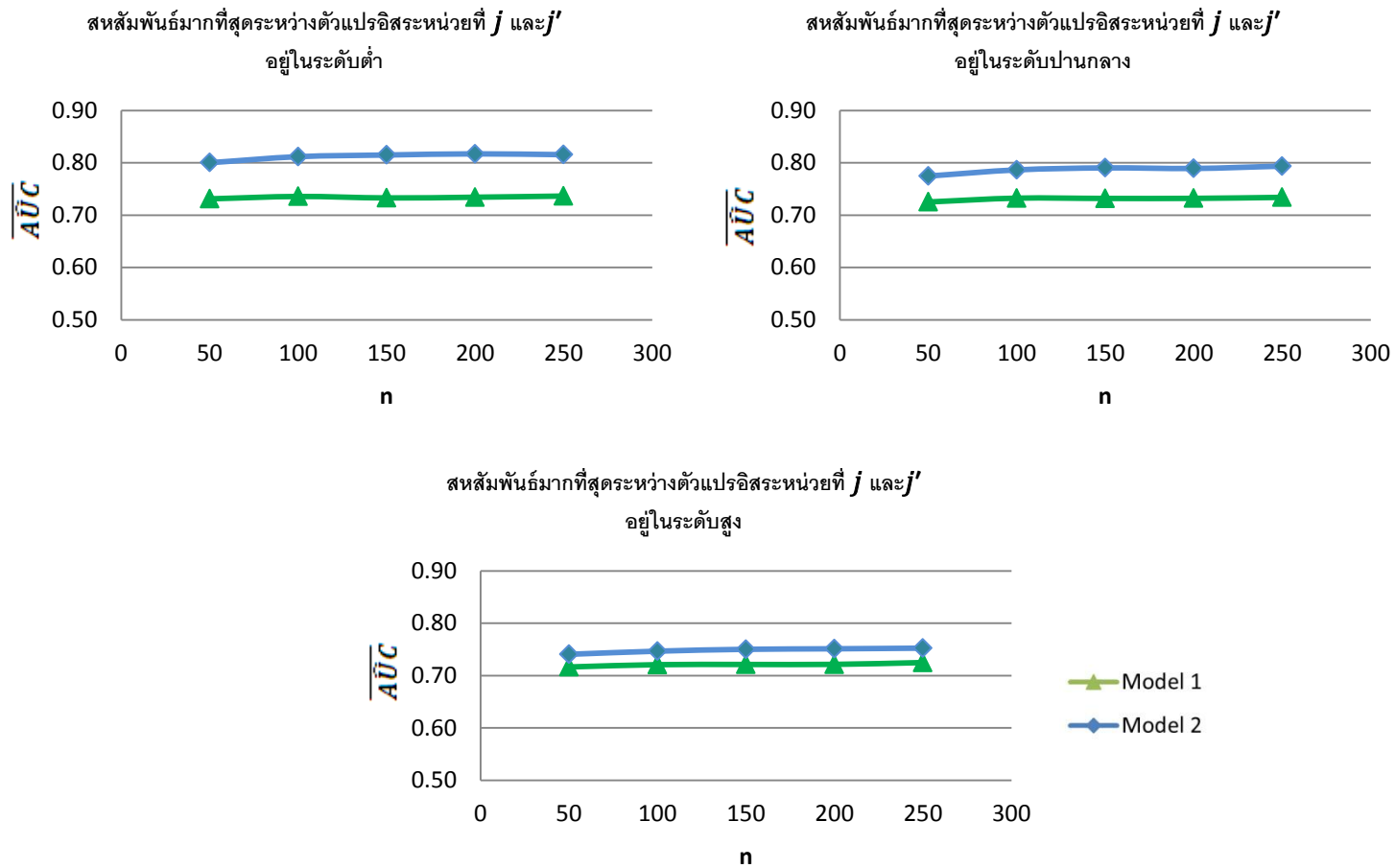
จากตารางที่ 4.7 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 4)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \geq \mu_2$ แยกกับ $H_1: \mu_1 < \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2

ภาพที่ 4.13 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.14 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 4)$, $\text{Max}\{r_{jj'}\}$. อยู่ใน ระดับต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



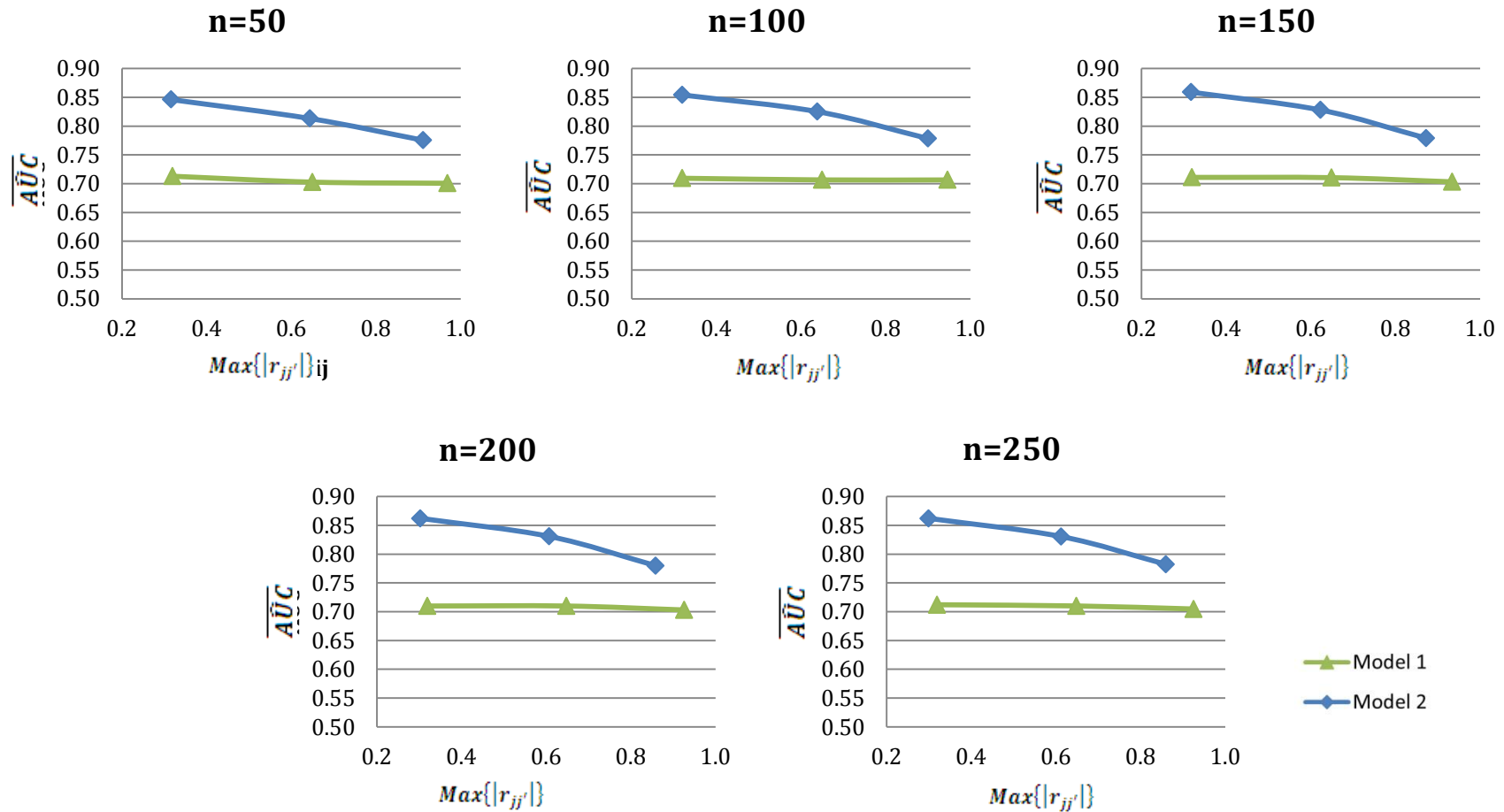
ตารางที่ 4.8 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 5)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
3	5	50	ต่ำ	0.31984	0.71296	0.31665	0.84614	ตัวแบบที่ 2
			ปานกลาง	0.64996	0.70272	0.64407	0.81323	ตัวแบบที่ 2
			สูง	0.96940	0.70073	0.91157	0.77547	ตัวแบบที่ 2
		100	ต่ำ	0.31998	0.70942	0.31945	0.85422	ตัวแบบที่ 2
			ปานกลาง	0.64975	0.70657	0.63883	0.82521	ตัวแบบที่ 2
			สูง	0.94182	0.70650	0.90025	0.77837	ตัวแบบที่ 2
		150	ต่ำ	0.31996	0.71095	0.31780	0.85899	ตัวแบบที่ 2
			ปานกลาง	0.64952	0.71046	0.62345	0.82798	ตัวแบบที่ 2
			สูง	0.93411	0.70318	0.872819	0.77909	ตัวแบบที่ 2
		200	ต่ำ	0.31992	0.70990	0.30326	0.86221	ตัวแบบที่ 2
			ปานกลาง	0.64808	0.70988	0.60743	0.83086	ตัวแบบที่ 2
			สูง	0.92652	0.70301	0.85873	0.78006	ตัวแบบที่ 2
		250	ต่ำ	0.31998	0.71209	0.29981	0.86224	ตัวแบบที่ 2
			ปานกลาง	0.64863	0.71000	0.61277	0.83056	ตัวแบบที่ 2
			สูง	0.92535	0.70473	0.85978	0.78234	ตัวแบบที่ 2

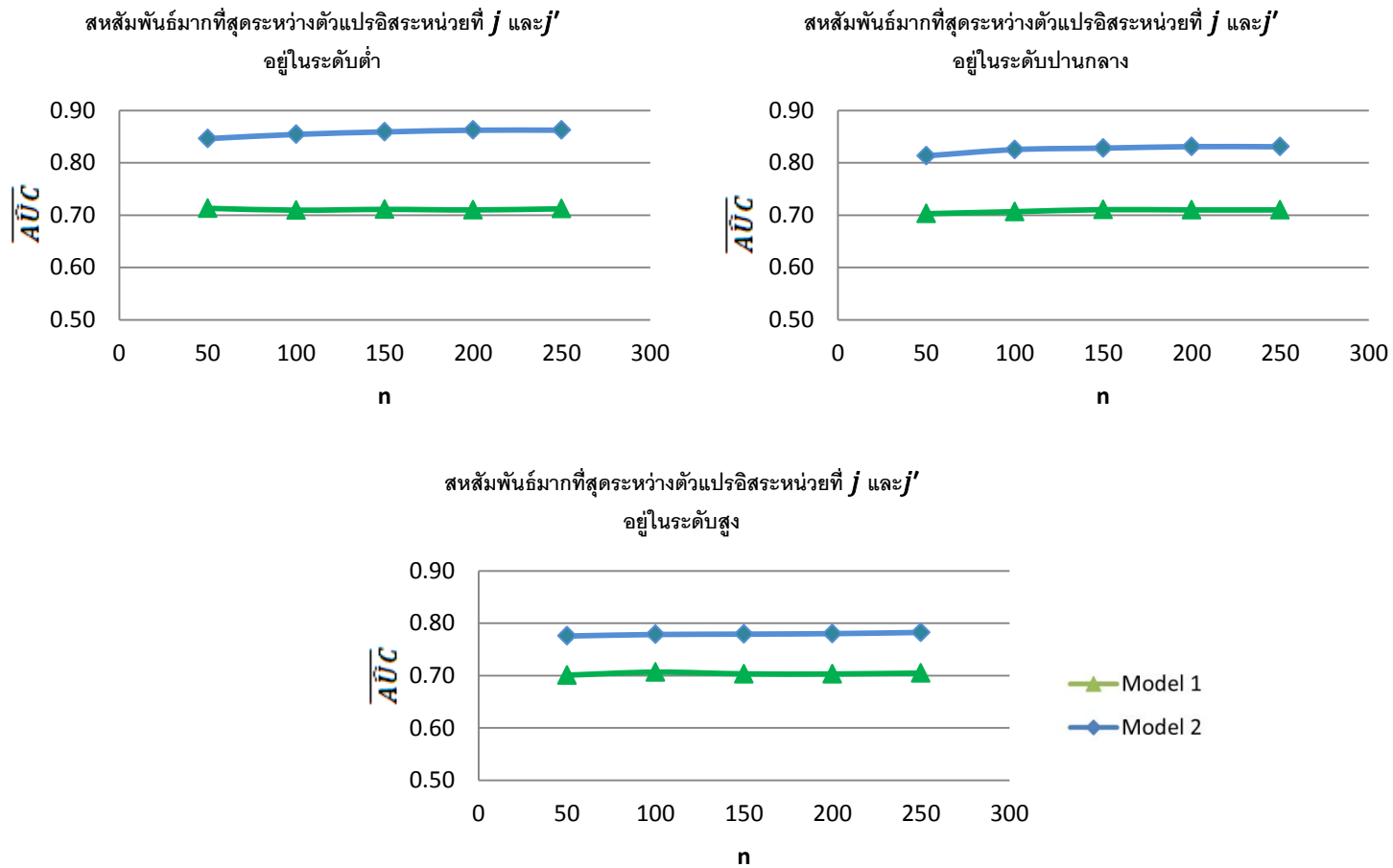
จากตารางที่ 4.8 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 5)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \geq \mu_2$ แยกกับ $H_1: \mu_1 < \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2

ภาพที่ 4.15 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.16 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (3, 5)$, $Max\{|r_{jj'}|\}$. อยู่ใน ระดับ ต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



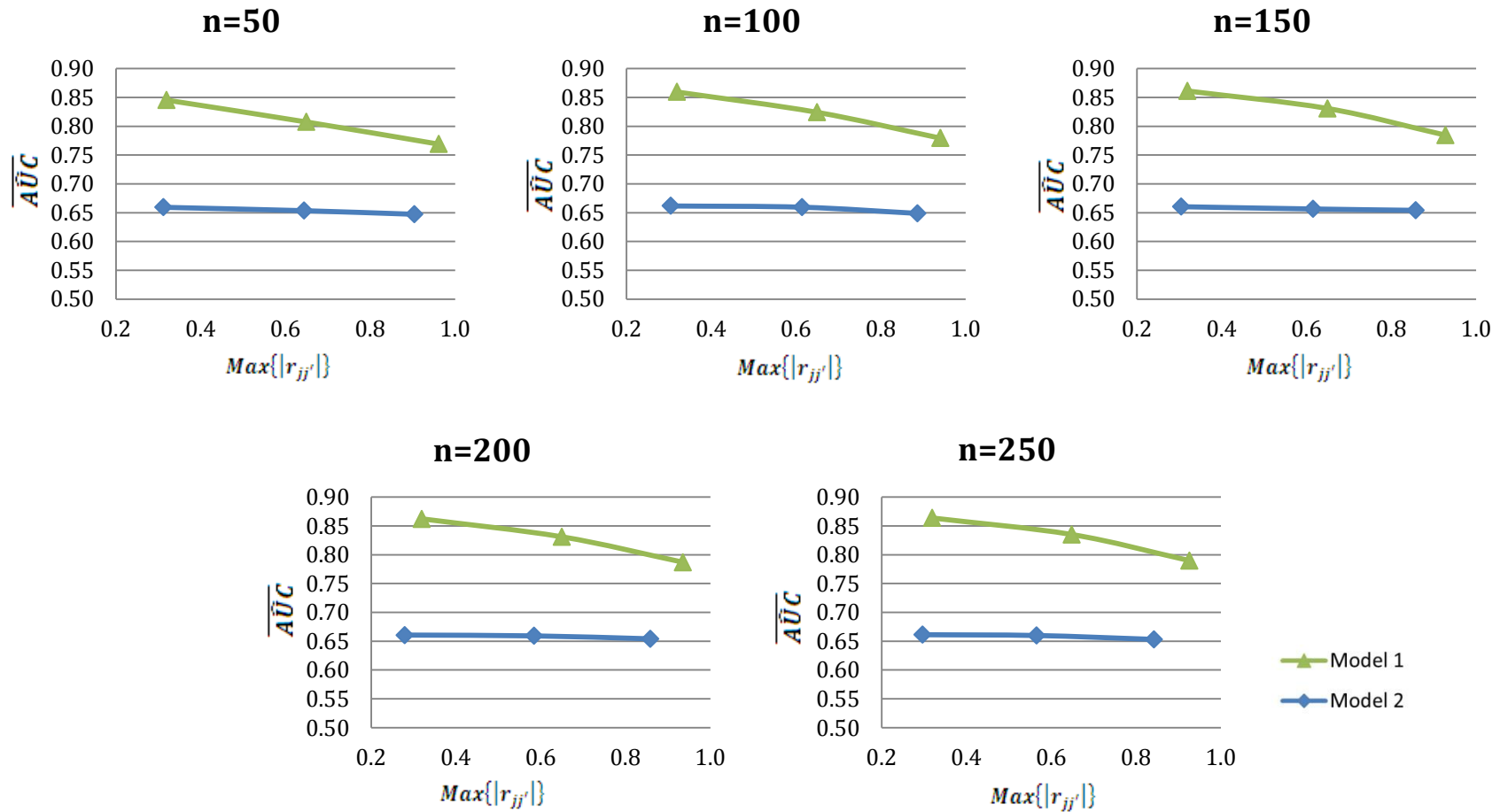
ตารางที่ 4.9 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 2)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
4	2	50	ต่ำ	0.32000	0.84542	0.31223	0.65933	ตัวแบบที่ 1
			ปานกลาง	0.64967	0.80742	0.64419	0.65341	ตัวแบบที่ 1
			สูง	0.96149	0.76896	0.90445	0.64719	ตัวแบบที่ 1
		100	ต่ำ	0.31986	0.85943	0.30475	0.66165	ตัวแบบที่ 1
			ปานกลาง	0.64984	0.82422	0.61446	0.65951	ตัวแบบที่ 1
			สูง	0.94131	0.77923	0.88690	0.64875	ตัวแบบที่ 1
		150	ต่ำ	0.31993	0.86125	0.30519	0.66036	ตัวแบบที่ 1
			ปานกลาง	0.64998	0.83085	0.61617	0.65640	ตัวแบบที่ 1
			สูง	0.92845	0.78449	0.85793	0.65402	ตัวแบบที่ 1
		200	ต่ำ	0.31992	0.86234	0.27973	0.66059	ตัวแบบที่ 1
			ปานกลาง	0.64998	0.83114	0.58477	0.65918	ตัวแบบที่ 1
			สูง	0.93566	0.78702	0.85874	0.65402	ตัวแบบที่ 1
		250	ต่ำ	0.31983	0.86384	0.29729	0.66118	ตัวแบบที่ 1
			ปานกลาง	0.64905	0.83509	0.56584	0.65982	ตัวแบบที่ 1
			สูง	0.92609	0.78987	0.84287	0.65307	ตัวแบบที่ 1

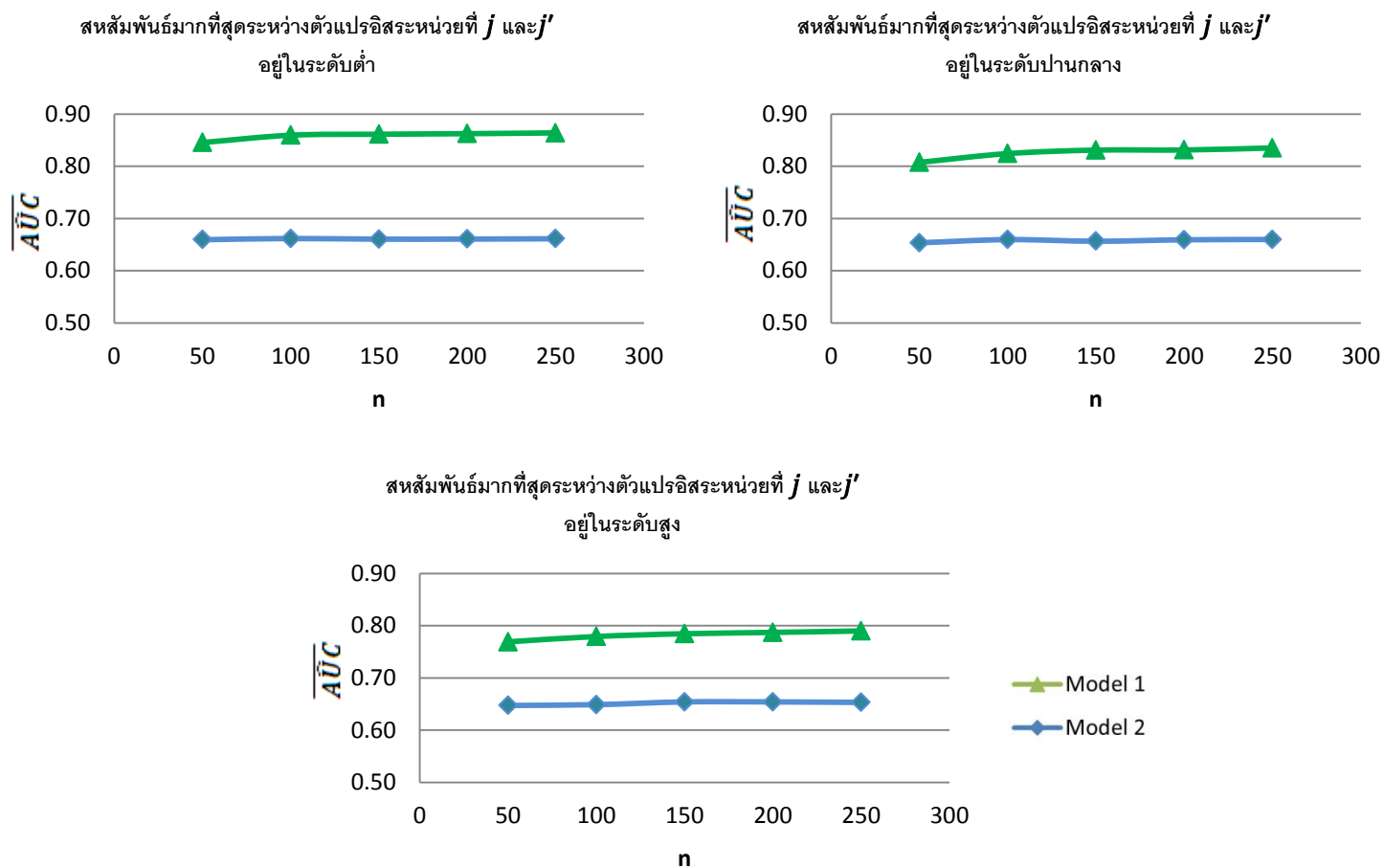
จากตารางที่ 4.9 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 2)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แอ้งกับ $H_1: \mu_1 > \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

ภาพที่ 4.17 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.18 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 2)$, $\text{Max}\{|r_{jj'}|\}$. อยู่ใน ระดับ ต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



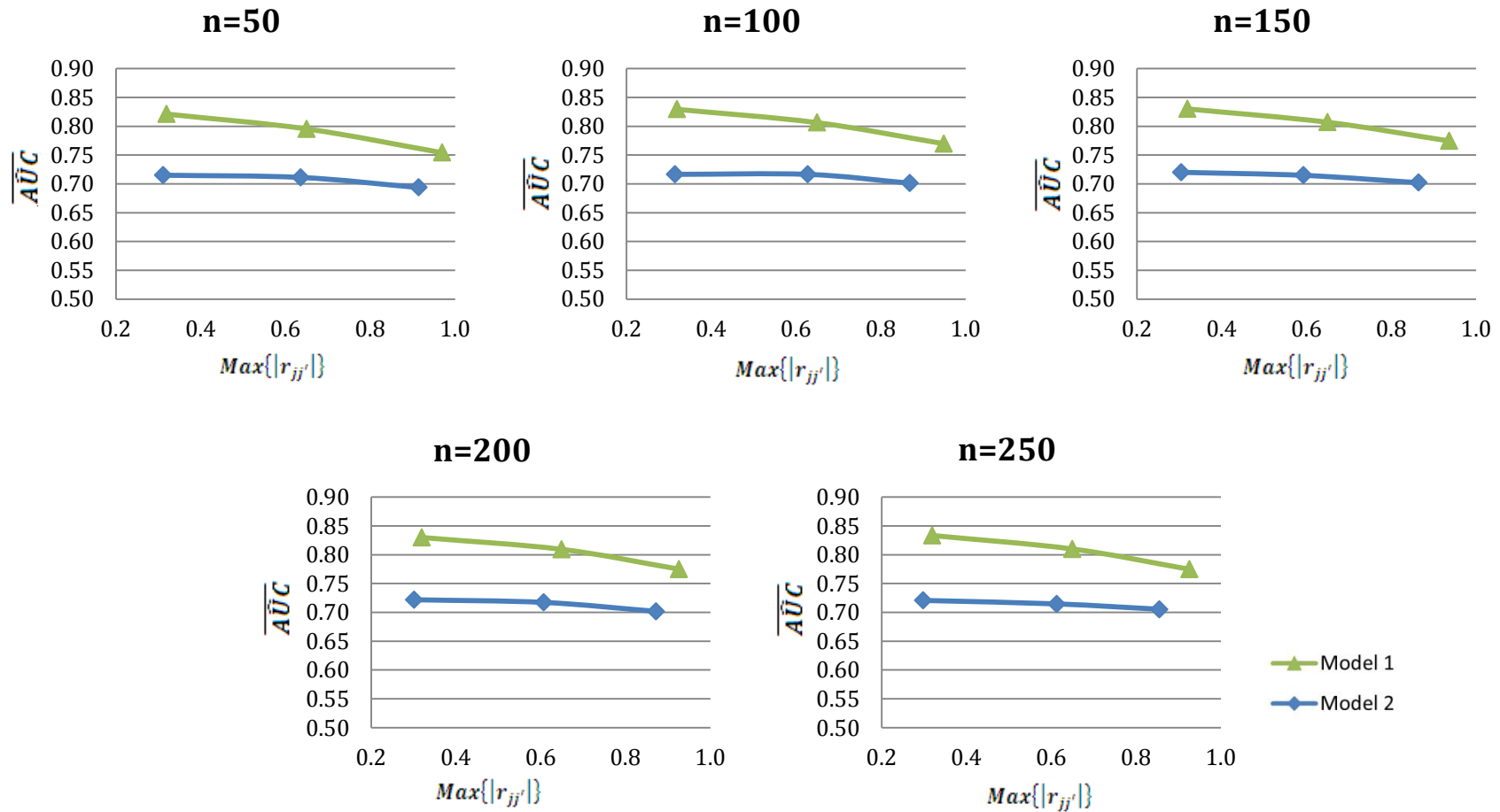
ตารางที่ 4.10 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 3)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
4	3	50	ต่ำ	0.31964	0.82113	0.31172	0.71504	ตัวแบบที่ 1
			ปานกลาง	0.64996	0.79539	0.63600	0.71116	ตัวแบบที่ 1
			สูง	0.96963	0.75412	0.91429	0.69362	ตัวแบบที่ 1
		100	ต่ำ	0.31972	0.82939	0.31518	0.71656	ตัวแบบที่ 1
			ปานกลาง	0.64981	0.80628	0.62821	0.71649	ตัวแบบที่ 1
			สูง	0.94873	0.76973	0.86876	0.70124	ตัวแบบที่ 1
		150	ต่ำ	0.31999	0.83012	0.30509	0.72004	ตัวแบบที่ 1
			ปานกลาง	0.64985	0.80694	0.59297	0.71480	ตัวแบบที่ 1
			สูง	0.93725	0.77441	0.86469	0.70203	ตัวแบบที่ 1
		200	ต่ำ	0.31994	0.82995	0.30183	0.72189	ตัวแบบที่ 1
			ปานกลาง	0.64961	0.80936	0.60743	0.71724	ตัวแบบที่ 1
			สูง	0.92652	0.77492	0.87212	0.70173	ตัวแบบที่ 1
		250	ต่ำ	0.31998	0.83338	0.29826	0.72068	ตัวแบบที่ 1
			ปานกลาง	0.64978	0.81003	0.61364	0.71468	ตัวแบบที่ 1
			สูง	0.92610	0.77478	0.85552	0.70539	ตัวแบบที่ 1

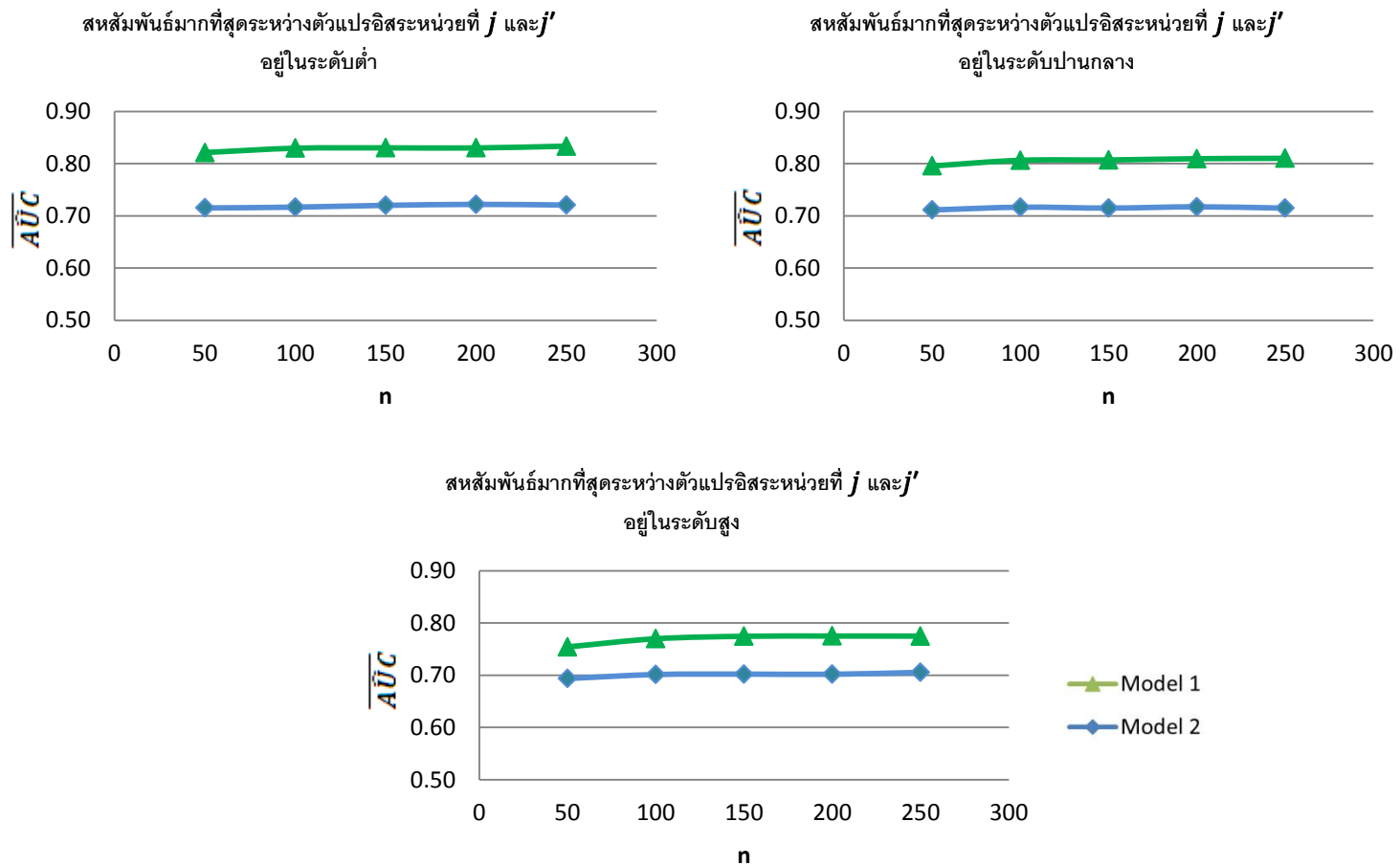
จากตารางที่ 4.10 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 3)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แยังกับ $H_1: \mu_1 > \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

ภาพที่ 4.19 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



รูปที่ 4.20 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4,3)$, $Max\{|r_{jj'}|\}$. อยู่ใน ระดับ ต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



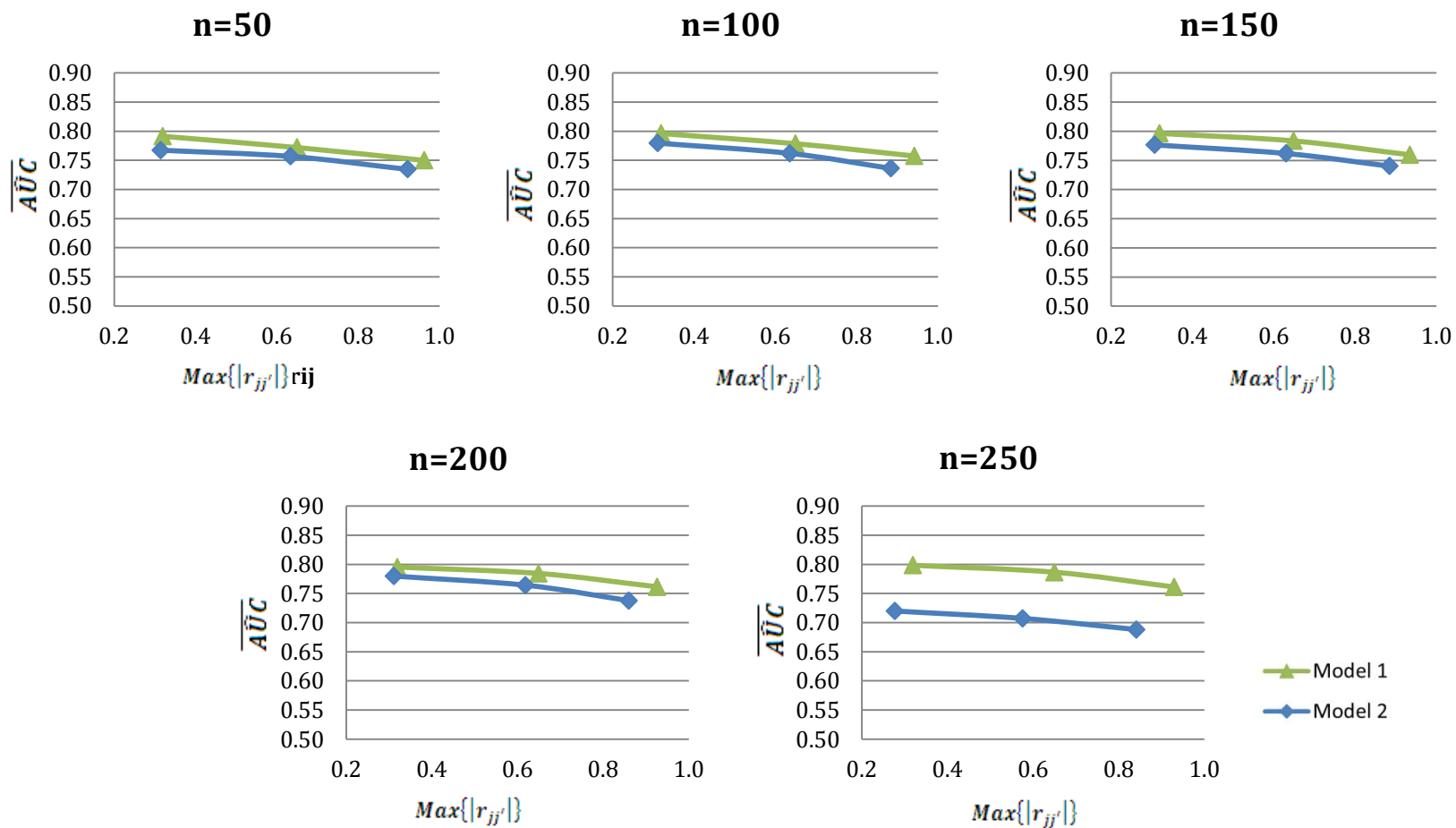
ตารางที่ 4.11 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 4)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
4	4	50	ต่ำ	0.31996	0.79126	0.31486	0.76739	ตัวแบบที่ 1
			ปานกลาง	0.64996	0.77196	0.63366	0.75675	ตัวแบบที่ 1
			สูง	0.96249	0.74985	0.92147	0.73453	ตัวแบบที่ 1
		100	ต่ำ	0.31998	0.79607	0.31150	0.77961	ตัวแบบที่ 1
			ปานกลาง	0.64993	0.77856	0.63558	0.76195	ตัวแบบที่ 1
			สูง	0.94182	0.75742	0.88446	0.73617	ตัวแบบที่ 1
		150	ต่ำ	0.31983	0.79632	0.30754	0.77662	ตัวแบบที่ 1
			ปานกลาง	0.64900	0.78348	0.63059	0.76189	ตัวแบบที่ 1
			สูง	0.93500	0.75962	0.88522	0.74018	ตัวแบบที่ 1
		200	ต่ำ	0.31987	0.79534	0.31139	0.77991	ตัวแบบที่ 1
			ปานกลาง	0.64969	0.78441	0.61859	0.76431	ตัวแบบที่ 1
			สูง	0.92652	0.76145	0.85987	0.73777	ตัวแบบที่ 1
250	ต่ำ	0.31998	0.79844	0.30013	0.77857	ตัวแบบที่ 1		
	ปานกลาง	0.64988	0.78653	0.61084	0.76591	ตัวแบบที่ 1		
	สูง	0.92960	0.76113	0.86046	0.73956	ตัวแบบที่ 1		

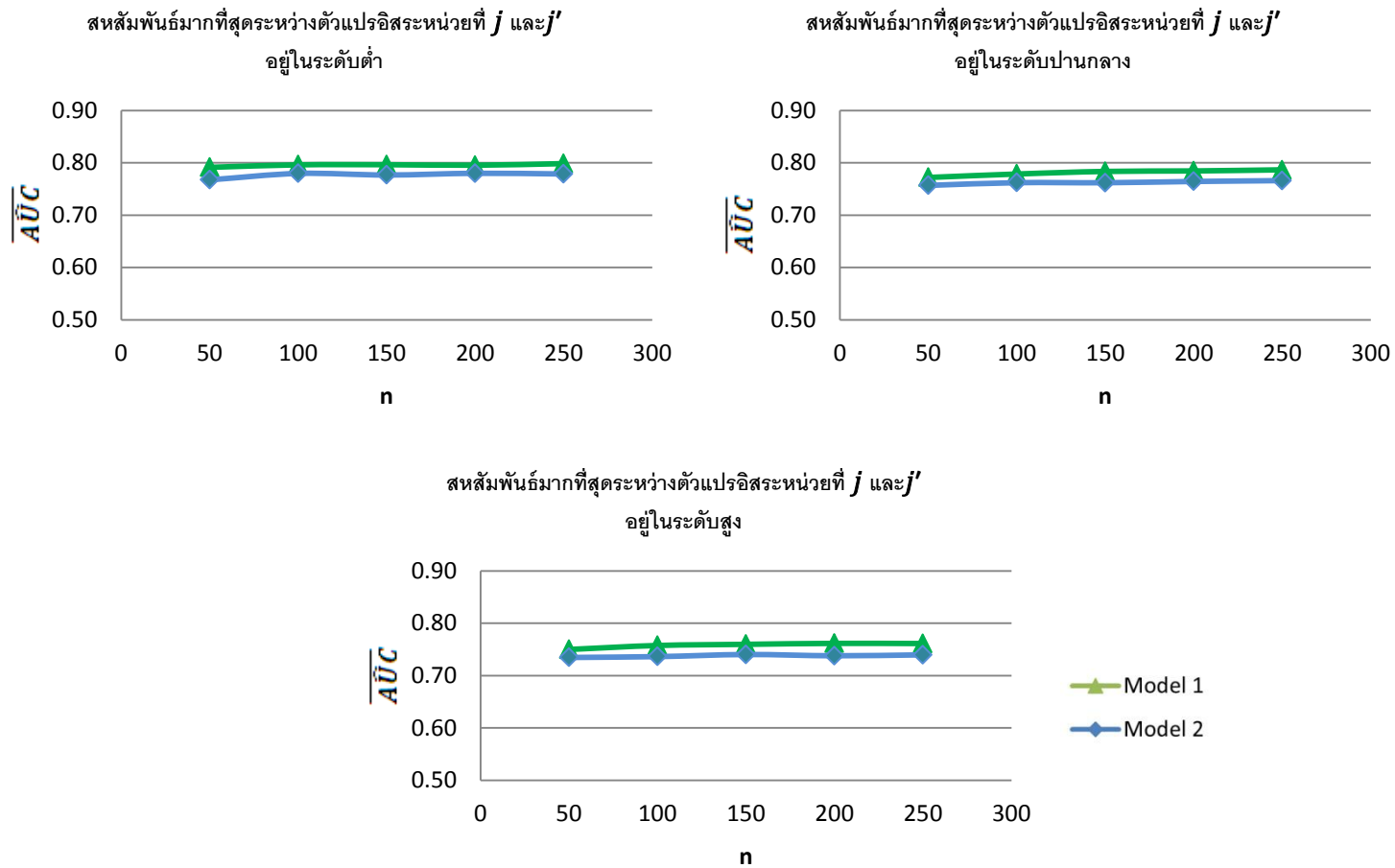
จากตารางที่ 4.11 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 4)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แยังกับ $H_1: \mu_1 > \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบ 1

ภาพที่ 4.21 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.22 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 4)$, $Max\{r_{jj'}\}$. อยู่ใน ระดับ ต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



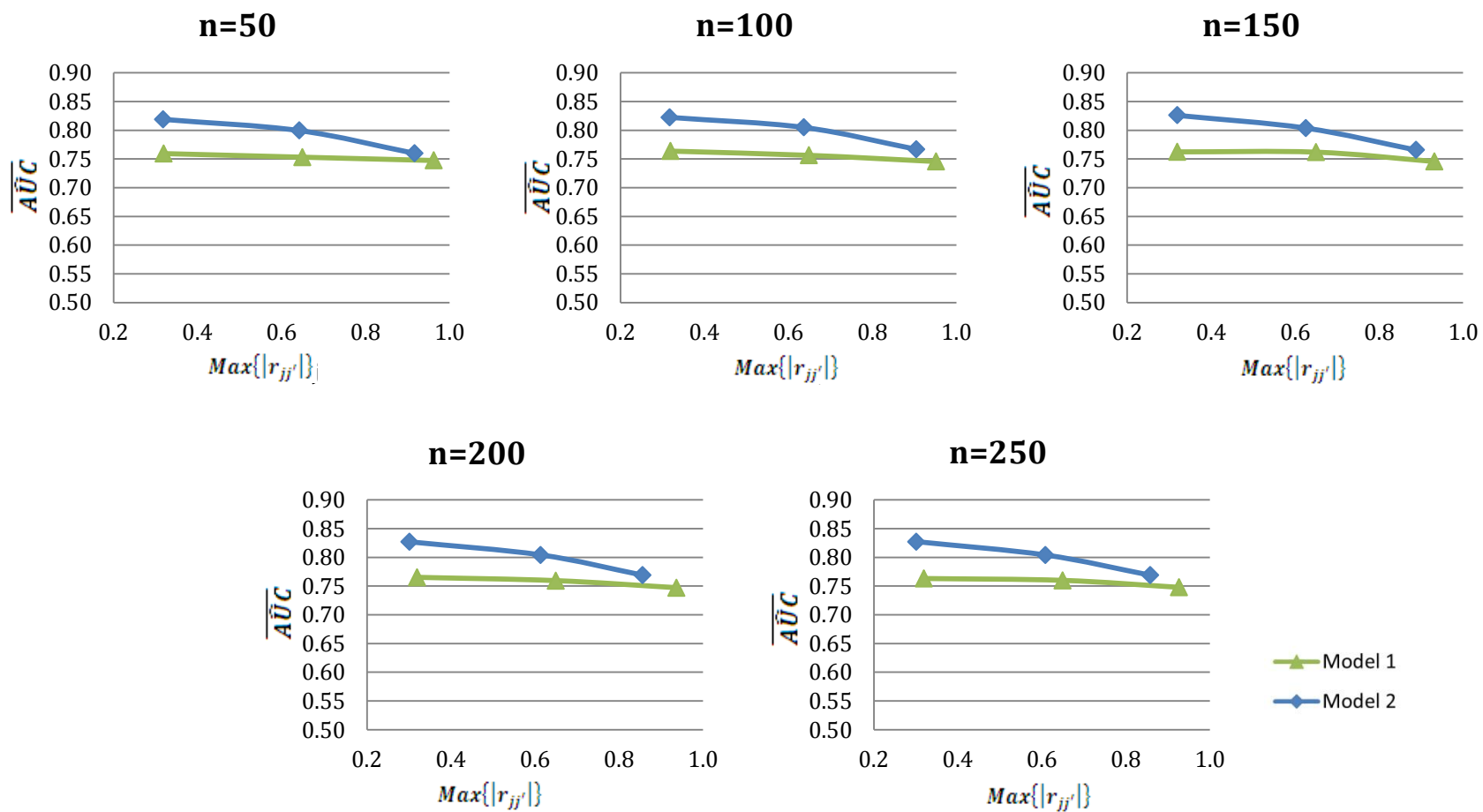
ตารางที่ 4.12 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 5)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{ r_{jj'} \}$	\overline{AUC}	$Max\{ r_{jj'} \}$	\overline{AUC}	
4	5	50	ต่ำ	0.31993	0.75941	0.31832	0.81896	Model 2
			ปานกลาง	0.64998	0.75311	0.64205	0.79946	Model 2
			สูง	0.96268	0.74736	0.91668	0.75977	Model 2
		100	ต่ำ	0.31996	0.76369	0.31741	0.82246	Model 2
			ปานกลาง	0.64929	0.75634	0.63718	0.80489	Model 2
			สูง	0.95169	0.74555	0.90429	0.76684	Model 2
		150	ต่ำ	0.31994	0.76241	0.31953	0.82599	Model 2
			ปานกลาง	0.64965	0.76189	0.6257	0.80376	Model 2
			สูง	0.93116	0.74562	0.88754	0.76564	Model 2
		200	ต่ำ	0.31996	0.76523	0.30128	0.82707	Model 2
			ปานกลาง	0.6498	0.7595	0.61366	0.80429	Model 2
			สูง	0.93691	0.74717	0.85587	0.76907	Model 2
		250	ต่ำ	0.31991	0.76318	0.30177	0.82736	Model 2
			ปานกลาง	0.64999	0.75992	0.609	0.80407	Model 2
			สูง	0.92686	0.74781	0.85814	0.76892	Model 2

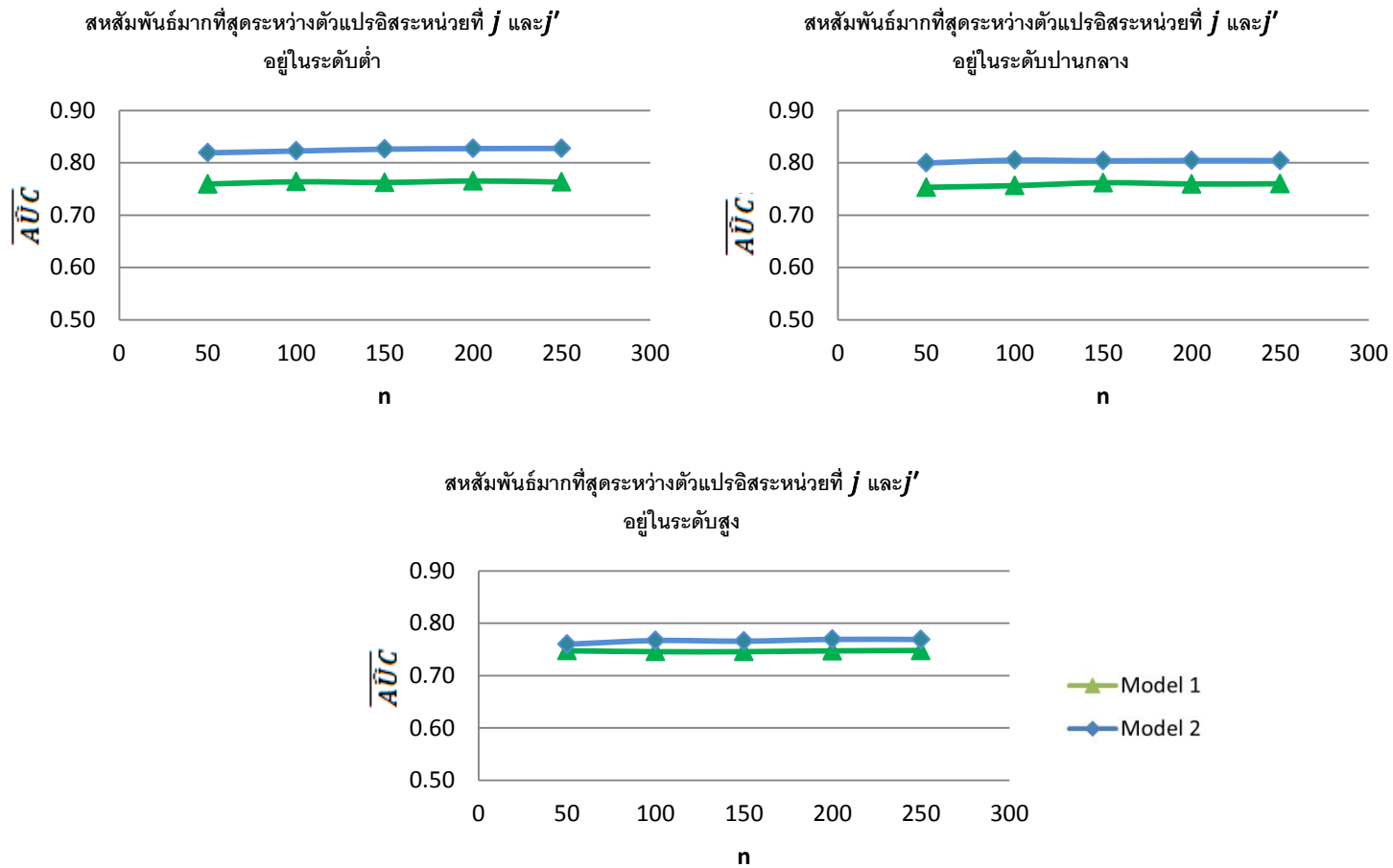
จากตารางที่ 4.12 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 5)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \geq \mu_2$ แยังกับ $H_1: \mu_1 < \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 2

ภาพที่ 4.23 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.24 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (4, 5)$, $\text{Max}\{|r_{jj'}|\}$. อยู่ใน ระดับ ต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



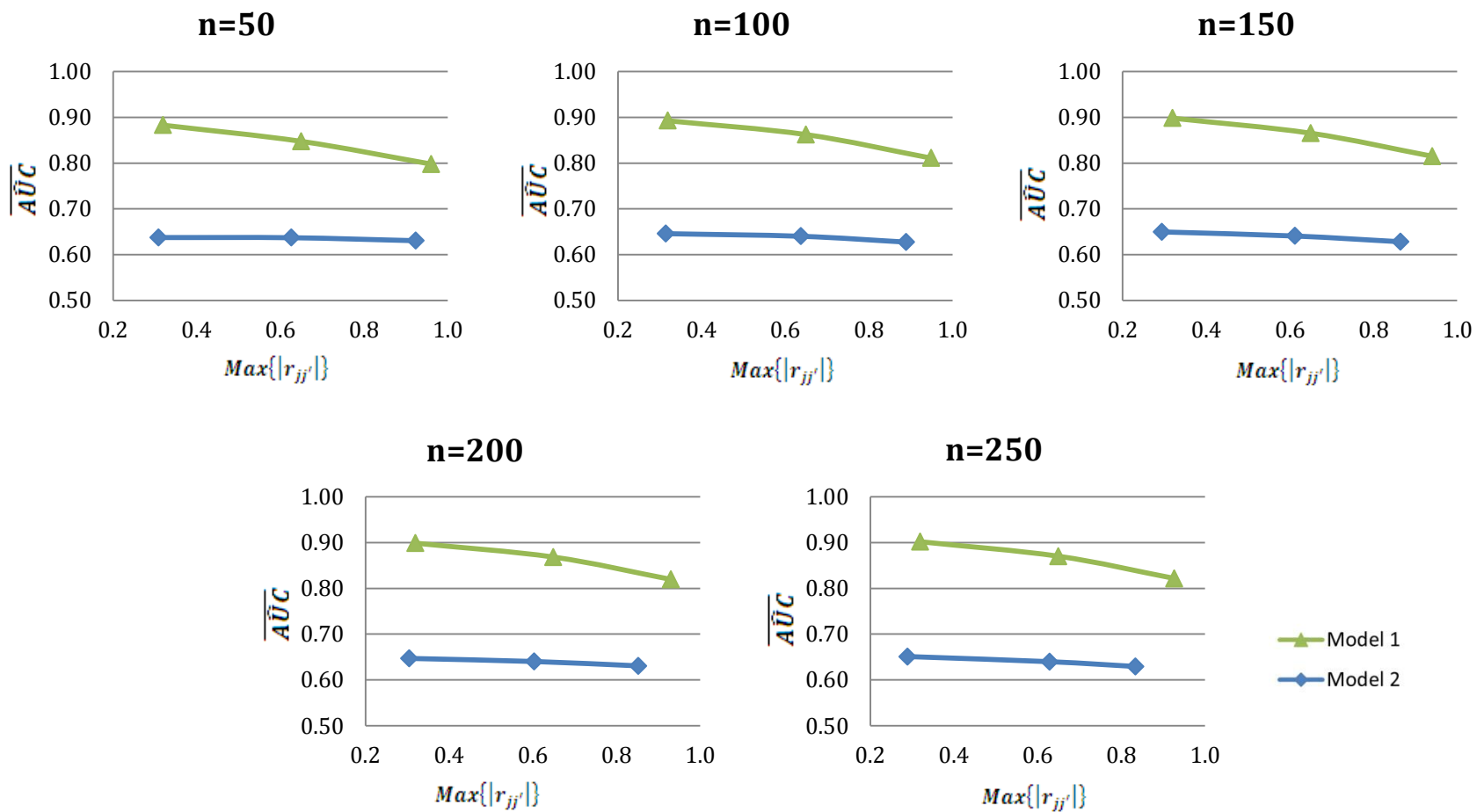
ตารางที่ 4.13 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 2)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
5	2	50	ต่ำ	0.31992	0.88297	0.30901	0.63730	ตัวแบบที่ 1
			ปานกลาง	0.64995	0.84755	0.62638	0.63690	ตัวแบบที่ 1
			สูง	0.96028	0.79793	0.92355	0.63038	ตัวแบบที่ 1
		100	ต่ำ	0.31962	0.89272	0.31513	0.64574	ตัวแบบที่ 1
			ปานกลาง	0.64968	0.86239	0.63747	0.64006	ตัวแบบที่ 1
			สูง	0.94873	0.81093	0.88933	0.62730	ตัวแบบที่ 1
		150	ต่ำ	0.31997	0.89831	0.29424	0.64958	ตัวแบบที่ 1
			ปานกลาง	0.64958	0.86546	0.61217	0.64065	ตัวแบบที่ 1
			สูง	0.93990	0.81510	0.86433	0.62811	ตัวแบบที่ 1
		200	ต่ำ	0.31988	0.89883	0.30491	0.64685	ตัวแบบที่ 1
			ปานกลาง	0.64926	0.86845	0.60363	0.64006	ตัวแบบที่ 1
			สูง	0.93042	0.81948	0.85185	0.63054	ตัวแบบที่ 1
		250	ต่ำ	0.31998	0.90194	0.28913	0.65094	ตัวแบบที่ 1
			ปานกลาง	0.64984	0.87020	0.62904	0.63967	ตัวแบบที่ 1
			สูง	0.92686	0.82167	0.83410	0.62919	ตัวแบบที่ 1

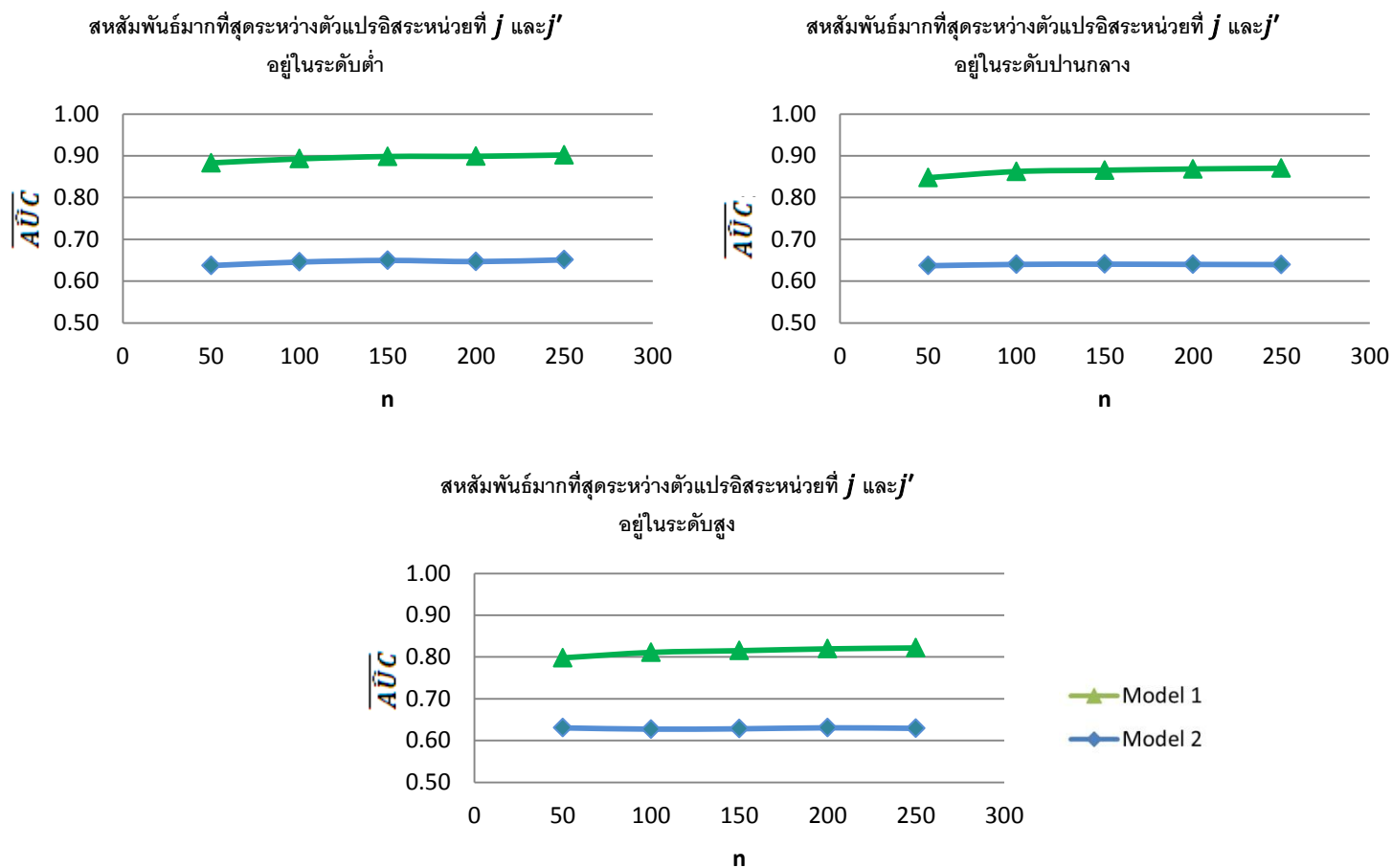
จากตารางที่ 4.13 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 2)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แยกกับ $H_1: \mu_1 > \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

ภาพที่ 4.25 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 2)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.26 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 2)$, $\text{Max}\{|r_{jj'}|\}$. อยู่ใน ระดับ ต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



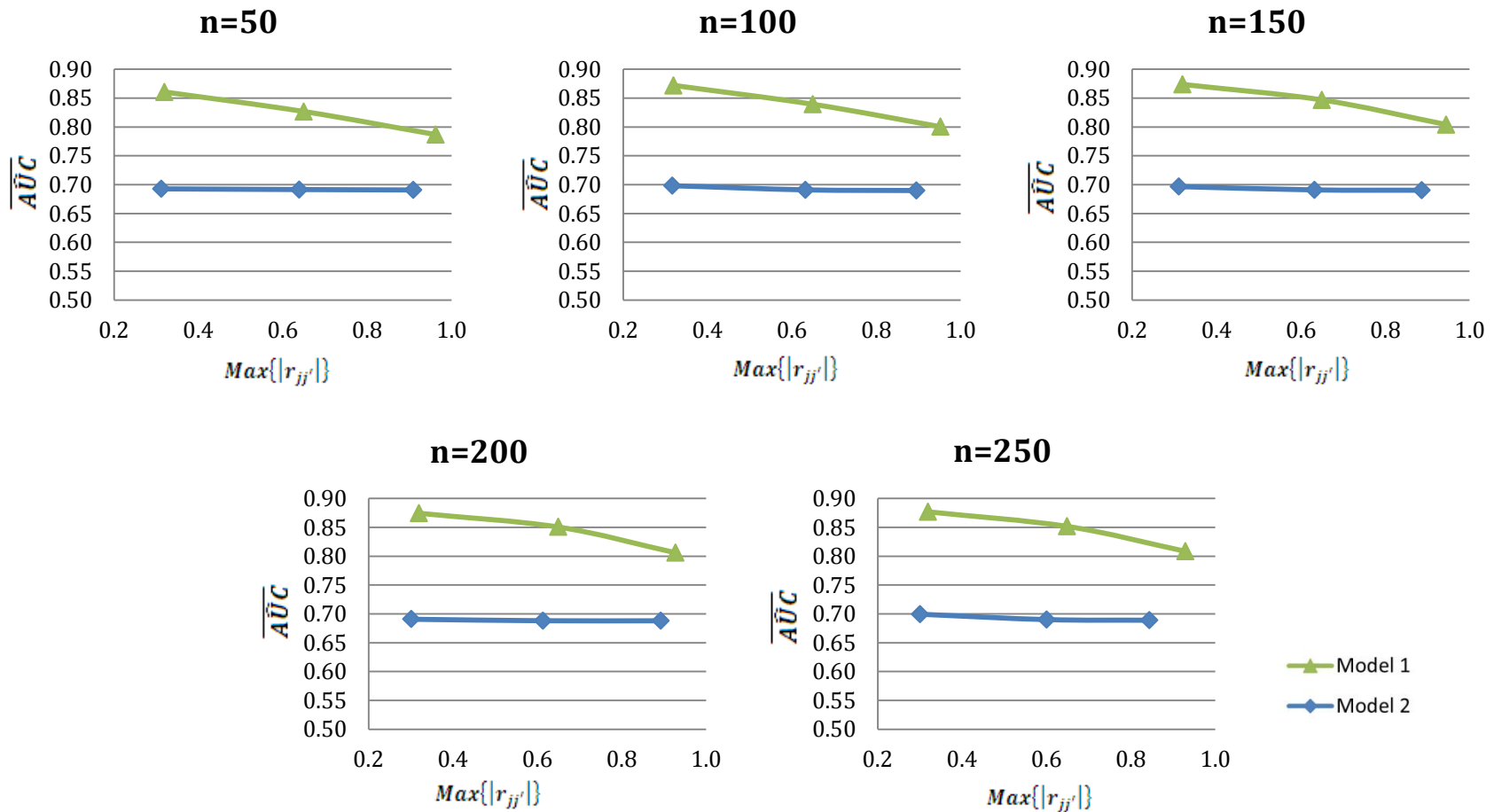
ตารางที่ 4.14 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 3)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
5	3	50	ต่ำ	0.31996	0.86071	0.31230	0.69271	ตัวแบบที่ 1
			ปานกลาง	0.64997	0.82680	0.63885	0.69151	ตัวแบบที่ 1
			สูง	0.96228	0.78684	0.90949	0.69076	ตัวแบบที่ 1
		100	ต่ำ	0.31986	0.87211	0.31691	0.69815	ตัวแบบที่ 1
			ปานกลาง	0.64986	0.83933	0.63238	0.69098	ตัวแบบที่ 1
			สูง	0.95214	0.80041	0.89538	0.68980	ตัวแบบที่ 1
		150	ต่ำ	0.31985	0.87383	0.31124	0.69658	ตัวแบบที่ 1
			ปานกลาง	0.64986	0.84699	0.63202	0.69090	ตัวแบบที่ 1
			สูง	0.94406	0.80426	0.88625	0.69032	ตัวแบบที่ 1
		200	ต่ำ	0.31998	0.87442	0.30150	0.69070	ตัวแบบที่ 1
			ปานกลาง	0.64988	0.85055	0.61358	0.68811	ตัวแบบที่ 1
			สูง	0.92755	0.80595	0.89226	0.68803	ตัวแบบที่ 1
		250	ต่ำ	0.31989	0.87692	0.30071	0.69936	ตัวแบบที่ 1
			ปานกลาง	0.64928	0.85165	0.60091	0.69006	ตัวแบบที่ 1
			สูง	0.92960	0.80839	0.84420	0.68902	ตัวแบบที่ 1

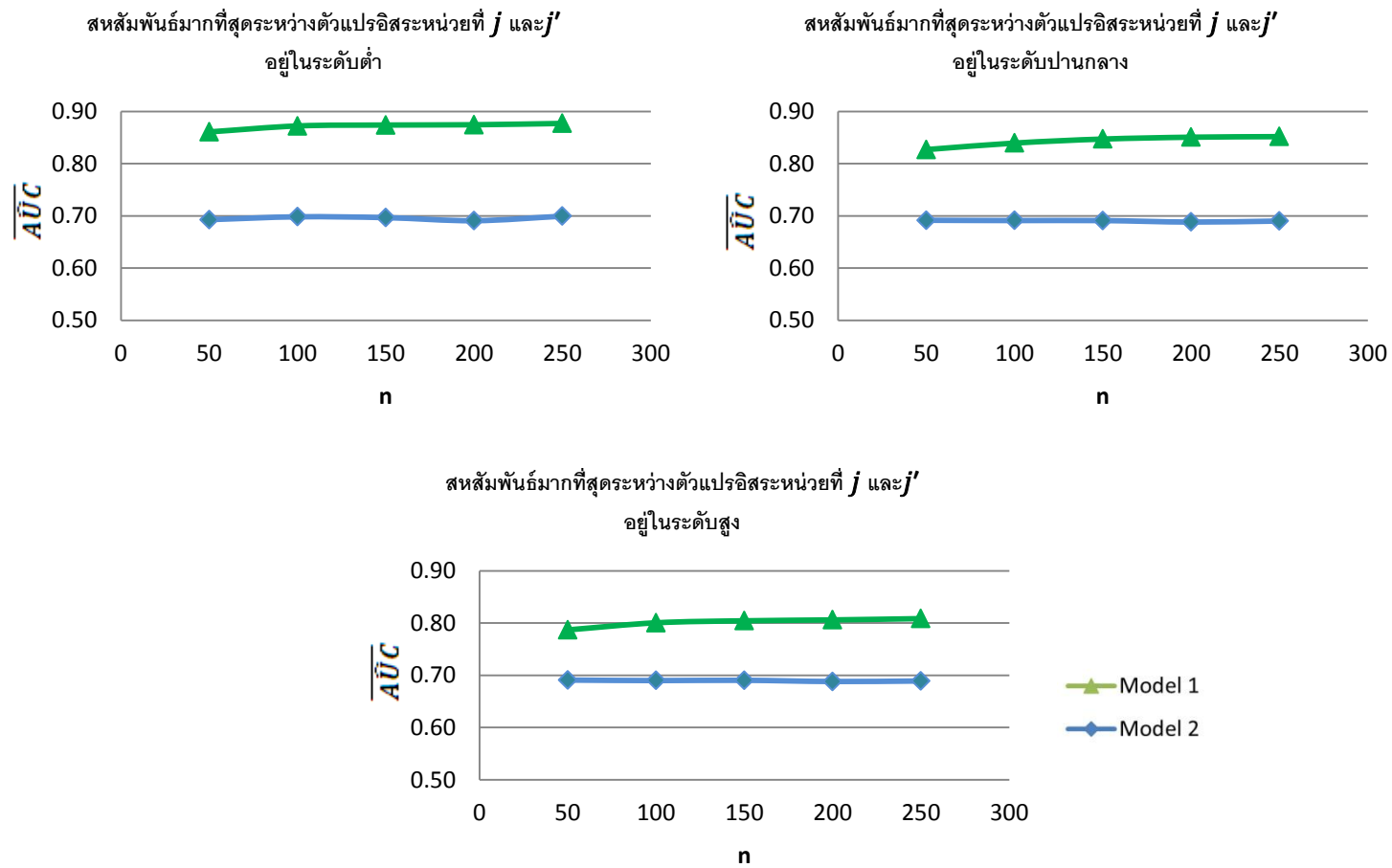
จากตารางที่ 4.14 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 3)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลงและเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แอ้งกับ $H_1: \mu_1 > \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

ภาพที่ 4.27 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 3)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.28 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 3)$, $\text{Max}\{|r_{jj'}|\}$. อยู่ใน ระดับ ต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



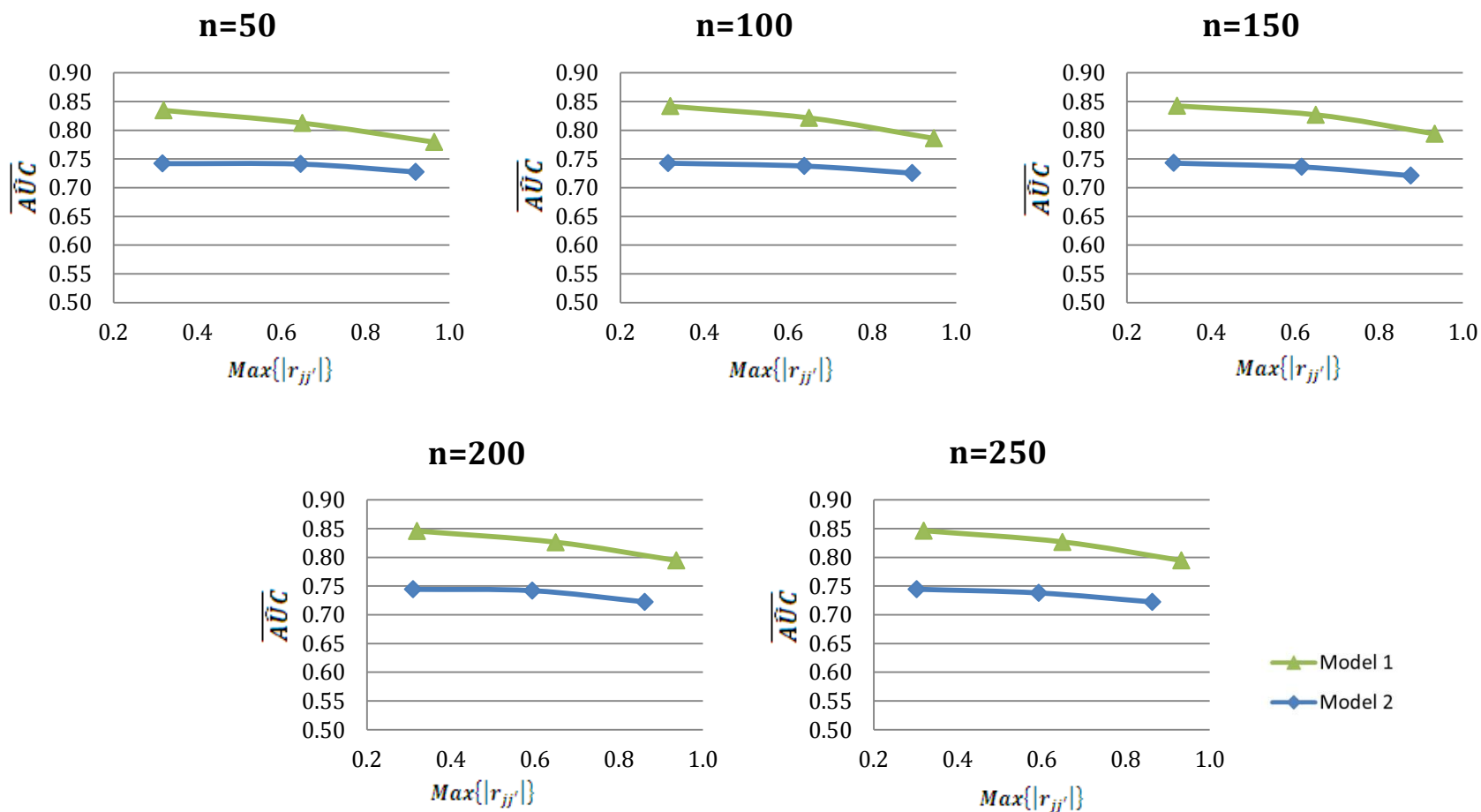
ตารางที่ 4.15 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 4)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{r_{jj'}\}$	\overline{AUC}	$Max\{r_{jj'}\}$	\overline{AUC}	
5	4	50	ต่ำ	0.31991	0.83462	0.31705	0.74192	ตัวแบบที่ 1
			ปานกลาง	0.64996	0.81246	0.64579	0.74107	ตัวแบบที่ 1
			สูง	0.96361	0.77933	0.91955	0.72734	ตัวแบบที่ 1
		100	ต่ำ	0.31993	0.84170	0.31432	0.74270	ตัวแบบที่ 1
			ปานกลาง	0.64990	0.82147	0.63844	0.73765	ตัวแบบที่ 1
			สูง	0.94671	0.78576	0.89529	0.72530	ตัวแบบที่ 1
		150	ต่ำ	0.31981	0.84237	0.31158	0.74281	ตัวแบบที่ 1
			ปานกลาง	0.64989	0.82692	0.61590	0.73616	ตัวแบบที่ 1
			สูง	0.93254	0.79372	0.87586	0.72091	ตัวแบบที่ 1
		200	ต่ำ	0.31990	0.84590	0.31055	0.74428	ตัวแบบที่ 1
			ปานกลาง	0.64993	0.82636	0.59379	0.74197	ตัวแบบที่ 1
			สูง	0.93691	0.79455	0.86172	0.72240	ตัวแบบที่ 1
250	ต่ำ	0.31998	0.84642	0.30275	0.74448	ตัวแบบที่ 1		
	ปานกลาง	0.64993	0.82687	0.59315	0.73801	ตัวแบบที่ 1		
	สูง	0.93268	0.79449	0.86361	0.72229	ตัวแบบที่ 1		

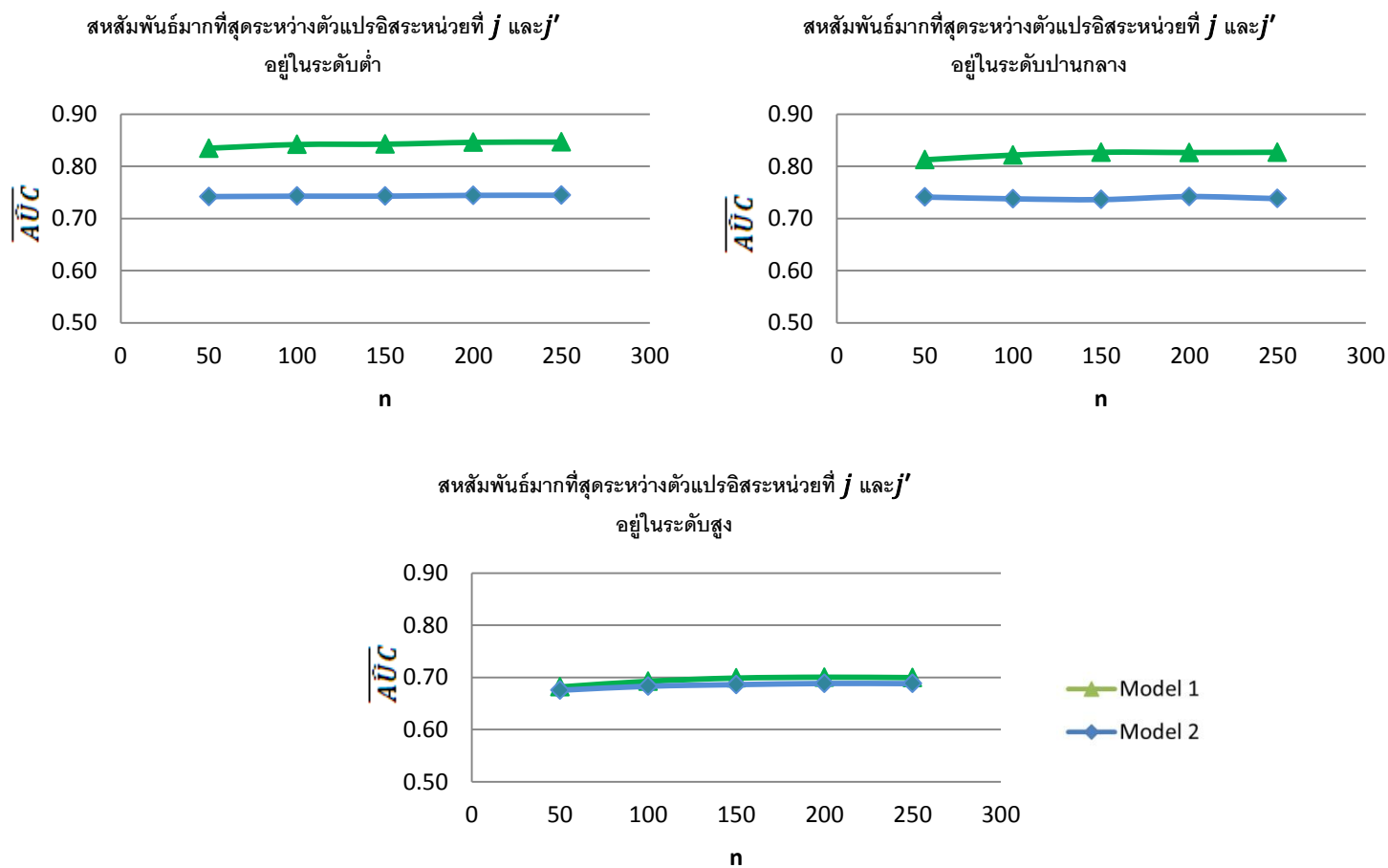
จากตารางที่ 4.15 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 4)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{r_{jj'}\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง และเมื่อทำการคัดเลือกตัวแบบจากการสมมติฐาน โดยต้องการทดสอบว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ $H_0: \mu_1 \leq \mu_2$ แยังกับ $H_1: \mu_1 > \mu_2$ ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 ในทุกกรณี (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

ภาพที่ 4.29 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 4)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับต่ำ, ปานกลาง และสูง



ภาพที่ 4.30 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 4)$, $\text{Max}\{|r_{jj'}|\}$. อยู่ในระดับ ต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



ตารางที่ 4.16 ค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 5)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

p_1	p_2	n	สหสัมพันธ์มากที่สุดอยู่ในระดับ	M_1		M_2		ตัวแบบที่ถูกเลือก
				$Max\{ r_{jj'} \}$	\overline{AUC}	$Max\{ r_{jj'} \}$	\overline{AUC}	
5	5	50	ต่ำ	0.31999	0.81261	0.31838	0.78964	ตัวแบบที่ 1
			ปานกลาง	0.64998	0.79222	0.64070	0.78334	ตัวแบบที่ 1
			สูง	0.96977	0.77043	0.92540	0.75358	ตัวแบบที่ 1
		100	ต่ำ	0.31992	0.81191	0.31719	0.79283	ตัวแบบที่ 1
			ปานกลาง	0.64993	0.80063	0.64102	0.77711	ตัวแบบที่ 1 หรือ ตัวแบบที่ 2
			สูง	0.94966	0.77790	0.88960	0.75677	ตัวแบบที่ 1
		150	ต่ำ	0.31999	0.82548	0.31412	0.79328	ตัวแบบที่ 1
			ปานกลาง	0.64977	0.81242	0.62389	0.78249	ตัวแบบที่ 1
			สูง	0.94401	0.80375	0.87290	0.77927	ตัวแบบที่ 1
		200	ต่ำ	0.31987	0.81203	0.30333	0.79285	ตัวแบบที่ 1
			ปานกลาง	0.64991	0.80257	0.64189	0.77734	ตัวแบบที่ 1
			สูง	0.92744	0.78229	0.86174	0.75262	ตัวแบบที่ 1
		250	ต่ำ	0.31954	0.81478	0.31011	0.79190	ตัวแบบที่ 1
			ปานกลาง	0.64991	0.80355	0.61292	0.78005	ตัวแบบที่ 1
			สูง	0.92608	0.78249	0.86534	0.75132	ตัวแบบที่ 1

จากตารางที่ 4.16 เมื่อพิจารณาค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 5)$, $n = 50, 100, 150, 200, 250$ เมื่อ $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง

พบว่า เมื่อระดับสหสัมพันธ์มากที่สุดระหว่างตัวแปรอิสระหน่วยที่ j และตัวแปรอิสระหน่วยที่ j' ; $j = 1, 2, \dots, p_1$ และ $j' = 1, 2, \dots, p_2$ เพิ่มมากขึ้น ค่าประมาณพื้นที่ใต้โค้ง ROC (\overline{AUC}) จะมีค่าลดลง

และเมื่อทำการคัดเลือกตัวแบบจากสมมติฐานว่าค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 หรือไม่ สมมติฐานคือ

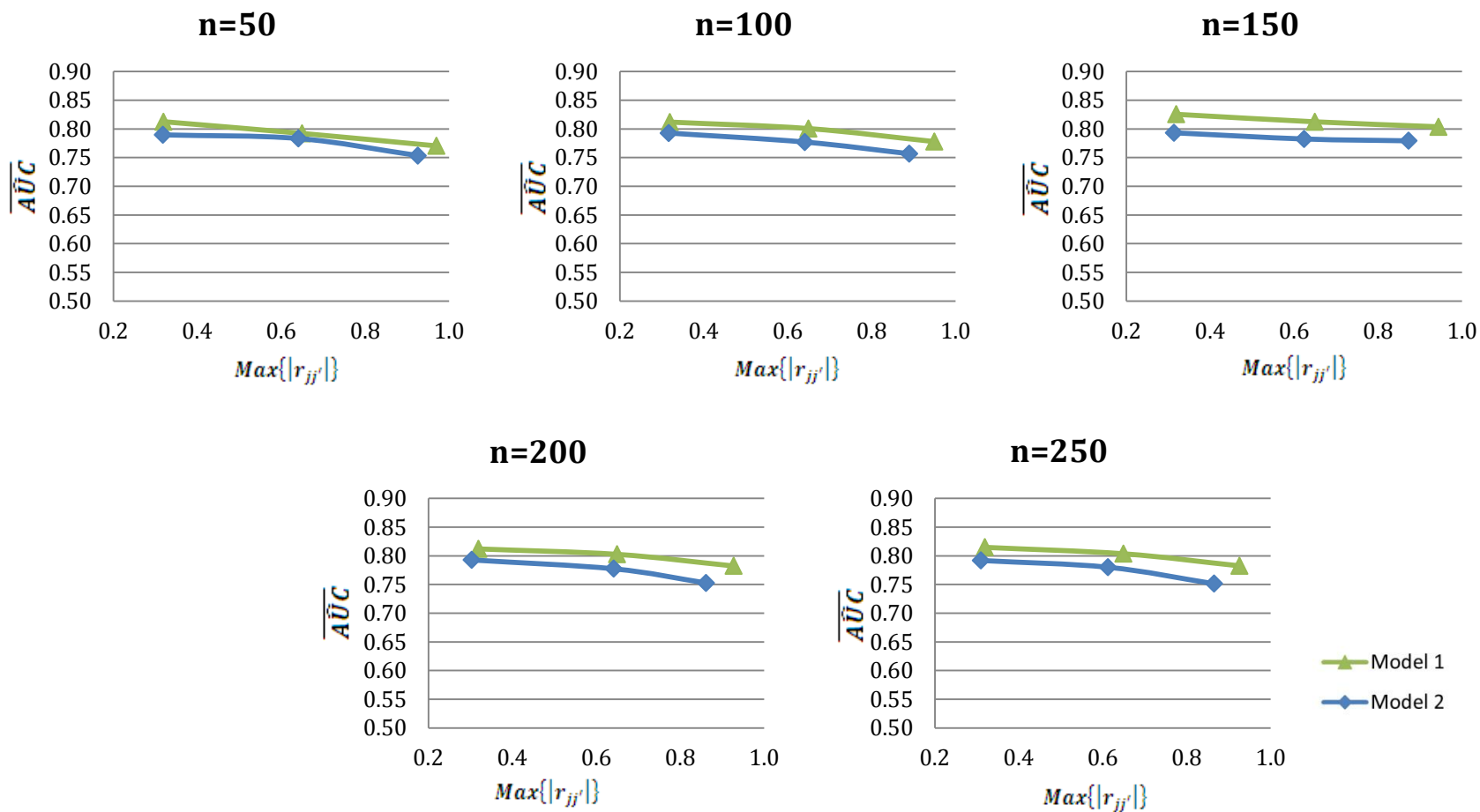
$$H_0: \mu_1 \leq \mu_2 \text{ แยกกับ } H_1: \mu_1 > \mu_2$$

ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 มากกว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 (P-Value < 0.05) ดังนั้นตัวแบบที่ถูกเลือกคือตัวแบบที่ 1

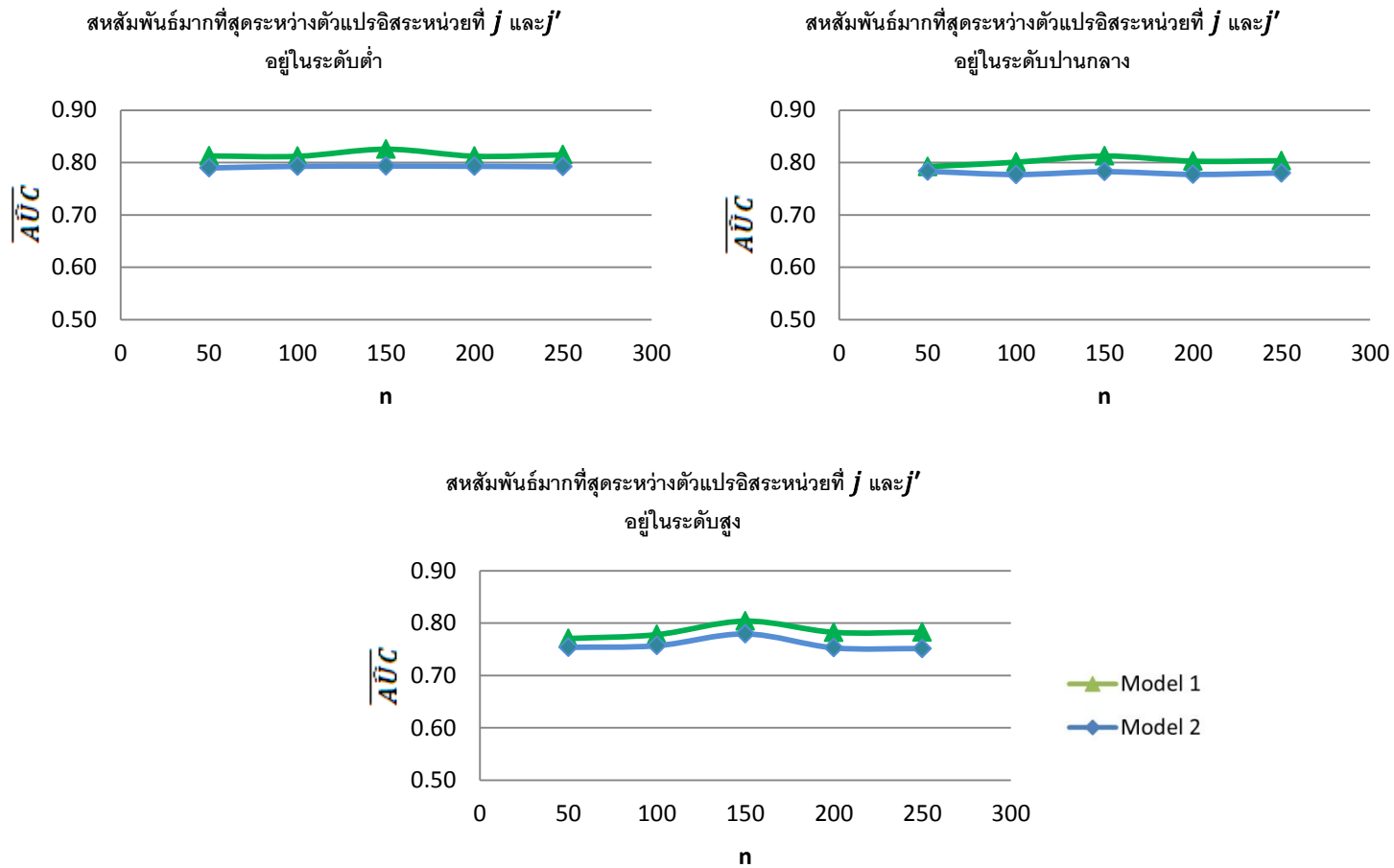
ยกเว้นกรณีที่ ขนาดตัวอย่างของเซตตัวแปรอิสระ $n = 100$ ที่ค่าสหสัมพันธ์มากที่สุดอยู่ในระดับปานกลาง

ถ้าทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$ ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 1 น้อยกว่าหรือเท่ากับ ค่าเฉลี่ยของค่าประมาณพื้นที่ใต้โค้ง ROC ตัวแบบที่ 2 (P-Value > 0.05) ดังนั้นตัวแบบที่ถูกเลือกจะเป็นตัวแบบที่ 1 หรือ ตัวแบบที่ 2 ก็ได้

ภาพที่ 4.31 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 5)$, $n = 50, 100, 150, 200, 250$ โดย $Max\{|r_{jj'}|\}$ เปลี่ยนแปลงไปอยู่ในระดับ ต่ำ, ปานกลาง และสูง



ภาพที่ 4.32 กราฟค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC เมื่อ $(p_1, p_2) = (5, 5)$, $Max\{|r_{jj'}|\}$. อยู่ใน ระดับต่ำ, ปานกลาง และสูง โดย n เปลี่ยนแปลงไปเท่ากับ 50, 100, 150, 200 และ 250



บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

จากจุดประสงค์การวิจัยที่ต้องการศึกษาอิทธิพลของปัจจัย ระดับสหสัมพันธ์ในแต่ละคู่ (Degree of pair-wise correlations) ของตัวแปรอิสระและจำนวนของตัวแปรอิสระ (p_1, p_2) ที่ส่งผลต่อการคัดเลือกตัวแบบนั้น ทำให้ผู้วิจัยสรุปได้ดังนี้

5.1.1 ปัจจัยที่ส่งผลต่อการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบถดถอยโลจิสติก ทวินามโดยใช้ฟังก์ชันคอมพลิเมนต์ารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง

ระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ

ระดับสหสัมพันธ์ของตัวแปรอิสระ ($Max\{|r_{ij}|\}$) โดยในการศึกษาครั้งนี้มีการแบ่งระดับของสหสัมพันธ์ออกเป็น 3 ระดับ คือ ระดับต่ำ กลางและสูง โดยระดับต่ำอยู่ในช่วง 0 ถึง 0.33 ระดับกลางอยู่ในช่วง 0.33 ถึง 0.66 และระดับสูงอยู่ในช่วง 0.66 ถึง 0.99 ซึ่งในการวิจัยครั้งนี้ใช้เกณฑ์การคัดเลือกตัวแบบ โดยพิจารณาจากค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีค่าสูงสุด ผลการศึกษาพบว่าเมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีค่าสหสัมพันธ์ในระดับต่ำ ค่าพื้นที่ใต้โค้ง ROC มีค่ามาก เมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีค่าสหสัมพันธ์ในระดับปานกลาง และระดับสูง ค่าพื้นที่ใต้โค้ง ROC มีค่าลดลง จึงสรุปได้ว่า ระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระมีค่าเพิ่มขึ้น ค่าพื้นที่ใต้โค้ง ROC จะมีค่าลดลง ซึ่งหมายถึงตัวแบบจะมีความน่าเชื่อถือลดลงด้วย

จำนวนตัวแปรอิสระของแต่ละตัวแบบ

จำนวนตัวแปรอิสระของแต่ละตัวแบบ ในการศึกษาครั้งนี้คือ (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) ผลการศึกษาพบว่าเมื่อจำนวนตัวแปรอิสระเพิ่มขึ้นจะทำให้ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC เพิ่มขึ้น ซึ่งหมายความว่า

ว่าตัวแบบจะมีความน่าเชื่อถือมากภายใต้ขนาดตัวอย่างและระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระคงที่

ขนาดตัวอย่าง

ขนาดตัวอย่างในการศึกษาครั้งนี้เท่ากับ 50, 100, 150, 200 และ 250 เกณฑ์การคัดเลือกตัวแบบพิจารณาจากพื้นที่ใต้โค้ง ROC มีค่าสูงสุด ผลการศึกษาพบว่าเมื่อขนาดตัวอย่างมีค่าเพิ่มมากขึ้น ค่าเฉลี่ยพื้นที่ใต้โค้ง ROC มีแนวโน้มเพิ่มสูงขึ้น ซึ่งหมายความว่า เมื่อขนาดตัวอย่างมีค่าเพิ่มมากขึ้น ตัวแบบจะมีความน่าเชื่อถือเพิ่มมากขึ้นภายใต้ระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระและจำนวนตัวแปรอิสระของแต่ละตัวแบบคงที่

ดังนั้นจึงสามารถสรุปได้ว่า ระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ ขนาดตัวอย่างและจำนวนตัวแปรอิสระเป็นปัจจัยที่มีผลต่อการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบ ถดถอยโลจิสติกทวินามโดยใช้ฟังก์ชันคอมพลีเมนทารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง

5.1.2 ผลการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ ของตัวแบบการถดถอยโลจิสติกแบบทวินามโดยใช้ฟังก์ชันคอมพลีเมนทารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง

กรณีจำนวนตัวแปรอิสระในตัวแบบแรกน้อยกว่าจำนวนตัวแปรอิสระในตัวแบบที่สอง ($p_1 < p_2$)

เมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 เท่ากับ ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 2 เนื่องจากตัวแบบที่ 2 มีค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 มากกว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 2 เนื่องจากตัวแบบที่ 2 มีค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC มากกว่า

กรณีจำนวนตัวแปรอิสระในตัวแบบแรกเท่ากับจำนวนตัวแปรอิสระในตัวแบบที่สอง ($p_1 = p_2$)

เมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 มากกว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC มากกว่า

กรณีจำนวนตัวแปรอิสระในตัวแบบแรกเท่ากับจำนวนตัวแปรอิสระในตัวแบบที่สอง ($p_1 > p_2$)

เมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 น้อยกว่า ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC มากกว่า

เมื่อระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 1 เท่ากับ ระดับความสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระในตัวแบบที่ 2 จะทำการคัดเลือกตัวแบบที่ 1 เนื่องจากตัวแบบที่ 1 มีค่าเฉลี่ยค่าประมาณพื้นที่ใต้โค้ง ROC มากกว่า

5.2 ข้อเสนอแนะ

ด้านการนำไปใช้ประโยชน์

เพื่อใช้แนวทางในการเลือกใช้ระดับสหสัมพันธ์ในแต่ละคู่ของตัวแปรอิสระ, ขนาดตัวอย่าง และจำนวนตัวแปรอิสระเพื่อให้ได้ตัวแบบที่ดีที่สุดจากการคัดเลือกตัวแบบ โลจิสติกทวินามแบบไม่ติดกลุ่ม สำหรับไปใช้ในการพยากรณ์ได้

ด้านการศึกษาวิจัย

1. ในการวิจัยครั้งนี้ได้ทำการศึกษาการคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกทวินาม โดยใช้ฟังก์ชันคอมพลีเมนทาลี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง โดยในการวิจัยนี้ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลที่ได้จากการจำลองขึ้นมา ในครั้งต่อไปผู้วิจัยอาจลองทำ

การวิจัยโดยใช้ข้อมูลที่ได้จากการสำรวจ ซึ่งลักษณะฟังก์ชันคอมพลิเมนทารี ล็อก-ล็อก เป็นลักษณะของข้อมูลที่พบเห็นมากโดยทั่วไปในปัจจุบัน

2. ในการวิจัยครั้งต่อไปอาจทำการศึกษาคัดเลือกตัวแบบไม่ติดกลุ่มอย่างสมบูรณ์ของตัวแบบการถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Model)

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา.การวิเคราะห์ข้อมูลหลายตัวแปร.พิมพ์ครั้งที่ 1 กรุงเทพมหานคร: สำนักพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย, 2548.

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

บุญจิรา มากอัน. การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยแบบไม่ติดกลุ่ม. วิทยานิพนธ์ ปริญญา มหาบัณฑิต, ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์ มหาวิทยาลัย, 2545.

วารุณี มูริกา. การเปรียบเทียบตัวแบบความถดถอยโลจิสติกแบบสถิติและแบบพลวัต. วิทยานิพนธ์ปริญญา มหาบัณฑิต, ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2551.

สุพล ศุรงค์วัฒนา.ตัวแบบและการวิเคราะห์ความถดถอยสำหรับการวิจัยขั้นสูง.กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2549.

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-Hall.

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

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศิริรัตน์ รสจันทร์ เกิดวันจันทร์ที่ 28 ตุลาคม พ.ศ. 2528 สำเร็จการศึกษา
หลักสูตรวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ประยุกต์ ภาควิชาคณิตศาสตร์และสถิติ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ในปีการศึกษา 2551 และเข้าศึกษาต่อ
ในหลักสูตรสถิติศาสตรมหาบัณฑิต (สถ.ม.) สาขาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการ
บัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปี พ.ศ. 2552