การวิเคราะห์ปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีสำหรับตัวแบบโลจิสติกพหุนาม

นางสาวพิมพลอย ภูววิเชียรฉาย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2555

HYPERVOLUME UNDER THE ROC MANIFOLD ANALYSIS FOR MULTINOMIAL LOGISTIC REGRESSION MODEL

Miss Pimploy Puwavicheanchai

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 2012

หัวข้อวิทยานิพนธ์	การวิเคราะห์ปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีสำหรับตัวแบบโลจิสติกพหุนาม
โดย	นางสาวพิมพลอย ภูววิเชียรฉาย
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

....คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

.... กรรมการ

(อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช)

.... กรรมการภายนอกมหาวิทยาลัย

(รองศาสตราจารย์ ดร. มนต์ทิพย์ เทียนสุวรรณ)

พิมพลอย ภูววิเชียรฉาย: การวิเคราะห์ปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีสำหรับตัวแบบโลจิสติกพหุนาม. (HYPERVOLUME UNDER THE ROC MANIFOLD ANALYSIS FOR MULTINOMIAL LOGISTIC REGRESSION MODEL) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. สุพล ดุรงค์วัฒนา, 83 หน้า.

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาผลกระทบที่มีต่อค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีสำหรับตัวแบบโลจิสติกพหุนาม กรณีที่มีจำนวนกลุ่มเท่ากับ 3, 4 และ 5 กลุ่ม ภายใต้เงื่อนไขว่า ตัวแปรอิสระ (X) ที่ทำการศึกษามีจำนวน 2, 3 และ 4 ตัว มาจากการแจกแจงแบบปกติที่ค่าเฉลี่ย (μ) เป็นค่าคงที่ และค่าความแปรปรวน (σ^2) เท่ากับ 1 มีค่าพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_0 = 0$ โดยงานวิจัยนี้ทำการศึกษาผลกระทบทั้งหมด 4 เรื่อง คือ เรื่องที่ 1 ศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์การถดถอย $\tilde{\beta}\ (\beta_1 = \cdots = \beta_p)$ ที่เปลี่ยนแปลงไป เรื่องที่ 2 ศึกษาผลกระทบจากค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ค่าระดับความสัมพันธ์เท่ากับ 0.5, 0.7 และ 0.9 เรื่องที่ 3 ศึกษาผลกระทบจากขนาดตัวอย่าง แบ่งเป็น 2 กรณี คือขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน และขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน เรื่องที่ 4 ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ยแต่ละกลุ่ม เมื่อความแปรปรวนของค่าเฉลี่ยแต่ละกลุ่มเปลี่ยนแปลงไป โดยทำการจำลองข้อมูลและวิเคราะห์ผลด้วยโปรแกรม R 2.15.0 และทดลองซ้ำในแต่ละกรณีจำนวน 1,000 รอบ

มีผลการศึกษาภายใต้ขอบเขตดังกล่าวซึ่งสรุปได้ดังนี้ เรื่องที่ 1 ค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้น ค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าสูงขึ้น เรื่องที่ 2 ระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น ค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าลดลง เรื่องที่ 3 กรณีขนาดตัวอย่างแต่ละกลุ่มมีจำนวนเท่ากัน เมื่อขนาดตัวอย่างแต่ละกลุ่มเพิ่มขึ้น ค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มเพิ่มขึ้น และกรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน เมื่อขนาดตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น จะมีค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีสูงขึ้น นอกจากนี้ยังพบว่า เมื่อเปรียบเทียบทั้ง 2 กรณีที่ขนาดตัวอย่างรวมมีจำนวนเท่ากัน ค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่ขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน จะมีค่าสูงกว่าค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่ขนาดตัวอย่างแต่ละกลุ่มเท่ากัน เรื่องที่ 4 เมื่อค่าความแปรปรวนของค่าเฉลี่ยเพิ่มขึ้น คือ ข้อมูลแต่ละกลุ่มมีส่วนที่ซ้อนทับกันน้อยลง ค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มเพิ่มขึ้น ซึ่งในกรณีของจำนวนกลุ่มเท่ากับ 3, 4 และ 5 กลุ่ม ให้ผลสอดคล้องไปในทิศทางเดียวกัน

ภาควิชา.........สถิติ.........ลายมือชื่อนิสิต..

สาขาวิชา.........สถิติ.........ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.................................

ปีการศึกษา.........2555.........

##5381850026: MAJOR STATISTICS

KEYWORDS: ROC ANALYSIS/ MULTICATEGORY/ HYPERVOLUME UNDER THE ROC

PIMPLOY PUWAVICHEANCHAI: HYPERVOLUME UNDER THE ROC MANIFOLD ANALYSIS FOR MULTINOMIAL LOGISTIC REGRESSION MODEL. ADVISOR: ASSOC.PROF.SUPOL DURONGWATTANA, Ph.D., 83 pp.

The objective of this research is to evaluate the effect of hypervolume under the ROC manifold for multinomial logistic regression model that dependent variable has 3, 4 and 5 groups. The experiment is simulated under the condition that 2, 3 and 4 independent variable x are generated normal distribution at constant mean (μ) and variance (σ^2) of the value 1. The regression coefficient β_0 is set to be equal to 0. The research is divided into 4 parts. Part 1: To evaluate the effect when the regression coefficient vector $\tilde{\beta}$ $(\beta_1 = \cdots = \beta_p)$ varies. Part 2: To appraise the effect of correlation between independent variable varies at 0.5, 0.7 and 0.9. Part 3: To evaluate the effect from the number of sample size n. This part is divided into 2 sub-parts, the equal sample size and the unequal sample size of each group. Part 4: To appraise the effect of the change of variance for each group mean. Simulation and analysis of generated data in each situation use R 2.15.0 and is repeated of 1,000 runs.

In conclusions, part 1: it is found that when the absolute of coefficient increase, the hypervolume under the ROC manifold value will increase. In part 2: if average correlation of independent variables increases, the hypervolume under the ROC manifold value will decrease. In part 3, in sub-parts 1(equal sample size), if sample size of each group increase the hypervolume under the ROC manifold value will increase. With unequal sample size, when the difference of sample size of each group increases, the hypervolume under the ROC manifold value will increase. Moreover, when we compare unequal with equal sample size that total sample size are the same, the hypervolume under the ROC manifold value of unequal sample size is higher than the hypervolume under the ROC manifold value of equal sample size. In part 4, when variance of each group increases, the hypervolume under the ROC manifold value will increase. All results are happened in the same regardless then member of groups for dependent variable.

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

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

Academic Year: …………2012………..…

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์ของบุคคลหลายท่าน ซึ่งไม่อาจจะนำมากล่าวได้ทั้งหมด ซึ่งผู้มีพระคุณท่านแรกที่ผู้วิจัยใคร่ขอกราบพระคุณคือ รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาสละเวลาให้ คำแนะนำตรวจทาน และแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่ทุกขั้นตอน เพื่อให้การค้นคว้าวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา และอาจารย์ ดร. อัครินทร์ ไพบูลย์พาณิชย์ ในฐานะประธานกรรมการและกรรมการสอบวิทยานิพนธ์ ที่กรุณาตรวจแก้ให้ข้อคิด และแนวทางที่ทำให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ให้โอกาสทางการศึกษา และประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ท้ายนี้ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา มารดา และเพื่อนๆ ที่อยู่เบื้องหลังในความสำเร็จที่ได้ให้ความช่วยเหลือสนับสนุนและให้กำลังใจแก่ผู้วิจัยเสมอมาจนกระทั่งสำเร็จการศึกษา

สารบัญ

หน้า

หน้า

สารบัญตาราง

ตารางที่ หน้า

ตารางที่ หน้า

สารบัญภาพ

ภาพที่ หน้า

ภาพที่ หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

Receiver Operating Characteristic (ROC) เป็นเครื่องวัดประสิทธิภาพของตัวแบบที่นิยมใช้กันมากวิธีหนึ่ง โดยการวิเคราะห์แบบใช้ในงานวิจัยที่มีผลของเหตุการณ์เกิดขึ้นได้สองเหตุการณ์ (dichotomous) แต่อย่างไรก็ตาม ในงานวิจัยหลายๆแขนง ไม่ว่าจะเป็นทางด้านการแพทย์ สังคมศาสตร์ เศรษฐศาสตร์ หรือธุรกิจต่างๆ มักจะเกิดผลของเหตุการณ์มากกว่าสองเหตุการณ์ขึ้นไป หรือแบ่งกลุ่มของข้อมูลที่นำมาวิจัยจำนวนมากกว่าสองกลุ่ม เช่น ในทางการแพทย์ เดิมอาจจะวิเคราะห์ความผิดปกติของโรคโดยแบ่งกลุ่มคนไข้เป็นสองกลุ่ม คือ กลุ่มที่ปรากฏโรค กับกลุ่มที่ไม่ปรากฏโรค ทั้งๆที่ในความเป็นจริงต้องการจะวินิจฉัยกลุ่มคนไข้ที่แตกต่างกันเป็นสามกลุ่ม คือ กลุ่มที่แสดงอาการอย่างรุนแรง กลุ่มที่แสดงอาการเล็กน้อย และกลุ่มที่ไม่มีอาการ ซึ่งจะช่วยให้แพทย์สามารถรักษาอาการของคนไข้ได้อย่างเหมาะสมกว่าแบ่งเป็นสองกลุ่ม เป็นต้น ในการใช้ Receiver Operating Characteristic ค่าที่ถูกใช้เป็นตัวแบบในการบอกความถูกต้องของการพยากรณ์คือพื้นที่ใต้โค้งอาร์โอซีซึ่งได้มาจากการพล็อตกราฟระหว่างค่า sensitivity และ 1-specificity จากจุดตัดต่างๆที่กำหนด โดยถ้าพื้นที่ใต้โค้งอาร์โอซีมีค่ามาก จะหมายถึงว่าตัวแบบนั้นมีอัตราความถูกต้องมาก และเป็นตัวแบบที่เหมาะสมจะนำไปใช้ในการพยากรณ์

DOUGLAS MOSSMAN (1999) ได้ทำการศึกษาถึงการใช้อาร์โอซีในกรณีที่แบ่งกลุ่มเหตุการณ์ที่เกิดขึ้นเป็นสามกลุ่ม ซึ่งผลการศึกษากล่าวว่า การใช้อาร์โอซีในการวิเคราะห์ความแม่นยำของการแบ่งกลุ่มในกรณีที่แบ่งกลุ่มของเหตุการณ์เป็นสามกลุ่มสามารถเป็นไปได้ ซึ่งสมมุติตัวอย่างอย่างง่ายว่า ให้มีรูปวงแหวนความละเอียดต่ำ โดยรูปดังกล่าวอาจเกิดจาก วงกลม (C), ห้าเหลี่ยม (P) หรือ สี่เหลี่ยม (S) และหาค่าส่วนผสมความน่าจะเป็นของรูปวงแหวนความละเอียดต่ำ คือ $(\mathrm{E_C}, \mathrm{E_P}, \mathrm{E_S})$ ซึ่งแสดงความน่าจะเป็นในการเป็นรูปต่างๆ โดยที่มีเงื่อนไขว่า $\mathrm{E_C}, \mathrm{E_P}, \mathrm{E_S} > 0$ และ $\mathrm{E_C} + \mathrm{E_P} + \mathrm{E_S} = 1$ และนำค่าส่วนผสมความน่าจะเป็นไปพล็อตกราฟอาร์โอซีบนพิกัดสามมิติ เพื่อหาปริมาตรที่เกิดขึ้นภายใต้พิกัดนั้น ซึ่งเป็นค่าความน่าจะเป็นของค่าการทดสอบที่จำแนกกลุ่มได้อย่างถูกต้อง โดยจะเรียกปริมาตรนั้นว่า ปริมาตรใต้ผิวโค้งอาร์โอซี (Volume Under the ROC Surface (VUS)) ในกรณีที่แบ่งกลุ่มเหตุการณ์ที่เกิดขึ้นเป็นสามกลุ่ม แทนพื้นที่ใต้โค้งอาร์โอซี (Area Under ROC Curve (AUS)) ในกรณีแบ่งกลุ่มเหตุการณ์ที่เกิดขึ้นเป็นสองกลุ่ม

XIN HE (2006) ได้ทำการศึกษาถึงการใช้อาร์โอซีในกรณีที่แบ่งกลุ่มเหตุการณ์ที่เกิดขึ้นเป็นสามกลุ่ม โดยใช้ทฤษฎีการตัดสินใจ (decision theory) ที่เขียนในรูปอรรถประโยชน์ ภายใต้สมมุติฐานว่า การพยากรณ์ผิดของแต่ละกลุ่มให้อรรถประโยชน์เท่ากัน และสร้างอัตราความควรจะเป็นเพื่อหาจุดที่ให้อรรถประโยชน์คาดหวังสูงสุด เพื่อไปพล็อตอาร์โอซี แล้วหาปริมาตรใต้ผิวโค้งอาร์โอซี (Volume Under the ROC Surface) และได้นำวิธีการนี้ไปประยุกต์ใช้กับการวัดค่ารังสีของผู้ป่วยโรคหัวใจ 3 กลุ่ม

THOMAS LANDGREBE AND ROBERT P.W. DUIN (2006) ขยายการวิเคราะห์อาร์โอซีจากกรณีที่มีจำนวนกลุ่มสองกลุ่มไปสู่กรณีที่มีจำนวนกลุ่มมากกว่าสองกลุ่ม โดยใช้ทฤษฎีเบย์ (Bayes theorem) ในการหาความน่าจะเป็นของการจำแนกกลุ่มแต่ละกลุ่ม และอธิบายการหาขอบเขตหรือเกณฑ์การตัดสินใจของค่าปริมาตรใต้ผิวโค้งอาร์โอซี ซึ่งมีค่าประมาณ $\frac{1}{k!}$ โดยที่ k คือจำนวนกลุ่มทั้งหมด

JIALIANG LI & JASON P. FINE (2008) ขยายการวิเคราะห์อาร์โอซีไปสู่กรณีที่เป็น multiple class หรือ แบ่งกลุ่มเหตุการณ์ที่เกิดขึ้นได้มากกว่าสองกลุ่ม โดยหาเวคเตอร์ความน่าจะเป็นของการจำแนกกลุ่มของแต่ละกลุ่ม $(E_1, E_2, \ldots, E_k)$ เพื่อนำไปคำนวณหาค่า Hypervolume Under the Manifold (HUM) ซึ่งเป็นค่าที่บ่งบอกประสิทธิภาพของตัวแบบในกรณีที่มีจำนวนกลุ่มมากกว่าหรือเท่ากับ 3 กลุ่ม และนำวิธีการดังกล่าวไปประยุกต์ใช้กับการทดสอบยีนที่ทำให้เกิดเนื้องอกที่มีมากกว่า 2 ประเภท

ในงานวิจัยนี้จะเลือกใช้วิธีการวิเคราะห์ข้อมูลเพื่อแบ่งกลุ่มของปัญหาต่างๆ ด้วยตัวแบบโลจิสติกพหุนาม (Multinomial Logistic Regression) ซึ่งเป็นตัวแบบที่สามารถแบ่งจำนวนกลุ่มของข้อมูลที่ได้ออกมามากกว่าสองกลุ่ม โดยผู้วิจัยจะนำตัวแบบที่ได้ไปพยากรณ์โอกาสในการเกิดเหตุการณ์ของแต่ละกลุ่ม แล้วจึงนำไปหาความแม่นยำของการพยากรณ์ของตัวแบบที่ได้ และเพื่อศึกษาหาความสัมพันธ์ของตัวแปรอิสระที่มีต่อโอกาสที่จะเกิดเหตุการณ์หรือตัวแปรตาม โดยที่ตัวแปรตามเป็นข้อมูลเชิงคุณภาพ ส่วนตัวแปรอิสระเป็นข้อมูลเชิงปริมาณ เนื่องจากสมการโล-จิสติกไม่ได้อยู่ในรูปเชิงเส้นทำให้ยากต่อการอธิบายความหมาย ดังนั้นจึงทำการตัวแบบให้อยู่ในรูปฟังก์ชันโลจิต (Logit Function) ซึ่งความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามอยู่ในรูปเชิงเส้นและง่ายต่อการวิเคราะห์มากกว่าฟังก์ชันโลจิสติก โดยในงานวิจัยนี้จะเลือกใช้ตัวแบบโลจิสติกแบบพหุนามที่มีตัวแปรอิสระจำนวนสอง, สาม และ สี่ ตัวแปร ที่มีระดับความสัมพันธ์ระหว่างตัวแปรอิสระตามที่กำหนดไว้ในขอบเขต เพื่อพิจารณาว่าค่าระดับความสัมพันธ์มีผลกระทบต่อค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีอย่างไร

โดยผู้วิจัยต้องการศึกษาผลกระทบจากปัจจัยต่างๆที่มีต่อค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี เมื่อขยายความคิดว่าสามารถใช้การวิเคราะห์อาร์โอซีในกรณีที่แบ่งกลุ่มเหตุการณ์ที่เกิดขึ้นได้มากกว่าสองเหตุการณ์ เพื่อนำไปประยุกต์ใช้กับการตัดสินใจเลือกตัวแบบที่เหมาะสมในกรณีที่มีเหตุการณ์เกิดขึ้นมากกว่าสองเหตุการณ์ โดยอาศัยค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (Hypervolume Under the ROC Manifold (HUM)) ในกรณีที่แบ่งกลุ่มเหตุการณ์ที่เกิดขึ้นได้มากกว่าหรือเท่ากับสามกลุ่ม

1.2 วัตถุประสงค์ของการศึกษา

เพื่อศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย, ค่าระดับความสัมพันธ์ของตัวแปรอิสระ, ขนาดตัวอย่าง และความแปรปรวนของค่าเฉลี่ย ที่มีต่อค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (Hypervolume Under the ROC Manifold (HUM)) ในกรณีที่มีตัวแปรตาม 3, 4 และ 5 กลุ่ม สำหรับตรวจสอบความถูกต้องของตัวแบบการวิเคราะห์ความถดถอยโลจิสติกแบบพหุนามที่ได้

1.3 ข้อตกลงเบื้องต้น

การศึกษาครั้งนี้มีข้อตกลงเบื้องต้นสำหรับการดำเนินงานวิจัย ดังนี้

1. ศึกษาตัวแบบการถดถอยโลจิสติกแบบพหุนาม (Multinomial Logistic Regression Model) ที่มีตัวแปรอิสระ 2, 3 และ 4 ตัว โดยมีรูปแบบ คือ

$$\begin{aligned} P(Y_i = j) &= \frac{e^{g_j(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} \\ &= \frac{e^{\beta_{j0}+\beta_{j1}X_{1i}+\beta_{j2}X_{2i}+\cdots+\beta_{jp}X_{pi}}}{1+e^{\beta_{20}+\beta_{21}X_{1i}+\cdots+\beta_{2p}X_{pi}}+\cdots+e^{\beta_{k0}+\beta_{k1}X_{1i}+\cdots+\beta_{kp}X_{pi}}} \\ &= Pr_j \qquad ; \ j = 1,2,\ldots,k \end{aligned}$$

โดยที่ Pr_j คือ ความน่าจะเป็นที่จะเป็นกลุ่มที่ j $(Y = j)$ เมื่อมีตัวแปรอิสระคือ $X_{1i}, \ldots, X_{pi}$

$\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp}$ คือ ค่าสัมประสิทธิ์การถดถอยโลจิสติก ของกลุ่มที่ j

$X_{1i}, \ldots, X_{pi}$ คือ ตัวแปรอิสระตัวที่ 1 ,…, p ของตัวอย่างหน่วยที่ i

n คือ ขนาดตัวอย่าง

และ กลุ่มที่ 1 เป็นกลุ่มฐาน

2. ศึกษาตัวแบบที่มีตัวแปรอิสระ 2, 3 และ 4 ตัว โดยเป็นข้อมูลเชิงปริมาณ ที่มีการแจกแจงแบบปกติ (Normal Distribution) นั่นคือ $x \sim N(\mu, \sigma^2)$ ซึ่งฟังก์ชันความน่าจะเป็นของ X อยู่ในรูปของ

$$f(x; \mu, \sigma^2) = \frac{2}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad ; X = (-\infty, \infty)\ , \mu = (-\infty, \infty), \sigma^2 = [0, \infty)$$

ซึ่งมีค่าเฉลี่ยและค่าความแปรปรวน คือ

$$E(x) = \mu$$
$$var(x) = \sigma^2$$

3. เรียกปริมาตรใต้ผิวโค้งอาร์โอซี ในกรณีที่เกิดเหตุการณ์ได้มากกว่า 2 เหตุการณ์ขึ้นไปรวมกันว่า Hypervolume Under the ROC Manifold (HUM) เนื่องจาก volume under ROC surface (VUS) จะใช้ในกรณีที่มีแบ่งเหตุการณ์ที่เกิดขึ้นได้ 3 กลุ่ม แต่ HUM สามารถใช้ได้กับกรณีที่แบ่งเหตุการณ์ที่เกิดขึ้นได้ตั้งแต่ 3 เหตุการณ์ขึ้นไป
4. $var(n) = \frac{\sum_{Y=1}^{k}(n_Y - \bar{n})^2}{k-1}$ ใช้สำหรับดูค่าการกระจายของขนาดตัวอย่าง ดังนี้
   - ค่า $var(n)$ มีค่าน้อย หมายถึง ขนาดตัวอย่างแต่ละกลุ่มมีจำนวนใกล้เคียงกัน
   - ค่า $var(n)$ มีค่ามาก หมายถึง ขนาดตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น

   โดยค่า $var(n) = [0, \infty)$ เมื่อค่า $var(n) = 0$ หมายถึง ขนาดตัวอย่างแต่ละกลุ่มมีจำนวนเท่ากัน ในการวิจัยครั้งนี้จะทำการศึกษาเฉพาะกรณีขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน $(var(n) = 0)$ และขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน $(var(n) \neq 0)$ โดยขนาดตัวอย่างจะเรียงจำนวนกลุ่มจากน้อยไปมากเท่านั้น
5. $var(\mu) = \frac{\sum_{Y=1}^{k}(\mu_Y - \bar{\mu})^2}{k-1}$ ใช้สำหรับดูค่าการกระจายของค่าเฉลี่ย ดังนี้
   - ค่า $var(\mu)$ มีค่าน้อย หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าใกล้เคียงกัน หรือตัวอย่างในแต่ละกลุ่มมีค่าใกล้เคียงกัน ข้อมูลแต่ละกลุ่มซ้อนทับกันมาก
   - ค่า $var(\mu)$ มีค่ามาก หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าแตกต่างกันมาก หรือตัวอย่างในแต่ละกลุ่มมีค่าแตกต่างกันมาก ข้อมูลแต่ละกลุ่มซ้อนทับกันน้อย

   โดยค่า $var(\mu) = [0, \infty)$ เมื่อค่า $var(\mu) = 0$ หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าเท่ากัน ในการวิจัยครั้งนี้จะทำการศึกษาเริ่มต้นที่ค่าเฉลี่ยของกลุ่มที่ 1 เท่ากับ 0 เสมอ และค่าเฉลี่ยของกลุ่มถัดไป มีค่าเพิ่มขึ้นเรื่อยๆ

6. ค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่าง (coefficient of variation: CV_n) คือ ค่าร้อยละที่ข้อมูลเบี่ยงเบนรอบๆค่าขนาดตัวอย่างเฉลี่ย ซึ่งใช้วัดการกระจายสัมพัทธ์ของข้อมูลเพื่อเปรียบเทียบการกระจายของข้อมูลตั้งแต่สองชุดขึ้นไป ซึ่งมีสูตร คือ

   $CV_n = \frac{SD}{\bar{n}} \times 100$ โดยที่

   - ค่า CV_n มีค่าน้อย หมายถึง ค่าการกระจายของข้อมูลรอบๆค่าขนาดตัวอย่างเฉลี่ยน้อย หรือตัวอย่างแต่ละกลุ่มมีจำนวนใกล้เคียงกัน
   - ค่า CV_n มีค่ามาก หมายถึง ค่าการกระจายของข้อมูลรอบๆค่าขนาดตัวอย่างเฉลี่ยมาก หรือ ตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น

1.4 ขอบเขตของการศึกษา

การวิจัยครั้งนี้มีขอบเขตของการวิจัยสำหรับการดำเนินวิจัย ดังนี้

1. ตัวแบบการถดถอยโลจิสติกพหุนาม มี 3 กรณี
   - กรณีมีตัวแปรอิสระ (X) 2 ตัว
   - กรณีมีตัวแปรอิสระ (X) 3 ตัว
   - กรณีมีตัวแปรอิสระ (X) 4 ตัว

   โดยตัวแปรอิสระเป็นตัวแปรเชิงปริมาณที่มีการแจกแจงเริ่มต้นมาจากการแจกแจงแบบปกติ ศึกษากรณี มีพารามิเตอร์ค่าเฉลี่ย คือ μ เป็นค่าคงที่ และมีค่าความแปรปรวน คือ σ^2 เท่ากับ 1 ตามระดับความสัมพันธ์ของตัวแปรอิสระที่กำหนดไว้ คือ 0.5, 0.7 และ 0.9
2. ตัวแปรตาม (Y) เป็นข้อมูลเชิงคุณภาพอยู่ในมาตรเรียงอันดับ (Ordinal Scale) ที่มีการแจกแจงแบบพหุนาม คือ มีค่าตัวแปรตามมากกว่า 2 ค่า หรือให้มี k ค่า คือ Y = 1, 2, … ,k โดยจะทำการศึกษาที่
   - กรณีจำนวนกลุ่มเท่ากับ 3 กลุ่ม คือ Y = 1, 2 และ 3 หรือ k=3
   - กรณีจำนวนกลุ่มเท่ากับ 4 กลุ่ม คือ Y = 1, 2, 3 และ 4 หรือ k=4
   - กรณีจำนวนกลุ่มเท่ากับ 5 กลุ่ม คือ Y = 1, 2, 3, 4 และ 5 หรือ k=5
3. ค่าความคลาดเคลื่อน (ε_i) มีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ μ และ σ โดยเงื่อนไขความคลาดเคลื่อนของตัวแบบโลจิสติก คือ $\mu = 0$ และ $\sigma^2 =$ ค่าคงที่ ดังนั้นในงานวิจัยจะทำการศึกษาที่ $\varepsilon_i \sim N(0,100)$
4. กำหนดค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น ในการจำลองข้อมูล ดังนี้

- กรณีมีตัวแปรอิสระ (X) 2 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ

  $\beta_1 = \beta_2 = -1,\quad \beta_1 = \beta_2 = -0.6,\quad \beta_1 = \beta_2 = -0.2,$
  $\beta_1 = \beta_2 = \ \ 1,\quad \beta_1 = \beta_2 = \ \ 0.6,\quad \beta_1 = \beta_2 = \ \ 0.2$
- กรณีมีตัวแปรอิสระ (X) 3 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ

  $\beta_1 = \beta_2 = \beta_3 = -1, \beta_1 = \beta_2 = \beta_3 = -0.6, \beta_1 = \beta_2 = \beta_3 = -0.2,$
  $\beta_1 = \beta_2 = \beta_3 = \ \ 1, \beta_1 = \beta_2 = \beta_3 = \ \ 0.6, \beta_1 = \beta_2 = \beta_3 = \ \ 0.2$
- กรณีมีตัวแปรอิสระ (X) 4 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ

  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = -1,\quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = -0.6,$
  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = -0.2,\quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1,$
  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \ \ 0.6,\quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = \ \ 0.2$

5. ศึกษาผลกระทบจากค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ กรณีมีตัวแปรอิสระ 2, 3และ 4 ตัว ขนาดตัวอย่าง $n_1 = 50, n_2 = 80, n_3 = 100, n_4 = 120$ และ $n_5 = 130$ ค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0, \beta_1 = \cdots = \beta_p = 1$ ดังนี้
   - กรณีมีตัวแปรอิสระ (X) 2 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_1, X_2) = 0.5, 0.7$ และ 0.9
   - กรณีมีตัวแปรอิสระ (X) 3 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_i, X_j) = 0.5, 0.7$ และ $0.9 ; i \neq j$ ให้ $\mathrm{corr}(X_1, X_2),\ \mathrm{corr}(X_1, X_3),\ \mathrm{corr}(X_2, X_3)$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดยทำการศึกษาภายใต้ระดับความสัมพันธ์ซึ่งมีทั้งหมด 24 วิธี จาก 27 วิธี ที่เป็น positive definite ซึ่งสามารถนำไปใช้ในการคำนวณได้ ดังนี้

     - 0.5, 0.5, 0.5
     - 0.5, 0.5, 0.7
     - 0.5, 0.5, 0.9
     - 0.5, 0.7, 0.5
     - 0.5, 0.7, 0.7
     - 0.5, 0.7, 0.9
     - 0.5, 0.9, 0.5
     - 0.5, 0.9, 0.7
     - 0.7, 0.5, 0.5
     - 0.7, 0.5, 0.7
     - 0.7, 0.5, 0.9
     - 0.7, 0.7, 0.5
     - 0.7, 0.7, 0.7
     - 0.7, 0.7, 0.9
     - 0.7, 0.9, 0.5
     - 0.7, 0.9, 0.7
     - 0.7, 0.9, 0.9
     - 0.9, 0.5, 0.5
     - 0.9, 0.5, 0.7
     - 0.9, 0.7, 0.5
     - 0.9, 0.7, 0.7
     - 0.9, 0.7, 0.9
     - 0.9, 0.9, 0.7
     - 0.9, 0.9, 0.9
   - กรณีมีตัวแปรอิสระ (X) 4 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_i, X_j) = 0.5, 0.7$ และ $0.9 ; i \neq j$ ให้ $\mathrm{corr}(X_1, X_2), \mathrm{corr}(X_1, X_3), \mathrm{corr}(X_1, X_4), \mathrm{corr}(X_2, X_3),$ $\mathrm{corr}(X_2, X_4), \mathrm{corr}(X_3, X_4)$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดย

ทำการศึกษาภายใต้ระดับความสัมพันธ์ซึ่งมีทั้งหมด 404 วิธี จาก 729 วิธี ที่เป็น positive definite ซึ่งสามารถนำไปใช้ในการคำนวณได้

7. จำนวนขนาดตัวอย่าง (n) จะให้มีอย่างน้อย 30 เท่าของจำนวนตัวแปรอิสระ โดยผู้วิจัยจะทำการศึกษาผลกระทบจากขนาดตัวอย่าง แบ่งเป็น 2 กรณี ได้แก่ กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน และกรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน โดยกำหนดค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0, \beta_1 = \cdots = \beta_p = 1$ และมีขนาดตัวอย่างของแต่ละกลุ่มดังนี้

- จำนวนกลุ่ม 3 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน

  - $n_1 = 30,\quad n_2 = 30,\quad n_3 = 30$
  - $n_1 = 50,\quad n_2 = 50,\quad n_3 = 50$
  - $n_1 = 80,\quad n_2 = 80,\quad n_3 = 80$
  - $n_1 = 100,\quad n_2 = 100,\quad n_3 = 100$
  - $n_1 = 120,\quad n_2 = 120,\quad n_3 = 120$

  กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน

  - $n_1 = 30,\quad n_2 = 40,\quad n_3 = 50$
  - $n_1 = 30,\quad n_2 = 50,\quad n_3 = 70$
  - $n_1 = 30,\quad n_2 = 60,\quad n_3 = 90$
  - $n_1 = 30,\quad n_2 = 70,\quad n_3 = 110$
  - $n_1 = 30,\quad n_2 = 80,\quad n_3 = 130$

- จำนวนกลุ่ม 4 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน

  - $n_1 = 30,\quad n_2 = 30,\quad n_3 = 30,\quad n_4 = 30$
  - $n_1 = 50,\quad n_2 = 50,\quad n_3 = 50,\quad n_4 = 50$
  - $n_1 = 80,\quad n_2 = 80,\quad n_3 = 80,\quad n_4 = 80$
  - $n_1 = 100,\quad n_2 = 100,\quad n_3 = 100,\quad n_4 = 100$
  - $n_1 = 120,\quad n_2 = 120,\quad n_3 = 120,\quad n_4 = 120$

  กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน

  - $n_1 = 30,\quad n_2 = 40,\quad n_3 = 50,\quad n_4 = 60$
  - $n_1 = 30,\quad n_2 = 50,\quad n_3 = 70,\quad n_4 = 90$
  - $n_1 = 30,\quad n_2 = 60,\quad n_3 = 90,\quad n_4 = 120$
  - $n_1 = 30,\quad n_2 = 70,\quad n_3 = 110,\quad n_4 = 150$
  - $n_1 = 30,\quad n_2 = 80,\quad n_3 = 130,\quad n_4 = 180$

- จำนวนกลุ่ม 5 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน

  - $n_1 = 30,\quad n_2 = 30,\quad n_3 = 30,\quad n_4 = 30,\quad n_5 = 30$
  - $n_1 = 50,\quad n_2 = 50,\quad n_3 = 50,\quad n_4 = 50,\quad n_5 = 50$
  - $n_1 = 80,\quad n_2 = 80,\quad n_3 = 80,\quad n_4 = 80,\quad n_5 = 80$
  - $n_1 = 100,\quad n_2 = 100,\quad n_3 = 100,\quad n_4 = 100,\quad n_5 = 100$
  - $n_1 = 120,\quad n_2 = 120,\quad n_3 = 120,\quad n_4 = 120,\quad n_5 = 120$

กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน

- $n_1 = 30$, $n_2 = 40$, $n_3 = 50$, $n_4 = 60$, $n_5 = 70$
- $n_1 = 30$, $n_2 = 50$, $n_3 = 70$, $n_4 = 90$, $n_5 = 110$
- $n_1 = 30$, $n_2 = 60$, $n_3 = 90$, $n_4 = 120$, $n_5 = 150$
- $n_1 = 30$, $n_2 = 70$, $n_3 = 110$, $n_4 = 150$, $n_5 = 190$
- $n_1 = 30$, $n_2 = 80$, $n_3 = 130$, $n_4 = 180$, $n_5 = 230$

8. ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย(var(μ)) หรือลักษณะการซ้อนทับกันของตัวอย่างในแต่ละกลุ่ม $Y \sim N(\mu, \sigma^2)$ มีตัวแปรอิสระ(X) 2, 3 และ 4 ตัว ค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0$, $\beta_1 = \cdots = \beta_p = 1$ ดังนี้

   - กรณีจำนวนกลุ่ม 3 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50$, $n_2 = 80$ และ $n_3 = 100$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้
     - $\mu_1 = 0$, $\mu_2 = 0$, $\mu_3 = 0$
     - $\mu_1 = 0$, $\mu_2 = 2$, $\mu_3 = 4$
     - $\mu_1 = 0$, $\mu_2 = 5$, $\mu_3 = 10$
     - $\mu_1 = 0$, $\mu_2 = 10$, $\mu_3 = 20$
   - กรณีจำนวนกลุ่ม 4 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50$, $n_2 = 80$, $n_3 = 100$ และ $n_4 = 120$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้
     - $\mu_1 = 0$, $\mu_2 = 0$, $\mu_3 = 0$, $\mu_4 = 0$
     - $\mu_1 = 0$, $\mu_2 = 2$, $\mu_3 = 4$, $\mu_4 = 6$
     - $\mu_1 = 0$, $\mu_2 = 5$, $\mu_3 = 10$, $\mu_4 = 15$
     - $\mu_1 = 0$, $\mu_2 = 10$, $\mu_3 = 20$, $\mu_4 = 30$
   - กรณีจำนวนกลุ่ม 5 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50$, $n_2 = 80$, $n_3 = 100$, $n_4 = 120$ และ $n_5 = 130$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้
     - $\mu_1 = 0$, $\mu_2 = 0$, $\mu_3 = 0$, $\mu_4 = 0$, $\mu_5 = 0$
     - $\mu_1 = 0$, $\mu_2 = 2$, $\mu_3 = 4$, $\mu_4 = 6$, $\mu_5 = 8$
     - $\mu_1 = 0$, $\mu_2 = 5$, $\mu_3 = 10$, $\mu_4 = 15$, $\mu_5 = 20$
     - $\mu_1 = 0$, $\mu_2 = 10$, $\mu_3 = 20$, $\mu_4 = 30$, $\mu_5 = 40$

9. การจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 1,000 รอบ (R=1,000)

1.5 คำจำกัดความที่ใช้ในการวิจัย

1. ตัวแบบการถดถอยโลจิสติกแบบพหุนาม (Multinomial Logistic Regression Model) หมายถึง ตัวแบบที่ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามซึ่งเป็นตัวแปรคุณภาพ กรณีที่มีจำนวนกลุ่มเท่ากับ 3, 4 และ 5 กลุ่ม ส่วนตัวแปรอิสระเป็นข้อมูลเชิงปริมาณ ซึ่งจะนำไปใช้ในการพยากรณ์โอกาสที่แต่ละหน่วยจะอยู่ในกลุ่มใดกลุ่มหนึ่ง

2. ฟังก์ชันโลจิต (logit function) หมายถึง ฟังก์ชันที่แปลงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามจากฟังก์ชันโลจิสติกให้อยู่ในรูปความสัมพันธ์เชิงเส้น ซึ่งเป็นการแปลงค่าความน่าจะเป็นจากช่วง (0,1) เป็นสมการโลจิตในช่วง $(-\infty, \infty)$
3. ค่าพื้นที่ใต้โค้งอาร์โอซี (Area Under ROC curve (AUS)) คือ ดัชนีในการบ่งชี้ความสามารถในการจำแนกกลุ่มหรือความเชื่อถือได้ของตัวแบบกรณีมีเหตุการณ์เกิดขึ้นได้ 2 เหตุการณ์
4. ค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (Hypervolume Under the ROC Manifold (HUM)) คือ ดัชนีในการบ่งชี้ความสามารถในการจำแนกกลุ่มหรือความเชื่อถือได้ของตัวแบบกรณีมีเหตุการณ์เกิดขึ้นได้มากกว่าหรือเท่ากับ 3 เหตุการณ์
5. $\mathrm{var(n)} = 0$ หมายถึง ขนาดตัวอย่างกลุ่มที่ i และ j เท่ากัน โดยที่ $\mathrm{i} \neq \mathrm{j}; \mathrm{i}, \mathrm{j} = 1, 2, \dots, \mathrm{k}$
   $\mathrm{var(n)} \neq 0$ หมายถึง ขนาดตัวอย่างกลุ่มที่ i และ j ไม่เท่ากัน โดยที่ $\mathrm{i} \neq \mathrm{j}; \mathrm{i}, \mathrm{j} = 1, 2, \dots, \mathrm{k}$
6. $\mathrm{var}(\mu)$ แสดงลักษณะการกระจายของค่าเฉลี่ยกลุ่มที่ i และ j
   โดยที่ $\mathrm{i} \neq \mathrm{j}; \mathrm{i}, \mathrm{j} = 1, 2, \dots, \mathrm{k}$

1.6 วิธีดำเนินการศึกษา

1. ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับงานวิจัย
2. กำหนดเงื่อนไขและขอบเขตของการวิจัย
   - กำหนดขนาดตัวอย่างแต่ละกลุ่ม $(\mathrm{n_i})$
   - กำหนดค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\tilde{\beta}$ $(\beta_0, \beta_1, \beta_2, \dots, \beta_p)$

   ตามกรณีที่ต้องการศึกษา
   - ศึกษาผลกระทบจากค่าพารามิเตอร์ β ณ ค่าต่างๆ
   - ศึกษาผลกระทบจากค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ
   - ศึกษาผลกระทบจากจำนวนขนาดตัวอย่าง
   - ศึกษาผลกระทบจากค่าความแปรปรวนของค่าเฉลี่ย
3. จำลองข้อมูลตามการแจกแจงและขอบเขตที่ต้องการศึกษา
   - กำหนดขนาดตัวอย่างแต่ละกลุ่ม
   - จำลองค่า X แต่ละกลุ่มตามการแจกแจงแบบปกติ โดยให้ $\mathrm{X_i}$ และ $\mathrm{X_j}$ มีความสัมพันธ์กันตามที่กำหนดขอบเขตไว้

- หาค่า $\tilde{Y}$ ของแต่ละกลุ่มตามที่กำหนดจำนวนไว้ จากค่าสัมประสิทธิ์ความถดถอยที่เริ่มต้น และค่า X ที่จำลองได้ของแต่ละกลุ่ม

4. หาค่าประมาณสัมประสิทธิ์การถดถอยของพารามิเตอร์ $\tilde{b}$ $(b_0, b_1, b_2, \dots, b_p)$ เพื่อสร้างตัวแบบสำหรับนำมาพยากรณ์
5. นำตัวแบบโลจิสติกที่ได้ไปพยากรณ์ข้อมูลย้อนกลับ เพื่อตรวจสอบผลการพยากรณ์เมื่อเทียบกับเหตุการณ์จริง
6. คำนวณหาค่าประมาณปริมาตรแบบไฮเปอร์ใต้ผิวโค้งอาร์โอซี (HUM)
7. ทำซ้ำในแต่ละการทดสอบจำนวน 1,000 รอบ เพื่อหาค่าเฉลี่ยของค่าปริมาตรแบบไฮเปอร์ใต้ผิวโค้งอาร์โอซี
8. เปรียบเทียบค่าปริมาตรแบบไฮเปอร์ใต้ผิวโค้งอาร์โอซีที่ได้ และสรุปผลที่ได้จากการวิจัย

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการใช้ค่าปริมาตรแบบไฮเปอร์ใต้ผิวโค้งอาร์โอซี สำหรับตัวแบบการถดถอยโลจิสติกที่มีตัวแปรตามมากกว่าหรือเท่ากับ 3 กลุ่ม
2. เพื่อใช้ในการวัดประสิทธิภาพของตัวแบบการวิเคราะห์ความถดถอยโลจิสติกแบบพหุนามโดยใช้ค่าปริมาตรแบบไฮเปอร์ใต้ผิวโค้งอาร์โอซี (Hypervolume Under the Manifold (HUM)) ในกรณีที่มีตัวแปรตามมากกว่าหรือเท่ากับ 3 กลุ่ม

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

2.1 การถดถอยโลจิสติกพหุนาม (Multinomial Logistic Regression)

เมื่อตัวแปรตาม (Y) เป็นตัวแปรเชิงกลุ่มที่มีค่ามากกว่า 2 ค่า คือ $Y_i = 1, 2, \ldots, k$ และตัวแปรอิสระ (X) อาจเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงกลุ่มก็ได้ โดยมีตัวแบบการถดถอยโลจิสติค ดังนี้

$$P(Y_i = j|X) = \frac{e^{g_j(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} = Pr_j \qquad ; j = 1, 2, \ldots, k \quad \ldots\ldots\ldots(1)$$

$$P(Y_i = 1|X) = \frac{e^{g_1(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} = \frac{e^{g_1(X)}}{1+e^{g_2(X)}+e^{g_3(X)}+\cdots+e^{g_k(X)}} = Pr_1$$

$$P(Y_i = 2|X) = \frac{e^{g_2(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} = \frac{e^{g_2(X)}}{1+e^{g_2(X)}+e^{g_3(X)}+\cdots+e^{g_k(X)}} = Pr_2$$

$$\vdots$$

$$P(Y_i = k|X) = \frac{e^{g_k(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} = \frac{e^{g_k(X)}}{1+e^{g_2(X)}+e^{g_3(X)}+\cdots+e^{g_k(X)}} = Pr_k$$

โดยที่ $\sum_{j=1}^{k} Pr = 1$

$$g_j(X) \quad = \beta_{j0} + \beta_{j1}X_{1i} + \beta_{j2}X_{2i} + \cdots + \beta_{jp}X_{pi} + \tilde{\varepsilon}_i$$

โดยที่ $Y_i = 1 = g_1(X) = \beta_{10} + \beta_{11}X_{1i} + \beta_{12}X_{2i} + \cdots + \beta_{1p}X_{pi} + \varepsilon_i = X'\tilde{\beta}_1 + \tilde{\varepsilon}_i$

$$Y_i = 2 = g_2(X) = \beta_{20} + \beta_{21}X_{1i} + \beta_{22}X_{2i} + \cdots + \beta_{2p}X_{pi} + \varepsilon_i = X'\tilde{\beta}_2 + \tilde{\varepsilon}_i$$

$$\vdots$$

$$Y_i = k = g_k(X) = \beta_{k0} + \beta_{k1}X_{1i} + \beta_{k2}X_{2i} + \cdots + \beta_{kp}X_{pi} + \varepsilon_i = X'\tilde{\beta}_k + \tilde{\varepsilon}_i$$

จากฟังก์ชันโลจิสติค (Logistic Response Function) หรือสมการที่ (1) จะพบว่าความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระไม่ได้อยู่ในรูปเชิงเส้น ดังนั้นจึงได้ทำการปรับสมการให้อยู่ในรูปเชิงเส้นที่เรียกว่าฟังก์ชันโลจิต (Logit Function) โดยเริ่มจากการหาอัตราส่วนโอกาสในการเกิดเหตุการณ์ที่ j โดยเทียบกับกลุ่มที่เป็นฐานคือ $Y_i = 1$

$$\text{Odds Ratio}_{j^*} = \frac{P(Y_i = j^*)}{P(Y_i = 1)} \qquad ; j^* = 2, \ldots, k$$

$$= \frac{e^{g_{j^*}(X)} \Big/ 1 + e^{g_2(X)} + e^{g_3(X)} + \cdots + e^{g_k(X)}}{1 \Big/ 1 + e^{g_2(X)} + e^{g_3(X)} + \cdots + e^{g_k(X)}}$$

$$\text{Odds Ratio}_{j^*} = e^{g_{j^*}(X)} \qquad \ldots\ldots\ldots\ldots\ldots(2)$$

โดยที่ Odds Ratio จะแสดงถึงโอกาสที่จะเกิดเหตุการณ์ที่ j* เป็นกี่เท่าของโอกาสที่จะเกิดเหตุการณ์ที่ $Y = 1$ และ Odds Ratio ของตัวแปรเชิงกลุ่มสองค่า จะต้องเป็นอิสระกับค่าอื่นๆ

จากนั้นทำสมการที่ (2) ให้อยู่ในรูปเชิงเส้น โดย

$$\begin{aligned} \ln \text{Odds Ratio}_{j^*} &= \ln e^{g_{j^*}(X)} \qquad ; j^* = 1, 2, \dots, k \\ &= g_{j^*}(X) \\ &= b_{j^*0} + b_{j^*1}X_{1i} + b_{j^*2}X_{2i} + \cdots + b_{j^*p}X_{pi} \qquad \dots\dots\dots\dots\dots (3) \end{aligned}$$

จะได้สมการที่ (3) เป็นฟังก์ชันโลจิต (Logit Function) ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามในรูปเชิงเส้น จากจำนวนกลุ่ม k กลุ่ม เราจะได้ฟังก์ชันโลจิตจำนวน k-1 สมการ นั่นคือ

$$\ln \text{Odds Ratio}_2 = \ln\left(\frac{P(Y_i=2)}{P(Y_i=1)}\right) = g_2(X) = b_{20} + b_{21}X_{1i} + b_{22}X_{2i} + \cdots + b_{2p}X_{pi}$$

$$\ln \text{Odds Ratio}_3 = \ln\left(\frac{P(Y_i=3)}{P(Y_i=1)}\right) = g_3(X) = b_{30} + b_{31}X_{1i} + b_{32}X_{2i} + \cdots + b_{3p}X_{pi}$$

$$\vdots$$

$$\ln \text{Odds Ratio}_k = \ln\left(\frac{P(Y_i=k)}{P(Y_i=1)}\right) = g_k(X) = b_{k0} + b_{k1}X_{1i} + b_{k2}X_{2i} + \cdots + b_{kp}X_{pi}$$

2.2 **ฟังก์ชันความควรจะเป็น** (Likelihood function)

การประมาณค่าสัมประสิทธิ์ความถดถอยโลจิสติค คือ การประมาณค่า $\beta_0, \beta_1, \dots, \beta_p$ โดยที่ตัวอย่างทั้ง n ตัวอย่างต้องเป็นอิสระกัน ฟังก์ชันความควรจะเป็น คือ

$$L = \prod_{i=1}^{n} P(Y_i = j)^{Y_i} = \prod_{i=1}^{n} (Pr_{Y_1})^{Y_1}(Pr_{Y_2})^{Y_2} \dots (Pr_{Y_n})^{Y_n} \qquad \dots\dots\dots\dots\dots (4)$$

$$; j = 1, 2, \dots, k$$

$$Pr_1 = \frac{1}{1 + e^{g_2(X)} + e^{g_3(X)} + \cdots + e^{g_k(X)}}$$

$$Pr_2 = \frac{e^{g_1(X)}}{1 + e^{g_2(X)} + e^{g_3(X)} + \cdots + e^{g_k(X)}}$$

$$\vdots$$

$$Pr_k = \frac{e^{g_k(X)}}{1 + e^{g_2(X)} + e^{g_3(X)} + \cdots + e^{g_k(X)}}$$

หาค่า L ของสมการที่ (4) จะได้ $\ln L$ ซึ่งเรียกว่า log-likelihood function

$$\ln L = \sum_{i=1}^{n} P(Y_i = j)^{Y_i}$$

$$= \sum_{i=1}^{n} \left(Y_1 \ln(\Pr_{Y_1}) + Y_2 \ln(\Pr_{Y_2}) + \cdots + Y_k \ln(\Pr_{Y_n}) \right) \qquad \ldots\ldots\ldots\ldots\ldots (5)$$

โดยการประมาณค่าสัมประสิทธิ์ $\beta_0, \beta_1, \ldots, \beta_p$ จะใช้หลักการความควรจะเป็นสูงสุด (Maximum Likelihood) หรือ ประมาณค่าสัมประสิทธิ์ $\beta_0, \beta_1, \ldots, \beta_p$ ที่ทำให้ $\ln L$ ในสมการที่ (5) มีค่ามากสุด โดยการหาอนุพันธ์ลำดับที่ 1 ของสมการที่ (4) เทียบกับ $\beta_0, \beta_1, \ldots, \beta_p$ แล้วให้เท่ากับศูนย์ แต่เนื่องจากสมการไม่ได้อยู่ในรูปเชิงเส้น จึงใช้เทคนิคการทำซ้ำของนิวตัน-ราฟสัน (Newton-Raphson method) ในการประมาณค่า $\beta_0, \beta_1, \ldots, \beta_p$ ด้วย $b_0, b_1, \ldots, b_p$

เมื่อประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดในตัวแบบการถดถอยโลจิสติกแบบพหุนามแล้วจะสามารถนำไปใช้ในการพยากรณ์การจำแนกกลุ่มของตัวแบบ ดังนี้

- หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มที่ 1 $(Y_i = 1)$ ถ้า

$$P(Y_i) = \frac{e^{g_j(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} \le c_1 = \frac{e^{g_1(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} \qquad \text{เมื่อ } 0 < c_1 < c_2$$

- หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มที่ 2 $(Y_i = 2)$ ถ้า

$$c_1 = \frac{e^{g_1(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} < P(Y_i) = \frac{e^{g_j(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} \le c_2 = \frac{e^{g_2(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} \qquad \text{เมื่อ } c_1 < c_2 < c_3$$

- หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มที่ 3 $(Y_i = 3)$ ถ้า

$$c_2 = \frac{e^{g_2(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} < P(Y_i) = \frac{e^{g_j(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} \le c_3 = \frac{e^{g_3(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} \qquad \text{เมื่อ } c_2 < c_3 < c_4$$

$$\vdots$$

- หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มที่ k $(Y_i = k)$ ถ้า

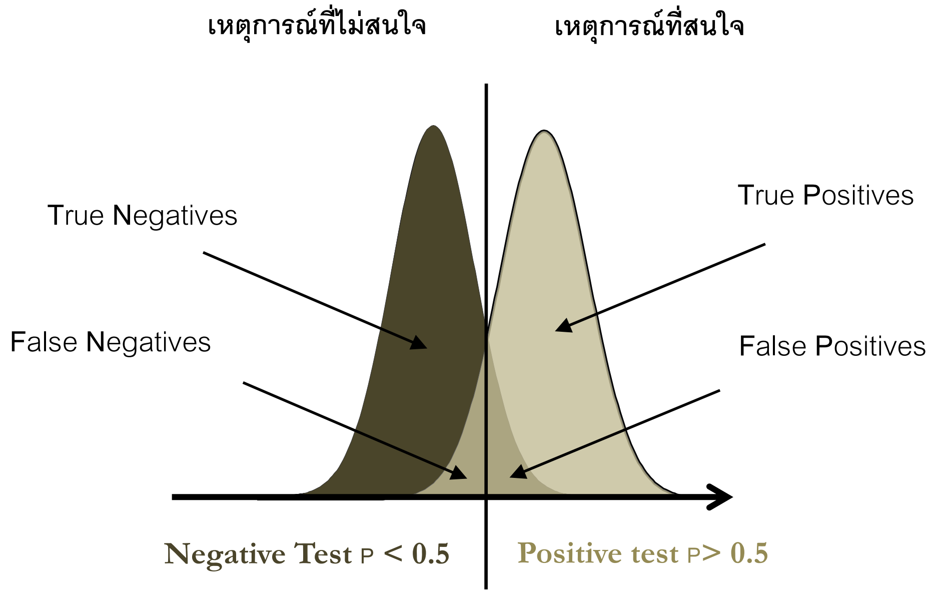
$$P(Y_i) = \frac{e^{g_j(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} > c_{k-1} = \frac{e^{g_k(X)}}{\sum_{j=1}^{k} e^{g_j(X)}} \qquad \text{เมื่อ } c_{k-2} < c_{k-1} < 1$$

เมื่อ c_i คือ จุดแบ่ง (Cut-off point) หรือระดับของความน่าจะเป็นที่ใช้ในการพิจารณาการจำแนกกลุ่มว่าแต่ละหน่วยจะอยู่ในกลุ่มใด เนื่องจากเรามีจำนวนกลุ่ม คือ k กลุ่ม จะมีจุดแบ่งจำนวน k-1 จุด

2.3 Receiver Operating Characteristic Curve (ROC Curve)

2.3.1 กรณีตัวแปรตามมี 2 เหตุการณ์ หรือแบ่งข้อมูลได้เป็น 2 กลุ่ม

จากการที่แบ่งเหตุการณ์ออกเป็น 2 เหตุการณ์ คือ เหตุการณ์ที่สนใจหรือผลการทดสอบเป็นบวก (Positive) ซึ่งจะแทนค่าตัวแปรตามด้วย 1 $(Y_i = 1)$ และเหตุการณ์ที่ไม่สนใจผลการทดสอบเป็นลบ (Negative) ซึ่งจะแทนค่าตัวแปรตามด้วย 0 $(Y_i = 0)$



รูปที่ 2.3.1 แสดงผลการพยากรณ์จำแนกประชากรออกเป็นกลุ่มเหตุการณ์ที่สนใจและกลุ่มเหตุการณ์ที่ไม่สนใจ

จากรูปที่ 2.3.1 หากใช้จุดตัด ซึ่งเป็นตำแหน่งตรงเส้นตรงเป็นเกณฑ์ในการจำแนกเหตุการณ์ออกเป็นกลุ่มของเหตุการณ์ที่ไม่สนใจ และกลุ่มของเหตุการณ์ที่สนใจ พบว่า

- ในเหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นบวก ผลดังกล่าวเป็นผลบวกจริง หรือเรียกว่า true positive (TP)
- ในเหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นบวก ผลดังกล่าวเป็นผลบวกลวง หรือเรียกว่า false positive (FP)
- ในเหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นลบ ผลดังกล่าวเป็นผลลบจริง หรือเรียกว่า true negative (TN)
- ในเหตุการณ์ที่สนใจบางเหตุการณ์ซึ่งมีผลการพยากรณ์เป็นลบ ผลดังกล่าวเป็นผลลบลวง หรือเรียกว่า false negative (FN)

ตารางที่ 2.3.1 Contingency Table for 2-class Classification

ค่าพยากรณ์ \ ค่าจริง	$Y_i = 1$	$Y_i = 2$
$\hat{Y}_i = 1$	$n_{1,1} = TP$	$n_{2,1} = FP$
$\hat{Y}_i = 2$	$n_{1,2} = FN$	$n_{2,2} = TN$
รวม	$TP + FN$	$TN + FP$

จะได้ค่า Sensitivity $= \frac{TP}{TP+FN}$ Specificity $= \frac{TN}{FP+TN}$

Prevalence $= \frac{TP+FN}{TP+FN+FP+TN}$ Accuracy $= \frac{TP+TN}{TP+FN+FP+TN}$

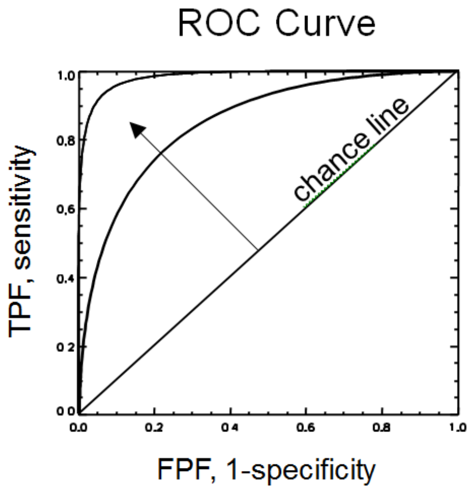
Positive Predict Value $= \frac{TP}{TP+FP}$ Negative Predict Value $= \frac{TN}{FN+TN}$

โดยเครื่องมือที่ใช้ในการวัดประสิทธิภาพของการพยากรณ์ของตัวแบบ ซึ่งผลที่เกิดจากการพยากรณ์มีเพียง 2 ค่า คือ

- Sensitivity (SN) คือสัดส่วนของเหตุการณ์ที่ให้ผลการพยากรณ์เป็นผลบวกจริง (TP) ต่อจำนวนเหตุการณ์ที่สนใจทั้งหมด นั่นคือ $\text{Sensitivity or TPF} = \frac{TP}{TP+FN}$
- Specificity (SP) คือสัดส่วนของเหตุการณ์ที่ให้ผลการพยากรณ์เป็นผลลบจริง (TN) ต่อจำนวนเหตุการณ์ที่ไม่สนใจทั้งหมด นั่นคือ $\text{Specificity or TNF} = \frac{TN}{FP+TN}$

จะเห็นได้ว่า ค่า Sensitivity จะขึ้นอยู่กับจำนวนเหตุการณ์ที่สนใจทั้งหมด ส่วนค่า Specificity ก็จะขึ้นอยู่กับจำนวนเหตุการณ์ที่ไม่สนใจทั้งหมด

โดย เส้นโค้งอาร์โอซีเป็นกราฟที่พล็อตระหว่างค่าของ Sensitivity และ 1-Specificity ซึ่งการพล็อตกราฟจะได้จากการกำหนดจุดตัด (c) ที่ระดับต่างๆ เพื่อแบ่งผลลัพธ์ของการพยากรณ์ออกเป็น 2 กลุ่ม คือกลุ่มที่เกิดเหตุการณ์ $P(Y = 1) \geq$ จุดตัด และกลุ่มที่ไม่เกิดเหตุการณ์ $P(Y = 1) <$ จุดตัด



รูปที่ 2.3.2 แสดงการพล็อตกราฟระหว่างค่า sensitivity กับ ค่า 1-spcificity ของกราฟ ROC

จากกราฟที่ได้จะนำมาหาค่าประมาณพื้นที่ใต้โค้งอาร์โอซี (Area under the Curve หรือ AUC)ออกมา โดยในการคำนวณ AUC จะใช้เทคนิคการประมาณค่าเกี่ยวกับการคำนวณอินทริกรัลจำกัดเขต ซึ่งแสดงดังนี้ $\int_a^b f(x)dx \,; a \le x \le b$

ค่า AUC ที่ได้มาจากการคำนวณ มีเกณฑ์ในการอธิบายความหมายดังนี้

- ถ้า $\mathrm{AUC} = 0.5$ ถือเป็นตัวแบบที่มีความเชื่อถือได้น้อย ไม่สามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้
- ถ้า $0.7 \le \mathrm{AUC} < 0.8$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้พอใช้
- ถ้า $0.8 \le \mathrm{AUC} < 0.9$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ดี โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้ดี
- ถ้า $\mathrm{AUC} \ge 0.9$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ในระดับดีมาก โดยสามารถจำแนกเหตุการณ์ที่สนใจออกจากกลุ่มเหตุการณ์ที่ไม่สนใจได้อย่างชัดเจน

2.3.2 กรณีตัวแปรตามมีมากกว่า 2 เหตุการณ์ หรือแบ่งข้อมูลได้มากกว่า 2 กลุ่ม

จากการที่แบ่งเหตุการณ์ออกเป็น k เหตุการณ์ คือ เหตุการณ์ที่ 1, 2, …, k ถ้าเหตุการณ์ที่เกิดขึ้นเป็นกลุ่มที่ 1 จะแทนค่าตัวแปรตามด้วย 1 $(Y_i = 1)$ ถ้าเหตุการณ์ที่เกิดขึ้นเป็นกลุ่มที่ 2 จะแทนค่าตัวแปรตามด้วย 2 $(Y_i = 2)$ และ ถ้าเหตุการณ์ที่เกิดขึ้นเป็นกลุ่มที่ k จะแทนค่าตัวแปรตามด้วย k $(Y_i = k)$ เป็นต้น

ตารางที่ 2.3.2 Contingency Table for k-class Classification

ค่าจริง / ค่าพยากรณ์	$Y_i = 1$	$Y_i = 2$	$\cdots$	$Y_i = k$
$\widehat{Y}_i = 1$	$n_{1,1}$	$n_{2,1}$	$\cdots$	$n_{k,1}$
$\widehat{Y}_i = 2$	$n_{1,2}$	$n_{2,2}$	$\cdots$	$n_{k,2}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\widehat{Y}_i = k$	$n_{1,k}$	$n_{2,k}$	$\cdots$	$n_{k,k}$
รวม	$n_{1,1} + n_{1,2} + \cdots + n_{1,k}$	$n_{2,1} + n_{2,2} + \cdots + n_{2,k}$	$\cdots$	$n_{k,1} + n_{k,2} + \cdots + n_{k,k}$

โดยที่ค่า $Y_i = i$ หมายถึง เหตุการณ์จริงอยู่กลุ่มที่ i

$\widehat{Y}_i = j$ หมายถึง พยากรณ์ว่าเหตุการณ์อยู่กลุ่มที่ j

$n_{i,j}$ หมายถึง จำนวนของเหตุการณ์จริงอยู่กลุ่มที่ i และพยากรณ์ว่าเป็นกลุ่มที่ j

$; i, j = 1, 2, \dots, k$

$$\text{Probability of correctly classifying class } 1(P_1) = \frac{n_{1,1}}{\sum_{j=1}^{k} n_{1,j}} = \frac{n_{1,1}}{n_{1,1} + n_{1,2} + \cdots + n_{1,k}}$$

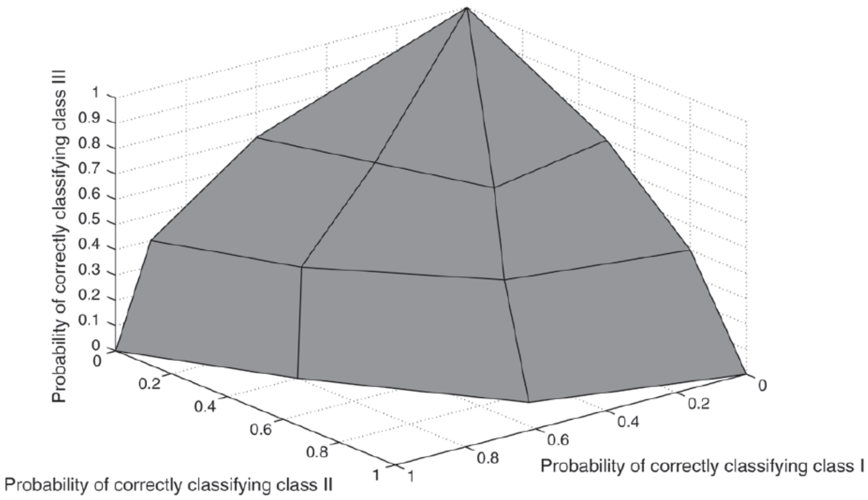
$$\text{Probability of correctly classifying class } 2(P_2) = \frac{n_{2,2}}{\sum_{j=1}^{k} n_{2,j}} = \frac{n_{2,2}}{n_{2,1} + n_{2,2} + \cdots + n_{2,k}}$$

$$\vdots$$

$$\text{Probability of correctly classifying class } k(P_k) = \frac{n_{k,k}}{\sum_{j=1}^{k} n_{k,j}} = \frac{n_{k,k}}{n_{k,1} + n_{k,2} + \cdots + n_{k,k}}$$

จากการสังเกตจะเห็นว่า ค่า Probability of correctly classifying class จะคล้าย กับค่า sensitivity และ ค่า specificity นั่นคือ เป็นค่าสัดส่วนของเหตุการณ์ที่ให้ผลการพยากรณ์เป็นกลุ่มที่ j ซึ่งตรงกับเหตุการณ์ที่เกิดขึ้นจริงเป็นกลุ่มที่ i (โดยที่ ค่า i=j) ต่อจำนวนเหตุการณ์ที่เกิดขึ้นจริงทั้งหมดที่เป็นกลุ่มที่ i จะได้ส่วนประกอบของความน่าจะเป็นของผลการพยากรณ์ที่ตรงกับกับเหตุการณ์ที่เกิดขึ้นจริงในแต่ละกลุ่ม คือ $(P_1, P_2, \dots, P_k)$ โดยที่ค่า $0 \le P_1, P_2, \dots, P_k \le 1$ และ $\sum_{i=1}^{k} P_{i,j} = 1$ โดยที่ $j = 1, 2, \dots, k$

ดังนั้น เส้นโค้งอาร์โอซี กรณีที่แบ่งกลุ่มได้มากกว่า 2 กลุ่ม จะเป็นกราฟที่พล็อตระหว่างค่า Probability of correctly classifying class ของแต่ละกลุ่ม ซึ่งการพล็อตกราฟจะได้จากการกำหนดจุดตัด (c_i) ที่ระดับต่างๆ เพื่อแบ่งผลลัพธ์ของการพยากรณ์ออกเป็น k กลุ่ม ดังนี้ ถ้า $p_1 > c_1$ หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มที่ 1 $(Y_i = 1)$, ถ้า $p_2 > c_2$ หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มที่ 2 $(Y_i = 2)$, …, ถ้า $p_{k-1} > c_{k-1}$ หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มที่ k-1 $(Y_i = k - 1)$ นอกเหนือจากนี้ หน่วยที่ i จะถูกจัดให้อยู่ในกลุ่มที่ k $(Y_i = k)$



รูปที่ 2.3.3 แสดงการพล็อตกราฟ 3 มิติ ระหว่างค่า Probability of correctly classifying class กลุ่มที่ 1, 2 และ 3 ของกราฟ ROC

จากกราฟ 3 มิติที่ได้จะนำมาหาค่าปริมาตรใต้โค้งอาร์โอซี (Volume Under the Curve หรือ VUS) ออกมา โดยในการคำนวณ VUS จะใช้เทคนิคการประมาณค่าเกี่ยวกับการคำนวณอินทริกรัลจำกัดเขตคล้ายกับกรณีมี 2 กลุ่ม ซึ่งแสดงดังนี้

$$\mathrm{VUS} = \int_0^1 \int_0^{f_1(P_1)} P_3 \, dP_2 dP_1 = \int_0^1 \int_0^{f_1(P_1)} f_2(P_1, P_2) \, dP_2 dP_1$$

ซึ่งจะขยายไปสู่กรณีที่มีมากกว่าหรือเท่ากับ 3 มิติ จะเรียกค่านี้ว่า Hypervolume Under the ROC Manifold หรือ HUM ออกมา โดยในการคำนวณ HUM จะใช้เทคนิคการประมาณค่าเกี่ยวกับการคำนวณอินทริกรัลจำกัดเขตคล้ายกับกรณีมี 2 กลุ่ม ซึ่งแสดงดังนี้

กำหนดให้ $P_2 = f_1(P_1)$, $P_3 = f_2(P_1, P_2)$, … , $P_k = f_{k-1}(P_1, P_2, \ldots, P_{k-1})$

ได้สมการดังนี้ $\mathrm{HUM} = \int_0^1 \int_0^{f_1(P_1)} \ldots \int_0^{f_{k-2}(P_1, P_2, \ldots, P_{k-2})} P_k \, dP_{k-1} \ldots dP_2 dP_1$

2.4 เกณฑ์ในการตัดสินใจ

ค่า Probability of correctly classifying จากตาราง Contingency Table Classification ซึ่งเป็นค่าอัตราส่วนที่ผลการทำนายตรงกับค่าจริงที่ได้จากการพยากรณ์ด้วยตัวแบบโลจิสติกแบบพหุนาม จะนำไปพล็อตบนกราฟที่มี k มิติ แล้วนำมาคำนวณหาค่าปริมาตรแบบ

ไฮเปอร์ใต้โค้งอาร์โอซี (HUM) โดยค่า HUM จะมีค่าอยู่ระหว่าง $[0,1]$

ถ้า $\mathrm{HUM} < \frac{1}{k!}$ ถือเป็นตัวแบบที่มีความเชื่อถือได้น้อย ไม่สามารถจำแนกเหตุการณ์ต่างๆ ออกจากกันได้

ถ้า $\mathrm{HUM} \geq \frac{1}{k!}$ ถือเป็นตัวแบบที่มีความเชื่อถือสามารถยอมรับได้ โดยสามารถจำแนกเหตุการณ์ต่างๆได้

ยิ่งค่า HUM เข้าใกล้ 1 ตัวแบบที่ได้ยิ่งสามารถจำแนกเหตุการณ์ต่างๆได้ดีขึ้น จากงานวิจัย[1] ค่าประมาณ $\frac{1}{k!}$ (k = จำนวนเหตุการณ์ทั้งหมดที่เกิด) เป็นค่าต่ำที่สุดที่ควรจะได้ซึ่งเป็นเกณฑ์ในการทำนายประสิทธิภาพของตัวแบบการจำแนกประเภท โดยจะมีความหมายว่า ถ้าได้ค่า HUM ต่ำกว่า $\frac{1}{k!}$ จะแสดงว่า ตัวแบบที่ใช้ในการพยากรณ์ไม่สามารถจำแนกเหตุการณ์ที่เกิดขึ้นออกจากกันได้

[1] Thomas Landgrebe, and Robert P.W. Duin, A simplified extension of the Area under the ROC to the multiclass domain (2006).

บทที่ 3

วิธีดำเนินการศึกษา

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย, ค่าระดับความสัมพันธ์ของตัวแปรอิสระ, ขนาดตัวอย่าง และความแปรปรวนของค่าเฉลี่ยที่มีต่อค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (Hypervolume Under the Manifold (HUM)) ในกรณีที่มีตัวแปรตาม 3, 4 และ 5 กลุ่ม สำหรับข้อมูลในงานวิจัยนี้ ทำการจำลองข้อมูลและวิเคราะห์ผลโดยใช้โปรแกรม R ซึ่งได้ทำการศึกษาภายใต้เงื่อนไขต่างๆดังนี้

3.1 ขอบเขตของการศึกษา

การวิจัยครั้งนี้มีขอบเขตของการวิจัยสำหรับการดำเนินวิจัย ดังนี้

1. ตัวแบบการถดถอยโลจิสติกพหุนาม มี 3 กรณี
   - กรณีมีตัวแปรอิสระ (X) 2 ตัว
   - กรณีมีตัวแปรอิสระ (X) 3 ตัว
   - กรณีมีตัวแปรอิสระ (X) 4 ตัว

   โดยตัวแปรอิสระเป็นตัวแปรเชิงปริมาณที่มีการแจกแจงเริ่มต้นมาจากการแจกแจงแบบปกติ ศึกษากรณี มีพารามิเตอร์ค่าเฉลี่ย คือ μ เป็นค่าคงที่ และมีค่าความแปรปรวน คือ σ^2 เท่ากับ 1 ตามระดับความสัมพันธ์ของตัวแปรอิสระที่กำหนดไว้ คือ 0.5, 0.7 และ 0.9
2. ตัวแปรตาม (Y) เป็นข้อมูลเชิงคุณภาพอยู่ในมาตรเรียงอันดับ (Ordinal Scale) ที่มีการแจกแจงแบบพหุนาม คือ มีค่าตัวแปรตามมากกว่า 2 ค่า หรือให้มี k ค่า คือ Y = 1, 2, … ,k โดยจะทำการศึกษาที่
   - กรณีจำนวนกลุ่มเท่ากับ 3 กลุ่ม คือ Y = 1, 2 และ 3 หรือ k=3
   - กรณีจำนวนกลุ่มเท่ากับ 4 กลุ่ม คือ Y = 1, 2, 3 และ 4 หรือ k=4
   - กรณีจำนวนกลุ่มเท่ากับ 5 กลุ่ม คือ Y = 1, 2, 3, 4 และ 5 หรือ k=5
3. ค่าความคลาดเคลื่อน (ε_i) มีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ μ และ σ โดยเงื่อนไขความคลาดเคลื่อนของตัวแบบโลจิสติก คือ $\mu = 0$ และ $\sigma^2 =$ ค่าคงที่ ดังนั้นในงานวิจัยจะทำการศึกษาที่ $\varepsilon_i \sim N(0,100)$

4. กำหนดค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น ในการจำลองข้อมูล ดังนี้
   - กรณีมีตัวแปรอิสระ (X) 2 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ
     $\beta_1 = \beta_2 = -1,\quad \beta_1 = \beta_2 = -0.6,\quad \beta_1 = \beta_2 = -0.2,$
     $\beta_1 = \beta_2 = 1,\quad \beta_1 = \beta_2 = 0.6,\quad \beta_1 = \beta_2 = 0.2$
   - กรณีมีตัวแปรอิสระ (X) 3 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ
     $\beta_1 = \beta_2 = \beta_3 = -1, \beta_1 = \beta_2 = \beta_3 = -0.6, \beta_1 = \beta_2 = \beta_3 = -0.2,$
     $\beta_1 = \beta_2 = \beta_3 = 1, \beta_1 = \beta_2 = \beta_3 = 0.6, \beta_1 = \beta_2 = \beta_3 = 0.2$
   - กรณีมีตัวแปรอิสระ (X) 4 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ
     $\beta_1 = \beta_2 = \beta_3 = \beta_4 = -1,\quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = -0.6,$
     $\beta_1 = \beta_2 = \beta_3 = \beta_4 = -0.2,\quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1,$
     $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.6,\quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.2$
5. ศึกษาผลกระทบจากค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ กรณีมีตัวแปรอิสระ 2, 3และ 4 ตัว ขนาดตัวอย่าง $n_1 = 50, n_2 = 80, n_3 = 100, n_4 = 120$ และ $n_5 = 130$ ค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0, \beta_1 = \cdots = \beta_p = 1$ ดังนี้
   - กรณีมีตัวแปรอิสระ (X) 2 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_1, X_2) = 0.5, 0.7$ และ 0.9
   - กรณีมีตัวแปรอิสระ (X) 3 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_i, X_j) = 0.5, 0.7$ และ $0.9\ ; i \neq j$ ให้ $\mathrm{corr}(X_1, X_2)$, $\mathrm{corr}(X_1, X_3)$, $\mathrm{corr}(X_2, X_3)$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดยทำการศึกษาภายใต้ระดับความสัมพันธ์ซึ่งมีทั้งหมด 24 วิธี จาก 27 วิธี ที่เป็น positive definite ซึ่งสามารถนำไปใช้ในการคำนวณได้ ดังนี้

     - 0.5, 0.5, 0.5
     - 0.5, 0.5, 0.7
     - 0.5, 0.5, 0.9
     - 0.5, 0.7, 0.5
     - 0.5, 0.7, 0.7
     - 0.5, 0.7, 0.9
     - 0.5, 0.9, 0.5
     - 0.5, 0.9, 0.7
     - 0.7, 0.5, 0.5
     - 0.7, 0.5, 0.7
     - 0.7, 0.5, 0.9
     - 0.7, 0.7, 0.5
     - 0.7, 0.7, 0.7
     - 0.7, 0.7, 0.9
     - 0.7, 0.9, 0.5
     - 0.7, 0.9, 0.7
     - 0.7, 0.9, 0.9
     - 0.9, 0.5, 0.5
     - 0.9, 0.5, 0.7
     - 0.9, 0.7, 0.5
     - 0.9, 0.7, 0.7
     - 0.9, 0.7, 0.9
     - 0.9, 0.9, 0.7
     - 0.9, 0.9, 0.9

   - กรณีมีตัวแปรอิสระ (X) 4 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_i, X_j) = 0.5, 0.7$ และ $0.9\ ; i \neq j$ ให้ $\mathrm{corr}(X_1, X_2), \mathrm{corr}(X_1, X_3), \mathrm{corr}(X_1, X_4), \mathrm{corr}(X_2, X_3),$

$\mathrm{corr}(X_2, X_4), \mathrm{corr}(X_3, X_4)$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดยทำการศึกษาภายใต้ระดับความสัมพันธ์ซึ่งมีทั้งหมด 404 วิธี จาก 729 วิธี ที่เป็น positive definite ซึ่งสามารถนำไปใช้ในการคำนวณได้

6. จำนวนขนาดตัวอย่าง (n) จะให้มีอย่างน้อย 30 เท่าของจำนวนตัวแปรอิสระ โดยผู้วิจัยจะทำการศึกษาผลกระทบจากขนาดตัวอย่าง แบ่งเป็น 2 กรณี ได้แก่ กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน และกรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน โดยกำหนดค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0, \beta_1 = \cdots = \beta_p = 1$ และมีขนาดตัวอย่างของแต่ละกลุ่มดังนี้

- จำนวนกลุ่ม 3 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน

  - $n_1 = 30, \quad n_2 = 30, \quad n_3 = 30$
  - $n_1 = 50, \quad n_2 = 50, \quad n_3 = 50$
  - $n_1 = 80, \quad n_2 = 80, \quad n_3 = 80$
  - $n_1 = 100, \quad n_2 = 100, \quad n_3 = 100$
  - $n_1 = 120, \quad n_2 = 120, \quad n_3 = 120$

  กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน

  - $n_1 = 30, \quad n_2 = 40, \quad n_3 = 50$
  - $n_1 = 30, \quad n_2 = 50, \quad n_3 = 70$
  - $n_1 = 30, \quad n_2 = 60, \quad n_3 = 90$
  - $n_1 = 30, \quad n_2 = 70, \quad n_3 = 110$
  - $n_1 = 30, \quad n_2 = 80, \quad n_3 = 130$

- จำนวนกลุ่ม 4 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน

  - $n_1 = 30, \quad n_2 = 30, \quad n_3 = 30, \quad n_4 = 30$
  - $n_1 = 50, \quad n_2 = 50, \quad n_3 = 50, \quad n_4 = 50$
  - $n_1 = 80, \quad n_2 = 80, \quad n_3 = 80, \quad n_4 = 80$
  - $n_1 = 100, \quad n_2 = 100, \quad n_3 = 100, \quad n_4 = 100$
  - $n_1 = 120, \quad n_2 = 120, \quad n_3 = 120, \quad n_4 = 120$

  กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน

  - $n_1 = 30, \quad n_2 = 40, \quad n_3 = 50, \quad n_4 = 60$
  - $n_1 = 30, \quad n_2 = 50, \quad n_3 = 70, \quad n_4 = 90$
  - $n_1 = 30, \quad n_2 = 60, \quad n_3 = 90, \quad n_4 = 120$
  - $n_1 = 30, \quad n_2 = 70, \quad n_3 = 110, \quad n_4 = 150$
  - $n_1 = 30, \quad n_2 = 80, \quad n_3 = 130, \quad n_4 = 180$

- จำนวนกลุ่ม 5 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน

  - $n_1 = 30, \quad n_2 = 30, \quad n_3 = 30, \quad n_4 = 30, \quad n_5 = 30$
  - $n_1 = 50, \quad n_2 = 50, \quad n_3 = 50, \quad n_4 = 50, \quad n_5 = 50$
  - $n_1 = 80, \quad n_2 = 80, \quad n_3 = 80, \quad n_4 = 80, \quad n_5 = 80$
  - $n_1 = 100, \quad n_2 = 100, \quad n_3 = 100, \quad n_4 = 100, \quad n_5 = 100$

- $n_1 = 120,\quad n_2 = 120,\quad n_3 = 120,\quad n_4 = 120,\quad n_5 = 120$

  กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน

- $n_1 = 30,\quad n_2 = 40,\quad n_3 = 50,\quad n_4 = 60,\quad n_5 = 70$
- $n_1 = 30,\quad n_2 = 50,\quad n_3 = 70,\quad n_4 = 90,\quad n_5 = 110$
- $n_1 = 30,\quad n_2 = 60,\quad n_3 = 90,\quad n_4 = 120,\quad n_5 = 150$
- $n_1 = 30,\quad n_2 = 70,\quad n_3 = 110,\quad n_4 = 150,\quad n_5 = 190$
- $n_1 = 30,\quad n_2 = 80,\quad n_3 = 130,\quad n_4 = 180,\quad n_5 = 230$

7. ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย(var(μ)) หรือลักษณะการซ้อนทับกันของตัวอย่างในแต่ละกลุ่ม $Y \sim N(\mu, \sigma^2)$ มีตัวแปรอิสระ(X) 2, 3 และ 4 ตัว ค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0,\ \beta_1 = \cdots = \beta_p = 1$ ดังนี้

   - กรณีจำนวนกลุ่ม 3 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50,\ n_2 = 80$ และ $n_3 = 100$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้

   - $\mu_1 = 0,\quad \mu_2 = 0,\quad \mu_3 = 0$
   - $\mu_1 = 0,\quad \mu_2 = 2,\quad \mu_3 = 4$
   - $\mu_1 = 0,\quad \mu_2 = 5,\quad \mu_3 = 10$
   - $\mu_1 = 0,\quad \mu_2 = 10,\quad \mu_3 = 20$

   - กรณีจำนวนกลุ่ม 4 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50,\ n_2 = 80,\ n_3 = 100$ และ $n_4 = 120$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้

   - $\mu_1 = 0,\quad \mu_2 = 0,\quad \mu_3 = 0,\quad \mu_4 = 0$
   - $\mu_1 = 0,\quad \mu_2 = 2,\quad \mu_3 = 4,\quad \mu_4 = 6$
   - $\mu_1 = 0,\quad \mu_2 = 5,\quad \mu_3 = 10,\quad \mu_4 = 15$
   - $\mu_1 = 0,\quad \mu_2 = 10,\quad \mu_3 = 20,\quad \mu_4 = 30$

   - กรณีจำนวนกลุ่ม 5 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50,\ n_2 = 80,\ n_3 = 100,\ n_4 = 120$ และ $n_5 = 130$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้

   - $\mu_1 = 0,\quad \mu_2 = 0,\quad \mu_3 = 0,\quad \mu_4 = 0,\quad \mu_5 = 0$
   - $\mu_1 = 0,\quad \mu_2 = 2,\quad \mu_3 = 4,\quad \mu_4 = 6,\quad \mu_5 = 8$
   - $\mu_1 = 0,\quad \mu_2 = 5,\quad \mu_3 = 10,\quad \mu_4 = 15,\quad \mu_5 = 20$
   - $\mu_1 = 0,\quad \mu_2 = 10,\quad \mu_3 = 20,\quad \mu_4 = 30,\quad \mu_5 = 40$

8. การจำลองในแต่ละสถานการณ์จะกระทำซ้ำ 1,000 รอบ (R=1,000)

3.2 วิธีดำเนินการศึกษา

1. ศึกษาค้นคว้าเอกสารและข้อมูลที่เกี่ยวข้องกับงานวิจัย
2. กำหนดเงื่อนไขและขอบเขตของการวิจัย
   - กำหนดขนาดตัวอย่างแต่ละกลุ่ม (n_i)
   - กำหนดค่าสัมประสิทธิ์ความถดถอยเริ่มต้น β $(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$

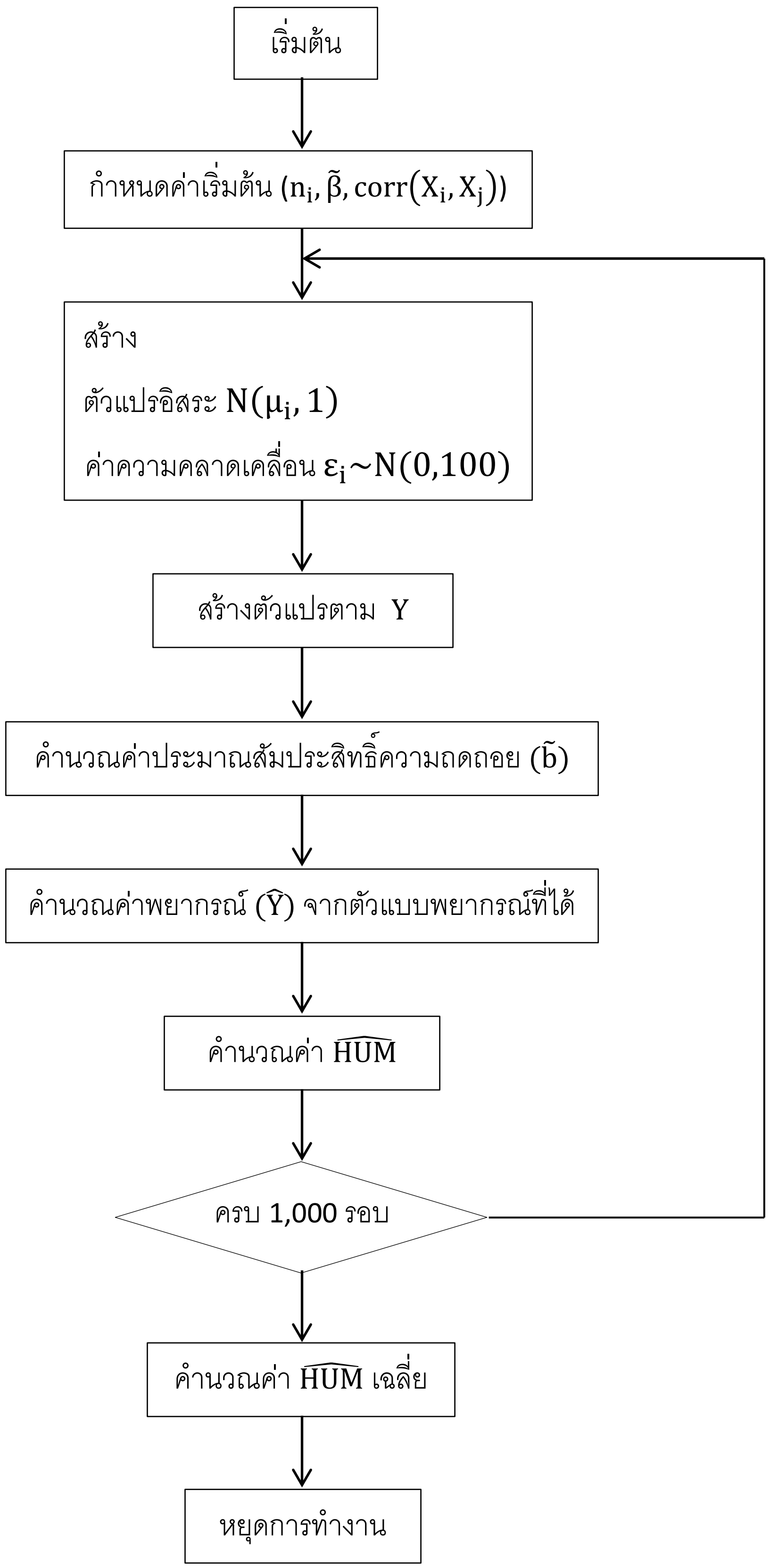
ตามกรณีที่ต้องการศึกษา

- ศึกษาผลกระทบจากค่าพารามิเตอร์ β ณ ค่าต่างๆ
- ศึกษาผลกระทบจากค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ
- ศึกษาผลกระทบจากจำนวนขนาดตัวอย่าง
- ศึกษาผลกระทบจากค่าความแปรปรวนของค่าเฉลี่ย

3. จำลองข้อมูลตามการแจกแจงและขอบเขตที่ต้องการศึกษา
   - กำหนดขนาดตัวอย่างแต่ละกลุ่ม
   - จำลองค่า X แต่ละกลุ่มตามการแจกแจงแบบปกติ โดยให้ X_i และ X_j มีความสัมพันธ์กันตามที่กำหนดขอบเขตไว้
   - หาค่า $\tilde{Y}$ ของแต่ละกลุ่มตามที่กำหนดจำนวนไว้ จากค่าสัมประสิทธิ์ความถดถอยที่เริ่มต้น และค่า X ที่จำลองได้ของแต่ละกลุ่ม
4. หาค่าประมาณสัมประสิทธิ์การถดถอยของพารามิเตอร์ $\tilde{b}$ $(b_0, b_1, b_2, \ldots, b_p)$ เพื่อสร้างตัวแบบสำหรับนำมาพยากรณ์
5. นำตัวแบบโลจิสติกที่ได้ไปพยากรณ์ข้อมูลย้อนกลับ เพื่อตรวจสอบผลการพยากรณ์เมื่อเทียบกับเหตุการณ์จริง
6. คำนวณหาค่าประมาณปริมาตรแบบไฮเปอร์ใต้ผิวโค้งอาร์โอซี (HUM)
7. ทำซ้ำในแต่ละการทดสอบจำนวน 1,000 รอบ เพื่อหาค่าเฉลี่ยของค่าปริมาตรแบบไฮเปอร์ใต้ผิวโค้งอาร์โอซี
8. เปรียบเทียบค่าปริมาตรแบบไฮเปอร์ใต้ผิวโค้งอาร์โอซีที่ได้ และสรุปผลที่ได้จากการวิจัย

จากขั้นตอนการศึกษาด้านบนสามารถเขียนแผนผังการดำเนินงานได้ ดังนี้

แผนผังการเขียนโปรแกรมในแต่ละกรณีที่ทำการศึกษา



บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย, ค่าระดับความสัมพันธ์ของตัวแปรอิสระ, ขนาดตัวอย่าง และความแปรปรวนของค่าเฉลี่ยที่มีต่อค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (Hypervolume Under the Manifold (HUM)) ในกรณีที่มีตัวแปรตามจำนวน 3, 4 และ 5 กลุ่ม สำหรับข้อมูลในงานวิจัยนี้ ทำการจำลองข้อมูลและวิเคราะห์ผลโดยใช้โปรแกรม R แบ่งผลการวิจัยออกเป็นกรณีต่างๆดังนี้

1. ผลการวิจัยกรณีมีจำนวนกลุ่ม 3 กลุ่ม

    1.1 ศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย $\tilde{\beta}$ ณ ระดับต่างๆ

    1.2 ศึกษาผลกระทบจากค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ

    1.3 ศึกษาผลกระทบจากจำนวนขนาดตัวอย่าง แบ่งเป็น 2 กรณี

        1.3.1 เมื่อขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน

        1.3.2 เมื่อขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน

    1.4 ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย $\mathrm{var}(\mu)$

2. ผลการวิจัยกรณีมีจำนวนกลุ่ม 4 กลุ่ม

    2.1 ศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย $\tilde{\beta}$ ณ ระดับต่างๆ

    2.2 ศึกษาผลกระทบจากค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ

    2.3 ศึกษาผลกระทบจากจำนวนขนาดตัวอย่าง แบ่งเป็น 2 กรณี

        2.3.1 เมื่อขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน

        2.3.2 เมื่อขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน

    2.4 ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย $\mathrm{var}(\mu)$

3.ผลการวิจัยกรณีมีจำนวนกลุ่ม 5 กลุ่ม

    3.1 ศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย $\tilde{\beta}$ ณ ระดับต่างๆ

    3.2 ศึกษาผลกระทบจากค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ

    3.3 ศึกษาผลกระทบจากจำนวนขนาดตัวอย่าง แบ่งเป็น 2 กรณี

        3.3.1 เมื่อขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน

        3.3.2 เมื่อขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน

    3.4 ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย $\mathrm{var}(\mu)$

ได้ผลการวิเคราะห์ดังนี้

1. **ผลการวิจัยกรณีมีจำนวนกลุ่ม 3 กลุ่ม ที่มีผลกระทบต่อปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (HUM) สำหรับตัวแบบโลจิสติกพหุนาม**

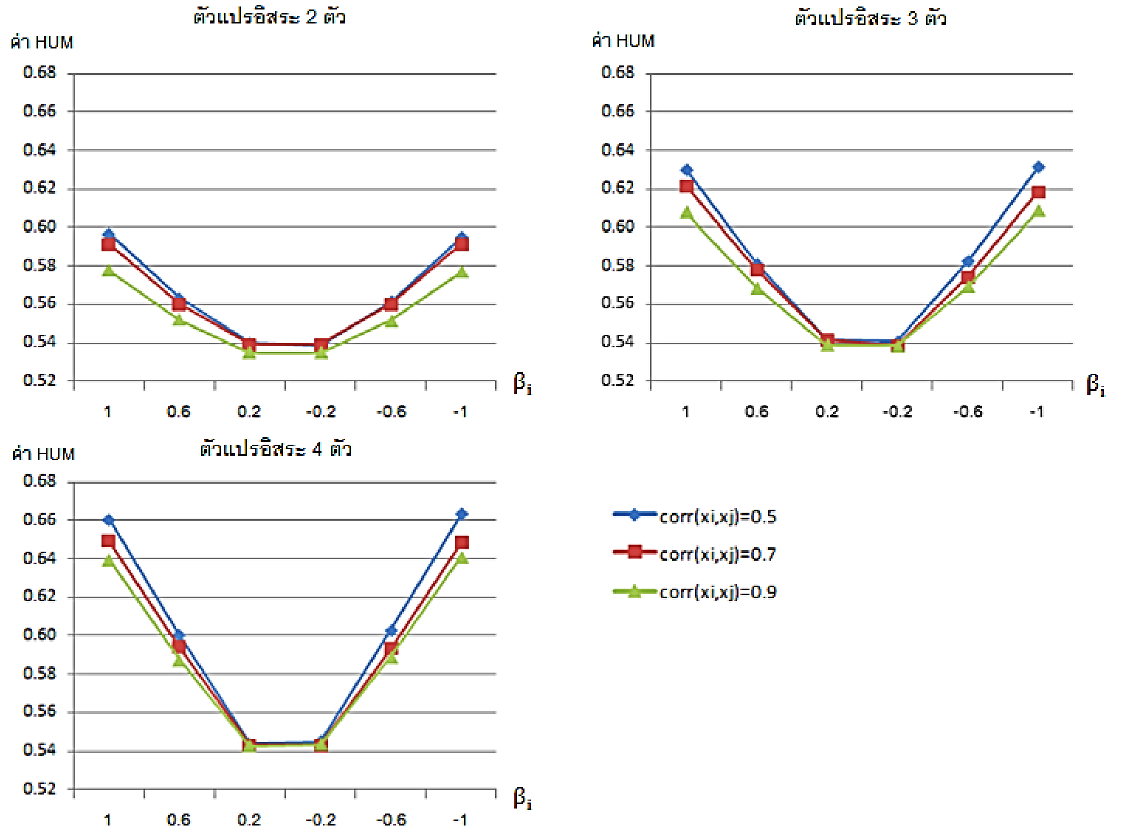
1.1 **ศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย $\tilde{\beta}$ ณ ระดับต่างๆ**

กรณี $n_1 = 50,\ n_2 = 80,\ n_3 = 100$ ที่ค่า $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})\ ; i \neq j\ ,\ i' \neq j'$ ที่ค่า β_0=0

ตารางที่ 1.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าสัมประสิทธิ์ความถดถอย $\tilde{\beta}$ ณ ระดับต่างๆ

		$\mathrm{corr}(X_i, X_j)$		
		0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	β_1=β_2= 1	0.5968	0.5913	0.5775
	β_1=β_2= 0.6	0.5632	0.5599	0.5518
	β_1=β_2= 0.2	0.5397	0.5389	0.5347
	β_1=β_2=-0.2	0.5387	0.5389	0.5349
	β_1=β_2=-0.6	0.5611	0.5601	0.5514
	β_1=β_2=-1	0.5947	0.5914	0.5767
ตัวแปรอิสระ 3 ตัว	β_1=β_2=β_3= 1	0.6301	0.6218	0.6081
	β_1=β_2=β_3= 0.6	0.5812	0.5778	0.5687
	β_1=β_2=β_3= 0.2	0.5414	0.5414	0.5386
	β_1=β_2=β_3=-0.2	0.5408	0.5384	0.5383
	β_1=β_2=β_3=-0.6	0.5824	0.5741	0.5691
	β_1=β_2=β_3=-1	0.6313	0.6184	0.6087
ตัวแปรอิสระ 4 ตัว	β_1=β_2=β_3=β_4= 1	0.6604	0.6492	0.6393
	β_1=β_2=β_3=β_4= 0.6	0.6002	0.5939	0.5875
	β_1=β_2=β_3=β_4= 0.2	0.5439	0.5427	0.5426
	β_1=β_2=β_3=β_4=-0.2	0.5449	0.5431	0.5432
	β_1=β_2=β_3=β_4=-0.6	0.6028	0.5933	0.5890
	β_1=β_2=β_3=β_4=-1	0.6633	0.6483	0.6409

รูปที่ 1.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าสัมประสิทธิ์ความถดถอย β ณ ระดับต่างๆ



จากตารางและรูปที่ 1.1 เมื่อค่าพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ ในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า เมื่อค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ เพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าเพิ่มขึ้น

จากผลการวิจัยข้างต้นพบว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีแปรผันตรงกับค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ คือ เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามมีค่าเพิ่มขึ้น ส่งผลให้ ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าเพิ่มขึ้น หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้มากขึ้น ทำให้ความสามารถในการทำนายกลุ่มสูงขึ้น

1.2 ศึกษาผลกระทบจากค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ

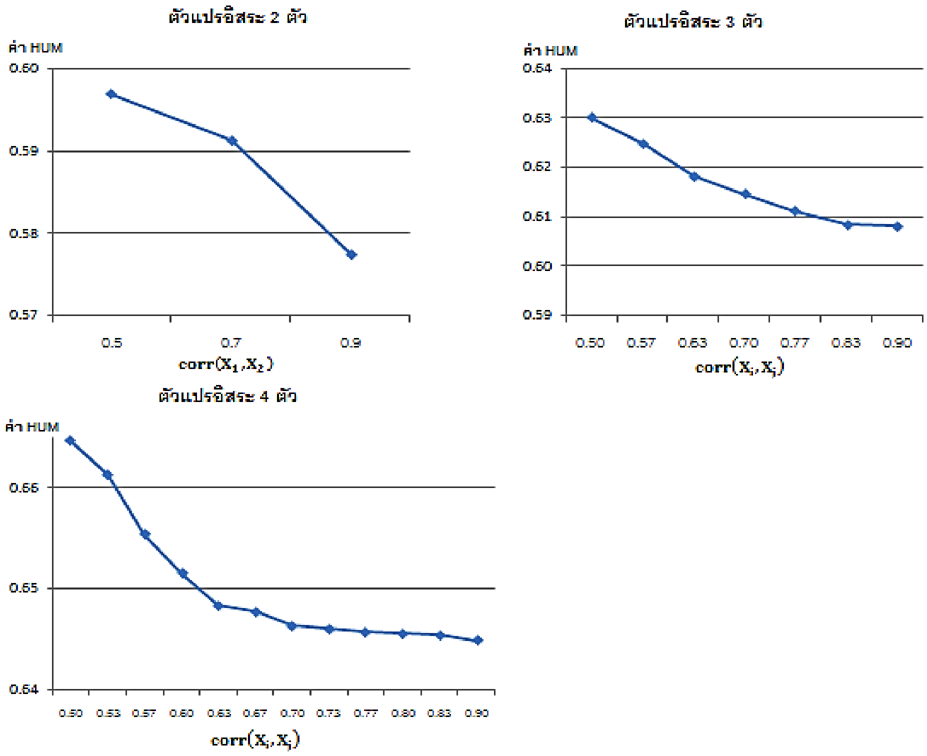
กรณี $n_1 = 50,\ n_2 = 80,\ n_3 = 100$ ที่ค่า $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})$

; $i \neq j$, $i' \neq j'$ ที่ค่า $\beta_0=0$ และ $\beta_1 = \cdots = \beta_p = 1$

ตารางที่ 1.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ระดับต่างๆ

ตัวแปรอิสระ 2 ตัว	$\mathrm{corr}(X_1, X_2)$	0.5	0.7	0.9				
	HUM	0.5968	0.5913	0.5775				
ตัวแปรอิสระ3 ตัว	$\mathrm{corr}(X_i, X_j)$	0.50	0.57	0.63	0.70	0.77	0.83	0.90
	HUM	0.6301	0.6250	0.6183	0.6146	0.6112	0.6084	0.6081
ตัวแปรอิสระ 4 ตัว	$\mathrm{corr}(X_i, X_j)$	0.50	0.53	0.57	0.60	0.63	0.67	0.70
	HUM	0.6596	0.6562	0.6504	0.6465	0.6434	0.6428	0.6413
	$\mathrm{corr}(X_i, X_j)$	0.73	0.77	0.80	0.83	0.90		
	HUM	0.6410	0.6407	0.6406	0.6404	0.6399		

รูปที่ 1.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ระดับต่างๆ



จากตารางและรูปที่ 1.2 เมื่อค่าความสัมพันธ์ระหว่างตัวแปรอิสระในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า ถ้าค่าความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้นจะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มลดลง

จากผลการวิจัยข้างต้นพบว่า ค่าความสัมพันธ์ระหว่างตัวแปรอิสระแปรผกผันกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี เนื่องจากการที่ ตัวแปรอิสระมีความสัมพันธ์กัน ผิดสมมุติฐานเบื้องต้นของการวิเคราะห์ความถดถอย เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูงขึ้นทำให้ความสามารถในการทำนายกลุ่มของสมการถดถอยโลจิสติกลดลง

1.3 **ศึกษาผลกระทบจากจำนวนขนาดตัวอย่าง** แบ่งเป็น 2 กรณี

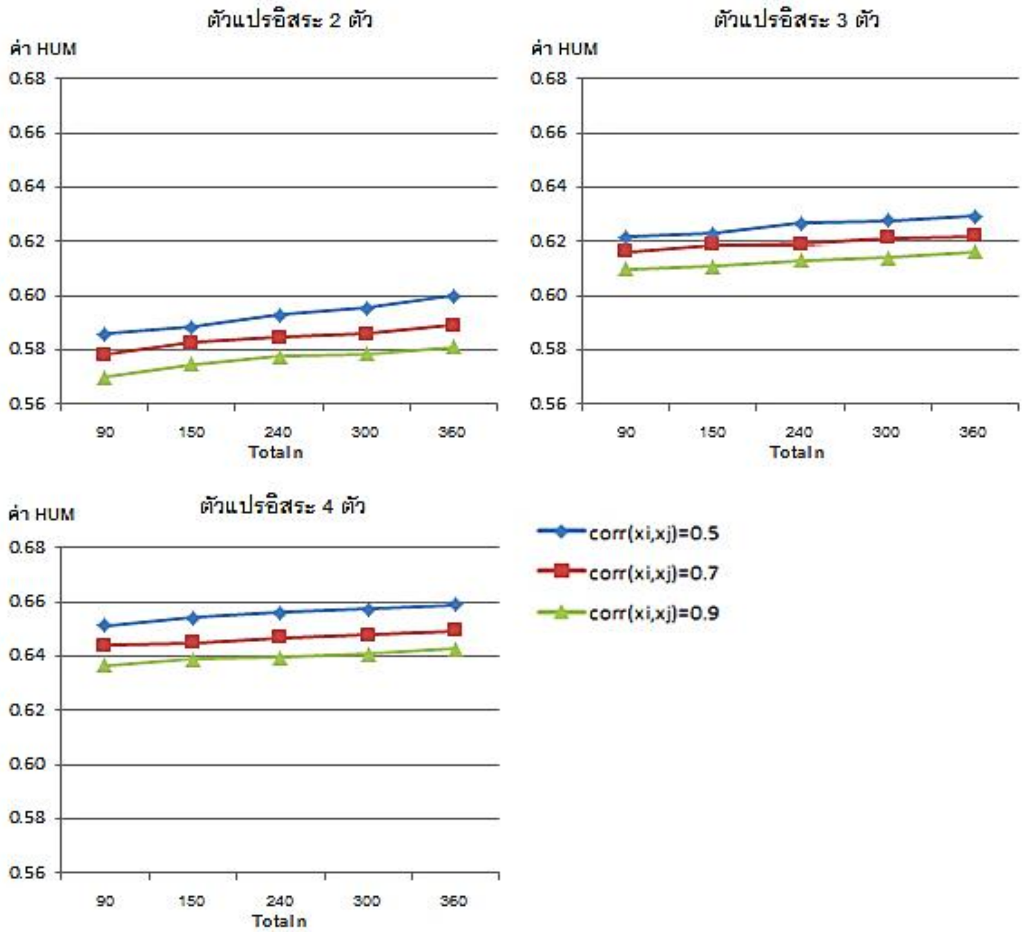
1.3.1 เมื่อขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน

กรณี $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})$; $i \neq j$, $i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$

<u>ตารางที่ 1.3.1</u> แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ขนาดตัวอย่างระดับต่างๆ

						$\mathrm{corr}(X_i, X_j)$		
	n_1	n_2	n_3	Total n	$\mathrm{var}(n)$	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	30	30	30	90	0	0.5860	0.5783	0.5700
	50	50	50	150	0	0.5886	0.5829	0.5746
	80	80	80	240	0	0.5933	0.5847	0.5776
	100	100	100	300	0	0.5953	0.5860	0.5786
	120	120	120	360	0	0.6000	0.5892	0.5812
ตัวแปรอิสระ 3 ตัว	30	30	30	90	0	0.6217	0.6162	0.6098
	50	50	50	150	0	0.6230	0.6187	0.6112
	80	80	80	240	0	0.6269	0.6192	0.6129
	100	100	100	300	0	0.6278	0.6212	0.6142
	120	120	120	360	0	0.6295	0.6221	0.6161
ตัวแปรอิสระ 4 ตัว	30	30	30	90	0	0.6513	0.6440	0.6366
	50	50	50	150	0	0.6542	0.6451	0.6388
	80	80	80	240	0	0.6563	0.6468	0.6394
	100	100	100	300	0	0.6579	0.6478	0.6408
	120	120	120	360	0	0.6590	0.6494	0.6427

รูปที่ 1.3.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ขนาดตัวอย่าง ระดับต่างๆ



จากตารางและรูปที่ 1.3.1 เมื่อขนาดตัวอย่างในตัวแบบจำลองข้อมูลที่ทุกกลุ่มมีจำนวนตัวอย่างเท่ากันเปลี่ยนแปลง จะพบว่า เมื่อขนาดตัวอย่างมีจำนวนเพิ่มขึ้นจะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น นอกจากนี้เมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าเพิ่มขึ้น โดย $\mathrm{var}(n) = 0$ หมายถึง ขนาดตัวอย่างกลุ่มที่ i และ j เท่ากัน โดยที่ $i \neq j; i, j = 1, 2, \ldots, k$

จากผลการวิจัยข้างต้นพบว่า ขนาดตัวอย่างกรณีที่จำนวนตัวอย่างเท่ากันทุกกลุ่มแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี

1.3.2 **เมื่อขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน**

กรณี $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})$; $i \neq j$, $i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$ โดยใช้ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ มาเปรียบเทียบกัน ซึ่งสามารถ

คำนวณหาความแปรปรวนของขนาดตัวอย่างได้จากสูตร $\mathrm{var(n)} = \frac{\sum_{Y=1}^{k}(n_Y-\bar{n})^2}{k-1}$ โดยที่

- ค่า $\mathrm{var(n)}$ มีค่าน้อย หมายถึง ตัวอย่างแต่ละกลุ่มมีจำนวนใกล้เคียงกัน
- ค่า $\mathrm{var(n)}$ มีค่ามาก หมายถึง ตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น

ค่า $\mathrm{var(n)} = [0,\infty)$ เมื่อค่า $\mathrm{var(n)} = 0$ หมายถึง ขนาดตัวอย่างแต่ละกลุ่มมีจำนวนเท่ากัน ในการวิจัยครั้งนี้จะทำการศึกษาเฉพาะกรณีขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน ($\mathrm{var(n)} = 0$) และขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน ($\mathrm{var(n)} \neq 0$) โดยขนาดตัวอย่างจะเรียงจำนวนกลุ่มจากน้อยไปมากเท่านั้น

ค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่าง (coefficient of variation: $\mathrm{CV_n}$) คือ ค่าร้อยละที่ข้อมูลเบี่ยงเบนรอบๆค่าขนาดตัวอย่างเฉลี่ย ซึ่งใช้วัดการกระจายสัมพัทธ์ของข้อมูลเพื่อเปรียบเทียบการกระจายของข้อมูลตั้งแต่สองชุดขึ้นไป ซึ่งมีสูตร คือ

$\mathrm{CV_n} = \frac{\mathrm{SD}}{\bar{n}} \times 100$ โดยที่

- ค่า $\mathrm{CV_n}$ มีค่าน้อย หมายถึง ค่าการกระจายของข้อมูลรอบๆค่าขนาดตัวอย่างเฉลี่ยน้อย หรือตัวอย่างแต่ละกลุ่มมีจำนวนใกล้เคียงกัน
- ค่า $\mathrm{CV_n}$ มีค่ามาก หมายถึง ค่าการกระจายของข้อมูลรอบๆค่าขนาดตัวอย่างเฉลี่ยมาก หรือ ตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น

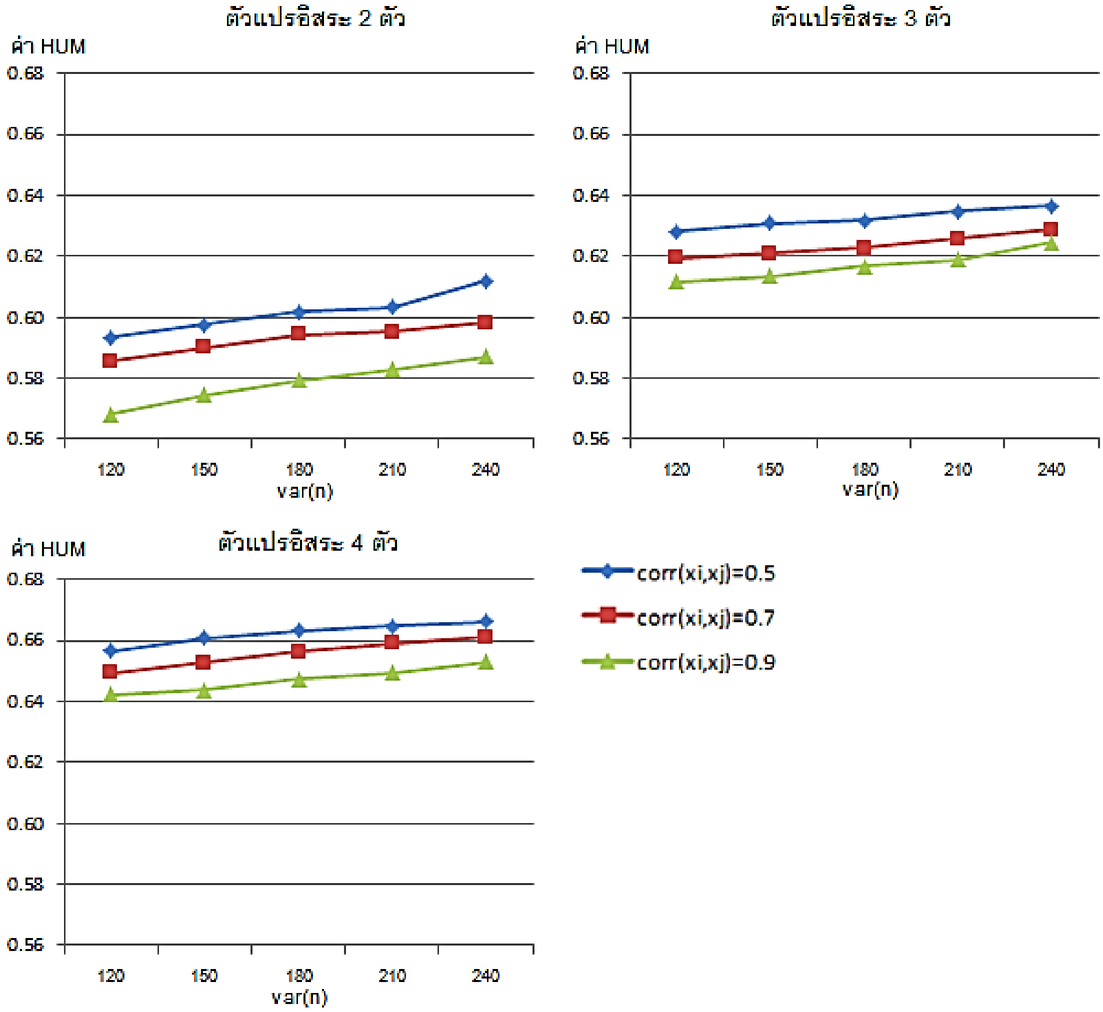
<u>ตารางที่ 1.3.2</u> แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ

								$\mathrm{corr}(X_i, X_j)$		
	n_1	n_2	n_3	Total n	$\bar{n}$	$\mathrm{var(n)}$	$\mathrm{CV_n}$	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	30	40	50	120	40	100	25	0.5933	0.5856	0.5683
	30	50	70	150	50	400	40	0.5974	0.5900	0.5745
	30	60	90	180	60	900	50	0.6016	0.5939	0.5793
	30	70	110	210	70	1600	57.14	0.6032	0.5950	0.5827
	30	80	130	240	80	2500	62.5	0.6119	0.5983	0.5871
ตัวแปรอิสระ 3 ตัว	30	40	50	120	40	100	25	0.6282	0.6195	0.6114
	30	50	70	150	50	400	40	0.6309	0.6209	0.6134
	30	60	90	180	60	900	50	0.6316	0.6226	0.6168
	30	70	110	210	70	1600	57.14	0.6347	0.6259	0.6185
	30	80	130	240	80	2500	62.5	0.6367	0.6286	0.6244

ตารางที่ 1.3.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ(ต่อ)

								$corr(X_i, X_j)$		
	n_1	n_2	n_3	Total n	$\bar{n}$	var(n)	CV_n	0.5	0.7	0.9
ตัวแปรอิสระ 4 ตัว	30	40	50	120	40	100	25	0.6567	0.6494	0.6421
	30	50	70	150	50	400	40	0.6605	0.6527	0.6439
	30	60	90	180	60	900	50	0.6633	0.6562	0.6473
	30	70	110	210	70	1600	57.14	0.6650	0.6590	0.6493
	30	80	130	240	80	2500	62.5	0.6661	0.6612	0.6530

รูปที่ 1.3.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ



จากตารางและรูปที่ 1.3.2 กรณีขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน เมื่อค่าความแปรปรวนของขนาดตัวอย่างในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า เมื่อค่าความแปรปรวนของขนาดตัวอย่างเพิ่มขึ้น เมื่อตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น หรือเมื่อพิจารณาจากค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่างที่เพิ่มขึ้น เมื่อมีค่าการกระจายของข้อมูลรอบๆค่าเฉลี่ยเพิ่มขึ้น จะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้ง อาร์โอซีมีค่าสูงขึ้น โดย $\mathrm{var}(n) \neq 0$ หมายถึง ขนาดตัวอย่างกลุ่มที่ i และ j ไม่เท่ากัน โดยที่ $i \neq j; i,\ j = 1, 2, \dots, k$

จากผลการวิจัยข้างต้นพบว่า กรณีที่ขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน ความแปรปรวนของขนาดตัวอย่างหรือค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่างแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี โดยการที่ขนาดตัวอย่างในแต่ละกลุ่มมีจำนวนแตกต่างกันเพิ่มขึ้นทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น ทั้งนี้อาจเนื่องมากจากการที่ขนาดตัวอย่างรวมมีจำนวนเพิ่มขึ้น เช่นเดียวกับกรณีที่ขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน นอกจากนี้ยังสังเกตเห็นว่า ในกรณีที่ขนาดตัวอย่างรวมเท่ากัน เมื่อเปรียบเทียบระหว่างกรณีที่มีตัวอย่างในแต่ละกลุ่มไม่เท่ากันและกรณีที่ตัวอย่างในแต่ละกลุ่มเท่ากัน ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่มีตัวอย่างในแต่ละกลุ่มไม่เท่ากันมีค่าสูงกว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่มีตัวอย่างในแต่ละกลุ่มเท่ากัน

1.4 ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย $(\mathrm{var}(\mu))$

กรณี $n_1 = 50,\ n_2 = 80,\ n_3 = 100$ ที่ค่า $\mathrm{corr}(X_I, X_j) = \mathrm{corr}(X_{i'}, X_{j'})\ ; i \neq j$, $i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$ ซึ่งสามารถคำนวณหาความแปรปรวนของค่าเฉลี่ยได้จากสูตร $\mathrm{var}(\mu) = \frac{\sum_{Y=1}^{k}(\mu_Y - \bar{\mu})^2}{k-1}$ โดยที่

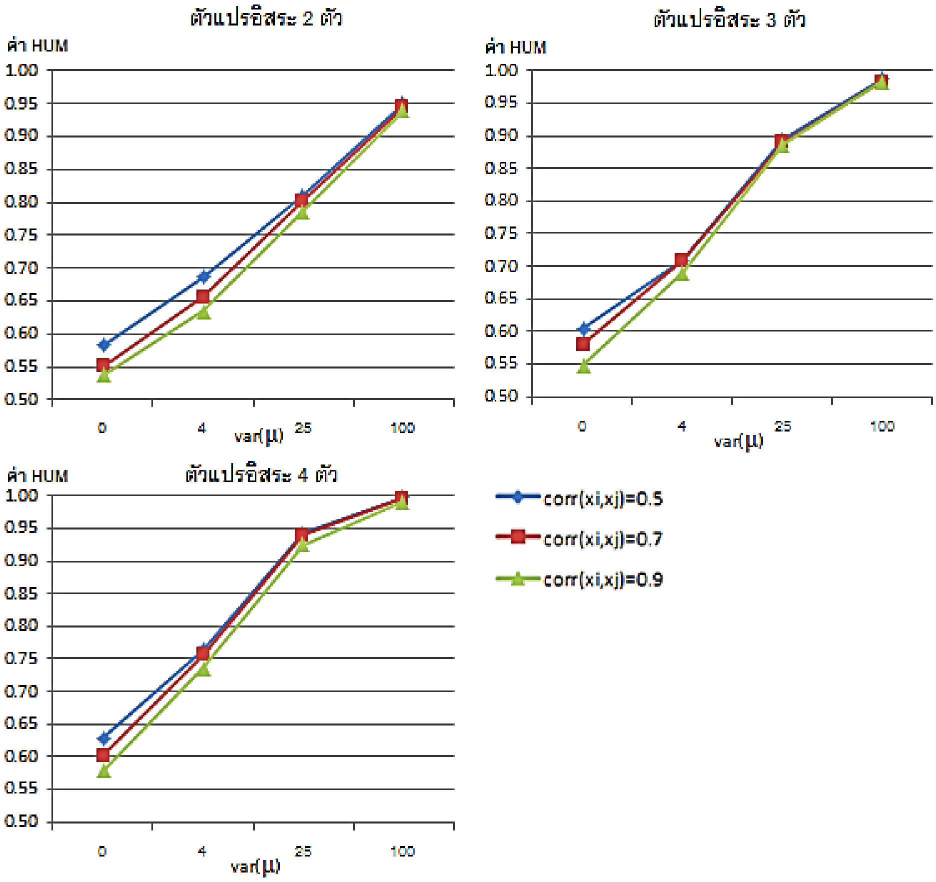
- ค่า $\mathrm{var}(\mu)$ มีค่าน้อย หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าใกล้เคียงกัน หรือ ตัวอย่างในแต่ละกลุ่มมีค่าใกล้เคียงกัน ข้อมูลแต่ละกลุ่มซ้อนทับกันมาก
- ค่า $\mathrm{var}(\mu)$ มีค่ามาก หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าแตกต่างกันมาก หรือ ตัวอย่างในแต่ละกลุ่มมีค่าแตกต่างกันมาก ข้อมูลแต่ละกลุ่มซ้อนทับกันน้อย

ค่า $\mathrm{var}(\mu) = [0, \infty)$ เมื่อค่า $\mathrm{var}(\mu) = 0$ หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าเท่ากัน ในการวิจัยครั้งนี้จะทำการศึกษาเริ่มต้นที่ค่าเฉลี่ยของกลุ่มที่ 1 เท่ากับ 0 เสมอ และค่าเฉลี่ยของกลุ่มถัดไป มีค่าเพิ่มขึ้นเรื่อยๆ

ตารางที่ 1.4 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของค่าเฉลี่ยระดับต่างๆ

					$\mathrm{corr}(X_i, X_j)$		
	μ_1	μ_2	μ_3	$\mathrm{var}(\mu)$	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	0	0	0	0	0.5819	0.5501	0.5369
	0	2	4	4	0.6863	0.6560	0.6343
	0	5	10	25	0.8100	0.8024	0.7857
	0	10	20	100	0.9499	0.9445	0.9391
ตัวแปรอิสระ 3 ตัว	0	0	0	0	0.6048	0.5813	0.5491
	0	2	4	4	0.7107	0.7086	0.6903
	0	5	10	25	0.8938	0.8903	0.8872
	0	10	20	100	0.9869	0.9845	0.9824
ตัวแปรอิสระ 4 ตัว	0	0	0	0	0.6281	0.6026	0.5776
	0	2	4	4	0.7633	0.7567	0.7353
	0	5	10	25	0.9422	0.9398	0.9255
	0	10	20	100	0.9978	0.9956	0.9910

<u>รูปที่ 1.4</u> แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของค่าเฉลี่ยระดับต่างๆ



จากตารางและรูปที่ 1.4 เมื่อความแปรปรวนของค่าเฉลี่ย เปลี่ยนแปลง จะพบว่า ถ้าความแปรปรวนของค่าเฉลี่ยเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น โดย $\mathrm{var}(\mu)$ แสดงลักษณะการกระจายของค่าเฉลี่ยกลุ่มที่ i และ j โดยที่ $i \neq j$ $; i,\ j = 1, 2, \ldots, k$ กรณี $\mathrm{var}(\mu) = 0$ หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าเท่ากันจะได้ว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าต่ำที่สุด ทุกระดับความแปรปรวน เนื่องจากข้อมูลทั้ง 3 กลุ่ม ซ้อนทับกันมาก ทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าน้อย และเมื่อค่าเฉลี่ยของแต่ละกลุ่มห่างกันมากขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะเพิ่มขึ้นเรื่อยๆ แต่ไม่เกิน 1

จากผลการวิจัยพบว่า ความแปรปรวนของค่าเฉลี่ยแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี คือ เมื่อค่าเฉลี่ยของแต่ละกลุ่มที่ค่าเท่ากัน ($\mathrm{var}(\mu) = 0$) ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าต่ำที่สุด และเมื่อค่าความแปรปรวนของค่าเฉลี่ยมากขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าเพิ่มขึ้น นอกจากนี้

ยังพบว่า เมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้นค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มเพิ่มขึ้น เนื่องมาจากการที่ตัวแปรอิสระเพิ่มมากทำให้ความสามารถในการอธิบายตัวแปรตามเพิ่มมากขึ้น

2. ผลการวิจัยกรณีมีจำนวนกลุ่ม 4 กลุ่ม ที่มีผลกระทบต่อปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (HUM) สำหรับตัวแบบโลจิสติกพหุนาม

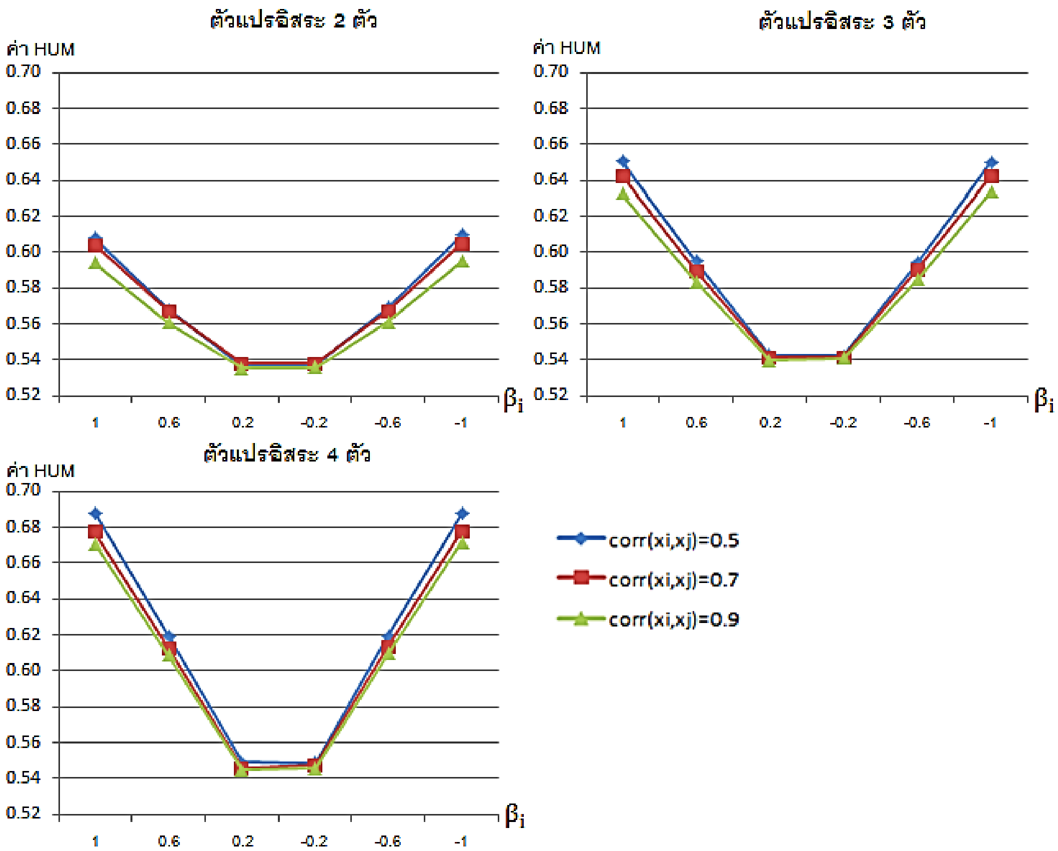
2.1 ศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย $\tilde{\beta}$ ณ ระดับต่างๆ

กรณี $n_1 = 50, n_2 = 80, n_3 = 100, n_4 = 120$ ที่ค่า $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})$; $i \neq j,\ i' \neq j'$ ที่ค่า β_0=0

<u>ตารางที่ 2.1</u> แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าสัมประสิทธิ์ความถดถอย $\tilde{\beta}$ ณ ระดับต่างๆ

		$\mathrm{corr}(X_i, X_j)$		
		0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	$\beta_1=\beta_2=$ 1	0.6079	0.6045	0.5941
	$\beta_1=\beta_2=$ 0.6	0.5680	0.5670	0.5608
	$\beta_1=\beta_2=$ 0.2	0.5367	0.5379	0.5355
	$\beta_1=\beta_2=$-0.2	0.5378	0.5385	0.5359
	$\beta_1=\beta_2=$-0.6	0.5694	0.5672	0.5611
	$\beta_1=\beta_2=$-1	0.6097	0.6048	0.5947
ตัวแปรอิสระ 3 ตัว	$\beta_1=\beta_2=\beta_3=$ 1	0.6507	0.6430	0.6327
	$\beta_1=\beta_2=\beta_3=$ 0.6	0.5947	0.5897	0.5837
	$\beta_1=\beta_2=\beta_3=$ 0.2	0.5427	0.5413	0.5397
	$\beta_1=\beta_2=\beta_3=$-0.2	0.5423	0.5415	0.5409
	$\beta_1=\beta_2=\beta_3=$-0.6	0.5941	0.5901	0.5848
	$\beta_1=\beta_2=\beta_3=$-1	0.6502	0.6427	0.6338
ตัวแปรอิสระ 4 ตัว	$\beta_1=\beta_2=\beta_3=\beta_4=$ 1	0.6879	0.6775	0.6712
	$\beta_1=\beta_2=\beta_3=\beta_4=$ 0.6	0.6196	0.6127	0.6085
	$\beta_1=\beta_2=\beta_3=\beta_4=$ 0.2	0.5490	0.5461	0.5448
	$\beta_1=\beta_2=\beta_3=\beta_4=$-0.2	0.5488	0.5472	0.5458
	$\beta_1=\beta_2=\beta_3=\beta_4=$-0.6	0.6194	0.6137	0.6093
	$\beta_1=\beta_2=\beta_3=\beta_4=$-1	0.6876	0.6781	0.6719

รูปที่ 2.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าสัมประสิทธิ์ความถดถอย β ณ ระดับต่างๆ



จากตารางและรูปที่ 2.1 เมื่อค่าพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ ในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า เมื่อค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ เพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าเพิ่มขึ้น

จากผลการวิจัยข้างต้นพบว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีแปรผันตรงกับค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ คือ เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามมีค่าเพิ่มขึ้น ส่งผลให้ ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าเพิ่มขึ้น หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้มากขึ้น ทำให้ความสามารถในการทำนายกลุ่มสูงขึ้น

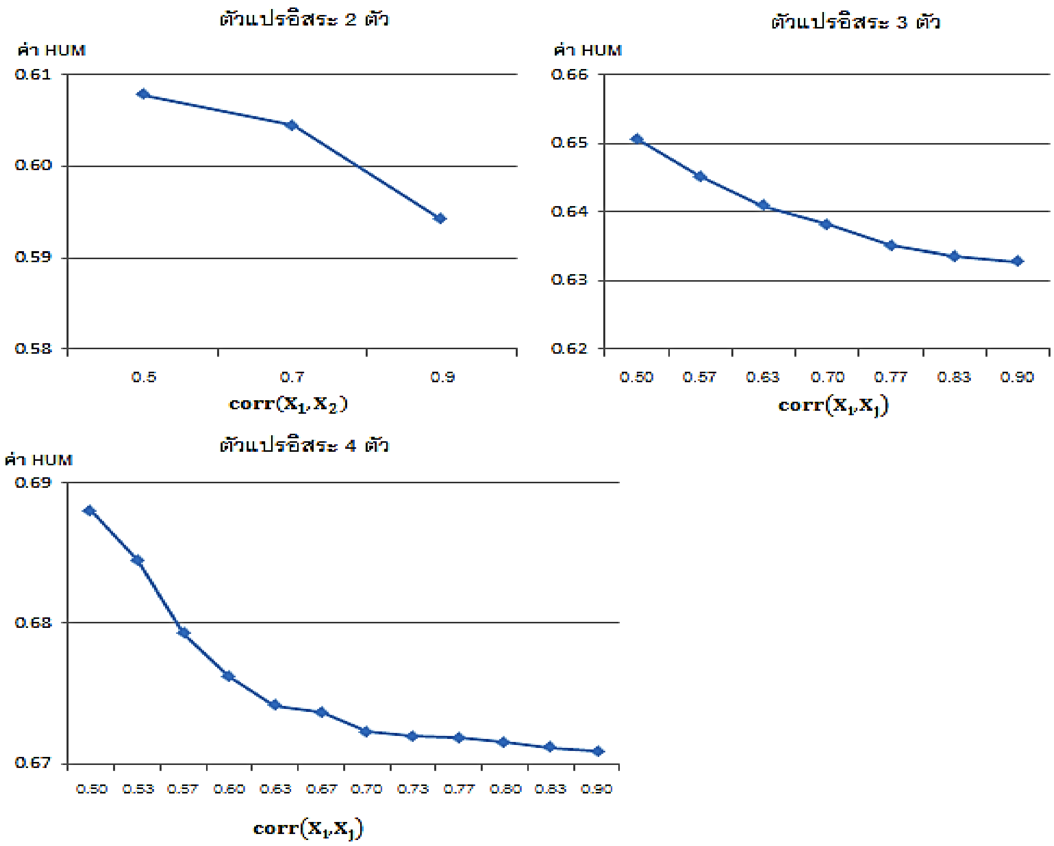
2.2 ศึกษาผลกระทบจากค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ระดับต่างๆ

กรณี $n_1 = 50,\ n_2 = 80,\ n_3 = 100,\ n_4 = 120$ ที่ค่า $\operatorname{corr}(X_I, X_j) = \operatorname{corr}(X_{i'}, X_{j'})$ $; i \neq j, i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$

ตารางที่ 2.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ระดับต่างๆ

ตัวแปรอิสระ 2 ตัว	$\mathrm{corr}(X_1, X_2)$	0.5	0.7	0.9				
	HUM	0.6079	0.6045	0.5941				
ตัวแปรอิสระ 3 ตัว	$\mathrm{corr}(X_i, X_j)$	0.50	0.57	0.63	0.70	0.77	0.83	0.90
	HUM	0.6507	0.6452	0.6409	0.6381	0.6350	0.6336	0.6327
ตัวแปรอิสระ 4 ตัว	$\mathrm{corr}(X_i, X_j)$	0.50	0.53	0.57	0.60	0.63	0.67	0.70
	HUM	0.6880	0.6845	0.6793	0.6762	0.6741	0.6736	0.6722
	$\mathrm{corr}(X_i, X_j)$	0.73	0.77	0.80	0.83	0.90		
	HUM	0.6719	0.6718	0.6715	0.6711	0.6709		

รูปที่ 2.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ระดับต่างๆ



จากตารางและรูปที่ 2.2 เมื่อค่าความสัมพันธ์ระหว่างตัวแปรอิสระในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า ถ้าค่าความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้นจะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มลดลง

จากผลการวิจัยข้างต้นพบว่า ค่าความสัมพันธ์ระหว่างตัวแปรอิสระแปรผกผันกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี เนื่องจากการที่ ตัวแปรอิสระมีความสัมพันธ์กัน ผิดสมมุติฐานเบื้องต้นของการวิเคราะห์ความถดถอย เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูงขึ้นทำให้ความสามารถในการทำนายกลุ่มของสมการถดถอยโลจิสติกลดลง

2.3 **ศึกษาผลกระทบจากจำนวนขนาดตัวอย่าง** แบ่งเป็น 2 กรณี

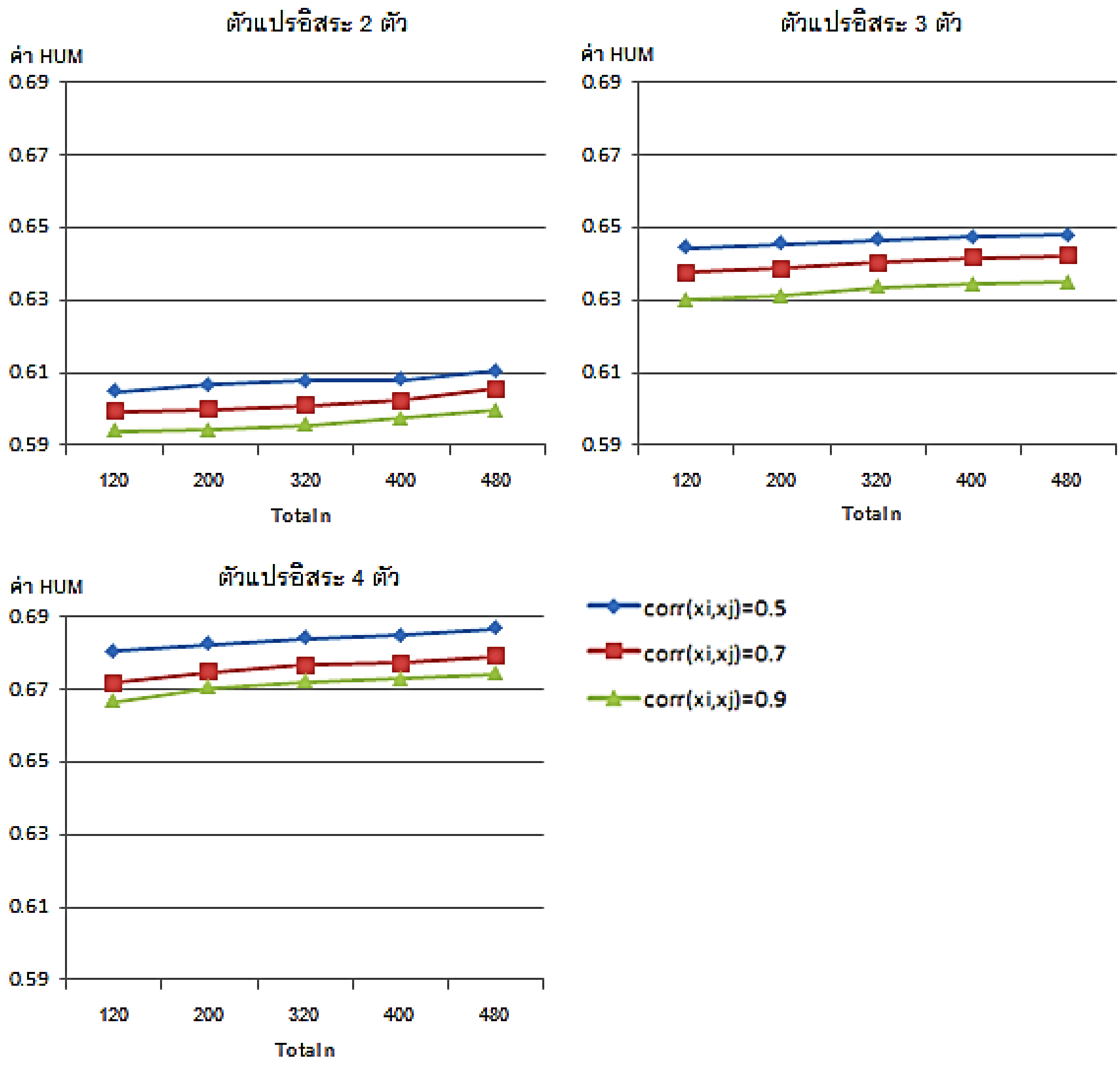
2.3.1 **เมื่อขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน**

กรณี $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})$; $i \neq j , i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$

ตารางที่ 2.3.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ขนาดตัวอย่างระดับต่างๆ

							$\mathrm{corr}(X_i, X_j)$		
	n_1	n_2	n_3	n_4	Total n	var(n)	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	30	30	30	30	120	0	0.6046	0.5992	0.5939
	50	50	50	50	200	0	0.6066	0.5997	0.5940
	80	80	80	80	320	0	0.6076	0.6009	0.5953
	100	100	100	100	400	0	0.6080	0.6023	0.5974
	120	120	120	120	480	0	0.6104	0.6054	0.5996
ตัวแปรอิสระ 3 ตัว	30	30	30	30	120	0	0.6445	0.6375	0.6302
	50	50	50	50	200	0	0.6455	0.6388	0.6311
	80	80	80	80	320	0	0.6465	0.6402	0.6332
	100	100	100	100	400	0	0.6474	0.6415	0.6346
	120	120	120	120	480	0	0.6482	0.6424	0.6350
ตัวแปรอิสระ 4 ตัว	30	30	30	30	120	0	0.6807	0.6720	0.6668
	50	50	50	50	200	0	0.6825	0.6748	0.6702
	80	80	80	80	320	0	0.6838	0.6767	0.6722
	100	100	100	100	400	0	0.6849	0.6774	0.6730
	120	120	120	120	480	0	0.6869	0.6789	0.6744

รูปที่ 2.3.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ขนาดตัวอย่าง ระดับต่างๆ



จากตารางและรูปที่ 2.3.1 เมื่อขนาดตัวอย่างในตัวแบบจำลองข้อมูลที่ทุกกลุ่มมีจำนวนตัวอย่างเท่ากันเปลี่ยนแปลง จะพบว่า เมื่อขนาดตัวอย่างมีจำนวนเพิ่มขึ้นจะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น นอกจากนี้เมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าเพิ่มขึ้น โดย $\mathrm{var}(n) = 0$ หมายถึง ขนาดตัวอย่างกลุ่มที่ i และ j เท่ากัน โดยที่ $i \neq j; i, j = 1, 2, \dots, k$

จากผลการวิจัยข้างต้นพบว่า ขนาดตัวอย่างกรณีที่จำนวนตัวอย่างเท่ากันทุกกลุ่มแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี

2.3.2 เมื่อขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน

กรณี $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})$; $i \neq j$, $i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$ โดยใช้ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ มาเปรียบเทียบกัน ซึ่งสามารถคำนวณหาความแปรปรวนของขนาดตัวอย่างได้จากสูตร $\mathrm{var}(n) = \frac{\sum_{Y=1}^{k}(n_Y - \bar{n})^2}{k-1}$ โดยที่

- ค่า $\mathrm{var}(n)$ มีค่าน้อย หมายถึง ตัวอย่างแต่ละกลุ่มมีจำนวนใกล้เคียงกัน
- ค่า $\mathrm{var}(n)$ มีค่ามาก หมายถึง ตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น

ค่า $\mathrm{var}(n) = [0, \infty)$ เมื่อค่า $\mathrm{var}(n) = 0$ หมายถึง ขนาดตัวอย่างแต่ละกลุ่มมีจำนวนเท่ากัน ในการวิจัยครั้งนี้จะทำการศึกษาเฉพาะกรณีขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน $(\mathrm{var}(n) = 0)$ และขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน $(\mathrm{var}(n) \neq 0)$ โดยขนาดตัวอย่างจะเรียงจำนวนกลุ่มจากน้อยไปมากเท่านั้น

ค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่าง (coefficient of variation: CV_n) คือ ค่าร้อยละที่ข้อมูลเบี่ยงเบนรอบๆค่าขนาดตัวอย่างเฉลี่ย ซึ่งใช้วัดการกระจายสัมพัทธ์ของข้อมูลเพื่อเปรียบเทียบการกระจายของข้อมูลตั้งแต่สองชุดขึ้นไป ซึ่งมีสูตร คือ

$CV_n = \frac{SD}{\bar{n}} \times 100$ โดยที่

- ค่า CV_n มีค่าน้อย หมายถึง ค่าการกระจายของข้อมูลรอบๆค่าขนาดตัวอย่างเฉลี่ยน้อย หรือตัวอย่างแต่ละกลุ่มมีจำนวนใกล้เคียงกัน
- ค่า CV_n มีค่ามาก หมายถึง ค่าการกระจายของข้อมูลรอบๆค่าขนาดตัวอย่างเฉลี่ยมาก หรือ ตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น

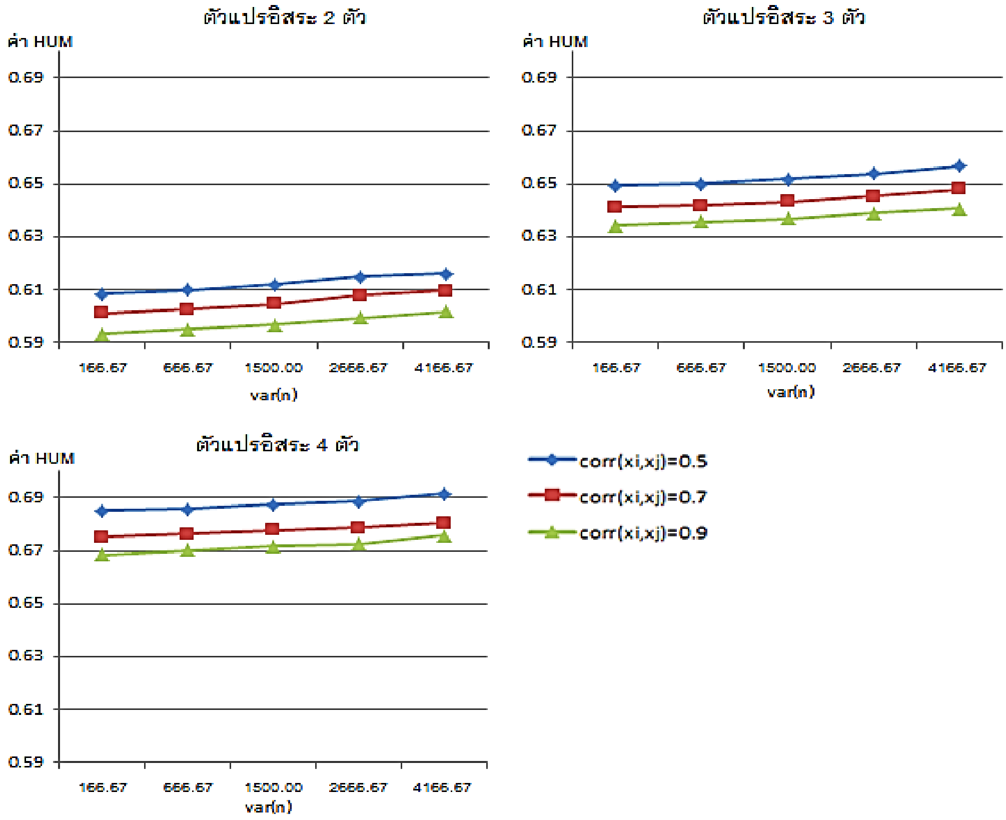
<u>ตารางที่ 2.3.2</u> แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ

									$\mathrm{corr}(X_i, X_j)$		
	n_1	n_2	n_3	n_4	Total n	$\bar{n}$	$\mathrm{var}(n)$	CV_n	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	30	40	50	60	180	45	166.67	28.69	0.6084	0.6011	0.5933
	30	50	70	90	240	60	666.67	43.03	0.6099	0.6026	0.5950
	30	60	90	120	300	75	1500.00	51.64	0.6120	0.6046	0.5969
	30	70	110	150	360	90	2666.67	57.38	0.6148	0.6079	0.5994
	30	80	130	180	420	105	4166.67	61.48	0.6162	0.6098	0.6017

ตารางที่ 2.3.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ(ต่อ)

									$corr(X_i, X_j)$		
	n_1	n_2	n_3	n_4	Total n	$\bar{n}$	var(n)	CV_n	0.5	0.7	0.9
ตัวแปรอิสระ 3 ตัว	30	40	50	60	180	45	166.67	28.69	0.6494	0.6412	0.6344
	30	50	70	90	240	60	666.67	43.03	0.6501	0.6419	0.6356
	30	60	90	120	300	75	1500.00	51.64	0.6517	0.6433	0.6369
	30	70	110	150	360	90	2666.67	57.38	0.6540	0.6453	0.6390
	30	80	130	180	420	105	4166.67	61.48	0.6567	0.6480	0.6406
ตัวแปรอิสระ 4 ตัว	30	40	50	60	180	45	166.67	28.69	0.6852	0.6753	0.6685
	30	50	70	90	240	60	666.67	43.03	0.6858	0.6765	0.6702
	30	60	90	120	300	75	1500.00	51.64	0.6874	0.6780	0.6716
	30	70	110	150	360	90	2666.67	57.38	0.6887	0.6788	0.6724
	30	80	130	180	420	105	4166.67	61.48	0.6915	0.6806	0.6757

รูปที่ 2.3.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ



จากตารางและรูปที่ 2.3.2 กรณีขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน เมื่อค่าความแปรปรวนของขนาดตัวอย่างในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า เมื่อค่าความแปรปรวนของขนาดตัวอย่างเพิ่มขึ้น เมื่อตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น หรือเมื่อพิจารณาจากค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่างที่เพิ่มขึ้น เมื่อมีค่าการกระจายของข้อมูลรอบๆค่าเฉลี่ยเพิ่มขึ้น จะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้ง อาร์โอซีมีค่าสูงขึ้น โดย $\mathrm{var}(\mathrm{n}) \neq 0$ หมายถึง ขนาดตัวอย่างกลุ่มที่ i และ j ไม่เท่ากัน โดยที่ $\mathrm{i} \neq \mathrm{j}; \mathrm{i},\ \mathrm{j} = 1, 2, \dots, \mathrm{k}$

จากผลการวิจัยข้างต้นพบว่า กรณีที่ขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน ความแปรปรวนของขนาดตัวอย่างหรือค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่างแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี โดยการที่ขนาดตัวอย่างในแต่ละกลุ่มมีจำนวนแตกต่างกันเพิ่มขึ้นทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น ทั้งนี้อาจเนื่องมากจากการที่ขนาดตัวอย่างรวมมีจำนวนเพิ่มขึ้น เช่นเดียวกับกรณีที่ขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน นอกจากนี้ยังสังเกตเห็นว่า ในกรณีที่ขนาดตัวอย่างรวมเท่ากัน เมื่อเปรียบเทียบระหว่างกรณีที่มีตัวอย่างในแต่ละกลุ่มไม่เท่ากันและกรณีที่ตัวอย่างในแต่ละกลุ่มเท่ากัน ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่มีตัวอย่างในแต่ละกลุ่มไม่เท่ากันมีค่าสูงกว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่มีตัวอย่างในแต่ละกลุ่มเท่ากัน

2.4 **ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย** $(\mathrm{var}(\mu))$

กรณี $\mathrm{n}_1 = 50,\ \mathrm{n}_2 = 80,\ \mathrm{n}_3 = 100,\ \mathrm{n}_4 = 120$ ที่ค่า $\mathrm{corr}(\mathrm{X_i}, \mathrm{X_j}) = \mathrm{corr}(\mathrm{X_{i'}}, \mathrm{X_{j'}})$

$; \mathrm{i} \neq \mathrm{j}, \mathrm{i}' \neq \mathrm{j}'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$

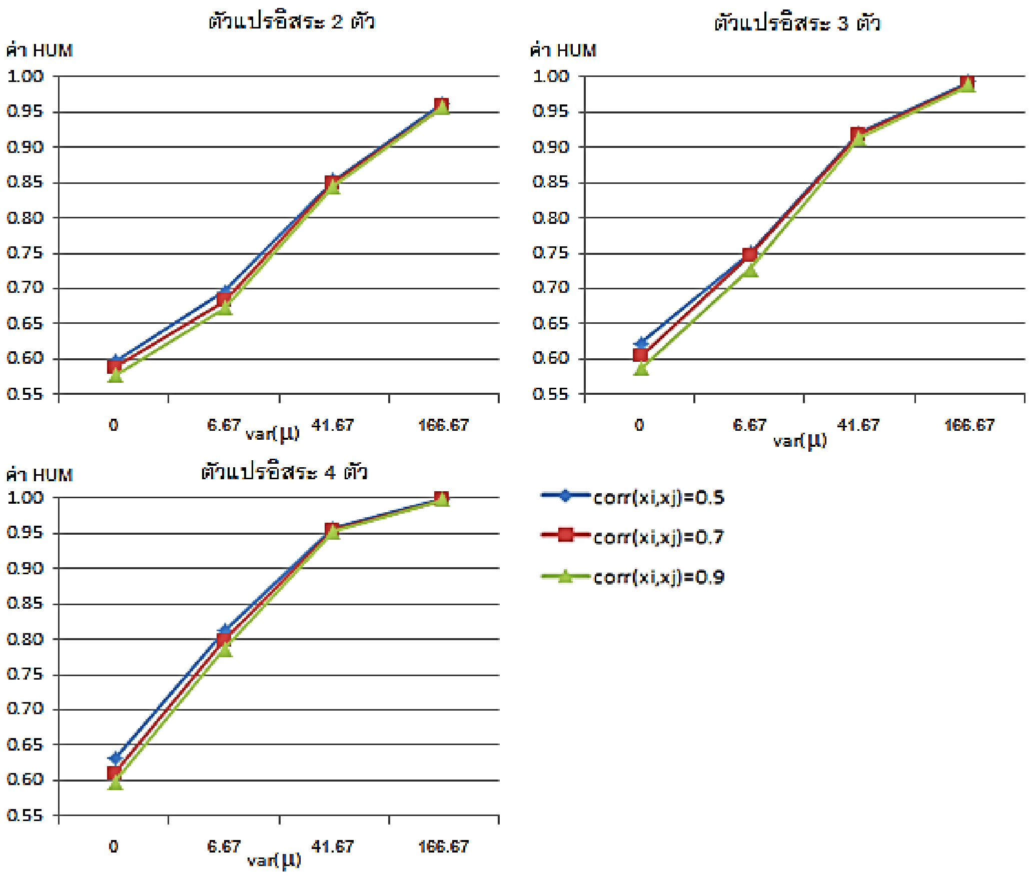
ตารางที่ 2.4 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของค่าเฉลี่ยระดับต่างๆ

						$\mathrm{corr}(\mathrm{X_i}, \mathrm{X_j})$		
	μ_1	μ_2	μ_3	μ_4	$\mathrm{var}(\mu)$	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	0	0	0	0	0	0.5970	0.5891	0.5777
	0	2	4	6	6.67	0.6954	0.6822	0.6723
	0	5	10	15	41.67	0.8534	0.8503	0.8462
	0	10	20	30	166.67	0.9618	0.9590	0.9571

ตารางที่ 2.4 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของค่าเฉลี่ยระดับต่างๆ(ต่อ)

						$corr(X_i, X_j)$		
	μ_1	μ_2	μ_3	μ_4	$var(\mu)$	0.5	0.7	0.9
ตัวแปรอิสระ 3 ตัว	0	0	0	0	0	0.6222	0.6045	0.5860
	0	2	4	6	6.67	0.7512	0.7473	0.7276
	0	5	10	15	41.67	0.9205	0.9179	0.9134
	0	10	20	30	166.67	0.9920	0.9902	0.9880
ตัวแปรอิสระ 4 ตัว	0	0	0	0	0	0.6310	0.6101	0.5983
	0	2	4	6	6.67	0.8113	0.7992	0.7871
	0	5	10	15	41.67	0.9572	0.9551	0.9520
	0	10	20	30	166.67	0.9985	0.9983	0.9980

รูปที่ 2.4 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของค่าเฉลี่ยระดับต่างๆ



จากตารางและรูปที่ 2.4 เมื่อความแปรปรวนของค่าเฉลี่ย เปลี่ยนแปลง จะพบว่า ถ้าความแปรปรวนของค่าเฉลี่ยเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น โดย $\mathrm{var}(\mu)$ แสดงลักษณะการกระจายของค่าเฉลี่ยกลุ่มที่ i และ j โดยที่ $i \neq j$; $i, j = 1,2, \dots, k$ กรณี $\mathrm{var}(\mu) = 0$ หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าเท่ากันจะได้ว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าต่ำที่สุด ทุกระดับความแปรปรวน เนื่องจากข้อมูลทั้ง 4 กลุ่ม ซ้อนทับกันมาก ทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าน้อย และเมื่อค่าเฉลี่ยของแต่ละกลุ่มห่างกันมากขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะเพิ่มขึ้นเรื่อยๆ แต่ไม่เกิน 1

จากผลการวิจัยพบว่า ความแปรปรวนของค่าเฉลี่ยแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี คือ เมื่อค่าเฉลี่ยของแต่ละกลุ่มที่ค่าเท่ากัน ($\mathrm{var}(\mu) = 0$) ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าต่ำที่สุด และเมื่อค่าความแปรปรวนของค่าเฉลี่ยมากขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าเพิ่มขึ้น นอกจากนี้ยังพบว่า เมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้นค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มเพิ่มขึ้น เนื่องมาจากการที่ตัวแปรอิสระเพิ่มมากทำให้ความสามารถในการอธิบายตัวแปรตามเพิ่มมากขึ้น

3. ผลการวิจัยกรณีมีจำนวนกลุ่ม 5 กลุ่ม ที่มีผลกระทบต่อปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (HUM) สำหรับตัวแบบโลจิสติกพหุนาม

3.1 ศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย $\boldsymbol{\beta}$ ณ ระดับต่างๆ

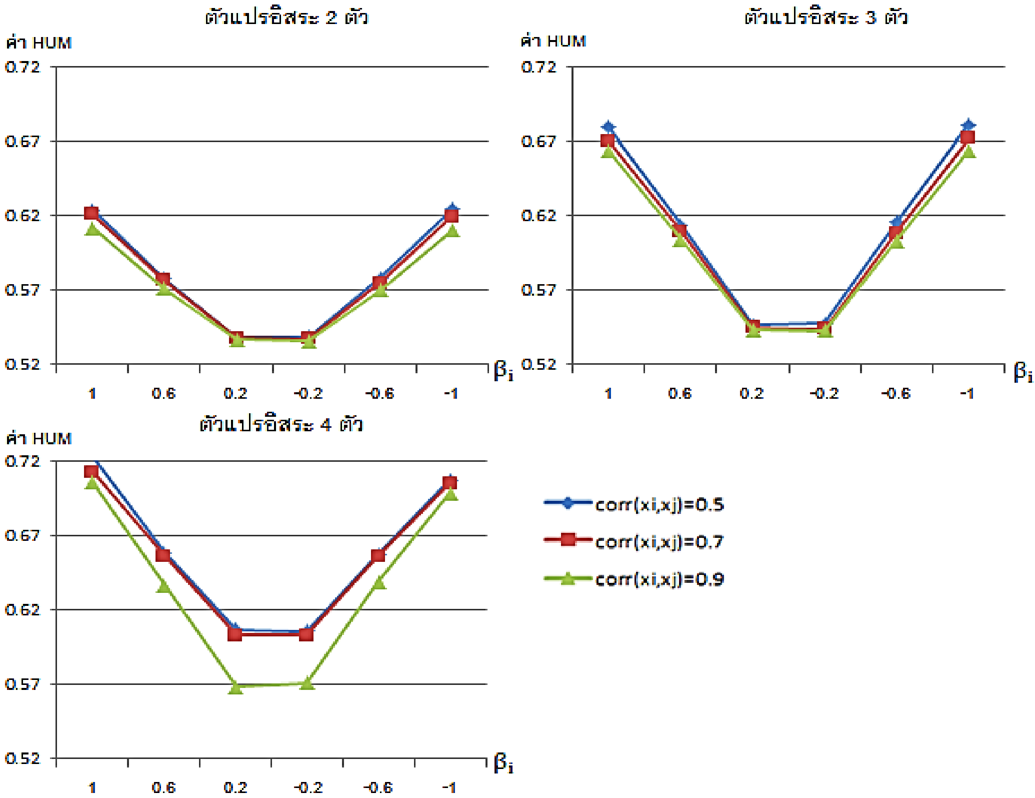
กรณี $n_1 = 50, n_2 = 80, n_3 = 100, n_4 = 120, n_5 = 130$

ที่ค่า $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})$; $i \neq j$, $i' \neq j'$ ที่ค่า β_0=0

ตารางที่ 3.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าสัมประสิทธิ์ความถดถอย β ณ ระดับต่างๆ

		$\mathrm{corr}(X_i, X_j)$		
		0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	$\beta_1=\beta_2=1$	0.6237	0.6211	0.6118
	$\beta_1=\beta_2=0.6$	0.5775	0.5765	0.5711
	$\beta_1=\beta_2=0.2$	0.5382	0.5381	0.5369
	$\beta_1=\beta_2=-0.2$	0.5387	0.5372	0.5360
	$\beta_1=\beta_2=-0.6$	0.5783	0.5750	0.5699
	$\beta_1=\beta_2=-1$	0.6244	0.6197	0.6105
ตัวแปรอิสระ 3 ตัว	$\beta_1=\beta_2=\beta_3=1$	0.6797	0.6707	0.6637
	$\beta_1=\beta_2=\beta_3=0.6$	0.6142	0.6100	0.6041
	$\beta_1=\beta_2=\beta_3=0.2$	0.5466	0.5453	0.5432
	$\beta_1=\beta_2=\beta_3=-0.2$	0.5478	0.5440	0.5427
	$\beta_1=\beta_2=\beta_3=-0.6$	0.6155	0.6089	0.6034
	$\beta_1=\beta_2=\beta_3=-1$	0.6815	0.6724	0.6634
ตัวแปรอิสระ 4 ตัว	$\beta_1=\beta_2=\beta_3=\beta_4=1$	0.7230	0.7130	0.7059
	$\beta_1=\beta_2=\beta_3=\beta_4=0.6$	0.6586	0.6568	0.6371
	$\beta_1=\beta_2=\beta_3=\beta_4=0.2$	0.6068	0.6034	0.5682
	$\beta_1=\beta_2=\beta_3=\beta_4=-0.2$	0.6059	0.6031	0.5712
	$\beta_1=\beta_2=\beta_3=\beta_4=-0.6$	0.6577	0.6567	0.6390
	$\beta_1=\beta_2=\beta_3=\beta_4=-1$	0.7076	0.7057	0.6989

รูปที่ 3.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าสัมประสิทธิ์ความถดถอย β ณ ระดับต่างๆ



จากตารางและรูปที่ 3.1 เมื่อค่าพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ ในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า เมื่อค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ เพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าเพิ่มขึ้น

จากผลการวิจัยข้างต้นพบว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีแปรผันตรงกับค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1, \ldots, \beta_p$ คือ เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามมีค่าเพิ่มขึ้น ส่งผลให้ ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าเพิ่มขึ้น หมายความว่าตัวแปรอิสระสามารถอธิบายตัวแปรตามได้มากขึ้น ทำให้ความสามารถในการทำนายกลุ่มสูงขึ้น

3.2 ศึกษาผลกระทบจากค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ

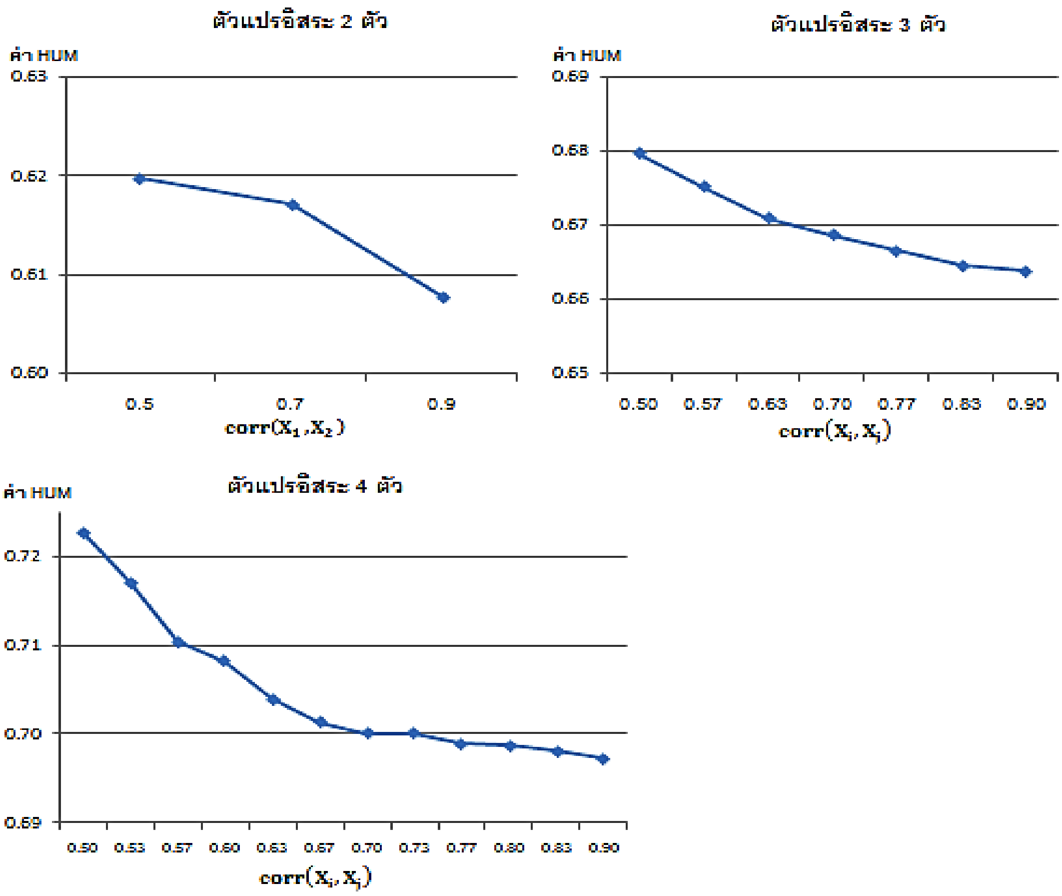
กรณี $n_1 = 50, n_2 = 80, n_3 = 100, n_4 = 120, n_5 = 130$ ที่ค่า

$\mathrm{corr}(X_I, X_j) = \mathrm{corr}(X_{i'}, X_{j'})\ ; i \neq j, i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$

ตารางที่ 3.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ระดับต่างๆ

ตัวแปร อิสระ 2 ตัว	$corr(X_1,X_2)$	0.5	0.7	0.9				
	HUM	0.6237	0.6211	0.6118				
ตัวแปร อิสระ3 ตัว	$corr(X_i,X_j)$	0.50	0.57	0.63	0.70	0.77	0.83	0.90
	HUM	0.6797	0.6751	0.6709	0.6686	0.6665	0.6645	0.6637
ตัวแปรอิสระ 4 ตัว	$corr(X_i,X_j)$	0.50	0.53	0.57	0.60	0.63	0.67	0.70
	HUM	0.7227	0.7172	0.7104	0.7081	0.7039	0.7013	0.7001
	$corr(X_i,X_j)$	0.73	0.77	0.80	0.83	0.90		
	HUM	0.7000	0.6990	0.6987	0.6981	0.6972		

รูปที่ 3.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความสัมพันธ์ระหว่างตัวแปรอิสระ ระดับต่างๆ



จากตารางและรูปที่ 3.2 เมื่อค่าความสัมพันธ์ระหว่างตัวแปรอิสระในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า ถ้าค่าความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้นจะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มลดลง

จากผลการวิจัยข้างต้นพบว่า ค่าความสัมพันธ์ระหว่างตัวแปรอิสระแปรผกผันกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี เนื่องจากการที่ ตัวแปรอิสระมีความสัมพันธ์กัน ผิดสมมุติฐานเบื้องต้นของการวิเคราะห์ความถดถอย เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูงขึ้นทำให้ความสามารถในการทำนายกลุ่มของสมการถดถอยโลจิสติกลดลง

3.3 **ศึกษาผลกระทบจากจำนวนขนาดตัวอย่าง** แบ่งเป็น 2 กรณี

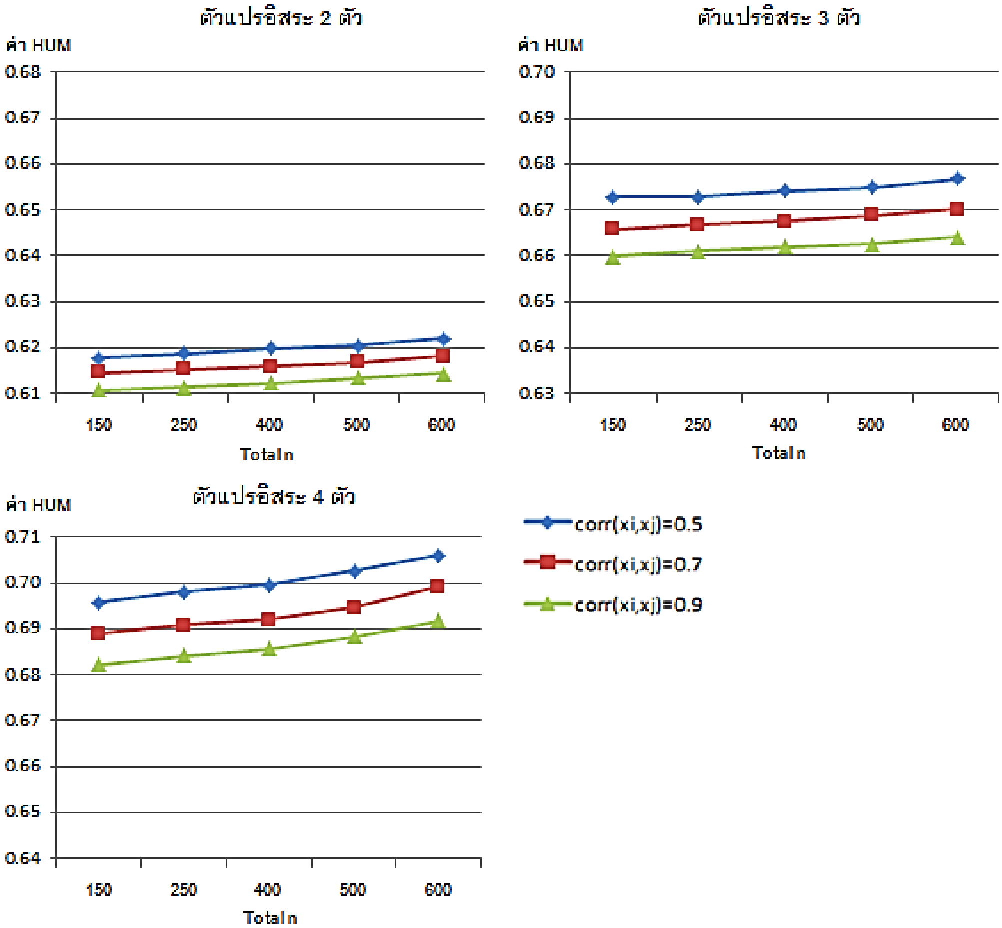
3.3.1 เมื่อขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน

กรณี $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})$; $i \neq j$, $i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$

ตารางที่ 3.3.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ขนาดตัวอย่างระดับต่างๆ

								$\mathrm{corr}(X_i, X_j)$		
	n_1	n_2	n_3	n_4	n_5	Total n	$\mathrm{var}(n)$	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	30	30	30	30	30	150	0	0.6180	0.6146	0.6108
	50	50	50	50	50	250	0	0.6187	0.6153	0.6115
	80	80	80	80	80	400	0	0.6198	0.6161	0.6123
	100	100	100	100	100	500	0	0.6205	0.6170	0.6134
	120	120	120	120	120	600	0	0.6221	0.6183	0.6143
ตัวแปรอิสระ 3 ตัว	30	30	30	30	30	150	0	0.6728	0.6659	0.6599
	50	50	50	50	50	250	0	0.6730	0.6668	0.6610
	80	80	80	80	80	400	0	0.6741	0.6676	0.6619
	100	100	100	100	100	500	0	0.6750	0.6689	0.6625
	120	120	120	120	120	600	0	0.6769	0.6703	0.6640
ตัวแปรอิสระ 4 ตัว	30	30	30	30	30	150	0	0.6957	0.6891	0.6822
	50	50	50	50	50	250	0	0.6980	0.6908	0.6840
	80	80	80	80	80	400	0	0.6996	0.6920	0.6858
	100	100	100	100	100	500	0	0.7026	0.6946	0.6884
	120	120	120	120	120	600	0	0.7060	0.6991	0.6916

รูปที่ 3.3.1 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ขนาดตัวอย่าง ระดับต่างๆ



จากตารางและรูปที่ 3.3.1 เมื่อขนาดตัวอย่างในตัวแบบจำลองข้อมูลที่ทุกกลุ่มมีจำนวนตัวอย่างเท่ากันเปลี่ยนแปลง จะพบว่า เมื่อขนาดตัวอย่างมีจำนวนเพิ่มขึ้นจะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น นอกจากนี้เมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าเพิ่มขึ้น โดย $\mathrm{var}(n) = 0$ หมายถึง ขนาดตัวอย่างกลุ่มที่ i และ j เท่ากัน โดยที่ $i \neq j; i, j = 1, 2, \dots, k$

จากผลการวิจัยข้างต้นพบว่า ขนาดตัวอย่างกรณีที่จำนวนตัวอย่างเท่ากันทุกกลุ่มแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี

3.3.2 **เมื่อขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน**

กรณี $\mathrm{corr}(X_i, X_j) = \mathrm{corr}(X_{i'}, X_{j'})\,; i \neq j\,, i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$ โดยใช้ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ มาเปรียบเทียบกัน ซึ่งสามารถ

คำนวณหาความแปรปรวนของขนาดตัวอย่างได้จากสูตร $\mathrm{var}(n) = \frac{\sum_{Y=1}^{k}(n_Y - \bar{n})^2}{k-1}$ โดยที่

- ค่า $\mathrm{var}(n)$ มีค่าน้อย หมายถึง ตัวอย่างแต่ละกลุ่มมีจำนวนใกล้เคียงกัน
- ค่า $\mathrm{var}(n)$ มีค่ามาก หมายถึง ตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น

ค่า $\mathrm{var}(n) = [0, \infty)$ เมื่อค่า $\mathrm{var}(n) = 0$ หมายถึง ขนาดตัวอย่างแต่ละกลุ่มมีจำนวนเท่ากัน ในการวิจัยครั้งนี้จะทำการศึกษาเฉพาะกรณีขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน $(\mathrm{var}(n) = 0)$ และขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน $(\mathrm{var}(n) \neq 0)$ โดยขนาดตัวอย่างจะเรียงจำนวนกลุ่มจากน้อยไปมากเท่านั้น

ค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่าง (coefficient of variation*: CV_n*) คือ ค่าร้อยละที่ข้อมูลเบี่ยงเบนรอบๆค่าขนาดตัวอย่างเฉลี่ย ซึ่งใช้วัดการกระจายสัมพัทธ์ของข้อมูลเพื่อเปรียบเทียบการกระจายของข้อมูลตั้งแต่สองชุดขึ้นไป ซึ่งมีสูตร คือ

$CV_n = \frac{SD}{\bar{n}} \times 100$ โดยที่

- ค่า CV_n มีค่าน้อย หมายถึง ค่าการกระจายของข้อมูลรอบๆค่าขนาดตัวอย่างเฉลี่ยน้อย หรือตัวอย่างแต่ละกลุ่มมีจำนวนใกล้เคียงกัน
- ค่า CV_n มีค่ามาก หมายถึง ค่าการกระจายของข้อมูลรอบๆค่าขนาดตัวอย่างเฉลี่ยมาก หรือ ตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น

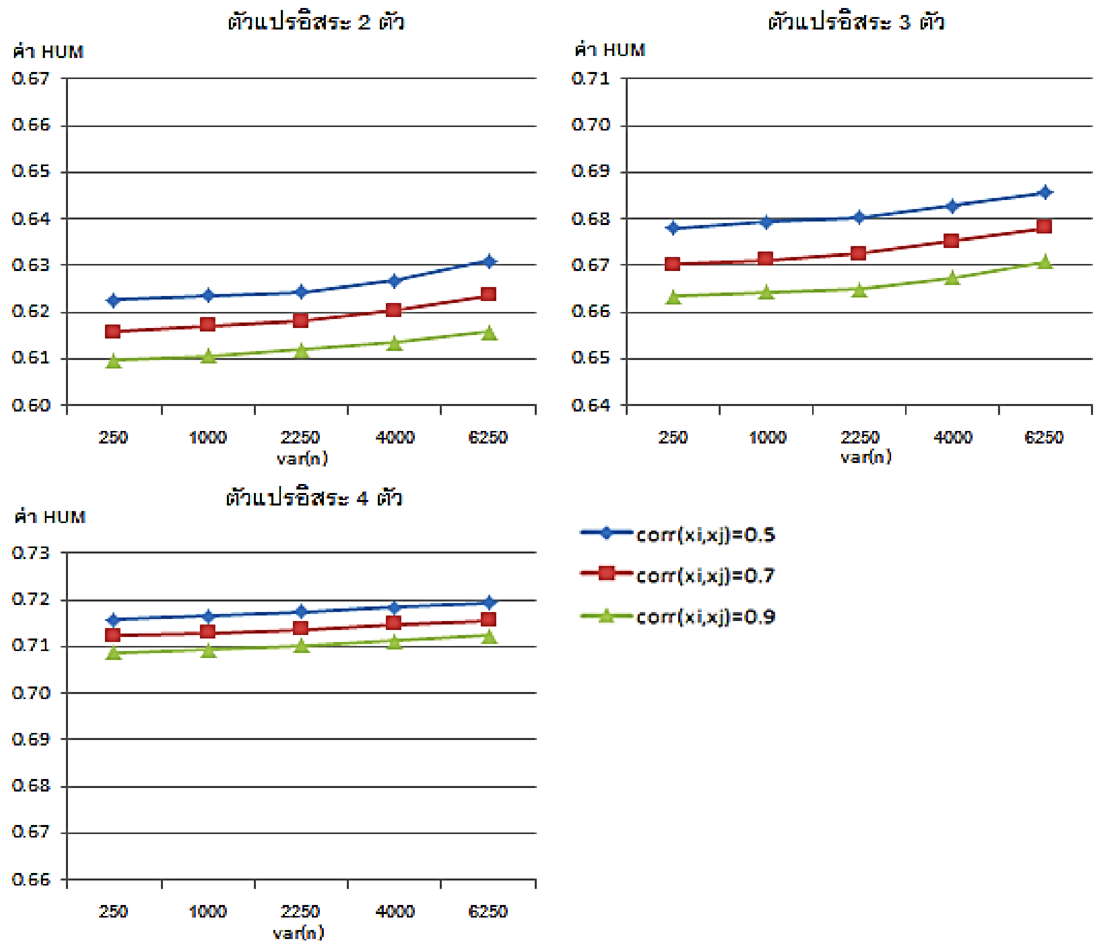
<u>ตารางที่ 3.3.2</u> แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ

										$\mathrm{corr}(X_i, X_j)$		
	n_1	n_2	n_3	n_4	n_5	Total n	$\bar{n}$	$\mathrm{var}(n)$	CV_n	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	30	40	50	60	70	250	50	250	31.62	0.6227	0.6158	0.6098
	30	50	70	90	110	350	70	1000	45.18	0.6235	0.6171	0.6107
	30	60	90	120	150	450	90	2250	52.70	0.6244	0.6182	0.6121
	30	70	110	150	190	550	110	4000	57.50	0.6269	0.6205	0.6136
	30	80	130	180	230	650	130	6250	60.81	0.6310	0.6236	0.6157
ตัวแปรอิสระ 3 ตัว	30	40	50	60	70	250	50	250	31.62	0.6780	0.6704	0.6634
	30	50	70	90	110	350	70	1000	45.18	0.6795	0.6712	0.6643
	30	60	90	120	150	450	90	2250	52.70	0.6804	0.6726	0.6650
	30	70	110	150	190	550	110	4000	57.50	0.6829	0.6751	0.6675
	30	80	130	180	230	650	130	6250	60.81	0.6858	0.6782	0.6708

ตารางที่ 3.3.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ(ต่อ)

										$corr(X_i, X_j)$		
	n_1	n_2	n_3	n_4	n_5	Total n	$\bar{n}$	$var(n)$	CV_n	0.5	0.7	0.9
ตัวแปรอิสระ 4 ตัว	30	40	50	60	70	250	50	250	31.62	0.7158	0.7124	0.7088
	30	50	70	90	110	350	70	1000	45.18	0.7165	0.7130	0.7093
	30	60	90	120	150	450	90	2250	52.70	0.7174	0.7138	0.7103
	30	70	110	150	190	550	110	4000	57.50	0.7185	0.7148	0.7112
	30	80	130	180	230	650	130	6250	60.81	0.7194	0.7157	0.7123

รูปที่ 3.3.2 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของขนาดตัวอย่างระดับต่างๆ



จากตารางและรูปที่ 3.3.2 กรณีขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน เมื่อค่าความแปรปรวนของขนาดตัวอย่างในตัวแบบจำลองข้อมูลเปลี่ยนแปลง จะพบว่า เมื่อค่าความแปรปรวนของขนาดตัวอย่างเพิ่มขึ้น เมื่อตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น หรือเมื่อพิจารณาจากค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่างที่เพิ่มขึ้น เมื่อมีค่าการกระจายของข้อมูลรอบๆค่าเฉลี่ยเพิ่มขึ้น จะทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้ง อาร์โอซีมีค่าสูงขึ้น โดย $\mathrm{var}(n) \neq 0$ หมายถึง ขนาดตัวอย่างกลุ่มที่ i และ j ไม่เท่ากัน โดยที่ $i \neq j; i,\ j = 1, 2, \dots, k$

จากผลการวิจัยข้างต้นพบว่า กรณีที่ขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน ความแปรปรวนของขนาดตัวอย่างหรือค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่างแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี โดยการที่ขนาดตัวอย่างในแต่ละกลุ่มมีจำนวนแตกต่างกันเพิ่มขึ้นทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น ทั้งนี้อาจเนื่องมากจากการที่ขนาดตัวอย่างรวมมีจำนวนเพิ่มขึ้น เช่นเดียวกับกรณีที่ขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน นอกจากนี้ยังสังเกตเห็นว่า ในกรณีที่ขนาดตัวอย่างรวมเท่ากัน เมื่อเปรียบเทียบระหว่างกรณีที่มีตัวอย่างในแต่ละกลุ่มไม่เท่ากันและกรณีที่ตัวอย่างในแต่ละกลุ่มเท่ากัน ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่มีตัวอย่างในแต่ละกลุ่มไม่เท่ากันมีค่าสูงกว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่มีตัวอย่างในแต่ละกลุ่มเท่ากัน

3.4 **ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย** $(\mathrm{var}(\mu))$

กรณี $n_1 = 50, n_2 = 80, n_3 = 100, n_4 = 120, n_5 = 130$ ที่ค่า

$\mathrm{corr}(X_I, X_j) = \mathrm{corr}(X_{i'}, X_{j'})\ ; i \neq j, i' \neq j'$ ที่ค่า β_0=0 และ $\beta_1 = \cdots = \beta_p = 1$

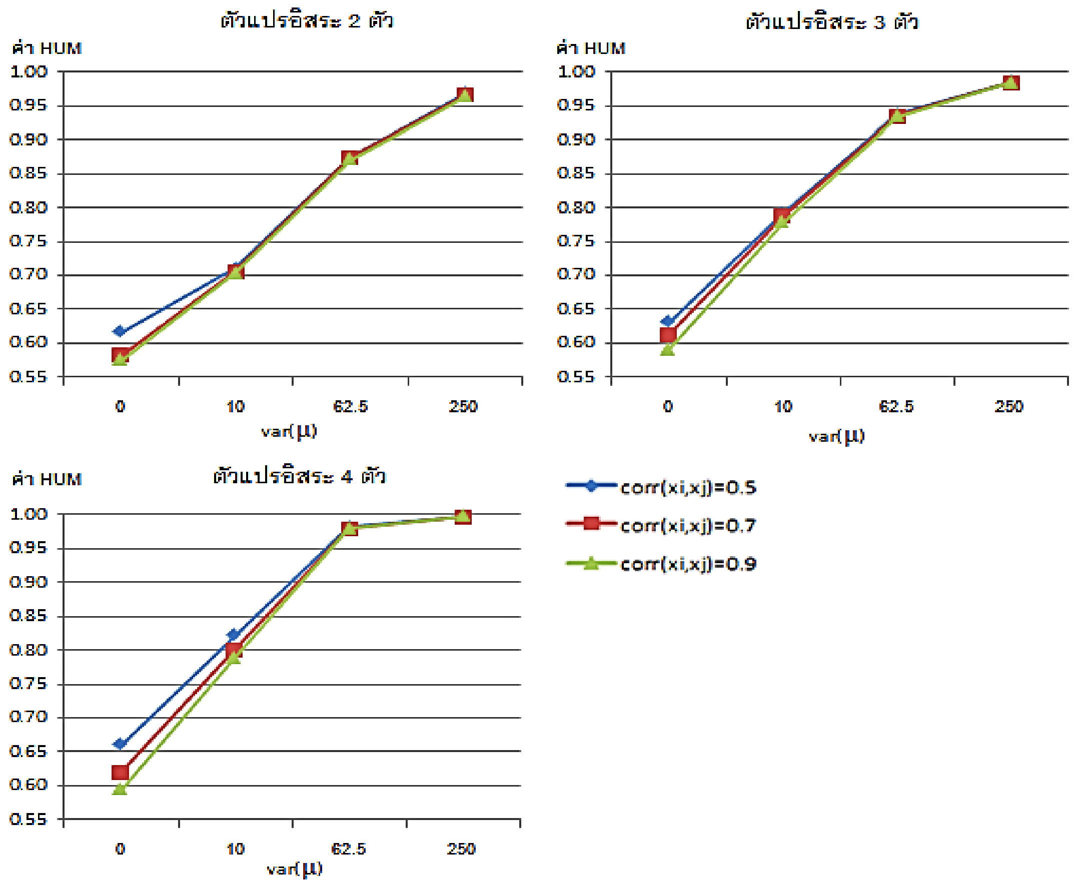
<u>ตารางที่ 3.4</u> แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของค่าเฉลี่ยระดับต่างๆ

							$\mathrm{corr}(X_i, X_j)$		
	μ_1	μ_2	μ_3	μ_4	μ_5	$\mathrm{var}(\mu)$	0.5	0.7	0.9
ตัวแปรอิสระ 2 ตัว	0	0	0	0	0	0	0.6168	0.5823	0.5751
	0	2	4	6	8	10	0.7112	0.7066	0.7031
	0	5	10	15	20	62.5	0.8757	0.8742	0.8712
	0	10	20	30	40	250	0.9682	0.9656	0.9634

ตารางที่ 3.4 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของค่าเฉลี่ยระดับต่างๆ(ต่อ)

							$\mathrm{corr}(X_i, X_j)$		
	μ_1	μ_2	μ_3	μ_4	μ_5	$\mathrm{var}(\mu)$	0.5	0.7	0.9
ตัวแปรอิสระ 3 ตัว	0	0	0	0	0	0	0.6311	0.6129	0.5900
	0	2	4	6	8	10	0.7917	0.7862	0.7780
	0	5	10	15	20	62.5	0.9386	0.9356	0.9335
	0	10	20	30	40	250	0.9858	0.9852	0.9846
ตัวแปรอิสระ 4 ตัว	0	0	0	0	0	0	0.6595	0.6196	0.5933
	0	2	4	6	8	10	0.8218	0.8009	0.7881
	0	5	10	15	20	62.5	0.9818	0.9799	0.9793
	0	10	20	30	40	250	0.9961	0.9967	0.9984

รูปที่ 3.4 แสดงค่า HUM กรณีมีตัวแปรอิสระ 2, 3 และ 4 ตัว ณ ค่าความแปรปรวนของค่าเฉลี่ยระดับต่างๆ



จากตารางและรูปที่ 3.4 เมื่อความแปรปรวนของค่าเฉลี่ย เปลี่ยนแปลง จะพบว่า ถ้าความแปรปรวนของค่าเฉลี่ยเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น โดย $\mathrm{var}(\mu)$ แสดงลักษณะการกระจายของค่าเฉลี่ยกลุ่มที่ i และ j โดยที่ $i \neq j$ $; i, j = 1,2, \dots, k$ กรณี $\mathrm{var}(\mu) = 0$ หมายถึง ค่าเฉลี่ยของแต่ละกลุ่มมีค่าเท่ากันจะได้ว่าค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าต่ำที่สุด ทุกระดับความแปรปรวน เนื่องจากข้อมูลทั้ง 5 กลุ่ม ซ้อนทับกันมาก ทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าน้อย และเมื่อค่าเฉลี่ยของแต่ละกลุ่มห่างกันมากขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะเพิ่มขึ้นเรื่อยๆ แต่ไม่เกิน 1

จากผลการวิจัยพบว่า ความแปรปรวนของค่าเฉลี่ยแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี คือ เมื่อค่าเฉลี่ยของแต่ละกลุ่มที่ค่าเท่ากัน ($\mathrm{var}(\mu) = 0$) ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าต่ำที่สุด และเมื่อค่าความแปรปรวนของค่าเฉลี่ยมากขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าเพิ่มขึ้น นอกจากนี้ยังพบว่า เมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้นค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มเพิ่มขึ้น เนื่องมาจากการที่ตัวแปรอิสระเพิ่มมากทำให้ความสามารถในการอธิบายตัวแปรตามเพิ่มมากขึ้น

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย, ค่าระดับความสัมพันธ์ของตัวแปรอิสระ, ขนาดตัวอย่าง และความแปรปรวนของค่าเฉลี่ย ที่มีต่อค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (Hypervolume Under the Manifold (HUM)) ในกรณีที่มีตัวแปรตาม 3, 4 และ 5 กลุ่ม สำหรับข้อมูลในงานวิจัยนี้ ทำการจำลองข้อมูลและวิเคราะห์ผลโดยใช้โปรแกรม R ซึ่งได้ทำการศึกษาภายใต้เงื่อนไขต่างๆดังนี้

1. ตัวแบบการถดถอยโลจิสติกพหุนาม มี 3 กรณี
   - กรณีมีตัวแปรอิสระ (X) 2 ตัว
   - กรณีมีตัวแปรอิสระ (X) 3 ตัว
   - กรณีมีตัวแปรอิสระ (X) 4 ตัว

   โดยตัวแปรอิสระเป็นตัวแปรเชิงปริมาณที่มีการแจกแจงเริ่มต้นมาจากการแจกแจงแบบปกติ ศึกษากรณี มีพารามิเตอร์ค่าเฉลี่ย คือ μ เป็นค่าคงที่ และมีค่าความแปรปรวน คือ σ^2 เท่ากับ 1 ตามระดับความสัมพันธ์ของตัวแปรอิสระที่กำหนดไว้ คือ 0.5, 0.7 และ 0.9
2. ตัวแปรตาม (Y) เป็นข้อมูลเชิงคุณภาพอยู่ในมาตรเรียงอันดับ (Ordinal Scale) ที่มีการแจกแจงแบบพหุนาม คือ มีค่าตัวแปรตามมากกว่า 2 ค่า หรือให้มี k ค่า คือ Y = 1, 2, … ,k โดยจะทำการศึกษาที่
   - กรณีจำนวนกลุ่มเท่ากับ 3 กลุ่ม คือ Y = 1, 2 และ 3 หรือ k=3
   - กรณีจำนวนกลุ่มเท่ากับ 4 กลุ่ม คือ Y = 1, 2, 3 และ 4 หรือ k=4
   - กรณีจำนวนกลุ่มเท่ากับ 5 กลุ่ม คือ Y = 1, 2, 3, 4 และ 5 หรือ k=5
3. ค่าความคลาดเคลื่อน (ε_i) มีการแจกแจงแบบปกติ ด้วยพารามิเตอร์ μ และ σ โดยเงื่อนไขความคลาดเคลื่อนของตัวแบบโลจิสติก คือ $\mu = 0$ และ $\sigma^2 =$ ค่าคงที่ ดังนั้นในงานวิจัยจะทำการศึกษาที่ $\varepsilon_i \sim N(0,100)$
4. กำหนดค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น ในการจำลองข้อมูล ดังนี้
   - กรณีมีตัวแปรอิสระ (X) 2 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ

     $\beta_1 = \beta_2 = -1,$ $\quad\beta_1 = \beta_2 = -0.6,$ $\quad\beta_1 = \beta_2 = -0.2,$

     $\beta_1 = \beta_2 = 1,$ $\quad\beta_1 = \beta_2 = 0.6,$ $\quad\beta_1 = \beta_2 = 0.2$

- กรณีมีตัวแปรอิสระ (X) 3 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ
  $\beta_1 = \beta_2 = \beta_3 = -1, \beta_1 = \beta_2 = \beta_3 = -0.6, \beta_1 = \beta_2 = \beta_3 = -0.2,$
  $\beta_1 = \beta_2 = \beta_3 = \ \ 1, \beta_1 = \beta_2 = \beta_3 = \ \ 0.6, \beta_1 = \beta_2 = \beta_3 = \ \ 0.2$

- กรณีมีตัวแปรอิสระ (X) 4 ตัว มีค่าสัมประสิทธิ์การถดถอยพารามิเตอร์เริ่มต้น $\beta_0 = 0$ และ
  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = -1, \qquad \beta_1 = \beta_2 = \beta_3 = \beta_4 = -0.6,$
  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = -0.2, \quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = \ \ 0.2,$
  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \ \ 0.6, \quad \beta_1 = \beta_2 = \beta_3 = \beta_4 = \ \ 1$

5. ศึกษาผลกระทบจากค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระ ณ ระดับต่างๆ กรณีมีตัวแปรอิสระ 2, 3และ 4 ตัว ขนาดตัวอย่าง $n_1 = 50, n_2 = 80, n_3 = 100, n_4 = 120$ และ $n_5 = 130$ ค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0, \beta_1 = \cdots = \beta_p = 1$ ดังนี้
   - กรณีมีตัวแปรอิสระ (X) 2 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_1, X_2) = 0.5, 0.7$ และ 0.9
   - กรณีมีตัวแปรอิสระ (X) 3 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_i, X_j) = 0.5, 0.7$ และ $0.9\ ; i \neq j$ ให้ $\mathrm{corr}(X_1, X_2)$, $\mathrm{corr}(X_1, X_3)$, $\mathrm{corr}(X_2, X_3)$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดยทำการศึกษาภายใต้ระดับความสัมพันธ์ซึ่งมีทั้งหมด 24 วิธี จาก 27 วิธี ที่เป็น positive definite ซึ่งสามารถนำไปใช้ในการคำนวณได้ ดังนี้

- 0.5, 0.5, 0.5
- 0.5, 0.5, 0.7
- 0.5, 0.5, 0.9
- 0.5, 0.7, 0.5
- 0.5, 0.7, 0.7
- 0.5, 0.7, 0.9
- 0.5, 0.9, 0.5
- 0.5, 0.9, 0.7
- 0.7, 0.5, 0.5
- 0.7, 0.5, 0.7
- 0.7, 0.5, 0.9
- 0.7, 0.7, 0.5
- 0.7, 0.7, 0.7
- 0.7, 0.7, 0.9
- 0.7, 0.9, 0.5
- 0.7, 0.9, 0.7
- 0.7, 0.9, 0.9
- 0.9, 0.5, 0.5
- 0.9, 0.5, 0.7
- 0.9, 0.7, 0.5
- 0.9, 0.7, 0.7
- 0.9, 0.7, 0.9
- 0.9, 0.9, 0.7
- 0.9, 0.9, 0.9

   - กรณีมีตัวแปรอิสระ (X) 4 ตัว ที่ระดับความสัมพันธ์ $\mathrm{corr}(X_i, X_j) = 0.5, 0.7$ และ $0.9\ ; i \neq j$ ให้ $\mathrm{corr}(X_1, X_2), \mathrm{corr}(X_1, X_3), \mathrm{corr}(X_1, X_4), \mathrm{corr}(X_2, X_3),$ $\mathrm{corr}(X_2, X_4), \mathrm{corr}(X_3, X_4)$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดยทำการศึกษาภายใต้ระดับความสัมพันธ์ซึ่งมีทั้งหมด 404 วิธี จาก 729 วิธี ที่เป็น positive definite ซึ่งสามารถนำไปใช้ในการคำนวณได้

6. จำนวนขนาดตัวอย่าง (n) จะให้มีอย่างน้อย 30 เท่าของจำนวนตัวแปรอิสระ โดยผู้วิจัยจะทำการศึกษาผลกระทบจากขนาดตัวอย่าง แบ่งเป็น 2 กรณี ได้แก่ กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน และกรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน โดยกำหนดค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0, \beta_1 = \cdots = \beta_p = 1$ และมีขนาดตัวอย่างของแต่ละกลุ่มดังนี้

- จำนวนกลุ่ม 3 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน
  - $n_1 = 30$, $n_2 = 30$, $n_3 = 30$
  - $n_1 = 50$, $n_2 = 50$, $n_3 = 50$
  - $n_1 = 80$, $n_2 = 80$, $n_3 = 80$
  - $n_1 = 100$, $n_2 = 100$, $n_3 = 100$
  - $n_1 = 120$, $n_2 = 120$, $n_3 = 120$

  กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน
  - $n_1 = 30$, $n_2 = 40$, $n_3 = 50$
  - $n_1 = 30$, $n_2 = 50$, $n_3 = 70$
  - $n_1 = 30$, $n_2 = 60$, $n_3 = 90$
  - $n_1 = 30$, $n_2 = 70$, $n_3 = 110$
  - $n_1 = 30$, $n_2 = 80$, $n_3 = 130$

- จำนวนกลุ่ม 4 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน
  - $n_1 = 30$, $n_2 = 30$, $n_3 = 30$, $n_4 = 30$
  - $n_1 = 50$, $n_2 = 50$, $n_3 = 50$, $n_4 = 50$
  - $n_1 = 80$, $n_2 = 80$, $n_3 = 80$, $n_4 = 80$
  - $n_1 = 100$, $n_2 = 100$, $n_3 = 100$, $n_4 = 100$
  - $n_1 = 120$, $n_2 = 120$, $n_3 = 120$, $n_4 = 120$

  กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน
  - $n_1 = 30$, $n_2 = 40$, $n_3 = 50$, $n_4 = 60$
  - $n_1 = 30$, $n_2 = 50$, $n_3 = 70$, $n_4 = 90$
  - $n_1 = 30$, $n_2 = 60$, $n_3 = 90$, $n_4 = 120$
  - $n_1 = 30$, $n_2 = 70$, $n_3 = 110$, $n_4 = 150$
  - $n_1 = 30$, $n_2 = 80$, $n_3 = 130$, $n_4 = 180$

- จำนวนกลุ่ม 5 กลุ่ม

  กรณีขนาดตัวอย่างแต่ละกลุ่มเท่ากัน
  - $n_1 = 30$, $n_2 = 30$, $n_3 = 30$, $n_4 = 30$, $n_5 = 30$
  - $n_1 = 50$, $n_2 = 50$, $n_3 = 50$, $n_4 = 50$, $n_5 = 50$
  - $n_1 = 80$, $n_2 = 80$, $n_3 = 80$, $n_4 = 80$, $n_5 = 80$
  - $n_1 = 100$, $n_2 = 100$, $n_3 = 100$, $n_4 = 100$, $n_5 = 100$
  - $n_1 = 120$, $n_2 = 120$, $n_3 = 120$, $n_4 = 120$, $n_5 = 120$

  กรณีขนาดตัวอย่างแต่ละกลุ่มไม่เท่ากัน
  - $n_1 = 30$, $n_2 = 40$, $n_3 = 50$, $n_4 = 60$, $n_5 = 70$
  - $n_1 = 30$, $n_2 = 50$, $n_3 = 70$, $n_4 = 90$, $n_5 = 110$

- $n_1 = 30,\quad n_2 = 60,\quad n_3 = 90,\quad n_4 = 120,\quad n_5 = 150$
- $n_1 = 30,\quad n_2 = 70,\quad n_3 = 110,\quad n_4 = 150,\quad n_5 = 190$
- $n_1 = 30,\quad n_2 = 80,\quad n_3 = 130,\quad n_4 = 180,\quad n_5 = 230$

7. ศึกษาผลกระทบจากความแปรปรวนของค่าเฉลี่ย(var(μ)) หรือลักษณะการซ้อนทับกันของตัวอย่างในแต่ละกลุ่ม $Y \sim N(\mu, \sigma^2)$ มีตัวแปรอิสระ(X) 2, 3 และ 4 ตัว ค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $\beta_0 = 0,\ \beta_1 = \cdots = \beta_p = 1$ ดังนี้

- กรณีจำนวนกลุ่ม 3 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50,\ n_2 = 80$ และ $n_3 = 100$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้
  - $\mu_1 = 0,\quad \mu_2 = 0,\quad \mu_3 = 0$
  - $\mu_1 = 0,\quad \mu_2 = 2,\quad \mu_3 = 4$
  - $\mu_1 = 0,\quad \mu_2 = 5,\quad \mu_3 = 10$
  - $\mu_1 = 0,\quad \mu_2 = 10,\quad \mu_3 = 20$
- กรณีจำนวนกลุ่ม 4 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50,\ n_2 = 80,\ n_3 = 100$ และ $n_4 = 120$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้
  - $\mu_1 = 0,\quad \mu_2 = 0,\quad \mu_3 = 0,\quad \mu_4 = 0$
  - $\mu_1 = 0,\quad \mu_2 = 2,\quad \mu_3 = 4,\quad \mu_4 = 6$
  - $\mu_1 = 0,\quad \mu_2 = 5,\quad \mu_3 = 10,\quad \mu_4 = 15$
  - $\mu_1 = 0,\quad \mu_2 = 10,\quad \mu_3 = 20,\quad \mu_4 = 30$
- กรณีจำนวนกลุ่ม 5 กลุ่ม ที่ขนาดตัวอย่าง $n_1 = 50,\ n_2 = 80,\ n_3 = 100,\ n_4 = 120$ และ $n_5 = 130$ มีค่าเฉลี่ยของแต่ละกลุ่ม ดังนี้
  - $\mu_1 = 0,\quad \mu_2 = 0,\quad \mu_3 = 0,\quad \mu_4 = 0,\quad \mu_5 = 0$
  - $\mu_1 = 0,\quad \mu_2 = 2,\quad \mu_3 = 4,\quad \mu_4 = 6,\quad \mu_5 = 8$
  - $\mu_1 = 0,\quad \mu_2 = 5,\quad \mu_3 = 10,\quad \mu_4 = 15,\quad \mu_5 = 20$
  - $\mu_1 = 0,\quad \mu_2 = 10,\quad \mu_3 = 20,\quad \mu_4 = 30,\quad \mu_5 = 40$

สามารถสรุปผลการศึกษาในกรณีต่างๆได้ดังนี้

5.1 สรุปผลการศึกษา

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย, ค่าระดับความสัมพันธ์ของตัวแปรอิสระ, ขนาดตัวอย่าง และความแปรปรวนของค่าเฉลี่ยที่มีต่อค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี (Hypervolume Under the Manifold (HUM)) ในกรณีที่มีตัวแปรตาม 3, 4 และ 5 กลุ่ม ซึ่งสรุปผลการวิเคราะห์ได้ดังนี้

จากการสังเกตจากตารางและกราฟข้อมูลจะพบว่า ในกรณีของจำนวนกลุ่มเป็น 3, 4 และ 5 กลุ่ม ให้ผลสอดคล้องไปในทิศทางเดียวกัน แต่เมื่อจำนวนกลุ่มเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะมีค่าสูงขึ้น เนื่องมาจากการที่จำนวนตัวอย่างรวมมีค่าสูงขึ้น นอกจากนี้ยังพบว่า ถ้าจำนวนตัวแปรอิสระเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าเพิ่มขึ้น โดยในแต่ละปัจจัยที่มีผลกระทบต่อค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีสามารถสรุปได้ดังนี้

1. ผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอย β

ค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอยแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี คือ เมื่อค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอยของมีค่าเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีจะเพิ่มขึ้น เนื่องจากการที่ค่าสัมบูรณ์ของพารามิเตอร์สัมประสิทธิ์การถดถอยมีค่าเพิ่มขึ้น ตัวแปรอิสระจะสามารถอธิบายตัวแปรตามได้มากขึ้น ทำให้ความสามารถในการทำนายกลุ่มแม่นยำขึ้น

2. ผลกระทบจากค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระ

ค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระแปรผกผันกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี คือ เมื่อค่าความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น ส่งผลให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีแนวโน้มลดลง เนื่องจากการที่ตัวแปรอิสระมีความสัมพันธ์กัน ผิดสมมุติฐานเบื้องต้นของการวิเคราะห์ความถดถอย เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูงขึ้นทำให้ความสามารถในการทำนายกลุ่มของสมการถดถอยโลจิสติกลดลง

3. ผลกระทบจากจำนวนขนาดตัวอย่าง

3.1 เมื่อขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน

ขนาดตัวอย่างกรณีที่ขนาดตัวอย่างทุกกลุ่มเท่ากันแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี คือ เมื่อขนาดตัวอย่างทุกกลุ่มมีจำนวนเพิ่มขึ้นอย่างเท่ากัน ทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น

เนื่องจากการที่ ขนาดตัวอย่างมีจำนวนเพิ่มขึ้นใกล้เคียงกับค่าประชากรมากขึ้น ทำให้ความสามารถในการทำนายกลุ่มของสมการถดถอยโลจิสติกมีความแม่นยำมากขึ้น

3.2 เมื่อขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน

กรณีที่ขนาดตัวอย่างในแต่ละกลุ่มไม่เท่ากัน ค่าความแปรปรวนของขนาดตัวอย่าง หรือค่าสัมประสิทธิ์การแปรผันของขนาดตัวอย่าง แปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี หมายถึง การที่ขนาดตัวอย่างแต่ละกลุ่มมีจำนวนแตกต่างกันมากขึ้น ทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น ทั้งนี้อาจเนื่องมาจากการที่ ขนาดตัวอย่างในแต่ละกลุ่มมีจำนวนเพิ่มขึ้นทำให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าสูงขึ้น เช่นเดียวกับกรณีที่ขนาดตัวอย่างในแต่ละกลุ่มเท่ากัน

นอกจากนี้ยังพบว่า ในกรณีที่ขนาดตัวอย่างรวมเท่ากัน เมื่อเปรียบเทียบระหว่างกรณีที่มีตัวอย่างในแต่ละกลุ่มไม่เท่ากันและกรณีที่ตัวอย่างในแต่ละกลุ่มเท่ากัน ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่มีตัวอย่างในแต่ละกลุ่มไม่เท่ากันมีค่าสูงกว่า ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีกรณีที่มีตัวอย่างในแต่ละกลุ่มเท่ากัน

4. ผลกระทบจากความแปรปรวนของค่าเฉลี่ย var(μ)

ความแปรปรวนของค่าเฉลี่ยแปรผันตรงกับค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี คือ ความแปรปรวนของค่าเฉลี่ยเพิ่มขึ้น ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าเพิ่มขึ้น เนื่องจาก เมื่อความแปรปรวนของค่าเฉลี่ยเพิ่มขึ้นแสดงว่า ข้อมูลของแต่ละกลุ่มซ้อนทับกันน้อย หรือข้อมูลส่วนใหญ่ของแต่ละกลุ่มอยู่ห่างกันเพิ่มขึ้น สมการพยากรณ์จึงสามารถพยากรณ์กลุ่มได้อย่างแม่นยำมากขึ้น แต่ถ้าข้อมูลส่วนใหญ่ของแต่ละกลุ่มอยู่ห่างกันลดลง ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีที่ได้จะมีแนวโน้มลดลงเรื่อยๆ โดยค่าต่ำที่สุดที่สามารถแบ่งกลุ่มได้ คือ $\frac{1}{k!}$ (k = จำนวนกลุ่มทั้งหมด) ถ้าค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีมีค่าต่ำกว่า $\frac{1}{k!}$ แสดงว่าไม่สามารถแบ่งกลุ่มออกจากกันได้

จากผลสรุปดังกล่าว ถ้าต้องการให้ค่าเฉลี่ยปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซีสูงๆหรือสามารถทำนายกลุ่มได้ดี ควรมีค่าระดับความสัมพันธ์ระหว่างตัวแปรอิสระต่ำๆ หรือทำการแก้ไขความสัมพันธ์ระหว่าตัวแปรอิสระก่อนนำมาวิเคราะห์ความถดถอย และเก็บข้อมูลให้มีจำนวนมากถ้ามีตัวแปรอิสระหลายตัว หรือลดจำนวนตัวแปรอิสระเท่าที่เป็นไปได้ให้อยู่ในเกณฑ์ที่สามารถยอมรับได้ กรณีที่ตัวอย่างมีน้อย

5.2 ข้อเสนอแนะ

1. ในงานวิจัยครั้งนี้ ศึกษาตัวแปรอิสระที่มีการแจงแจงแบบปกติ และตัวแปรตามมีลักษณะเป็นสเกลอันดับจากน้อยไปมาก ดังนั้นในงานวิจัยต่อไปอาจทำการศึกษาในกรณีที่ตัวแปรอิสระมาจากการแจกแจงแบบอื่นๆ
2. ในงานวิจัยครั้งนี้ ศึกษาผลกระทบจากค่าพารามิเตอร์สัมประสิทธิ์ความถดถอยที่มีค่าเท่ากันทุกตัวและเท่ากันทุกกลุ่ม ดังนั้นในงานวิจัยต่อไปอาจทำการศึกษาในกรณีที่ค่าพารามิเตอร์สัมประสิทธิ์ความถดถอยแต่ละกลุ่มมีค่าไม่เท่ากัน
3. ศึกษาผลกระทบจากกรณีที่ขนาดตัวอย่างรวมเท่าเดิม แต่อัตราส่วนขนาดตัวอย่างของแต่ละกลุ่มเปลี่ยนแปลงไป

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 4. กรุงเทพฯ : ธรรมสาร, 2552.

ภาษาอังกฤษ

C.Ferri, J.Hernandez-Orallo, and M.A. Salido, Volume under the roc surface for multi-class problems. *Proc. Of 14th European Conference on Machine Learning*, pp. 108-120, 2003.

Thomas Landgrebe, and Robert P.W. Duin. A simplified extension of the Area under the ROC to the multiclass domain. *Proc. Of the 17th annual symposium pattern recognition association*, South Africa. 2006.

บรรณานุกรม

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 4. กรุงเทพฯ : ธรรมสาร, 2552.

กัลยา วานิชย์บัญชา. การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows. พิมพ์ครั้งที่ 7. กรุงเทพฯ : ธรรมสาร, 2552.

ปฐมาภรณ์ สานุกูล. อำนาจการทดสอบของตัวสถิติสำหรับพื้นที่ใต้โค้ง ROC บนตัวแบบโพรบิท. วิทยานิพนธ์ปริญญามหาบัณฑิต, ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2552.

ภาษาอังกฤษ

Douglas Mossman. Three-way ROCs. *Medical Decision Making*, vol.19, pp.78-89,1999.

He, X., Metz, C. E., Tsui, B. M.W., Links, J. M., & Frey, E. C.. Three-Class Roc Analysis-A Decision Theoretic Approach Under the Ideal Observer Framework. *IEEE Transactions on Medical Imaging*, 25, pp.571-581.

J. LI & J. P. FINE. ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics*, **9**, 3, *pp.* 566–576, 2008.

Rosa Bersabe & Teresa Rivas. A General Equation to Obtain Multiple Cut-off Scores on a Test from Multinomial Logistic Regression. *The Spanish Journal of Psychology*, vol.13, No.1, pp.494-502, 2010.

Thomas Landgrebe & Robert P.W. Duin. A simplified extension of the Area under the ROC to the multiclass domain. *Elect.Eng., Maths and Comp. Sc.*, Delft University of Technology, The Netherlands.

ภาคผนวก

คำสั่งที่ใช้ในการวิเคราะห์ข้อมูลจำลองจากโปรแกรม R

ยกตัวอย่างกรณี มีตัวแปรตาม 3 กลุ่ม ตัวแปรอิสระ 2 ตัว ซึ่งมีระดับความสัมพันธ์ระหว่างตัวแปรอิสระ 0.5 จำนวนตัวอย่าง $n_1 = 50, n_2 = 80, n_3 = 100$ ที่ค่า β_0=0 $\beta_1 = \beta_2 = 1$ มีค่าเฉลี่ยของแต่ละกลุ่ม คือ $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 2$

```
library(Formula)
library(statmod)
library(lmtest)
library(zoo)
library(miscTools)
library(maxLik)
library(MASS)
library(mlogit)
library(plyr)
library(pROC)

hum<-c()
for(i in 1:1000){
        z1<-rnorm(50)
        z2<-rnorm(50)
        z3<-rnorm(80)
        z4<-rnorm(80)
        z5<-rnorm(100)
        z6<-rnorm(100)
```

```
m<-matrix(c(1,0.5,0.5,1),2,2)
cm<-chol(m)
t(cm)
Z1<-data.frame(z1,z2)
Z2<-data.frame(z3,z4)
Z3<-data.frame(z5,z6)

MZ1<-as.matrix(Z1)
MZ2<-as.matrix(Z2)
MZ3<-as.matrix(Z3)

MZN1<-MZ1%*%t(cm)
MZN2<-MZ2%*%t(cm)
MZN3<-MZ3%*%t(cm)

mu1<-0
mu2<-1
mu3<-2
var<-1
MXg1<-(sqrt(var)*MZN1)+mu1
MXg2<-(sqrt(var)*MZN2)+mu2
MXg3<-(sqrt(var)*MZN3)+mu3

MX1g1<-MXg1[,1]
MX2g1<-MXg1[,2]
MX1g2<-MXg2[,1]
```

```
    MX2g2<-MXg2[,2]
    MX1g3<-MXg3[,1]
    MX2g3<-MXg3[,2]

MX1<-c(MXg1[,1],MXg2[,1],MXg3[,1])
MX2<-c(MXg1[,2],MXg2[,2],MXg3[,2])
X<-data.frame(MX1,MX2)
MX<-as.matrix(X)

    er1<-rnorm(50,0,sqrt(100))
    Mer1<-matrix(c(er1),50,1)
    er2<-rnorm(80,0,sqrt(100))
    Mer2<-matrix(c(er2),80,1)
    er3<-rnorm(100,0,sqrt(100))
    Mer3<-matrix(c(er3),100,1)

    Y1<-1*MX1g1+1*MX2g1+Mer1
    Y2<-1*MX1g2+1*MX2g2+Mer2
    Y3<-1*MX1g3+1*MX2g3+Mer3

    Y1<-c(Y1)
    Y2<-c(Y2)
    Y3<-c(Y3)
    Y<-c(Y1,Y2,Y3)

    n<-length(Y)
```

```
	n1<-length(Y1)
	n2<-length(Y2)
	n3<-length(Y3)

r1<-Y[1:n1]
r2<-Y[(n1+1):(n1+n2)]
r3<-Y[(n1+n2+1):n]

g1<-c()
for(i in 1:length(r1)){
d<-ifelse(r1[i]==r1[i],1,0)
g1<-c(g1,d)
}

g2<-c()
for(i in 1:length(r2)){
d<-ifelse(r2[i]==r2[i],2,0)
g2<-c(g2,d)
}

g3<-c()
for(i in 1:length(r3)){
d<-ifelse(r3[i]==r3[i],3,0)
g3<-c(g3,d)
}
```

```
YG<-c(g1,g2,g3)

          YGd<-data.frame(YG)
          data<-data.frame(YGd,X)
          names(data)

          data$YG<-as.factor(data$YG)

          mldata<-mlogit.data(data, varying=NULL, choice="YG", shape="wide")
          mlogit.model<-mlogit(YG ~ 1| MX1+MX2, data = mldata, reflevel="1")
          summary(mlogit.model)

          eq<-summary(mlogit.model)

          newdata<-data.frame(cbind(MX1,MX2))
          logit1<-rep(0,n)
          logit2<-(eq$coef[1])+(eq$coef[3]*MX1)+(eq$coef[5]*MX2)
          logit3<-(eq$coef[2])+(eq$coef[4]*MX1)+(eq$coef[6]*MX2)

          m1<-mlogit(YG~1|MX1+MX2,data=mldata,reflevel="1")

          logits<-cbind(logit1,logit2,logit3)
          p.unscaled<-exp(logits)
          p<-cbind(newdata,(p.unscaled/rowSums(p.unscaled)))
          colnames(p)<- c("X1","X2","pred.1","pred.2","pred.3")
```

```
        pred.1<-p[,3]
        pred.2<-p[,4]
        pred.3<-p[,5]
        pred<-matrix(c(pred.1,pred.2,pred.3),n,3)

        YGhat<-max.col(pred)

        cTab<-table(YG,YGhat)
        addmargins(cTab)

        data.class(model.frame(YGhat~Y,data=data))
        multiclass.roc(YGhat~Y,data)
        multiclass.roc(YGhat~Y,data)->cc

hum<-c(hum,cc$auc)
}
```

ตัวอย่างในการคำนวณหาค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี

ข้อมูลภาวะน้ำหนักเกินและอ้วน จากสำนักงานสำรวจสุขภาพประชาชนไทย โดยทำการสำรวจภาวะอ้วนในประชากรหญิง อายุ 15 - 29 ปี มีจำนวนทั้งสิ้น 1,306 ราย จำแนกกลุ่มออกเป็น 3 กลุ่ม คือ กลุ่มน้ำหนักน้อย $(Y = 1)$, กลุ่มน้ำหนักปกติ $(Y = 2)$ และกลุ่มน้ำหนักเกิน $(Y = 3)$

โดยพิจารณาจาก

น้ำหนักตัว(Weight: W) หน่วยเป็นกิโลกรัม

การวัดเส้นรอบวงเอว(Waist Circumference: WC) เป็นค่าที่ได้จากการวัดรอบเอวด้วยสายวัดมาตรฐาน โดยวัดรอบเอวระดับตำแหน่งกึ่งกลางของข้างเอวระหว่างขอบล่างของซี่โครงล่างกับขอบบนของ iliac crest ให้สายรอบเอวแนบรอบเอวและอยู่ในแนวขนานกับพื้น

ซึ่งมีข้อมูลดังนี้

Y	W	WC
1	46	68
2	51	67
2	58	75
2	66	74
2	61	78
3	87	102
2	62	80
2	60	76
3	69	78
2	58	77
3	59	75
2	55	72
2	60	76
1	47	66
2	55	75
2	61	76
2	52	78
1	41	58
2	54	65
3	68	85

Y	W	WC
2	46	59
2	58	69
3	72	69
2	59	76
2	63	84
2	54	76
2	55	63
1	41	62
1	45	70
2	63	78
3	69	79
2	63	82
2	54	67
2	62	60
2	63	77
2	55	82
2	56	81
1	39	60
3	73	79
2	49	70

Y	W	WC
3	61	81
2	53	83
3	75	78
2	53	76
2	53	75
2	54	64
2	65	77
2	54	68
3	67	97
2	59	72
2	51	73
2	62	86
3	67	76
1	34	60
1	44	56
2	58	89
2	67	70
1	42	60
2	62	70
2	50	66

Y	W	WC
2	67	74
2	60	78
2	62	67
1	44	65
2	55	64
2	48	65
2	61	84
2	58	82
3	69	87
2	51	69
1	48	63
1	43	57
2	62	76
3	75	94
2	61	75
2	61	78
3	67	91
1	38	63
2	52	74
2	67	73

Y	W	WC
2	60	80
2	51	78
2	58	80
2	60	69
2	61	80
2	60	76
2	49	63
3	68	90
3	66	79
3	74	78
2	60	74
2	57	73
2	62	74
3	76	85
2	57	71
2	53	76
2	53	68
3	64	73
2	51	69
3	64	86
2	67	83
3	64	87
2	56	76
2	66	76
2	50	76
1	45	75
1	43	69
2	60	72
1	43	59
2	61	83
2	59	77
3	77	88
2	47	68
2	47	59
2	48	69
3	62	87
2	62	75

Y	W	WC
1	49	79
1	46	66
1	44	57
2	63	83
2	60	63
1	48	64
1	40	66
1	40	58
2	62	67
2	48	76
1	48	61
1	39	60
1	49	66
1	42	64
1	44	64
2	48	69
2	58	74
2	53	69
1	36	59
2	53	72
3	80	76
1	30	63
2	56	60
3	67	89
2	52	62
2	59	72
3	73	90
2	58	74
2	53	69
3	72	69
3	73	94
2	56	69
1	44	66
1	33	61
2	58	91
3	62	82
2	54	79

Y	W	WC
2	58	70
2	53	65
2	45	76
2	49	64
2	66	66
2	63	68
2	58	77
2	54	70
2	61	73
3	62	83
2	56	68
2	60	78
2	47	77
3	81	86
2	54	73
2	54	66
2	61	70
2	51	69
2	54	65
3	72	70
2	47	65
2	50	75
2	53	67
3	70	77
2	61	77
3	66	79
2	60	87
2	51	76
2	55	68
2	54	68
2	50	58
3	73	83
2	61	70
2	55	78
3	79	84
2	48	60
2	59	72

Y	W	WC
2	58	84
2	51	65
1	39	73
3	65	78
1	40	63
2	62	78
2	54	69
2	50	79
3	80	80
2	56	71
1	46	62
2	58	62
3	66	73
2	56	83
2	47	60
2	59	72
2	63	85
2	55	72
1	47	64
2	44	71
3	72	77
2	57	64
3	65	87
1	45	56
2	55	76
3	70	65
2	53	70
2	57	70
3	75	87
3	69	84
3	71	79
1	46	72
2	57	79
2	56	71
2	50	83
2	56	81
2	56	77

Y	W	WC
2	66	77
2	60	68
2	61	78
3	65	82
2	58	73
3	74	77
2	65	72
3	71	73
1	47	73
1	31	57
2	52	60
2	61	85
2	60	81
2	58	62
2	55	74
2	52	75
2	60	79
2	53	71
3	75	91
2	60	74
2	67	87
2	51	72
2	47	65
2	53	66
2	56	78
2	52	74
2	62	79
1	45	71
3	71	73
1	47	60
2	59	85
2	54	70
2	60	73
3	62	87
2	57	72
2	53	74
1	48	66

Y	W	WC
2	61	71
2	54	84
2	53	77
2	57	73
1	44	55
1	45	69
1	44	62
2	58	60
2	63	73
2	63	76
2	53	73
3	65	80
1	43	65
2	58	80
3	76	85
3	72	76
2	63	76
2	52	57
2	56	55
2	59	61
3	73	72
2	52	74
2	59	66
3	69	78
2	61	77
3	66	85
3	69	85
2	61	82
2	50	78
2	63	79
1	48	64
1	47	66
3	64	74
3	79	92
2	60	83
3	78	77
1	42	58

Y	W	WC
1	47	75
3	68	86
2	51	64
1	44	63
2	50	72
1	45	57
2	63	65
3	69	94
2	54	72
1	45	63
2	59	76
2	53	79
1	42	58
1	44	71
1	49	60
1	46	58
3	66	79
2	61	80
2	56	82
2	59	68
2	62	78
3	69	81
3	71	85
2	47	70
2	52	76
2	50	64
2	60	82
3	65	74
3	64	84
1	44	62
3	97	112
3	79	90
2	49	62
3	63	86
1	40	56
2	61	67
2	55	62

Y	W	WC
1	40	61
3	63	86
2	51	68
3	76	83
1	43	60
2	64	79
2	51	72
2	62	75
2	52	70
2	58	84
2	51	75
1	43	68
2	63	77
1	38	58
2	61	66
2	65	74
2	60	75
1	40	67
2	64	78
1	43	53
2	48	69
2	63	69
2	53	72
3	64	77
2	51	65
2	60	70
3	71	92
1	44	65
2	53	72
2	59	76
2	59	90
2	49	76
2	52	77
1	44	79
2	65	86
2	50	60
2	55	71

Y	W	WC
2	57	84
1	46	72
2	59	85
1	37	62
1	42	72
2	59	82
2	55	71
2	67	87
3	69	75
2	59	80
3	66	85
3	68	82
3	62	82
2	55	77
2	47	66
1	34	52
2	62	81
2	49	64
2	62	77
3	68	93
1	44	70
3	70	65
2	60	72
2	54	61
2	51	69
2	53	70
2	62	72
1	47	66
3	61	80
2	53	61
1	33	72
1	40	56
2	56	68
2	63	80
3	84	99
3	71	83
2	60	71

Y	W	WC
2	58	73
2	60	81
2	63	75
2	59	84
2	60	80
2	52	66
2	48	59
2	48	70
2	61	75
3	64	84
2	57	69
1	47	65
2	56	68
3	70	89
2	65	75
2	56	81
2	59	79
2	55	72
2	51	68
3	74	77
3	68	78
1	38	64
3	75	84
3	71	80
3	68	75
2	52	71
3	70	92
3	66	80
2	63	79
2	60	88
2	66	81
2	54	84
1	47	66
2	60	61
1	45	59
2	51	56
2	52	67

Y	W	WC
3	64	83
3	75	91
2	54	72
1	38	71
2	56	67
2	55	64
2	59	77
2	61	75
2	61	74
2	52	75
2	64	81
2	53	66
3	64	85
2	64	88
2	55	79
2	59	74
2	61	74
2	63	73
2	49	74
2	61	82
1	40	52
2	56	72
2	56	77
3	68	71
2	54	72
2	55	69
1	46	61
2	48	65
2	48	62
3	70	88
3	75	87
2	49	58
2	59	72
2	59	83
2	49	74
3	69	64
3	67	82

Y	W	WC
2	51	65
3	82	89
2	60	84
2	62	80
3	59	81
2	51	74
2	58	83
2	50	77
3	62	89
3	65	79
2	55	77
1	39	66
3	65	73
2	59	82
3	73	84
2	61	77
2	51	64
3	60	90
2	55	67
2	59	67
3	77	79
2	49	68
1	42	67
1	48	66
3	66	84
2	58	81
2	57	74
2	59	78
2	60	83
2	48	82
2	55	72
1	41	63
2	65	87
2	66	80
2	59	75
1	44	58
2	58	76

Y	W	WC
3	64	83
2	52	71
1	36	63
1	43	62
3	74	85
2	45	64
1	48	59
3	67	78
2	52	71
2	61	87
3	87	98
2	53	64
2	52	71
2	63	82
2	61	69
2	55	82
2	63	73
2	56	68
3	75	86
2	62	71
2	57	77
2	56	85
2	48	65
1	31	59
2	68	71
2	65	70
1	33	63
2	58	79
2	55	67
2	63	68
1	41	58
2	49	62
2	54	73
2	59	74
2	55	77
2	57	70
2	58	90

Y	W	WC
2	60	64
2	53	77
2	52	67
2	58	73
3	72	92
2	49	76
1	25	50
2	51	78
2	63	69
2	50	68
3	79	72
1	49	65
2	55	78
2	54	76
2	55	76
1	40	56
2	58	74
2	58	75
2	58	80
2	50	76
2	57	84
2	48	69
1	44	74
1	35	57
2	60	82
2	44	60
2	59	65
2	64	84
2	46	56
1	46	62
3	62	81
1	46	61
2	56	73
1	44	66
2	53	70
1	47	64
2	63	71

Y	W	WC
3	71	81
2	49	73
2	61	87
2	53	70
1	38	59
2	59	68
2	47	75
2	64	76
2	64	86
2	60	83
1	45	66
2	55	90
1	45	66
1	37	61
2	44	85
1	42	64
3	68	71
2	54	75
2	56	67
1	45	61
2	50	75
2	58	72
2	51	74
2	57	75
2	53	65
2	62	79
3	74	79
2	53	74
1	43	69
2	59	76
2	55	77
2	62	76
2	52	78
3	68	84
2	54	80
2	56	71
1	52	72

Y	W	WC
2	53	80
2	61	76
3	68	85
1	45	58
2	61	78
2	59	73
2	55	84
2	60	81
2	64	89
2	57	85
2	55	73
2	63	76
3	74	83
3	68	87
3	69	87
2	60	73
2	56	74
2	52	61
2	57	64
3	64	76
2	54	65
3	71	85
2	54	64
3	78	87
2	56	67
3	65	88
1	44	58
2	59	94
2	68	76
2	54	71
1	40	61
2	68	79
2	58	76
2	46	73
2	49	70
2	61	75
2	50	63

Y	W	WC
2	54	76
2	59	62
3	72	75
2	55	79
1	34	71
2	48	66
3	72	82
2	63	81
2	60	71
2	56	78
2	64	72
2	55	72
2	55	71
2	62	68
2	59	70
1	33	57
2	56	64
3	70	75
2	57	73
2	54	69
3	63	81
3	72	99
2	58	77
2	51	78
2	55	70
2	54	75
3	58	80
2	51	64
3	68	86
3	92	103
1	37	59
2	59	65
1	45	64
2	57	75
3	67	80
1	38	61
2	63	75

Y	W	WC
2	57	68
3	71	90
2	50	80
2	65	77
2	62	72
3	69	80
2	58	74
2	52	75
2	49	75
2	55	80
3	62	79
2	55	63
2	53	63
2	64	69
1	49	61
2	66	74
3	65	79
1	45	73
2	54	55
1	43	57
2	60	61
3	70	84
3	73	88
1	40	55
3	66	87
1	44	61
2	56	77
1	41	72
3	65	86
2	63	70
3	66	87
2	55	74
2	59	84
3	86	83
2	50	64
2	59	76
2	53	82

Y	W	WC
1	45	71
3	63	83
2	58	74
2	53	78
3	64	88
2	46	68
2	58	78
2	67	71
2	61	84
1	34	65
2	52	78
2	62	82
3	67	82
2	62	68
2	58	72
2	62	77
2	58	57
1	40	74
3	70	80
2	60	70
3	69	84
3	68	92
3	65	83
2	51	75
3	70	90
3	70	84
1	46	57
2	44	66
2	52	71
3	75	84
2	54	63
2	59	66
2	67	86
3	67	80
3	71	83
2	57	63
3	69	74

Y	W	WC
2	55	60
2	55	86
2	59	88
3	75	83
2	62	78
2	55	75
3	63	81
1	40	55
2	65	83
2	65	71
2	58	71
2	57	87
2	52	74
2	46	75
2	52	82
2	57	75
2	58	69
1	38	68
3	64	86
2	53	73
1	42	76
2	61	71
1	46	78
2	47	74
2	48	72
2	54	64
2	55	69
3	65	76
2	61	83
2	52	73
3	72	90
1	45	55
1	44	67
3	64	69
3	68	82
1	45	62
2	65	70

Y	W	WC
3	72	92
2	58	72
2	61	89
3	72	87
2	48	71
2	51	61
2	53	65
3	77	97
2	56	66
2	53	70
2	64	76
2	55	76
2	54	80
2	46	64
2	54	90
1	50	70
2	56	67
1	33	66
3	73	76
3	57	84
2	54	65
1	43	75
2	44	56
2	48	70
2	43	77
1	45	64
2	62	84
3	69	89
1	37	59
2	59	76
2	59	70
3	73	80
2	56	69
2	58	71
3	64	83
3	82	85
2	52	68

Y	W	WC
2	56	79
3	71	87
2	52	79
1	41	60
2	60	73
2	52	77
2	49	78
3	61	88
3	79	83
2	54	73
2	63	67
2	60	68
3	62	74
1	36	62
2	53	72
1	39	72
2	57	86
2	53	67
2	52	63
3	73	82
1	44	59
2	47	72
2	59	72
2	54	83
3	70	98
2	64	69
2	60	72
2	55	67
2	67	84
2	60	72
2	53	72
2	51	80
2	47	62
2	53	66
2	57	72
1	40	68
2	66	75

Y	W	WC
2	60	79
3	62	69
3	62	89
2	56	73
3	65	80
2	52	67
3	68	94
3	69	93
2	54	76
1	35	55
1	38	64
2	56	73
1	39	57
2	61	63
3	68	91
2	67	79
3	68	87
2	65	74
2	56	73
2	56	61
3	69	86
1	52	75
2	53	88
2	67	74
2	64	76
1	40	63
2	49	71
1	41	69
2	64	72
2	57	68
2	53	76
2	59	81
1	37	59
2	58	79
2	50	70
2	51	67
2	57	72

Y	W	WC
2	58	70
1	45	59
2	50	68
2	45	62
2	63	75
2	62	75
2	56	80
2	61	84
2	49	77
1	40	54
2	56	73
2	46	69
2	51	68
2	59	85
2	53	83
1	48	68
1	38	67
3	75	78
2	55	67
2	54	81
2	58	86
3	64	81
1	45	60
3	68	87
2	59	80
1	46	56
2	63	82
2	56	65
3	74	88
3	87	78
2	62	77
2	57	78
2	61	81
2	57	69
2	67	76
1	50	76
2	59	71

Y	W	WC
1	47	58
2	63	71
2	52	82
2	57	78
2	56	86
2	51	75
2	62	71
2	45	65
2	60	105
1	42	60
1	37	70
1	39	66
2	61	83
2	52	75
3	73	88
2	63	80
2	57	69
2	52	76
2	57	68
2	54	58
2	65	76
2	58	83
3	66	85
2	60	86
3	77	88
1	49	65
2	49	76
2	54	67
2	53	71
2	64	82
3	72	87
2	62	78
1	43	70
1	48	55
2	54	78
2	48	73
2	66	77

Y	W	WC
3	71	91
2	59	80
2	52	69
2	60	79
2	50	61
2	60	72
1	44	58
2	49	69
2	60	75
2	55	67
2	57	82
2	51	68
2	60	80
3	77	89
2	54	60
2	54	74
3	64	84
3	79	89
1	41	69
3	70	69
3	66	85
2	60	78
1	35	67
3	65	75
1	32	56
2	56	75
2	55	69
3	62	88
3	67	81
3	62	73
2	56	83
3	65	90
3	62	84
2	65	69
2	50	67
2	65	82
2	52	65

Y	W	WC
2	57	73
1	37	65
2	55	76
2	57	84
3	62	82
2	63	71
3	63	78
1	46	65
2	59	86
2	48	71
1	35	61
2	59	76
2	67	81
2	65	73
3	84	108
1	37	59
2	63	72
2	60	70
2	66	85
2	66	70
2	52	73
3	74	78
2	50	62
2	59	70
2	54	70
1	43	67
3	74	86
2	48	68
2	56	74
2	63	80
1	34	65
1	40	64
2	51	68
2	53	64
1	31	59
2	56	83
2	56	87

Y	W	WC
3	75	81
3	76	102
2	58	80
3	66	80
2	66	70
3	69	87
2	55	80
3	71	71
2	53	67
1	49	66
2	48	65
2	60	70
1	36	63
3	76	79
2	57	82
1	24	55
2	50	68
3	75	93
2	58	76
1	39	65
2	53	76
2	49	83
1	41	60
2	55	77
2	56	64
1	37	60
3	62	75
3	66	84
2	67	85
1	32	55
2	62	77
3	73	81
2	59	71
1	51	65
2	58	84
3	67	81
1	38	56

Y	W	WC
2	57	87
3	68	99
2	62	93
1	45	60
2	61	72
1	44	68
2	59	74
2	50	76
2	57	82
2	62	79
2	47	73
2	45	73
2	49	59
3	64	89
3	79	90
1	44	57
1	44	65
3	73	80
2	59	82
2	55	93
2	50	65
3	64	85
1	40	70
2	45	57
1	46	61
3	81	97
1	44	64
3	67	73
2	58	76
3	65	86
1	40	68
1	44	61
3	69	79
2	46	71
1	42	59
1	49	69
2	51	67

Y	W	WC
1	39	69
3	65	82
2	59	82
2	50	58
2	50	72
2	63	69
2	57	77
3	70	78
2	51	73
3	70	85
3	71	70
2	61	80
2	65	66
2	61	78
2	56	70
1	38	67
3	66	80
1	45	66
1	47	69
2	52	72
2	52	71
2	52	70
1	47	73
2	59	79
1	46	59
3	64	76
1	46	64
2	58	79
2	49	71
3	73	75
3	61	78
1	44	62
2	60	70
2	50	68
2	57	73
3	73	76
2	63	77

Y	W	WC
3	69	77
2	52	79
2	62	80
2	65	73
2	56	75
2	50	75
2	61	64
2	58	68
2	55	68
3	69	76
2	55	69
2	57	69
2	57	90
1	40	68
2	59	66
2	61	74
2	57	69
2	57	74
2	48	68
3	71	83
2	57	70
2	61	66
2	57	71
2	60	79
3	82	104
2	65	84
2	65	74
1	41	71
2	50	66
3	63	87
2	55	68
2	50	69
2	56	78
3	69	84
3	65	74
1	40	63
2	49	80

Y	W	WC
2	49	65
1	40	72
2	58	78
2	59	60
1	46	59
3	66	86
2	55	70
2	55	78
3	69	73
2	51	62
2	53	68
2	61	80
3	68	82
2	55	72
3	67	93
3	74	87
2	63	74
2	56	80
3	66	89
2	56	69
2	56	86
2	62	80
3	69	74
2	50	80
3	80	82
2	51	77
2	58	73
3	64	83
2	58	64
3	70	83
1	50	70
2	60	79
2	50	70
2	53	73
3	69	83
1	45	64
3	70	79

Y	W	WC
1	46	67
1	45	64
2	54	76
2	49	73
1	46	58
2	54	66
2	48	78
1	46	67
3	80	82
2	56	66
1	41	60

Y	W	WC
2	55	54
2	63	83
2	58	85
2	53	56
2	53	63
2	54	64
2	60	71
2	55	78
2	48	57
1	45	70
3	63	83

Y	W	WC
2	48	68
2	55	76
2	49	58
1	47	69
1	47	55
2	58	76
3	61	85
1	48	65
2	54	66
2	52	67
2	53	80

Y	W	WC
1	43	75
2	64	74
1	43	59
2	57	75
2	55	70
1	46	72
2	51	89
2	65	74

ซึ่งมีระดับความสัมพันธ์ระหว่างค่าน้ำหนักตัวและค่ารอบวงเอว เท่ากับ 0.6914

ได้สมการพยากรณ์ดังนี้

$$P(Y_i = 1) = \frac{1}{1 + e^{-39.124+0.715W+0.083WC} + e^{-87.769+1.282W+0.233WC}}$$

$$P(Y_i = 2) = \frac{e^{-39.124+0.715W+0.083WC}}{1 + e^{-39.124+0.715W+0.083WC} + e^{-87.769+1.282W+0.233WC}}$$

$$P(Y_i = 3) = \frac{e^{-87.769+1.282W+0.233WC}}{1 + e^{-39.124+0.715W+0.083WC} + e^{-87.769+1.282W+0.233WC}}$$

ได้ค่าการพยากรณ์ดังนี้

Observed	Predicted			
	$\hat{Y} = 1$	$\hat{Y} = 2$	$\hat{Y} = 3$	Percent Correct
$Y = 1$	196	29	0	87.1%
$Y = 2$	26	753	33	92.7%
$Y = 3$	0	47	222	82.5%
Overall Percent	17%	63.5%	19.5%	89.7 %

จากสมการที่ได้จะพบว่าตัวแปรที่มีความสัมพันธ์คือตัวแปรอิสระ 2 ตัว ได้แก่ น้ำหนักและรอบวงเอว สามารถอธิบายระดับความอ้วน มีความแม่นยำในการพยากรณ์ 89.7%

เมื่อใช้คำสั่งในการวิเคราะห์ข้อมูลจากโปรแกรม R คำนวณหาค่าปริมาตรแบบไฮเปอร์ใต้โค้งอาร์โอซี หรือค่าประสิทธิภาพของตัวแบบได้ค่าเท่ากับ 0.837

เนื่องจากค่า HUM $\geq \frac{1}{3!}$ หมายความว่า ตัวแบบที่ได้ มีความน่าเชื่อถือสามารถยอมรับได้และสามารถจำแนกเหตุการณ์ต่างๆได้

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวพิมพลอย ภูววิเชียรฉาย เกิดวันที่ 21 สิงหาคม พ.ศ. 2530 สำเร็จการศึกษาปริญญาเศรษฐศาสตรบัณฑิต (ศ.บ.) สาขาวิชาการเงิน คณะเศรษฐศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2551 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต (สต.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2553