เรื่อง

ปัจจัยที่มีอิทธิพลต่ออัตราการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

Factors Affecting the Drop-out Rate in Faculty of Science and Technology, Thammasat University

โดย

นางสาวณิชาพัชร์ ธนาวุฒิ	**เลขทะเบียน**	**6309682752**
นางสาวบงกชทิพ คู่วัฒนา	**เลขทะเบียน**	**6309682851**
นายศิริเดช เจริญศิริ	**เลขทะเบียน**	**6309683040**

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชา ส.495 โครงงานพิเศษ 2
ตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาสถิติ สาขาวิชาคณิตศาสตร์และสถิติ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
ปีการศึกษา 2566

คำนำ

รายงานเรื่อง ปัจจัยที่มีอิทธิพลต่อการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ เป็นส่วนหนึ่งของวิชา ส.495 โครงงานพิเศษ 2 จัดทำขึ้นโดยที่งานวิจัยนี้สนใจศึกษาปัจจัยที่จะส่งผลต่ออัตราการพ้นสภาพของนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยธรรมศาสตร์ โดยใช้การวิเคราะห์การถดถอยทวินาม ภายใต้ฟังก์ชันการเชื่อมโยงที่แตกต่างกันสามฟังก์ชัน ได้แก่ ฟังก์ชันเชื่อมโยงลอจิต, ฟังก์ชันเชื่อมโยงโพรบิต และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก - ล็อก และเปรียบเทียบประสิทธิภาพฟังก์ชันเชื่อมโยงทั้งสามฟังก์ชันเชื่อมโยงโดยใช้เกณฑ์สารสนเทศของอะกะอิเกะ และ เกณฑ์สารสนเทศของเบส์

ผู้จัดทำหวังเป็นอย่างยิ่งว่ารายงานฉบับนี้จะเป็นประโยชน์กับผู้อ่าน และผู้ที่สนใจศึกษาเกี่ยวกับเรื่องนี้ทุกท่าน และสามารถนำไปใช้ต่อยอดในการศึกษาเรื่องอื่น ๆ ต่อไปในอนาคต หากรายงานฉบับนี้มีข้อผิดพลาดประการใด คณะผู้จัดทำต้องขออภัยไว้ ณ ที่นี้ด้วย

คณะผู้จัดทำ

หัวข้อโครงงานพิเศษ	ปัจจัยที่มีอิทธิพลต่ออัตราการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
คณะผู้จัดทำ	นางสาวณิชาพัชร์ ธนาวุฒิ นางสาวบงกชทิพ คู่วัฒนา นายศิริเดช เจริญศิริ
ชื่อปริญญา	วิทยาศาสตรบัณฑิต
หลักสูตร/สาขา	หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาสถิติ
คณะ/มหาวิทยาลัย	คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษางานวิจัย	ผู้ช่วยศาสตราจารย์.ดร. มณฑิรา ดวงสาพล
ปีการศึกษา	2566

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่ส่งผลต่ออัตราการพ้นสภาพการเป็นนักศึกษาของนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ จำนวน 20 หลักสูตรในปีการศึกษา 2561 โดยใช้การวิเคราะห์การถดถอยทวินาม โดยผู้วิจัยได้สร้างตัวแบบเพื่อวิเคราะห์ปัจจัยที่มีผลต่ออัตราการพ้นสภาพการเป็นนักศึกษา ภายใต้ฟังก์ชันการเชื่อมโยงที่แตกต่างกัน ได้แก่ ฟังก์ชันเชื่อมโยงลอจิต, ฟังก์ชันเชื่อมโยงโพรบิต และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก จากการศึกษาพบว่าตัวแบบการถดถอยทวินาม ผ่านฟังก์ชันการเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ให้ค่า AIC และ BIC ต่ำที่สุด และผลจากการวิเคราะห์หาปัจจัยที่ส่งผลต่ออัตราการพ้นสภาพการเป็นนักศึกษา จากทั้งหมด 10 ปัจจัย พบว่ามี 3 ปัจจัยที่ส่งผลต่ออัตราการพ้นสภาพการเป็นนักศึกษา คือ คะแนน Admission (รอบ 3) สูงสุด จำนวนนักศึกษาที่เข้ามารอบ 3 และคะแนน Admission (รอบ 4) สูงสุด นอกจากนี้ผู้วิจัยได้ศึกษาตัวแบบการถดถอยทวินามผ่านการจำลองข้อมูล ซึ่งการจำลองข้อมูลในงานวิจัยนี้ใช้เทคนิคมอนติคาร์โล ภายใต้ขนาดตัวอย่างที่แตกต่างกัน คือ 20, 50, 200 และ 750 ผลการศึกษาโดยสรุปจากการจำลองพบว่าค่าเฉลี่ย AIC และ BIC ให้ประสิทธิภาพที่ดีทั้งสามฟังก์ชันเชื่อมโยง ส่วนค่าร้อยละของจำนวนครั้งที่ให้ค่า AIC และ BIC ต่ำสุด ฟังก์ชันเชื่อมโยงโพรบิตและคอมพลีเมนทารีล็อก – ล็อก ให้ประสิทธิภาพที่ดี ในขณะที่ฟังก์ชันเชื่อมโยงลอจิตมีประสิทธิภาพที่ดีเมื่อขนาดตัวอย่างเท่ากับ 750

คำสำคัญ : ตัวแบบการถดถอยทวินาม, ฟังก์ชันการเชื่อมโยง, เทคนิคมอนติคาร์โล

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงได้ด้วยความกรุณาช่วยเหลือและคำแนะนำเป็นอย่างดีจาก ผู้ช่วยศาสตราจารย์ ดร.มณฑิรา ดวงสาพล อาจารย์ที่ปรึกษางานวิจัยนี้ ผู้จุดประกายแนวคิดรวมทั้งการให้คำแนะนำข้อคิดเห็น ต่าง ๆ ที่เป็นประโยชน์ในการวิจัย ตลอดจนช่วยเหลือและแก้ไขข้อบกพร่องต่าง ๆ อย่างละเอียดทุกขั้นตอนเป็นอย่างดี อีกทั้งยังเป็นผู้ให้กำลังใจที่ดีตลอดจนกระทั่งงานวิจัยนี้เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ผู้วิจัยขอขอบพระคุณคณาจารย์ภาควิชาคณิตศาสตร์และสถิติทุกท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้แก่ผู้วิจัย ขอขอบคุณเพื่อน ๆ และพี่น้องภาควิชาคณิตศาสตร์และสถิติ ที่เป็นกำลังใจและให้ความช่วยเหลือผู้วิจัยเป็นอย่างดี

คณะผู้จัดทำ

สารบัญ

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ปัญหาการพ้นสภาพของนักศึกษามีความสำคัญในหลายประเทศ รวมทั้งประเทศไทยเนื่องด้วยการศึกษาเป็นรากฐานสำคัญในการสร้างความเจริญก้าวหน้าทางสังคม เศรษฐกิจของประเทศ เพราะการศึกษาเป็นกระบวนการในการพัฒนาคนให้เป็นมนุษย์ที่สมบูรณ์ เป็นการสร้างความเข้มแข็งให้แก่ชุมชน ภาครัฐได้ตระหนักถึงความสำคัญและเล็งเห็นถึงผลกระทบที่จะเกิดขึ้นจากการที่ประชากรภายในประเทศมีการศึกษาน้อย ซึ่งจะก่อให้เกิดความสูญเสียต่อประเทศทั้งทางตรงและทางอ้อม รวมทั้งส่งผลกระทบต่อเศรษฐกิจอย่างร้ายแรง อัตราการพ้นสภาพอาจนำไปสู่การว่างงานที่เพิ่มขึ้น รายได้ลดลง และความคล่องตัวทางสังคมลดลง นอกจากนี้อัตราการพ้นสภาพที่สูงอาจส่งผลเสียต่อคุณภาพการศึกษาและนำไปสู่การสูญเสียทรัพยากร โดยมีผลกระทบด้านลบต่อคุณภาพการศึกษาและโอกาสในการทำงานในอนาคต ตลอดจนเศรษฐกิจและสังคมซึ่งคณะวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยธรรมศาสตร์มีการรายงานอัตราการพ้นสภาพที่สูงในช่วงไม่กี่ปีที่ผ่านมา โดยเฉพาะในช่วงปีการศึกษา 2563 ซึ่งเป็นช่วงที่มีการระบาดของไวรัสโคโรนา โควิด-19 อย่างรุนแรง ทำให้มหาวิทยาลัยธรรมศาสตร์รวมทั้งมหาวิทยาลัยอื่น ๆ เองก็มีความจำเป็นที่จะต้องปรับเปลี่ยนรูปแบบการเรียนการสอนจากเดิมที่เป็นการเรียนแบบออนไซต์ในมหาวิทยาลัย 100% เปลี่ยนมาเป็นการเรียนในรูปแบบออนไลน์ 100% ในช่วงภาคเรียนที่ 2 ปีการศึกษา 2563 ทำให้การใช้ชีวิตของนักศึกษาเปลี่ยนแปลงไปรวมถึงการเรียนเช่นกัน เนื่องด้วยวิกฤตการณ์ที่ไม่เคยเกิดขึ้นจึงทำให้นักศึกษาจำนวนไม่น้อยที่ไม่สามารถปรับตัวรับมือได้ นั่นเป็นเหตุผลที่ทำให้รูปแบบการเรียนที่เปลี่ยนไปเกิดเป็นปัญหาใหญ่ นอกจากนี้ปัญหาการพ้นสภาพอาจมีปัจจัยอื่นๆที่เกี่ยวข้อง เช่น ปัจจัยด้านหลักสูตรและการเรียนการสอนในแต่ละสาขามีหลักสูตรวิชาที่ต้องศึกษาแตกต่างกัน จำนวนหน่วยกิตก็ต่างกันด้วย อาจส่งผลให้นักศึกษาที่เข้ามาเรียนในแต่ละหลักสูตรรู้สึกว่าไม่ตอบโจทย์หรือไม่เหมาะสมกับตนเอง รวมถึงปัญหาด้านสถานศึกษา ด้านสภาพแวดล้อม หรือด้านเศรษฐกิจที่ชะลอตัว เกิดปัญหาการเงินใดๆ

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีทางสถิติที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตอบสนองกับตัวแปรอธิบายชุดหนึ่งโดยมีจุดมุ่งหมายเพื่อที่จะอธิบายหรือทำนายตัวแปรตอบสนอง ซึ่งตัวแปรตอบสนองอาจจะเป็นตัวแปรสุ่มไม่ต่อเนื่องหรือต่อเนื่อง ในกรณีที่ตัวแปรตอบสนองเป็นตัวแปรสุ่มไม่ต่อเนื่องจะเรียกว่า การวิเคราะห์การถดถอยแบบนับ (Count Regression Analysis) ยกตัวอย่างเช่น การวิเคราะห์การถดถอยลอจิสติก (Logistic Regression Analysis) การวิเคราะห์การถดถอยปัวซง (Poisson Regression analysis) การวิเคราะห์การถดถอยทวินามเชิงลบ (Negative Regression Analysis) การถดถอยทวินาม (Binomial Regression Analysis) จากข้อมูลการรายงานอัตราการพ้นสภาพการเป็นนักศึกษาของนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ในแต่ละหลักสูตรนั้น สามารถนำตัวแบบ

การถดถอยทวินามมาประยุกต์ใช้เพื่อการศึกษาปัจจัยต่าง ๆ ที่อาจจะส่งผลต่ออัตราการพ้นสภาพของนักศึกษาได้

ดังนั้นในงานวิจัยนี้สนใจศึกษาข้อมูลอัตราการพ้นสภาพการเป็นนักศึกษาและปัจจัยที่จะส่งผลต่ออัตราการพ้นสภาพโดยรวบรวมจากนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีการศึกษา 2561 โดยใช้การวิเคราะห์การถดถอยทวินาม (Binomial Regression Analysis) ภายใต้ฟังก์ชันการเชื่อมโยงที่แตกต่างกันสามฟังก์ชัน ได้แก่ ฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function), ฟังก์ชันเชื่อมโยงโพรบิต (Probit Link Function) และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก (Complementary Log-Log Link Function) และเปรียบเทียบประสิทธิภาพฟังก์ชันเชื่อมโยงทั้งสาม โดยใช้เกณฑ์สารสนเทศของอะกะอิเกะ (AIC) และ เกณฑ์สารสนเทศของเบส์(BIC) นอกจากนี้ผู้วิจัยได้ศึกษาเชิงจำลองเนื่องจากข้อมูลคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ประกอบด้วยสาขาวิชา 20 สาขา ซึ่งเป็นหน่วยตัวอย่างที่มีขนาดเล็กซึ่งอาจจะส่งผลต่อการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยทวินาม ดังนั้นเพื่อให้เห็นถึงประสิทธิภาพของการประมาณในสถานการณ์ต่างๆ ที่มีขนาดตัวอย่างที่แตกต่างกันและฟังก์ชันเชื่อมโยงที่แตกต่างกัน ซึ่งเมื่อเพิ่มขนาดตัวอย่างมากขึ้นอาจจะส่งผลให้ประสิทธิภาพของตัวประมาณมีความแม่นยำ

1.2 วัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาปัจจัยที่มีอิทธิพลต่ออัตราการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีการศึกษา 2561 โดยใช้ตัวแบบการถดถอยทวินาม

2. เพื่อเปรียบเทียบประสิทธิภาพของฟังก์ชันเชื่อมโยงที่แตกต่างกันในตัวแบบการถดถอยทวินามจากการจำลองข้อมูลในสถานการณ์ต่าง ๆ สำหรับการทำนายอัตราการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยธรรมศาสตร์

3. เพื่อให้ข้อมูลเชิงลึกและคำแนะนำแก่สาขาวิชาในการลดอัตราการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ จากการวิเคราะห์แบบจำลองการถดถอยทวินาม

1.3 ขอบเขตของการศึกษา

ในงานวิจัยนี้เป็นการประยุกต์ใช้กับข้อมูลจริงและการศึกษาเชิงจำลอง โดยประมวลผลจากโปรแกรม Rstudio เวอร์ชัน 2023.03.0+386 ภายใต้ข้อมูลที่มีสถานการณ์ต่าง ๆ ซึ่งมีขอบเขตการศึกษาดังนี้

1.3.1 ตัวแบบการถดถอยทวินาม มีรูปแบบดังนี้

1.3.1.1 ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต

$$\log it(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = X_i'\beta = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}$$

เมื่อ

X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1)\times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

p คือ จำนวนตัวแปรอธิบาย กำหนดเท่ากับ 3 ตัว

กล่าวได้ว่า π_i ยังคงอยู่ในช่วง [0,1] มีรูปแบบดังนี้

$$\pi_i = \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}$$

จะได้ฟังก์ชันมวลความน่าจะเป็นแบบมีเงื่อนไข Y_i เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{y_i}\left(\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)^{y_i}\left(1-\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)^{n_i-y_i}$$

1.3.1.2 ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต

$$Probit(\pi_i) = \phi^{-1}(\pi_i) = X_i'\beta = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}$$

เมื่อ

X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1)\times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

p คือ จำนวนตัวแปรอธิบาย กำหนดเท่ากับ 3 ตัว

กล่าวได้ว่า π_i ยังคงอยู่ในช่วง [0,1] มีรูปแบบดังนี้

$$\pi_i = \phi(X_i'\beta)$$

จะได้ฟังก์ชันมวลความน่าจะเป็นแบบมีเงื่อนไข Y_i เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{y_i}(\phi(X_i'\beta))^{y_i}(1-\phi(X_i'\beta))^{n_i-y_i}$$

1.3.1.3 ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารี ล็อก-ล็อก

$$\log(-\log(1-\pi_i)) = X_i'\beta = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip}$$

เมื่อ

X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1)\times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

p คือ จำนวนตัวแปรอธิบาย กำหนดเท่ากับ 3 ตัว

กล่าวได้ว่า π ยังคงอยู่ในช่วง [0,1] มีรูปแบบดังนี้

$$\pi_i = 1 - \exp(-\exp(X_i'\beta))$$

จะได้ฟังก์ชันมวลความน่าจะเป็นแบบมีเงื่อนไข Y_i เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{y_i}(1-\exp(-\exp(X_i'\beta)))^{y_i}(\exp(-\exp(X_i'\beta)))^{n_i - y_i}$$

1.3.2 การวิเคราะห์ข้อมูลจริง

1.3.2.1 ประยุกต์ใช้กับข้อมูลการพ้นสภาพการเป็นนักศึกษาของนักศึกษาปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ จำนวน 20 หลักสูตร ในปีที่เข้าศึกษา 2561

ตัวแปรตอบสนอง คือ จำนวนนักศึกษาที่พ้นสภาพในแต่ละหลักสูตร (Y_{i})

ตัวแปรอธิบาย คือ หลักสูตรที่มีวิชาบังคับ C และหลักสูตร (X_1), จำนวนหน่วยกิต แต่ละหลักสูตร (X_2), ค่าเทอมในแต่ละหลักสูตร (X_3), ประเภทหลักสูตร (X_4), คะแนน Admission รอบ 3 สูงสุดในแต่ละหลักสูตร (X_5), คะแนน Admission รอบ 3 ต่ำสุด ในแต่ละหลักสูตร (X_6), จำนวนนักศึกษาที่เข้ามารอบ 3 ในแต่ละหลักสูตร (X_7), คะแนน Admission รอบ 4 สูงสุด ในแต่ละหลักสูตร (X_8), คะแนน Admission รอบ 4 ต่ำสุด ในแต่ละหลักสูตร (X_9) และจำนวนนักศึกษาที่เข้ามารอบ 4 ในแต่ละหลักสูตร (X_{10})

ในการศึกษาโดยใช้ข้อมูลจริง Y_{i} มีการแจกแจงทวินาม ซึ่งเขียนแทนด้วย $Y_{\mathrm{i}} \sim Binomial\,(n_i, \pi_i)$ โดยที่ Y_{i} หมายถึง จำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร, n_i หมายถึง จำนวนนักศึกษาที่เข้ามาศึกษาในแต่ละหลักสูตรและ π_i หมายถึง อัตราการพ้นสภาพนักศึกษาในแต่ละหลักสูตร โดยวิเคราะห์ผ่านฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function), ฟังก์ชันเชื่อมโยงโพรบิต (Probit Link Function) และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก (Complementary Log-Log Link Function)

1.3.2 การศึกษาเชิงจำลอง

เนื่องจากข้อมูลคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ประกอบด้วยสาขาวิชา 20 สาขา ซึ่งเป็นหน่วยตัวอย่างที่มีขนาดเล็กซึ่งอาจจะส่งผลต่อการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยทวินาม ดังนั้นเพื่อให้เห็นถึงประสิทธิภาพของการประมาณในสถานการณ์ต่างๆที่มีขนาดตัวอย่างที่แตกต่างกันและฟังก์ชันเชื่อมโยงที่แตกต่างกัน ซึ่งเมื่อเพิ่มขนาดตัวอย่างมากขึ้นอาจจะส่งผลให้ประสิทธิภาพของตัวประมาณมีความแม่นยำ โดยในการศึกษาเชิงจำลองจะกำหนดสถานการณ์ต่างๆและสถานการณ์ที่ใกล้เคียงกับกรณีศึกษาอัตราการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

1.3.2.1 ฟังก์ชันมวลความน่าจะเป็นของตัวแปรสุ่ม $Y_{\mathrm{i}} \sim Binomial\ (n_i, \pi_i)$ ที่มีการแจกแจงทวินาม มีรูปแบบดังนี้

$$f(Y_i|\ n_i, \pi_i) = \binom{n_i}{\pi_i} \pi_i{}^{y_i}(1-\pi_i)^{n_i - y_i}\ ;\ y_{\mathrm{i}}\ =\ 0,1,2,3, \ldots, n_i$$

เมื่อ Y_i หมายถึง จำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร

n_i หมายถึง จำนวนนักศึกษาที่เข้ามาศึกษาในแต่ละหลักสูตร

π_i หมายถึง อัตราการพ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร

1.3.2.2 กำหนดขนาดตัวอย่างที่ทำการศึกษา มี 4 ระดับ คือ 20, 50, 200, และ 750

1.3.2.3 สร้างตัวแปรอธิบายจำนวน 3 ตัวแปร โดยจำลองจากการแจกแจงเอกรูป (Uniform distribution) และการแจกแจงปัวซง (Poisson distribution) โดยค่าพารามิเตอร์เหล่านี้จะถูกกำหนดให้ใกล้เคียงจากข้อมูลจริง ตามลำดับดังนี้

โดยที่ $X_1 \sim Uni(30.33, 75.87)$ คือ คะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร ปี 2561 ที่มีค่าต่ำสุดเท่ากับ 30.33 และค่าสูงสุดเท่ากับ 75.87

โดยที่ $X_2 \sim Uni(12127.1,\ 17239.5)$ คือ คะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร ปี 2561 ที่มีค่าต่ำสุดเท่ากับ 12127.1 และค่าสูงสุดเท่ากับ 17239.5

โดยที่ $X_3 \sim Poisson(50)$ คือ จำนวนนักศึกษาที่เข้ามารอบ 3 ในแต่ละหลักสูตร ปี 2561 ที่มีค่าเฉลี่ยเท่ากับ 50

1.3.2.4 กำหนดค่าพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta = (\beta_0, \beta_1, \beta_2, \beta_3\ )$ ตัวแปรอธิบายให้ใกล้เคียงกับผลจากการวิเคราะห์อัตราการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีเข้ารับการศึกษา 2561 โดยแบ่งการกำหนดค่าพารามิเตอร์ ดังนี้

กรณีที่ 1 ตัวแบบการถดถอยทวินาม ที่มีตัวแปรตอบสนอง Y จะอยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_i|X_i)$ โดยที่ π_i จะสัมพันธ์กับตัวแปรอธิบาย X ผ่านฟังก์ชันเชื่อมโยงลอจิต (Logit-Link) กำหนดค่าพารามิเตอร์ ดังนี้

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (0.133000, -0.040800, 0.000130, -0.021500)$$

กรณีที่ 2 ตัวแบบการถดถอยทวินาม ตัวแปรตอบสนอง Y จะอยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_i|X_i)$ โดยที่ π_i จะสัมพันธ์กับตัวแปรอธิบาย X ผ่านฟังก์ชันเชื่อมโยงโพรบิต (Probit-link) กำหนดค่าพารามิเตอร์ ดังนี้

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (0.069100, -0.024500, 0.000078, -0.012900)$$

กรณีที่ 3 ตัวแบบการถดถอยทวินาม ตัวแปรตอบสนอง Y จะอยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_i|X_i)$ โดยที่ π_i จะสัมพันธ์กับตัวแปรอธิบาย X ผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารี ล็อก-ล็อก (Complementary Log-Log Link) กำหนดค่าพารามิเตอร์ ดังนี้

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (-0.246524, -0.033570, 0.000107, -0.017955)$$

1.3.2.5 กำหนดค่า $n_i \sim Poisson\ (54)$ โดยที่ในงานวิจัยนี้จะกำหนดค่า λ คือ ค่าเฉลี่ยของจำนวนนักศึกษาที่รับเข้ามาในแต่ละหลักสูตร ปี 2561 โดยมีค่าเท่ากับ $\lambda = 54$

1.3.2.6 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพตัวแบบ คือ เกณฑ์สารสนเทศของอะกะอิเกะ (AIC) และเกณฑ์สารสนเทศของเบส์ (Bayesian Information Criterion: BIC)

1.3.2.7 ในการจำลองจะทำซ้ำ 1000 รอบ ในทุกเงื่อนไขที่กำหนดในการศึกษานี้

1.4 นิยามศัพท์เฉพาะและสัญลักษณ์ที่ใช้

1.4.1 MLE หมายถึงการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุด (Method of Maximum Likelihood Estimation)

1.4.2 β หมายถึงเวกเตอร์ค่าพารามิเตอร์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบายกับตัวแปรตอบสนองของตัวแบบเชิงเส้นนัยทั่วไปที่ตัวแปรมีการแจกแจงทวินาม

1.4.3 AIC หมายถึง ค่าเกณฑ์สารสนเทศของอะกะอิเกะ

1.4.4 BIC หมายถึง ค่าเกณฑ์สารสนเทศของเบส์

1.4.5 ตัวแบบเชิงเส้นนัยทั่วไปที่ตัวแปรมีการแจกแจงทวินาม (Generalized Model with Binomial Distribution) หมายถึงตัวแบบที่ใช้ในการอธิบายโครงสร้างความสัมพันธ์ระหว่างตัวแปรอธิบายกับตัวแปรตอบสนอง โดยที่ส่วนประกอบเชิงสุ่มมีการแจกแจงทวินามเมื่อฟังก์ชันลอจิต, โพรบิต และคอมพลีเมนทารี ล็อก-ล็อก เป็นฟังก์ชันเชื่อมโยง

1.4.6 นักศึกษา หมายถึง นักศึกษาปริญญาตรีที่พ้นสภาพการเป็นนักศึกษา ปีรับเข้าศึกษา 2561 ระหว่างชั้นปีที่ 1 ภาคเรียนที่ 1 ถึง ชั้นปีที่ 4 ภาคเรียนที่ 1

1.4.7 สถานภาพนักศึกษา มหาวิทยาลัยจะจำแนกสถานภาพนักศึกษาในทุกภาคการศึกษา ทั้งนี้ไม่นับภาคการศึกษาที่ได้ลาพักหรือถูกให้พัก จำแนกได้ 6 ประเภท

1) นักศึกษาที่มีคะแนนเฉลี่ยสะสม 2.00 ขึ้นไป มีสถานภาพวิชาการปกติ (Normal)
2) นักศึกษาที่มีคะแนนเฉลี่ยสะสมต่ำกว่า 2.00 มีสถานภาพทางวิชาการเตือนครั้งที่ 1 (Warning 1) เว้นแต่กรณีเป็นภาคการศึกษาแรกที่เข้าศึกษา ให้มีสถานภาพทางวิชาการเตือนพิเศษ (Warning)
3) นักศึกษาซึ่งอยู่ในสถานภาพทางวิชาการเตือนพิเศษในภาคการศึกษาแรกที่เข้าศึกษาตาม (2) และมีคะแนนเฉลี่ยสะสมต่ำกว่า 1.50 ในภาคการศึกษาถัดมา ต้องถูกถอนชื่อออกจากทะเบียนนักศึกษา (Dismissed)
4) นักศึกษาซึ่งอยู่ในสถานภาพทางวิชาการเตือนพิเศษ หรือเตือนครั้งที่ 1 ตาม (2) ในภาคการศึกษาที่ผ่านมา และมีคะแนนเฉลี่ยต่ำกว่า 2.00 ในภาคการศึกษาถัดมา ให้มีสถานภาพทางวิชาการเตือนครั้งที่ 2 (Warning 2)
5) นักศึกษาซึ่งอยู่ในสถานภาพทางวิชาการเตือนครั้งที่ 2 ตาม (4) ในภาคการศึกษาที่ผ่านมา และมีคะแนนเฉลี่ยสะสมต่ำกว่า 2.00 ในภาคการศึกษาถัดมา ให้มีสถานภาพทางวิชาการภาวะรอพินิจ (Probation)
6) นักศึกษาซึ่งอยู่ในสถานภาพทางวิชาการภาวะรอพินิจ ตาม (5) ในภาคการศึกษาที่ผ่านมา และมีคะแนนเฉลี่ยสะสมต่ำกว่า 2.00 ในภาคการศึกษาถัดมา ต้องถูกถอนชื่อออกจากทะเบียนนักศึกษา (Dismissed)

1.4.9 การพ้นสภาพการเป็นนักศึกษา จำแนกได้ ดังนี้

1) ตายหรือลาออก
2) ต้องโทษทางวินัยให้พ้นสภาพการเป็นนักศึกษา
3) ไม่ได้ลงทะเบียนเรียนภายใน 30 วัน นับจากวันเปิดภาคเรียนปกติ โดยมิได้รับการอนุมัติให้ลาพักการศึกษา หรือไม่ได้รักษาสถานภาพ
4) โดนสถานภาพทางวิชาการให้พ้นสภาพการเป็นนักศึกษา

1.4.8 รอบที่ 3 รับตรงร่วมกัน (Admission 1) หมายถึง การรับตรงร่วมกัน เป็นการรับตรงของแต่ละมหาวิทยาลัย ซึ่งโครงการรับตรงอย่าง กสพท. ก็รวมอยู่ในรอบนี้ด้วย โดยที่ ทปอ. จะเป็นส่วนกลางในการรับสมัคร และมหาวิทยาลัยจะพิจารณาผลการคัดเลือก ผ่านระบบ การคัดเลือกของ TCAS

1.4.9 รอบที่ 4 รับกลางร่วมกัน (Admission 2) หมายถึง การคัดเลือกแบบ Admission โดยใช้ทั้งคะแนน GPAX, O-NET, GAT/PAT หรือคะแนนอื่นๆที่ทางมหาวิทยาลัยเป็นผู้กำหนด

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อประเมินฟังก์ชันเชื่อมโยงใดเหมาะสมที่สุด สำหรับการสร้างแบบจำลองอัตราการพ้นสภาพการเป็นนักศึกษาในคณะวิทยาศาสตร์และเทคโนโลยี ซึ่งจะช่วยให้เข้าใจความสัมพันธ์ได้ดีขึ้น ระหว่างฟังก์ชันการเชื่อมโยงต่างๆ กับประสิทธิภาพของตัวแบบการถดถอยทวินามโดยทั่วไป

2. เพื่อนำผลการวิจัยที่ได้ ไปใช้เป็นข้อมูลพื้นฐานประกอบการปรับปรุง พัฒนา หรือแก้ไขปัญหาในการจัดการเรียนการสอนของคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

3. เพื่อเป็นแนวทางในการประยุกต์ใช้ข้อมูลอื่น ๆ ภายใต้ตัวแบบการถดถอยทวินาม

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีพื้นฐานต่าง ๆ ได้แก่ ตัวแบบเชิงเส้นนัยทั่วไป การแจกแจงทวินาม ตัวแบบการถดถอยทวินาม วิธีภาวะน่าจะเป็นสูงสุด การประมาณค่าพารามิเตอร์ด้วยวิธีน่าจะเป็นสูงสุดสำหรับการถดถอยทวินาม เกณฑ์ที่ใช้ในการเปรียบเทียบ รวมถึงวรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

วีรานันท์ (2555) ได้กล่าวว่าตัวแบบเชิงเส้นนัยทั่วไปเป็นตัวแบบที่ขยายมาจากตัวแบบเชิงเส้นแบบคลาสสิก (Classical linear model) โดยขยายใน 3 ส่วน คือ (1) ขยายส่วนประกอบเชิงสุ่ม (Random component) ของตัวแปรตอบสนอง จากเดิมที่ใช้สำหรับการแจกแจงปรกติไปสู่การแจกแจงในวงศ์เลขชี้กำลัง (2) ขยายส่วนประกอบเชิงระบบ (Systematic component) จากเดิมที่เป็นผลรวมเชิงเส้นบนเทอมพารามิเตอร์ของตัวแปรอธิบายแบบต่อเนื่องให้ใช้ได้สำหรับตัวแปรอธิบายแบบจำแนกประเภททั้งหมด หรือเป็นแบบผสมทั้งแบบจำแนกประเภทหรือเชิงกลุ่มและแบบต่อเนื่อง และ (3) ขยายส่วนประกอบฟังก์ชันเชื่อมโยง (Link function) ที่เดิม เคยใช้เฉพาะฟังก์ชันเชื่อมโยงเอกลักษณ์ให้ใช้ได้กับฟังก์ชันเชื่อมโยงแบบอื่นที่เป็นฟังก์ชันทางเดียวที่หาอนุพันธ์ได้ (Monotonic differentiable function) มีรายละเอียดของแต่ละส่วนประกอบดังนี้

2.1.1 ส่วนประกอบเชิงสุ่ม (Random Component)

ส่วนประกอบเชิงสุ่มเป็นส่วนของลักษณะการแจกแจงของตัวแปรตอบสนองเชิงสุ่ม (Y) โดยมีค่าสังเกตของ Y ขนาด N หน่วยคือ $y_i\,; i = 1, \dots, N$ ที่เป็นอิสระต่อกันและ $Y = (Y_1, Y_2, \dots, Y_N)$
มีการแจกแจงในวงศ์เลขชี้กำลัง (Exponential Family) ภายใต้ฟังก์ชันน่าจะเป็นดังนี้

$$f_y(y_i; \theta, \phi) = \exp\left\{\frac{[y_i\theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi\,)\right\} \tag{2.1}$$

เมื่อ $a(\cdot)\,, b(\cdot)$ และ $c(\cdot)$แทนฟังก์ชันต่าง ๆ และ (θ, ϕ) แทนพารามิเตอร์ของการแจกแจงที่มี θ เป็นพารามิเตอร์ธรรมชาติ (Natural Parameter) และ ϕ เป็นพารามิเตอร์การกระจาย (Dispersion Parameter) โดยทั่วไปแล้วฟังก์ชัน $a(\theta)$ มีรูปแบบเป็น $a(\theta) = \phi/w_i$ โดยที่ w_i แทนน้ำหนักที่ทราบค่า เช่น $\bar{y}_i$ แทนค่าเฉลี่ยของ N_i หน่วยที่เป็นอิสระต่อกันและนิยมใช้ $w_i = N_i$ สำหรับการแจกแจงใด ๆ ที่สามารถจัดให้อยู่ในวงศ์เลขชี้กำลังได้แล้ว
กำหนดให้

$$L = \sum_i L_i$$

จะได้ว่า

$$L_i = [[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i) \quad (2.2)$$

ซึ่งสามารถหาอนุพันธ์ของ (2.2) ได้ดังนี้

$$\frac{\partial L_i}{\partial \theta_i} = \frac{[y_i \; - \; b'(\theta_i)]}{a(\phi)} \quad (2.3)$$

$$\frac{\partial^2 L_i}{\partial \theta_i^2} = -\frac{b''(\theta_i)}{a(\phi)} \quad (2.4)$$

เมื่อ $b'(\theta_i)$ และ $b''(\theta_i)$ แทน ค่าอนุพันธ์ลำดับที่หนึ่งและค่าอนุพันธ์ลำดับที่สองของ $b(\cdot)$ สำหรับค่าที่ θ_i ตามลำดับ จากพื้นฐานความรู้ของภาวะน่าจะเป็นทั่วไป (General Likelihood)

$$E\left(\frac{\partial L}{\partial \theta}\right) = 0 \quad (2.5)$$

และ

$$-E\left(\frac{\partial^2 L}{\partial \theta^2}\right) = E\left(\frac{\partial L}{\partial \theta}\right)^2 \quad (2.6)$$

สามารถหาค่าเฉลี่ยของค่าสังเกตใด ๆ ได้จากการแทน (2.3) ลงใน (2.5) ดังนี้

$$E[y_i - b'(\theta_i)]/\, a(\phi) \; = \; 0$$

ดังนั้นค่าเฉลี่ยของค่าสังเกตใด ๆ มีค่าเท่ากับ

$$\mu_i = E(y_i) = b'(\theta_i) \quad (2.7)$$

และสามารถหาค่าความแปรปรวนของค่าสังเกตใด ๆ ได้จากการแทนค่า (2.4) ลงใน (2.6) ดังนี้

$$\frac{b''(\theta_i)}{a(\phi)} = E\left[\frac{y_i - b'(\theta_i)}{a(\phi)}\right]^2 = \frac{Var(y_\mathrm{i})}{[a(\phi)]^2}$$

ดังนั้น ค่าความแปรปรวนของค่าสังเกตใด ๆ คือ

$$Var(y_i) = \; b''(\theta_i)\, a(\phi) \quad (2.8)$$

2.1.2 ส่วนประกอบเชิงระบบ (Systematic Component)

ส่วนประกอบเชิงระบบเป็นส่วนของเซตของตัวแปรอธิบายที่ระบบหรือรูปแบบเชิงเส้นในเทอมของพารามิเตอร์ หรือผลรวมเชิงเส้น Xβ ซึ่งทำหน้าที่เชื่อมกับเวกเตอร์ η เรียก Xβ ว่าเป็นส่วนประกอบเชิงระบบ ดังนี้

$$\eta = X\beta \tag{2.9}$$

หรือ

$$\eta_i \ = \sum_{j=1}^{p} \beta_j X_{ij}\,, i = 1, \dots, N$$

เมื่อ η แทน ตัวพยากรณ์เชิงเส้น (Linear Predictor) หรือ $\eta = (\eta_1, \dots, \eta_N)'$

X แทน เมทริกซ์ของตัวแปรอธิบาย ที่มีขนาด $N \times p$

β แทน เมทริกซ์ของพารามิเตอร์ $\beta = \left(\beta_1, \dots, \beta_p\right)'$

2.1.3 ส่วนประกอบฟังก์ชันเชื่อมโยง (Link Functions)

ส่วนประกอบฟังก์ชันเชื่อมโยง คือฟังก์ชันของค่าเฉลี่ยของตัวแปรตอบสนองเชิงสุ่ม Y ที่ใช้เป็นฟังก์ชันเชื่อมโยงระหว่างส่วนประกอบเชิงสุ่มและส่วนประกอบเชิงระบบเข้าด้วยกัน

ให้ $\mu_i = E(y_i), i = 1, \dots, N$ โดยที่ μ_i มีความสัมพันธ์กับ η_i ในรูปฟังก์ชันของค่าเฉลี่ย ดังนี้

$$\eta_i \ = g(\mu_i) = \sum_{j=1}^{p} \beta_j X_{ij}\,, i = 1, \dots, N$$

เมื่อ $g(\mu_i)$ แทน ฟังก์ชันที่สามารถหาค่าอนุพันธ์ได้ทางเดียว (Monotonic Differentiable Function) ซึ่งเรียกว่าฟังก์ชันเชื่อมโยง (Link Function) และ p แทน จำนวนตัวแปรอธิบาย ตัวอย่างของฟังก์ชันเชื่อมโยงได้แก่ ฟังก์ชันเชื่อมโยงเอกลักษณ์ (Identity Link) ฟังก์ชันเชื่อมโยงล็อก (Log Link) ฟังก์ชันเชื่อมโยงโพรบิต (Probit Link) ฟังก์ชันเชื่อมโยงโคชี (Cauchy Link) และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก (Complementary Log-Log Link) เป็นต้น เนื่องจากฟังก์ชันเชื่อมโยงฟังก์ชันหนึ่งต่อหนึ่ง (One-to-One Function) ดังนั้นสามารถใช้วิธีการแปลงผกผัน (Inverse Transformation Method) ฟังก์ชันให้อยู่ในรูปได้ดังนี้

$$\mu_i \ = \ g^{-1}(X_i'\beta)$$

โดยปกติแล้วตัวแบบสำหรับ μ_i จะมีความซับซ้อนมากกว่าตัวแบบสำหรับ η_i สำหรับตัวแบบที่ต้องการเชื่อมโยงระหว่างฟังก์ชันค่าเฉลี่ยของ Y กับเซตของตัวแปรอธิบาย X จะมีรูปแบบทั่วไปคือ

$$g(\mu_i) = \sum_{j=1}^{p} \beta_j X_{ij}\,, i = 1, \dots, N$$

สำหรับฟังก์ชันเชื่อมโยงที่ได้จากการพิจารณาพารามิเตอร์ธรรมชาติ (Natural Parameter) จะเรียกฟังก์ชันเชื่อมโยงเหล่านี้ว่าเป็นฟังก์ชันเชื่อมโยงคานอนิคัล (Canonical Link) ซึ่งจะทำการแปลงค่าเฉลี่ยให้อยู่ในเทอมพารามิเตอร์ นั่นคือ

$$g(\mu_i) = \eta_i = \sum_{j=1}^{p} \beta_j X_{ij}, i = 1, \dots, N$$

ตัวอย่างฟังก์ชันเชื่อมโยงคานอนิคัลของแต่ละการแจกแจงได้แก่ การแจกแจงปรกติ (Normal Distribution) จะมีฟังก์ชันเชื่อมโยงเอกลักษณ์เป็นฟังก์ชันเชื่อมโยงคานอนิคัล สำหรับการแจกแจงทวินาม (Binomial Distribution) จะมีฟังก์ชันเชื่อมโยงลอจิตเป็นฟังก์ชันเชื่อมโยงคานอนิคัล และสำหรับการแจกแจงปัวซง (Poisson Distribution) จะมีฟังก์ชันเชื่อมโยงล็อกเป็นฟังก์ชันเชื่อมโยงคานอนิคัล เป็นต้น สำหรับประโยชน์ของการใช้ฟังก์ชันเชื่อมโยงคานอนิคัล คือ จะให้ตัวสถิติพอเพียงที่ต่ำที่สุด (Minimal Sufficient Statistics) สำหรับพารามิเตอร์ β แต่อย่างไรก็ตาม ยังสามารถใช้ฟังก์ชันเชื่อมโยงอื่น ๆ โดยที่ไม่จำเป็นจะต้องใช้เฉพาะฟังก์ชันเชื่อมโยงคานอนิคัลดังข้างต้นที่กล่าวมา

2.2 การประยุกต์ใช้ข้อมูลกับตัวแบบเชิงเส้นนัยทั่วไปสำหรับการแจกแจงต่าง ๆ

Peter K Dunn (Department of Mathematics and Computing University of Southern Queensland) ได้กล่าวถึงการแจกแจงที่เหมาะสมในแต่ละประเภทข้อมูลที่สนใจศึกษา โดยใช้ตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) สำหรับการแจกแจงเกาส์เซียน (Gaussian Distribution) การแจกแจงปัวซง (Poisson Distribution) การแจกแจงทวินาม (Binomial Distribution) การแจกแจงแกมมา (Gamma Distribution) และการแจกแจงผกผันเกาส์เซียน (Inverse Gaussian Distribution) โดยมีรายละเอียดดังตารางที่ 2.1 ดังนี้

ตาราง 2.1 แสดงการแจกแจงที่เหมาะสมในแต่ละประเภทข้อมูล

การแจกแจง	ประเภทข้อมูล
การแจกแจงเกาส์เซียน (Gaussian Distribution)	Normal Distribution
การแจกแจงปัวซง (Poisson Distribution)	Counts
การแจกแจงทวินาม (Binomial Distribution)	Proportions
การแจกแจงแกมมา (Gamma Distribution)	Positive Continuous data
การแจกแจงผกผันเกาส์เซียน (Inverse Gaussian Distribution)	Positive Continuous data

นอกจากนี้ Peter K Dunn ได้กล่าวถึงฟังก์ชันเชื่อมโยงที่เหมาะสมสำหรับตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) สำหรับการแจกแจงเกาส์เซียน (Gaussian Distribution) การแจกแจงปัวซง (Poisson Distribution) การแจกแจงทวินาม (Binomial Distribution) การแจกแจงแกมมา (Gamma Distribution) และการแจกแจงผกผันเกาส์เซียน (Inverse Gaussian Distribution) แสดงดังตาราง 2.2 ดังนี้

ตาราง 2.2 แสดงฟังก์ชันเชื่อมโยงที่เหมาะสมสำหรับตัวแบบเชิงเส้นนัยทั่วไปของการแจกแจง ต่าง ๆ

ฟังก์ชันเชื่อมโยง		Gaussian Distribution	Poisson Distribution	Binomial Distribution	Gamma Distribution	Inverse Gaussian Distribution
Identity	$\mu = \eta$	✓	✓		✓	✓
Log	$\log(\mu) = \eta$	✓	✓		✓	✓
Inverse	$\frac{1}{\mu} = \eta$	✓			✓	✓
Sqrt	$\sqrt{\mu} = \eta$		✓			
Logit	$logit(\mu) = \eta$			✓		
Probit	$probit(\mu) = \eta$			✓		
cauchit	$cauchit(\mu) = \eta$			✓		
Cloglog	$cloglog(\mu) = \eta$			✓		
1/mu^2	$\frac{1}{\mu^2} = \eta$					✓

หมายเหตุ : กำหนดให้สัญลักษณ์ ✓ แทน ฟังก์ชันเชื่อมโยงที่เหมาะสมสำหรับแต่ละการแจกแจง

จากตาราง 2.2 จะเห็นได้ว่าแต่ละการแจกแจงจะมีฟังก์ชันเชื่อมโยงที่เหมาะสม ตัวอย่างเช่น ฟังก์ชันเชื่อมโยงลอจิตที่ถูกใช้สำหรับการถดถอยทวินาม (Binomial Regression) จะให้ผลที่ดีและง่ายต่อการอธิบายผลของพารามิเตอร์ การถดถอย (Regression Parameters) แต่ก็ไม่สามารถยืนยันได้ว่าจะให้ค่าประมาณที่ดีที่สุดเสมอไปโดยทั่วไปแล้วการเลือกฟังก์ชันเชื่อมโยงขึ้นอยู่กับวิจารณญาณของผู้วิจัย อย่างไรก็ตามการเลือกฟังก์ชันเชื่อมโยงทีไม่เหมาะสมจะส่งผลต่อความเอนเอียง (Bias) เป็นอย่างมากต่อพารามิเตอร์การถดถอยและค่าประมาณของตัวแปรตอบสนอง ดังนั้นการเลือกฟังก์ชันเชื่อมโยงที่เหมาะสมยังคงเป็นสิ่งสำคัญ

2.3 การแจกแจงทวินาม (Binomial Distribution)

การแจกแจงแบบทวินามเป็นการแจกแจงความน่าจะเป็นที่อธิบายความน่าจะเป็นของความสำเร็จจำนวนหนึ่งในการทดลองอิสระในจำนวนคงที่ โดยมีผลลัพธ์ที่เป็นไปได้สองอย่าง เช่น สำเร็จหรือล้มเหลว

หัวหรือก้อย ใช่หรือไม่ใช่ เป็นแนวคิดพื้นฐานในทฤษฎีความน่าจะเป็นและมีการนำไปใช้มากมายในด้านสถิติ วิทยาศาสตร์ และวิศวกรรม เป็นต้น

เมื่อทำการทดลองแบบแบร์นูลลีซ้ำ ๆ กัน จำนวน n ครั้ง ซึ่งแต่ละครั้งเป็นอิสระต่อกัน และเราสนใจครั้งทั้งหมดของการเกิดความสำเร็จ ความน่าจะเป็นที่จะเกิดผลลัพธ์สำเร็จเท่ากับ π และความน่าจะเป็นที่จะเกิดผลลัพธ์ไม่สำเร็จคือ $1-\pi$ จะต้องคงที่ตลอดของการทดลองแบบแบร์นูลลีแต่ละครั้ง

กำหนดให้ Y เป็นตัวแปรสุ่มที่มีการแจกแจงทวินาม ที่มีพารามิเตอร์ n และ π เขียนแทนด้วย $Y \sim Binomial(n,\pi)$ ซึ่งมีฟังก์ชันมวลความน่าจะเป็น ดังนี้

$$f(Y;n,p) = \binom{n}{y}\pi^{y}(1-\pi)^{n-y}\,; y = 0,1,\dots,n$$

โดย n หมายถึง จำนวนครั้งของการทดลอง

π หมายถึง ความน่าจะเป็นที่จะเกิดผลลัพธ์สำเร็จ

ค่าคาดหวังและความแปรปรวนของการแจกแจงทวินาม คือ

$$E(Y) = n\pi \text{ และ } Var(Y) = n\pi(1-\pi)$$

2.4 ฟังก์ชันเชื่อมโยง (Link Function)

ใน GLMs ฟังก์ชันเชื่อมโยงที่ได้จากการพิจารณาพารามิเตอร์ธรรมชาติจะถูกเรียกว่าฟังก์ชันคานอนิคัล (Canonical Links) เช่น การใช้ฟังก์ชันเชื่อมโยงลอจิตสำหรับการแจกแจงทวินาม ฟังก์ชันเชื่อมโยงล็อกสำหรับการแจกแจงปัวซง และฟังก์ชันเชื่อมโยงอินเวอร์สกำลังสอง (Inverse Squared Link) สำหรับการแจกแจงอินเวอร์สเกาส์เซียน (Inverse Gaussian Distribution) อย่างไรก็ตาม ยังมีฟังก์ชันเชื่อมโยงอื่น ๆ มากกว่าฟังก์ชันคานอนิคัลเหล่านี้ที่สามารถเชื่อมโยงระหว่าง ส่วนประกอบเชิงเส้นเชิงระบบเส้นเชิงระบบ (Systematic Linear Component) ไปสู่ช่วง [0,1] นอกจากนี้ แม้ว่าฟังก์ชันเชื่อมโยงคานอนิคัล (Canonical Link) ใน GLM เช่น ฟังก์ชันเชื่อมโยงลอจิตที่ถูกใช้สำหรับการถดถอยทวินาม (Binomial Regression) จะให้ผลที่ดีและง่ายต่อการอธิบายผลของพารามิเตอร์ การถดถอย (Regression Parameters) แต่ก็ไม่สามารถยืนยันได้ว่าจะให้ค่าประมาณที่ดีที่สุดเสมอไปโดยทั่วไปแล้วการเลือกฟังก์ชันเชื่อมโยงขึ้นอยู่กับวิจารณญาณของผู้วิจัย อย่างไรก็ตามการเลือกฟังก์ชันเชื่อมโยงที่ผิดจะส่งผลต่อความเอนเอียง (Bias) เป็นอย่างมากต่อพารามิเตอร์การถดถอยและค่าประมาณของตัวแปรตอบสนอง ดังนั้นการเลือกฟังก์ชันเชื่อมโยงที่เหมาะสมยังคงเป็นสิ่งสำคัญ

งานวิจัยนี้ได้สนใจฟังก์ชันเชื่องโยงสำหรับจำนวน 3 ฟังก์ชัน ได้แก่ ฟังก์ชันเชื่อมโยงลอจิต ฟังก์ชันเชื่อมโยงโพรบิต และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก โดยมีรายละเอียดของแต่ละฟังก์ชันดังนี้

ตาราง 2.1 แสดงรายละเอียดฟังก์ชันเชื่อมโยงลอจิต ฟังก์ชันเชื่อมโยงโพรบิต และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก

ฟังก์เชื่อมโยง	$\boldsymbol{\eta}$	$\boldsymbol{c.d.f.}$
ฟังก์ชันเชื่อมโยงลอจิต	$\log\left[\frac{\pi_i}{(1-\pi_i)}\right]$	$\pi_i = \frac{\exp(\eta)}{1+\exp(\eta)}$
ฟังก์ชันเชื่อมโยงโพรบิต	$\phi^{-1}(\pi_i)$	$\pi_i = \phi(\eta)$
ฟังก์ชันเชื่อมโยงคอมพลีเมนทนรีล็อก-ล็อก	$\log[-log(1-\pi_i)]$	$\pi_i = 1 - \exp\,(-\exp\,(\eta))$

โดยปกติแล้วฟังก์ชันเชื่อมโยงลอจิตจะถูกใช้สำหรับข้อมูลแบบทวิภาค (Binary Data) นอกจากนี้ยังมีฟังก์ชันเชื่อมโยงอื่น ๆ ที่สามารถนำมาพิจารณาแทนฟังก์ชันเชื่อมโยงลอจิตได้ ซึ่งในความเป็นจริงนั้นการแปลงข้อมูลใด ๆ ที่ส่งต่อค่าความน่าจะเป็นสู่ค่าจริงควรจะใช้ตัวแบบนัยเชิงเส้นทั่วไป (GLMs) โดยที่การแปลงยังคงเป็นแบบหนึ่งต่อหนึ่งอยู่ (One-to-one) และเป็นค่าต่อเนื่องที่สามารถหาอนุพันธ์ได้

สมมติให้ $F(\cdot)$ คือ ฟังก์ชันการแจกแจงสะสม (Cumulative Distribution Function) ของตัวแปรสุ่มบนจำนวนจริง (Real Line) ที่สามารถเขียนได้ดังนี้

$$\pi_i = F(\eta_i), -\infty < \eta_i < \infty \tag{2.10}$$

เมื่อใช้การแปลงผกผันจะทำให้ได้ฟังก์ชันเชื่อมโยงดังนี้

$$\eta_i = F^{-1}(\pi_i), \qquad 0 < \pi_i < 1 \tag{2.11}$$

เมื่อกำหนดให้ตัวแปรตามขึ้นเวกเตอร์ของตัวแปรร่วม X จากความสัมพันธ์ดังกล่าวให้สามารถใช้ตัวแบบเชิงเส้นแบบธรรมดา (Ordinary Linear Model) สำหรับตัวแปรแฝง (Latent Variable) ซึ่งสามารถเขียนได้ดังนี้

$$z_i = x_i'\beta + U_i$$

เมื่อ z_i คือ ตัวแปรแฝง

β คือ เวกเตอร์ของสัมประสิทธิ์สหสัมพันธ์ของตัวแปรร่วม x_i

U_i คือ ส่วนของความคลาดเคลื่อน (Error Term) และสมมติให้มีการแจกแจงสะสม (c.d.f)

$F(u)$ ซึ่งไม่จำเป็นว่าจะต้องเป็นการแจกแจงปรกติ

2.4.1 ฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function)

อีกทางเลือกหนึ่งสำหรับการแจกแจงปรกติ คือ การแจกแจงลอจิสติกมาตรฐาน (Strandard Logistic Distribution) ซึ่งมีรูปร่างคล้ายการแจกแจงปรกติแต่มีข้อดี คือ สามารถเขียนให้อยู่ในรูปปิดได้ดังนี้

$$\pi_i = F(\eta_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \tag{2.12}$$

การแจกแจงลอจิสติกมาตรฐานมีลักษณะสมมาตรที่มีค่าเฉลี่ยเท่ากับศูนย์และความแปรปรวนเท่ากับ $\pi^2/3$ โดยรูปร่างของการแจกแจงดังกล่าวคล้ายคลึงการแจกแจงปรกติเป็นอย่างมากเพียงแต่มีลักษณะหางที่มากกว่า การแปลงผกผัน (Inverse Transformation) (2.12) จะให้ฟังก์ชันเชื่อมโยงลอจิต (2.13) ดังนี้

$$\eta_i = F^{-1}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) \tag{2.13}$$

ดังนั้นค่าสัมประสิทธิ์สหสัมพันธ์ในตัวแบบถดถอยลอจิต นอกจากจะสามารถอธิบายในรูปแบบของล็อก-ออดส์ (Log-Odds) แล้วยังสามารถอธิบายในส่วนของอิทธิพลของตัวแปรร่วมสำหรับตัวแปรแฝงที่มาจากตัวแบบเชิงเส้นกับส่วนของความคลาดเคลื่อนลอจิสติกได้อีกด้วย

2.4.2 ฟังก์ชันเชื่อมโยงโพรบิต (Probit Link Function)

ในกรณีที่มีการแจกแจงของส่วนของความคลาดเคลื่อนเป็นการแจกแจงปรกติมาตรฐาน (Standard Normal Distribution) $U_i \sim N(0,1)$ จะสามารถเขียนให้อยู่ในรูปแบบของ (2.14) ได้ดังนี้

$$\pi_i = \phi(\eta_i) \tag{2.14}$$

เมื่อ ϕ คือ ฟังก์ชันการแจกแจงสะสมปรกติมาตรฐาน(Standard Normal Cumulative Density Function) ซึ่งมีลักษณะสมมาตร และเมื่อจัดตามรูปแบบ (2.14) จะได้ฟังก์ชันเชื่อมโยง (2.15) ซึ่งเรียกได้ว่าฟังก์ชันเชื่อมโยงโพรบิต (Probit Link)

$$\eta_i = \phi^{-1}(\pi_i) \tag{2.15}$$

สำหรับกรณีทั่วไปเมื่อส่วนของความคลาดเคลื่อน $U_i \sim N(0,\sigma^2)$ มีแจกแจงปรกติที่มีความแปรปรวนเท่ากับ σ^2 ซึ่งสามารถหาความน่าจะเป็น π_i ได้โดยการหาร σ เพื่อทำให้อยู่ในรูปตัวแปรปรกติมาตรฐาน ดังนี้

$$\begin{aligned} \pi_i &= P\{z_i > 0\} \\ &= P\{U_i > -x_i'\beta\} \\ &= P\{U_i/\sigma > -x_i'\beta\} \end{aligned}$$

$$= 1 - P\{U_i/\sigma \leq -x_i'\beta\}$$

$$= 1 - \phi(-x_i'\beta)$$

$$= \phi(x_i'\beta)$$

ดังนั้นเมื่อ $\sigma = 1$ จะได้ว่า $\pi_i = \phi(x_i'\beta)$

ข้อเสียสำหรับการใช้การแจกแจงปรกติเป็นฟังก์ชันเชื่อมโยงสำหรับตัวแบบทวิภาค (Binary Response Models) คือ การแจกแจงสะสมไม่มีรูปแบบปิด (Closed Form) แม้ว่าการประมาณเชิงตัวเลข (Numerical Approximations) และขั้นตอนวิธีการทางคอมพิวเตอร์จะสามารถคำนวณได้ทั้งการหาอนุพันธ์และส่วนกลับ (โพรบิต) ได้ก็ตาม

ฟังก์ชันเชื่อมโยงลอจิตและฟังก์ชันเชื่อมโยงโพรบิตมีลักษณะที่เกือบเป็นฟังก์ชันเชิงเส้นของกันและกันสำหรับค่าของ π_i ที่อยู่ในช่วง 0.1 ถึง 0.9 ดังนั้นทั้งสองฟังก์ชันจึงมีแนวโน้มที่จะให้ผลลัพธ์ที่ใกล้เคียงกัน การเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ของทั้งสองฟังก์ชันดังกล่าวนั้น ควรทำการปรับค่าก่อนเนื่องจากทั้งสองฟังก์ชันมีความแปรปรวนที่แตกต่างกันโดยตัวแบบโพรบิตจะทำการกำหนด $\sigma = 1$ ในขณะที่ตัวแบบลอจิตจะกำหนดให้ $\sigma = \pi/\sqrt{3}$ ดังนั้นค่าสัมประสิทธิ์สหสัมพันธ์ในตัวแบบลอจิตควรถูกปรับค่าให้มาตรฐาน (Standardized) ด้วยการหารด้วย $\sigma = \pi/\sqrt{3}$ ก่อนจะทำการเปรียบเทียบกับค่าสัมประสิทธิ์สหสัมพันธ์ในตัวแบบโพรบิต (Rodríguez, G. (2007))

2.4.3 ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก (Complementary Log-Log Link)

ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อกหรือตัวแบบล็อก-ล็อกเติมเต็ม เป็นส่วนกลับของฟังก์ชันการแจกแจงสะสมของค่าสุดขีด (Extreme Value) หรือเรียกอีกชื่อหนึ่งว่า การแจกแจงกอมเพอร์ซ (Gompertz Distribution) ซึ่งมีฟังก์ชันการแจกแจงสะสม คือ

$$\pi_i = F(\eta_i) = 1 - \exp(-\exp(\eta_i)) \tag{2.16}$$

ซึ่งส่วนกลับของฟังก์ชัน (2.16) ทำให้ได้ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก (2.17) ดังนี้

$$\eta_i = \log(-log(1 - \pi_i)) \tag{2.17}$$

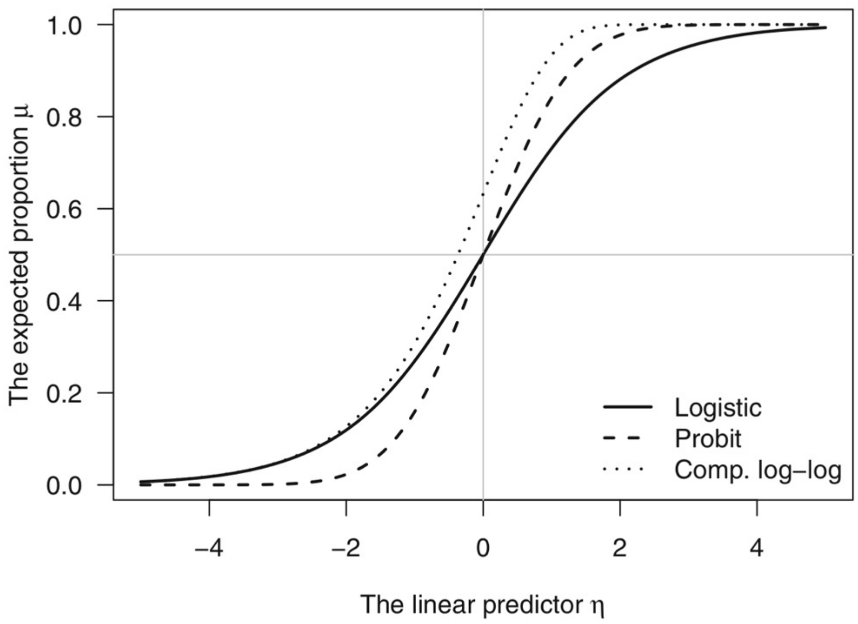
เมื่อค่า π_i ของฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อกมีค่าน้อยจะทำให้มีลักษณะใกล้เคียงกับฟังก์ชันเชื่อมโยงลอจิตและเมื่อค่า π_i เพิ่มมากขึ้นจะทำให้การเข้าสู่ค่าอนันต์ของฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก ช้ากว่าฟังก์ชันเชื่อมโยงลอจิต

สำหรับการเลือกฟังก์ชันเชื่อมโยงที่ให้เฉพาะเจาะจงมากขึ้นจากรูปแบบของตัวแฝงทั่วไป เมื่อสมมติให้ U_i มีการแจกแจงค่าสุดขีดมาตรฐาน (Standard Extreme Value Distribution) ที่มีฟังก์ชันการแจกแจงสะสมเป็น

$$F(U_i) = \exp(-\exp(-U_i))$$

โดยส่วนกลับของการแจกแจงค่าสุดขีดมีลักษณะไม่สมมาตรที่มีหางยาวทางขวาและมีค่าเฉลี่ยเท่ากับ ค่าคงที่ออยเลอร์ (Euler’s constrant) 0.577 และมีค่าความแปรปรวนเท่ากับ $\pi^2/6 = 1.645$ ค่ามัธยฐานเท่ากับ $-\log(log(2)) = 0.367$ และมีค่าควอร์ไทล์เท่ากับ -0.327 และ 12.46

การแปลงผกผันของการแจกแจงสะสมค่าสุดขีดส่วนกลับและการประยุกต์ทำให้พบว่า ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก สามารถใช้ได้กับทั้งการแจกแจงของความคลาดเคลื่อนแบบสมมาตรและไม่สมมาตร ดังนั้นค่าสัมประสิทธิ์สหสัมพันธ์ในตัวแบบเชิงเส้นนัยทั่วไปที่มีตัวแปรตอบสนองแบบทวิภาคและฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อกสามารถอธิบายอิทธิพลของตัวแปรร่วมสำหรับตัวแปรแฝงที่อยู่ในตัวแบบเชิงเส้นกับค่าความคลาดเคลื่อนของค่าสุดขีดส่วนกลับ (Reverse Extreme Value Error) ได้เช่นกัน สำหรับการเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์กับค่าประมาณภายใต้ตัวแบบโพรบิตควรทำการปรับค่าให้เป็นมาตรฐานก่อน โดยการหารด้วย $\pi/6$ และหารด้วย $\sqrt{2}$ สำหรับการเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์กับค่าประมาณภายใต้ตัวแบบลอจิต หรืออาจจะทำการปรับค่าให้เป็นมาตรฐาน ทั้งค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแบบคอมพลีเมนทารีล็อก-ล็อกและตัวแบบลอจิต



ภาพที่ 2.1 กราฟฟังก์ชันการแจกแจงสะสมของฟังก์ชันเชื่อมโยง

2.5 ตัวแบบการถดถอยทวินาม (The Binomial Regression Model)

วิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรตามหรือตัวแปรตอบสนอง (Response Variable) กับตัวแปรอธิบาย (Independent Variable) โดยที่ตัวแปรตอบสนองเป็นตัวแปรไม่ต่อเนื่อง ที่มีค่าเป็นจำนวนเต็มไม่มีค่าติดลบ คือ การวิเคราะห์การถดถอยสำหรับข้อมูลจำนวนนับ โดยมีจุดมุ่งหมายเพื่อประมาณค่าพารามิเตอร์สัมประสิทธิ์การถดถอยของตัวแปรอธิบายที่ต้องการศึกษาส่งผลกระทบต่อตัวแปรตอบสนองและสามารถสร้างตัวแบบที่เหมาะสมกับข้อมูลเพื่ออธิบายความสัมพันธ์ของข้อมูลได้ การศึกษาครั้งนี้จะกล่าวถึงตัวแบบถดถอยข้อมูลเชิงนับสำหรับตัวแปรตามเป็นตัวแปรสุ่มไม่ต่อเนื่อง ซึ่ง Y_i มีการแจกแจงทวินาม โดยมีพารามิเตอร์ n_i และ π_i ซึ่งพารามิเตอร์ π_i มีความเกี่ยวข้องกับตัวแปรอธิบายความสัมพันธ์ดังกล่าวได้พัฒนาเป็นการถดถอยสำหรับ $Y_i|X_i$

2.5.1 ตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต

$$\log it(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = X_i'\beta = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} \qquad (2.18)$$

เมื่อ X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1)\times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

p คือ จำนวนตัวแปรอธิบาย

พารามิเตอร์ π ยังคงอยู่ในช่วง [0,1] จากสมการ (2.18) สามารถหา π ดังนี้

$$\pi_i = \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)} \qquad (2.19)$$

ฟังก์ชันมวลความน่าจะเป็นของ y_i เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{y_i}\left(\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)^{Y_i}\left(1-\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)^{n_i-Y_i} \qquad (2.20)$$

2.5.2 ตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงโพรบิต

$$Probit(\pi_i) = \phi^{-1}(\pi_i) = X_i'\beta = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} \qquad (2.21)$$

เมื่อ X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1)\times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

p คือ จำนวนตัวแปรอธิบาย

พารามิเตอร์ π ยังคงอยู่ในช่วง [0,1] จากสมการ (2.21) สามารถหา π ดังนี้

$$\pi_i \;=\; \phi(X_i'\beta) \tag{2.22}$$

ฟังก์ชันมวลความน่าจะเป็นของ Y_i เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{y_i}(\phi(X_i'\beta))^{Y_i}(1-\ \phi(X_i'\beta))^{n-\ Y_i} \tag{2.23}$$

2.5.3 ตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก

$$\log(-log(1-\pi_i\,))\;=\;X_i'\beta=\;\beta_0+\;\beta_1X_{i1}+..+\;\beta_pX_{ip} \tag{2.24}$$

เมื่อ X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1)\times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

p คือ จำนวนตัวแปรอธิบาย

พารามิเตอร์ π ยังคงอยู่ในช่วง [0,1] จากสมการ (2.24) สามารถหา π ดังนี้

$$\pi_i\;=1-\exp(-exp(\beta_0+\;\beta_1X_{i1}+..+\;\beta_{10}X_{i10})) \tag{2.25}$$

ฟังก์ชันมวลความน่าจะเป็นของ Y_i เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{Y_i}(1-\exp\,(-\exp\,(X_i'\beta)))^{Y_i}(\exp\,(-\exp\,(X_i'\beta)))^{n_i-\ Y_i} \tag{2.26}$$

2.6 วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation)

2.6.1 หลักการของวิธีภาวะน่าจะเป็นสูงสุด

การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด เป็นวิธีการที่ใช้กันแพร่หลายมากที่สุด มีแนวคิดมานานตั้งแต่คริสต์ศตวรรษที่ 18 คาร์ล ฟรีดริด เกาส์ (Karl Friedrich Gauss 1777-1855) และแดเนียล แบร์นูลลี (Daniel Bernoulli) ได้ใช้วิธีการนี้มาแล้ว ต่อมาในต้นศตวรรษ ที่ 20 นัลด์ ไอล์เมอร์ ฟิชเชอร์ (Ronald Aylmer Fisher 1890-1962) ได้ทำการศึกษาคุณสมบัติของวิธีการนี้ ทำให้มีผู้ใช้กันกว้างขึ้นและถือได้ว่าวิธีการนี้เป็นผลงานของฟิชเชอร์ โดยเขาได้นำเสนอผลงานเกี่ยวกับวิธีการนี้ในปี ค.ศ. 1912

พร้อมทั้งมีการปรับปรุงแก้ไขส่วนที่เกี่ยวข้องให้เหมาะสมขึ้นอีกด้วย นักสถิติคนอื่น ๆ ก็มีส่วนทำให้วิธีการนี้เป็นที่นิยมแพร่หลายยิ่งขึ้นด้วย

นิยาม 2.1 ให้ $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta), \theta \epsilon \Omega$ ฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) ของตัวอย่างสุ่ม คือ ฟังก์ชันความหนาแน่นร่วม $L = L(\theta; x_1, x_2, \dots, x_n)$ ของตัวอย่างสุ่มนั้นที่ถือว่าเป็นฟังก์ชันของพารามิเตอร์ θ นั่นคือ

$$L = L(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i;\theta)$$

นิยาม 2.2 ค่าของพารามิเตอร์ θ ในเทอมค่าสังเกตของตัวอย่างสุ่ม $X_1, X_2, \dots, X_n$ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็นมีค่าสูงสุด เรียกว่าตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE) ของ θ นั่นคือค่าของ $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ เป็น MLE ของ θ ก็ต่อเมื่อ
$L(\hat{\theta}) = L(\hat{\theta}(X_1, X_2, \dots, X_n))$ มีค่าสูงสุด

2.6.2 วิธีหาตัวประมาณภาวะน่าจะเป็นสูงสุด

เป็นวิธีการหาค่าของพารามิเตอร์ θ ที่ทำให้ฟังก์ชัน $L(\theta)$ สูงสุด ในการนี้มีข้อควรสังเกตดังต่อไปนี้

1. เป้าหมายในการหา MLE ของ θ คือการหาค่า θ เรียกว่า
$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ ที่ทำให้ $L(\hat{\theta}; x_1, x_2, \dots, x_n) \geq L = L(\theta; x_1, x_2, \dots, x_n)$ เมื่อ $\theta \epsilon \Omega$ และ $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$

2. ถ้าฟังก์ชันภาวะน่าจะเป็น $L(\theta)$ เป็นฟังก์ชันที่หาอนุพันธ์ได้ (Differentiable Function) เมื่อเทียบกับ θ อาจใช้อนุพันธ์หา MLE ของ θ ได้ เมื่อเรนจ์ของ $f(x;\theta)$ ไม่ได้ขึ้นอยู่กับ θ และ θ อยู่ในช่วงจำนวนจริงช่วงหนึ่ง ในกรณีดังกล่าว $\hat{\theta}$ คือรากของสมการ $\frac{\partial L}{\partial \theta} = 0$

เงื่อนไขพอเพียง (Sufficient Condition) ที่ $\hat{\theta}$ ทำให้ $L(\hat{\theta}) \geq L = L(\theta)$ เมื่อ $\theta \epsilon \Omega$ คือ $\frac{\partial^2 L}{\partial \theta^2} < 0$ เมื่อ $\theta = \hat{\theta}$

3. การใช้อนุพันธ์หา MLE ในหลายกรณีใช้ $\ln L$ จะสะดวกกว่าที่จะใช้ L

สังเกตว่า $\frac{\partial \ln L}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta}$

และ $L > 0$ ดังนั้น เมื่อ $\frac{\partial \ln L}{\partial \theta} = 0$ จะได้ $\frac{\partial L}{\partial \theta} = 0$ ด้วย

นอกจากนั้น เมื่อ $\frac{\partial^2 \ln L}{\partial \theta^2} < 0$ ก็จะทำให้ $\frac{\partial^2 L}{\partial \theta^2} < 0$ ด้วย

นิยาม 2.3 สมการที่ใช้หา MLE คือ $\frac{\partial L}{\partial \theta} = 0$ หรือ $\frac{\partial \ln L}{\partial \theta} = 0$ เรียกว่า สมการภาวะน่าจะเป็น (Likelihood Equation)

4. ในบางกรณี อาจไม่สามารถใช้อนุพันธ์ในการหา MLE เช่นเมื่อเรนจ์ของ $f(x;\theta)$ ขึ้นอยู่กับ θ

2.6.3 ปัญหาในการหาตัวประมาณภาวะน่าจะเป็นสูงสุด

ในการหาตัวประมาณภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ θ จะพยายามหาค่าของ θ ในเทอมของค่าสังเกต $X_1, X_2, \ldots, X_n$ ที่ทำให้ฟังก์ชันภาวะน่าจะเป็น $L(\theta)$ มีค่าสูงสุด เมื่อใช้อนุพันธ์นักสถิติจะหาค่าของ θ ดังกล่าวจากสมการภาวะน่าจะเป็น $\frac{\partial L}{\partial \theta} = 0$ หรือ $\frac{\partial \ln L}{\partial \theta} = 0$ คือหาจุดวิกฤติ (Critical Point) ที่จะทำให้ L สูงสุดนั้นเอง บางครั้งการแก้สมการดังกล่าวอาจทำได้ยาก เช่น เมื่อสมการภาวะน่าจะเป็นนั้นเป็นสมการระดับสูง ๆ หรือเป็นสมการเศษส่วนที่ซับซ้อน

ในบางกรณีไม่อาจใช้อนุพันธ์เพื่อหาค่าของพารามิเตอร์ θ ในเทอมของค่าสังเกต $X_1, X_2, \ldots, X_n$ ได้เช่นเมื่อ L หรือ $\ln L$ เป็นฟังก์ชันของ θ ที่หาอนุพันธ์ไม่ได้หรือเมื่อเรนจ์ของฟังก์ชันขึ้นอยู่กับ θ หรืออนุพันธ์ไม่มี θ อยู่ด้วย จึงต้องใช้วิธีการอื่น ๆ เช่น การสังเกตว่า $L(\theta)$ จะสูงสุดเมื่อไรหรือการเปรียบเทียบค่าของ $L(\theta)$ เมื่อค่าของ θ เปลี่ยนไป

ในกรณีที่ใช้อนุพันธ์ในการหา θ ที่ทำให้ $L(\theta)$ มีค่าสูงสุด แต่ไม่อาจแก้สมการภาวะน่าจะเป็นหรือแก้ได้ยาก อาจจะใช้การประมาณ (Approximation) ค่าของ $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ โดยอาศัยค่าสังเกต $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ ตามวิธีการของนิวตัน (Newton's approximation) ได้เมื่อ n มีค่ามาก

ทฤษฎีบท 2.1 ค่าประมาณของรากของสมการภาวะน่าจะเป็น $\frac{\partial \ln L}{\partial \theta} = 0$ ได้แก่

$$\hat{\theta} = \theta_0 + \left(\frac{\partial \ln L}{\partial \theta}\right)_{\theta=\theta_0} V(\hat{\theta}) \tag{2.27}$$

โดยที่ $V(\hat{\theta}) = -1/E(\frac{\partial^2 \ln L}{\partial \theta^2})$

เราอาจใช้สมการ (2.28) ซ้ำ ๆ โดยให้ θ_0 เป็นค่าประมาณเริ่มต้น $\hat{\theta}$ โดยที่ค่าของ $\hat{\theta}$ หาได้ในรอบแรกจะใช้เป็น θ_0 ในรอบที่สอง เช่นนี้ซ้ำ ๆ ได้หลายรอบจนกระทั่งได้ค่าของ $\hat{\theta}$ ที่ค่อนข้างนิ่ง คือ มีค่าต่างจากที่หาได้ในรอบก่อนไม่มากนัก

2.6.4 ตัวประมาณภาวะน่าจะเป็นสูงสุดของพารามิเตอร์หลายตัวแปร

ในกรณีที่มีพารามิเตอร์ $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ อาจหาตัวประมาณภาวะน่าจะเป็นสูงสุด ของ $\theta_1, \theta_2, \ldots, \theta_k$ ที่ปรากฏอยู่ในฟังก์ชันความหนาแน่น $f(x;\theta_1, \theta_2, \ldots, \theta_k)$ ได้ โดยใช้หลักเกณฑ์เดิม นอกจากนั้น ตัวแปรสุ่มแต่ละตัวของตัวอย่างสุ่ม $X_1, X_2, \ldots, X_n$ ยังอาจเป็นเวกเตอร์ได้ด้วย

นิยาม 2.4 เมื่อ $X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันความหนาแน่น $f(x;\theta_1, \theta_2, \ldots, \theta_k)$ ค่าของ $\theta_1, \theta_2, \ldots, \theta_k$ ในเทอมของค่าสังเกตของตัวอย่างสุ่มทำให้ฟังก์ชันภาวะน่าจะเป็น $L(\theta_1, \theta_2, \ldots, \theta_k)$ มีค่าสูงสุด เรียกว่า **ตัวประมาณภาวะน่าจะเป็นสูงสุด** ของ $\theta_1, \theta_2, \ldots, \theta_k$ นั่นคือ

$\hat{\theta}_1(X_1, X_2, \dots, X_n), \hat{\theta}_2(X_1, X_2, \dots, X_n), \dots, \hat{\theta}_k(X_1, X_2, \dots, X_n)$ เป็นตัวประมาณภาวะน่าจะเป็นสูงสุดของ $\theta_1, \theta_2, \dots, \theta_k$ ตามลำดับ ก็ต่อเมื่อ $L(\theta_1, \theta_2, \dots, \theta_k)$ มีค่าสูงสุด

เงื่อนไขจำเป็นที่ทำให้ $L(\theta_1, \theta_2, \dots, \theta_k)$ มีค่าสูงสุดได้แก่ $\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_i} = 0, i = 1, \dots, k$ หรือสมมูลกันก็คือ $\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_j} = 0, j = 1, \dots, k$ และเงื่อนไขพอเพียงที่จะได้ L สูงที่สุด ได้แก่การที่เมทริกซ์ $\left(\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}\right)$ เป็น **เมทริกซ์นิเสธแน่นอน** (Negative definite matrix) ขนาด $k \times k$ เมื่อสามารถใช้อนุพันธ์ในการหาตัวประมาณภาวะน่าจะเป็นสูงสุดของพารามิเตอร์ $\theta_1, \theta_2, \dots, \theta_k$ ได้ เราเรียกระบบสมการ $\frac{\partial L}{\partial \theta_j} = 0$ หรือ $\frac{\partial \ln L}{\partial \theta_j} = 0$ ว่า **ระบบสมการภาวะน่าจะเป็น** (ประชุม สุวัติถี, 2553)

2.7 การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดสำหรับการถดถอยวินาม

ในการประมาณค่าพารามิเตอร์ที่ไม่ทราบค่าใช้หลักการภาวะน่าจะเป็นสูงสุด อธิบายได้ดังนี้ ให้ x_i คือเวกเตอร์ค่าสังเกตของตัวแปรอธิบาย และ y_i คือ ค่าสังเกตของตัวแปรตอบสนองเมื่อ n คือ ขนาดตัวอย่าง สามารถหาฟังก์ชันภาวะน่าจะเป็นของค่าสังเกต แต่ในหลายสถานการณ์พบว่า ไม่สามารถใช้วิธีการประมาณค่าดังกล่าวได้โดยตรง เนื่องจากสมการที่พบมีรูปแบบไม่เป็นเชิงเส้น (Non Linear) ในเทอมของพารามิเตอร์ โดยวิธีภาวะน่าจะเป็นสูงสุดจะใช้การคำนวณวิธีย้อนซ้ำเชิงตัวเลข (Numerical Iteratin) ทำให้ได้วิธีที่มีประสิทธิภาพมากขึ้น ซึ่งจะเรียกวิธีดังกล่าว การประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดแบบย้อนซ้ำมีหลายวิธี แต่ในที่นี้จะกล่าวถึงเฉพาะวิธีฟิชเชอร์สกอริง (Fisher’s Scoring Method) ซึ่งเป็นวิธีที่ใช้สำหรับการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิตในคำสั่ง glm ของโปรแกรม R
จากสมการที่ 2.21 สามารถหาฟังก์ชันภาวะน่าจะเป็นของค่าสังเกต ดังนี้

$$L(\beta|x,y) = \prod_{i=1}^{n}\left[\binom{n_i}{y_i}\left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)^{y_i}\left(1-\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)^{n_i-y_i}\right] \tag{2.28}$$

ผ่านฟังก์ชันเชื่อมโยงลอจิต

$$L(\beta|x,y) = \sum_{i=1}^{n}\left[\ln\binom{n_i}{y_i} + y_i\ln\left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right) + (n_i - y_i)\ln\left(\frac{1-\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)\right] \tag{2.29}$$

เมื่อหาอนุพันธ์อันดับ 1 เทียบกับพารามิเตอร์ที่ไม่ทราบค่า β และกำหนดค่าให้เท่ากับศูนย์ อนุพันธ์อับดับ 1 ของฟังก์ชัน L (Loglikelihood Function) เทียบกับพารามิเตอร์ β

$$\frac{\partial L(\beta|x,y)}{\partial\beta} = \sum_{i=1}^{n}\left[y_i - \left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)\right]x_i - \sum_{i=1}^{n}\left[(n_i - y_i) - \left(\frac{n_i}{(n_i - y_i)}\right)\left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)\right]$$

จะเห็นว่าไม่สามารถที่จะหาอนุพันธ์ได้โดยตรงเนื่องจากสมการอนุพันธ์อันดับ 1 ไม่มีรูปแบบปิด จึงนำหลักการแก้ปัญหาหาค่าตัวประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด (MLE) ตามหลักการระเบียบวิธีเชิงตัวเลข โดยในการประมาณค่าผ่านฟังก์ชันเชื่อมโยงโพรบิตและคอมพลีเมนทารีล็อก-ล็อก สามารถทำได้รูปแบบเดียวกัน

2.8 วิธีการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยทีละขั้น (Stepwise Selection) ของตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

เป็นวิธีการคัดเลือกตัวแปรอธิบายใดๆผสมผสานระหว่างวิธีการคัดเลือกตัวแปรอธิบายทั้งแบบการเพิ่มตัวแปร (Forward Selection) และการลดตัวแปรเข้าด้วยการ (Backward Elimination) ในส่วนของตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model) เกณฑ์ที่จะใช้ในการคัดเลือกตัวแปรที่จะเข้าหรือตัวแปรที่จะออกคือ AIC ซึ่งมีขั้นตอนสามารถทำได้ ดังนี้

1) เริ่มต้นด้วยการกำหนดตัวแบบเริ่มต้นที่ไม่มีตัวแปรอธิบาย (“null model”) หรือขั้นต่ำสุดที่ต้องพิจารณา คือ ตัวแบบที่ไม่มีตัวแปรอธิบาย และกำหนดตัวแบบสูงสุดที่มีตัวแปรอธิบายทุกตัว (“full model”) หรือขั้นสูงสุดที่ต้องพิจารณาคือตัวแบบที่มีตัวแปรอธิบายทุกตัว
2) ทุกตัวแปรอิสระสามารถถูกคัดเลือกเพิ่มเข้าไปในตัวแบบเริ่มต้น และตัวแปรอธิบายที่อยู่ภายในตัวแบบสามารถคัดตัวแปรอธิบายออกได้ทีละตัวในแต่ละขั้นตอนจนกว่าค่า AIC จะไม่ดีขึ้น โดยเรียกใช้ฟังก์ชัน stepAIC( ) ซึ่งอยู่ใน library(MASS) ในโปรแกรม R
3) สุดท้ายจะได้ตัวแบบที่ดีที่สุด หรือตัวแบบที่ให้ค่า AIC ดีที่สุด

ขั้นตอนหรืออัลกอริทึมในโปรแกรม R มีดังนี้

```
> fullModel_cloglog <- glm(cbind(Yi, ni - Yi) ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10, family = binomial(link = "cloglog"))
> nullModel_cloglog <- glm(cbind(Yi, ni - Yi) ~ 1, family = binomial(link = "cloglog"))
> library(MASS)
> bothways_cloglog <- stepAIC(nullModel_cloglog, direction = "both", scope = list(upper = fullModel_cloglog, lower = nullModel_cloglog))
```

Start: AIC=161.44

cbind(Yi, ni - Yi) ~ 1

	Df	Deviance	AIC
+ x5	1	58.288	141.92
+ x9	1	66.980	150.62
+ x7	1	67.096	150.73
+ x1	1	74.750	158.38
<none>		79.804	161.44
+ x6	1	78.083	161.72
+ x8	1	79.392	163.03
+ x4	1	79.437	163.07

+ x10	1	79.508	163.14
+ x3	1	79.559	163.19
+ x2	1	79.746	163.38

Step: AIC=141.92

cbind(Yi, ni - Yi) ~ x5

	Df	Deviance	AIC
+ x7	1	45.303	130.94
+ x8	1	55.081	140.72
<none>		58.288	141.92
+ x6	1	56.388	142.02
+ x2	1	56.903	142.54
+ x9	1	57.106	142.74
+ x1	1	57.775	143.41
+ x4	1	58.221	143.86
+ x3	1	58.273	143.91
+ x10	1	58.277	143.91
- x5	1	79.804	161.44

Step: AIC=130.94

cbind(Yi, ni - Yi) ~ x5 + x7

	Df	Deviance	AIC
+ x8	1	41.787	129.42
+ x4	1	42.166	129.80
+ x3	1	43.223	130.86
<none>		45.303	130.94
+ x1	1	43.941	131.57
+ x10	1	45.179	132.81
+ x6	1	45.190	132.82
+ x2	1	45.243	132.88
+ x9	1	45.301	132.94
- x7	1	58.288	141.92
- x5	1	67.096	150.73

Step: AIC=129.42

cbind(Yi, ni - Yi) ~ x5 + x7 + x8

	Df	Deviance	AIC
<none>		41.787	129.42
+ x10	1	40.222	129.86
+ x2	1	41.217	130.85

+ x4	1	41.258	130.89
- x8	1	45.303	130.94
+ x1	1	41.347	130.98
+ x9	1	41.378	131.01
+ x3	1	41.571	131.21
+ x6	1	41.573	131.21
- x7	1	55.081	140.72
- x5	1	67.080	152.72

```
> formula(bothways_cloglog)
cbind(Yi, ni - Yi) ~ x5 + x7 + x8
```

2.9 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของตัวแบบ

2.9.1 เกณฑ์สารสนเทศของอะกะอิเกะ (Akaike’s Information Criterion: AIC)

เกณฑ์ AIC ถูกพัฒนาขึ้นโดย Hirotsuga Akaike ในปี ค.ศ 1974 ในขณะนั้น เกณฑ์ AIC ยังไม่เป็นที่นิยมมากนัก จนกระทั่งในศตวรรษที่ 21 เกณฑ์ AIC จึงถูกนำมาใช้มากขึ้นจนถึงปัจจุบัน เกณฑ์ AIC มักถูกใช้สถิติเพื่อแสดงผลในส่วนของตัวแบบทางสถิติ โดยทั่วไปแล้วเกณฑ์ AIC จะถูกพบได้ 2 รูปแบบดังสมการที่ (2.30) และ (2.31)

$$AIC = \frac{-2\ln(\hat{L}) + 2k}{n} = \frac{-2(\ln(\hat{L}) - k)}{n} \tag{2.30}$$

และ

$$AIC = -2\ln(\hat{L}) + 2k = -2\left(\ln(\hat{L}) - k\right) \tag{2.31}$$

เมื่อ $\hat{L}$ คือ ค่าสูงสุดของฟังก์ชันความน่าจะเป็นของตัวแบบการถดถอยที่ประมาณได้จากตัวประมาณภาวะน่าจะเป็นสูงสุด, k คือ จำนวนพารามิเตอร์และ n คือ จำนวนของค่าสังเกตในตัวแบบ

2.9.2 เกณฑ์สารสนเทศของเบส์ (Bayesian Information Criterion: BIC)

Gideon E. Schwarz นำเสนอเกณฑ์ Bayesian Information Criterion (BIC) เป็นครั้งแรกในปี 1978 ในเอกสารของเขาเรื่อง "Estimating the Dimensions of a Model" Schwarz ได้นำเสนอ BIC เพื่อเป็นเกณฑ์สำหรับการเลือกแบบจำลองที่ "ดีที่สุด" ต่อมา Bayesian Information Criterion (BIC) เป็นมาตรการทางสถิติที่ใช้โดยนักวิจัยและนักวิเคราะห์ในสาขาต่างๆ เช่น สถิติ แมชชีนเลิร์นนิง และวิทยาศาสตร์ข้อมูล สามารถใช้เพื่อเปรียบเทียบและเลือกรุ่นที่เหมาะสมที่สุดจากชุดของรุ่นตัวเลือก และเพื่อป้องกันการใช้งานเกินพอดี โดยการปรับรุ่นที่มีพารามิเตอร์จำนวนมาก BIC มักใช้ในด้านต่างๆ เช่น เศรษฐมิติ การเงิน ชีววิทยา จิตวิทยา และวิศวกรรม เป็นต้น โดยทั่วไปแล้วเกณฑ์ BIC มีรูปแบบสมการต่อไปนี้

$$BIC = -2\ln\left(\hat{L}\right) + k\ln(n) \tag{2.32}$$

เมื่อ $\hat{L}$ คือ ค่าสูงสุดของฟังก์ชันความน่าจะเป็นของตัวแบบการถดถอยที่ประมาณได้จากตัวประมาณภาวะน่าจะเป็นสูงสุด, k คือจำนวนพารามิเตอร์และ n คือ จำนวนของค่าสังเกตในตัวแบบ

2.10 ระบบการคัดเลือกบุคคลเข้าศึกษาต่อในระดับมหาวิทยาลัย (TCAS)

ระบบการคัดเลือกบุคคลเข้าศึกษาต่อในระดับมหาวิทยาลัย ชื่อเต็มว่ามีชื่อเต็มว่า Thai University Center Admission System ทั้งนี้ระบบการสอบแบบนี้เพื่อลดการสอบลง เหลือแค่การสอบกลางที่จัดสอบโดย สทศ. เพียง 3 อย่างเท่านั้น ได้แก่ GAT PAT , 9 วิชาสามัญ, ONET และกสพท. ซึ่งจะเริ่มนำมาใช้ในปีการศึกษา 2561 เป็นระบบที่ออกแบบโดยที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) การคัดเลือกของ TCAS จะมีด้วยกันทั้งหมด 5 รอบ โดยจะใช้เกณฑ์ดังต่อไปนี้

รอบที่ 1 คัดเลือกโดยการส่งแฟ้มสะสมผลงาน (Portfolio) ไม่มีการสอบข้อเขียน และไม่ได้เป็นการรับทั่วไป แต่จะดูผลงานและความสามารถเป็นหลัก ซึ่งแต่ละมหาวิทยาลัยจะคัดเลือกนักเรียนจำนวนหนึ่ง อาจจะมีการสัมภาษณ์หรือทดสอบทักษะเฉพาะทาง ช่วงที่เปิดรับสมัคร เดือนธันวาคม ถึง มกราคม

รอบที่ 2 สมัครโควตาแบบมีสอบข้อเขียน สำหรับนักเรียนในพื้นที่ ที่ทางมหาวิทยาลัยกำหนดและสามารถจัดสอบเองได้เลย หรือจะใช้ข้อสอบส่วนกลาง เช่น 9วิชาสามัญ หรือ GAT/PAT เพื่อคัดเลือกบุคคลเข้าศึกษา ช่วงที่เปิดรับสมัคร เดือนกุมภาพันธ์ – เมษายน

รอบที่ 3 การรับตรงร่วมกัน เป็นการรับตรงของแต่ละมหาวิทยาลัย ซึ่งโครงการรับตรงอย่าง กสพท. ก็รวมอยู่ในรอบนี้ด้วย โดยที่ทปอ. จะเป็นส่วนกลางในการรับสมัคร และมหาวิทยาลัยจะพิจารณาผลการคัดเลือก โดยผู้สมัครสามารถเลือกได้ 4 สาขาวิชา โดยไม่มีการเลือกอันดับ ช่วงที่เปิดรับสมัคร เดือนเมษายน ถึง พฤษภาคม

รอบที่ 4 การรับแบบ Admission เป็นการใช้เกณฑ์การคัดเลือกแบบ Admission โดยใช้ทั้งคะแนน GPAX, O-NET, GAT/PAT หรือคะแนนอื่นๆที่ทางมหาวิทยาลัยเป็นผู้กำหนด ซึ่งผู้สมัครสามารถเลือกได้ 4 สาขาวิชา โดยมีการเลือกลำดับ ช่วงที่เปิดรับสมัคร เดือนพฤษภาคม ถึง มิถุนายน

รอบที่ 5 การรับตรงแบบอิสระ ทางมหาวิทยาลัยเป็นผู้กำหนดขึ้นเองหรือการสอบวิชาเฉพาะ และส่งผลการคัดเลือกให้ทาง ทปอ.

2.11 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

Rindang Bangun Prasetyo และคณะ (ค.ศ.2019) ได้ทำการศึกษาถึงอัตราการลาออกกลางคันในชวาตะวันออก ประเทศอินโดนีเซียโดยนำมาประยุกต์ใช้กับแบบจำลองถดถอยทวินาม กล่าวคือแบบจำลองถดถอยเชิงเส้นคลาสสิกไม่เพียงพอ เมื่อตัวแปรตอบสนองเป็นจำนวนของความสำเร็จ จึงได้นำแบบจำลองถดถอยทวินามซึ่งวิเคราะห์ผ่านแบบจำลองเชิงเส้นโดยนัยทั่วไปกับฟังก์ชันเชื่อมโยง โดยฟังก์ชันเชื่อมโยงที่นำมาใช้ในการวิเคราะห์ข้อมูลอัตราการพ้นสภาพการเป็นนักศึกษาคือฟังก์ชัน logit, probit, complementary log-log (cloglog) โดย logit และ probit เป็นฟังก์ชันเชื่อมโยงสมมาตร และ cloglog เป็นฟังก์ชันเชื่อมโยงอสมมาตร โดยจะใช้เกณฑ์ AIC และ BIC ในการประเมินฟังก์ชันเชื่อมโยงทั้งสามประเภท โดยผู้วิจัยจะใช้ทั้งข้อมูลจริงและข้อมูลจำลอง จากการวิจัยพบว่า cloglog เป็นฟังก์ชันเชื่อมโยงที่ดีที่สุด

นารีรัตน์ ณ นุวงศ์ และ แสงหล้า ชัยมงคล (2552) ได้ศึกษาอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงที่ไม่ถูกต้องที่มีผลต่อสัมประสิทธิ์การตัดสินใจที่ปรับค่าสำหรับการวิเคราะห์การถดถอยลอจิสติกทวินามของฟังก์ชันเชื่อมโยงโพรบิตและฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก เมื่อกำหนดฟังก์ชันเชื่อมโยงลอจิตเป็นฟังก์ชันเชื่อมโยงที่แท้จริง ทำการศึกษาด้วยวิธีการจำลองข้อมูลและใช้เกณฑ์การพิจารณาความเอนเอียงสัมพัทธ์ของค่าประมาณมัธยฐานกับค่าสัมประสิทธิ์การตัดสินใจที่แท้จริงรวมถึงร้อยละของค่าประมาณที่อยู่นอกช่วง [0,1] จากการศึกษาพบว่า การกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องแบบโพรบิตมีอิทธิพลต่อสัมประสิทธิ์การตัดสินใจที่ปรับค่า R_{adj}^2 ไม่แตกต่างจากฟังก์ชันเชื่อมโยงลอจิต ในขณะที่ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อกจะมีผลให้ค่าประมาณ R_{adj}^2 เป็นค่าประมาณที่เอนเอียงทุกตัว ยกเว้น $R_{l,adj,SAS_{DEV}}^2$ โดยความเอนเอียงนี้จะขึ้นอยู่กับจำนวนตัวแปรอธิบาย ขนาดตัวอย่างและประเภทค่า R_{adj}^2 โดยค่า R_{adj}^2 ที่คำนวณด้วยวิธีกำลังสองน้อยที่สุดแบบสามัญจะได้รับอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องแบบคอมพลีเมนทารีล็อก-ล็อก น้อยกว่าค่า R_{adj}^2 ที่คำนวณวิธีความน่าจะเป็นสูงสุด

Gunduz และ Fokoue (2013) ได้ศึกษาเกี่ยวกับความแตกต่างและความเหมือนกันของฟังก์ชันเชื่อมโยงโพรบิตและฟังก์ชันเชื่อมโยงลอจิต และนำเสนอนิยามของโครงสร้างและความเท่าเทียมกันในด้านการทำนายสำหรับตัวแบบถดถอยทวินามภายใต้ฟังก์ชันเชื่อมโยงที่ศึกษา จากการศึกษาพบว่า ฟังก์ชันเชื่อมโยงโพรบิตและฟังก์ชันเชื่อมโยงลอจิตสามารถทำนายได้ถูกต้อง นอกจากนี้ยังมีฟังก์ชันเชื่อมโยงโคจิตและฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก ที่สามารถทำนายได้ถูกต้องเช่นกัน ซึ่งผลที่ได้จากการศึกษาทั้งแบบจำลองข้อมูลและการใช้ข้อมูลจริงมีความคล้ายคลึงกันและเป็นไปตามหลักทฤษฎีที่ได้พิสูจน์ไว้

Li (2014) ทำการศึกษาการเลือกฟังก์ชันเชื่อมโยงที่เหมาะสมสำหรับข้อมูลทวิภาค โดยใช้ฟังก์ชันเชื่อมโยงในการศึกษา 3 ฟังก์ชัน คือ ฟังก์ชันเชื่อมโยงลอจิต ฟังก์ชันเชื่อมโยงโพรบิตและ ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก ใช้ข้อมูลจริงในการศึกษาจำนวน 2 ชุดที่มีลักษณะสมมาตร และไม่สมมาตร พิจารณาความเหมาะสมของตัวแบบจากเกณฑ์ AIC และเกณฑ์ BIC จากผลการศึกษา พบว่าฟังก์ชันเชื่อมโยงลอจิตและฟังก์ชันเชื่อมโยงโพรบิตเหมาะสมกับข้อมูลที่มีลักษณะสมมาตร ในขณะที่ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อกเหมาะสมกับข้อมูลที่มีลักษณะไม่สมมาตร

Saddam Adams Damisa และคณะ (2017) ทำการเปรียบเทียบฟังก์ชันเชื่อมโยงที่แตกต่างกัน 3 ฟังก์ชัน ได้แก่ ฟังก์ชันเชื่อมโยงลอจิต ฟังก์ชันเชื่อมโยงโพรบิตและฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก สำหรับข้อมูลทวิภาคที่มีตัวอย่างขนาดเล็ก (< 1,000) โดยทำการจำลองข้อมูลตัวอย่างขนาด 50 ภายใต้สมมติฐานของความสมมาตรและไม่สมมาตรของข้อมูล เมื่อใช้เกณฑ์ AIC ใน การพิจารณาพบว่าฟังก์ชันเชื่อมโยงโพรบิตควรใช้เมื่อข้อมูลมีลักษณะสมมาตร ในขณะที่ควรจะใช้ ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก เมื่อข้อมูลมีลักษณะไม่สมมาตร

Wu และ Lord (2017) ได้ทำการทดสอบอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงที่ไม่ถูกต้องในตัวแบบถดถอยของ CMFs (Crash Modification Factors) จากการจำลองข้อมูลพบว่า การใช้ฟังก์ชันเชื่อมโยงผิดไม่ว่าจะสำหรับ 1 ตัวแปรหรือหลายตัวแปรจะทำให้การประมาณค่าเกิดความเอนเอียงได้

Roger Koenker and Jungmo Yoon (2009) ฟังก์ชันเชื่อมโยงลอจิตและโพรบิตมักจะนำมาประยุกต์ใช้กับการตอบสนองแบบไบนารีจำนวนมาก แต่ในฟังก์ชันเชื่อมโยงที่คลาสใหญ่กว่าอาจนำมาใช้เพียงบางครั้งมีการตรวจสอบสองพารามิเตอร์ของฟังก์ชันเชื่อมโยงได้ว่า gosset link ที่อิงตาม student t latent โมเดลตัวแปรแฝงที่มีองศาเสรีควบคุมลักษณะของหางและ pregibon link อิงจากวงศ์ Tukey ด้วยสองพารามิเตอร์ที่ควบคุมความเบ้และลักษณะหางมีการสำรวจเปรียบเทียบและอนุมานวิธีภาวะน่าจะเป็นสูงสุดและเบส์ พบว่าการระบุฟังก์ชันเชื่อมโยงที่ไม่ถูกต้องอาจทำให้เกิดข้อผิดพลาดได้ การประมาณค่าจุดแบบเบส์ผ่าน MCMC ทำได้ค่อนข้างใกล้เคียง MLE

Necla Gunduz และ Ernest Fokoue (2015) ได้ให้เหตุผลทางทฤษฎีและการคำนวณเพื่อสนับสนุนการอ้างว่าฟังก์ชันเชื่อมโยง probit และ logit มักใช้ในการจำแนกประเภทแบบไบนารี แม้จะมีการรับรู้อย่างกว้างขวางถึงความคล้ายคลึงกันอย่างมากระหว่างฟังก์ชันเชื่อมโยงทั้งสองนี้ แต่มีนักวิจัยเพียงไม่กี่คนที่ทุ่มเทเวลาเพื่อศึกษาอย่างเป็นทางการโดยมุ่งเป้าไปที่การสร้างและระบุคุณสมบัติทั้งหมดของความคล้ายคลึงกันและความแตกต่างให้เห็นอย่างชัดเจน โดยเสนอคำนิยามของทั้งความเทียบเท่าเชิงโครงสร้างและเชิงพยากรณ์ของแบบจำลองการถดถอยไบนารีตามฟังก์ชันเชื่อมโยงสองแบบ และสำรวจวิธีต่างๆ ที่คล้ายคลึงกันหรือแตกต่างกัน จากมุมมองของการวิเคราะห์เชิงพยากรณ์ ปรากฏว่าไม่เพียงแต่ probit และ logit จะสอดคล้องกันในการทำนายอย่างสมบูรณ์แบบเท่านั้น แต่ฟังก์ชันเชื่อมโยงอื่น ๆ เช่น Cauchit และ Complementary Log-Log Link ยังมีเปอร์เซ็นต์ความเทียบเท่าในการทำนายที่สูงมาก

บทที่ 3
วิธีการดำเนินงานวิจัย

งานวิจัยนี้สนใจศึกษาข้อมูลอัตราการพ้นสภาพการเป็นนักศึกษาและปัจจัยที่จะส่งผลต่ออัตราการพ้นสภาพโดยรวบรวมจากนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีการศึกษา 2561 โดยใช้การวิเคราะห์การถดถอยทวินาม (Binomial Regression Analysis) ภายใต้ฟังก์ชันการเชื่อมโยงที่แตกต่างกันสามฟังก์ชัน ได้แก่ ฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function), ฟังก์ชันเชื่อมโยงโพรบิต (Probit Link Function) และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก (Complementary Log-Log Link Function) และเปรียบเทียบประสิทธิภาพฟังก์ชันเชื่อมโยงทั้งสามโดยใช้เกณฑ์สารสนเทศของอะกะอิเกะ (AIC) และ เกณฑ์สารสนเทศของเบส์(BIC) นอกจากนี้ผู้วิจัยได้ศึกษาเชิงจำลองเนื่องจากข้อมูลคณะวิทยาศาสตร์และเทคโนโลยีมหาวิทยาลัยธรรมศาสตร์ประกอบด้วย 20 หลักสูตร ซึ่งเป็นหน่วยตัวอย่างที่มีขนาดเล็กซึ่งอาจจะส่งผลต่อการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยทวินามดังนั้นเพื่อให้เห็นถึงประสิทธิภาพของการประมาณในสถานการณ์ต่างๆที่มีขนาดตัวอย่างที่แตกต่างกันและฟังก์ชันเชื่อมโยงที่แตกต่างกัน ซึ่งเมื่อเพิ่มขนาดตัวอย่างมากขึ้นอาจจะส่งผลให้ประสิทธิภาพของตัวประมาณมีความแม่นยำ

3.1 แผนการดำเนินงานวิจัย

3.1.1 การประยุกต์ใช้กับข้อมูลจริง

ประยุกต์ใช้กับข้อมูลการพ้นสภาพการเป็นนักศึกษาของนักศึกษาปริญญาตรี
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ จำนวน 20 หลักสูตร ในปีที่เข้าศึกษา 2561

ตัวแปรตอบสนอง คือ จำนวนนักศึกษาที่พ้นสภาพในแต่ละหลักสูตร ($\mathrm{y_i}$)

ตัวแปรอธิบาย คือ หลักสูตรที่มีวิชาบังคับ C และหลักสูตร (X_1), จำนวนหน่วยกิตแต่ละหลักสูตร (X_2), ค่าเทอมในแต่ละหลักสูตร (X_3), ประเภทหลักสูตร(X_4), คะแนน Admission รอบ 3 สูงสุด ในแต่ละหลักสูตร (X_5), คะแนน Admission รอบ 3 ต่ำสุด ในแต่ละหลักสูตร(X_6), จำนวนนักศึกษาที่เข้ามารอบ 3 ในแต่ละหลักสูตร (X_7), คะแนน Admission รอบ 4 สูงสุด ในแต่ละหลักสูตร (X_8), คะแนน Admission รอบ 4 ต่ำสุด ในแต่ละหลักสูตร (X_9) และจำนวนนักศึกษาที่เข้ามารอบ 4 ในแต่ละหลักสูตร (X_{10})

ในการศึกษาโดยใช้ข้อมูลจริง Y_{i} มีการแจกแจงทวินาม ซึ่งเขียนแทนด้วย $Y_{\mathrm{i}} \sim Binomial\ (n_i, \pi_i)$ โดยที่ Y_{i} หมายถึง จำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร, n_i หมายถึง จำนวนนักศึกษาที่เข้ามาศึกษาในแต่ละหลักสูตรและ π_i หมายถึง อัตราการพ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร โดยวิเคราะห์ผ่านฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function), ฟังก์ชันเชื่อมโยงโพรบิต (Probit Link Function) และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก (Complementary Log-Log Link Function)

3.1.2 การศึกษาเชิงจำลอง

1. กำหนดขนาดตัวอย่าง

2. จำลองข้อมูลของตัวแปรอธิบาย (X)

3. กำหนดให้เวกเตอร์สัมประสิทธิ์การถดถอย

4. สร้างตัวแปรตอบสนอง Y_i มีการแจกแจงทวินาม $f(Y_i|X_i)$ เมื่อ π_i จะมีความสัมพันธ์กับตัวแปรอิสระ X_i ผ่านฟังก์ชันเชื่อมโยงลอจิต ฟังก์ชันเชื่อมโยงโพรบิต และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก จากการกำหนดสถานการณ์ข้างต้น p คือจำนวนตัวแปรอธิบาย กำหนดเท่ากับ 3 ตัวแปร และ k คือจำนวนพารามิเตอร์เท่ากับ 4

5. ประมาณค่าพารามิเตอร์ของตัวแบบวิธีภาวะน่าจะเป็นสูงสุด

6. เก็บรวบรวมและคำนวณค่าที่ใช้เป็นเกณฑ์ในการเปรียบเทียบมี 2 เกณฑ์ คือ ค่าเฉลี่ย AIC และ BIC โดยเฉลี่ย 1000 รอบ และร้อยละจำนวนครั้งที่ AIC และ BIC ต่ำสุด ในแต่ละรอบ จำนวน 1000 รอบ

3.2 ขั้นตอนการดำเนินงานวิจัย

ในงานวิจัยนี้เป็นการประยุกต์ใช้กับข้อมูลจริงและการศึกษาเชิงจำลองโดยประมวลผลจากโปรแกรม Rstudio เวอร์ชัน 2023.03.0+386 ภายใต้ข้อมูลที่มีสถานการณ์ต่าง ๆ โดยมีขั้นตอนในการดำเนินงานดังนี้

3.2.1 ขั้นตอนการประยุกต์ใช้กับข้อมูลจริง

1. ประยุกต์ใช้กับข้อมูลอัตราการพ้นสภาพของนักศึกษาปริญญาตรี คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ จำนวน 20 หลักสูตร โดยแบ่งเป็น โครงการปกติ จำนวน 15 หลักสูตร และ โครงการพิเศษ จำนวน 5 หลักสูตร โดยเข้ารับการศึกษาปี 2561 โดยมีหลักสูตรดังนี้

โครงการปกติจำนวน 15 หลักสูตร และโครงการพิเศษจำนวน 5 หลักสูตร ได้แก่

1. หลักสูตรสาขาวิชาสถิติ (โครงการปกติ)
2. หลักสูตรสาขาวิชาคณิตศาสตร์ (โครงการปกติ)
3. หลักสูตรสาขาวิชาคณิตศาสตร์ประยุกต์ (โครงการปกติ)
4. หลักสูตรสาขาวิชาสถิติ (โครงการพิเศษ)
5. หลักสูตรสาขาวิชาคณิตศาสตร์ (โครงการพิเศษ)
6. หลักสูตรสาขาวิชาวิทยาการประกันภัย (โครงการพิเศษ)
7. หลักสูตรสาขาวิชาวิทยาศาสตร์สิ่งแวดล้อม (โครงการปกติ)
8. หลักสูตรสาขาวิชาเทคโนโลยีการเกษตร (โครงการปกติ)
9. หลักสูตรสาขาวิชาเคมี (โครงการปกติ)
10. หลักสูตรสาขาวิชาวิทยาศาสตร์และเทคโนโลยีการอาหาร (โครงการปกติ)
11. หลักสูตรสาขาวิชาฟิสิกส์อิเล็กทรอนิกส์ (โครงการปกติ)

12. หลักสูตรสาขาวิชาเทคโนโลยีชีวภาพ (โครงการปกติ)
13. หลักสูตรสาขาวิชาฟิสิกส์ (โครงการปกติ)
14. หลักสูตรสาขาวิชาวัสดุศาสตร์ (โครงการปกติ)
15. หลักสูตรสาขาวิชาวิทยาศาสตร์และเทคโนโลยีสิ่งทอ (โครงการปกติ)
16. หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการปกติ)
17. หลักสูตรสาขาวิชาเทคโนโลยีเพื่อการพัฒนายั่งยืน (โครงการปกติ)
18. หลักสูตรสาขาวิชาเทคโนโลยีและนวัตกรรมทางอาหาร (โครงการปกติ)
19. หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการพิเศษ)
20. หลักสูตรสาขาวิชาเทคโนโลยีพลังงานชีวภาพและการแปรรูปเคมีชีวภาพ (โครงการพิเศษ)

ตัวแปรตอบสนอง คือ จำนวนนักศึกษาที่พ้นสภาพในแต่ละหลักสูตร (Y_i)
ตัวแปรอธิบาย คือ หลักสูตรที่มีวิชาบังคับ C และหลักสูตร (X_1), จำนวนหน่วยกิต แต่ละหลักสูตร (X_2), ค่าเทอมในแต่ละหลักสูตร (X_3), ประเภทหลักสูตร(X_4), คะแนน Admission รอบ 3 สูงสุด ในแต่ละหลักสูตร (X_5), คะแนน Admission รอบ 3 ต่ำสุด ในแต่ละหลักสูตร(X_6), จำนวนนักศึกษาที่เข้ามารอบ 3 ในแต่ละหลักสูตร (X_7), คะแนน Admission รอบ 4 สูงสุด ในแต่ละหลักสูตร (X_8), คะแนน Admission รอบ 4 ต่ำสุด ในแต่ละหลักสูตร (X_9) และจำนวนนักศึกษาที่เข้ามารอบ 4 ในแต่ละหลักสูตร (X_{10})

2. สร้างตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต, ฟังก์ชันเชื่อมโยงโพรบิต และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก

3. คัดเลือกตัวแปรโดยใช้วิธีการคัดเลือกตัวแปรอธิบายด้วยวิธีการถดถอยทีละขั้น (Stepwise Selection) ของตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Model)

4. คำนวณค่า AIC และ BIC ในแต่ละตัวแบบทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต, ฟังก์ชันเชื่อมโยงโพรบิต และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก

5. คัดเลือกตัวแบบทวินามโดยใช้ค่า AIC และ BIC ต่ำสุด

6. ประมาณค่าอัตราการพ้นสภาพนักศึกษาโดยใช้ข้อมูลปีรับเข้าศึกษา 2563

7. สรุปผล

3.2.2 ขั้นตอนการศึกษาเชิงจำลอง

1. กำหนดขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 20, 50, 200, 750

2. กำหนดค่าพารามิเตอร์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบายกับตัวแปรตอบสนองของตัวแบบทวินาม โดยค่าพารามิเตอร์เหล่านี้จะถูกกำหนดให้ใกล้เคียงกับ กรณีศึกษาอัตราการพ้นสภาพการเป็นนักศึกษา มหาวิทยาลัยธรรมศาสตร์ ปีเข้ารับการศึกษา 2561 โดยแบ่งการกำหนดค่าพารามิเตอร์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบายกับตัวแปรตอบสนองของตัวแบบทวินามเป็น 3 กรณี ดังนี้

กรณีที่ 1 เมื่อตัวแบบการถดถอยทวินาม ตัวแปรตอบสนอง Y_i จะอยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_i|X_i)$ เมื่อ π_i จะสัมพันธ์กับตัวแปรอธิบาย X_i ผ่านฟังก์ชันเชื่อมโยงลอจิต (Logit-Link) กำหนดค่าพารามิเตอร์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบายกับตัวแปรตอบสนองของตัวแบบทวินาม ดังนี้

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (0.133000, -0.040800, 0.000130, -0.021500)$$

กรณีที่ 2 เมื่อตัวแบบการถดถอยทวินาม ตัวแปรตอบสนอง Y_i จะอยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_i|X_i)$ เมื่อ π_i จะสัมพันธ์กับตัวแปรอธิบาย X_i ผ่านฟังก์ชันเชื่อมโยงโพรบิต (Probit-link) กำหนดค่าพารามิเตอร์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบายกับตัวแปรตอบสนองของตัวแบบทวินาม ดังนี้

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (0.069100, -0.024500, 0.000078, -0.012900)$$

กรณีที่ 3 เมื่อตัวแบบการถดถอยทวินาม ตัวแปรตอบสนอง Y_i จะอยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_i|X_i)$ เมื่อ π_i จะสัมพันธ์กับตัวแปรอธิบาย X_i ผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารี ล็อก-ล็อก (Complementary Log-Log Link) กำหนดค่าพารามิเตอร์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบายกับตัวแปรตอบสนองของตัวแบบทวินาม ดังนี้

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (-0.246524, -0.033570, 0.000107, -0.017955)$$

3. สร้างตัวแปรอธิบายจำนวนจำนวน 3 ตัวแปร โดยจำลองจากการแจกแจงเอกรูป (Uniform distribution) และการแจกแจงปัวซง (Poisson distribution) โดยค่าพารามิเตอร์เหล่านี้จะถูกกำหนดให้ใกล้เคียงจากข้อมูลจริง ตามลำดับดังนี้

โดยที่ $X_1 \sim Uni(30.33, 75.87)$ คือ คะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร ปี 2561 ที่มีค่าต่ำสุดเท่ากับ 30.33 และค่าสูงสุดเท่ากับ 75.87

โดยที่ $X_2 \sim Uni(12127.1,\ 17239.5)$ คือ คะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร ปี 2561 ที่มีค่าต่ำสุดเท่ากับ 12127.1 และค่าสูงสุดเท่ากับ 17239.5

โดยที่ $X_3 \sim Poisson(50)$ คือ จำนวนนักศึกษาที่เข้ามารอบ 3 ในแต่ละหลักสูตร ปี 2561 ที่มีค่าเฉลี่ยเท่ากับ 50

4. กำหนดค่า $n_i \sim Poisson\ (54)$ โดยที่ในงานวิจัยนี้จะกำหนดค่า λ คือ ค่าเฉลี่ยของจำนวนนักศึกษาที่รับเข้ามาในแต่ละหลักสูตร โดยมีค่าเท่ากับ $\lambda = 54$

5. สร้างตัวแปรตอบสนอง Y มีการแจกแจงทวินาม

5.1 ตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงลอจิต

ตัวแบบการถดถอยทวินาม ตัวแปรตาม Y อยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_i|X_i)$ เมื่อ π_i จะสัมพันธ์กับตัวแปรอิสระ X ผ่านฟังก์ชันเชื่อมโยงลอจิต ดังนี้

$$\log it(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$$

เมื่อ

X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1) \times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1) \times 1$

p คือ จำนวนตัวแปรอธิบาย กำหนดเท่ากับ 3 ตัว

กล่าวได้ว่า π_i ยังคงอยู่ในช่วง [0,1] มีรูปแบบดังนี้

$$\pi_i = \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}$$

จะได้ฟังก์ชันมวลความน่าจะเป็นแบบมีเงื่อนไข Y_i เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{Y_i}\left(\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)^{Y_i}\left(1-\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)^{n_i-Y_i}$$

5.2 ตัวแบบการถดถอยผ่านฟังก์ชันโพรบิต

ตัวแบบการถดถอยทวินาม ตัวแปรตอบสนอง Y อยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_i|X_i)$ เมื่อ π_i จะสัมพันธ์กับตัวแปรอธิบาย X ผ่านฟังก์ชันเชื่อมโยงโพรบิต ดังนี้

$$Probit(\pi_i) = \phi^{-1}(\pi_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$$

เมื่อ

X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1) \times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1) \times 1$

p คือ จำนวนตัวแปรอธิบาย กำหนดเท่ากับ 3 ตัว

กล่าวได้ว่า π_i ยังคงอยู่ในช่วง [0,1] มีรูปแบบดังนี้

$$\pi_i \;=\; \phi(X_i'\beta)$$

จะได้ฟังก์ชันมวลความน่าจะเป็นแบบมีเงื่อนไข y_{i} เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{Y_i}(\phi(X_i'\beta))^{Y_i}(1-\phi(X_i'\beta))^{n_i-Y_i}$$

5.3 ตัวแบบการถดถอยผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก

ตัวแบบการถดถอยทวินาม ตัวแปรตาม Y อยู่ในรูปการแจกแจงแบบมีเงื่อนไขทวินาม $f(Y_{\mathrm{i}}|X_i)$ เมื่อ π_i จะสัมพันธ์กับตัวแปรอิสระ X ผ่านฟังก์ชันเชื่อมคอมพลีเมนทารีล็อก-ล็อก ดังนี้

$$\log(-log(1-\pi_i)) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$$

เมื่อ

X_i คือ เวกเตอร์ของตัวแปรอธิบายที่ค่าสังเกตที่ i ขนาด $(p+1)\times 1$

Y_i คือ ตัวแปรตอบสนองที่ค่าสังเกตที่ i

β คือ เวกเตอร์สัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

p คือ จำนวนตัวแปรอธิบาย กำหนดเท่ากับ 3 ตัว

กล่าวได้ว่า π_i ยังคงอยู่ในช่วง [0,1] มีรูปแบบดังนี้

$$\pi_i \;= 1-\exp(-\exp(X_i'\beta))$$

จะได้ฟังก์ชันมวลความน่าจะเป็นแบบมีเงื่อนไข Y_i เมื่อกำหนด X_i มีรูปแบบดังนี้

$$f(Y_i|X_i) = \binom{n_i}{Y_i}(1-\exp(-\exp(X_i'\beta)))^{Y_i}(\exp(-\exp(X_i'\beta)))^{n_i-Y_i}$$

6. ประมาณค่าพารามิเตอร์ของตัวแบบด้วยวิธีภาวะน่าจะเป็นสูงสุด

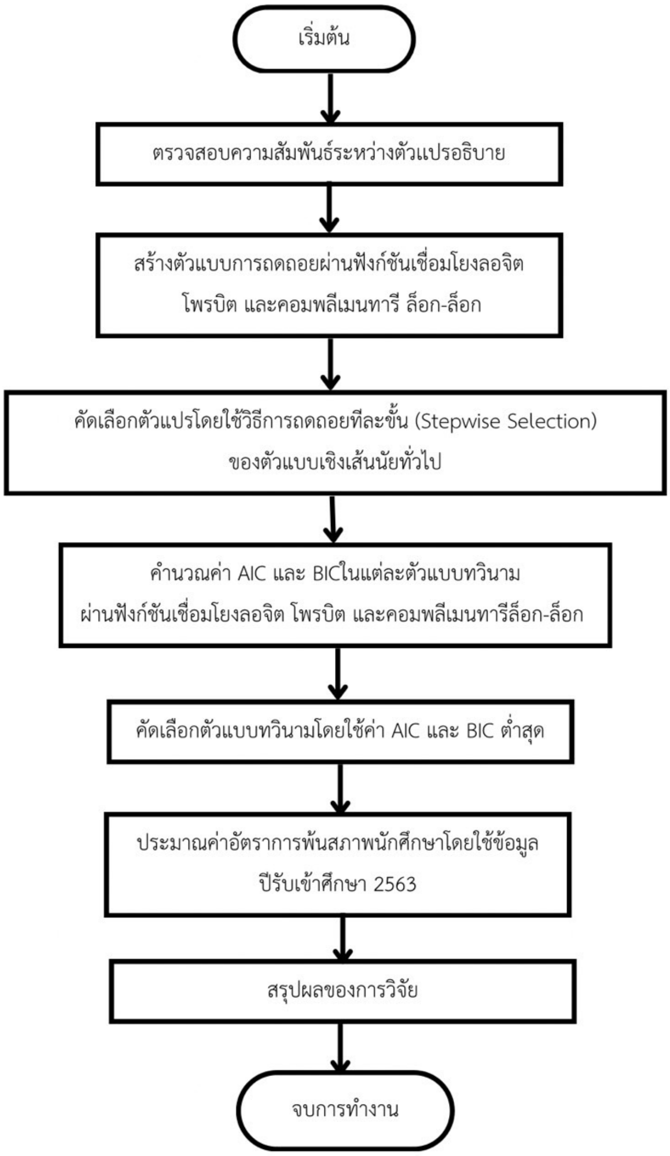
ในที่นี้การประมาณค่าพามิเตอร์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบายกับตัวแปรตอบสนองผ่านฟังก์ชันเชื่อมโยงจะใช้วิธีภาวะน่าจะเป็นวิธีภาวะน่าจะเป็นสูงสุดในทุกกรณี ดังมีรายละเอียดของการประมาณค่า ในบทที่ 2 หัวข้อการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดสำหรับการถดถอยวินาม

7. เก็บรวบรวมและคำนวณค่าที่ใช้เป็นเกณฑ์ในการเปรียบเทียบมี 2 เกณฑ์ คือ ค่าเฉลี่ย AIC และ BIC โดยเฉลี่ย 1000 รอบ และร้อยละจำนวนครั้งที่ AIC และ BIC ต่ำสุด ในแต่ละรอบ จำนวน 1000 รอบ

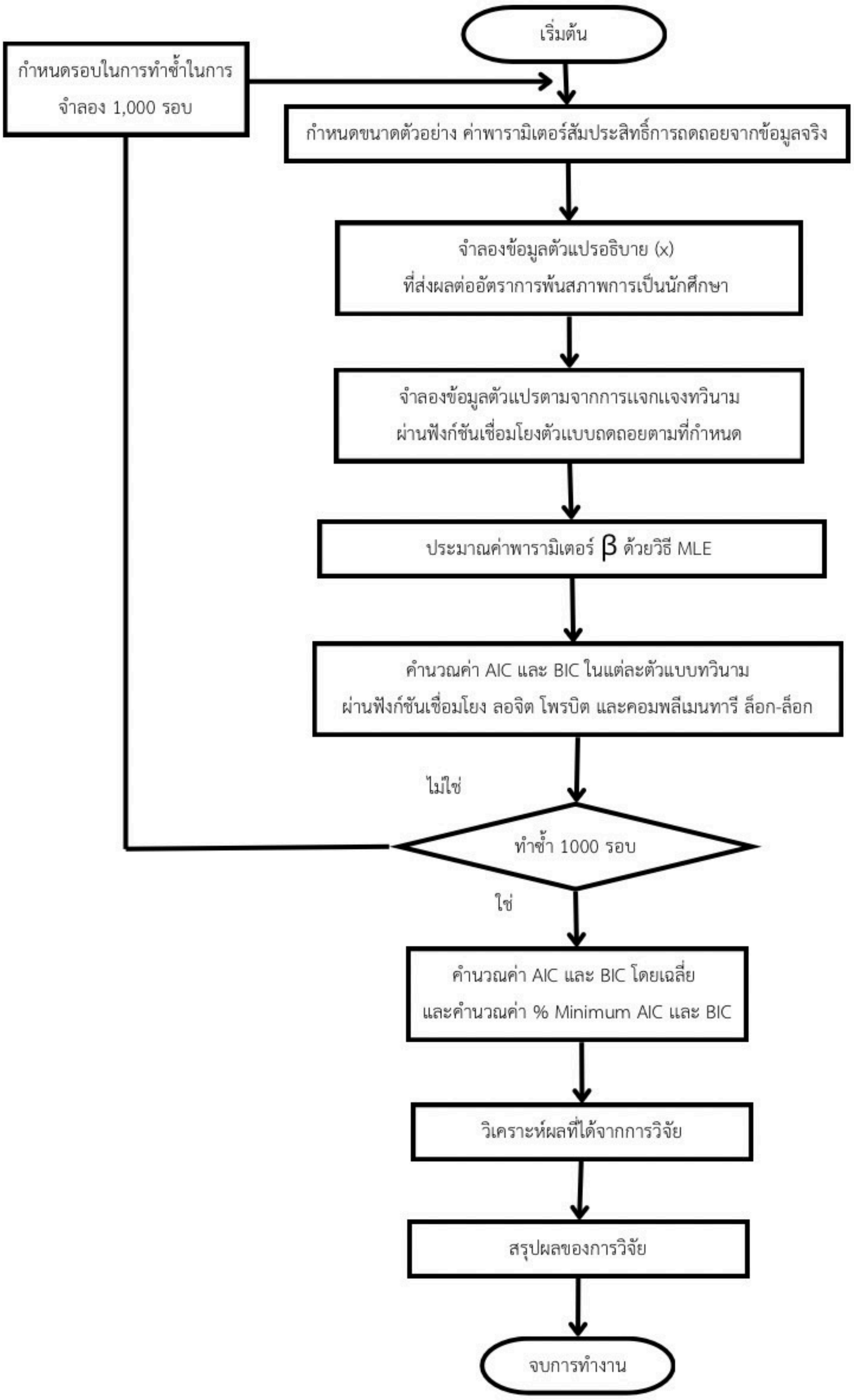
8. วิเคราะห์และสรุปผลการวิจัยจากการจำลอง

3.3 แผนภาพแสดงขั้นตอนการดำเนินงานวิจัย

แผนภาพที่ 3.3.1 ขั้นตอนการดำเนินงานวิจัยในการประยุกต์ใช้กับข้อมูลจริง



แผนภาพที่ 3.3.2 ขั้นตอนการดำเนินงานวิจัยในการศึกษาจำลองข้อมูล



บทที่ 4
ผลการวิจัยและอภิปรายผล

งานวิจัยนี้สนใจศึกษาข้อมูลอัตราการพ้นสภาพการเป็นนักศึกษาและปัจจัยที่จะส่งผลต่ออัตราการพ้นสภาพโดยรวบรวมจากนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีการศึกษา 2561 โดยใช้การวิเคราะห์การถดถอยทวินาม (Binomial Regression Analysis) ภายใต้ฟังก์ชันการเชื่อมโยงที่แตกต่างกันสามฟังก์ชัน ได้แก่ ฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function), ฟังก์ชันเชื่อมโยงโพรบิต (Probit Link Function) และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก (Complementary Log-Log Link Function) และเปรียบเทียบประสิทธิภาพฟังก์ชันเชื่อมโยงทั้งสามโดยใช้เกณฑ์สารสนเทศของอะกะอิเกะ (AIC) และ เกณฑ์สารสนเทศของเบส์(BIC) นอกจากนี้ผู้วิจัยได้ศึกษาเชิงจำลอง เนื่องจากข้อมูลคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ประกอบด้วย 20 หลักสูตร ซึ่งเป็นหน่วยตัวอย่างที่มีขนาดเล็กซึ่งอาจจะส่งผลต่อการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยทวินาม ดังนั้นเพื่อให้เห็นถึงประสิทธิภาพของการประมาณในสถานการณ์ต่างๆที่มีขนาดตัวอย่างที่แตกต่างกันและฟังก์ชันเชื่อมโยงที่แตกต่างกัน ซึ่งเมื่อเพิ่มขนาดตัวอย่างมากขึ้นอาจจะส่งผลให้ประสิทธิภาพของตัวประมาณมีความแม่นยำ สำหรับในบทนี้ผู้วิจัยจะนำเสนอผลการวิจัย โดยแยกเป็น 2 ส่วน คือ ส่วนที่ 1 ผลการวิจัยจากการประยุกต์ใช้ข้อมูลจริง และส่วนที่ 2 ผลการวิจัยจากการศึกษาเชิงจำลองของตัวแบบการถดถอยทวินาม ดังรายละเอียดต่อไปนี้

4.1 ผลการวิจัยจากการประยุกต์ใช้ข้อมูลจริง

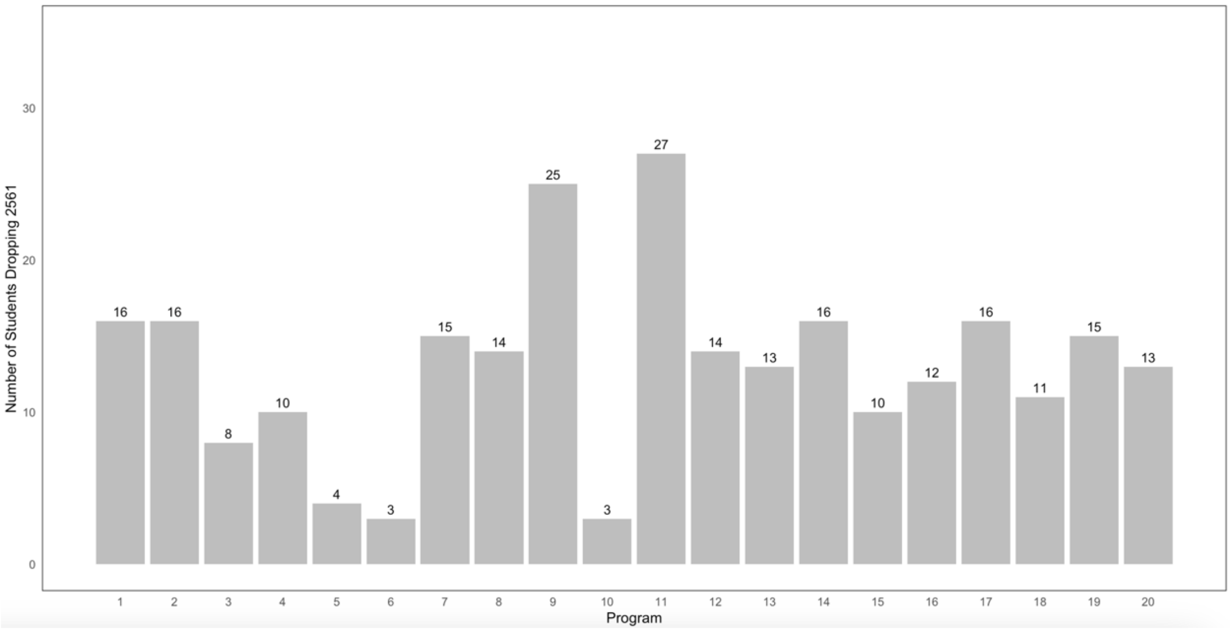
จากข้อมูลจริงจำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร จำนวน 20 หลักสูตร ปีเข้าศึกษา 2561 โดยเก็บข้อมูลระหว่าง ปี 1 ภาคเรียนที่ 1 ถึง ปี 4 ภาคเรียนที่ 1 และจำนวนนักศึกษาที่รับเข้ามา ปีการเข้าศึกษา 2561 ซึ่งข้อมูลตัวแปรตอบสนอง คือจำนวนนักศึกษาที่พ้นสภาพเป็นนักศึกษาในแต่ละหลักสูตร โดยเข้ารับการศึกษาปี 2561 คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ โดยมีหลักสูตรดังนี้

โครงการปกติจำนวน 15 หลักสูตร และโครงการพิเศษจำนวน 5 หลักสูตร ได้แก่

1. หลักสูตรสาขาวิชาสถิติ (โครงการปกติ)
2. หลักสูตรสาขาวิชาคณิตศาสตร์ (โครงการปกติ)
3. หลักสูตรสาขาวิชาคณิตศาสตร์ประยุกต์ (โครงการปกติ)
4. หลักสูตรสาขาวิชาสถิติ (โครงการพิเศษ)
5. หลักสูตรสาขาวิชาคณิตศาสตร์ (โครงการพิเศษ)
6. หลักสูตรสาขาวิชาวิทยาการประกันภัย (โครงการพิเศษ)
7. หลักสูตรสาขาวิชาวิทยาศาสตร์สิ่งแวดล้อม (โครงการปกติ)

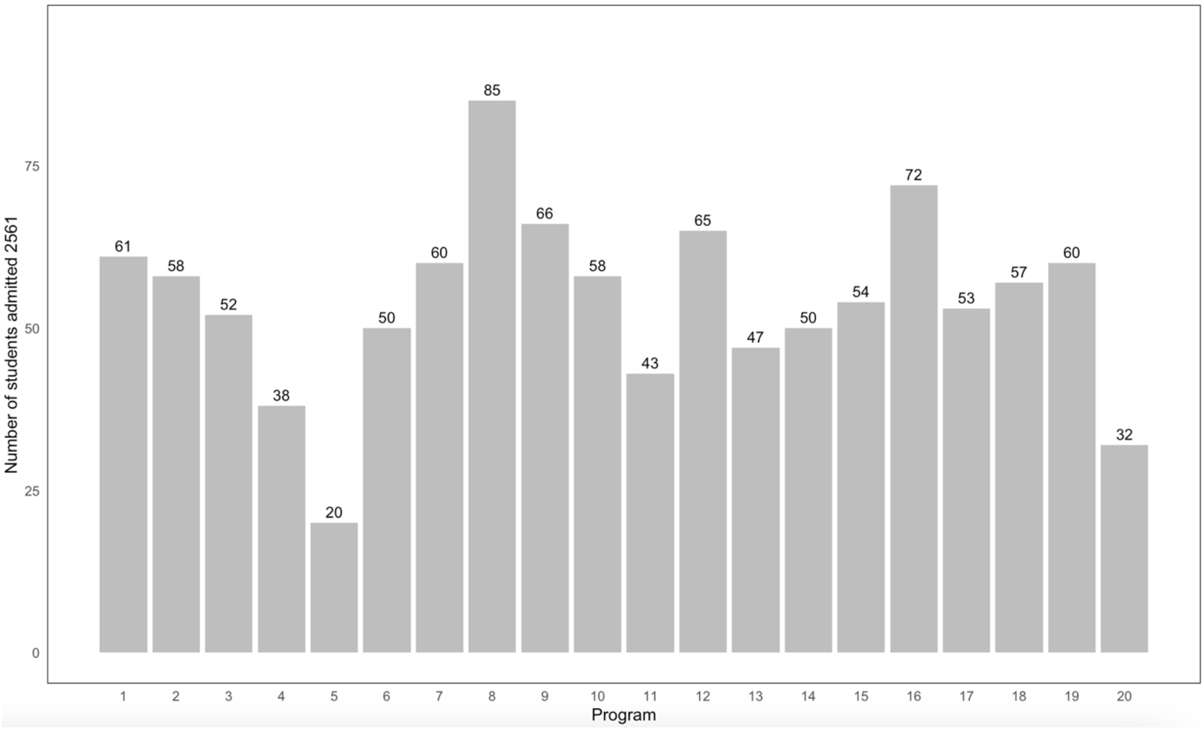
8. หลักสูตรสาขาวิชาเทคโนโลยีการเกษตร (โครงการปกติ)
9. หลักสูตรสาขาวิชาเคมี (โครงการปกติ)
10. หลักสูตรสาขาวิชาวิทยาศาสตร์และเทคโนโลยีการอาหาร (โครงการปกติ)
11. หลักสูตรสาขาวิชาฟิสิกส์อิเล็กทรอนิกส์ (โครงการปกติ)
12. หลักสูตรสาขาวิชาเทคโนโลยีชีวภาพ (โครงการปกติ)
13. หลักสูตรสาขาวิชาฟิสิกส์ (โครงการปกติ)
14. หลักสูตรสาขาวิชาวัสดุศาสตร์ (โครงการปกติ)
15. หลักสูตรสาขาวิชาวิทยาศาสตร์และเทคโนโลยีสิ่งทอ (โครงการปกติ)
16. หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการปกติ)
17. หลักสูตรสาขาวิชาเทคโนโลยีเพื่อการพัฒนายั่งยืน (โครงการปกติ)
18. หลักสูตรสาขาวิชาเทคโนโลยีและนวัตกรรมทางอาหาร (โครงการปกติ)
19. หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการพิเศษ)
20. หลักสูตรสาขาวิชาเทคโนโลยีพลังงานชีวภาพและการแปรรูปเคมีชีวภาพ (โครงการพิเศษ)

แผนภาพที่ 4.1 จำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร ปีเข้าศึกษา 2561 (Y_i)



จากแผนภาพที่ 4.1 ข้อมูลจำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปีเข้าศึกษา 2561 โดยที่หลักสูตรที่มีจำนวนนักศึกษาพ้นสภาพมากที่สุด คือ หลักสูตรสาขาวิชาฟิสิกส์อิเล็กทรอนิกส์ (โครงการปกติ) มีจำนวน 27 คน รองลงมา คือหลักสูตรสาขาวิชาเคมี (โครงการปกติ) มีจำนวน 25 คน

แผนภาพที่ 4.2 จำนวนนักศึกษาที่รับเข้ามา ปีการเข้าศึกษา 2561 ในแต่ละหลักสูตร (n_i)



จากแผนภาพที่ 4.2 ข้อมูลจำนวนนักศึกษาที่รับเข้ามา ปีการศึกษา 2561 ในแต่ละหลักสูตร โดยที่หลักสูตรที่รับเข้ามากที่สุด คือ หลักสูตรสาขาวิชาเทคโนโลยีการเกษตร (โครงการปกติ) มีจำนวน 85 คน รองลงมา คือ หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการปกติ) มีจำนวน 72 คน

4.1.1 ผลการวิเคราะห์ข้อมูลเชิงพรรณนา

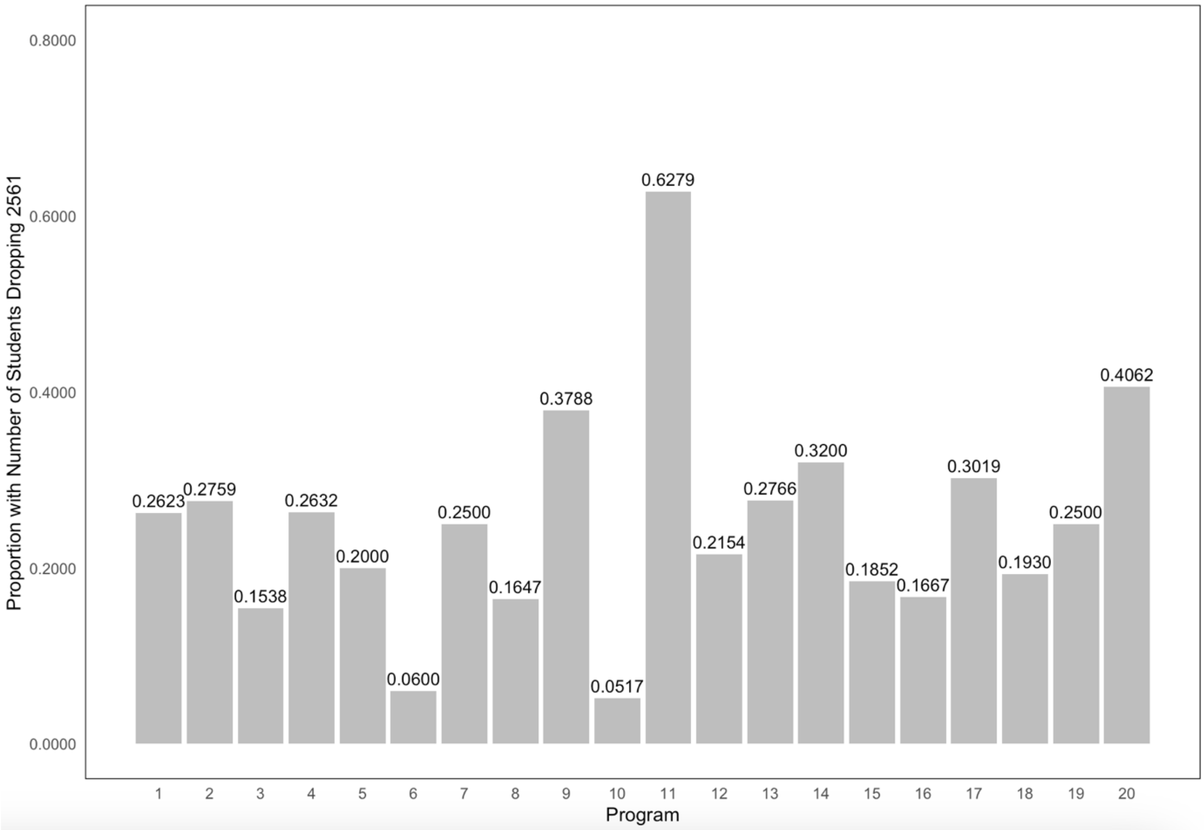
กำหนดให้ π_i แทนอัตราการพ้นสภาพนักศึกษาในแต่ละหลักสูตร จากจำนวนทั้งสิ้น 20 หลักสูตร ปีการเข้าศึกษา 2561 ซึ่งหาได้จากจำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษาส่วนด้วยจำนวนนักศึกษาที่รับเข้ามา (Y_i/n_i) ผลการวิเคราะห์ข้อมูลเชิงพรรณนาของอัตราการพ้นสภาพนักศึกษาแสดงดังตารางที่ 4.1 และแผนภาพที่ 4.3 ได้ดังนั้น

ตารางที่ 4.1 สถิติพรรณนาของอัตราการพ้นสภาพนักศึกษา

ค่าสถิติ	อัตราการพ้นสภาพนักศึกษา
เฉลี่ย	0.2502
ค่าสูงสุด	0.6279
ค่าต่ำสุด	0.0517
ส่วนเบี่ยงเบนมาตรฐาน	0.1263
ความเบ้	1.1503

จากตารางที่ 4.1 ค่าสูงสุดและค่าต่ำสุดเท่ากับ 0.6279 และ 0.0517 ตามลำดับ มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.2501 และ 0.1263 และที่สำคัญพบว่าสัมประสิทธิ์ความเบ้มีค่าเป็นบวก ซึ่งเท่ากับ 1.1503 ชี้ให้เห็นว่าอัตราการพ้นสภาพการเป็นนักศึกษามีการกระจายลักษณะเบ้ขวา

แผนภาพที่ 4.3 อัตราการพ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร ปีเข้ารับการศึกษา 2561



จากแผนภาพที่ 4.3 ข้อมูลอัตราการพ้นสภาพการเป็นนักศึกษา ปีเข้ารับศึกษา 2561 โดยที่หลักสูตรที่อัตราการพ้นสภาพมากที่สุด คือหลักสูตรสาขาวิชาเคมี (โครงการปกติ) รองลงมาคือ หลักสูตรสาขาวิชาเทคโนโลยีพลังงานชีวภาพและการแปรรูปเคมีชีวภาพ (โครงการพิเศษ)

ตารางที่ 4.2 ชื่อของตัวแปรอธิบายที่ใช้ในการศึกษา

ตัวแปร	ชื่อของตัวแปร
X_1	หลักสูตรที่มีวิชาบังคับ C (Compulsory subjects C) $X_1 = 1$ คือ หลักสูตรที่มีวิชาบังคับ C, $X_1 = 0$ คือ หลักสูตรที่ไม่มีวิชาบังคับ C
X_2	จำนวนหน่วยกิต แต่ละหลักสูตร (Number of credits)
X_3	ค่าเทอมในแต่ละหลักสูตร (Tuition fees)
X_4	ประเภทหลักสูตร (Program) $X_4 = 1$ คือ หลักสูตรปกติ C, $X_4 = 0$ คือ หลักสูตรพิเศษ
X_5	คะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร (Highest Admission Scores in the 3rd round)
X_6	คะแนน Admission (รอบ 3) ต่ำสุด แต่ละหลักสูตร (Lowest Admission Scores in the 3rd round)
X_7	จำนวนนักศึกษาที่เข้ามารอบ 3 ในแต่ละหลักสูตร (Number of students accepted in the 3rd round)
X_8	คะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร (Highest Admission Scores in the 4th round)
X_9	คะแนน Admission (รอบ 4) ต่ำสุด แต่ละหลักสูตร (Lowest Admission Scores Round in the 4th round)
X_{10}	จำนวนนักศึกษาที่เข้ามารอบ 4 ในแต่ละหลักสูตร (Number of students accepted into the 4th round)

โดยตัวแปรอธิบายดังข้างต้นมีเมทริกซ์สหสัมพันธ์แสดงดังตารางที่ 4.3

4.1.2 ตรวจสอบความสัมพันธ์ระหว่างตัวแปรอธิบาย

การตรวจสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบาย โดยใช้ The Point-Biserial Correlation coefficient [1] ซึ่งเป็นกรณีพิเศษของ The Pearson Correlation coefficient สำหรับการวิเคราะห์ตัวแปรอธิบายเชิงกลุ่มกับตัวแปรอธิบายเชิงปริมาณ และใช้ The Pearson Correlation coefficient สำหรับการวิเคราะห์ตัวแปรอธิบายอื่น ๆ ในการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอธิบายทั้ง 10 ตัวแปร ได้ผลดังต่อไปนี้

ตารางที่ 4.3 เมทริกซ์สหสัมพันธ์ระหว่างตัวแปรอธิบายทั้ง 10 ตัวแปร

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	1									
X_2	-0.4671	1								
X_3	0.3880	-0.3453	1							
X_4	-0.4714	0.3078	-0.9596	1						
X_5	0.3086	-0.1584	0.0551	-0.034	1					
X_6	-0.1413	-0.3138	-0.4266	0.5182	0.1027	1				
X_7	-0.1065	0.1144	-0.4320	0.4988	0.1428	0.5173	1			
X_8	-0.1926	0.3057	-0.5029	0.6013	0.3467	0.3964	0.3000	1		
X_9	-0.0935	-0.2156	-0.1944	0.2857	0.6435	0.6434	0.3333	0.3994	1	
X_{10}	-0.1188	-0.0099	-0.0741	0.2652	-0.1612	0.2901	0.0317	0.3816	-0.0882	1

จากตารางที่ 4.3 เมทริกซ์สหสัมพันธ์ระหว่างตัวแปรอธิบายทั้ง 10 ตัวแปรและจากสถิติทดสอบสหสัมพันธ์ พบว่าตัวแปรอธิบาย X_1 กับ X_2 , X_1 กับ X_4 , X_3 กับ X_4 , X_3 กับ X_8 , X_4 กับ X_6 , X_4 กับ X_7, X_4 กับ X_8 , X_5 กับ X_9 , X_6 กับ X_7 และ X_6 กับ X_9 มีความสัมพันธ์กัน ส่วนตัวแปรอธิบายคู่อื่นนั้นไม่มีความสัมพันธ์กัน ที่ระดับนัยสำคัญ 0.05 แต่เมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ที่ค่าสูงสุดคือ -0.9596 ซึ่งมาจากระหว่างตัวแปรอธิบาย X_3 กับ X_4 มีความสัมพันธ์กันสูง ซึ่งเมื่อมานำมาวิเคราะห์หาปัจจัยพบว่า X_3 กับ X_4 ไม่มีผลต่ออัตราการพ้นสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

[1] สัมประสิทธิสหาสัมพันธ์แบบพอยท์ไบซีเรียล (The Point-Biserial Correlation coefficient) ใช้สัญญาลักษณ์ r_{pb} เป็นวิธีที่ใช้วัดความสัมพันธ์ระหว่างตัวแปร หรือข้อมูล 2 ชุด โดยที่ตัวแปรหนึ่งเป็นตัวแปรต่อเนื่อง อีกตัวหนึ่งมี 2 ลักษณะจริง (true dichotomous)

4.1.3 การสร้างตัวแบบทำนายการถดถอยทวินาม

สำหรับการสร้างตัวแบบทำนายการถดถอยทวินามภายใต้ฟังก์ชันการเชื่อมโยงที่แตกต่างกันสามฟังก์ชัน ได้แก่ ฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function), ฟังก์ชันเชื่อมโยงโพรบิต (Probit Link Function) และ ฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก (Complementary Log-Log Link) โดยใช้การคัดเลือกตัวแปรแบบ Stepwise Selection ผลลัพธ์การวิเคราะห์ของตัวแบบการถดถอยภายใต้ฟังก์ชันการเชื่อมโยงที่แตกต่างกันสามฟังก์ชัน แสดงดังนี้

ตารางที่ 4.4 ค่าสัมประสิทธิ์การถดถอยสำหรับตัวแบบทำนายอัตราการพ้นสภาพการเป็นนักศึกษาได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต (Logit Link Function) โดยวิธีการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยทีละขั้น (Stepwise Selection)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.1328000	0.9490000	0.1400000	0.8887760
X_5	-0.0408400	0.0083890	-4.8680000	0.0000010 ***
X_7	-0.0215500	0.0060960	-3.5300000	0.0004080 ***
X_8	0.0001298	0.0000677	1.9150000	0.0554840 .

หมายเหตุ : Signif. ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

จากตารางที่ 4.4 เขียนตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต ในรูปสมการเชิงเส้นได้สามารถเขียนได้ดังนี้

$$\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = 0.1328000 - 0.0408400X_{i5} - 0.0215500X_{i7} + 0.0001298X_{i8}$$

หรือตัวแบบทำนายอัตราการพ้นสภาพนักศึกษาได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิตได้ดังนี้

$$\hat{\pi}_i = \frac{\exp(0.1328000 - 0.0408400X_{i5} - 0.0215500X_{i7} + 0.0001298X_{i8})}{1+\exp(0.1328000 - 0.0408400X_{i5} - 0.0215500X_{i7} + 0.0001298X_{i8})}$$

เมื่อ $\hat{\pi}_i$ แทน ค่าประมาณอัตราการพ้นสภาพการเป็นนักศึกษา แต่ละหลักสูตร

X_{i5} แทน คะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร

X_{i7} แทน จำนวนนักศึกษาที่เข้ามารอบ 3 แต่ละหลักสูตร

X_{i8} แทน คะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร

ตารางที่ 4.5 ค่าสัมประสิทธิ์การถดถอยสำหรับตัวแบบทำนายอัตราการพ้นสภาพนักศึกษาได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต (Probit Link Function) โดยวิธีการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยทีละขั้น (Stepwise Selection)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.0690800	0.5545000	0.1250000	0.9008670
X_5	-0.0245400	0.0049190	-4.9890000	0.0000006 ***
X_7	-0.0129100	0.0035900	-3.5960000	0.0003240 ***
X_8	0.0000780	0.0000391	1.9930000	0.0462230 *

หมายเหตุ : Signif. ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

จากตารางที่ 4.5 ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต ในรูปสมการเชิงเส้นสามารถเขียนได้ดังนี้

$$\phi^{-1}(\hat{\pi}_i) = 0.0690800 - 0.0245400X_{i5} - 0.0129100X_{i7} + 0.0000780X_{i8}$$

หรือตัวแบบทำนายอัตราการพ้นสภาพนักศึกษาได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิตได้ดังนี้

$$\hat{\pi}_i = \phi(0.0690800 - 0.0245400X_{i5} - 0.0129100X_{i7} + 0.0000780X_{i8})$$

เมื่อ $\hat{\pi}_i$ แทน ค่าประมาณอัตราการพ้นสภาพการเป็นนักศึกษา แต่ละหลักสูตร

X_{i5} แทน คะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร

X_{i7} แทน จำนวนนักศึกษาที่เข้ามารอบ 3 แต่ละหลักสูตร

X_{i8} แทน คะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร

ϕ แทน ฟังก์ชันสะสมของการแจกแจงปรกติมาตรฐาน

ϕ^{-1} แทน ฟังก์ชันผกผันสะสมของการแจกแจงปรกติมาตรฐาน

ตารางที่ 4.6 ค่าสัมประสิทธิ์การถดถอยสำหรับตัวแบบทำนายอัตราการพ้นสภาพนักศึกษาได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่องโยงคอมพลีเมนทารีล็อก – ล็อก (Complementary Log-Log Link) โดยวิธีการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยทีละขั้น (Stepwise Selection)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.2465240	0.823502	-0.300000	0.7646665
X_5	-0.0335700	0.006526	-5.140000	0.0000002 ***
X_7	-0.0179550	0.005180	-3.470000	0.0005290 ***
X_8	0.0001074	0.059008	1.820000	0.0687333 .

หมายเหตุ : Signif. ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

จากตารางที่ 4.6 ตัวแบบถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ในรูปแบบเชิงเส้นสามารถเขียนได้ดังนี้

$$\log(-log(1-\hat{\pi}_i)) = -0.2465240 - 0.0335700X_{i5} - 0.0179550X_{i7} + 0.0001074X_{i8}$$

หรือเขียนตัวแบบทำนายอัตราการพ้นสภาพนักศึกษาได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อกได้ดังนี้

$$\hat{\pi}_i = 1 - \exp(-exp(-0.2465240 - 0.0335700X_{i5} - 0.0179550X_{i7} + 0.0001074X_{i8}))$$

เมื่อ $\hat{\pi}_i$ แทน ค่าประมาณอัตราการพ้นสภาพการเป็นนักศึกษา แต่ละหลักสูตร

X_{i5} แทน คะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร

X_{i7} แทน จำนวนนักศึกษาที่เข้ามารอบ 3 แต่ละหลักสูตร

X_{i8} แทน คะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร

ตารางที่ 4.7 เปรียบเทียบประสิทธิภาพตัวแบบทำนายอัตราการพ้นสภาพการเป็นนักศึกษาได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงทั้งสามภายใต้ฟังก์ชันการเชื่อมโยงที่แตกต่างกันสามฟังก์ชัน ได้แก่ ลอจิต (Logit Link Function), โพรบิต (Probit Link Function) และคอมพลีเมนทารีล็อก – ล็อก (Complementary Log-Log Link Function) โดยใช้การคัดเลือกตัวแปรแบบ Stepwise Selection เกณฑ์สารสนเทศของอะกะอิเกะ (AIC) และ เกณฑ์สารสนเทศของเบส์ (BIC) แสดงดังนี้

ตัวแบบ	**AIC**	**BIC**
Logit	129.9280	133.9109
Probit	130.0159	133.9988
Cloglog	**129.4223**	**133.4053**

หมายเหตุ : ตัวหนา หมายถึง ค่า AIC และ BIC ของตัวแบบที่ให้ค่าต่ำที่สุด

จากตารางที่ 4.7 พบว่าตัวแบบทำนายอัตราการพ้นสภาพนักศึกษาได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงของคอมพลีเมนทารีล็อก – ล็อก ให้ค่า AIC และ BIC ต่ำที่สุด ดังนั้นตัวแบบทำนายอัตราการพ้นสภาพผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก จะมีประสิทธิภาพดีกว่าตัวแบบทำนายอัตราการพ้นสภาพการเป็นนักศึกษาที่ได้จากการวิเคราะห์การถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต และโพรบิต

4.1.4 การประมาณค่าอัตราการพ้นสภาพนักศึกษาโดยใช้ข้อมูลปีรับเข้าศึกษา 2563

ทั้งนี้ในการรับนักศึกษาเข้าศึกษาต่อของคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ถ้ายังใช้เกณฑ์เดิมในการรับเข้าในปีถัดถัดไป คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ สามารถทำนายอัตราการพ้นสภาพการเป็นนักศึกษาในแต่ละหลักสูตร หรือจำนวนนักศึกษาที่พ้นสภาพในแต่ละหลักสูตร เพื่อให้เป็นตัวอย่างผู้วิจัยได้นำข้อมูลการรับนักศึกษาเข้าศึกษาต่อ ปี 2563 คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ซึ่งยังใช้เกณฑ์เดิมในการรับนักศึกษาเข้าศึกษาต่อ เพื่อให้ดูว่าตัวแบบดังกล่าวมีประสิทธิภาพผู้วิจัยได้นำข้อมูลปีรับเข้า 2561 มาทำนายอัตราการพ้นสภาพจะตรงกับข้อมูลจริงหรือไม่ และเพื่อเป็นตัวอย่างที่จะใช้ในปีอื่น ๆ ผู้วิจัยได้นำข้อมูลปี 2563 มาทำนายอัตราการพ้นสภาพนักศึกษา โดยใช้ตัวแบบการทำนายทวินาม ผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ซึ่งเป็นตัวแบบที่ให้ แสดงดังตารางต่อไปนี้

ตารางที่ 4.8 ทำนายอัตราการพ้นสภาพนักศึกษา โดยใช้ตัวแบบการทำนายทวินาม ผ่านฟังก์ชันการเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก โดยใช้ข้อมูลปีเข้ารับการศึกษา 2563

ชื่อหลักสูตร	คะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร	จำนวนนักศึกษาที่เข้ามา รอบ 3 แต่ละหลักสูตร	คะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร	ค่าประมาณอัตราการพ้นสภาพการเป็นนักศึกษา
หลักสูตรสาขาวิชาสถิติ (โครงการปกติ)	58.08	50	18478.5	0.2810
หลักสูตรสาขาวิชาคณิตศาสตร์ (โครงการปกติ)	51.93	35	17747.5	0.3878
หลักสูตรสาขาวิชาคณิตศาสตร์ประยุกต์ (โครงการปกติ)	55.21	50	15856.5	0.2397
หลักสูตรสาขาวิชาสถิติ (โครงการพิเศษ)	41.55	50	14514.5	0.3129
หลักสูตรสาขาวิชาคณิตศาสตร์ (โครงการพิเศษ)	40.07	40	15895.5	0.4216
หลักสูตรสาขาวิชาวิทยาการประกันภัย (โครงการพิเศษ)	65.95	40	18259.2	0.2562
หลักสูตรสาขาวิชาวิทยาศาสตร์สิ่งแวดล้อม (โครงการปกติ)	45.33	40	15789.5	0.3647
หลักสูตรสาขาวิชาเทคโนโลยีการเกษตร (โครงการปกติ)	39.85	45	15276	0.3760
หลักสูตรสาขาวิชาเคมี (โครงการปกติ)	53.6	35	18051	0.3808
หลักสูตรสาขาวิชาวิทยาศาสตร์และเทคโนโลยีการอาหาร (โครงการปกติ)	50.82	40	17707	0.3710
หลักสูตรสาขาวิชาฟิสิกส์อิเล็กทรอนิกส์ (โครงการปกติ)	42.07	25	16805	0.5223
หลักสูตรสาขาวิชาเทคโนโลยีชีวภาพ (โครงการปกติ)	53.73	40	17000.3	0.3227
หลักสูตรสาขาวิชาฟิสิกส์ (โครงการปกติ)	45.53	25	16312.8	0.4642
หลักสูตรสาขาวิชาวัสดุศาสตร์ (โครงการปกติ)	42.92	25	16045.4	0.4841
หลักสูตรสาขาวิชาวิทยาศาสตร์และเทคโนโลยีสิ่งทอ (โครงการปกติ)	49.88	25	15487	0.3895
หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการปกติ)	58.83	40	19288.5	0.3429
หลักสูตรสาขาวิชาเทคโนโลยีเพื่อการพัฒนายั่งยืน (โครงการปกติ)	44.25	25	15150.5	0.4372
หลักสูตรสาขาวิชาวิทยาศาสตร์และนวัตกรรมทางอาหาร (โครงการปกติ)	52.55	25	16894.4	0.4083
หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการพิเศษ)	65.05	40	16828.5	0.2302
หลักสูตรสาขาวิชาเทคโนโลยีพลังงานชีวภาพและการแปรรูปเคมีชีวภาพ (โครงการพิเศษ)	45.52	20	16727	0.5102

ตารางที่ 4.9 ทำนายจำนวนนักศึกษาที่พ้นสภาพ โดยใช้ตัวแบบการทำนายทวินาม ผ่านฟังก์ชันการเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก โดยใช้ข้อมูลปีเข้ารับการศึกษา 2563

ชื่อหลักสูตร	จำนวนนักศึกษาที่รับเข้ามา	ค่าประมาณอัตราการพ้นสภาพการเป็นนักศึกษา	ค่าประมาณจำนวนนักศึกษาที่พ้นสภาพ
หลักสูตรสาขาวิชาสถิติ (โครงการปกติ)	92	0.2810	26
หลักสูตรสาขาวิชาคณิตศาสตร์ (โครงการปกติ)	44	0.3878	17
หลักสูตรสาขาวิชาคณิตศาสตร์ประยุกต์ (โครงการปกติ)	42	0.2397	10
หลักสูตรสาขาวิชาสถิติ (โครงการพิเศษ)	55	0.3129	17
หลักสูตรสาขาวิชาคณิตศาสตร์ (โครงการพิเศษ)	43	0.4216	18
หลักสูตรสาขาวิชาวิทยาการประกันภัย (โครงการพิเศษ)	48	0.2562	12
หลักสูตรสาขาวิชาวิทยาศาสตร์สิ่งแวดล้อม (โครงการปกติ)	59	0.3647	22
หลักสูตรสาขาวิชาเทคโนโลยีการเกษตร (โครงการปกติ)	74	0.3760	28
หลักสูตรสาขาวิชาเคมี (โครงการปกติ)	68	0.3808	26
หลักสูตรสาขาวิชาวิทยาศาสตร์และเทคโนโลยีการอาหาร (โครงการปกติ)	61	0.3710	23
หลักสูตรสาขาวิชาฟิสิกส์อิเล็กทรอนิกส์ (โครงการปกติ)	44	0.5223	23
หลักสูตรสาขาวิชาเทคโนโลยีชีวภาพ (โครงการปกติ)	64	0.3227	21
หลักสูตรสาขาวิชาฟิสิกส์ (โครงการปกติ)	39	0.4642	18
หลักสูตรสาขาวิชาวัสดุศาสตร์ (โครงการปกติ)	42	0.4841	20
หลักสูตรสาขาวิชาวิทยาศาสตร์และเทคโนโลยีสิ่งทอ (โครงการปกติ)	48	0.3895	19
หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการปกติ)	144	0.3429	49
หลักสูตรสาขาวิชาเทคโนโลยีเพื่อการพัฒนายั่งยืน (โครงการปกติ)	45	0.4372	20
หลักสูตรสาขาวิชาวิทยาศาสตร์และนวัตกรรมทางอาหาร (โครงการปกติ)	39	0.4083	16
หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการพิเศษ)	96	0.2302	22
หลักสูตรสาขาวิชาเทคโนโลยีพลังงานชีวภาพและการแปรรูปเคมีชีวภาพ (โครงการพิเศษ)	35	0.5102	18

หมายเหตุ : ค่าประมาณอัตราการพ้นสภาพการเป็นนักศึกษามาจากตารางที่ 4.8 และค่าประมาณจำนวนนักศึกษาที่พ้นสภาพมาจากผลคูณของจำนวนนักศึกษาที่รับเข้ามากับค่าประมาณอัตราการ

จากตารางที่ 4.8 แสดงการทำนายอัตราการพ้นสภาพการเป็นนักศึกษา โดยใช้ตัวแบบการทำนายทวินาม ผ่านฟังก์ชันการเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก โดยใช้ข้อมูล ปีเข้ารับการศึกษา 2563 พบว่าหลักสูตรที่ค่าประมาณอัตราการพ้นสภาพการเป็นนักศึกษามากสุด คือ หลักสูตรสาขาวิชาฟิสิกส์อิเล็กทรอนิกส์ (โครงการปกติ) รองลงมาคือ หลักสูตรสาขาวิชาวัสดุศาสตร์ (โครงการปกติ) และจากตารางที่ 4.9 แสดงการทำนายจำนวนนักศึกษาที่พ้นสภาพ โดยใช้ตัวแบบการทำนายทวินาม ผ่านฟังก์ชันการเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก โดยใช้ข้อมูล ปีเข้ารับการศึกษา 2563 พบว่าหลักสูตรที่ค่าประมาณจำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษามากสุดคือ หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ (โครงการปกติ) รองลงมาคือหลักสูตรสาขาวิชาเทคโนโลยีการเกษตร (โครงการปกติ)

4.2 ผลการวิจัยจากการศึกษาเชิงจำลองของตัวแบบการถดถอยทวินาม

ในหัวข้อนี้จะกล่าวถึงผลการวิจัยจากการจำลองข้อมูลให้มีการแจกแจงทวินามเพื่อศึกษาการเอาชนะข้อจำกัดในขนาดตัวอย่าง โดยมีความสามารถในการเพิ่มขนาดตัวอย่างให้มากกว่าที่มีอยู่ในปัจจุบันในชุดข้อมูลจริง ขนาดตัวอย่างที่ใหญ่ขึ้นเป็นสิ่งสำคัญยิ่งในการได้รับประสิทธิภาพทางสถิติที่เพียงพอด้วยการสร้างชุดข้อมูลจำลองที่มีขนาดตัวอย่างเพิ่มขึ้น และตรวจสอบการทำงานของฟังก์ชันเชื่อมโยงว่ามีความถูกต้องหรือมีประสิทธิภาพมากน้อยแค่ไหน เพื่อทำให้เกิดความมั่นใจมากขึ้นในงานวิจัยของผู้วิจัย ภายใต้สถานการณ์ที่ข้อมูลมีขนาดตัวอย่างที่แตกต่างกัน กระทำซ้ำ 1,000 รอบ เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพฟังก์ชันเชื่อมโยงทั้งสามโดยใช้เกณฑ์สารสนเทศของอะกะอิเกะ (AIC) และ เกณฑ์สารสนเทศของเบส์ (BIC)

ตารางที่ 4.10 เปรียบเทียบตัวแบบจำลองโดยใช้ค่าเฉลี่ยของ AIC และ BIC ที่สอดคล้องกับฟังก์ชันเชื่อมโยงเมื่อขนาดตัวอย่างเท่ากับ 20

ตัวแบบ	จำลองผ่านฟังก์ชันเชื่อมโยง					
	Logit		Probit		Cloglog	
	Average of AIC	Average of BIC	Average of AIC	Average of BIC	Average of AIC	Average of BIC
Logit	**104.3726**	**108.3556**	104.0287	108.0116	104.8454	108.8283
Probit	104.4025	108.3854	**103.9768**	**107.9598**	104.9503	108.9333
Cloglog	104.4582	108.4411	104.1996	108.1826	**104.8106**	**108.7936**

หมายเหตุ : ตัวหนา หมายถึง ค่าเฉลี่ย 1,000 รอบ AIC (Average of AIC) และค่าเฉลี่ย 1000 รอบ BIC (Average of BIC) ของตัวแบบที่ให้ค่าต่ำที่สุด

ตารางที่ 4.11 เปรียบเทียบตัวแบบจำลองโดยใช้เปอร์เซ็นต์ของ AIC และ BIC ต่ำสุด ที่สอดคล้องกับฟังก์ชันเชื่อมโยงเมื่อขนาดตัวอย่างเท่ากับ 20

ตัวแบบ	จำลองผ่านฟังก์ชันเชื่อมโยง					
	Logit		Probit		Cloglog	
	% *Minimum* AIC	% *Minimum* BIC	% *Minimum* AIC	% *Minimum* BIC	% *Minimum* AIC	% *Minimum* BIC
Logit	10.8%	10.8%	11.4%	11.4%	9.4%	9.4%
Probit	**45.6%**	**45.6%**	**52.4%**	**52.4%**	38.4%	38.4%
Cloglog	43.6%	43.6%	36.2%	36.2%	**52.2%**	**52.2%**

หมายเหตุ : ตัวหนา หมายถึง % Minimum AIC และ % Minimum BIC ของตัวแบบที่ให้ค่าสูงที่สุด

จากตารางที่ 4.8 และ 4.9 แสดงการเปรียบเทียบตัวแบบจำลองโดยใช้ค่าเฉลี่ย 1,000 รอบ ของ AIC (Average of AIC) และค่าเฉลี่ย 1,000 รอบ BIC (Average of BIC) ผลลัพธ์ที่แสดงในตารางที่ 4.8 เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงโพรบิต ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด และเมื่อจำลองผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด และเพื่อเป็นการยืนยันผู้วิจัยได้หา % Minimum ของตัวแบบจำลองผ่านฟังก์ชันเชื่อมโยงทั้ง 3 ฟังก์ชัน ที่ให้ค่า AIC และ BIC ต่ำสุด จากกระทำซ้ำ 1,000 รอบ แสดงดังตารางที่ 4.9 พบว่า เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต จะได้ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต ให้ค่า % Minimum มากสุดเท่ากับ 45.6% ซึ่งขัดแย้งกับค่าเฉลี่ย 1,000 รอบ แต่เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงโพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่าสอดคล้องกับ ค่าเฉลี่ย 1,000 รอบ

ตารางที่ 4.12 เปรียบเทียบตัวแบบจำลองโดยใช้ค่าเฉลี่ยของ AIC และ BIC ที่สอดคล้องกับฟังก์ชันเชื่อมโยง กรณีขนาดตัวอย่าง 50

ตัวแบบ	จำลองผ่านฟังก์ชันเชื่อมโยง					
	Logit		Probit		Cloglog	
	Average of AIC	Average of BIC	Average of AIC	Average of BIC	Average of AIC	Average of BIC
Logit	**254.3310**	**261.9791**	254.9003	262.5484	255.4422	263.0903
Probit	254.4817	262.1298	**254.7603**	**262.4084**	255.8260	263.4741
Cloglog	254.4901	262.1382	255.4417	263.0898	**255.2764**	**262.9245**

หมายเหตุ : ตัวหนา หมายถึง ค่าเฉลี่ย 1,000 รอบ AIC (Average of AIC) และค่าเฉลี่ย 1000 รอบ BIC (Average of BIC) ของตัวแบบที่ให้ค่าต่ำที่สุด

ตารางที่ 4.13 เปรียบเทียบตัวแบบจำลองโดยใช้เปอร์เซ็นต์ของ AIC และ BIC ต่ำสุด ที่สอดคล้องกับฟังก์ชันเชื่อมโยง กรณีขนาดตัวอย่าง 50

ตัวแบบ	จำลองผ่านฟังก์ชันเชื่อมโยง					
	Logit		Probit		Cloglog	
	% $Minimum$ AIC	% $Minimum$ BIC	% $Minimum$ AIC	% $Minimum$ BIC	% $Minimum$ AIC	% $Minimum$ BIC
Logit	16.3%	16.3%	16.6%	16.6%	13.5%	13.5%
Probit	**42.4%**	**42.4%**	**55.8%**	**55.8%**	30.8%	30.8%
Cloglog	41.3%	41.3%	27.6%	27.6%	**55.7%**	**55.7%**

หมายเหตุ : ตัวหนา หมายถึง % Minimum AIC และ % Minimum BIC ของตัวแบบที่ให้ค่าสูงที่สุด

จากตารางที่ 4.10 และ 4.11 แสดงการเปรียบเทียบตัวแบบจำลองโดยใช้ค่าเฉลี่ย 1,000 รอบ ของ AIC (Average of AIC) และค่าเฉลี่ย 1,000 รอบ BIC (Average of BIC) ผลลัพธ์ที่แสดงในตารางที่ 4.10 เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงโพรบิต ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด และเมื่อจำลองผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด และเพื่อเป็นการยืนยันผู้วิจัยได้หา % Minimum ของตัวแบบจำลองผ่านฟังก์ชันเชื่อมโยงทั้ง 3 ฟังก์ชัน ที่ให้ค่า AIC และ BIC ต่ำสุด จากกระทำซ้ำ 1,000 รอบ แสดงดังตารางที่ 4.11 พบว่า เมื่อจำลองผ่านฟังก์ชัน

เชื่อมโยงลอจิต จะได้ ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพตบิต ให้ค่า % Minimum มากสุดเท่ากับ 42.4% ซึ่งขัดแย้งกับค่าเฉลี่ย 1,000 รอบ แต่เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงโพรบิตและคอมพลีเมนทารีล็อก – ล็อก ให้ค่าสอดคล้องกับ ค่าเฉลี่ย 1,000 รอบ

ตารางที่ 4.14 เปรียบเทียบตัวแบบจำลองโดยใช้ค่าเฉลี่ยของ AIC และ BIC ที่สอดคล้องกับฟังก์ชันเชื่อมโยง กรณีขนาดตัวอย่าง 200

ตัวแบบ	จำลองผ่านฟังก์ชันเชื่อมโยง					
	Logit		Probit		Cloglog	
	Average of AIC	Average of BIC	Average of AIC	Average of BIC	Average of AIC	Average of BIC
Logit	**1007.431**	**1020.624**	1005.307	1018.500	1010.458	1023.651
Probit	1008.009	1021.203	**1004.562**	**1017.755**	1012.076	1025.270
Cloglog	1008.320	1021.513	1007.755	1020.948	**1009.724**	**1022.917**

หมายเหตุ : ตัวหนา หมายถึง ค่าเฉลี่ย 1,000 รอบ AIC (Average of AIC) และค่าเฉลี่ย 1000 รอบ BIC (Average of BIC) ของตัวแบบที่ให้ค่าต่ำที่สุด

ตารางที่ 4.15 เปรียบเทียบตัวแบบจำลองโดยใช้เปอร์เซ็นต์ของ AIC และ BIC ต่ำสุด ที่สอดคล้องกับฟังก์ชันเชื่อมโยง กรณีขนาดตัวอย่าง 200

ตัวแบบ	จำลองผ่านฟังก์ชันเชื่อมโยง					
	Logit		Probit		Cloglog	
	% *Minimum* AIC	% *Minimum* BIC	% *Minimum* AIC	% *Minimum* BIC	% *Minimum* AIC	% *Minimum* BIC
Logit	32.8%	32.8%	21%	21%	21.9%	21.9%
Probit	**36.6%**	**36.6%**	**68.6%**	**68.6%**	11.5%	11.5%
Cloglog	30.6%	32.1%	10.4%	10.4%	**66.6%**	**66.6%**

หมายเหตุ : ตัวหนา หมายถึง % Minimum AIC และ % Minimum BIC ของตัวแบบที่ให้ค่าสูงที่สุด

จากตารางที่ 4.12 และ 4.13 แสดงการเปรียบเทียบตัวแบบจำลองโดยใช้ค่าเฉลี่ย 1,000 รอบของ AIC (Average of AIC) และค่าเฉลี่ย 1,000 รอบ BIC (Average of BIC) ผลลัพธ์ที่แสดงในตารางที่ 4.12 เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงโพรบิต ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยง

โพรบิต ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด และเมื่อจำลองผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด และเพื่อเป็นการยืนยันผู้วิจัยได้หา % Minimum ของตัวแบบจำลองผ่านฟังก์ชันเชื่อมโยงทั้งสามฟังก์ชัน ที่ให้ค่า AIC และ BIC ต่ำสุด จากกระทำซ้ำ 1,000 รอบ แสดงดังตารางที่ 4.13 พบว่า เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต จะได้ ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต ให้ค่า % Minimum มากสุด เท่ากับ 36..6% ซึ่งขัดแย้งกับค่าเฉลี่ย 1,000 รอบ แต่เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงโพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่าสอดคล้องกับ ค่าเฉลี่ย 1,000 รอบ

ตารางที่ 4.16 เปรียบเทียบตัวแบบจำลองโดยใช้ค่าเฉลี่ยของ AIC และ BIC ที่สอดคล้องกับฟังก์ชันเชื่อมโยง กรณีขนาดตัวอย่าง 750

ตัวแบบ	จำลองผ่านฟังก์ชันเชื่อมโยง					
	Logit		Probit		Cloglog	
	Average of AIC	Average of BIC	Average of AIC	Average of BIC	Average of AIC	Average of BIC
Logit	**3766.872**	**3785.352**	3758.549	3777.030	3780.437	3798.918
Probit	3769.215	2787.696	**3756.055**	**3774.535**	3786.716	3805.196
Cloglog	3770.062	3788.542	3767.669	3786.149	**3777.551**	**3796.031**

หมายเหตุ : ตัวหนา หมายถึง ค่าเฉลี่ย 1,000 รอบ AIC (Average of AIC) และค่าเฉลี่ย 1000 รอบ BIC (Average of BIC) ของตัวแบบที่ให้ค่าต่ำที่สุด

ตารางที่ 4.17 เปรียบเทียบตัวแบบจำลองโดยใช้เปอร์เซ็นต์ของ AIC และ BIC ต่ำสุด ที่สอดคล้องกับฟังก์ชันเชื่อมโยง กรณีขนาดตัวอย่าง 750

ตัวแบบ	จำลองผ่านฟังก์ชันเชื่อมโยง					
	Logit		Probit		Cloglog	
	% *Minimum* AIC	% *Minimum* BIC	% *Minimum* AIC	% *Minimum* BIC	% *Minimum* AIC	% *Minimum* BIC
Logit	**59.6%**	**59.6%**	21.2%	21.2%	18.2%	18.2%
Probit	21.4%	21.4%	**77.8%**	**77.8%**	0.6%	0.6%
Cloglog	19%	19%	1%	1%	**81.2%**	**81.2%**

หมายเหตุ : ตัวหนา หมายถึง % Minimum AIC และ % Minimum BIC ของตัวแบบที่ให้ค่าสูงที่สุด

จากตารางที่ 4.14 และ 4.15 แสดงการเปรียบเทียบตัวแบบจำลองโดยใช้ค่าเฉลี่ย 1,000 รอบ ของ AIC (Average of AIC) และค่าเฉลี่ย 1,000 รอบ BIC (Average of BIC) ผลลัพธ์ที่แสดงในตารางที่ 4.8 เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงโพรบิต ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด และเมื่อจำลองผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ให้ค่าเฉลี่ย AIC และ BIC ต่ำที่สุด และเพื่อเป็นการยืนยันผู้วิจัยได้หา % Minimum ของตัวแบบจำลองผ่านฟังก์ชันเชื่อมโยงทั้งสามฟังก์ชัน ที่ให้ค่า AIC และ BIC ต่ำสุด จากกระทำซ้ำ 1,000 รอบ แสดงดังตารางที่ 4.11 พบว่า เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต จะได้ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต ให้ค่า % Minimum มากสุด เท่ากับ 59.6% ซึ่งสอดคล้องกับค่าเฉลี่ย 1,000 รอบ เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงโพรบิตจะได้ ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงโพรบิต ให้ค่า % Minimum มากสุด เท่ากับ 77.8% ซึ่งสอดคล้องกับค่าเฉลี่ย 1,000 รอบ และเมื่อจำลองผ่านฟังก์ชันเชื่อมโยงผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก จะได้ ตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ให้ค่า % Minimum มากสุด เท่ากับ 81.6% ซึ่งสอดคล้องกับค่าเฉลี่ย 1,000 รอบ

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

5.1.1 สรุปผลจากการประยุกต์ใช้กับข้อมูลจริง

การวิจัยนี้มีวัตถุประสงค์หลักในการศึกษาและวิเคราะห์ข้อมูลเกี่ยวกับอัตราการพ้นสภาพการเป็นนักศึกษาและปัจจัยที่อาจมีผลต่อจำนวนนักศึกษาที่พ้นสภาพ รวบรวมจากนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ จำนวน 20 หลักสูตร และถูกวิเคราะห์โดยใช้ตัวแบบการถดถอยทวินาม (Binomial Regression Analysis) โดยผู้วิจัยได้สร้างตัวแบบเพื่อวิเคราะห์ปัจจัยที่มีผลต่ออัตราการพ้นสภาพการเป็นนักศึกษา การวิเคราะห์นี้ได้ดำเนินการโดยใช้ฟังก์ชันการเชื่อมโยงที่แตกต่างกัน ได้แก่ ลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก พบว่าตัวแบบการทำนายทวินาม ผ่านฟังก์ชันการเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ให้ค่า AIC และ BIC ต่ำที่สุด และผลจากการวิเคราะห์หาปัจจัยที่ส่งผลต่ออัตราการพ้นสภาพการเป็นนักศึกษา จากทั้งหมด 10 ปัจจัย พบว่ามี 3 ปัจจัยที่ส่งผลต่ออัตราการพ้นสภาพการเป็นนักศึกษา คือ ปัจจัยคะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร ปัจจัยจำนวนนักศึกษาที่เข้ามารอบ 3 แต่ละหลักสูตร และปัจจัยคะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร ซึ่งทั้ง 3 ปัจจัยสามารถนำไปทำนายอัตราการพ้นสภาพนักศึกษาได้ โดยใช้ตัวแบบการทำนายทวินาม ผ่านฟังก์ชันการเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก ดังนี้

$$\hat{\pi}_i = 1 - \exp(-exp(-0.2465240 - 0.0335700X_{i5} + -0.0179550X_{i7} + 0.0001074X_{i8})) \quad (5.1)$$

เมื่อ $\hat{\pi}_i$ แทน ค่าประมาณอัตราการพ้นสภาพการเป็นนักศึกษา แต่ละหลักสูตร

X_{i5} แทน คะแนน Admission (รอบ 3) สูงสุด แต่ละหลักสูตร

X_{i7} แทน จำนวนนักศึกษาที่เข้ามารอบ 3 แต่ละหลักสูตร

X_{i8} แทน คะแนน Admission (รอบ 4) สูงสุด แต่ละหลักสูตร

จากตัวแบบการทำนายทวินาม ผ่านฟังก์ชันการเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก (5.1) เป็นตัวแบบทำนายอัตราการพ้นสภาพนักศึกษา ทั้งนี้ในการรับนักศึกษาเข้าศึกษาต่อของคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ยังใช้เกณฑ์เดิมในการรับนักศึกษาเข้าศึกษาในปีถัดถัดไป จะสามารถทำนายอัตราการพ้นสภาพนักศึกษาในแต่ละหลักสูตร หรือจำนวนนักศึกษาที่พ้นสภาพในแต่ละหลักสูตร อย่างเช่นปี 2563 สามารถมาคำนวณอัตราการพ้นสภาพนักศึกษาได้ เนื่องจากยังใช้เกณฑ์เดิมในการรับนักศึกษาเข้าศึกษา ซึ่งข้อมูลการทำนายอัตราการพ้นสภาพเป็นประโยชน์กับคณะวิทยาศาสตร์และเทคโนโลยีที่จะวางแผนรับมือกับจำนวนนักศึกษาที่พ้นสภาพในแต่ละหลักสูตร

5.1.2 สรุปผลจากการศึกษาเชิงจำลอง

จากการศึกษาเชิงจำลองเป็นการยืนยันเชิงประจักษ์ว่าการเลือกใช้ฟังก์เชื่อมโยงที่ไม่เหมาะสมอาจทำให้ตัวแบบการถดถอยทวินามไม่เหมาะสม ซึ่งการเลือกฟังก์ชันเชื่อมโยงที่ผิดจะส่งผลต่อความเอนเอียง (Bias) เป็นอย่างมากต่อพารามิเตอร์การถดถอยและค่าประมาณของตัวแปรตอบสนอง ดังนั้นการเลือกฟังก์ชันเชื่อมโยงที่เหมาะสมยังคงเป็นสิ่งสำคัญ และการจำลองข้อมูลโดยกำหนดสถานการณ์ให้ใกล้เคียงกับกรณีศึกษาอัตราการพ้นสภาพการเป็นนักศึกษาคณะวิทยาศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ซึ่งกำหนดขนาดตัวอย่างที่แตกต่างกัน คือ 20, 50, 200 และ 750 สรุปผลได้ดังต่อไปนี้

ขนาดตัวอย่างเท่ากับ 20 เมื่อผู้วิจัยใช้เกณฑ์การเปรียบเทียบตัวแบบจำลอง โดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ พบว่าเมื่อจำลองฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ผ่านตัวแบบทวินามโดยฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่าเฉลี่ย AIC และ BIC 1,000 รอบ ต่ำสุด เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงตัวแบบเดียวกัน นอกจากนี้เพื่อเป็นการยืนยันการศึกษาเชิงจำลองผู้วิจัยได้ใช้เกณฑ์การเปรียบเทียบตัวแบบจำลอง โดยใช้ค่าร้อยละของจำนวนครั้งที่ตัวแบบทวินามจำลองผ่านฟังก์ชันเชื่อมลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่า AIC และ BIC ต่ำสุดในแต่ละรอบ จำนวน 1,000 รอบ พบว่าเมื่อจำลองฟังก์ชันเชื่อมโยงโพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ร้อยละ AIC และ BIC ของจำนวนครั้งสอดคล้องกัน โดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ แต่เมื่อจำลองฟังก์ชันเชื่อมโยงลอจิต ให้ร้อยละ AIC และ BIC ของจำนวนครั้งขัดแย้งกันโดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ แสดงให้เห็นว่าตัวแบบทวินามที่จำลองผ่านฟังก์ชันเชื่อมโยงลอจิต อาจไม่เหมาะสมที่จะนำมาใช้กับตัวแบบ เมื่อขนาดตัวอย่างเท่ากับ 20

ขนาดตัวอย่างเท่ากับ 50 เมื่อผู้วิจัยใช้เกณฑ์การเปรียบเทียบตัวแบบจำลอง โดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ พบว่าเมื่อจำลองฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ผ่านตัวแบบทวินามโดยฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่าเฉลี่ย AIC และ BIC 1,000 รอบ ต่ำสุดเมื่อจำลองผ่านฟังก์ชันเชื่อมโยงตัวแบบเดียวกัน นอกจากนี้เพื่อเป็นการยืนยันการศึกษาเชิงจำลองผู้วิจัยได้ใช้เกณฑ์การเปรียบเทียบตัวแบบจำลอง โดยใช้ค่าร้อยละของจำนวนครั้งที่ตัวแบบทวินามจำลองผ่านฟังก์ชันเชื่อมลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่า AIC และ BIC ต่ำสุดในแต่ละรอบ จำนวน 1,000 รอบ พบว่าเมื่อจำลองฟังก์ชันเชื่อมโยงโพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ร้อยละ AIC และ BIC ของจำนวนครั้งที่สอดคล้องกัน โดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ แต่เมื่อจำลองฟังก์ชันเชื่อมโยงลอจิต ให้ร้อยละ AIC และ BIC ของจำนวนครั้งขัดแย้งกัน โดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ แสดงให้เห็นว่าตัวแบบทวินามที่จำลองผ่านฟังก์ชันเชื่อมโยงลอจิต อาจไม่เหมาะสมที่จะนำมาใช้กับตัวแบบ เมื่อขนาดตัวอย่างเท่ากับ 50

ขนาดตัวอย่างเท่ากับ 200 เมื่อผู้วิจัยใช้เกณฑ์การเปรียบเทียบตัวแบบจำลอง โดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ พบว่าเมื่อจำลองฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ผ่านตัวแบบทวินาม โดยฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่าเฉลี่ย AIC และ

BIC 1,000 รอบ ต่ำสุด เมื่อจำลองผ่านฟังก์ชันเชื่อมโยงตัวแบบเดียวกัน นอกจากนี้เพื่อเป็นการยืนยันการศึกษาเชิงจำลองผู้วิจัยได้ใช้เกณฑ์การเปรียบเทียบตัวแบบจำลอง โดยใช้ค่าร้อยละของจำนวนครั้งที่ตัวแบบทวินามจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่า AIC และ BIC ต่ำสุดในแต่ละรอบ จำนวน 1,000 รอบ พบว่าเมื่อจำลองฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ร้อยละ AIC และ BIC ของจำนวนครั้งที่สอดคล้องกัน โดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ แต่เมื่อจำลองฟังก์ชันเชื่อมโยงลอจิต ให้ร้อยละ AIC และ BIC ของจำนวนครั้งที่ขัดแย้งกัน โดยใช้ค่า AIC และBIC เฉลี่ย 1,000 รอบ แสดงให้เห็นว่าตัวแบบทวินามที่จำลองผ่านฟังก์ชันเชื่อมโยงลอจิต อาจไม่เหมาะสมที่จะนำมาใช้กับตัวแบบ เมื่อขนาดตัวอย่างเท่ากับ 200

ขนาดตัวอย่างเท่ากับ 750 เมื่อผู้วิจัยใช้เกณฑ์การเปรียบเทียบตัวแบบจำลอง โดยใช้ค่า AIC และ BIC เฉลี่ย 1,000 รอบ พบว่าเมื่อจำลองเมื่อฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ผ่านตัวแบบทวินามโดยฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่าเฉลี่ย AIC และ BIC 1,000 รอบ ต่ำสุดเมื่อจำลองผ่านฟังก์ชันเชื่อมโยงตัวแบบเดียวกัน นอกจากนี้เพื่อเป็นการยืนยันการศึกษาเชิงจำลองผู้วิจัยได้ใช้เกณฑ์การเปรียบเทียบตัวแบบจำลอง โดยใช้ค่าร้อยละของจำนวนครั้งที่ตัวแบบทวินามจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ค่า AIC และ BIC ต่ำสุด ในแต่ละรอบ จำนวน 1,000 รอบ พบว่าเมื่อจำลองฟังก์ชันเชื่อมลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก ให้ร้อยละ AIC และ BIC ของจำนวนครั้งที่สอดคล้องกัน โดยใช้ค่า AIC และBIC เฉลี่ย 1,000 รอบ แสดงให้เห็นว่าตัวแบบทวินามที่จำลองผ่านฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก – ล็อก เหมาะสมที่จะนำมาใช้กับตัวแบบ เมื่อขนาดตัวอย่างเท่ากับ 750

5.2 วิจารณ์ผลและข้อเสนอแนะ

1. ในการประยุกต์ใช้กับข้อมูลจริงผู้วิจัยได้ใช้วิธีการประมาณค่าพารามิเตอร์ คือ วิธีภาวะน่าจะเป็นสูงสุด แต่เนื่องจากขนาดข้อมูลจริงที่น้อย สำหรับงานวิจัยครั้งต่อไปอาจศึกษาวิธีในการประมาณค่าพารามิเตอร์ด้วยวิธีอื่น เช่น วิธีประมาณค่าพารามิเตอร์แบบบูตสแตรป
2. ในงานวิจัยครั้งนี้ผู้วิจัยใช้ฟังก์ชันเชื่อมโยงเพียง 3 ฟังก์ชัน ประกอบไปด้วย ฟังก์ชันเชื่อมลอจิต, ฟังก์ชันเชื่อมโยงโพรบิต และฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก – ล็อก แต่สำหรับตัวแบบการถดถอยทวินามยังมีฟังก์ชันเชื่อมโยงที่เหมาะสมอีกหนึ่งฟังก์ชัน คือ ฟังก์ชันเชื่อมโยงคอชิต สามารถใช้เป็นฟังก์ชันเชื่อมโยงสำหรับงานวิจัยครั้งต่อไปได้
3. ในงานวิจัยครั้งต่อไปควรจะเพิ่มขนาดตัวอย่างให้มากขึ้น โดยการขยายขอบเขตการเก็บข้อมูลจากหลักสูตรในคณะวิทยาศาสตร์ เป็นทุกหลักสูตรของทุกคณะในมหาวิทยาลัยธรรมศาสตร์

บรรณานุกรม

หนังสือและบทความในหนังสือ

วีรานันท์ พงศาภักดี. (2555). *การวิเคราะห์ข้อมูลจำแนกประเภท: ทฤษฎีและการประยุกต์ด้วย GLIM, SPSS, SAS และ MTB* (พิมพ์ครั้งที่ 3). โรงพิมพ์มหาวิทยาลัยศิลปากร.

P. McCullagh and J.A. Nelder, *Generalized Linear Models 2nd Ed.* (Chapman and Hall , London, 1989)

The University of Sydney. (2020). *Variable Selection: Stepwise, ALC and BIC*

บทความวารสาร

Gunduz N, Fokoue E. (2013). *On the predictive analytics of the probit and logit link functions*.
RIT Scholar Works.

Wu, L., & Lord, D. (2017). Examining the influence of link function misspecification in c onventional regression models for developing crash modification factors. Accident Analysis & Prevention,102, 123-135.

Prasetyo, Rindang Bangun, et al. “A Comparison of Some Link Functions for Binomial Regression Models with Application to School Drop-out Rates in East Java.” AIP Conference Proceedings, 2019, https://doi.org/10.1063/1.5139815.

Kushagra Jain. (2022) Beginner’s Guide to Algorithmic Trading in R (Part 5/6) — Machine Learning Backtesting, from https://medium.com/ @kushagrajain7augtrading/beginners-guide-to-algorithmic-trading-in-r-part-5-6- machine-learning-backtesting-c21da20fbe30

Gunduz N, Fokoue E. (2015). *On the Predictive Properties of Binary Link Function*.

Koenker R, Yoon J. (2009). Parametric links for binary choice models: A Fisherian–Bayesian colloquy. *Journal of Econometrics*, 152(2), 1-25.

วิทยานิพนธ์

ปวีณ์กร มิ่งเชื้อ (2564). *การคัดเลือกตัวแปรสำหรับตัวแบบการถดถอยไวบูลไม่ต่อเนื่อง.* มหาวิทยาลัยธรรมศาสตร์.

Li, Jingwei. (2014). Choosing the proper link function for binary data (master's thesis). University of Texas.

สุนิสา จันทร์น้ำท่วม (2561). *ผลกระทบของฟังก์ชันเชื่อมโยงและตัวแบบที่มีผลต่อช่วงความเชื่อมั่นสำหรับพารามิเตอร์ส่วนประกอบแบร์นูลลี*. มหาวิทยาลัยธรรมศาสตร์.

จันทิรา แย้มสรวล (2559). *การประมาณขนาดประชากรภายใต้การแจกแจงคอนเวย์แมกซ์เวลล์ปัวซง.* มหาวิทยาลัยธรรมศาสตร์.

โศจิรา พรประดิษฐ์พันธุ์ (2552). *การประมาณค่าพารามิเตอร์การกระจายภายใต้ตัวแบบเชิงเส้นนัยทั่วไปที่ตัวแปรมีการแจกแจงทวินามลบ เมื่อตัวอย่างขนาดเล็ก.* มหาวิทยาลัยธรรมศาสตร์

พัชรพรรณ ขุมแร่ (2556). *การพัฒนาชุดคำสั่งโปรแกรม R ที่ทำงานร่วมกับโปรแกรมประยุกต์บนเว็บสำหรับการวิเคราะห์การถดถอยลอจิสติกทวินาม.* มหาวิทยาลัยธรรมศาสตร์.

เว็บไซต์

Rodríguez, G. (2007). *Lecture Notes on Generalized Linear Models*. Retrieved January 15, 2019, from http://data.princeton.edu/wws509/notes/.

Sherry Towers. (2018). *Logistic (Binomial) regression*. https://sherrytowers.com/2018/03/07/logistic-binomial-regression/

Mustafa, A. (2023). *A Gentle Introduction to Complementary Log-Log Regression*. Retrieved October 31, 2023, from https://towardsdatascience.com/a-gentle-introduction-to-complementary-log-log-regression-8ac3c5c1cd83

ภาคผนวก

1
โปรแกรมคอมพิวเตอร์ที่ใช้ในงานวิจัย

การประยุกต์ใช้กับข้อมูลจริงและการศึกษาเชิงจำลองสามารถทำได้โดยการเขียนคำสั่งในโปรแกรม RStudio เวอร์ชัน 2023.03.0+386 ในการจำลองตัวแปรอธิบาย ตัวแปรตอบสนอง และการประมาณค่าพารามิเตอร์ของตัวแบบเชิงเส้นทั่วไปที่ตัวแปรมีการแจกแจงทวินามด้วย วิธีภาวะน่าจะเป็นสูงสุด (MLE) โดยการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุดจะเป็นวิธีฟิชเชอร์สกอริง (Fisher’s Scoring Method) ซึ่งเป็นวิธีที่ใช้สำหรับการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก-ล็อก ในคำสั่ง glm ของโปรแกรม Rstudio

ก.1 โปรแกรมสำหรับประยุกต์ใช้กับข้อมูลจริง

```
# Set Yi Number of students dropping out of each program Variables
Yi <- c(16, 16, 8, 10, 4, 3, 15, 14, 25, 3, 27, 14, 13, 16, 10, 12, 16, 11, 15, 13)
# Set Number of students admitted in Year 1, Semester 1
ni <- c(61, 58, 52, 38, 20, 50, 60, 85, 66, 58, 43, 65, 47, 50, 54, 72, 53, 57, 60, 32)
# Calculate the proportion of students dropping out for each program
pi <- Yi/ni
# Set X1  compulsory subjects C
x1 <- c(1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0)
# Set X2 a few credits each course
x2 <- c(137, 133, 133, 137, 133, 122, 137, 138, 132, 138, 129, 138, 129, 138, 138, 129, 138, 138, 129, 138)
# Set X3 college tuition fees
x3 <- c(17300, 17300, 17300, 33900, 33800, 44300, 17300, 17300, 17300, 17300, 17300, 17300, 17300, 17300,
17300, 17300, 17300, 17300, 45700, 47900)
# Set X4 Normal Project (1) or Special Project (0)
x4 <- c(1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0)
# Set X5 Highest Score Admission Round 3
x5 <- c(65.24, 67.71, 59.34, 57.41, 56.7, 75.87, 62.11, 42.04, 65.02, 72.45, 30.33, 60.41, 63.46,
        56.61, 54.97, 66.65, 58.43, 60.37, 54.29, 55.25)
# Set X6 Lowest score Admission Round 3
x6 <- c(51.42, 48.02, 45.96, 33.40, 32.06, 56.56, 48.56, 55.66, 53.14, 54.01, 56.12, 46.76, 46.03,
        42.99, 40.26, 52.34, 35.04, 48.83, 37.67, 35.6)
# Set X7 The number of students admitted in Round 3
x7 <- c(48, 47, 45, 31, 18, 42, 52, 74, 46, 41, 24, 44, 43, 32, 36, 44, 46, 42, 40, 23)
# Set X8 Highest Score Admission Round 4
x8 <- c(16644.10, 16597.90, 16867.00, 14842.90, 12127.10, 14125.50, 16567.00, 14247.40,
```

```
       17075.70, 17017.10, 14920.40, 16750.50, 14484.20, 14870.10, 15926.30, 17239.50,
        15236.80, 17017.10, 14030.20, 15976.00)
# Set X9 Lowest score Admission Round 4
x9 <- c(14784.70, 14225.30, 13786.80, 10962.10, 12089.10, 17821.00, 14030.20, 13229.50 ,14863.90,
        15318.80, 12230.70, 13893.10, 14022.40, 13095.40, 13163.60, 13867.20, 13416.70, 15318.8,
        10578.40, 13158.7)
# Set X10 The number of students admitted in Round 4
x10 <- c(12, 7, 7, 7, 2, 8, 7, 11, 16, 12, 18, 21,  2, 18, 18, 27, 7, 15, 20, 9)
##############################  คัดเลือกตัวแปร stepwise #####################
# Define the full model logit
fullModel_logit <- glm(cbind(Yi, ni - Yi) ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10, family =
binomial(link = "logit"))
nullModel_logit <- glm(cbind(Yi, ni - Yi) ~ 1, family = binomial(link = "logit")) model with the intercept only
# PerformStepwise elimination
bothways_logit <- stepAIC(nullModel_logit,
                    direction = "both", # runStepwise selection
                    scope = list(upper = fullModel_logit,
                              lower = nullModel_logit))
# Print the final selected model
summary(bothways_logit)
AIC(bothways_logit)
BIC(bothways_logit)
# Define the full model probit
fullModel_probit <- glm(cbind(Yi, ni - Yi) ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10, family =
binomial(link = "probit"))
nullModel_probit <- glm(cbind(Yi, ni - Yi) ~ 1, family = binomial(link = "probit")) ### model with the
intercept only
# Perform Stepwise elimination
bothways_probit <- stepAIC(nullModel_probit,
                    direction = "both",
                    scope = list(upper = fullModel_probit,
                              lower = nullModel_probit))
# Print the final selected model
summary(bothways_probit)
AIC(bothways_probit)
BIC(bothways_probit)
```

```
# Define the full model cloglog
fullModel_cloglog <- glm(cbind(Yi, ni - Yi) ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10, family =
binomial(link = "cloglog"))
nullModel_cloglog <- glm(cbind(Yi, ni - Yi) ~ 1, family = binomial(link = "cloglog")) ### model with the
intercept only
# Perform Stepwise elimination
bothways_cloglog <- stepAIC(nullModel_cloglog,
                        direction = "both", # run Stepwise selection
                        scope = list(upper = fullModel_cloglog,
                                    lower = nullModel_cloglog))
# Print the final selected model
formula(bothways_cloglog)
summary(bothways_cloglog)
AIC(bothways_cloglog)
BIC(bothways_cloglog)
bothways_cloglog <- stepAIC(nullModel_cloglog, direction = "both",  scope = list(upper =
fullModel_cloglog, lower = nullModel_cloglog))
```

ก.2 โปรแกรมสำหรับการจำลองข้อมูลของตัวแปรอธิบายและตัวแปรตอบสนอง เมื่อตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก-ล็อก สำหรับขนาดตัวอย่างเท่ากับ 200 โดยจำลองผ่านฟังก์ชันเชื่อมโยงลอจิต

```
####################################Link function logit is pi ###########################
n2 = 200
# Set the number of iterations
niter_n200_linklogit <- 1000
# Create empty vectors to store AIC and BIC values
AIC_n200_logit_linklogit <- rep(0, niter_n200_linklogit)
BIC_n200_logit_linklogit <- rep(0, niter_n200_linklogit)
AIC_n200_probit_linklogit <- rep(0, niter_n200_linklogit)
BIC_n200_probit_linklogit <- rep(0, niter_n200_linklogit)
AIC_n200_cloglog_linklogit <- rep(0, niter_n200_linklogit)
BIC_n200_cloglog_linklogit <- rep(0, niter_n200_linklogit)
for (i in 1:niter_n200_linklogit) {
  # Define the coefficients for the binomial regression model
  #beta <- c(0.13279, -0.04084, 0.12983, -0.02155) # SetNew_logit
  beta <- c(0.1328000, -0.0408400, 0.0001298, -0.0215500) # SetNewNew_logit
  # Set the Poisson parameter
```

```
  lambda1 <- 54  # Average number of students admitted
  lambda2 <- 50 # Average of number of students accepted in the 3rd round
  # Simulate independent variables x1 and x2 from a normal distribution
  x1 <- runif(n2, min = 30.33,   max = 75.87)      # Highest Score Admission Round 3 runif(n, min, max)
  x2 <- runif(n2, min = 12127.1, max = 17239.5)    # Highest score Admission Round 4 runif(n, min, max) X8
  # Simulate independent variables x3 from a Bernoulli distribution
  x3 <- rpois(n2, lambda2) # Number of students accepted in the 3rd round
  # Combine independent variables into a matrix X
  X <- matrix(0, nrow = n2, ncol = 4)
  X[, 1] <- t(rep(1, n2))  # intercept term
  X[, 2] <- x1
  X[, 3] <- x2
  X[, 4] <- x3
  # Generate the Poisson counts
  ni <- rpois(n2, lambda1) # Number of students admitted
  # Calculate the predicted probabilities pi for each observation in X
  PropY_linklogit <- exp(X %*% beta) / (1 + exp(X %*% beta))
  # Simulate the dependent variable y for each observation from a binomial distribution
  Y_logit <- rbinom(n2, ni, PropY_linklogit)
  Y_logit
  # Fit the binomial regression model using glm function
  binomial_model_logit <- glm( cbind(Y_logit, ni-Y_logit) ~ x1 + x2 + x3, family = binomial(link = "logit"))
  binomial_model_probit <- glm( cbind(Y_logit, ni-Y_logit) ~ x1 + x2 + x3, family = binomial(link = "probit"))
  binomial_model_cll <- glm( cbind(Y_logit, ni-Y_logit) ~ x1 + x2 + x3, family = binomial(link = "cloglog"))
  # View the summary of the model
  summary(binomial_model_logit)
  summary(binomial_model_probit)
  summary(binomial_model_cll)
  # Store the AIC and BIC values
  AIC_n200_logit_linklogit[i] <- AIC(binomial_model_logit)
  BIC_n200_logit_linklogit[i] <- BIC(binomial_model_logit)
  AIC_n200_probit_linklogit[i] <- AIC(binomial_model_probit)
  BIC_n200_probit_linklogit[i] <- BIC(binomial_model_probit)
  AIC_n200_cloglog_linklogit[i] <- AIC(binomial_model_cll)
  BIC_n200_cloglog_linklogit[i] <- BIC(binomial_model_cll)
}
```

```
# set mean
meanAIC_logit_linklogit_n200 <- mean(AIC_n200_logit_linklogit)
meanBIC_logit_linklogit_n200 <- mean(BIC_n200_logit_linklogit)
meanAIC_probit_linklogit_n200 <- mean(AIC_n200_probit_linklogit)
meanBIC_probit_linklogit_n200 <- mean(BIC_n200_probit_linklogit)
meanAIC_cloglog_linklogit_n200 <- mean(AIC_n200_cloglog_linklogit)
meanBIC_cloglog_linklogit_n200 <- mean(BIC_n200_cloglog_linklogit)
# Create a data frame to store the confidence intervals and model fit statistics
model_comparison_n200_linklogit <- data.frame(
  Model = c("Logit", "Probit", "Complementary Log-Log"),
  AIC = c(meanAIC_logit_linklogit_n200, meanAIC_probit_linklogit_n200 , meanAIC_cloglog_linklogit_n200),
  BIC = c(meanBIC_logit_linklogit_n200, meanBIC_probit_linklogit_n200, meanBIC_cloglog_linklogit_n200)
)
# Print the table
print(model_comparison_n200_linklogit)
# Count the lowest number of times
numberAIC_lesslogit_n200_logit   <- 0
numberAIC_lessprobit_n200_logit  <- 0
numberAIC_lesscloglog_n200_logit <- 0
numberBIC_lesslogit_n200_logit   <- 0
numberBIC_lessprobit_n200_logit  <- 0
numberBIC_lesscloglog_n200_logit <- 0
#################### Check lower AIC ##########################
for (i in 1:1000) {
  if (AIC_n200_logit_linklogit[i] < AIC_n200_probit_linklogit[i] &&
      AIC_n200_logit_linklogit[i] < AIC_n200_cloglog_linklogit[i]) {
    numberAIC_lesslogit_n200_logit = numberAIC_lesslogit_n200_logit + 1
  } else if (AIC_n200_probit_linklogit[i] < AIC_n200_logit_linklogit[i] &&
             AIC_n200_probit_linklogit[i] < AIC_n200_cloglog_linklogit[i]) {
    numberAIC_lessprobit_n200_logit = numberAIC_lessprobit_n200_logit + 1
  } else {
    numberAIC_lesscloglog_n200_logit = numberAIC_lesscloglog_n200_logit + 1
  }
}
cat("Total number of AIC (n = 200):", numberAIC_lesslogit_n200_logit + numberAIC_lessprobit_n200_logit
+numberAIC_lesscloglog_n200_logit, "\n")
```

```
cat("Minimum number of AIC logit  : ",  numberAIC_lesslogit_n200_logit,"or
",(numberAIC_lesslogit_n200_logit/1000)*100, "%" ,"\n")
cat("Minimum number of AIC probit : ",  numberAIC_lessprobit_n200_logit,"or
",(numberAIC_lessprobit_n200_logit/1000)*100,"%" , "\n")
cat("Minimum number of ACI cll    : ",  numberAIC_lesscloglog_n200_logit,"or
",(numberAIC_lesscloglog_n200_logit/1000)*100,"%", "\n")
################################ Check lower BIC ##############################
for (i in 1:1000) {
  if (BIC_n200_logit_linklogit[i] < BIC_n200_probit_linklogit[i] &&
      BIC_n200_logit_linklogit[i] < BIC_n200_cloglog_linklogit[i]) {
    numberBIC_lesslogit_n200_logit = numberBIC_lesslogit_n200_logit + 1
  } else if (BIC_n200_probit_linklogit[i] < BIC_n200_logit_linklogit[i] &&
             BIC_n200_probit_linklogit[i] < BIC_n200_cloglog_linklogit[i]) {
    numberBIC_lessprobit_n200_logit = numberBIC_lessprobit_n200_logit + 1
  } else {
    numberBIC_lesscloglog_n200_logit = numberBIC_lesscloglog_n200_logit + 1
  }
}
cat("Total number of BIC (n = 200):", numberBIC_lesslogit_n200_logit + numberBIC_lessprobit_n200_logit +
numberBIC_lesscloglog_n200_logit, "\n")
cat("Minimum number of BIC logit  : ",  numberBIC_lesslogit_n200_logit,"or
",(numberBIC_lesslogit_n200_logit/1000)*100, "%" ,"\n")
cat("Minimum number of BIC probit : ",  numberBIC_lessprobit_n200_logit,"or
",(numberBIC_lessprobit_n200_logit/1000)*100,"%" , "\n")
cat("Minimum number of BCI cll    : ",  numberBIC_lesscloglog_n200_logit,"or
",(numberBIC_lesscloglog_n200_logit/1000)*100,"%", "\n")
```

ก.3 โปรแกรมสำหรับการจำลองข้อมูลของตัวแปรอธิบายและตัวแปรตอบสนอง เมื่อตัวแบบการถดถอยทวินามผ่านฟังก์ชันเชื่อมโยงลอจิต, โพรบิต และคอมพลีเมนทารีล็อก-ล็อก สำหรับขนาดตัวอย่างเท่ากับ 200 โดยจำลองผ่านคอมพลีเมนทารีล็อก-ล็อก

```
####################### Link function Cloglog is pi ####################
n2 <- 200
# Set the number of iterations
niter_n200_linkcll <- 1000
# Create empty vectors to store AIC and BIC values
AIC_n200_logit_linkcll <- rep(0, niter_n200_linkcll)
```

```
BIC_n200_logit_linkcll <- rep(0, niter_n200_linkcll)
AIC_n200_probit_linkcll <- rep(0, niter_n200_linkcll)
BIC_n200_probit_linkcll <- rep(0, niter_n200_linkcll)
AIC_n200_cloglog_linkcll <- rep(0, niter_n200_linkcll)
BIC_n200_cloglog_linkcll <- rep(0, niter_n200_linkcll)
for (i in 1:niter_n200_linkcll) {
  # Define the coefficients for the binomial regression model
  #beta <- c(-0.24652, -0.03357, 0.10740, -0.01795) ## SetNew_Cloglog
  beta <- c(-0.2465240, -0.0335700, 0.0001074, -0.0179550) ## SetNewNew_Cloglog
  # Set the Poisson parameter
  lambda1 <- 54  # Average number of students admitted
  lambda2 <- 50 # Average of number of students accepted in the 3rd round
  # Simulate independent variables x1 and x2 from a normal distribution
  x1 <- runif(n2, min = 30.33,   max = 75.87)       # Highest Score Admission Round 3 runif(n, min, max)
  x2 <- runif(n2, min = 12127.1, max = 17239.5)    # Highest score Admission Round 4 runif(n, min, max) X8
  # Simulate independent variables x3 from a Bernoulli distribution
  x3 <- rpois(n2, lambda2) # Number of students accepted in the 3rd round
  # Combine independent variables into a matrix X
  X <- matrix(0, nrow = n2, ncol = 4)
  X[, 1] <- t(rep(1, n2))  # intercept term
  X[, 2] <- x1
  X[, 3] <- x2
  X[, 4] <- x3
  # Generate the Poisson counts
  ni <- rpois(n2, lambda1) # Number of students admitted
  # Calculate the predicted probabilities pi for each observation in X
  pi_cloglog <- 1 - exp(-exp(X %*% beta))
  # Simulate the dependent variable y for each observation from a binomial distribution
  Y_cloglog <- rbinom(n2, ni, pi_cloglog)
  Y_cloglog
  # Fit the binomial regression model using glm function
  binomial_model_logit <- glm( cbind(Y_cloglog, ni-Y_cloglog) ~ x1 + x2 + x3, family = binomial(link = 
"logit"))
  binomial_model_probit <- glm( cbind(Y_cloglog, ni-Y_cloglog) ~ x1 + x2 + x3, family = binomial(link = 
"probit"))
```

```
  binomial_model_cll <- glm( cbind(Y_cloglog, ni-Y_cloglog) ~ x1 + x2 + x3, family = binomial(link =
"cloglog"))
  # View the summary of the model
  summary(binomial_model_logit)
  summary(binomial_model_probit)
  summary(binomial_model_cll)
  AIC_n200_logit_linkcll[i] <- AIC(binomial_model_logit)
  BIC_n200_logit_linkcll[i] <- BIC(binomial_model_logit)
  AIC_n200_probit_linkcll[i] <- AIC(binomial_model_probit)
  BIC_n200_probit_linkcll[i] <- BIC(binomial_model_probit)
  AIC_n200_cloglog_linkcll[i] <- AIC(binomial_model_cll)
  BIC_n200_cloglog_linkcll[i] <- BIC(binomial_model_cll)

}
# Set mean
meanAIC_logit_linkcll_n200 <- mean(AIC_n200_logit_linkcll)
meanBIC_logit_linkcll_n200 <- mean(BIC_n200_logit_linkcll)
meanAIC_probit_linkcll_n200 <- mean(AIC_n200_probit_linkcll)
meanBIC_probit_linkcll_n200 <- mean(BIC_n200_probit_linkcll)
meanAIC_cloglog_linkcll_n200 <- mean(AIC_n200_cloglog_linkcll)
meanBIC_cloglog_linkcll_n200 <- mean(BIC_n200_cloglog_linkcll)
# Create a data frame to store the confidence intervals and model fit statistics
model_comparison_n200_linkcll <- data.frame(
  Model = c("Logit", "Probit", "Complementary Log-Log"),
  AIC = c(meanAIC_logit_linkcll_n200, meanAIC_probit_linkcll_n200 , meanAIC_cloglog_linkcll_n200),
  BIC = c(meanBIC_logit_linkcll_n200, meanBIC_probit_linkcll_n200, meanBIC_cloglog_linkcll_n200)
)
# Print the table
print(model_comparison_n200_linkcll)
# Count the lowest number of times
numberAIC_lesslogit_n200_cll   <- 0
numberAIC_lessprobit_n200_cll  <- 0
numberAIC_lesscloglog_n200_cll <- 0
numberBIC_lesslogit_n200_cll   <- 0
numberBIC_lessprobit_n200_cll  <- 0
numberBIC_lesscloglog_n200_cll <- 0
```

```
################################ Check lower AIC ##############################
for (i in 1:1000) {
  if (AIC_n200_logit_linkcll[i] < AIC_n200_probit_linkcll[i] &&
      AIC_n200_logit_linkcll[i] < AIC_n200_cloglog_linkcll[i]) {
    numberAIC_lesslogit_n200_cll = numberAIC_lesslogit_n200_cll + 1
  } else if (AIC_n200_probit_linkcll[i] < AIC_n200_logit_linkcll[i] &&
             AIC_n200_probit_linkcll[i] < AIC_n200_cloglog_linkcll[i]) {
    numberAIC_lessprobit_n200_cll = numberAIC_lessprobit_n200_cll + 1
  } else {
    numberAIC_lesscloglog_n200_cll = numberAIC_lesscloglog_n200_cll + 1
  }
}
cat("Total number of AIC (n = 20):", numberAIC_lesslogit_n200_cll + numberAIC_lessprobit_n200_cll
+numberAIC_lesscloglog_n200_cll, "\n")
cat("Minimum number of AIC logit  : ",  numberAIC_lesslogit_n200_cll,"or
",(numberAIC_lesslogit_n200_cll/1000)*100, "%" ,"\n")
cat("Minimum number of AIC probit : ",  numberAIC_lessprobit_n200_cll,"or
",(numberAIC_lessprobit_n200_cll/1000)*100,"%" , "\n")
cat("Minimum number of ACI cll    : ",  numberAIC_lesscloglog_n200_cll,"or
",(numberAIC_lesscloglog_n200_cll/1000)*100,"%", "\n")
########################## Check lower BIC #########################
for (i in 1:1000) {
  if (BIC_n200_logit_linkcll[i] < BIC_n200_probit_linkcll[i] &&
      BIC_n200_logit_linkcll[i] < BIC_n200_cloglog_linkcll[i]) {
    numberBIC_lesslogit_n200_cll = numberBIC_lesslogit_n200_cll + 1
  } else if (BIC_n200_probit_linkcll[i] < BIC_n200_logit_linkcll[i] &&
             BIC_n200_probit_linkcll[i] < BIC_n200_cloglog_linkcll[i]) {
    numberBIC_lessprobit_n200_cll = numberBIC_lessprobit_n200_cll + 1
  } else {
    numberBIC_lesscloglog_n200_cll = numberBIC_lesscloglog_n200_cll + 1
  }
}
cat("Total number of BIC (n = 20):", numberBIC_lesslogit_n200_cll + numberBIC_lessprobit_n200_cll +
numberBIC_lesscloglog_n200_cll, "\n")
cat("Minimum number of BIC logit  : ",  numberBIC_lesslogit_n200_cll,"or
",(numberBIC_lesslogit_n200_cll/1000)*100, "%" ,"\n")
```

```
cat("Minimum number of BIC probit : ",  numberBIC_lessprobit_n200_cll,"or
",(numberBIC_lessprobit_n200_cll/1000)*100,"%" , "\n")
cat("Minimum number of BCI cll     : ",  numberBIC_lesscloglog_n200_cll,"or
",(numberBIC_lesscloglog_n200_cll/1000)*100,"%", "\n")
```