

## การถดถอยโลจิสติก (Logistic Regression)

### ๑. หลักการ

การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) เป็นเทคนิคการวิเคราะห์ตัวแปรเชิงพหุที่มีวัตถุประสงค์เพื่อประมาณค่าหรือทำนายเหตุการณ์ที่สนใจว่าจะเกิดหรือไม่เกิดเหตุการณ์นั้นภายใต้อิทธิของตัวปัจจัย แบบจำลองโลจิสติกประกอบด้วยตัวแปรตาม (หรือตัวแปรเกณฑ์) ที่ต้องเป็นตัวแปรแบบทวินาม (Dichotomous Variable) กล่าวคือมีได้สองค่า เช่น “เกิด” กับ “ไม่เกิด” หรือ “เสี่ยง” กับ “ไม่เสี่ยง” เป็นต้น และตัวแปรอิสระ (หรือตัวแปรทำนาย) ที่อาจมีตัวเดียวหรือหลายตัวที่เป็นได้ทั้งตัวแปรเชิงกลุ่ม (Categorical Variable) หรือตัวแปรแบบต่อเนื่อง (Continuous Variable) การวิเคราะห์การถดถอยโลจิสติกเกี่ยวข้องกับทฤษฎีความน่าจะเป็นทวินามถูกเรียกว่า Binomial Logistic Regression ถ้าตัวแปรตามเป็นพหุนามจะเรียกว่า Multinomial Logistic Regression การถดถอยโลจิสติกจัดเป็นเครื่องมือวิเคราะห์ข้อมูลในการศึกษาวิจัยที่มีวัตถุประสงค์เพื่อทำนายเหตุการณ์ หรือประเมินความเสี่ยง จึงมีการประยุกต์ใช้ในงานวิจัยหลากหลายสาขา ทั้งสาขาทางการแพทย์ วิศวกรรมศาสตร์ นิเวศวิทยา เศรษฐศาสตร์ และสังคมศาสตร์

### ๒. แบบจำลองการถดถอยโลจิสติก

การวิเคราะห์การถดถอยโลจิสติกเป็นการประมาณค่าความน่าจะเป็นของการเกิดเหตุการณ์ มีตัวแบบมาจากฟังก์ชันโลจิสติก หากมีตัวแปรอิสระเพียงตัวเดียวฟังก์ชันโลจิสติก (Fig. 1a) จะแสดงความน่าจะเป็นของเหตุการณ์ที่มีรูปดังต่อไปนี้

$$Prob(event) = \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}} \quad \text{----- 1}$$

หรือ

$$Prob(event) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \quad \text{----- 2}$$

เมื่อ

$\beta_0$  คือ ค่าคงที่ (เมื่อไม่มีอิทธิพลจากตัวแปรอิสระใด)

$\beta_1$  คือ ค่าสัมประสิทธิ์ของตัวแปรอิสระ (ประมาณได้จากข้อมูลสังเกต)

$X$  คือ ตัวแปรอิสระ (ตัวแปรทำนาย)

$e$  คือ ลอการิทึมธรรมชาติ (มีค่าประมาณ 2.71828....)

ในกรณีที่มีตัวแปรทำนายหลายตัว ( $n$ ) ฟังก์ชันโลจิสติกจะแสดงในรูปดังนี้

$$Prob(event) = \frac{e^Z}{1 + e^Z} \quad \text{----- 3}$$

หรือ

$$Prob(event) = \frac{1}{1 + e^{-Z}} \quad \text{----- 4}$$

เมื่อ  $Z$  คือ Linear Combination ที่อยู่ในรูป

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad \text{----- 5}$$

เมื่อการเกิดเหตุการณ์แสดงเป็นความน่าจะเป็น ดังนั้นความน่าจะเป็นของการไม่เกิดเหตุการณ์ ก็คือ

$$Prob(no event) = 1 - Prob(event) \quad \text{----- 6}$$

ถ้าแทน  $Prob(event)$  ด้วย  $P_y$  และแทนค่า  $Z$  ตามสมการที่ 5 เข้าไปในสมการที่ 4 ก็จะได้รูปฟังก์ชัน (Fig. 1b) ดังนี้

$$P_y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad \text{----- 7}$$

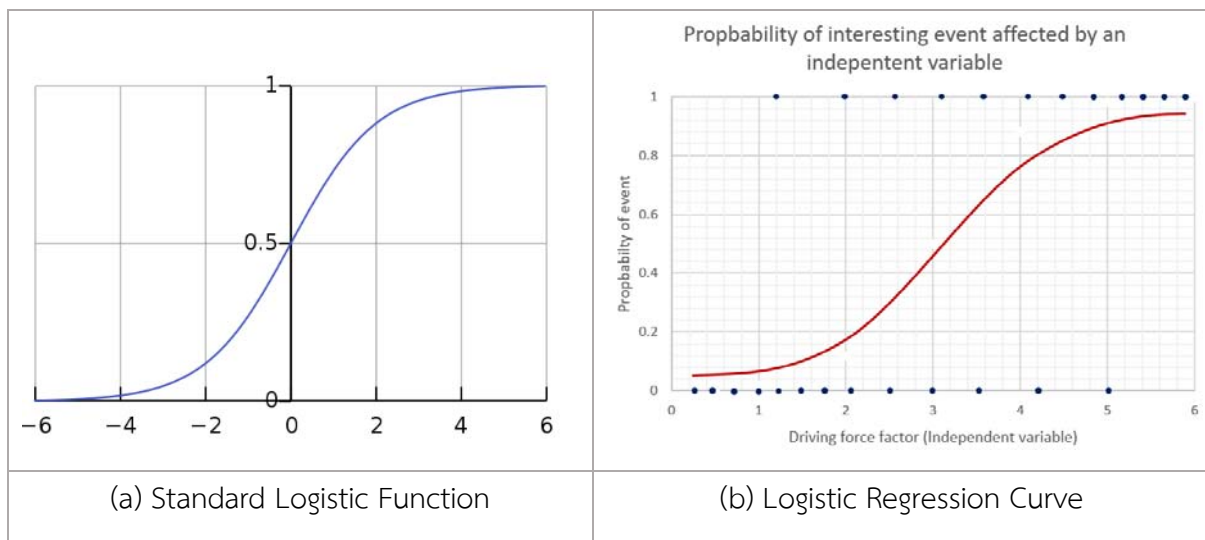


Fig. 1

ความน่าจะเป็นจะมีค่าอยู่ระหว่าง 0 ถึง 1 เช่น กล้าไม้พะยุงที่รากแก้วขาดมีความน่าจะเป็นที่จะตายเท่ากับ 0.25 หมายถึง กล้าไม้มีโอกาสตาย 25 เปอร์เซ็นต์ แต่ในการถดถอยโลจิสติกที่ใช้ทำนายเหตุการณ์ที่น่าสนใจที่เป็นไปได้สองทางเช่นนี้ จะใช้ค่า odds แทนความน่าจะเป็น โดย odds คืออัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์กับโอกาสที่จะไม่เกิดเหตุการณ์ ซึ่งจะมีค่าเป็น 0 หรือมากกว่า และให้ความหมายว่า เหตุการณ์ที่จะเกิดนั้นมีโอกาสเป็นกี่เท่าของเหตุการณ์ที่จะไม่เกิด (odds โดยทั่วไปก็คือการแสดงคู่ของตัวเลขในลักษณะอัตราต่อรอง) ดังนั้นจากตัวอย่างข้างต้นก็จะได้ว่า odds มีค่าเท่ากับ  $0.25/0.75$  ซึ่งเท่ากับ  $0.333$  เท่า (หรือ 1 ใน 3) นั่นคือโอกาสที่กล้าไม้พะยุงที่รากแก้วขาดมีโอกาส (ความเสี่ยง) ที่จะตายเป็น  $1/3$  เท่าของกล้าที่ไม่ตาย ในการวิเคราะห์การถดถอยโลจิสติก อัตราส่วนของโอกาสดังกล่าวแสดงในรูปต่อไปนี้

$$odds = \frac{P_y}{1-P_y} \quad \text{-----} 8$$

สมการที่ 8 แสดงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรทำนายที่ไม่ได้อยู่ในรูปสมการเชิงเส้น จึงต้องแปลงฟังก์ชันให้อยู่ในรูปเชิงเส้น (Logistic Transformation) จะได้ Log of odds ที่เรียกว่า logit ซึ่งก็คือแบบจำลองโลจิสติกที่แสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรทำนายที่อยู่ในรูปต่อไปนี้

$$\log(odds) = \ln\left(\frac{P_y}{1-P_y}\right) \quad \text{-----} 9$$

โดยสามารถแสดงให้อยู่ในรูปสมการถดถอยเชิงเส้นได้ดังนี้

$$\log(odds) = \text{logit} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad \text{-----} 10$$

การสร้างแบบจำลองหรือสมการถดถอยโลจิสติก กระทำโดยการประมาณค่าสัมประสิทธิ์จากชุดข้อมูลสังเกตที่เก็บวัดมาได้ โดยใช้วิธีการประมาณความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation; MLE) ซึ่งต่างจากการวิเคราะห์การถดถอยทั่วไปที่ใช้วิธีกำลังสองน้อยที่สุด (Least Square Method) ที่มีแนวคิดที่ต้องให้ได้ค่าส่วนเหลือ (Residuals) มีการกระจายแบบปกติ ดังนั้นการวิเคราะห์แบบ MLE เพื่อให้ได้ตัวแบบที่เหมาะสมที่สุดจึงใช้กระบวนการวิเคราะห์แบบเวียนซ้ำ (Iterative Process) โดยเริ่มจากการประมาณค่าสัมประสิทธิ์ของตัวแปรเพื่อให้ได้สมการตั้งต้น หลังจากนั้นก็ใช้สมการเพื่อทำนายค่าแล้วนำมาคำนวณซ้ำเพื่อหาค่าสัมประสิทธิ์ใหม่ที่ทำให้ได้ความน่าจะเป็นสูงสุดเพื่อให้สามารถทำนายค่าของตัวแปรตามได้ใกล้เคียงค่าของข้อมูลจริงมากที่สุด จากแบบจำลองโลจิสติก แสดงให้อยู่ในรูปสมการประมาณค่าได้ดังนี้

$$\hat{P}_y = \frac{e^{\beta_0 + \beta_1 + \beta_2 + \dots + \beta_n}}{1 + e^{\beta_0 + \beta_1 + \beta_2 + \dots + \beta_n}} \quad \text{-----} 11$$

หากทำการ antilog จะได้ odds ratio (OR) ที่มีรูปดังนี้

$$\frac{\widehat{P}_y}{1-\widehat{P}_y} = e^{\beta_0+\beta_1+\beta_2+\dots+\beta_n} \quad \text{-----} 12$$

ค่าสัมประสิทธิ์ ( $\beta_i$ ) ใน linear combination ในแบบจำลองโลจิสติกนั้นมักใช้ตีความหมายที่เข้าใจได้ในเชิงทิศทางมากกว่าการบอกขนาด เพราะไม่ได้สะท้อนสิ่งที่สนใจในการวิเคราะห์นั้นๆ เท่าไรนัก โดยทั่วไปจึงใช้การตีความหมายจากค่า odds ratio เพราะให้ความหมายที่สะท้อนสิ่งที่สนใจในการวิเคราะห์ได้ดีกว่า (รายละเอียดแสดงในหัวข้อที่ ๔)

### ๓. ความเหมาะสมของแบบจำลอง

แบบจำลองที่ได้จะต้องมีการทดสอบความเหมาะสมโดยพิจารณาความเป็นสารรูปสนิทธิ (Goodness of Fit) การทดสอบและประเมินความเหมาะสมของแบบจำลองโลจิสติกสามารถกระทำได้ด้วยเทคนิควิธีต่าง ๆ ดังต่อไปนี้

#### ๓.๑ Likelihood Ratio Test

การทดสอบอัตราส่วนภาวะน่าจะเป็น เป็นการพิจารณาค่า Log-Likelihood (LL) ของแบบจำลองโลจิสติกกรณีที่ไม่มีตัวแปรอิสระ (baseline) กับกรณีที่มีตัวแปรอิสระ โดยการเปรียบเทียบผลต่างระหว่างค่า -2LLs ที่มีการแจกแจงแบบ Chi-square หรืออาจเรียกว่า ค่าพฤติกรรมเบี่ยงเบน (Deviance) หากมีค่าสูงแสดงว่าการเพิ่มตัวแปรอิสระเข้าไปในแบบจำลองมีส่วนช่วยลดความคลาดเคลื่อนที่ยังไม่สามารถอธิบายได้ จะทำการทดสอบไคสแคว์เพื่อประเมินว่า ค่า -2LLs ระหว่างโมเดลฐานกับโมเดลที่มีตัวแปรอิสระมีความแตกต่างอย่างมีนัยสำคัญหรือไม่ ใน SPSS การทำสอบนี้จะปรากฏอยู่ในตารางผลการทดสอบ Omnibus Test of Model Coefficients

$$LL = \sum_{i=1}^N [Y_i \ln(\hat{Y}_i) + (1 - Y_i) \ln(1 - \hat{Y}_i)] \quad \text{-----} 13$$

$$\chi^2 = 2[(LL_{base\ model} - LL_{new\ model})] \quad \text{-----} 14$$

#### ๓.๒ Wald Statistics

เป็นการทดสอบด้วย t-statistic เพื่อพิจารณาว่าค่าสัมประสิทธิ์ของตัวแปรอิสระต่างจากค่าศูนย์หรือไม่ หาได้โดยนำค่าสัมประสิทธิ์ของตัวแปรอิสระหารด้วยค่า standard error (SE) ของตัวแปรตาม ซึ่งสูตรนี้เมื่อยกกำลังสองก็จะมีกระจายแบบ Chi-square ผลการทดสอบหากพบว่ามีนัยสำคัญก็แสดงว่าตัวแปรอิสระนั้น ๆ มีอิทธิพลต่อตัวแปรตาม

#### ๓.๓ R Square

เนื่องจากการประมาณค่าสัมประสิทธิ์แบบ MLE จะไม่มีสถิติ  $R^2$  ที่แท้จริงมีแต่ตัวที่เทียบเคียงได้เรียกว่า Pseudo  $R^2$  ซึ่งมีการพัฒนาจากหลายแนวคิด ดังเช่น

- ๑) Cox & Snell's  $R^2$  เป็นตัวสถิติที่บ่งบอกถึงความกลมกลืนของแบบจำลองในแง่ของการเปรียบเทียบคุณภาพของแบบจำลองที่สร้างขึ้นกับแบบจำลองฐานที่เหมาะสมน้อยที่สุด (คือแบบจำลองที่ยังไม่มีตัวแปรอิสระใดๆ) แต่มีข้อเสียคือ ค่าสูงสุดจะมีค่าได้ไม่ถึง 1 มีสูตร ดังนี้

$$R_{Cs}^2 = 1 - e^{\left[-\frac{2}{n}(LL(new) - LL(baseline))\right]} \quad \text{-----} 15$$

- ๒) Nagelkerke's  $R^2$  เป็นตัวสถิติที่ปรับวิธีการคำนวณจาก  $R^2$  ของ Cox and Snell เพื่อให้มีค่าได้ถึง 1 มีสูตรการหาดังนี้

$$R_N^2 = \frac{R_{Cs}^2}{1 - e^{\left[\frac{2(LL(baseline))}{n}\right]}} \quad \text{-----} 16$$

- ๓) McFadden's  $R^2$  เป็นตัวสถิติที่ถูกพัฒนามาจากแนวคิดที่ต้องการให้สามารถอธิบายความแปรปรวนของตัวแปรตามได้ และให้ความกลมกลืนในมุมมองของการเปรียบเทียบกับแบบจำลองฐานที่ไม่มีตัวแปรอิสระ มีสูตรการหาดังนี้

$$R_{Mc}^2 = 1 - \frac{LL(new)}{LL(baseline)} \quad \text{-----} 17$$

#### ๓.๔ Hosmer and Lemeshow Test

เป็นการทดสอบแบบจำลองโดยใช้สถิติ Chi-Square เพื่อดูว่าแบบจำลอง Logistic ที่สร้างขึ้นนั้นสามารถให้ค่าทำนายความน่าจะเป็นของการเกิดเหตุการณ์สอดคล้องกับการเกิดเหตุการณ์จริงที่วัดได้จากข้อมูลที่เก็บมาหรือไม่ ถ้าผลการทดสอบสถิติ Chi-Square ไม่นัยสำคัญ แสดงว่าแบบจำลอง Logistic ที่ได้นั้นยอมรับได้เพราะค่าที่พยากรณ์กับค่าจริงไม่แตกต่างกัน

#### ๔. การตีความหมายค่าสัมประสิทธิ์

แบบจำลองโลจิสติกเป็นการวัดโอกาสของลอการิทึมของ odds (ล็อกของอัตราส่วนความน่าจะเป็นของเหตุการณ์) ค่าสัมประสิทธิ์ของสมการจึงมีความหมายดังนี้

- ๑) ค่าสัมประสิทธิ์การถดถอย (Regression Coefficients) จะสะท้อนไปที่ค่า log odds ไม่ได้สะท้อนค่า  $Prob(event)$  โดยตรง จึงตีความหมายให้เป็นเข้าใจได้ยาก แต่สามารถบอกนัยของทิศทางอิทธิพลได้ (ในตารางผลการวิเคราะห์ของ SPSS แสดงด้วยค่า B)
- ๒) เอ็กโปเนนเชียลของค่าสัมประสิทธิ์ (Exponential Coefficients) ในตารางผลการวิเคราะห์ของ SPSS จะแสดงด้วยค่า  $Exp(B)$  เป็นการสะท้อนค่าของ odds ด้วย odds ratio (OR) ทำให้การตีความหมายเป็นที่เข้าใจได้ง่าย อัตราส่วน odds จะบอกให้ทราบว่า odds ของเหตุการณ์หนึ่งจะสัมพันธ์กับอีกเหตุการณ์หนึ่งอย่างไรทำให้สามารถเปรียบเทียบชุดข้อมูลสองกลุ่มที่มีลักษณะเฉพาะบางประการหรือมีทรีตเมนต์ต่างกันว่าจะแสดงผลต่อตัวแปรตามที่น่าสนใจได้มากน้อยเป็นกี่เท่าของกันและกัน แสดงได้ดังสมการต่อไปนี้

$$OR = \frac{\text{odds of a certain event}}{\text{odds of the another event}} \quad \text{-----} \quad 18$$

ในการวิเคราะห์การถดถอยโลจิสติก จะได้สมการพยากรณ์ในรูปแบบการถดถอยแบบพหุ (กรณีมีตัวแปรอิสระหลายตัว) (ดังรูปสมการที่ 10) การอธิบายตัวแปรอิสระแต่ละตัว ( $X_i$ ) สามารถตีความหมายทำนองเดียวกับสมการถดถอยแบบปกติ กล่าวคือ ค่าสัมประสิทธิ์แสดงทั้งขนาดและทิศทาง (เครื่องหมาย  $+/-$ ) สามารถระบุเป็นเปอร์เซ็นต์อิทธิพลของตัวแปรอิสระตัวนั้น ๆ ที่มีต่อตัวแปรตาม เพียงแต่ต้องตระหนักว่า ตัวแปรตามในที่นี้คือ logit หรือ ล็อกของอัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์กับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ ดังนั้น ถ้าต้องการทราบอิทธิพลของตัวแปรอิสระที่มีต่อตัวแปรตามคือโอกาสเกิดเหตุการณ์ที่สนใจ ให้พิจารณาที่ค่า  $e^{\beta_i}$  (หรือ  $\exp(\beta_i)$ ) ซึ่งจะแสดงแสดงการเปลี่ยนแปลงของ odds โดยมีความหมายว่า เมื่อ  $X_i$  เพิ่มขึ้น 1 หน่วย จะทำให้ค่า odds เปลี่ยนแปลงไป  $e^{\beta_i}$  เท่าจากค่าอัตราส่วนเดิม โดยทิศทางการเปลี่ยนแปลงนั้นก็ให้พิจารณาที่เครื่องหมาย บวก-ลบ

ในการวิเคราะห์การถดถอยโลจิสติก พิจารณาที่ odds แทน Probability จึงมีความหมายในการทำนองอัตราต่อรอง หรือความเป็นต่อที่จะ เกิด กับ ไม่เกิด คือต้องการคำตอบเป็น dichotomous คือมีแค่สองทางเลือก ความน่าจะเป็นมีค่าอยู่ระหว่าง 0 ถึง 1 เรามักตัดค่า (cut value) ที่ 0.5 เพื่อระบุเหตุการณ์ ดังนั้น ค่า 0 – 0.5 ก็จะถูกตีความว่าเป็นกรณี ไม่เกิดเหตุการณ์ ถ้าเกิน 0.5 ขึ้นไปก็หมายถึงการเกิดเหตุการณ์ที่สนใจ

## ๕. การประเมินประสิทธิภาพของแบบจำลอง

การประเมินประสิทธิภาพของสมการในการให้ผลการทำนายเป็นสิ่งจำเป็นถ้าเราต้องการนำแบบจำลองที่ได้นั้นไปทำนายเหตุการณ์ที่มีค่าปัจจัย (ตัวแปรอิสระ) แปรผันแตกต่างออกไปจากค่าที่ใช้สร้างแบบจำลอง ควรที่จะต้องมีการทดสอบการทำนายของแบบจำลองว่าให้ความถูกต้องเพียงใด โดยใช้ข้อมูลการสำรวจเหตุการณ์จริงจำนวนหนึ่งมาเปรียบเทียบกับสิ่งที่แบบจำลองทำนายได้ ซึ่งในแบบจำลองถดถอยโลจิสติก

เป็นการทำนายเหตุการณ์ว่าจะเกิดหรือไม่เกิด ดังนั้น การทดสอบก็จะเป็นการหาข้อมูลเหตุการณ์อ้างอิงมาเพื่อเปรียบเทียบ ข้อมูลตัวอย่างนี้จะสามารถแบ่งได้เป็นสองประชากรคือ กลุ่มที่เกิดเหตุการณ์ กับ กลุ่มที่ไม่เกิดเหตุการณ์ ซึ่งในทางปฏิบัติการจำแนกสองประชากรนี้โดยสมบูรณ์เป็นสิ่งที่กระทำได้ยาก ผลของการทดสอบหากนำมาแสดงเป็นกราฟการกระจายตัวของข้อมูลจะมีการซ้อนทับกันบางส่วน ดังรูป (Fig. 2)

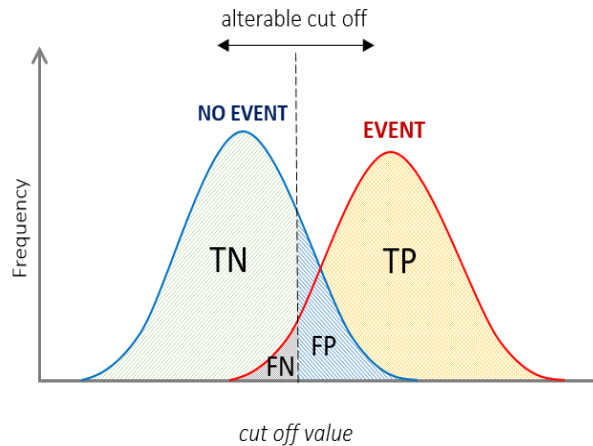


Fig. 2 The distribution of the test results with a cut off value

แบบจำลองที่มีตัวแปรอิสระที่เราจัดกลุ่มไม่ว่าจะเป็นลำดับหรือไม่ก็ตามมักจะมีผลต่อประสิทธิภาพของการทำนายเหตุการณ์ ในทำนองเดียวกันจุดที่เรากำหนดเป็นค่าเกณฑ์หรือค่าที่ตัดเหตุการณ์ (cut off) เพื่อแยกความแตกต่างกันระหว่างสองประชากรดังกล่าว บางครั้งสามารถส่งผลให้การทำนายเหตุการณ์ที่เกิดขึ้นได้ถูกต้องซึ่งถือว่าผลการทำนายออกมาเป็นบวก (หรือเชิงบวก) แต่บางครั้งก็มีผลการทำนายเป็นลบได้ ในทำนองเดียวกันกรณีที่ไม่เกิดเหตุการณ์นั้นเกิดขึ้นจริงผลการทำนายของแบบจำลองก็อาจให้ผลเป็นบวกหรือเป็นลบก็ได้เช่นกัน ดังนั้นในการทดสอบ จึงมีตัวสถิติหลายตัวที่ใช้อธิบายประสิทธิภาพของแบบจำลอง โดยต้องนำข้อมูลผลการทำสอบมาสร้างเป็นเมตริกเพื่อการประเมินในรูปต่อไปนี้

		ค่าอ้างอิง (เป็นจริง)		รวม
		เกิดเหตุการณ์	ไม่เกิดเหตุการณ์	
ผลการทำนาย จากแบบจำลอง	เกิดเหตุการณ์	True Positive (TP)	False Positive (FP)	TP + FP
	ไม่เกิดเหตุการณ์	False Negative (FN)	True Negative (TN)	FN + TN
	รวม	TP + FN	FP + TN	TP+FP+FN+TN

ค่าจากเมตริกสามารถนำมาคำนวณหาความไว (sensitivity) ความจำเพาะ (specificity) ค่าความถูกต้อง (Accuracy) ค่าทำนายเป็นบวก (Positive Predictive Value) ค่าทำนายเป็นลบ (Negative Predictive Value) อัตราส่วนความน่าจะเป็นเชิงบวก (Positive Likelihood Ratio) และ อัตราส่วนความน่าจะเป็นเชิงลบ (Negative Likelihood Ratio) ได้ดังสูตรการหา ต่อไปนี้

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Specificity} = \text{TN} / (\text{FP} + \text{TN})$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

$$\text{Positive Predictive Value} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Negative Predictive Value} = \text{TN} / (\text{FN} + \text{TN})$$

$$\text{Positive Likelihood Ratio} = \text{Sensitivity} / (1 - \text{Specificity})$$

$$\text{Negative Likelihood Ratio} = (1 - \text{Sensitivity}) / \text{Specificity}$$

โดย

- Sensitivity คือ ความน่าจะเป็นที่ผลการทำนายจะเป็นบวกเมื่อเหตุการณ์ที่สนใจนั้นเกิดขึ้นจริง (TP แสดงเป็นเปอร์เซ็นต์)
- Specificity คือ ความน่าจะเป็นที่ผลการทำนายจะเป็นลบเมื่อเหตุการณ์ที่สนใจนั้นไม่เกิดขึ้น (TN แสดงเป็นเปอร์เซ็นต์)
- Positive Predictive Value คือ ความน่าจะเป็นที่เกิดเหตุการณ์ที่สนใจเมื่อผลทดสอบออกมาเป็นบวก (แสดงเป็นเปอร์เซ็นต์)
- Negative Predictive Value คือ ความน่าจะเป็นที่ไม่เกิดเหตุการณ์ที่สนใจเมื่อผลทดสอบออกมาเป็นลบ (แสดงเป็นเปอร์เซ็นต์)
- Positive Likelihood Ratio คือ อัตราส่วนระหว่างความน่าจะเป็นของผลการทดสอบที่เป็นบวกเมื่อเกิดเหตุการณ์ที่สนใจ กับ ความน่าจะเป็นของผลการทดสอบที่เป็นบวกเมื่อไม่เกิดเหตุการณ์ที่สนใจ
- Negative Likelihood Ratio คือ อัตราส่วนระหว่างความน่าจะเป็นของผลการทดสอบที่เป็นลบเมื่อเกิดเหตุการณ์ที่สนใจ กับ ความน่าจะเป็นของผลการทดสอบที่เป็นลบเมื่อไม่เกิดเหตุการณ์ที่สนใจ

สร้างกราฟ Receiver Operator Characteristics (ROC) จากค่าความไวและค่าความจำเพาะจะได้ดังภาพ (Fig. 3) ROC เป็นตัวประเมินที่ช่วยให้เราสามารถเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองได้ ROC สร้างจากการพล็อตค่าที่มีความถูกต้องเชิงบวก (true positive rate) กับค่าที่ความถูกต้องเชิงลบ (true negative rate) ณ ค่า threshold ที่กำหนด (cut off value) โดย ROC จะมีค่าพื้นที่ใต้เส้นกราฟ (AUC) อยู่ระหว่าง 0 – 1 โดยหากค่าเข้าใกล้ 1 คือ ดี โดยทั่วไปในการประยุกต์ใช้งานด้านภูมิสารสนเทศค่า AUC ตั้งแต่ 0.7 ขึ้นไปถือว่าแบบจำลองนั้นมีประสิทธิภาพที่ยอมรับได้



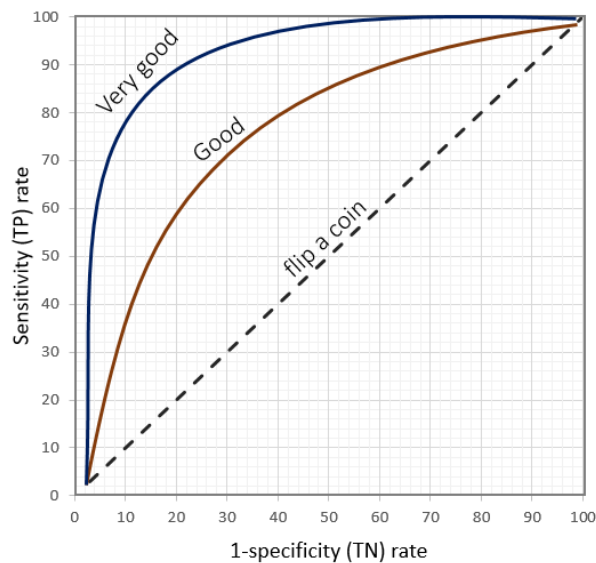


Fig. 3 Receiver Operator Characteristics Curve

### ๕. การใช้การถดถอยโลจิสติกในงานเชิงพื้นที่

ในงานวิจัยที่มีการประมาณค่าแบบเชิงพื้นที่ที่จะใช้แบบจำลองถดถอยโลจิสติกที่อยู่ในรูปสมการ logit ที่ประมาณค่าสัมประสิทธิ์ของพารามิเตอร์ของสมการได้แล้ว โดยเป็นการประยุกต์ที่นอกจากจะใช้อธิบายตัวแปร (ตัวปัจจัยที่มีผลต่อเหตุการณ์ที่สนใจ) แล้ว ยังต้องการใช้เป็นแบบจำลองเพื่อทำนายค่าตัวแปรตามเมื่อค่าตัวแปรอิสระเปลี่ยนแปลงไป ในการประยุกต์เชิงพื้นที่จึงจำเป็นต้องมีข้อมูลเชิงพื้นที่ของตัวแปรอิสระทุกตัวสำหรับพื้นที่ที่สนใจ โดยแต่ละตัวต้องมีค่าหรือมีการจัดชั้นให้สอดคล้องกับสเกลของตัวแปรอิสระในแบบจำลองนั้น ซึ่งอาจเป็นค่าจริง ค่าจัดชั้น หรือค่าคัมมัลหลังจากนั้นก็ใช้สมการที่ 3 หรือ 4 หรือ 7 เพื่อคำนวณค่าโอกาสของการเกิดเหตุการณ์ ซึ่งจะได้ค่าอยู่ระหว่าง 0 ถึง 1 หลังจากนั้นก็ทำการกำหนดอันตรภาคชั้นเพื่อแบ่งช่วงค่าออกเป็นระดับซึ่งอาจอิง Likert Scale เพื่อให้มีความหมายในเชิงคุณภาพ จากต่ำสุดไปสูงสุด เพื่อประโยชน์ในการจัดทำเป็นแผนที่ เช่น การทำแผนที่เสี่ยงภัย แผนที่การกระจายชนิดพันธุ์ เป็นต้น

ตัวอย่างการศึกษาเพื่อวิเคราะห์หาพื้นที่เสี่ยงต่อการถูกทำลายป่าในพื้นที่แห่งหนึ่ง โดยมีปัจจัยคือ ระยะห่างจากถนน (X1) ทิศด้านลาด (X2) ความลาดชัน (X3) จำนวนประชากรของหมู่บ้านในและรอบพื้นที่ป่า (X4) โดยผู้วิจัยได้ใช้การจัดชั้นแบบมีอันดับของตัวแปรแต่ละตัว ได้ผลการวิเคราะห์ดังตารางต่อไปนี้

Var.	B	S.E.	Wald	df	Sig	Exp(B)	Lower	Upper
distRoad	-0.360	0.079	20.574	1	0.000	0.698	0.597	0.815
aspect	-0.119	0.052	5.176	1	0.230	0.888	0.801	0.984
slope	-0.295	0.071	14.063	1	0.000	0.767	0.668	0.881
population	-0.592	0.132	20.153	1	0.000	0.553	0.427	0.716
constant	3.007	0.475	40.038	1	0.000	20.236		

ได้สมการทำนายดังนี้

$$P_y = \frac{1}{1 + e^{-(3.007 - 0.360 \text{distRoad} - 0.119 \text{aspect} - 0.295 \text{slope} - 0.592 \text{population})}}$$

ผู้วิจัยได้จัดทำชั้นข้อมูลแผนที่เชิงภาพในระบบสารสนเทศภูมิศาสตร์ของทุกตัวแปร โดยจัดค่าข้อมูลของแต่ละตัวแปรให้มีค่าตามลำดับชั้นที่กำหนดที่ความละเอียดของจุดภาพที่ 10x10 เมตร ดังนั้น ถ้าจุดภาพใดอ่านค่าของตัวแปรได้ X1, X2, X3, X4 เท่ากับ 1, 3, 3, 2 เมื่อแทนค่าในสมการทำนายก็จะได้  $P_y$  เท่ากับ 0.930

อย่างไรก็ตามต้องระมัดระวังในเรื่องของความแม่นยำของข้อมูลผลลัพธ์ที่ได้จัดชั้นความเสี่ยง เพราะแบบจำลองการถดถอยโลจิสติกที่ใช้ นั้น อาจมีอำนาจในการพยากรณ์ต่ำอันเนื่องจากตัวแปรอิสระไม่เพียงพอ หรือมีสเกลของค่าไม่เหมาะสม ดังนั้นการประเมินประสิทธิภาพของสมการในการทำนายเป็นสิ่งจำเป็น