

การวิเคราะห์การถดถอยโลจิสติกแบบไบนารีสำหรับการวิจัยทางสังคมศาสตร์

Binary logistic regression analysis for social science research

อรรถัย เจริญสิทธิ์

สาขาวิทยาการคอมพิวเตอร์ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยเอเชียอาคเนย์

บทคัดย่อ : บทความนี้มีวัตถุประสงค์เพื่อนำเสนอแนวคิด สมการ และค่าสถิติที่สำคัญของการวิเคราะห์การถดถอยโลจิสติกแบบไบนารี ซึ่งการวิเคราะห์การถดถอยโลจิสติกมีวัตถุประสงค์ เพื่อทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ในการวิจัยทางสังคมศาสตร์สามารถนำการวิเคราะห์การถดถอยโลจิสติกมาประยุกต์เพื่อศึกษาถึงปัจจัยที่ส่งผลกระทบต่อ การเกิดเหตุการณ์หรือพฤติกรรมที่สนใจ ซึ่งนอกจากจะทราบถึงความน่าจะเป็นของการเกิดเหตุการณ์หรือพฤติกรรมที่สนใจ แล้ว สมการพยากรณ์ยังทำให้ทราบถึงอิทธิพลในเชิงปริมาณของปัจจัยต่างๆ ได้ ซึ่งสามารถใช้เป็นแนวทางในการเพิ่ม หรือลดโอกาสของการเกิดเหตุการณ์หรือพฤติกรรมที่สนใจได้ ดังเช่น ในการศึกษาปัจจัยที่ส่งผลต่อพฤติกรรม การสูบบุหรี่ของนักเรียนมัธยมศึกษาตอนปลายของBoamah (2011) ที่พบว่า พฤติกรรมการไม่มีกิจกรรมทางร่างกาย และ ความรู้สึกเศร้าหรือสิ้นหวัง เป็นปัจจัยที่เพิ่มความน่าจะเป็นที่จะมีพฤติกรรม การสูบบุหรี่ ซึ่งหากลดพฤติกรรมดังกล่าวลง ความน่าจะเป็นที่จะมีพฤติกรรม การสูบบุหรี่ก็จะลดลง

คำสำคัญ : การถดถอยโลจิสติกแบบไบนารี, การวิจัยทางสังคมศาสตร์

Abstract : The objective of this article is to present concept, model and statistics of binary logistic regression analysis. Logistic regression purposed to forecast the probability of interest event. In social science research, logistic regression applied to study the factors that affect the occurrence of events or interested behavior. This will determine the probability of the occurrence of events or behavior that already. The forecast also indicates the influence of the quantitative factors, which can be adopted to increase or decrease the likelihood of the occurrence of events or interested behavior such as the prediction of high school students' cigarette smoking behavior, the feeling sad or hopeless, and physically inactive behavior are increased high school students' cigarette smoking behavior.

Keywords : Binary logistic regression, Social science research

1. บทนำ

การวิเคราะห์การถดถอยโลจิสติก (Logistic regression analysis) เป็นการวิเคราะห์การถดถอยแบบหนึ่งที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพ (Qualitative variable) การวิเคราะห์การถดถอยโลจิสติกแบ่งออกเป็น 2 ประเภท คือ การวิเคราะห์การถดถอยโลจิสติกแบบไบนารี (Binary logistic regression analysis) คือการวิเคราะห์การถดถอยที่ตัวแปรตามมี 2 ค่า คือ ไม่เกิดเหตุการณ์ ($y = 0$) หรือเกิดเหตุการณ์ ($y = 1$) และการวิเคราะห์การถดถอยโลจิสติกพหุกลุ่ม (Multinomial logistic regression analysis) เป็นการวิเคราะห์ที่ตัวแปรตามที่มีค่ามากกว่า 2 กลุ่ม (Polychromous variable) การวิเคราะห์การถดถอยโลจิสติกมีวัตถุประสงค์ เพื่อทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจ จึงทำให้ถูกประยุกต์ใช้ในหลายสาขาวิชา เช่น ทางด้านการแพทย์ใช้สมการโลจิสติกทำนายโอกาสของการเกิดโรคต่างๆ ทางด้านธุรกิจใช้ในการพยากรณ์ภาวะทางด้านเศรษฐกิจ เช่น โอกาสของการเกิดหนี้เสีย เช่นเดียวกับทางสังคมศาสตร์ก็มีการประยุกต์ใช้การวิเคราะห์การถดถอยโลจิสติกในหลายๆ ด้าน เช่น การทำนายระดับความเครียดในสถานการณ์ต่างๆ การเกิดหรือไม่เกิดพฤติกรรมบางอย่างที่สนใจ เห็นได้ว่าการวิเคราะห์การถดถอยโลจิสติกสามารถนำมาประยุกต์ใช้ในหลายศาสตร์ ในบทความนี้จะขอกล่าวถึงการวิเคราะห์การถดถอยโลจิสติกแบบไบนารีสำหรับการวิจัยทางสังคมศาสตร์

วัตถุประสงค์การวิเคราะห์การถดถอยโลจิสติก

วัตถุประสงค์ของการวิเคราะห์การถดถอยโลจิสติก มีดังนี้

1. เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ พร้อมทั้งศึกษาระดับความสัมพันธ์ของตัวแปรตามและตัวแปรอิสระ

2. เพื่อพยากรณ์โอกาสของการเกิดเหตุการณ์ที่สนใจ โดยใช้สมการที่สร้างขึ้นด้วยปัจจัยหรือตัวแปรที่มีผลต่อตัวแปรตาม ทำให้สามารถพยากรณ์โอกาสที่จะเกิดเหตุการณ์ได้

ข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยโลจิสติก

การวิเคราะห์การถดถอยโลจิสติกมีข้อตกลงเบื้องต้น ดังนี้

1. ตัวแปรอิสระ X 's อาจจะเป็นข้อมูลชนิด Dichotomous (มีค่าได้ 2 ค่า) หรือเป็นสเกลอันดับ (Interval Scale) และสเกลอัตราส่วน (Ratio Scale)
2. ค่าคาดหวังของค่าความคลาดเคลื่อนเป็นศูนย์หรือ $E(e) = 0$
3. e_i และ e_j เป็นอิสระต่อกัน
4. e_i และ X_i เป็นอิสระต่อกัน
5. ตัวแปรอิสระไม่ควรมีความสัมพันธ์กันหรือไม่ควรเกิดปัญหา Multicollinearity ซึ่งหากพิจารณาจากเกณฑ์ของ บริการร้านค้าปลีก (1996) ค่าสหสัมพันธ์ (r) ควรมีค่าไม่เกิน 0.8

1. โมเดลการวิเคราะห์การถดถอยโลจิสติกแบบไบนารี (Binary logistic regression analysis model)

โมเดลการวิเคราะห์การถดถอย คือ

$$y_i = x_i' \beta + \varepsilon_i \quad (1.1)$$

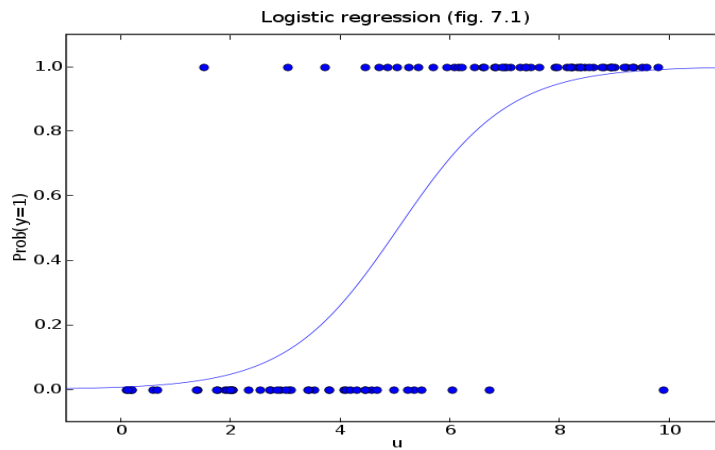
เมื่อ $x_i' = [1, x_{i1}, x_{i2}, \dots, x_{ik}]$

$$\beta' = [\beta_0, \beta_1, \dots, \beta_k]$$

ε_i คือ ความคลาดเคลื่อน

สำหรับการวิเคราะห์โลจิสติกที่เป็นแบบไบนารี ตัวแปรตาม (y) มีค่า คือ 0 และ 1 ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรทำนายจึงไม่อยู่ในรูป

เชิงเส้น ซึ่งความสัมพันธ์ของตัวแปรตามและตัวแปรทำนายในการวิเคราะห์โลจิสติกจะอยู่ในรูปคล้ายตัว S ดังรูปที่ 1



รูปที่ 1 ความสัมพันธ์ของตัวแปรตามและตัวแปรทำนายในการวิเคราะห์โลจิสติก

ที่มา : Douglas C. (2006) Introduction to Regression Analysis

ดังนั้น

$$P(y) = \frac{e^{b_0+b_1x}}{1+e^{b_0+b_1x}} \quad (1.2)$$

สำหรับการนิยามตัวแปรทำนายมากกว่า 1 ตัว จะได้ฟังก์ชันดังนี้

$$P(y) = \frac{e^{b_0+b_1x_1+\dots+b_px_p}}{1+e^{b_0+b_1x_1+\dots+b_px_p}} \quad (1.3)$$

เมื่อ

$P(y)$ = ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

$Q(y)$ = ความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ โดยที่ $Q(y) = 1 - P(y)$ โดยที่

$P(y) \geq 0.5$ สรุปว่าเกิดเหตุการณ์ที่สนใจ

$P(y) < 0.5$ สรุปว่าไม่เกิดเหตุการณ์ที่สนใจ

โดยค่า 0.5 นี้ เป็นค่าความน่าจะเป็นที่จะใช้เป็น Cutting Score ซึ่งอาจกำหนดให้เป็นค่า 0.7 หรือ 0.4 หรือค่าอื่นได้ โดยทั่วไปมักนิยมใช้ 0.5 (กลยา นิษฐ์ บัญชา, 2552)

จากความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรทำนายที่ไม่เป็นเชิงเส้น จึงทำการปรับตัวอยู่ในรูปเชิงเส้น โดยเขียนให้อยู่ในรูปของ odds หรือ odd ratio ซึ่งหมายถึงอัตราส่วนระหว่างความน่าจะเป็นของการเกิดเหตุการณ์กับความน่าจะเป็นของการไม่เกิดเหตุการณ์ ดังนี้

$$odd = \frac{P(y)}{Q(y)} \quad (1.4)$$

ค่าของ odds แสดงถึงโอกาสที่จะเกิดเหตุการณ์ที่สนใจ เป็นกึ่งแรกของโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ การเขียนโมเดลโลจิสติกจะอยู่ในรูป $\log(odds)$ เรียกว่า logit หรือ logistic response function ซึ่งเขียนในรูปสมการดังนี้

$$\log\left(\frac{P(y)}{Q(y)}\right) = \log\left(\frac{P(y)}{1-P(y)}\right) = b_0 + b_1x_1 + \dots + b_px_p \quad (1.5)$$

ดังนั้น

$$\log(odds) \text{ หรือ } logit = b_0 + b_1x_1 + \dots + b_px_p \quad (1.6)$$

ในการประมาณค่าสัมประสิทธิ์การถดถอย (b.) ของของสมการถดถอยโลจิสติก จะใช้วิธีภาวะน่าจะเป็นสูงสุด หรือ Maximum likelihood โดยใช้วิธีการคำนวณซ้ำๆ (Iteration) โดยเริ่มต้นจากการประมาณค่าสัมประสิทธิ์ในสมการถดถอยโลจิสติก เพื่อให้สามารถแก้สมการได้ แล้วพิจารณาผลการทำนายเพื่อนำมาประมาณค่าสัมประสิทธิ์ใหม่ที่จะทำให้เกิดความน่าจะเป็นสูงสุดเพื่อที่จะสามารถทำนายค่าของตัวแปรตามได้ถูกต้องใกล้เคียงกับข้อมูลจริงมากที่สุด

2. การตรวจสอบความเหมาะสมของสมการถดถอยโลจิสติก

การตรวจสอบความเหมาะสมของสมการถดถอยโลจิสติก มีการตรวจสอบหลายวิธี ดังนี้

2.1. พิจารณาค่าความเป็นไปได้ (likelihood value) การวิเคราะห์การถดถอยแบบปกติ จะประเมินคุณภาพของโมเดลหรือสมการพยากรณ์จากค่า Residual Sum of Squares แต่สำหรับการวิเคราะห์การถดถอยโลจิสติก จะประเมินจากค่า Log-Likelihood (LL) ซึ่งเทียบเคียงได้กับค่า Residuals นั่นก็คือ ยิ่งค่า LL น้อยยิ่งแสดงว่าสมการโลจิสติกที่สร้างขึ้นมีคุณภาพหรือกลมกลืนกับข้อมูลดี โดยจะศึกษาจากค่า 2LL (-2 log likelihood) ซึ่งเป็นค่ามาจาก log likelihood ที่คูณด้วย -2 เพื่อต้องการให้ค่าที่ได้มีการแจกแจง χ^2 สำหรับการทดสอบนัยสำคัญทางสถิติการพิจารณาค่า 2LL ถ้ามีค่าต่ำ สมการโลจิสติกมีความเหมาะสม ในการทดสอบนัยสำคัญความเหมาะสมของสมการโลจิสติก ใช้สถิติ χ^2 ที่ $df = p$ (จำนวนตัวแปรทำนาย) ทดสอบสมมติฐาน ดังนี้

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \beta_i \neq 0 \text{ อย่างน้อย 1 ค่า; } i = 1, 2, \dots, p$$

โดยที่

$$\text{LogLikelihood} = \sum_{i=1}^N [y_i \ln(\hat{y}_i) + (1 - y_i) \ln((1 - \hat{y}_i))] \dots \dots \dots (2.1.1)$$

และ

$$\chi^2 = 2[LL(B) - LL(0)] \dots \dots \dots (2.1.2)$$

การทดสอบถ่วงค่าสถิติ χ^2 มีนัยสำคัญ หรือยอมรับ H_1 แสดงว่า กลุ่มของตัวแปรทำนายสามารถร่วมกันทำนายโอกาสของการเกิดเหตุการณ์ที่สนใจ ($y = 1$) ได้ด้วยความเชื่อมั่น $(1 - \alpha) \times 100\%$

2.2 พิจารณาสถิติทดสอบความเหมาะสม

ของ Hosmer and Lemeshow จะใช้ทดสอบความเหมาะสมของโมเดล โดยทดสอบสมมติฐานดังนี้

$$H_0 : \text{โมเดลเหมาะสม}$$

$$H_1 : \text{โมเดลไม่เหมาะสม}$$

ในการทดสอบ หากค่าสถิติ χ^2 ไม่มีนัยสำคัญหรือยอมรับ H_0 แสดงว่า โมเดลมีความเหมาะสม

3. การทดสอบสัมประสิทธิ์การถดถอยโลจิสติก

การทดสอบของสัมประสิทธิ์ถดถอยโลจิสติก เป็นการทดสอบว่าตัวแปรทำนายมีผลต่อการเปลี่ยนแปลงของตัวแปรเกณฑ์หรือไม่ ทดสอบด้วยสถิติทดสอบ 2 ตัว ได้แก่

3.1 สถิติทดสอบของวอลด์ (Wald statistic)

ค่าสถิติวอลด์ (Wald statistic) เป็นการทดสอบสมมติฐาน ดังนี้

$$H_0 : \beta_i = 0; i = 1, 2, \dots, p \text{ (ตัวแปรทำนายไม่มีผลต่อการเปลี่ยนแปลงของ odds ratio)}$$

$$H_1 : \beta_i \neq 0 \text{ อย่างน้อย 1 ค่า; } i = 1, 2, \dots, p$$

Wald หรือ

$$w = \left(\frac{b_i}{SE(b_i)} \right)^2 \dots \dots \dots (3.1.1)$$

ในการทดสอบถ้าผลการทดสอบไม่มีนัยสำคัญ หรือยอมรับ H_0 แสดงว่า ตัวแปรทำนาย i ไม่มีผลต่อการเปลี่ยนแปลง odds ratio ดังนั้นตัวแปรทำนาย i จึงไม่มีผลต่อความน่าจะเป็นของการเกิดเหตุการณ์นั้น และหากมีนัยสำคัญหรือยอมรับ H_1 โดยสถิติทดสอบของวอลด์ (Wald test) จะมีการแจกแจงแบบ χ^2 และมี $df = 1$

ในงานวิจัยของศศิธร แทนรินทร์ (2554) ที่ศึกษาปัจจัยที่ส่งผลต่อการตัดสินใจศึกษาต่อในมหาวิทยาลัยราชภัฏชัยภูมิ ได้แสดงค่าสัมประสิทธิ์การถอยถอยโลจิสติกและสถิติวอลด์ (Wald test) ของปัจจัยที่สามารถพยากรณ์การตัดสินใจศึกษาต่อในมหาวิทยาลัยราชภัฏชัยภูมิ ดังนี้

ตัวแปรพยากรณ์	B	S.E	Wald	p-value	EXP(B)
1. การแนะนำของบิดา มารดา (AD_{FM})	.666	.174	14.656**	.000	1.945
2. การแนะนำของเพื่อน (AD_{FR})	.421	.162	6.739**	.009	1.523
3. ชื่อเสียงของมหาวิทยาลัย (IN_{WE})	.851	.312	7.453**	.006	2.344
4. คุณภาพของคณาจารย์ (IN_{TEA})	-.756	.275	7.566**	.002	0.469
5. คุณภาพการสอน (IN_{INS})	.836	.275	9.235**	.000	2.308
6. ทำเล/สถานที่ตั้งของมหาวิทยาลัย (ENV_{LO})	.932	.239	15.203**	.000	2.540

$R^2 = .441$, Constant = -10.531

สามารถเขียนสมการทำนายได้ดังนี้

$$\text{Logit}(y) = -10.53 + 0.67 AD_{FM} + 0.42 AD_{FR} + 0.85 IN_{WE} - 0.76 IN_{TEA} + 0.84 IN_{INS} + 0.93 ENV_{LO} \quad (3.1.2)$$

จากสมการดังกล่าวสามารถพยากรณ์ค่าความน่าจะเป็นของการศึกษาต่อในมหาวิทยาลัยราชภัฏชัยภูมิ ดังนี้

$$P(\text{ศึกษาต่อ}) = \frac{e^{-10.53 + 0.67 AD_{FM} + 0.42 AD_{FR} + 0.85 IN_{WE} - 0.76 IN_{TEA} + 0.84 IN_{INS} + 0.93 ENV_{LO}}}{1 + e^{-10.53 + 0.67 AD_{FM} + 0.42 AD_{FR} + 0.85 IN_{WE} - 0.76 IN_{TEA} + 0.84 IN_{INS} + 0.93 ENV_{LO}}} \quad (3.1.3)$$

$P(\text{ศึกษาต่อ}) < 0.5$ จะสรุปว่าไม่ต้องการศึกษาต่อในมหาวิทยาลัยราชภัฏชัยภูมิ

$P(\text{ศึกษาต่อ}) \geq 0.5$ จะสรุปว่าต้องการศึกษาต่อในมหาวิทยาลัยราชภัฏชัยภูมิ

ในงานวิจัยนี้ได้นำสมการถอยถอยโลจิสติกมาประยุกต์ใช้ในการทำนายโอกาสที่จะศึกษาต่อที่มหาวิทยาลัยราชภัฏชัยภูมิ โดยผู้วิจัยศึกษาตัวแปรตามคือการเลือก/ไม่เลือก ศึกษาต่อในมหาวิทยาลัยราชภัฏชัยภูมิ และตัวแปรพยากรณ์ซึ่งเป็นตัวแปรทางสังคมศาสตร์ คือ ปัจจัยที่ส่งผลต่อการตัดสินใจศึกษาต่อในมหาวิทยาลัยราชภัฏชัยภูมิทั้งหมด 6 ด้าน โดยตัว

แปรนี้มีมาตราส่วนประมาณค่า (Rating Scale) แบ่งเป็น 5 ระดับ ตามแนวคิดของลิเคิร์ท (Likert) จำนวน 35 ข้อ ซึ่งสมการทำนายที่ได้นี้ กลุ่มของตัวแปรพยากรณ์สามารถอธิบายตัวแปรตามได้ 44.1% ($R^2 = 0.441$) ซึ่งในการวิจัยทางสังคมศาสตร์ถือว่าการทำนายพฤติกรรมมนุษย์ โดยทั่วไปค่า R^2 ที่ได้มักจะมีค่าต่ำกว่า 50% เนื่องด้วยพฤติกรรมมนุษย์เป็นสิ่งที่ยากจะคาดเดาได้

เมื่อเทียบกับกระบวนการทางกายภาพ (สุวดี น้าพา
เจริญ และชลทิชา จำรัสพร, 2558)

3.2 สถิติทดสอบความเหมาะสมสัมประสิทธิ์ ของสมการ

การทดสอบความเหมาะสมสัมประสิทธิ์ของ
สมการ เป็นการทดสอบค่าสัมประสิทธิ์การถดถอยโล
จิสติกในรูปของค่าสถิติ χ^2 ซึ่งมีด้วยกัน 3 ค่า ได้แก่
ค่า Step, Block และ Model ถ้าพบว่าค่าสถิติ χ^2 ใน
Step Block และ Model มีค่าเท่ากันทั้ง 3 ค่า และมี
นัยสำคัญทางสถิติแสดงว่า ตัวแปรทำนายที่เพิ่มเข้าไป
ในสมการนั้นมีความเหมาะสม สมมติฐานที่กำหนด
เป็นการทดสอบความเหมาะสมของสัมประสิทธิ์นี้คือ

$$H_0 : \beta_1 = \beta_2 = \dots \beta_p$$

H_1 : สมการขึ้นอยู่กับตัวแปรทำนายอย่าง
น้อย 1 ตัว

ผลการทดสอบ ถ้าค่าสถิติ χ^2 มีนัยสำคัญ
จะยอมรับ H_1

โดยการศึกษาของอมรา ดันประวดี (2544)
ที่ศึกษาปัจจัยที่มีผลต่อการชำระคืนเงินโครงการ
กองทุนให้กู้ยืมเพื่อการศึกษา โดยสร้างสมการโลจิสติก
เพื่อพยากรณ์เหตุการณ์ที่ผู้กู้จะชำระคืนเงิน (1 = ชำระ,
0 = ไม่ชำระ) ได้สมการพยากรณ์ดังนี้

$$\text{Logit}(y) = -8.93 + 3.09\text{INC} + 1.01\text{INFOR} + 0.19\text{KNO} + 3.73\text{CH} + 0.27\text{GRAD} \quad (3.2.1)$$

โดยที่ INC แทนรายได้ต่อเดือน
INFOR แทนการได้รับข่าวสารเกี่ยวกับการชำระเงินคืน
KNO แทนความเข้าใจในหลักเกณฑ์ของการกู้ยืม และการผ่อนชำระ
CH แทนความมีจิตสำนึกในการชำระเงินกู้
GRAD แทนผู้กู้ยืมเป็นผู้ที่สำเร็จการศึกษาในระดับอุดมศึกษา

โดยได้แสดงผลการวิเคราะห์ความเหมาะสม
ของสมการถดถอยโลจิสติกแบบไบนารีโดยใช้ค่าสถิติ
 χ^2 ดังนี้

	Chi-Square	df	Significance
Model	54.6390	5	0.000
Black	54.6390	5	0.000
Step	54.6390	5	0.000

ผลการวิเคราะห์ พบว่าค่าสถิติ χ^2 ใน Step
Block และ Model มีค่าเท่ากันคือ 54.6390 และมี
นัยสำคัญทางสถิติ ซึ่งแสดงว่า ตัวแปรทำนายที่เพิ่มเข้า
ไปในสมการนั้นมีความเหมาะสม

ในงานวิจัยนี้ได้้นำการถดถอยโลจิสติกมา
ประยุกต์ใช้ในการพยากรณ์โอกาสที่ผู้กู้ยืมจะชำระคืน

เงินโครงการกองทุนให้กู้ยืมเพื่อการศึกษาหลังจากจบ
การศึกษา โดยผู้วิจัยศึกษาตัวแปรตามคือการชำระคืน/
การไม่ชำระคืน และตัวแปรพยากรณ์เป็นตัวแปรทาง
สังคมศาสตร์ คือ ปัจจัยที่มีผลต่อการชำระเงินเงิน
ทั้งหมด 3 ด้าน (ยกเว้นรายได้ต่อเดือน และการที่ผู้กู้ยืม
เป็นผู้ที่สำเร็จการศึกษาในระดับอุดมศึกษา) โดย
สมการพยากรณ์นี้มีค่า Rcs^2 (Cox & Snell R Squared)
เท่ากับ 38.90% และค่า Nagelkerke R Squared หรือ
 RN^2 เท่ากับ 61.7

4 สถิติทดสอบระดับความสัมพันธ์

สถิติทดสอบระดับความสัมพันธ์ ดังนี้

4.1 สถิติทดสอบ Cox & Snell R square หรือ Rcs² สถิติทดสอบ Cox & Snell R square เป็นการพิจารณาหรือตรวจสอบความสอดคล้องของ model หรือเปอร์เซ็นต์ที่สามารถอธิบายความแปรปรวนหรือความผันแปรในการวิเคราะห์ถดถอยโลจิสติก โดยปกติค่า Cox & Snell R square หรือ Rcs² มีค่าน้อยกว่า 1 เสมอ สถิตินี้จะคล้ายกับค่า R² ในการวิเคราะห์ถดถอยพหุปกติ

4.2 สถิติทดสอบ Nagelkerke R square

(RN²) สถิตินี้ค่า RN² จะมีลักษณะเหมือนกับ Rcs² แต่จะมีค่ามากกว่า Rcs² เสมอ ค่า RN² และ ค่า Rcs² ของ Cox & Shell และ Nagelkerke เป็นค่า R² เทียม (Pseudo R²) ซึ่งเป็นค่าเปอร์เซ็นต์ที่สามารถอธิบายความผันแปรในการวิเคราะห์การถดถอยโลจิสติก

งานวิจัยของ Boamah (2011) ที่ศึกษาปัจจัยที่ส่งผลต่อพฤติกรรมการสูบบุหรี่ของนักเรียนมัธยมศึกษาตอนปลาย โดยใช้ข้อมูลitudinal จาก CDC's 2009 National Youth Risk Behavior Survey (YRBS) ผลการวิจัยเป็นดังนี้

	Predictor	B	S.E	Wald	Df	p-value	Odds
	Race White	1.62	.18	77.75	1	.00	5.4
	Hispanic	-.46	.27	2.95	1	.09	.63
	FCU	1.26	.07	285.62	1	.00	.352
	PIB	.54	.08	41.4	1	.00	1.71
	FSH	.53	.08	43.65	1	.00	1.70
	ICSA	.49	.02	712.41	1	.00	1.64
	Constant	-7.23	.22	1050.87	1	.00	0.00

โดยที่	Race White	แทน	Race
	Hispanic	แทน	Hispanic
	FCU	แทน	Frequent cocaine use
	PIB	แทน	Physically Inactive Behavior
	FSH	แทน	Felt Sad or Hopeless
	ICSA	แทน	Initial Cigarette Smoking Age

โดยสมการพยากรณ์นี้มีค่า Rcs² (Cox & Snell R Squared) เท่ากับ 13% และค่า Nagelkerke R Squared หรือ RN² เท่ากับ 31% สำหรับค่า odds พบว่านักเรียนที่มีผิวขาวมีความน่าจะเป็นที่จะมีพฤติกรรมการสูบบุหรี่สูงกว่ากลุ่มอื่นๆ ถึง 5.4 เท่า แต่หากพิจารณาทางด้านพฤติกรรมศาสตร์ พบว่า พฤติกรรม

การไม่มีกิจกรรมทางร่างกาย (PIB) และความรู้สึกเศร้าหรือสิ้นหวัง (FSH) เป็นปัจจัยที่เพิ่มความน่าจะเป็นที่จะมีพฤติกรรมการสูบบุหรี่สูงกว่าปัจจัยอื่นๆ ซึ่งหากลดพฤติกรรมดังกล่าวลงได้ความน่าจะเป็นที่จะมีพฤติกรรมการสูบบุหรี่ก็จะลดลง

ข้อเสนอแนะในการนำการถดถอยโลจิสติกไปใช้ในการวิจัยทางสังคมศาสตร์

สิ่งที่สำคัญก่อนการวิเคราะห์ข้อมูลทางสถิติ ผู้วิจัยควรตรวจสอบข้อมูลว่าเป็นไปตามข้อตกลงเบื้องต้น (Assumption) หรือไม่ ซึ่งข้อดีของการวิเคราะห์การถดถอยโลจิสติกคือข้อตกลงเบื้องต้นค่อนข้างมีความยืดหยุ่น คือไม่มีข้อกำหนดเกี่ยวกับการแจกแจงปกติของตัวแปรตาม (Normality) และการกระจายของตัวแปรตาม (Homoscedasticity) อีกทั้งยังสามารถวิเคราะห์ข้อมูลตัวแปรอิสระได้ทุกระดับการวัด ซึ่งข้อมูลในการวิจัยทางสังคมศาสตร์ส่วนใหญ่มักมีข้อจำกัดในเรื่องดังกล่าว สำหรับข้อจำกัดของการวิเคราะห์การถดถอยโลจิสติกคือการแปลผลค่าสถิติต่างๆ ที่ค่อนข้างยุ่งยาก

สรุป

การวิเคราะห์การถดถอยโลจิสติก เป็นเทคนิคทางสถิติที่มีความน่าสนใจ เพราะสามารถพยากรณ์การเกิดเหตุการณ์ที่สนใจ อีกทั้งการวิเคราะห์การถดถอยโลจิสติกยังมีข้อตกลงเบื้องต้นที่ค่อนข้างมีความยืดหยุ่นเหมาะสมสำหรับการวิเคราะห์ข้อมูลทางสังคมศาสตร์ ในการวิจัยทางสังคมศาสตร์สามารถนำการวิเคราะห์การถดถอยโลจิสติกมาประยุกต์เพื่อศึกษาถึงปัจจัยที่ส่งผลกระทบต่อการศึกษา หรือพฤติกรรมที่สนใจ โดยนอกจากจะทราบถึงความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจแล้ว ยังสามารถทราบถึงอิทธิพลในเชิงปริมาณของปัจจัยต่างๆ ได้จากสมการพยากรณ์ ซึ่งสามารถใช้เป็นแนวทางในการเพิ่มหรือลดโอกาสของการเกิดเหตุการณ์หรือพฤติกรรมที่สนใจได้ด้วย ดังเช่นในการศึกษาปัจจัยที่ส่งผลกระทบต่อพฤติกรรมการสูบบุหรี่ของนักเรียนมัธยมศึกษาตอนปลายของBoamah (2011) ที่พบว่า พฤติกรรมการไม่มีกิจกรรมทางร่างกาย และความรู้สึกเศร้าหรือสิ้นหวัง เป็นปัจจัยที่เพิ่มความน่าจะเป็นที่จะมีพฤติกรรมการสูบบุหรี่

ซึ่งหากลดพฤติกรรมดังกล่าวลงก็จะส่งผลให้ความน่าจะเป็นที่จะมีพฤติกรรมการสูบบุหรี่ลดลง

เอกสารอ้างอิง

- [1] กัลยา วานิชย์บัญชา.(2552). การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 4. ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- [2] ศศิธร แพนรินทร์. (2554). ปัจจัยที่ส่งผลต่อการตัดสินใจศึกษาต่อในมหาวิทยาลัยราชภัฏชัยภูมิ ระดับปริญญาตรีภาคปกติ ปีการศึกษา (2553). วิทยานิพนธ์ครุศาสตรมหาบัณฑิต. มหาวิทยาลัยราชภัฏชัยภูมิ.
- [3] สุวดี น้าพาเจริญ และชลธิชา จำรัสพร. (2558) การวิเคราะห์สมการถดถอย.เข้าถึงจาก http://www.tpa.or.th/writer/read_this_book_topic.php?bookID=3086&read=true&count=true
- [4] อมรา ต้นประวดี. (2544). ปัจจัยที่มีผลต่อการชำระคืนเงินกู้โครงการกองทุนเงินให้กู้ยืมเพื่อการศึกษา: กรณีศึกษา นิสิต นักศึกษาของสถาบันอุดมศึกษาในเขตกรุงเทพมหานคร วิทยานิพนธ์. วท.ม. (เศรษฐศาสตร์ธุรกิจ) มหาวิทยาลัยเกษตรศาสตร์.
- [5] Boamah, J. (2011). Multiple logistic regression analysis of cigarette use among high school students, Journal of Case Studies in Education, เข้าถึงจาก <http://www.aabri.com/manuscripts/10617.pdf>
- [6] Montgomery D., Peck E., Vining G. (2006) Introduction to Regression Analysis, A John Wiley & Sons Inc Publication, New Jersey, page 428-447.
- [7] Stevens, J. (1996). Applied multivariate statistics for the social science. Lawrence Erlbaum Associate, Inc., Mahwah, New Jersey page 558-586.

ประวัติผู้วิจัย



นางอรทัย เจริญสิทธิ์

Ph.D

อาจารย์ประจำสาขาวิทยาการ

คอมพิวเตอร์

คณะศิลปศาสตร์และวิทยาศาสตร์

มหาวิทยาลัยเอเชียอาคเนย์

การศึกษา

: ปี พ.ศ. 2536, วิทยาศาสตรบัณฑิต (วท.บ.)

สถิติประยุกต์มหาวิทยาลัยหอการค้าไทย

: ปี พ.ศ. 2540, สถิติศาสตรมหาบัณฑิต(สถม.)

มหาวิทยาลัยจุฬาลงกรณ์มหาวิทยาลัย

: ปี พ.ศ. 2559, ปรัชญาดุษฎีบัณฑิต (วิจัย วัดผลและ

สถิติการศึกษา) มหาวิทยาลัยบูรพา