

การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis)

มะแอน ราโอบ

หลักการของการวิเคราะห์ความถดถอยแบบพหุคูณ ตัวแปรตาม (ตัวแปรเกณฑ์) จะเป็นตัวแปรเชิงปริมาณ ในขณะที่ตัวแปรอิสระ (ตัวแปรพยากรณ์) จะเป็นตัวแปรเชิงปริมาณเพียงอย่างเดียว หรืออาจมีตัวแปรบางตัวที่เป็นตัวแปรเชิงปริมาณและตัวแปรบางตัวเป็นตัวแปรเชิงกลุ่มหรือเชิงคุณภาพ แต่ถ้าตัวแปรตามเป็นตัวแปรเชิงกลุ่ม จะต้องใช้เทคนิคการวิเคราะห์การถดถอยโลจิสติก ซึ่งยังคงมีวัตถุประสงค์และแนวคิดเหมือนกับการวิเคราะห์การถดถอยเชิงเส้น คือ เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ และนำเสนอการที่ได้ไปประมาณหรือพยากรณ์ตัวแปรตาม เมื่อกำหนดค่าตัวแปรอิสระ

ประเภทของการวิเคราะห์การถดถอยโลจิสติก (กัลยา วานิชย์บัญชา, 2546)

การวิเคราะห์การถดถอยโลจิสติกแบ่งออกเป็น 2 ประเภท

1. Binary Logistic จะใช้เมื่อตัวแปรตาม Y เป็นตัวแปรเชิงกลุ่มที่มีค่าได้เพียง 2 ค่า (Dichotomous Variable) เช่น

$$Y = \begin{cases} 1 & \text{ถ้าคนไข้เป็นโรคหัวใจ} \\ 0 & \text{ถ้าคนไข้ไม่ได้เป็นโรคหัวใจ} \end{cases}$$

โดยพิจารณาจากตัวแปรต้น หรือตัวแปรอิสระ เช่น อาการ การออกกำลังกาย การสูบบุหรี่ ระดับคอเลสเตอรอล เป็นต้น

หรือ

$$Y = \begin{cases} 1 & \text{ถ้าลูกค้าซื้อสินค้า} \\ 0 & \text{ถ้าลูกค้าไม่ซื้อสินค้า} \end{cases}$$

โดยพิจารณาตัวแปรต้นหรือตัวแปรอิสระ อาจจะเป็นรายได้ อาชีพ และอายุ เป็นต้น

2. Multinomial Logistic จะใช้เมื่อตัวแปรตาม Y เป็นตัวแปรเชิงกลุ่มที่มีค่ามากกว่า 2 ค่า เช่น

ด้านการตลาด ในตลาดภาพยนตร์ บริษัทผู้สร้างภาพยนตร์ต้องการพยากรณ์ว่าภาพยนตร์ประเภทใด จะเป็นที่นิยม โดยที่

$$Y = \begin{cases} 1 & \text{ถ้าเป็นภาพยนตร์ประเภทบู๊} \\ 2 & \text{ถ้าเป็นภาพยนตร์ประเภทชีวิต} \\ 3 & \text{ถ้าเป็นภาพยนตร์ประเภทตลก} \end{cases}$$

โดยตัวแปรอิสระอาจจะเป็น อายุของกลุ่มเป้าหมาย เพศ เป็นต้น

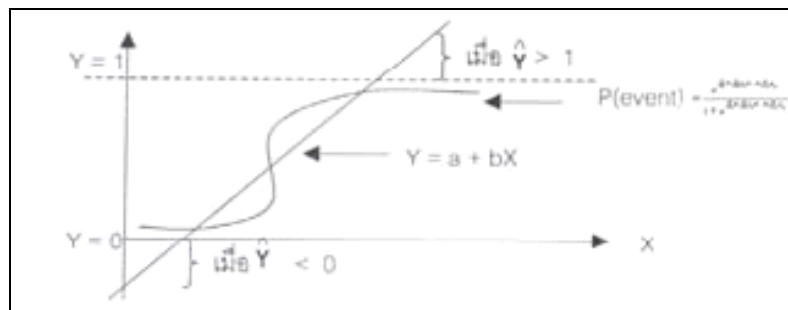
ด้านสังคมศาสตร์ Y หมายถึงระดับที่มีส่วนร่วมในการพัฒนาชุมชน $Y = 1, 2, 3, 4, 5$ โดยที่ $Y=1$ เมื่อไม่มีส่วนร่วมเลย, ..., $Y=5$ มีส่วนร่วมมากที่สุด

ด้านการแพทย์ Y หมายถึงระดับการเป็นโรคมะเร็ง เช่น $Y = 1, 2, 3, 4, 5$ โดยที่ $Y=1$ หมายถึงไม่เป็นโรคมะเร็ง, $Y=2$ หมายถึงการเป็นมะเร็งขั้นต้น, ..., $Y=5$ หมายถึงการเป็นมะเร็งขั้นสุดท้าย

เหตุผลที่ใช้การวิเคราะห์การถดถอยโลจิสติกแทนการวิเคราะห์การถดถอยเชิงเส้น

(สวัสดีชัย ศรีพนมธนากร : 2548) ได้กล่าวว่า

1. เมื่อ Y มีค่าได้เพียง 2 ค่า ทำให้ค่าประมาณของ Y เป็นโอกาสที่เหตุการณ์ที่สนใจจะเกิด ซึ่งมีค่าระหว่าง 0 ถึง 1 ถ้าใช้สมการการถดถอยเชิงเส้นตรง คือ $Y' = a + bX$ ค่า Y' ที่ได้อาจจะไม่ได้อยู่ในช่วง 0 ถึง 1 หรือ อาจมีค่าน้อยกว่า 0 หรือ มากกว่า 1



จะพบว่ากราฟของโลจิสติกไม่ใช่เส้นตรง และมีค่าอยู่ในช่วง 0 ถึง 1

2. Nonnormal Error Terms ในการวิเคราะห์การถดถอยเชิงเส้น มีเงื่อนไขว่า ค่าความคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ แต่เมื่อ Y มีค่าเพียง 2 ค่า คือ 0 กับ 1 จะทำให้ค่าความคลาดเคลื่อน e มีค่าได้เพียง 2 ค่าด้วย ซึ่งเป็นไปไม่ได้ที่ e จะมีการแจกแจงแบบปกติ จึงทำให้ไม่สามารถใช้การวิเคราะห์การถดถอยเชิงเส้นตรงโดนทั่วไป

3. Nonconstant Error Variance เนื่องจากเงื่อนไขของการวิเคราะห์การถดถอย คือ ค่าแปรปรวนของค่าความคลาดเคลื่อน หรือ $V(e)$ ต้องคงที่ทุกค่าของ X แต่ใน logistic นั้น เมื่อ Y มีค่าได้เพียง 2 ค่า และ Y มีการแจกแจงแบบเบอร์นูลลี ซึ่งทำให้ค่าแปรปรวนและค่าเฉลี่ยมีความสัมพันธ์กัน จึงทำให้เงื่อนไขที่ว่า $V(e)$ คงที่ไม่เป็นจริง ซึ่งทำให้ไม่สามารถใช้การวิเคราะห์การถดถอยเชิงเส้นตรงตามปกติได้

วัตถุประสงค์ของการวิเคราะห์การถดถอยโลจิสติก (กัลยา วานิชย์บัญชา, 2546)

1. เพื่อศึกษาความสัมพันธ์ของตัวแปรอิสระที่มีต่อโอกาสที่จะเกิดเหตุการณ์ (ตัวแปรตาม) พร้อมทั้งศึกษาระดับความสัมพันธ์ของตัวแปรอิสระแต่ละตัว
2. เพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ จากสมการที่เหมาะสม หรือใช้สมการ โดยการเลือกตัวแปรอิสระที่เหมาะสมเพื่อทำให้เปอร์เซ็นต์ของความถูกต้องในการพยากรณ์มีค่าสูงสุด

เงื่อนไขของการวิเคราะห์การถดถอยโลจิสติก (กัลยา วานิชย์บัญชา, 2546)

การวิเคราะห์การถดถอยโลจิสติก จะมีเงื่อนไขน้อยกว่าการวิเคราะห์การถดถอยแบบปกติ แต่อย่างไรก็ตาม การวิเคราะห์การถดถอยโลจิสติกก็ยังมีเงื่อนไขหลายข้อดังนี้

1. ตัวแปรอิสระ X 's อาจจะเป็นข้อมูลชนิด Dichotomous (มีค่าได้ 2 ค่า) หรือเป็นสเกลอันตรภาค (Interval Scale) และสเกลอัตราส่วน (Ratio Scale) ก็ได้
 2. ค่าคาดหวังของค่าความคลาดเคลื่อนเป็นศูนย์หรือ $E(e) = 0$
 3. e_i และ e_j เป็นอิสระกัน
 4. e_i และ X_i เป็นอิสระกัน
 5. ตัวแปรอิสระไม่ควรมีความสัมพันธ์กัน หรือไม่ควรเกิดปัญหา Multicollinearity
- สำหรับเงื่อนไขของการวิเคราะห์การถดถอยแบบปกติ นอกจากจะมีเงื่อนไขทั้ง 5 ข้อข้างต้น จะต้องเพิ่มเงื่อนไขอีก 2 ข้อ คือ
1. ค่าคลาดเคลื่อนมีการแจกแจงแบบปกติ
 2. ค่าแปรปรวนของค่าคลาดเคลื่อนคงที่

หมายเหตุ การวิเคราะห์การถดถอยโลจิสติกจะต้องใช้ขนาดตัวอย่าง n มากกว่าการวิเคราะห์การถดถอยแบบปกติ โดยทั่วไป $n \geq 30p$ โดยที่ p เป็นจำนวนตัวแปรอิสระ

โมเดลการวิเคราะห์การถดถอยโลจิสติก (สวัสดีชัย ศรีพนมธนากร, 2548)

ในกรณีที่มีตัวแปรอิสระเพียงตัวเดียว โมเดลการวิเคราะห์การถดถอยโลจิสติก สามารถเขียนได้ดังสมการ

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

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

เมื่อ β_0 และ β_1 เป็นสัมประสิทธิ์ที่ประมาณได้จากข้อมูล
X เป็นตัวแปรอิสระ

e เป็นล็กอธรรมชาติ (natural logarithms) มีค่าประมาณ 2.718

จากสมการข้างต้น เราสามารถเขียนสมการใหม่ได้เป็น

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

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

โดย $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$

และโอกาสของการไม่เกิดเหตุการณ์จะประมาณได้จากสมการ

$$\text{Prob(noevent)} = 1 - \text{Prob(event)}$$

ในการวิเคราะห์การถดถอยเชิงเส้น จะประมาณค่าพารามิเตอร์ใน โมเดลโดยใช้กำลังสองน้อยที่สุด โดยคัดเลือกสัมประสิทธิ์การถดถอย ที่ทำให้ค่าผลรวมของกำลังสองของความคลาดเคลื่อนในการทำนาย $(\sum (Y - Y')^2)$ มีค่าน้อยที่สุด ส่วนการวิเคราะห์การถดถอยโลจิสติก จะประมาณค่าพารามิเตอร์โดยวิธี maximum-likelihood อันเป็นการคำนวณทวนซ้ำ (iterative algorithm) เพื่อให้ได้ค่าประมาณของพารามิเตอร์ที่ใกล้เคียงกับข้อมูลเชิงประจักษ์มากที่สุด

เพื่อให้เห็นภาพของการวิเคราะห์การถดถอยโลจิสติก จะเสนอผลการวิจัยของบราวน์ (Brown, 1980 cited in norusis, 1990: อ้างถึงใน ศิริเดช สุชีวะ, 2541) เกี่ยวกับการพยากรณ์การเป็นมะเร็ง โดยมี

ที่มาของปัญหาวิจัยว่า การรักษาและการพยากรณ์การเป็นมะเร็งนั้นขึ้นอยู่กับการลุกลามของเซลล์มะเร็ง ซึ่งมักเกิดในต่อมน้ำเหลืองก่อนการตรวจสอบมะเร็งในระยะเริ่มแรกจึงต้องอาศัยการผ่าตัดต่อมน้ำเหลือง เพื่อสำรวจว่ามีเซลล์มะเร็งอยู่หรือไม่ แล้วจึงให้การรักษาที่จำเป็นต่อไป ถ้าเราสามารถพยากรณ์การมีเซลล์มะเร็งในต่อมน้ำเหลืองได้ โดยอาศัยข้อมูลการวินิจฉัยที่มีอยู่ ก็ไม่ต้องทำการผ่าตัดต่อมน้ำเหลือง และลดค่าใช้จ่ายต่าง ๆ ออกไปได้มาก บรรานันท์ จึงเก็บข้อมูลจากผู้เข้ารับการตรวจมะเร็งระยะเริ่มแรกจำนวน 53 คน โดยเก็บข้อมูล อายุ serum acid phosphatase (สารที่จะมีค่าสูงขึ้นเมื่อเซลล์ที่เจริญผิดปกติได้แพร่กระจายไปในพื้นที่นั้น) ระยะของโรค (stage) ระดับการผิดปกติของเซลล์ (grade) และผลการ X-ray รวมทั้งผลการผ่าตัดพิสูจน์มะเร็งในต่อมน้ำเหลือง จากนั้นจึงสร้างสมการการถดถอยโลจิสติก เพื่อพยากรณ์การเป็นมะเร็งต่อมน้ำเหลืองจากข้อมูลกลุ่มตัวอย่างทั้ง 53 คน

ในการสร้างสมการถดถอยโลจิสติกเพื่อพยากรณ์ จะให้คำสั่งในโปรแกรม SPSS ดังนี้

LOGISTIC REGRESSION VARIABLES = dependent WITH independents.

ซึ่งจะให้ได้ผลการวิเคราะห์ข้อมูลดังนี้

Parameter estimates for the logistic regression model

คำสั่ง

LOGISTIC REGRESSION VARIABLES = dependent WITH independents.

ผลลัพธ์

Variables in the Equation

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
AGE	-.0693	.0579	1.4320	1	.2314	.0000	.9331
ACID	.0243	.0132	3.4229	1	.0643	.1423	1.0246
XRAY	2.0453	.8072	6.4207	1	.0113	.2509	7.7317
GRADE	.7614	.7708	.9758	1	.3232	.0000	2.1413
STAGE	1.5641	.7740	4.0835	1	.0433	.1722	4.7783
Constant	.0618	3.4599	.0003	1	.9357		

ค่าพารามิเตอร์ที่ประมาณได้จากโมเดลการถดถอยโลจิสติก

จากตาราง แสดงค่าสัมประสิทธิ์ที่ประมาณได้ (คอตมันน์ B) และค่าสถิติที่เกี่ยวข้องกับการวิเคราะห์การถดถอยโลจิสติก เพื่อทำนายการเป็นมะเร็งต่อมไทรอยด์จากตัวแปร AGE, ACID, XRAY, STAGE และ GRADE ตัวแปรอิสระ 3 ตัวสุดท้ายนี้ เป็นตัวแปรเชิงปริมาณหรือเรียกอีกอย่างว่าตัวแปรบ่งชี้ (indicator variable) มีรหัสเป็น 0 กับ 1 โดย XRAY มีค่าเป็น 1 หมายถึง ให้ผลการ X-ray เป็นบวก STAGE เป็น 1 หมายถึง อยู่ในระยะรุนแรง และ GRADE เป็น 1 หมายถึง การผิดปกติของเซลล์อยู่ในระดับร้ายแรง

จากค่าสัมประสิทธิ์ที่ได้ เราสามารถเขียนสมการการถดถอยโลจิสติก ทำนายโอกาสของการเป็นมะเร็งต่อมไทรอยด์ได้ดังนี้

$$\text{Prob}(\text{การเป็นมะเร็งต่อมไทรอยด์}) = 1 / 1 + e^{-z}$$

$$\text{โดย } z = 0.0618 - 0.0693(\text{AGE}) + 0.0243(\text{ACID}) + 2.0453(\text{XRAY}) \\ + 1.5641(\text{STAGE}) + 0.7614(\text{GRADE})$$

ถ้าลองพยากรณ์การเป็นมะเร็งของชายอายุ 66 ปี ที่มีค่า serum acid phosphatase เป็น 48 และค่าตัวแปรที่เหลือเป็น 0 จะได้ว่า

$$z = 0.0618 - 0.0693(66) + 0.0243(48) = -3.346$$

$$\text{Prob}(\text{การเป็นมะเร็งต่อมไทรอยด์}) = 1 / 1 + e^{-(-3.346)} = 0.0340$$

จากผลการวิเคราะห์นี้เราพยากรณ์ว่า เขาไม่น่าจะเป็นมะเร็งต่อมไทรอยด์ โดยการพยากรณ์ยึดเกณฑ์ความน่าจะเป็น 0.5

การทดสอบความถูกต้อง (goodness of fit) ของโมเดล (สวัสดีชัย ศรีพนมธนากร, 2548)

การทดสอบความสอดคล้องระหว่างโมเดลกับข้อมูลในการวิเคราะห์การถดถอยโลจิสติกทำได้หลายวิธี เช่น การใช้ตารางจัดจำพวก ช่วยในการพิจารณาเปรียบเทียบผลการพยากรณ์จากโมเดลกับข้อมูลเชิงประจักษ์ หรือ การพิจารณาฮิสโตแกรมของค่าประมาณความน่าจะเป็น ซึ่งจะแสดงค่าจริงเปรียบเทียบกับค่าประมาณ หรือการพิจารณาภาวะเนบสนิตีของโมเดล โดยพิจารณาค่า $-2 \log$ likelihood ($-2LL$) ซึ่งถ้าค่า $-2LL$ มีค่าน้อย หรือเข้าใกล้ศูนย์ แสดงว่า โมเดลเหมาะสม และถ้าโมเดลเหมาะสม 100% ค่า likelihood จะเท่ากับ 1 และ $(-2LL)$ จะเท่ากับ 0

ส่วนอีกวิธีหนึ่งที่ใช้คือ Hosmer-Lemeshow goodness-of fit ลักษณะการตรวจสอบความเหมาะสมของวิธีนี้ จะแบ่ง case ออกเป็น 10 กลุ่ม ๆ ละ พอ กัน โดยการแบ่งพิจารณาจากค่าประมาณ

ของโอกาสที่เหตุการณ์จะเกิด โดยใช้สถิติทดสอบไคสแควร์ซึ่งขนาดตัวอย่างต้องมากพอที่จะทำให้จำนวน case ที่คาดหวังไว้ (Expected value : E_i ; $i = 1, \dots, 10$) จะต้องมากกว่า 15 และไม่มีกลุ่มใดที่มี $E_i = 1$

$$\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i}$$

ใช้ทดสอบ H_0 : โมเดล เหมาะสม

แต่ถ้าขนาดตัวอย่างใหญ่ ค่าสถิติทดสอบ Hosmer-Lemeshow จะมีค่ามากด้วย ทำให้อาจสรุปว่า ปฏิเสธ H_0 ทั้งที่โมเดลเหมาะสม ผู้ใช้จึงควรระมัดระวังการแปลผลด้วย

การทดสอบสมมติฐานเกี่ยวกับค่าสัมประสิทธิ์ของโมเดลการวิเคราะห์การถดถอยโลจิสติก

การทดสอบสมมติฐานว่า ค่าสัมประสิทธิ์ไม่เท่ากับ 0 จะใช้ Wald statistic ซึ่งมีการแจกแจงแบบไคสแควร์ Wald statistic เป็นกำลังสองของอัตราส่วนระหว่างค่าสัมประสิทธิ์กับค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์นั้น ถ้าเป็นตัวแปรจัดกลุ่ม (categorical variable) ซึ่งเป็นตัวแปรเชิงคุณภาพ Wald statistic มี degree of freedom เท่ากับจำนวนกลุ่มลบด้วย 1 ตัวอย่างเช่น สัมประสิทธิ์ของตัวแปร AGE เท่ากับ -0.0693 และความคลาดเคลื่อนมาตรฐานเป็น 0.0579 (ในคอลัมน์ S.E.) Wald statistic จะเท่ากับ $(-0.0693 / 0.0579)$ หรือประมาณ 1.432^2 นัยสำคัญของ Wald statistic แสดงในคอลัมน์ Sig. จากตารางข้างต้น สัมประสิทธิ์ของ XRAY และ STAGE เท่านั้นที่ไม่เท่ากับ 0 อย่างมีนัยสำคัญที่ระดับ .05

สหสัมพันธ์บางส่วน (partial correlation) ในโมเดลการวิเคราะห์การถดถอยโลจิสติก

ในการพิจารณาการมีส่วนร่วมของตัวแปรอิสระ แต่ละตัวที่มีคือการพยากรณ์ด้วยตัวแปรตามนั้นจะดูจากค่าสหสัมพันธ์บางส่วน ระหว่างตัวแปรตามกับตัวแปรอิสระ แต่ละตัวดังในคอลัมน์ R statistic ในตารางข้างต้น R มีค่าตั้งแต่ -1 ถึง +1 ค่า R ที่เป็นบวก หมายถึง ถ้าค่าของตัวแปรเพิ่มค่า likelihood ของการเกิดเหตุการณ์นั้น ก็จะเพิ่มขึ้นด้วย ค่า R ที่เป็นลบจะแปลความหมายในทางตรงกันข้าม

การแปลความหมายของสัมประสิทธิ์การถดถอย

การวิเคราะห์การถดถอยเชิงเส้นตรง จะแปลความหมายสัมประสิทธิ์การถดถอยได้โดยตรงว่าเป็นขนาดการเปลี่ยนแปลงของตัวแปรตาม เมื่อตัวแปรอิสระ เปลี่ยนไปหนึ่งหน่วย แต่การแปลความหมายสัมประสิทธิ์โลจิสติกจะแตกต่างกันไปงานนี้ ก่อนอื่นขออธิบายเกี่ยวกับโมเดลโลจิสติกก่อนว่า โมเดลโลจิสติกสามารถเขียนในรูปของ odd ของการเกิดเหตุการณ์ได้ (odd ของการเกิดเหตุการณ์ หมายถึง อัตราส่วนระหว่างโอกาสที่จะเกิดกับโอกาสที่จะไม่เกิดเหตุการณ์ เช่น odd ของการออกหัวในการโยนเหรียญ 1 ครั้ง เท่ากับ $0.5/0.5 = 1$ เป็นต้น) การเขียนโมเดลโลจิสติกในรูป log ของ odd (ซึ่งเรียกว่า logit) เป็นดังนี้

$$\log(\text{Prob(event)} / \text{Prob(no event)}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

จากสมการจะเห็นว่าสัมประสิทธิ์โลจิสติก สามารถแปลความได้ว่าเป็นการเปลี่ยนแปลงของ log odd ตามการเปลี่ยนแปลงหนึ่งหน่วยของตัวแปรอิสระ ในตารางข้างต้น สัมประสิทธิ์ของ GRADE เปลี่ยนจาก 0 เป็น 1 และควบคุมตัวแปรอิสระ ที่เหลือ log odd ของการเป็นมะเร็งต่อมน้ำเหลืองจะเพิ่มขึ้น 0.76 แต่การแปลความหมายในรูปของ odd จะง่ายกว่า log odd ดังนั้น จึงเขียนสมการ โลจิสติกใหม่ในเทอมของ odd ได้เป็น

$$\left(\frac{\text{Prob(event)}}{\text{Prob(noevent)}} \right) = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p} = e^{\beta_0} e^{\beta_1 X_1} \dots e^{\beta_p X_p}$$

e ยกกำลัง β_i เป็นค่า odd ที่เปลี่ยนแปลง เมื่อตัวแปรอิสระ ตัวที่ i มีค่าเพิ่มขึ้น 1 หน่วย β_i เป็นลบ เทอมนี้จะน้อยกว่า 1 หมายความว่า odd จะลดลง ถ้า $\beta_i = 0$ เทอมนี้จะมีค่าเท่ากับ 1 ซึ่งหมายความว่า odd จะไม่เปลี่ยนแปลง เช่น เมื่อ GRADE เปลี่ยนจาก 0 เป็น 1 ค่า odd จะเพิ่มขึ้น 2.14 ดังแสดงในคอลัมน์ $\text{Exp}(\beta)$ ของตารางข้างต้น

ถ้าลองคำนวณ odd ของการเป็นมะเร็งต่อมน้ำเหลืองของชายอายุ 60 ปี ที่มี seruma acid phosphatase ระดับ 62 ผล X-ray เป็น 1 และผลการวินิจฉัย STAGE และ GRADE เป็น 0 ขึ้นแรกต้องคำนวณโอกาสโดยประมาณของการเป็นมะเร็งต่อมน้ำเหลือง

$$\text{Estimated prob (การเป็นมะเร็งต่อมน้ำเหลือง)} = 1 / 1 + e^{-Z}$$

$$\begin{aligned} \text{โดย } Z &= 0.0618 - 0.0693(60) + 0.0243(62) + 2.0453(1) + 0.7614(0) + 1.5641(0) \\ &= -0.54 \end{aligned}$$

แทนค่า Z ในสมการได้ $1 / 1 + e^{-(-0.54)} = 0.37$

เมื่อโอกาสในการเป็นมะเร็งต่อมน้ำเหลืองเท่ากับ 0.37 และโอกาสที่จะไม่เป็นเท่ากับ 0.63
odds ของการเป็นมะเร็งประมาณได้ดังนี้

$$\text{odds} = \text{Prob}(\text{event}) / \text{Prob}(\text{no event}) = 0.37 / 1-0.37 = 0.59$$

และ log odd เท่ากับ -0.53

สัมประสิทธิ์การพยากรณ์ (R^2)

ค่าสัมประสิทธิ์การพยากรณ์ (R^2) เป็นค่าที่บอกสัดส่วน หรือเปอร์เซ็นต์ที่สามารถอธิบายความผันแปรใน logistic regression model ซึ่งจะมีลักษณะคล้ายกับค่า (R^2) ในการวิเคราะห์การถดถอยเชิงเส้นตรง ในโปรแกรม SPSS จะนำเสนอ ค่า (R^2) ของCox&Snell และNahelkerke หรือเรียกว่า Pseudo

$$\text{Cox \& Snell } R^2 = 1 - \left[\frac{L(O)}{L(B)} \right]^{2/n}$$

โดยที่L(O) คือ likelihood สำหรับ โมเดลที่มีเพียงค่าคงที่

L(B) คือ likelihood สำหรับ โมเดลที่มีตัวแปรอิสระตามที่กำหนด

$$\text{Nahelkerke } R^2 = \frac{\text{Cox \& Snell } R^2}{\text{Cox \& Snell } R^2_{\max}}$$

$$\text{Cox \& Snell } R^2_{\max} = 1 - [L(O)]^{2/n}$$

การคัดเลือกตัวแปรพยากรณ์

การวิเคราะห์การถดถอยโลจิสติก สามารถคัดเลือกตัวแปรพยากรณ์ (ตัวแปรอิสระ) ที่ดีได้ เช่นเดียวกับการวิเคราะห์การถดถอยเชิงเส้น ซึ่งมีหลายวิธี ดังนี้

1. Enter เป็นเทคนิควิธีการเลือกตัวแปรอิสระเข้าสมการในขั้นตอนเดียว โดยผู้ใช้จะต้องเป็นผู้ตัดสินใจในเองว่าตัวแปรอิสระตัวใดบ้างที่มีความสัมพันธ์กับตัวแปรตาม หรือควรที่อยู่ในสมการการถดถอยโลจิสติก โดยอาจพิจารณาจากค่าสถิติทดสอบ Significance ของสถิติทดสอบหรืออาจอาศัยวรรณกรรมและทฤษฎีที่เกี่ยวข้องกับตัวแปรนั้นเป็นพื้นฐาน

2. forward : conditional เป็นเทคนิค Forward stepwise เทคนิคหนึ่ง โดยที่การทดสอบเพื่อเลือกตัวแปรของสมการถดถอยโลจิสติกจะพิจารณาจากค่าความน่าจะเป็นของ likelihood ratio statistic ที่ขึ้นกับ conditional parameter estimates

3. Forward : LR เป็นเทคนิค Forward stepwise เทคนิคหนึ่ง โดยที่การทดสอบเพื่อเลือกตัวแปรของสมการถดถอยโลจิสติก จะต้องพิจารณาจากค่าความน่าจะเป็นของ likelihood ratio statistic ที่ขึ้นอยู่กับค่าประมาณ โดยวิธี maximum likelihood

4. Forward : Wald เป็นเทคนิค Forward stepwise เทคนิคหนึ่ง โดยที่การทดสอบเพื่อเลือกตัวแปรของสมการถดถอยโลจิสติก จะต้องพิจารณาจากค่าความน่าจะเป็นของสถิติ Wald

5. Backward : Conditional เป็นเทคนิค Backward stepwise เทคนิคหนึ่ง โดยที่การทดสอบตัวแปรของสมการถดถอยโลจิสติก จะขึ้นอยู่กับค่าความน่าจะเป็นของ likelihood ratio statistic ที่ขึ้นกับค่า Conditional parameter estimates

6. Backward : LR เป็นเทคนิค Backward เทคนิคหนึ่ง โดยที่การทดสอบตัวแปรของสมการถดถอยโลจิสติก จะขึ้นอยู่กับค่าความน่าจะเป็นของ likelihood ratio statistic ที่ขึ้นกับค่าประมาณ โดยวิธี maximum likelihood

7. Backward : Wald เป็นเทคนิค Backward stepwise เทคนิคหนึ่ง โดยที่การทดสอบตัวแปรของสมการถดถอยโลจิสติก จะขึ้นกับค่าความน่าจะเป็นสถิติ Wald

บทสรุป (สมหวัง พิธิยานุวัฒน์, 2541)

จากที่กล่าวมาทั้งหมด คงจะทำให้ผู้อ่านมองเห็นภาพของการวิเคราะห์การถดถอยโลจิสติกและการแปลความหมายได้อย่างชัดเจนขึ้น การเลือกใช้สถิติให้เหมาะสมกับเป้าหมายของการวิจัยเป็นเรื่องที่สำคัญสำหรับการสร้างสรรคงานวิจัยที่มีคุณภาพ การวิเคราะห์การถดถอยโลจิสติก เป็นทางเลือกหนึ่งสำหรับการวิจัยที่มีเป้าหมายเพื่อพยากรณ์ตัวแปรตามที่มีลักษณะเป็น dichotomous โดยอาศัยหลักการของการวิเคราะห์ถดถอยเชิงเส้น แต่มีการแปลความหมายสัมประสิทธิ์การถดถอย ที่แตกต่างออกไป โดยต้องแปลความหมายในรูปของอัตราส่วนแถมต่อ (odd ratio) และโมเดลการถดถอยโลจิสติกสามารถใช้วิเคราะห์ข้อมูลที่มีลักษณะตามข้อตกลงเบื้องต้นของการวิเคราะห์จำแนก (discriminant analysis) ได้เป็นอย่างดีด้วย (Homse, D.W. Lemeshow, S., 1989) การวิเคราะห์ถดถอยโลจิสติกจึงเป็นทางเลือกที่น่าสนใจอีกทางเลือกหนึ่ง

ขั้นตอนของการวิเคราะห์การถดถอยโลจิสติก

ขั้นที่ 1 เลือกตัวแปรอิสระที่คาดว่าจะมีความสัมพันธ์กับตัวแปรตาม (โอกาสที่เหตุการณ์จะเกิด) โดยที่ตัวแปรอิสระมีมากกว่า 1 ตัวก็ได้

ขั้นที่ 2 ตรวจสอบค่าที่ผิดปกติของตัวแปรอิสระแต่ละตัว

ขั้นที่ 3 สร้างสมการ Logistic Response Function แล้วตรวจสอบความถูกต้องเหมาะสมของสมการโดยพิจารณาจากค่า pseudo R^2 และค่า Wald Statistics

ขั้นที่ 4 ตรวจสอบเงื่อนไขของการวิเคราะห์การถดถอยโลจิสติก

ขั้นที่ 5 ถ้าวัตถุประสงค์เพื่อพยากรณ์ case ใหม่ว่าจะเกิดเหตุการณ์ที่สนใจหรือไม่ เช่น ลูกค่าจะซื้อสินค้าหรือไม่ หรือผู้อาศัยในชุมชนจะมีส่วนร่วมในการพัฒนาชุมชนหรือไม่ คนไข้จะเป็นโรคหัวใจหรือไม่ จะใช้สมการที่ \hat{p} ในการพยากรณ์ หรือประมาณค่าความน่าจะเป็นที่จะเกิดเหตุการณ์ของ case นั้น เมื่อทราบค่าตัวแปรอิสระ X's ถ้า

\hat{p} (เหตุการณ์) \rightarrow < 0.5 จะได้ $Y = 0$ หรือไม่เกิดเหตุการณ์

\hat{p} (เหตุการณ์) \rightarrow ≥ 0.5 จะได้ $Y = 1$ หรือเกิดเหตุการณ์

ค่า 0.5 เป็นค่าความน่าจะเป็นที่ใช้เป็น Cutting Score ที่ผู้วิเคราะห์อาจจะกำหนดค่า 0.7 หรือ 0.4 หรือค่าอื่น ๆ แล้วแต่ผู้วิเคราะห์จะเห็นสมควรในแต่ละเรื่องที่จะโลจิสติกไปประยุกต์ใช้ แต่โดยทั่วไปมักนิยมใช้ 0.5

ตัวอย่างของ Binary Logistic

ผู้เขียนได้สมมติข้อมูลโดยใช้ตัวแปรต่าง ๆ ตามวิทยานิพนธ์ของคุณ ภิญ โย วรณสุข เรื่อง “การประยุกต์ใช้ในการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์อิทธิพลในการศึกษาปัจจัยที่มีอิทธิพลต่อความคาดหวังในการศึกษาต่อของนักเรียนชั้นประถมศึกษาในเขตการศึกษา 11” โดยมีรายละเอียดเกี่ยวกับตัวแปรดังต่อไปนี้

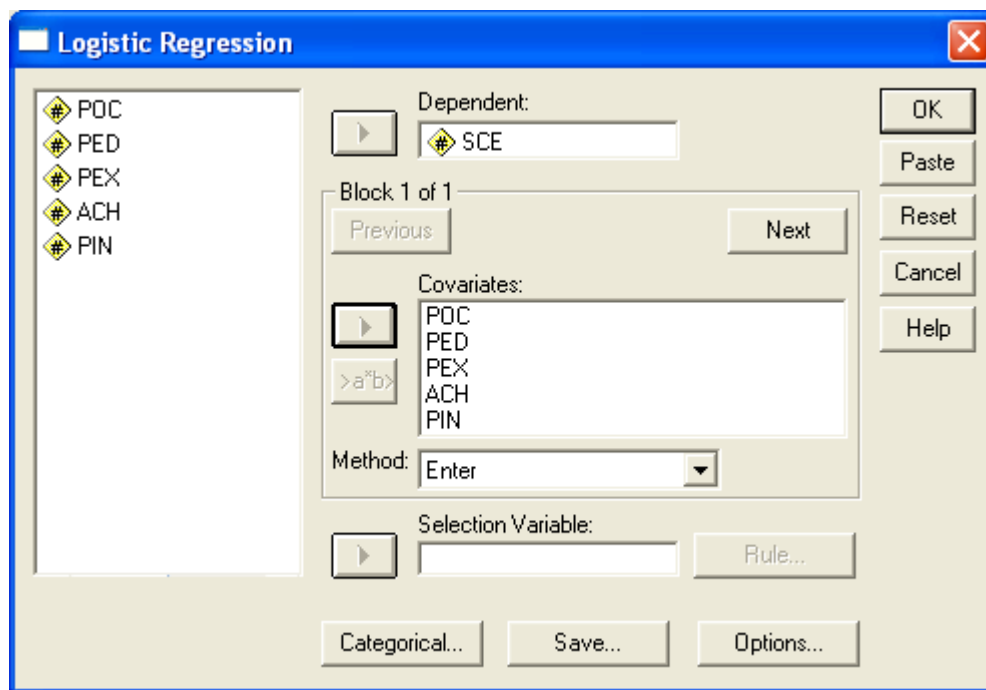
POC	parents' occupation	อาชีพของผู้ปกครอง (1 คือ อาชีพเกษตรกรรม 0 คืออาชีพอื่นจากเกษตรกรรม)
PED	parents' education	ระดับการศึกษาของผู้ปกครอง (0 คือ ไม่ได้เรียนหนังสือ 6, 9, 12, 14, 16 และ 18 ในกรณีที่จบการศึกษาระดับประถมศึกษา มัธยมศึกษาตอนต้น มัธยมศึกษาตอนปลาย อนุปริญญา ปริญญาตรี และสูงกว่าปริญญาตรี ตามลำดับ)
PIN	parents' income	รายได้ของผู้ปกครอง (1 มีรายได้รวมปีละประมาณต่ำกว่า 60,000 บาท ให้คะแนน 2, 3, และ 4 เมื่อมีรายได้ 60,001-80,000 บาท 80,001-100,000 บาท และมากกว่า 100,000 บาท ตามลำดับ)
PEX	parents' expectation	ความมุ่งหวังที่จะให้บุตรได้รับการศึกษาชั้นสูง (6, 9, 12, 14, 16 และ 18 ในกรณีที่จบการศึกษาระดับประถมศึกษา มัธยมศึกษาตอนต้น มัธยมศึกษาตอนปลาย อนุปริญญา ปริญญาตรี และสูงกว่าปริญญาตรี ตามลำดับ)
ACH	achievement	ผลสัมฤทธิ์ทางการเรียน (ให้คะแนนตามระดับผลการเรียนเฉลี่ยของนักเรียน)
SCE	students' continuing education expectation	ความคาดหวังในการศึกษาต่อของนักเรียน (1 คือ คาดหวังว่าจะศึกษาต่อ 0 คือ คาดหวังว่าจะไม่ศึกษาต่อ)

ในที่นี้ตัวแปร SCE เป็นตัวแปรตาม เนื่องจากมีค่าได้เพียง 2 ค่า คือ 0 และ 1 ส่วนที่เหลือเป็นตัวแปรอิสระ

โดยจะใช้คำสั่ง

Analyze → Regression → Binary Logistic...

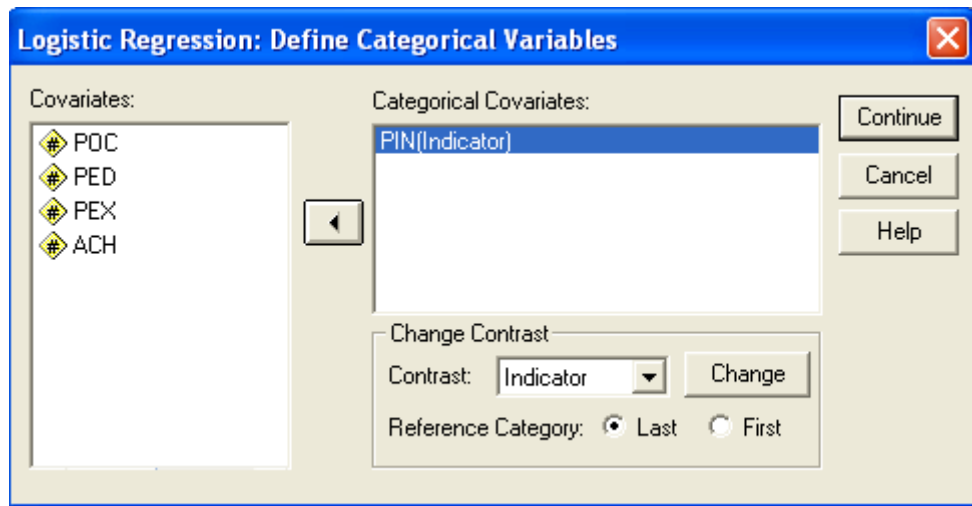
จะได้หน้าจอดังรูป



รูปที่ 1

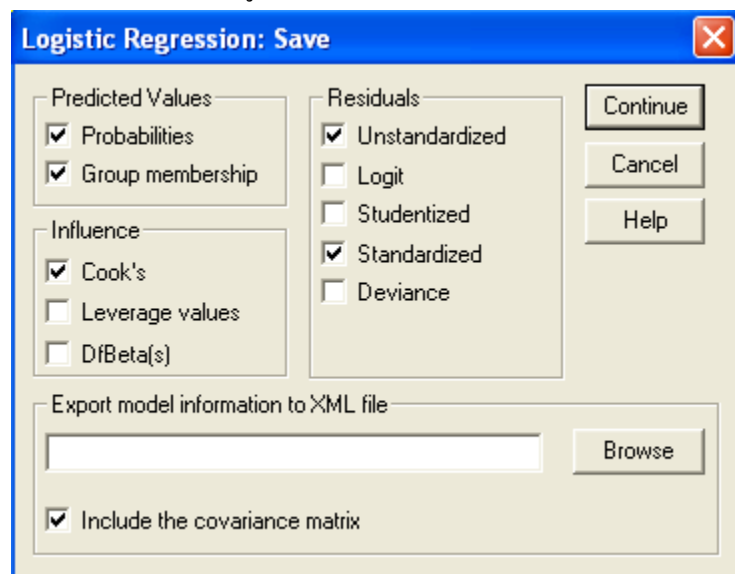
จากรูปข้างต้น จะต้องทำตามขั้นตอนดังต่อไปนี้

1. เลือกตัวแปร SCE ใส่ใน box ของ Dependent
2. เลือกตัวแปร POC, PED, PEX ACH, และ PIN ใส่ใน box Co
3. ในส่วนของ method เลือก Enter
4. คลิกปุ่ม **Categorical...** จะได้หน้าดังรูปที่ 2



รูปที่ 2

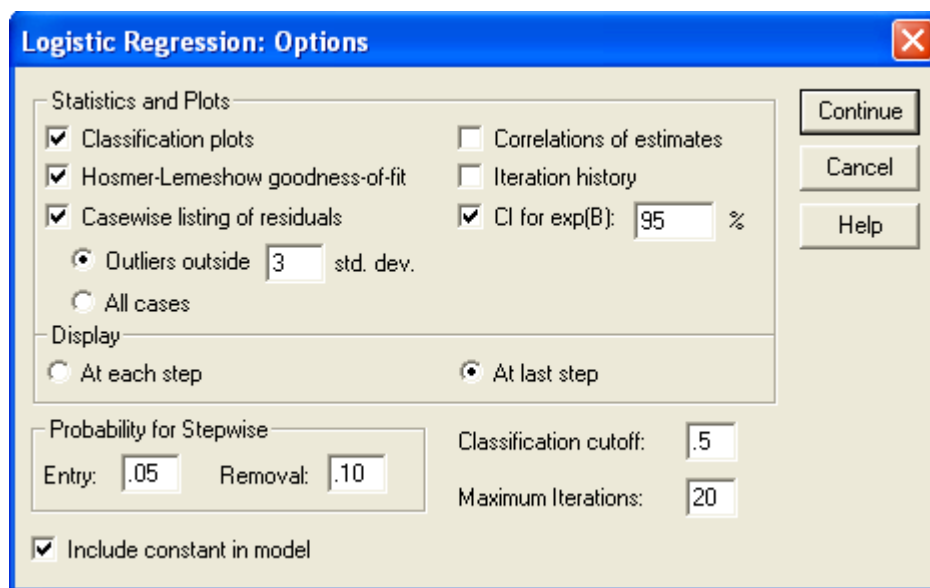
- ในรูปที่ 2 เป็นการระบุตัวแปรเชิงกลุ่ม ในที่นี้เลือกตัวแปร PIN ใส่ใน box ของ Categorical Covariates เนื่องจากตัวแปร PIN เป็นตัวแปรเชิงกลุ่ม
 - คลิก **Continue** กลับไปหน้าจอรูปที่ 1
5. คลิกปุ่ม **Save...** จะได้หน้าจอรูปที่ 3



รูปที่ 3

- ในส่วนของ Predicted Value เลือก Probabilities และ Group Membership
- ในส่วนของ Residuals Unstandardized และ Standardized
- ในส่วนของ Influences เลือก Cook's
- คลิก **Continue** กลับไปหน้าจอรูปที่ 1

- คลิกปุ่ม **Options...** จะได้น้ำจอรูปร่างที่ 4



รูปที่ 4

- เลือก Classification plots, Hosmer-Lemeshow goodness of-fit และ Casewise listing of residuals เป็น **3** std. dev.
 - เลือก CI for exp (B): 95%
 - ในส่วนของ Display เลือก At last step
7. คลิก **OK** จะได้ผลลัพธ์ดังแสดงในตารางที่ 1-11

ตารางที่ 1

Dependent Variable Encoding

Original Value	Internal Value
no	0
study	1

ความหมายของผลลัพธ์ตารางที่ 1

เป็นการระบุค่าของตัวแปร SCE เป็นตัวแปรตาม โดยกำหนดให้

$$SCE = \begin{cases} 1 & \text{ถ้าคาดหวังว่าจะศึกษาต่อ} \\ 0 & \text{ถ้าคาดหวังว่าจะไม่ศึกษาต่อ} \end{cases}$$

ตารางที่ 2

Categorical Variables Codings

	Frequency	Parameter coding		
		(1)	(2)	(3)
PIN > 60000	32	1.000	.000	.000
60001-80000	46	.000	1.000	.000
80001-100000	13	.000	.000	1.000
<100000	9	.000	.000	.000

ความหมายของผลลัพธ์ตารางที่ 2

เป็นการระบุว่าตัวแปรอิสระที่เป็นตัวแปรเชิงกลุ่ม (Categorical Variables) มี 1 ตัว คือ PIN โดยกำหนดว่า

$$PIN = \begin{cases} 1\ 0\ 0 & \text{ถ้ามีรายได้น้อยกว่า 60,000 บาท} \\ 0\ 1\ 0 & \text{ถ้ามีรายได้ 60,001-80,000 บาท} \\ 0\ 0\ 1 & \text{ถ้ามีรายได้ 80,001-100,000 บาท} \\ 0\ 0\ 0 & \text{ถ้ามีรายได้มากกว่า 100,000 บาท} \end{cases}$$

Block 0: Beginning Block

ตารางที่ 3

Classification Table^{a,b}

Observed	Predicted			
	SCE		Percentage Correct	
	no	study		
Step 0 SCE	no	0	41	.0
	study	0	59	100.0
Overall Percentage				59.0

a. Constant is included in the model.

b. The cut value is .500

ตารางที่ 4

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	.364	.203	3.204	1	.073	1.439

ตารางที่ 5

Variables not in the Equation

Step	Variables	Score	df	Sig.
0	POC	14.044	1	.000
	PED	14.183	1	.000
	PEX	52.372	1	.000
	ACH	43.441	1	.000
	PIN	30.891	3	.000
	PIN(1)	22.489	1	.000
	PIN(2)	.576	1	.448
	PIN(3)	10.384	1	.001
Overall Statistics		61.117	7	.000

ความหมายของผลลัพธ์ตารางที่ 3-5

ตารางที่ 7-8 แสดงรายละเอียดใน Block 0 หรือ Beginning Block หรือเรียกว่า Step 0 ซึ่งหมายถึงยังไม่ได้นำตัวแปรอิสระทั้ง 7 ตัวเข้าสมการ ดังนั้นในสมการจึงมีเฉพาะค่าคงที่

สำหรับความหมายของตารางที่ 7 เป็นการตรวจสอบความเชื่อเชื่อถือได้ในการพยากรณ์ของ Step 0 ซึ่งหมายถึงมีเฉพาะค่าคงที่จะพบว่า จากข้อมูลจริงความคาดหวังในการศึกษาต่อแบบ

Competitive 59 ฉบับ และเมื่อใช้สมการที่มีเฉพาะค่าคงที่พยากรณ์ จะพยากรณ์ว่าความคาดหวังดังกล่าว พยากรณ์ถูก 100% ในขณะที่ความคาดหวังในการศึกษาต่อแบบ Fixed มีจำนวน 41 ฉบับ แต่เมื่อ พยากรณ์โดยใช้สมการที่มีเฉพาะค่าคงที่ จะพยากรณ์ความคาดหวังทั้ง 50 ฉบับ หมายความว่าพยากรณ์ ถูก 0% จึงได้

$$\text{เปอร์เซ็นต์เฉลี่ยของการพยากรณ์ถูก} = \frac{59}{100}(100) = 59\%$$

Block 1: Method = Enter

ความหมายของ Block 1: Method = Enter หมายถึง จาก Block 0 หรือ Step 0 จะมีเฉพาะ ค่าคงที่ แต่ใน Block 1 จะนำตัวแปรเข้าสมการเพื่อใช้พยากรณ์โอกาสที่เหตุการณ์จะเกิด 7 ตัว คือ POC, PED, PEX, ACH, PIN (1), PIN (2) และ PIN (3)

ตารางที่ 6

Model Summary

Step	-2 Log ¹ likelihood	Cox & Snell R Square ²	Nagelkerke R Square ³
1	50.397 ^a	.572	.772

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

ความหมายของ ผลลัพธ์ตารางที่ 6

ผลลัพธ์	ความหมาย
1	เป็นการทดสอบความเหมาะสมของ Model (Goodness of fit) นั่นคือ -2Log Likelihood (-2LL) มีค่าเท่ากับ 50.397
2	ค่า Cox & Snell R ² = .572
3	สมการที่สร้างขึ้นมีค่า R ² เท่ากับ .772 หมายความว่า สมการนี้สามารถอธิบายความผันแปรของตัวแปรตามได้ถึงร้อยละ 77.2

ตารางที่ 7

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.560	7	.286

ตารางที่ 8

Contingency Table for Hosmer and Lemeshow Test

		SCE = no		SCE = study		Total
		Observed	Expected	Observed	Expected	
Step 1	1	11	10.925	0	.075	11
	2	11	10.482	0	.518	11
	3	9	9.043	2	1.957	11
	4	4	5.958	6	4.042	10
	5	6	2.941	4	7.059	10
	6	0	1.403	12	10.597	12
	7	0	.242	11	10.758	11
	8	0	.005	10	9.995	10
	9	0	.000	14	14.000	14

ความหมายของตารางที่ 7-8

Hosmer and Lemeshow ใช้ทดสอบความเหมาะสมของ Model

$$P(\text{Fixedcontract}) = \frac{1}{1 + e^{-w}}$$

โดยที่ $w = \beta_0 + \beta_1 POC + \beta_2 PEX + \beta_3 PEX + \beta_4 ACH + \beta_5 PIN(1) + \beta_6 PIN(2) + \beta_7 PIN(3)$

สมมติฐานของการทดสอบ คือ

H_0 : เหมาะสม

H_1 : ไม่เหมาะสม

จากค่าทดสอบ Chi-Square ในตารางที่ 11 ได้ค่า Chi-Square เป็น 8.560 และค่า Significance = .286 ซึ่ง $> .05$ จึงสรุปได้ว่า Model เหมาะสม

ตารางที่ 9

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	84.974	7	.000
	Block	84.974	7	.000
	Model	84.974	7	.000

ความหมายของตารางที่ 9

ใน Step 1 จะให้ค่าสถิติทดสอบไคสแควร์ 3 ค่า คือค่า Chi-Square ของ Model, Block และ Step ตามลำดับ

Model: ค่า Model Chi-Square

*ใช้ทดสอบสมมติฐาน

H_0 : โอกาสที่จะศึกษาต่อไม่ขึ้นกับตัวแปรอิสระทั้ง 7 ตัว

H_1 : โอกาสที่จะศึกษาต่อขึ้นกับตัวแปรอิสระอย่างน้อย 1 ตัว

*สถิติทดสอบ Model Chi-Square = 84.974 และค่า Sig. = .000 นั่นคือ ปฏิเสธ H_0 แสดงว่า โอกาสที่จะศึกษาต่อขึ้นกับตัวแปรอิสระอย่างน้อย 1 ตัว

ตารางที่ 10

Classification Table^a

Observed			Predicted		Percentage Correct
			SCE		
			no	study	
Step 1	SCE	no	33	8	① 80.5
		study	8	51	② 86.4
	Overall Percentage				③ 84.0

a. The cut value is .500

ความหมายของตารางที่ 10

ผลลัพธ์	ความหมาย
❶	จะพบว่าข้อมูลจริงของความคาดหวังในการไม่ศึกษาต่อ จำนวน 41 ฉบับ แต่เมื่อใช้สมการจะสามารถพยากรณ์ จำนวน 33 ฉบับ นั่นคือ พยากรณ์ถูกต้อง 80.5%
❷	จะพบว่าข้อมูลจริงของความคาดหวังในการศึกษาต่อ จำนวน 59 ฉบับ แต่เมื่อใช้สมการจะสามารถพยากรณ์ จำนวน 51 ฉบับ นั่นคือ พยากรณ์ถูกต้อง 86.4%
❸	เปอร์เซ็นต์รวมของการพยากรณ์ถูกต้อง คือ 84.0%

//

ตารางที่ 11

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)		
							Lower	Upper	
Step 1	POC	-.841	1.221	.474	1	.491	.431	.039	4.719
	PED	-.163	.148	1.211	1	.271	.849	.635	1.136
	PEX	.663	.182	13.275	1	.000	1.940	1.358	2.771
	ACH	2.156	.941	5.247	1	.022	8.632	1.365	54.592
	PIN			4.995	3	.172			
	PIN(1)	-19.377	12960.891	.000	1	.999	.000	.000	.
	PIN(2)	-17.381	12960.891	.000	1	.999	.000	.000	.
	PIN(3)	1.133	16400.938	.000	1	1.000	3.105	.000	.
	Constant	5.172	12960.892	.000	1	1.000	176.331		

a. Variable(s) entered on step 1: POC, PED, PEX, ACH, PIN.

จากตารางที่ 11 แสดงว่า มีเพียงค่าสัมประสิทธิ์ (B) ของ PEX และ ACH ที่สามารถพยากรณ์ได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

$$Z = 5.172 + .663PEX + 2.156ACH$$

อย่างไรก็ตามสามารถเขียนเป็นสมการในการพยากรณ์ความคาดหวังในการศึกษาต่อ จากตัวแปรทั้งหมดได้ดังนี้

$$Z = 5.172 - .841POC - .163PED + .663PEX + 2.156ACH - 19.377PIN(1) - 17.381PIN(2) + 1.133PIN(3)$$

จากค่าสัมประสิทธิ์ที่ได้ เราสามารถเขียนสมการการถดถอยโลจิสติก ทำนายโอกาสของการศึกษาต่อ ได้ดังนี้

$$\text{Prob}(\text{การศึกษาต่อ}) = 1/1 + e^{-z}$$

$$\text{โดย } Z = 5.172 - .841POC - .163PED + .663PEX + 2.156ACH \\ - 19.377PIN(1) - 17.381PIN(2) + 1.133PIN(3)$$

ถ้าลองพยากรณ์การศึกษาต่อของนักเรียนที่ผู้ปกครองมีความมุ่งหวังที่จะให้บุตรได้รับการศึกษาขั้นสูงระดับปริญญาตรี เป็น 6 นักเรียนมีผลสัมฤทธิ์ทางการเรียนเป็น 3.50 และค่าตัวแปรที่เหลือเป็น 0 จะได้ว่า

$$Z = 5.172 + .663(6) + 2.156(3.50) = 14.71$$

$$\text{Prob}(\text{การศึกษาต่อ}) = 1/1 + e^{-(14.71)} = 0.99$$

จากผลการวิเคราะห์นี้เราพยากรณ์ว่า เขาน่าจะมีโอกาสได้รับการศึกษาต่อ โดยการพยากรณ์ยึดเกณฑ์ความน่าจะเป็น 0.5

ถ้าลองพยากรณ์การศึกษาต่อของนักเรียนที่ผู้ปกครองมีอาชีพเกษตรกร เป็น 1 ผลสัมฤทธิ์ทางการเรียนของนักเรียนเป็น 2.15 และรายได้ของผู้ปกครอง 45,000 บาท เป็น 1 และ และค่าตัวแปรที่เหลือเป็น 0 จะได้ว่า

$$Z = 5.172 - .841(1) + 2.156(2.45) - 19.377(1) = -9.76$$

$$\text{Prob}(\text{การศึกษาต่อ}) = 1/1 + e^{-(-9.76)} = 0.00005$$

จากผลการวิเคราะห์นี้เราพยากรณ์ว่า นักเรียนคนนี้น่าจะไม่ค่อยมีโอกาสได้รับการศึกษาต่อ โดยการพยากรณ์ยึดเกณฑ์ความน่าจะเป็น 0.5

บรรณานุกรม

กัลยา วานิชย์บัญชา. 2546. การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows. กรุงเทพฯ ฯ : ศูนย์ หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.

กัญญา วรรณสุข. 2540. การประยุกต์ใช้ในการวิเคราะห์การถดถอยโลจิสติกและการวิเคราะห์อิทธิพลในการศึกษาปัจจัยที่มีอิทธิพลต่อความคาดหวังในการศึกษาต่อของนักเรียนชั้นประถมศึกษาในเขตการศึกษา 11. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต สาขาวิจัยการศึกษา จุฬาลงกรณ์มหาวิทยาลัย.

สมหวัง พิธิยานุวัฒน์. 2541 รวมบทความทางวิธีวิทยาการวิจัย. กรุงเทพฯ ฯ : ศูนย์ หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.

สวัสดิ์ชัย ศรีพนมธนากร. 2548 .การวิเคราะห์การถดถอยโลจิสติก. (ออนไลน์). สืบค้นจาก [http:// http://www.thairenu.com/logistic.htm](http://www.thairenu.com/logistic.htm) [15 สิงหาคม 2549]