Veridian E-Journal, Silpakorn University
ISSN 1906 - 3431

การศึกษาการรายงานผลการวิเคราะห์ถดถอยโลจิสติคสำหรับงานวิจัย

The Study on the Report of the Logistic Regression Analysis Results in Research Work

Received:	*April*	*1, 2019*
Revised:	*July*	*28, 2019*
Accepted:	*July*	*30, 2019*

ธีระดา ภิญโญ (Terada Pinyo)*

บทคัดย่อ

การรายงานผลการวิเคราะห์ถดถอยโลจิสติคซึ่งเป็นกระบวนการวิจัยวิธีหนึ่งที่นิยมใช้สำหรับงานวิจัย มีวัตถุประสงค์เพื่อถ่ายทอดองค์ความรู้ให้เข้าใจวิธีการทุกขั้นตอน ตั้งแต่วัตถุประสงค์ของการใช้กระบวนการวิเคราะห์ข้อมูล การอ่านผล และการรายงานผล การวิเคราะห์ถดถอยโลจิสติคเป็นการวิเคราะห์เพื่อค้นหาตัวแปรอิสระหรือตัวแปรทำนายที่อธิบายโอกาสการเกิดเหตุการณ์ตามตัวแปรตามหรือตัวแปรเกณฑ์ และสร้างสมการถดถอยโลจิสติคทำนายโอกาสที่จะเกิดเหตุการณ์ที่ผู้วิจัยสนใจ โดยคัดเลือกตัวแปรทำนายที่เหมาะสมเพื่อทำให้เปอร์เซ็นต์ความถูกต้องของการทำนายมีค่าสูงสุด โดยมี 2 โมเดลขึ้นอยู่จำนวนตัวแปรเกณฑ์หรือตัวแปรตาม ได้แก่ โมเดลการวิเคราะห์ถดถอยโลจิสติคแบบทวิที่ตัวแปรเกณฑ์ มี 2 ค่า เช่นการเป็นโรคหัวใจและไม่เป็นโรคหัวใจ และโมเดลการวิเคราะห์ถดถอยโลจิสติคแบบพหุที่ตัวแปรเกณฑ์ มีมากกว่า 2 ค่าขึ้นไป เช่นขนาดสถานศึกษา แบ่งเป็น เล็ก กลาง ใหญ่ และใหญ่พิเศษ ส่วนตัวแปรอิสระอาจจะเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพหรืออาจจะมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพก็ได้ เนื่องจากการวิเคราะห์ถดถอยโลจิสติคเป็นการหาความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ที่ไม่มีความสัมพันธ์เชิงเส้น ดังนั้นเพื่อให้การสรุปรายงานผลการวิจัยมีความชัดเจนและง่ายต่อการทำความเข้าใจจึงจำเป็นต้องมีการปรับให้อยู่ในรูปแบบความสัมพันธ์เชิงเส้นในรูป log ของ odds ซึ่งเรียกว่า logit

คำสำคัญ : การวิเคราะห์ถดถอยโลจิสติค งานวิจัย

* รองศาสตราจารย์ (สาขาสถิติประยุกต์)
Associate Professor (applied statistics) , therada.pi@hotmail.co.th. ,086-7745266

Abstract

The Report of the Logistic Regression Analysis Results is one of the popular research processes used for the research work. It aimed to pass on the knowledge of understanding in all processes which consist of the purpose of use, data analysis processes, results reading and results report to the students. The logistic regression analysis for research work aims to explore independent variables or predictor variables, describing the probability of event affected by dependent variables or criterion variables; and to form regression logistic model to predict the probability mentioned. The predictor variables are properly selected to maximize the accuracy percentage. Two models are specified: binary logistic regression model, having two criterion variables, such as ones with heart disease and ones without heart problem; multiple logistic regression analysis model, having more than two criterion variables, such as size of schools (small, medium, large, and extra-large.) Regarding free variables, either quantitative variables or qualitative variables can be applied or both. Since the regression logistic analysis, investigating the relationship between the predictor variables and the criterion variables, has no linear relation, it is necessary to adjust the conclusion in order to bring about clarity and easy understanding by applying linear relation in the form of log Odds called logit.

Keywords: ตัวอย่าง *21st Century*, Globalization, Educational Management

บทนำ

การวิเคราะห์ถดถอยโลจิสติคเป็นเทคนิคการวิเคราะห์ความถดถอยอีกวิธีหนึ่งที่นักศึกษาระดับบัณฑิตศึกษาหรือนักวิจัยสามารถเลือกใช้ได้ในกรณีที่ข้อมูลที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพหรือตัวแปรเชิงกลุ่ม (qualitative variable) การวิเคราะห์นี้มีวัตถุประสงค์เพื่อทำนายว่าเหตุการณ์หนึ่งจะเกิดขึ้นได้หรือไม่ หรือมีโอกาสเกิดขึ้นได้มากน้อยเพียงใด ทำให้ผู้วิจัยบอกเหตุผลของการเกิดเหตุการณ์นั้นๆ อาทิ สาเหตุการเป็นโรคความดันโลหิตสูง การดื่มแอลกอฮอล์ การเลือกทำงาน เป็นต้น เป็นการวิเคราะห์ข้อมูลโดยการสร้างสมการเพื่อหาความสัมพันธ์ระหว่างตัวแปรตาม (dependent variable) หรือตัวแปรเกณฑ์ และตัวแปรอิสระ (independent variable) หรือตัวแปรทำนาย แล้วนำสมการถดถอยที่ได้ไปพยากรณ์ค่าตัวแปรตามเมื่อกำหนดค่าตัวแปรอิสระ โดยมีแนวคิดใกล้เคียงกับการวิเคราะห์ถดถอยพหุคูณแบบปกติ เพียงแต่สมการถดถอยแบบโลจิสติคไม่จำเป็นต้องมีเงื่อนไขเกี่ยวกับค่าความคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ และค่าความแปรปรวนของค่าความคลาดเคลื่อนต้องคงที่อันเนื่องจากตัวแปรตามเป็นตัวแปรเชิงคุณภาพมีลักษณะไม่ต่อเนื่อง ขนาดตัวอย่างที่ใช้ในงานวิจัยสำหรับการวิเคราะห์ถดถอยโลจิสติคควรมีค่ามากกว่าการวิเคราะห์ถดถอยแบบปกติ โดยควรมากกว่า 30p เมื่อ p คือจำนวนตัวแปรอิสระ

วัตถุประสงค์ของการวิเคราะห์ถดถอยโลจิสติคสำหรับงานวิจัย

1. เพื่อค้นหาตัวแปรอิสระหรือตัวแปรทำนายที่อธิบายโอกาสการเกิดเหตุการณ์ตามตัวแปรตามหรือตัวแปรเกณฑ์ และศึกษาระดับความสัมพันธ์ของตัวแปรทำนายแต่ละตัว

2. เพื่อสร้างสมการถดถอยโลจิสติคทำนายโอกาสที่จะเกิดเหตุการณ์ที่สนใจ โดยคัดเลือกตัวแปรทำนายที่เหมาะสมเพื่อทำให้เปอร์เซ็นต์ความถูกต้องของการทำนายมีค่าสูงสุด

ข้อตกลงเบื้องต้นของการวิเคราะห์ถดถอยโลจิสติคสำหรับงานวิจัย (กัลยา วานิชย์บัญชา,2551; เพชรน้อย สิงห์ช่างชัย,2549; ยุทธ ไกยวรรณ์,2555; Stevens,1996)

1. ตัวแปรอิสระหรือตัวแปรทำนายเป็นตัวแปรอยู่ในมาตรวัดระดับอันตรภาค (interval scale) ขึ้นไป แต่ถ้าเป็นตัวแปรเชิงกลุ่มให้เปลี่ยนเป็นตัวแปรหุ่น (dummy variable) ที่มีค่าเป็น 0 หรือ 1 เท่านั้น

2. ตัวแปรตามหรือตัวแปรเกณฑ์ กรณีมี 2 ค่าคือ 0 และ 1 เรียกว่าการวิเคราะห์ถดถอยโลจิสติคแบบทวิ (binary logistic regression) ส่วนกรณีแบ่งตัวแปรเกณฑ์มากกว่า 2 ค่าขึ้นไป เช่น 3 ค่า (1,2,3) เรียกว่า การวิเคราะห์ถดถอยโลจิสติคแบบพหุ (multinomial logistic regression)

3. ค่าเฉลี่ยของความคลาดเคลื่อนเป็นศูนย์ E(e) = 0 หรือไม่มีความสัมพันธ์กัน

4. ตัวแปรอิสระหรือตัวแปรทำนายทุกตัวไม่มีความสัมพันธ์เชิงพหุ หรือไม่เกิดความสัมพันธ์สูงมาก (multicollinearity) โดยพิจารณาจากเกณฑ์ของ Burns and Grove (1993) ใช้ค่าความสัมพันธ์ไม่เกิน 0.65 หรือใช้เกณฑ์ของ Stevens (1996) ใช้ค่าความสัมพันธ์ไม่เกิน 0.80

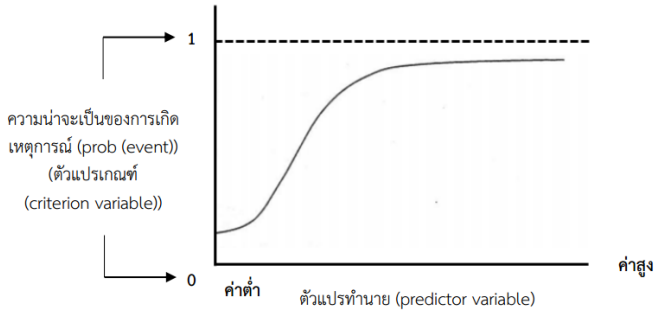
5. ขนาดของกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ถดถอยโลจิสติค ควรมากกว่าหรือเท่ากับ 30p ขึ้นไป เมื่อ p คือจำนวนตัวแปรอิสระหรือตัวแปรทำนาย

โมเดลของการวิเคราะห์ถดถอยโลจิสติคสำหรับงานวิจัย

แบ่งได้ 2 รูปแบบ ได้แก่

1. โมเดลการวิเคราะห์ถดถอยโลจิสติคแบบทวิ เป็นการวิเคราะห์ข้อมูลที่ตัวแปรตามหรือตัวแปรเกณฑ์ (y) มี 2 ค่า คือการไม่เกิดเหตุการณ์ (y=0) และการเกิดเหตุการณ์ (y=1) ซึ่งแบ่งได้เป็น 2 กรณี

1.1 กรณีที่ใช้ตัวแปรอิสระหรือตัวแปรทำนาย 1 ตัว เป็นการวิเคราะห์ถดถอยอย่างง่าย มีความสัมพันธ์ของตัวแปรอยู่ในรูปคล้ายตัว S ดังแสดงในภาพที่ 1



ภาพที่ 1 ฟังก์ชั่นโลจิสติค (logistic function)

โดยที่ $p(y) = \frac{1}{1+e^{-f(x)}} = \frac{1}{1+e^{-(b_0+b_1x)}} = \frac{e^{b_0+b_1x}}{1+e^{b_0+b_1x}}$

เมื่อ p(y) = ความน่าจะเป็นของการเกิดเหตุการณ์ y , e = 2.71828 , f(x) = ฟังชั่นของตัวแปรทำนาย

ให้ P(y) = ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ (y=1)

Q(y) = ความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ (y=0) = 1 - P(y)

$\therefore$ $P(y) = \frac{e^{b_0+b_1x}}{1+e^{b_0+b_1x}}$ และ $Q(y) = \frac{1}{1+e^{b_0+b_1x}}$

1.2 กรณีที่ใช้ตัวแปรอิสระหรือตัวแปรทำนายมากกว่า 1 ตัว (p ตัว) (กัลยา วานิชย์บัญชา,2551) จะได้ค่า

$P(y) = \frac{e^{b_0+b_1x_1+.....+b_px_p}}{1+e^{b_0+b_1x_1+.....+b_px_p}}$ และ $Q(y) = \frac{1}{1+e^{b_0+b_1x_1+.....+b_px_p}}$

2. โมเดลการวิเคราะห์ถดถอยโลจิสติคแบบพหุ (Multinomial logistic regression analysis model) เป็นการวิเคราะห์ข้อมูลที่ตัวแปรตามหรือตัวแปรเกณฑ์ (y) มีค่ามากกว่า 2 ค่า เช่น ระดับการศึกษา 3 ระดับ คือ ปริญญาตรี (y=1) ปริญญาโท (y=2) และปริญญาเอก (y=3) ซึ่งแบ่งได้เป็น 2 กรณีเช่นเดียวกัน กรณีใช้ตัวแปรอิสระหรือตัวแปรทำนาย 1 ตัว หรือ มากกว่า 1ตัวขึ้นไป

ขั้นตอนของการวิเคราะห์ถดถอยโลจิสติคสำหรับงานวิจัย

ประกอบด้วยขั้นตอนหลัก 3 ขั้นตอน

ขั้นที่ 1 ตรวจสอบความสัมพันธ์ของตัวแปรอิสระทุกตัวว่ามีความสัมพันธ์กันสูงหรือไม่ (Multicollinearity) สำหรับตัวแปรต่อเนื่อง ใช้คำสั่งของโปรแกรมสำเร็จรูปทางสถิติ Analyze ⟶ Correlate ⟶ Bivariate และตัวแปรเชิงกลุ่มให้ใช้คำสั่ง Analyze ⟶ Descriptive ⟶ Crosstabs โดยตัวแปรอิสระทุกตัวควรมีระดับความสัมพันธ์กันไม่เกิน 0.80 (Stevens,1996) ทำให้ตัวแปรอิสระทุกตัวในงานวิจัยไม่เกิดความสัมพันธ์เชิงพหุระหว่างปัจจัย จึงทำให้ผู้วิจัยสามารถใช้ตัวแปรอิสระเหล่านี้ในการวิเคราะห์ความถดถอยโลจิสติคประเภทไบนารีโลจิสติคได้ต่อไป

การเกิดปัญหา Multicollinearity มีผลทำให้ผลการทดสอบความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามจากการทดสอบ F และ t ขัดแย้งกัน ซึ่งทำให้สัมประสิทธิ์ความถดถอยเปลี่ยนไปเมื่อมีตัวแปรอิสระในสมการเพิ่มขึ้น ซึ่งการแก้ไขปัญหานี้อาจทำได้โดยเพิ่มจำนวนตัวอย่างให้มากขึ้น หรือตัดตัวแปรอิสระบางตัวออกไป หรือใช้วิธีการรวมตัวแปรที่มีความสัมพันธ์ไว้ด้วยกันด้วยวิธีการทางสถิติที่เรียกว่า การวิเคราะห์องค์ประกอบ (Factor Analysis)

ขั้นที่ 2 การวิเคราะห์ถดถอยโลจิสติค โดยใช้คำสั่ง Analyze ⟶ Regression ⟶ Binary Logistic โดยมีขั้นตอน 4 ขั้นตอนย่อย ดังนี้

(1) เลือกตัวแปรอิสระที่คาดว่าจะมีความสัมพันธ์กับตัวแปรตามเข้าสมการถดถอยโลจิสติค มีวิธีการเลือกได้ 3 วิธีหลัก ได้แก่

- **Enter Method** เป็นเทคนิคการเลือกตัวแปรอิสระ(ตัวแปรทำนาย)ทุกตัวเข้าสมการถดถอยโลจิสติคในขั้นตอนเดียว ผู้วิจัยต้องเป็นผู้ตัดสินใจเองว่า ตัวแปรอิสระตัว Significance ของสถิติทดสอบ ถ้ามีนัยสำคัญทางสถิติก็ถือว่าตัวแปรอิสระนั้นควรจะอยู่ในสมการ

- **Forward Method** เป็นเทคนิคการเลือกตัวแปรอิสระ(ตัวแปรทำนาย)ที่สามารถอธิบายความผันแปรของตัวแปรตามหรือตัวแปรเกณฑ์ได้สูงสุดและมีนัยสำคัญทางสถิติเข้าสมการก่อน จากนั้นจึงเลือกตัวแปรทำนายอันดับรองเข้าสมการ ตามลำดับ ทำเช่นนั้นไปเรื่อยๆไปจนกระทั่งไม่มีตัวแปรทำนายใดที่อธิบายความผันแปรของตัวแปรเกณฑ์อย่างมีนัยสำคัญทางสถิติอีกแล้ว วิธีนี้มี 3 แบบย่อยโดยแตกต่างที่ค่าสถิติที่ใช้ในการพิจารณา ได้แก่ (1) Forward : LR (likelihood ratio) พิจารณาจากอัตราส่วนความเป็นไปได้หรือการเปลี่ยนแปลงของ – 2LL (- 2 likelihood-ratio statistic (2) Forward : Wald พิจารณาจากค่าสถิติของ Wald และ (3) Forward : Conditional เหมือนกับ Forward : LR ทุกประการเพียงแต่ให้ใช้กับตัวอย่างที่มีขนาดเล็ก

- **Backward Method** เป็นเทคนิคการนำตัวแปรอิสระ(ตัวแปรทำนาย)ทุกตัวเข้าสมการพร้อมกัน จากนั้นจึงคัดตัวแปรทำนายออกมาทีละตัวโดยพิจารณาจากการที่ตัวแปรทำนายอธิบายความผันแปรของตัวแปรเกณฑ์ได้น้อยที่สุดออกจากสมการ ทำเช่นนี้ไปเรื่อยๆจนกระทั่งเหลือตัวแปรทำนายที่อธิบายความผันแปรของตัวแปรเกณฑ์ได้อย่างมีนัยสำคัญทางสถิติอีกแล้ว วิธีนี้มี 3 แบบย่อยโดยแตกต่างที่ค่าสถิติที่ใช้ในการพิจารณา ได้แก่ (1) Backward : LR (likelihood ratio) พิจารณาจากอัตราส่วนความเป็นไปได้หรือการเปลี่ยนแปลงของ – 2LL (- 2 likelihood-ratio statistic (2) Backward : Wald พิจารณาจากค่าสถิติของ Wald และ (3) Backward : Conditional เหมือนกับ Backward : LR ทุกประการเพียงแต่ให้ใช้กับตัวอย่างที่มีขนาดเล็ก

(2) ตรวจสอบค่าผิดปกติของตัวแปรทำนายแต่ละตัว (p ตัว) ทุกตัว ($x_1, x_2, x_3, x_4, \ldots\ldots, x_p$)

(3) สร้างสมการถดถอยโลจิสติค และตรวจสอบความเหมาะสมของสมการ ได้ 2 วิธี ดังนี้

(3.1) พิจารณาจากค่าความเป็นไปได้ (likelihood value) โดยวัดค่า -2LL(- 2 log likelihood) ถ้าค่า -2LLมีค่าน้อยหรือเข้าใกล้ศูนย์แสดงว่าสมการถดถอยโลจิสติคมีความเหมาะสม (ถ้าสมการมีความเหมาะสม 100%จะได้ค่า -2LL = 0) ซึ่งทดสอบความมีนัยสำคัญทางสถิติได้โดยใช้ Model Chi – Square ที่ df = จำนวนตัวแปรทำนายที่ใช้ในสมการ ถ้า χ^2 มีนัยสำคัญทางสถิติ แสดงว่าตัวแปรทำนายอย่างน้อย 1 ตัวร่วมกันทำนายโอกาสของการเกิดเหตุการณ์ที่สนใจ (y = 1) ด้วยความเชื่อมั่น (1 - α)100%

นอกจากนี้ยังมีค่า R^2 ของ Cox & Snell และ Nagelkerke ซึ่งเป็นค่าที่แสดงสัดส่วนหรือร้อยละของการที่ตัวแปรอิสระหรือตัวแปรทำนายทุกตัวในสมการสามารถอธิบายการเกิดเหตุการณ์ที่สนใจในสมการ คล้ายกับกับค่า R^2 ของการวิเคราะห์ถดถอยเชิงเส้น เช่น Cox & Snell R^2 (R^2_{cs}) = 0.584 หมายความว่า ตัวแปรอิสระ 7 ตัว สามารถอธิบายโอกาสที่จะเป็นโรคหัวใจได้ร้อยละ 58.4 และค่า Nagelkerke R^2 (R^2_N) = 0.795 หมายความว่า ตัวแปรอิสระ 7 ตัว สามารถอธิบายโอกาสที่จะเป็นโรคหัวใจได้ร้อยละ 79.5

โดย Cox & Snell R^2 (R^2_{cs}) = 1 - $\left[\frac{L(0)}{L(B)}\right]^{2/n}$ และ Nagelkerke R^2 (R^2_N) = $\frac{Cox\ \&\ SnellR^2}{Cox\ \&\ SnellR^2_{\max}}$

เมื่อ $R^2_{\max}$ = 1 - $[L(0)]^{2/n}$

L(0) = likelihood ของโมเดลที่มีเพียงค่าคงที่

L(B) = likelihood ของโมเดลที่มีตัวแปรอิสระตามที่เข้าในสมการ

ค่า R^2_{cs} มีค่าน้อยกว่า 1 และค่า R^2_N มีค่ามากกว่า R^2_{cs} เสมอ

Omnibus Tests of Model Coefficients

		Chi – square	df	Sig
Step 1	Step	23.399	3	0.000
	Block	23.399	3	0.000
	Model	23.399	3	0.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	456.488	0.065	0.087

จากตาราง output ข้างต้น Model Chi – Square ที่ df = 3 (จำนวนตัวทำนาย) และค่า χ^2 มีนัยสำคัญทางสถิติ แสดงว่าตัวแปรทำนายอย่างน้อย 1 ตัวร่วมกันทำนายโอกาสของการเกิดเหตุการณ์ที่สนใจ (y = 1) ด้วยความเชื่อมั่น 95% นอกจากนั้น Cox & Snell R^2 (R^2_{cs}) = 0.065 หมายความว่า ตัวแปรอิสระ 3 ตัว สามารถอธิบายโอกาสที่เกิดเหตุการณ์ที่สนใจได้ร้อยละ 6.5 และค่า Nagelkerke R^2 (R^2_N) = 0.087 หมายความว่า ตัวแปรอิสระ 3 ตัว สามารถอธิบายโอกาสที่เกิดเหตุการณ์ที่สนใจได้ร้อยละ 8.7

(3.2) พิจารณาจากสถิติทดสอบความเหมาะสมของ Hosmer and Lemeshow เมื่อค่า χ^2 ไม่มีนัยสำคัญทางสถิติ แสดงว่าโมเดลมีความเหมาะสม โดยผู้วิจัยตั้งสมมุติฐาน ดังนี้

H_o : สมการถดถอยโลจิสติคเหมาะสมในการพยากรณ์

H_1 : สมการถดถอยโลจิสติคไม่เหมาะสมในการพยากรณ์

Hosmer and Lemeshow Test

Step	Chi – square	df	Sig
1	13.119	8	0.108

จากตาราง output ข้างต้น ค่า χ^2 ไม่มีนัยสำคัญทางสถิติ (sig = .108) แสดงว่า สมการถดถอยโลจิสติคที่สร้างขึ้นมีความเหมาะสมในการพยากรณ์ข้อมูลต่อไป

Veridian E-Journal, Silpakorn University
ISSN 1906 - 3431

(4) ตรวจสอบค่านัยสำคัญทางสถิติของสัมประสิทธิ์ถดถอยโลจิสติคของตัวแปรทำนายแต่ละตัว มีวิธีการ 2 วิธี

(4.1) ใช้สถิติทดสอบของ Wald (Wald Statistics) ที่มีการแจกแจงแบบ χ^2 และ df = 1 ซึ่ง $W = \left[\frac{b_0}{SE(b_0)}\right]^2$ โดยผู้วิจัยตั้งสมมุติฐาน ดังนี้

H_o : ตัวแปรทำนายไม่มีผลต่อการเปลี่ยนแปลงของ odds ratio หรือ $\beta_i = 0$; I = 1,2,3, ... p

H_1 : ตัวแปรทำนายมีผลต่อการเปลี่ยนแปลงของ odds ratio หรือ $\beta_i \neq 0$

ซึ่งถ้าการทดสอบมีนัยสำคัญทางสถิติและค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรทำนายนั้นมีผลต่อการเพิ่มความน่าจะเป็นของการเกิดเหตุการณ์ แต่ถ้าค่าสัมประสิทธิ์เป็นลบ แสดงว่าตัวแปรทำนายนั้นมีผลต่อการลดความน่าจะเป็นของการเกิดเหตุการณ์

Variables in the Equation

		B	S.E	Wald	df	Sig	Exp(B)	95% C.I. for Exp(B)	
								Lower	Upper
Step	FEEL	-1.087	0.812	1.790	1	0.181	0.337	0.069	1.657
	EAT	-0.431	0.206	4.437	1	0.660	0.650	0.435	0.970
	GOOD	0.976	0.227	18.463	1	2.655	2.655	1.701	4.144
	Constant	-0.096	1.103	0.007	1	0.910	0.910		

จากตาราง output ข้างต้น ค่า Wald ของตัวแปร EAT และ GOOD มีนัยสำคัญทางสถิติ (sig = .036 และ .000) แสดงว่าตัวแปร EAT และ GOOD มีผลต่อการทำนายการเปลี่ยนแปลงของตัวแปรเกณฑ์ที่ระดับนัยสำคัญ .05

(4.2) ใช้สถิติทดสอบความเหมาะสมโดยพิจารณาจากค่า χ^2 ของ Step ,Block และ Model ต้องมีค่าเท่ากันทั้ง 3 ค่า และมีมีนัยสำคัญทางสถิติ แสดงว่าตัวแปรทำนายที่เพิ่มเข้าไปในสมการมีความเหมาะสม ดังแสดงในตาราง 0utput ข้างล่าง

Omnibus Tests of Model Coefficients

		Chi – square	df	Sig
Step 1	Step	23.399	3	0.000
	Block	23.399	3	0.000
	Model	23.399	3	0.000

ขั้นที่ 3 ขั้นสรุป เนื่องจากการวิเคราะห์ถดถอยโลจิสติคเป็นการหาความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ที่ไม่เป็นความสัมพันธ์เชิงเส้น ดังนั้นในการสรุปให้เข้าใจง่ายขึ้นจึงจำเป็นต้องมีการปรับให้อยู่ในรูปแบบเชิงเส้นในรูปของ odds ดังนั้นการเขียนสมการถดถอยโลจิสติคจึงเขียนอยู่ในรูป log ของ odds ซึ่งเรียกว่า logit

Odds หรือ odd ratio หมายถึงอัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจ (y = 1) กับ โอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ (y = 0)

Odds = $\frac{P_y}{Q_y}$ เมื่อ P_y = โอกาสที่จะเกิดเหตุการณ์ที่สนใจ และ Q_y = โอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

$$= \frac{P(Y=1)}{P(y=0)} = \frac{p}{1-p} = \frac{\frac{e^{\beta_0+\beta_1x_1+.....+\beta_px_p}}{1+e^{\beta_0+\beta_1x_1+....+\beta_px_p}}}{\frac{1}{1+e^{\beta_0+\beta_1x_1+.....+\beta_px_p}}} = e^{\beta_0+\beta_1x_1+.....+\beta_px_p}$$

$Log_e(OR) = ln(OR) = \beta_0+\beta_1x_1+ .....+\beta_px_p$

หรือ $log_e(\hat{OR}) = b_0 + b_1x_1 + ... + b_px_p$ สำหรับกลุ่มตัวอย่าง

ค่าของ Odds จะแสดงถึงโอกาสที่จะเกิดเหตุการณ์ที่สนใจเป็นกี่เท่าของโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

- ถ้า OR หรือ Odds เท่ากับ 1 ($b_0 = 0$ ทำให้ $e^{b_0} = e^0 = 1$) หมายถึง โอกาสที่จะเกิดเหตุการณ์ที่สนใจเท่ากับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

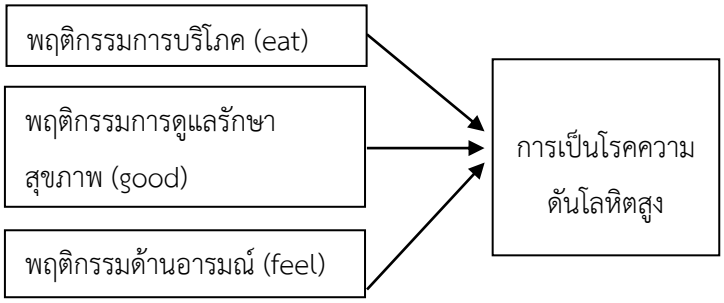
- ถ้า OR หรือ Odds Odds มีค่าน้อยกว่า 1 ($b_1 < 0$ ทำให้ $e^{b_1} < 1$) แสดงว่าโอกาสที่จะเกิดเหตุการณ์ที่สนใจน้อยกว่าโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

-แต่ถ้า OR หรือ Odds Odds มีค่ามากกว่า 1 ($b_1 > 0$ ทำให้ $e^{b_1} > 1$) แสดงว่าโอกาสที่จะเกิดเหตุการณ์ที่สนใจนั้นมากกว่าโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

6. ตัวอย่างรายงานการวิเคราะห์ถดถอยโลจิสติคของงานวิจัย

ตัวอย่างที่ 1 กรณีตัวแปรทำนายทุกตัวเป็นตัวแปรต่อเนื่อง

ผู้วิจัยศึกษาปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคความดันโลหิตสูงของผู้ป่วยนอก (ศุภกิจ ขจรกิตติมศักดิ์ และธีระดา ภิญโญ, 2553) จำนวน 3 ตัวแปร กำหนดกรอบแนวคิดการวิจัย และรายงานผลดังตาราง 1- 2



ตาราง 1 ผลการตรวจสอบปัจจัยเสี่ยงทุกปัจจัยที่มีอิทธิพลต่อการเป็นโรคความดันโลหิตสูงของผู้ป่วยนอก

ปัจจัย	Constant	β	S.E.	Exp(β)	χ^2	df	P-value	Cox & Snell R Square
1.ปัจจัยพฤติกรรมการบริโภค	-0.637	0.886	2.425	0.22	17.461*	1	0	0.049
2. ปัจจัยพฤติกรรมการดูแลรักษาสุขภาพ	-1.062	1.293	3.642	0.353	15.303*	1	0	0.043
3. ปัจจัยพฤติกรรมด้านอารมณ์	0.137	-0.061	0.941	0.638	0.009	1	0.924	0

*P < .05

จากตาราง 1 ปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคความดันโลหิตสูงของผู้ป่วยนอก คือ ปัจจัยพฤติกรรมการบริโภค (χ^2 = 17.461, p - value = .000) และปัจจัยพฤติกรรมการดูแลรักษาสุขภาพ (χ^2 = 15.303 , p - value = .000) ส่วนปัจจัยเสี่ยงที่ไม่มีอิทธิพลต่อการเป็นโรคความดันโลหิตสูง คือ ปัจจัยพฤติกรรมด้านอารมณ์ (χ^2 = .009, p - value = .924)

จากการตรวจสอบระดับอิทธิพลของปัจจัยเสี่ยง พบว่า ปัจจัยพฤติกรรมการบริโภค และปัจจัยพฤติกรรมการดูแลรักษาสุขภาพ มีความสัมพันธ์ทางบวกต่อการเป็นโรคความดันโลหิตสูง (β = .886 และ 1.293) ตามลำดับ โดยปัจจัยพฤติกรรมการบริโภคเป็นปัจจัยที่มีอิทธิพลสูงสุดต่อการเป็นโรคความดันโลหิตสูง (Cox & Snell R^2 = 0.049) และปัจจัยพฤติกรรมการดูแลรักษาสุขภาพ (Cox & Snell R^2 = 0.043) ตามลำดับ

ตาราง 2 การวิเคราะห์ตัวแบบที่เหมาะสมของปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคความดันโลหิตสูงของผู้ป่วยนอก

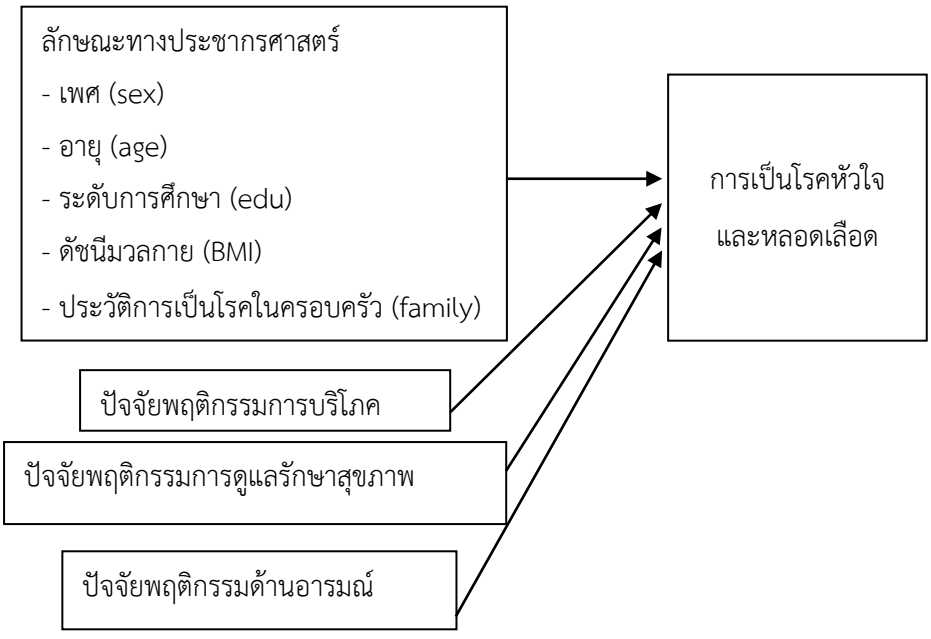
ปัจจัย	B	S.E.	Exp (β)	P - value
ปัจจัยพฤติกรรมการบริโภค (Eat)	.707	.238	2.029	.003
ปัจจัยพฤติกรรมการดูแลรักษาสุขภาพ (Good)	1.403	.382	4.068	.000
ค่าคงที่ (Constant)	-2.819	.789	.060	.000

จากตาราง 2 ตัวแบบของปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคความดันโลหิตสูงของผู้ป่วยนอก ได้แก่ Log (odds) = -2.819 +.707 (Eat) + 1.403(Good) โดยที่ log (odds) = $\log\left[\frac{p}{1-p}\right]$ เมื่อ p คือ โอกาสหรือความเสี่ยงของการเป็นโรคความดันโลหิตสูง และ 1–p คือ โอกาสหรือความเสี่ยงที่จะไม่เป็นโรคความดันโลหิตสูง

เมื่อพิจารณาแต่ละปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคความดันโลหิตสูงของผู้ป่วยนอก พบว่า ผู้ป่วยนอกที่มีปัจจัยพฤติกรรมการดูแลรักษาสุขภาพเพิ่มขึ้น 1 หน่วย มีความเสี่ยงต่อการเป็นโรคความดันโลหิตสูงเพิ่มขึ้น 4.1 เท่า (Exp (β) = 4.068) และผู้ป่วยนอกที่มีปัจจัยพฤติกรรมการบริโภคเพิ่มขึ้น 1 หน่วย มีความเสี่ยงต่อการเป็นโรคความดันโลหิตสูงเพิ่มขึ้น 2.0 เท่า (Exp (β) = 2.029)

ตัวอย่างที่ 2 กรณีตัวแปรทำนายเป็นตัวแปรต่อเนื่องและไม่ต่อเนื่อง

ผู้วิจัยศึกษาปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก (ตรีนุช นุ่มสีหมอก และธีระดา ภิญโญ, 2553).โดยกำหนดกรอบแนวคิดในการวิจัยและรายงานผลดังตาราง 3– 5



ตาราง 3 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างปัจจัยเสี่ยงทุกปัจจัยที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือด

ปัจจัยเสี่ยง	age	BMI	eat	good	feel	sex	edu	Family
1.age	1.000							
2.BMI	.205◆**	1.000						
3.eat	-.070◆	.026◆	1.000					
4.good	-144◆**	.134◆*	.191◆**	1.000				
5.feel	.104◆	.048◆	-.028◆	.058◆	1.000			
6. sex	.379●	.768●	.102●	.103●	.294●	1.000		
7. edu	.421●	.794●	.157●	.113●	.324●	.095●	1.000	
8.Family	.374●	.779●	.098●	.063●	.301●	.011●	.141●	1.000

* P < .05, ** P< .01

เมื่อ ◆ = สถิติค่าสหสัมพันธ์อย่างง่ายเพียร์สัน (Pearson Product Moment Correlation Coefficient)
● = สถิติค่าสัมประสิทธิ์คราเมอร์สวี (Creamer's V coefficient)

จากตาราง 3 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก พบว่า ความสัมพันธ์ระหว่างปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอกมีระดับความสัมพันธ์ไม่เกิน .80 (Stevens,1996) งานวิจัยนี้จึงไม่มีปัญหาตัวแปรอิสระมีความสัมพันธ์กันสูงมาก (multicolinearity) ดังนั้นในการศึกษาครั้งนี้จึงไม่จำเป็นต้องตัดตัวแปรอิสระออกไปจากการวิเคราะห์ตัวแบบความสัมพันธ์

ตาราง 4 ผลการวิเคราะห์การตรวจสอบปัจจัยเสี่ยงทุกปัจจัยที่ส่งผลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก

ปัจจัย	Constant	β	S.E.	Exp(β)	χ^2	df	P-value	Cox & Snell R Square
1.เพศ(sex)	-	-	-	-	-	-	-	-
- ชาย	-.975	.143	.270	1.154	.282	1	.596	.001
- หญิง(ตัวแปรอ้างอิง)	=	=	=	=	=	=	=	=
2.อายุ(Age)	-.969	-.081	.013	.922	49.631*	1	.000	.143

ตาราง 4 (ต่อ)

ปัจจัย	Constant	β	S.E.	Exp(β)	χ^2	df	P-value	Cox & Snell R Square
3.ระดับการศึกษา	-	-	-	-	-	-	-	-
- ไม่ได้รับการศึกษา (Edu1)	-.224	- 1.607	.613	.200	10.293*	1	.009	.032
- ต่ำกว่าปริญญาตรี (Edu2)	-.625	- 1.756	.290	.173	40.673*	1	.000	.119
- สูงกว่าปริญญาตรี(ตัวแปรอ้างอิง)	=	=	=	=	=	=	=	=
4.ดัชนีมวลกาย(BMI)	-.980	-.147	.034	.863	21.242*	1	.000	.064
5.ประวัติโรคหัวใจและหลอดเลือดในครอบครัว (history)	-	-	-	-	-	-	=	-
- มี	-.656	-.352	.270	.703	1.719	1	.192	.005
- ไม่มี	=	=	=	=	=	=	=	=
6.ปัจจัยพฤติกรรมการบริโภค(Eat)	-.977	.624	.247	1.867	6.427*	1	.011	.020
7.ปัจจัยพฤติกรรมการดูแลรักษาสุขภาพ (Good)	-.933	-.850	.399	.427	4.576*	1	.033	.014
8.ปัจจัยพฤติกรรมด้านอารมณ์(Feel)	-.992	-.066	.029	.936	5.446*	1	.024	.017

*P < .05

จากตาราง 4 ผลการวิเคราะห์การตรวจสอบปัจจัยเสี่ยงทุกปัจจัยที่ส่งผลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก พบว่า ปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือด คือ อายุ (χ^2 = 49.631, p-value =.000) ระดับการศึกษาต่ำกว่าปริญญาตรี (χ^2 = 40.673, p-value=.000) ดัชนีมวลกาย (χ^2 =21.242,p-value=.000) ไม่ได้รับการศึกษา (χ^2 = 10.293, p-value=.009) ปัจจัยพฤติกรรมการบริโภค (χ^2 = 6.427, p-value =.011) ปัจจัยพฤติกรรมด้านอารมณ์ (χ^2 =5.446, p-value=.024) และปัจจัยพฤติกรรมการดูแลรักษาสุขภาพ (χ^2 = 4.576, p-value=.033) ตามลำดับ ส่วนปัจจัยเสี่ยงที่ไม่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือด คือ ประวัติการเป็นโรคหัวใจและหลอดเลือดในครอบครัว (χ^2 = 1.719, p-value=.192) และเพศ(χ^2 = .282, p-value=.596)

จากการตรวจสอบระดับอิทธิพลของปัจจัยเสี่ยงสูงสุดที่ส่งผลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก พบว่า ปัจจัยพฤติกรรมการบริโภค (β = .624) มีความสัมพันธ์ทางบวกส่วนปัจจัยที่มีระดับอิทธิพลทางลบต่อการเกิดโรคหัวใจและหลอดเลือด ได้แก่ อายุ (β = -.081) รองลงมา ระดับการศึกษาต่ำกว่าปริญญาตรี (β = .-1.756) ที่ไม่ได้รับการศึกษา (β = -1.607) ดัชนีมวลกาย (β = -.147) ปัจจัยพฤติกรรมด้านอารมณ์ (β = -.066) และปัจจัยพฤติกรรมการดูแลสุขภาพ (β = -.850) ตามลำดับ

ตาราง 5 ผลการวิเคราะห์ตัวแบบปัจจัยเสี่ยงที่ส่งผลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก

ปัจจัย	B	S.E.	Exp (β)
อายุ(Age)	-.060	.014	.942
ระดับการศึกษา	-	-	-
- ไม่ได้รับการศึกษา(Edu1)	-3.001	.677	.050
- ต่ำกว่าปริญญาตรี(Edu2)	-2.287	.359	.102
ดัชนีมวลกาย(BMI)	-.072	.040	.930
ปัจจัยพฤติกรรมการบริโภค(Eat)	.207	.318	1.230
ปัจจัยพฤติกรรมการดูแลรักษาสุขภาพ (Good)	-.801	.515	.449
ปัจจัยพฤติกรรมด้านอารมณ์(Feel)	-.024	.032	.976
Constant	6.559	1.905	705.654

จากตาราง 5 ผลการวิเคราะห์ตัวแบบของปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก พบว่า ตัวแบบของปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือด คือ

Log(odds) = 6.559-.060(Age)–3.001(Edu1)-2.287(Edu2)-.072(BMI) +.207(Eat)-.801(Good)-.024(Feel)

ตัวแบบของปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอกที่ไม่ได้รับการศึกษา ได้แก่ Log (odds) = 3.558-.060(Age)-.072(BMI)+. 207(Eat)-.801(Good)-.024(Feel)

ตัวแบบของปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก ระดับการศึกษาต่ำกว่าปริญญาตรี ได้แก่Log(odds)= 4.272 -.060(Age)- .072(BMI)+.207(Eat) -.801(Good)-.024(Feel)

ตัวแบบของปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอกระดับการศึกษาสูงกว่าปริญญาตรี ได้แก่ Log(odds)= 6.559 -.060(Age)- .072(BMI)+.207(Eat) -.801(Good)-.024(Feel)

เมื่อ Log (odds) $= \log\left[\frac{p}{1-p}\right]$ ซึ่ง p คือ โอกาสการเป็นโรคหัวใจและหลอดเลือด และ 1 – p คือ โอกาสที่ไม่เป็นโรคหัวใจและหลอดเลือด

เมื่อพิจารณาแต่ละปัจจัยเสี่ยงที่มีอิทธิพลต่อการเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอก พบว่า ผู้ป่วยนอกที่มีปัจจัยพฤติกรรมการบริโภคเพิ่มขึ้น 1 หน่วย มีความเสี่ยงต่อการเป็นโรคหัวใจและหลอดเลือดเพิ่มขึ้น 1.2 เท่า (Exp (β) = 1.230) รองลงมา ผู้ป่วยนอกที่มีปัจจัยพฤติกรรมด้านอารมณ์เพิ่มขึ้น 1 หน่วยมีความเสี่ยงต่อการเป็นโรคหัวใจและหลอดเลือดลดลง .976 เท่า (Exp (β) = .976) ผู้ป่วยนอกที่มีอายุของผู้ป่วยนอกเพิ่มขึ้น 1 ปี มีความเสี่ยงต่อการเป็นโรคหัวใจและหลอดเลือดลดลง .942 เท่า (Exp (β) = .942) ผู้ป่วยนอกที่มีดัชนีมวลกายเพิ่มขึ้น 1 หน่วย มีความเสี่ยงต่อการเป็นโรคหัวใจและหลอดเลือดลดลง .930 เท่า (Exp (β) = .930) และผู้ป่วยนอกที่มีปัจจัยพฤติกรรมการดูแลรักษาสุขภาพเพิ่มขึ้น 1 หน่วย มีความเสี่ยงต่อการเป็นโรคหัวใจและหลอดเลือดลดลง .449 เท่า (Exp (β) = .449)

โอกาสหรือความเสี่ยงที่จะเป็นโรคหัวใจและหลอดเลือดของผู้ป่วยนอกที่ไม่ได้รับการศึกษาและต่ำกว่าปริญญาตรีลดลง .050 เท่า และ .102 เท่า ของผู้ป่วยนอกที่มีระดับการศึกษาสูงกว่าปริญญาตรี

บทสรุป

การวิเคราะห์ถดถอยโลจิสติค เป็นวิธีการทางสถิติที่ใช้วิเคราะห์ข้อมูลเชิงปริมาณ เพื่อทำนายว่าเหตุการณ์หนึ่งๆจะเกิดขึ้นได้หรือไม่ เกิดขึ้นได้มากน้อยเพียงใด รวมทั้งบอกตัวแปรทำนายที่ส่งผลให้เกิดเหตุการณ์นั้นด้วย เป็นวิธีการที่ไม่มีข้อกำหนดเกี่ยวกับตัวแปรตามและค่าความคลาดเคลื่อนที่ต้องมีการแจกแจงแบบปกติ ไม่ต้องมีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรทำนายและตัวแปรเกณฑ์ และสามารถใช้วิเคราะห์ข้อมูลได้ทุกระดับการวัด มีนักวิจัยนำวิธีการวิเคราะห์แบบนี้ไปใช้ประโยชน์ในงานด้านต่างๆได้ อาทิ งานทางการแพทย์ งานด้านการวางแผน งานด้านการบริหารจัดการ เป็นต้น การนำเสนอรายงานครั้งนี้เพื่อทำให้นักวิจัยมีความรู้ความเข้าใจและนำไปใช้ประโยชน์ในการใช้วางแผนการวิจัยได้อีกทางเลือกหนึ่ง

References

Kaiyawan,Y.(2012). "lakkān læ kānchai kān wikhroʿ kān thotthoj̄ lochittik samrap kānwichai". [Principle and Using Logistic Regression Analysis for Research]. Rajamangda University of Technology Srivijaya Research Journal 4(1) p. 1-12

Khajornkittimasak,S., & Pinyo, T. (2010). "patchai thī mī 'itthiphon toʿ kān pen rok khwām dan lohit sung khoʿng phupuainoʿk".[Rick Factors Influencing High Blood Pressure Diseases to Outpatients]. (Research Project). Faculty of Science and Technology, Suan Sunandha Rajabhat University

Numsimok,T.,& Pinyo,T. (2010). "patchai thī mī 'itthiphon toʿ kān pen rok huachai læ loʿtluat khoʿng phupuainoʿk". [Rick Factors Influencing Heart and Vascular Diseases to Outpatients]. (Research Project) Faculty of Science and Technology, Suan Sunandha Rajabhat University

Singchogchai, P. (2006). "lakkān læ kānchai sathiti kān wikhroʿ tuapræ lai tua samrap kānwichai thangkān phayaban". [Principles and Using Multivariate Statistics Analysis for Nursing Research]. (3rd ed.). Songkhla: Chan Muang Printing.

Stevens, J. 1996. **Applied multivariate statistics for the social science. Lawrence Erlbaum Associate, Inc.,** New Jersey : Mahwah.Vanichbuncha, K. (2008). "kān wikhroʿ khoʿmun lai tuapræ". [Multivariate Analysis]. (3rd ed.). Statistics Department. Chulalongkorn Business School Bangkok : Chulalongkorn University.