อิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องที่มีผลต่อสัมประสิทธิ์การตัดสินใจที่ปรับค่า สำหรับการถดถอยโลจิสติกทวิภาค

An Effect of Link Function Misspecification on the Adjusted Coefficients of Determination for Binary Logistic Regression

นารีรัตน์ ณ นุวงศ์ และ แสงหล้า ชัยมงคล

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี 12121

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องที่มีผลต่อสัมประสิทธิ์การตัดสินใจที่ปรับค่า $\left(R^2_{adj}\right)$ สำหรับการวิเคราะห์การถดถอยโลจิสติกทวิภาค โดยศึกษาเฉพาะค่า R^2_{adj} ที่ใช้หลักการคำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญ $\left(R^2_{O,adj,MS}, R^2_{O,adj,LM}\right)$ และวิธีความควรจะเป็นสูงสุด $(R^2_{l,adj,MS}, R^2_{l,adj,LM}$ $R^2_{l,adj,SAS_{DEV}}, R^2_{l,adj,SAS_{AIC}})$ และการวิจัยนี้จะศึกษาการกำหนดฟังก์ชันเชื่อมโยงของตัวแบบไม่ถูกต้อง (misspecified link function) 2 รูปแบบคือ ฟังก์ชันโพรบิท (probit) และฟังก์ชันคอมพลีเมนทารีล็อก-ล็อก (complementary log-log) เมื่อกำหนดฟังก์ชันเชื่อมโยงโลจิท (logit) เป็นฟังก์ชันที่แท้จริง การตรวจสอบอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องนั้นจะใช้วิธีการจำลองข้อมูลที่เปลี่ยนค่าของจำนวนตัวแปรอธิบาย (1, 5 และ 10 ตัว) และขนาดตัวอย่าง (50, 100, 1000 และ 2000) โดยจะพิจารณาความเอนเอียงสัมพัทธ์ของค่าประมาณมัธยฐานกับค่า R^2 ที่แท้จริง และร้อยละของค่าประมาณที่อยู่นอกช่วง [0, 1] ที่ได้จากการทำซ้ำ 1,000 ครั้ง

ผลที่ได้จากการศึกษา พบว่าการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องแบบโพรบิทมีอิทธิพลต่อค่า R^2_{adj} ไม่แตกต่างจากฟังก์ชันเชื่อมโยงโลจิท ในขณะที่ฟังก์ชันคอมพลีเมนทารีล็อก-ล็อกจะมีผลทำให้ค่าประมาณ R^2_{adj} ทุกตัว ยกเว้น $R^2_{l,adj,SAS_{DEV}}$ เป็นค่าประมาณที่เอนเอียง โดยความเอนเอียงนี้จะขึ้นอยู่กับจำนวนตัวแปรอธิบาย ขนาดตัวอย่าง และประเภทของค่า R^2_{adj} โดยค่า R^2_{adj} ที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญจะได้รับอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องแบบคอมพลีเมนทารีล็อก-ล็อกน้อยกว่า ค่า R^2_{adj} ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุด

คำสำคัญ: ตัวแบบเชิงเส้นวางนัยทั่วไป โลจิท โพรบิท คอมพลีเมนทารีล็อก-ล็อก กำลังสองน้อยสุดแบบสามัญ ความควรจะเป็นสูงสุด

Abstract

The objective of this research was to study the effect of link function misspecification on the adjusted coefficients of determination, R^2_{adj} for binary logistic regression. We focused on several R^2_{adj} types which were based on ordinary least square $\left(R^2_{O,adj,MS}, R^2_{O,adj,LM}\right)$ and maximum likelihood estimators $(R^2_{l,adj,MS}, R^2_{l,adj,LM}, R^2_{l,adj,SAS_{DEV}}, R^2_{l,adj,SAS_{AIC}})$. We studied two types of misspecified link function, probit link and complementary log-log link, when we assumed that the logit link was the true link function. We investigated the effect of incorrect link function though a simulation study under different number of predictors (1, 5, and 10) and different number of observations (50, 100, 1000, and 2000). The effect of link function was determined by the relative bias and the percentage that the R^2_{adj} estimated lies outside [0, 1] calculated from 1,000 runs.

The results revealed that the effect of probit link on the studied R^2_{adj} measures was similar to the logit link. On the other hand, the complementary log-log link function induced bias in all R^2_{adj} measures, except for the $R^2_{l,adj,SAS_{DEV}}$. The bias varied across the number of predictors, number of observations, and type of R^2_{adj}. The R^2_{adj} measures based on ordinary least square was affected by the complementary log-log link less than the R^2_{adj} measures based on maximum likelihood.

Keywords: Generalized linear models, logit, probit, complementary log-log, ordinary least square, maximum likelihood

1. ที่มาและความสำคัญของปัญหา

การวิเคราะห์การถดถอยโลจีสติกทวิภาค (binary logistic regression analysis) เป็นวิธีทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างกลุ่มตัวแปรอธิบาย (explain variable) และตัวแปรตามที่มีค่าเป็นไปได้ 2 ค่า (binary responses variable) โดยมีวัตถุประสงค์เพื่อหาตัวแบบที่แสดงความสัมพันธ์ระหว่างตัวแปรเหล่านี้ และนำค่าตัวแปรอธิบายไปพยากรณ์ค่าความน่าจะเป็นที่จะเกิดความสำเร็จ กำหนดให้ $(Y_i, x_i), 1 \le i \le n$ เมื่อ Y_i มีค่า 0 หรือ 1 และ $x_i = (x_{i1}, x_{i2}, ..., x_{ip})'$ เป็นเวกเตอร์ของตัวแปรอธิบาย การสร้างตัวแบบแสดงความสัมพันธ์ของข้อมูลนี้จะอยู่ในรูป

$$P(Y_i = 1 \mid x_i) \equiv \pi_i = F(\beta_0 + \beta'_p x_i) \quad (1)$$

เมื่อ $\left[\beta_0, \beta'_p\right]$ คือเวกเตอร์ขนาด $(p+1)$ ของสัมประสิทธิ์การถดถอยที่ไม่ทราบค่า และ $F(\cdot)$ คือฟังก์ชันการแจกแจงสะสมที่ทราบรูปแบบและเรียก $F(\cdot)$ ว่าฟังก์ชันเชื่อมโยง โดยฟังก์ชันเชื่อมโยงแบบบัญญัติ (canonical link function) สำหรับตัวแปรตามที่เป็นทวิภาคคือฟังก์ชันโลจิท (logit) และตัวแบบของสมการ (1) สามารถเขียนได้ในรูป

$$logit(\pi_i) \equiv \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_p' x_i$$

เมื่อ $\pi_i = P(Y_i = 1 | x_i)$ ที่สามารถเขียนในรูปของฟังก์ชันโลจีสติกได้คือ

$$P(Y_i = 1 | x_i) = \frac{\exp(\beta_0 + \beta_p' x_i)}{1 + \exp(\beta_0 + \beta_p' x_i)}$$

ดังนั้นสมการที่ใช้ฟังก์ชันเชื่อมโยงแบบโลจิท จึงมีชื่อเรียกว่าสมการถดถอยโลจีสติก แต่ยังมีฟังก์ชันเชื่อมโยงแบบอื่นๆ ที่สามารถใช้ได้กับตัวแปรแบบทวิภาค เช่นฟังก์ชันเชื่อมโยงโพรบิท (probit link) ที่มีพื้นฐานมาจากฟังก์ชันความน่าจะเป็นสะสมของการแจกแจงปกติมาตรฐาน (standard normal cumulative distribution function) คือ

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{1}{2}z^2\right) dz$$

และตัวแบบในสมการ (1) สามารถเขียนได้ในรูป

$$\Phi^{-1}(\pi_i) = \beta_0 + \beta_p' x_i$$

นอกจากนั้นฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก (complementary log-log) เป็นอีกฟังก์ชันหนึ่งที่สามารถใช้ได้กับตัวแปรตามแบบทวิภาค โดยฟังก์ชันชนิดนี้เป็นฟังก์ชันเชื่อมโยงที่ปรับจากฟังก์ชันโลจิท เมื่อค่าของ π_i เพิ่มจาก 0 ค่อนข้างช้า แต่มีค่าเข้าใกล้ 1 ค่อนข้างเร็ว [1] ในกรณีนี้ตัวแบบในสมการ (1) สามารถเขียนได้ในรูป

$$\log[-\log(1-\pi_i)] = \beta_0 + \beta_p' x_i$$

การเลือกใช้ฟังก์ชันเชื่อมโยงรูปแบบใดนั้นไม่มีกฎเกณฑ์ที่แน่นอน การเลือกใช้ฟังก์ชันเชื่อมโยงที่ไม่ถูกต้องกับลักษณะของข้อมูล ทำให้ได้ค่าประมาณพารามิเตอร์ และค่าพยากรณ์ความน่าจะเป็นที่มีความเอนเอียง [2] โดย Czado and Santner ได้ศึกษาอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องในตัวแบบการถดถอยโลจีสติก โดยได้ศึกษาฟังก์ชันเชื่อมโยงในรูปแบบฟังก์ชันโลจีสติก ฟังก์ชัน Box-Cox ฟังก์ชัน Cauchy และฟังก์ชัน Burr พบว่าความเบ้ (skewness) ของฟังก์ชันเชื่อมโยงจะมีผลกระทบต่อค่าประมาณของตัวแบบการถดถอยโลจีสติกมากกว่าความโด่ง (kurtosis) ของฟังก์ชันเชื่อมโยง สำหรับในที่นี้จะศึกษาอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องที่มีต่อค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่า (adjusted coefficient of determination: R^2_{adj}) ของตัวแบบการถดถอยโลจีสติก เมื่อกำหนดให้ตัวแบบที่แท้จริงมีฟังก์ชันเชื่อมโยงในรูปแบบของฟังก์ชันโลจิท โดยศึกษาฟังก์ชันเชื่อมโยงที่ไม่ถูกต้อง ใน 2 รูปแบบคือฟังก์ชันโพรบิท และฟังก์ชันคอมพลีเมนทารีล็อก-ล็อก

2. ค่าสัมประสิทธิการตัดสินใจที่ปรับค่าสำหรับตัวแบบการถดถอยโลจิสติก $\left(R^2_{adj}\right)$

ค่าสัมประสิทธิ์การตัดสินใจ (coefficient of determination) $\left(R^2\right)$ ในการวิเคราะห์การถดถอยเชิงเส้นทั่วไป เป็นค่าที่แสดงถึงสัดส่วนความผันแปรของตัวแปรตามที่สามารถอธิบายได้ด้วยตัวแบบการถดถอย แต่ในบางครั้งเกิดปัญหาคือ สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอธิบายมากกว่า 1 ตัว ค่า R^2 จะให้ค่าสูงกว่าที่ควรจะเป็น แม้ว่าตัวแปรอธิบายที่เพิ่มเข้าไปในตัวแบบจะไม่มีความสัมพันธ์กับตัวแปรตาม ดังนั้นจึงทำการปรับค่าของสัมประสิทธิ์การตัดสินใจ หรือที่เรียกว่าสัมประสิทธิ์การตัดสินใจที่ปรับค่า R^2_{adj} และเช่นเดียวกับการวิเคราะห์การถดถอยโลจีสติก ควรมีการใช้ค่า R^2_{adj} ในกรณีที่มีตัวแปรอธิบายจำนวนมาก และตัวอย่างมีขนาดเล็ก การคำนวณค่า R^2

และ R^2_{adj} สำหรับการวิเคราะห์การถดถอยโลจิสติกนั้น มีได้หลายรูปแบบ ในที่นี้จะศึกษาการคำนวณ R^2 และ R^2_{adj} โดยใช้วิธีกำลังสองน้อยสุดแบบสามัญ (ordinary least square, OLS: R^2_O) และวิธีความควรจะเป็นสูงสุด (maximum likelihood, ML: R^2_l) เมื่อ

$$R^2_O = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{\pi}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \qquad (2)$$

และ

$$R^2_l = 1 - \frac{l\left(y,\hat{\pi}\right)}{l\left(y,\hat{\pi}^0\right)} \qquad (3)$$

เมื่อ n คือขนาดตัวอย่าง $l\left(y,\hat{\pi}\right)$ และ $l\left(y,\hat{\pi}^0\right)$ คือฟังก์ชันล็อกควรจะเป็นสูงสุดสำหรับตัวแบบที่มีตัวแปรอธิบาย p ตัว และตัวแบบที่มีเพียงเทอมค่าคงที่ซึ่งคำนวณจาก

$$\sum_{i=1}^{n}\left[y_i \log\left(\hat{\pi}_i\right) + \left(1-y_i\right)\log\left(1-\hat{\pi}_i\right)\right]$$

และ $\sum_{i=1}^{n}\left[y_i \log\left(\bar{y}\right) + \left(1-y_i\right)\log\left(1-\bar{y}\right)\right]$

ตามลำดับ

ค่า R^2_{adj} ที่นำมาศึกษาในงานวิจัยนี้ได้แก่

2.1 ค่า R^2_{adj} ที่ศึกษาโดย Mittlböck และ Schemper [3] ที่ทำการปรับค่าด้วยการใช้องศาเสรี (degree of freedom) และสถิติ deviance แทนด้วย $R^2_{O,adj,MS}$, $R^2_{l,adj,MS}$ ตามลำดับ โดยที่

$$R^2_{O,adj,MS} = 1 - \frac{(n-1)\sum_{i=1}^{n}\left(y_i - \hat{\pi}_i\right)^2}{(n-p-1)\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \qquad (4)$$

และ

$$R^2_{l,adj,MS} = 1 - \frac{l\left(y,\hat{\pi}\right) - (p+1)/2}{l\left(y,\hat{\pi}^0\right) - 1/2} \qquad (5)$$

2.2 ค่า R^2_{adj} ที่เสนอโดย Liao และ McGee [4] ที่มีการปรับค่าด้วยค่าคลาดเคลื่อนของการพยากรณ์ที่ไม่สามารถแยกได้ (inherent prediction error, IPE) แทนด้วย $R^2_{O,adj,LM}$, $R^2_{l,adj,LM}$ โดยที่

$$R^2_{O,adj,LM} = 1 - \frac{I\hat{P}E^p_o}{I\hat{P}E^o_o} \qquad (6)$$

และ

$$R^2_{l,adj,LM} = 1 - \frac{I\hat{P}E^p_l}{I\hat{P}E^o_l} \qquad (7)$$

เมื่อ $I\hat{P}E^P_o$ และ $I\hat{P}E^o_o$ คือค่าประมาณของ IPE สำหรับตัวแบบที่มีตัวแปรอธิบาย p ตัว และตัวแบบที่มีเพียงเทอมค่าคงที่เท่านั้น ตามลำดับ เมื่อคำนวณโดยวิธีกำลังสองน้อยสุดแบบสามัญ

$I\hat{P}E^P_l$ และ $I\hat{P}E^o_l$ คือค่าประมาณของ IPE สำหรับตัวแบบที่มีตัวแปรอธิบาย p ตัว และตัวแบบมีเพียงเทอมค่าคงที่เท่านั้น ตามลำดับ เมื่อคำนวณโดยวิธีความควรจะเป็นสูงสุด

2.3 ค่า R^2_{adj} ที่ใช้ในโปรแกรม SAS เวอร์ชัน 8.0 ที่แนะนำโดย Shtatland [5] จำนวน 2 ตัว ที่ใช้หลักการของสถิติ deviance และ AIC (Akaike's information criterion) มาทำการปรับค่า แทนด้วย $R^2_{l,adj,SAS_{DEV}}$, $R^2_{l,adj,SAS_{AIC}}$ ตามลำดับ โดยที่

$$R^2_{l,adj,SAS_{DEV}} = \frac{1 - \exp\left\{\frac{-2\left[l\left(y,\hat{\pi}\right) - l\left(y,\hat{\pi}^0\right)\right]}{n}\right\}}{1 - \exp\left[\frac{2l\left(y,\hat{\pi}^0\right)}{n}\right]} \qquad (8)$$

และ

$$R^2_{l,adj,SAS_{AIC}} = 1 - \frac{l\left(y,\hat{\pi}\right) - \left[\dfrac{(p+1)(n-1)}{(n-p-1)}\right]}{l\left(y,\hat{\pi}^0\right) - 1} \quad (9)$$

เมื่อ $l\left(y,\hat{\pi}\right)$ และ $l\left(y,\hat{\pi}^0\right)$ อธิบายได้เช่นเดียวกับ (3)

3. วิธีการดำเนินการศึกษา

ศึกษาด้วยการจำลองข้อมูล (simulation data) โดยใช้เทคนิคมอนติคาร์โล ที่ใช้โปรแกรมสำเร็จรูป SAS กำหนดเงื่อนไขของการศึกษาดังต่อไปนี้

3.1 กำหนดให้ขนาดตัวอย่าง (n) เท่ากับ 50, 100, 1000, และ 2000

3.2 กำหนดให้จำนวนตัวแปรอธิบาย (p) เท่ากับ 1, 5, และ 10 ตัว โดยกำหนดให้ตัวแปรอธิบายมีทั้งตัวแปรที่เป็นเชิงปริมาณและเชิงคุณภาพ ตัวแปรที่เป็นเชิงปริมาณคือ X_1, X_3, X_5, X_7, และ X_9 โดยตัวแปรเหล่านี้จะถูกสุ่มจากการแจกแจงปรกติดังต่อไปนี้

$X_1 \sim N(2,4), X_3 \sim N(20,4), X_5 \sim N(8,3)$, $X_7 \sim N(0,1)$, และ $X_9 \sim N(15,2)$ สำหรับตัวแปรเชิงคุณภาพนั้น จะถูกสุ่มจากการแจกแจงเอกรูปที่ไม่ต่อเนื่อง (discrete uniform) โดยมีรายละเอียดดังนี้

$X_2 \sim U[0,1]$, $X_4 \sim U[0,3]$, $X_6 \sim U[0,2]$, $X_8 \sim U[0,1]$, และ $X_{10} \sim U[0,4]$

3.3 ตัวแปรตามแบบทวิภาคสร้างจากตัวแบบ 3 ตัวแบบ ที่ใช้ฟังก์ชันเชื่อมโยงโลจิท ดังต่อไปนี้

$$logit\left(\pi_i\right) = 0.5 - 0.3x_{i1} \quad (10)$$

$$\begin{aligned} logit\left(\pi_i\right) = {} & 0.5 - 0.3x_{i1} + 3.525x_{i2} - 0.075x_{i3} \\ & - 3.725x_{i4} + 0.05x_{i5} \end{aligned} \quad (11)$$

$$\begin{aligned} logit\left(\pi_i\right) = {} & 0.5 - 0.3x_{i1} + 3.525x_{i2} - 0.075x_{i3} \\ & - 3.725x_{i4} + 0.05x_{i5} - 0.3x_{i6} \\ & + 3.525x_{i7} + 0.07x_{i8} - 0.3x_{i9} \\ & - 3.725x_{i10} \end{aligned} \quad (12)$$

โดยที่ค่าสัมประสิทธิ์ของตัวแบบจะกำหนดขึ้นเพื่อให้ค่าของตัวแปรตามมีค่า 0 และ 1 ในสัดส่วนเท่าๆ กัน

3.4 กำหนดฟังก์ชันเชื่อมโยงที่ไม่ถูกต้อง 2 รูปแบบ คือฟังก์ชันโพรบิท และฟังก์ชันคอมพลีเมนทารีล็อก-ล็อก

3.5 การเปรียบเทียบความแกร่งของค่า R^2_{adj} ทั้ง 6 ตัว จะเปรียบเทียบความเอนเอียงสัมพัทธ์ (relative biased, RB.) ของค่าประมาณมัธยฐาน $\left(\tilde{R}^2\right)$ ที่ได้จากการทำซ้ำจำนวน 1,000 ครั้ง โดยที่

$$\text{RB.} = \left[\left(\tilde{R}^2 - R^2_{true}\right) \Big/ R^2_{true}\right] \times 100\%$$

และค่าที่แท้จริงหาได้จาก

$$R^2_{O,true} = 1 - \frac{\sum_{i=1}^{n}\left[\pi_i\left(1-\pi_i\right)\right]}{\sum_{i=1}^{n}\left[\bar{\pi}\left(1-\bar{\pi}\right)\right]} \quad (13)$$

และ

$$R^2_{l,true} = 1 - \frac{\sum_{i=1}^{n}\left[\pi_i \log\pi_i + \left(1-\pi_i\right)\log\left(1-\pi_i\right)\right]}{\sum_{i=1}^{n}\left[\bar{\pi}\log\bar{\pi} + \left(1-\bar{\pi}\right)\log\left(1-\bar{\pi}\right)\right]} \quad (14)$$

เมื่อ $R^2_{O,true}$, $R^2_{l,true}$ คือค่า R^2 ที่แท้จริง ที่คำนวณโดยวิธีกำลังสองน้อยสุดแบบสามัญ และวิธีความควรจะเป็นสูงสุด ตามลำดับ π_i คือค่าเริ่มต้นที่ได้จากการจำลอง $logit\left(\pi_i\right)$ และ $\bar{\pi}$ คือค่าเฉลี่ยของ π_i เริ่มต้น

นอกจากนั้น ยังเปรียบเทียบร้อยละของค่าประมาณสัมประสิทธิ์การตัดสินใจ ที่มีค่าอยู่นอกช่วง [0,1] จากการทำซ้ำจำนวน 1,000 ครั้ง

4. ผลการวิจัย

จากตารางที่ 1 เมื่อพิจารณาค่ามัธยฐานของค่าประมาณ R_{adj}^2 เมื่อใช้ฟังก์ชันเชื่อมโยงโลจิท ซึ่งเป็นฟังก์ชันเชื่อมโยงที่ถูกต้อง และตัวแบบมีตัวแปรอธิบายเพียงตัวเดียว พบว่าค่า $R_{l,adj,SAS_{DEV}}^2$ เป็นค่าประมาณที่มีความเอนเอียง (biased estimator) โดยจะให้ค่าเอนเอียงสัมพัทธ์ (relative biased) สูงที่สุดในทุกขนาดตัวอย่าง (มีค่าอยู่ระหว่าง 68.9 ถึง 79.1) ในขณะที่ค่า R_{adj}^2 ตัวอื่นๆ เป็นค่าประมาณที่มีความเอนเอียงเล็กน้อย โดยแต่ละค่าจะให้ค่าเอนเอียงสัมพัทธ์น้อยมาก และมีค่าไม่แตกต่างจากค่า R^2 ที่ไม่ทำการปรับค่า และเมื่อพิจารณาร้อยละของค่า R_{adj}^2 ที่อยู่นอกช่วง [0,1] พบว่าค่า R_{adj}^2 ที่อยู่นอกช่วงนั้นจะเกิดขึ้นเฉพาะในกรณีที่มีขนาดตัวอย่าง 50 และ 100 เท่านั้น โดยที่ค่า R_{adj}^2 ที่เสนอโดย Liao และ McGee ทั้งที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญ และวิธีความควรจะเป็นสูงสุด $\left(R_{O,adj,LM}^2, R_{l,adj,LM}^2\right)$ จะให้ร้อยละของค่าที่อยู่นอกช่วงสูงกว่าค่า R_{adj}^2 ของ Mittlböck และ Schemper $\left(R_{O,adj,MS}^2, R_{l,adj,MS}^2\right)$ นอกจากนั้นค่าที่อยู่นอกช่วงยังพบในค่า R_{adj}^2 ที่ใช้ในโปรแกรม SAS ที่ใช้การปรับค่าโดยอาศัยหลักการของAIC (Akaike's information criterion: $R_{l,adj,SAS_{AIC}}^2$) อย่างไรก็ตามร้อยละของค่าเหล่านี้ถือว่ามีค่าน้อย (ค่ามากที่สุดเพียง 12%)

เมื่อพิจารณากรณีที่กำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องเมื่อฟังก์ชันเชื่อมโยงโลจิทเป็นฟังก์ชันเชื่อมโยงที่แท้จริง พบว่าฟังก์ชันเชื่อมโยงโพรบิทให้ค่าเอนเอียงสัมพัทธ์ของค่า R_{adj}^2 และร้อยละของค่า R_{adj}^2 ที่อยู่นอกช่วง [0,1] ไม่แตกต่างจากฟังก์ชันเชื่อมโยงโลจิทมากนัก ในทางตรงกันข้ามกับฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อกมีผลทำให้ค่า R_{adj}^2 ทุกตัว (ยกเว้น $R_{l,adj,SAS_{DEV}}^2$) เป็นค่าประมาณที่มีความเอนเอียง โดยจะให้ค่าเอนเอียงสัมพัทธ์ที่สูงมากในทุกขนาดตัวอย่าง (มีค่าอยู่ระหว่าง 38.4 ถึง 79.6) โดยค่าเอนเอียงสัมพัทธ์ที่เกิดขึ้นจะไม่ขึ้นอยู่กับขนาดตัวอย่าง และวิธีการที่ใช้ในการคำนวณค่า R_{adj}^2 นอกจากนั้นฟังก์ชันเชื่อมโยงคอมพลีเมนทารีล็อก-ล็อก ยังมีผลต่อร้อยละของค่าประมาณ R_{adj}^2 ที่อยู่นอกช่วง [0,1] โดยจะให้ร้อยละที่อยู่นอกช่วงสูง เมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 โดยค่า R_{adj}^2 ที่เสนอโดย Liao และ McGee ทั้งที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญ และวิธีความควรจะเป็นสูงสุด $\left(R_{O,adj,LM}^2, R_{l,adj,LM}^2\right)$ จะมีร้อยละสูงสุด (ประมาณ 30-40%) รองลงมาคือ R_{adj}^2 ที่ใช้ในโปรแกรม SAS ที่ใช้การปรับค่าโดยอาศัยหลักการของ AIC $\left(R_{l,adj,SAS_{AIC}}^2\right)$ ค่า R_{adj}^2 ของ Mittlböck และ Schemper $\left(R_{O,adj,MS}^2, R_{l,adj,MS}^2\right)$ และ $R_{l,adj,SAS_{DEV}}^2$ ตามลำดับ ค่าร้อยละที่อยู่นอกช่วงนี้จะลดลงอย่างมากจนเกือบไม่มีเมื่อขนาดตัวอย่างเพิ่มขึ้นเป็น 1000 และ 2000

เมื่อมีจำนวนตัวแปรอธิบายในตัวแบบ 5 และ 10 ตัว (ตารางที่ 2 และ 3 ตามลำดับ) ผลสรุปที่ได้จะไปในแนวทางเดียวกัน นั่นคือเมื่อใช้ฟังก์ชันเชื่อมโยงโลจิทค่า $R_{l,adj,SAS_{DEV}}^2$ จะให้ค่าเอนเอียงสัมพัทธ์มากสุดในทุก ขนาดตัวอย่าง ในขณะที่ค่า $R_{l,adj,SAS_{AIC}}^2$ ให้ค่าเอนเอียงสัมพัทธ์ที่สูงเฉพาะกรณีที่ขนาดตัวอย่างเท่ากับ 50 และ 100 เท่านั้น และให้ค่าเอนเอียงสัมพัทธ์พอๆ กับค่า R^2 ที่ไม่ได้ทำการปรับ

ค่า โดยที่ค่าเอนเอียงสัมพัทธ์ของค่าเหล่านี้จะมีค่าเพิ่มขึ้นเมื่อมีจำนวนตัวแปรอธิบายในตัวแบบมากขึ้น สำหรับค่า R^2_{adj} ของ Mittlböck และ Schemper และ Liao และ McGee นั้นจะให้ค่าเอนเอียงสัมพัทธ์ที่สูงเฉพาะในกรณีที่ขนาดตัวอย่างเล็ก (n=50) และจะมีผลต่อค่า R^2_{adj} ที่ใช้คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญมากกว่าวิธีความควรจะเป็นสูงสุด สำหรับร้อยละของค่า R^2_{adj} ที่อยู่นอกช่วง [0,1] นั้นสรุปได้เช่นเดียวกับกรณีที่มีตัวแปรอธิบายจำนวน1 ตัว

อิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้อง เมื่อฟังก์ชันเชื่อมโยงที่แท้จริงคือฟังก์ชันโลจิท พบว่าอิทธิพลของฟังก์ชันเชื่อมโยงโพรบิททั้งในกรณีที่มีตัวแปรอธิบายในตัวแบบจำนวน 5 และ 10 ตัว สรุปได้เช่นเดียวกับกรณีที่มีตัวแปรอธิบายจำนวน 1 ตัว นั่นคือฟังก์ชันโพรบิทให้ค่าเอนเอียงสัมพัทธ์ของค่า R^2_{adj} และร้อยละของค่า R^2_{adj} ที่อยู่นอกช่วง [0,1] ไม่แตกต่างจากฟังก์ชันโลจิทมากนัก ในขณะที่อิทธิพลของฟังก์ชันคอม-พลีเมนทารีล็อก-ล็อกนั้น จะทำให้ค่าประมาณ R^2_{adj} ทุกตัวยกเว้น $R^2_{l,adj,SAS_{DEV}}$ เป็นค่าประมาณที่เอนเอียงทั้งในกรณีที่มีตัวแปรอธิบาย 5 และ 10 ตัว ในขณะที่ค่า $R^2_{l,adj,SAS_{AIC}}$ ให้ค่าเอนเอียงสัมพัทธ์สูงเฉพาะในกรณีที่มีตัวแปรอธิบาย 10 ตัวเท่านั้น และอิทธิพลของฟังก์ชันคอมพลีเมนทารีล็อก-ล็อกนั้น จะมีผลต่อ R^2_{adj} ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดมากกว่าที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญ นอกจากนั้นยังพบว่าในกรณีที่ขนาดตัวอย่างใหญ่ (n=1000, 2000) ค่า R^2_{adj} ที่เสนอโดย Liao และ McGee นั้นจะมีค่าเอนเอียงสัมพัทธ์น้อยที่สุด แต่ถ้าขนาดตัวอย่างเท่ากับ 50 แล้ว ค่า R^2 ที่ไม่ได้ทำการปรับค่าจะมีค่าเอนเอียงสัมพัทธ์น้อยที่สุด ทั้งในกรณีที่มีตัวแปรอธิบายจำนวน 5 และ 10 ตัว แต่ในกรณีที่ขนาดตัวอย่างเท่ากับ 100 เมื่อตัวแปรอธิบายมีจำนวน 5 ตัว ค่า R^2_{adj} ที่เสนอโดย Liao และ McGee นั้นจะมีค่าเอนเอียงสัมพัทธ์น้อยที่สุด แต่เมื่อตัวแปรอธิบายมีจำนวน 10 ตัว ค่า R^2 ที่ไม่ได้ทำการปรับค่าจะมีค่าเอนเอียงสัมพัทธ์น้อยที่สุด และค่า R^2_{adj} ของ Mittlböck และ Schemper จะมีค่าเอนเอียงสัมพัทธ์น้อยกว่าค่า R^2_{adj} ที่เสนอโดย Liao และ McGee เมื่อพิจารณาร้อยละของค่า R^2_{adj} ที่อยู่นอกช่วง [0,1] ทั้งในกรณีที่มีตัวแปรอธิบายจำนวน 5 และ 10 ตัว (ตารางที่ 2 และ 3 ตามลำดับ) พบว่าค่า R^2_{adj} ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดจะมีค่าร้อยละที่อยู่นอกช่วง [0,1] สูงกว่าค่า R^2_{adj} ที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญมากๆ โดยเรียงลำดับจากมากไปหาน้อยได้ดังนี้ ค่า $R^2_{l,adj,SAS_{AIC}}$ รองลงมาคือ $R^2_{l,adj,MS}$, $R^2_{l,adj,LM}$, และ $R^2_{l,adj,SAS_{DEV}}$ ตามลำดับ โดยที่ร้อยละที่อยู่นอกช่วง [0,1] ของ $R^2_{l,adj,SAS_{DEV}}$ มีค่าไม่แตกต่างจาก R^2_l ที่ไม่ปรับค่า

5. สรุปผลการวิจัยและข้อเสนอแนะ

เมื่อกำหนดฟังก์ชันเชื่อมโยงที่แท้จริงเป็นฟังก์ชันโลจิท พบว่าค่า $R^2_{l,adj,SAS_{DEV}}$ มีประสิทธิภาพน้อยที่สุด เมื่อพิจารณาจากค่าเอนเอียงสัมพัทธ์ของค่าประมาณมัธยฐานกับค่า R^2 ที่แท้จริงและร้อยละของค่า R^2_{adj} ที่อยู่นอกช่วง [0,1] ในทุกขนาดตัวอย่างและทุกจำนวนตัวแปรอธิบายที่มีอยู่ในตัวแบบ สำหรับประสิทธิภาพของค่า R^2_{adj} อื่นๆ นั้นขึ้นอยู่กับขนาดตัวอย่างและจำนวนตัวแปรอธิบาย โดยที่เมื่อมีตัวแปรอธิบายเพียงตัวเดียวค่า R^2_{adj} แบบต่างๆ มีประสิทธิภาพดีไม่แตกต่างกันในทุกขนาดตัวอย่าง แต่เมื่อมีจำนวนตัวแปรอธิบายเพิ่มมากขึ้นและมีขนาด

ตัวอย่างจำนวน 50 และ 100 ค่า $R^2_{l,adj,SAS_{AIC}}$ มีประสิทธิภาพไม่แตกต่างจากค่า R^2 ที่ไม่ปรับค่าที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญและวิธีความควรจะเป็นสูงสุด ซึ่งมีประสิทธิภาพต่ำกว่าค่า R^2_{adj} ของ Mittlböck และ Schemper และของ Liao และ McGee ในกรณีที่มีตัวแปรอธิบายจำนวน 10 ตัว และขนาดตัวอย่างเท่ากับ 50 ประสิทธิภาพของ R^2_{adj} ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดจะมีประสิทธิภาพดีกว่า R^2_{adj} ที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญ และวิธีการของ Liao และ McGee มีประสิทธิภาพสูงกว่าวิธีการของ Mittlböck และ Schemper แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็น 1000 และ 2000 ค่า R^2 ที่ไม่ปรับค่าและปรับค่าทุกตัว (ยกเว้น $R^2_{l,adj,SAS_{DEV}}$) มีประสิทธิภาพไม่แตกต่าง ทั้งในกรณีที่มีตัวแปรอธิบายจำนวน 5 และ 10 ตัว ดังนั้นจากผลการวิจัยที่ได้ สามารถสรุปได้ว่าในกรณีที่มีตัวแปรอธิบายจำนวนมาก และขนาดตัวอย่างเล็ก สมควรที่จะเลือกใช้ค่า R^2_{adj} ที่เสนอโดย Liao และ McGee เพื่อใช้อธิบายสัดส่วนของความผันแปรที่เกิดขึ้นในความน่าจะเป็นที่จะเกิดความสำเร็จของตัวแปรตาม ที่เกิดเนื่องจากอิทธิพลของตัวแปรอธิบาย หรือใช้ในการคัดเลือกตัวแบบที่เหมาะสม

เมื่อพิจารณาอิทธิพลของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้อง เมื่อฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยงที่แท้จริง พบว่าฟังก์ชันโพรบิทมีอิทธิพลต่อค่า R^2_{adj} ต่างๆ ไม่แตกต่างจากฟังก์ชันโลจิท ซึ่งเป็นไปตามที่คาดไว้ ทั้งนี้เพราะฟังก์ชันโลจิทและฟังก์ชันโพรบิทมีลักษณะที่คล้ายกันเพียงแต่ฟังก์ชันโพรบิทจะมีค่าที่เข้าใกล้ 0 หรือ 1 ค่อนข้างเร็วกว่าฟังก์ชันโลจิท และฟังก์ชันทั้งสองจะมีลักษณะสมมาตร ในขณะที่ฟังก์ชันคอมพลีเมนทารีล็อก-ล็อกนั้นมีลักษณะไม่สมมาตร จึงทำให้ฟังก์ชันคอมพลีเมนทารีล็อก-ล็อกมีอิทธิพลต่อค่า R^2_{adj} ต่างจากฟังก์ชันโลจิท เนื่องจากความเบ้ (skewness) ของฟังก์ชันเชื่อมโยงจะมีผลกระทบต่อค่าประมาณของตัวแบบการถดถอยโลจิสติก ซึ่งเป็นไปตามการศึกษาของ Czado and Santner [2] และอิทธิพลที่มีต่อค่า R^2_{adj} ต่างๆ นั้นขึ้นอยู่กับจำนวนตัวแปรอธิบาย ขนาดตัวอย่าง และประเภทของค่า R^2_{adj} โดยที่จะมีผลทำให้ค่าประมาณ R^2_{adj} ทุกตัว ยกเว้น $R^2_{l,adj,SAS_{DEV}}$ เป็นค่าประมาณที่เอนเอียง และ R^2_{adj} ที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญจะมีความแกร่งต่อการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องแบบคอมพลีเมนทารีล็อก-ล็อกมากกว่า R^2_{adj} ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุด เมื่อมีตัวแปรอธิบายมากกว่า 1 ตัว และเมื่อขนาดตัวอย่างเท่ากับ 1000 และ 2000 ค่า R^2_{adj} ที่เสนอโดย Liao และ McGee เป็นค่า R^2_{adj} ที่มีความแกร่งมากสุด แต่เมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 และมีตัวแปรอธิบายจำนวน 10 ตัว ค่า R^2 ที่ไม่ปรับค่าจะมีความแกร่งมากสุด รองลงมาคือค่า R^2_{adj} ที่เสนอโดย Liao และ McGee และเมื่อมีตัวแปรอธิบายจำนวน 5 ตัว ค่า R^2_{adj} ที่เสนอโดย Liao และ McGee มีความแกร่งมากสุด

ดังนั้นเมื่อพิจารณาทั้งประสิทธิภาพ และความแกร่งของการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้อง สามารถสรุปได้ว่าค่า R^2_{adj} ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดของ Liao และ McGee เป็นค่าที่ดีและเหมาะกับสถานการณ์ทุกสถานการณ์ของขนาดตัวอย่าง และจำนวนตัวแปรอธิบาย โดยเฉพาะอย่างยิ่งในกรณีที่มีขนาดตัวอย่างเล็ก แต่มีจำนวนตัวแปรอธิบายจำนวนมาก ซึ่งเป็นไปตามการศึกษาของ Liao และ McGee [4].

6. เอกสารอ้างอิง

[1]. Aldrich, J. H., and Nelson, F.D., Linear Probability Logit and Probit Model, Beverly Hills: Sage, 1984.

[2]. Czado, C., and Santner, T. J., The Effect of Link Misspecification on Binary Regression Inference, Journal of Statistics Planning and Inference, Vol. 33; pp. 213-231, 1992.

[3]. Mittlböck , M., and Schemper, M., Explained Variation for Logistic Regression, Statistics in Medicine, Vol. 15; pp. 1987-1997, 1996.

[4]. Liao, J. G., and McGee, D., Adjusted Coefficients of Determination for Logistic Regression, The American Statistician, Vol. 57; pp.161-165, 2003.

[5]. Shtatland, E., Moore, S., and Mary, B., Why We Need an Measure of Fit (and not only one) in PROC LOGISTIC and PROC GENMOD Paper, pp.256-225, Online Store., www2.sas.com/proceedings/sugi25/25/st/25p256.pdf.

ตารางที่ 1 ค่าสถิติของสัมประสิทธิ์การตัดสินใจแบบต่างๆ ที่ได้จากการทำซ้ำ 1,000 ครั้ง เมื่อใช้ตัวแบบ $logit(\pi_i) = 0.5 - 0.45x_{i1}$ ภายใต้ขนาดตัวอย่าง 50, 100, 1000 และ 2000

n	R^2	R^2_{true}	logit link			probit link			complementary link		
			$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$
50	R^2_O	0.1431	0.1508	5.38	0	0.1507	5.31	0	0.0837	-41.51	20.7
	$R^2_{O,adj,MS}$		0.1331	-6.99	2.9	0.1330	-7.06	2.9	0.0646	-54.86	26.7
	$R^2_{O,adj,LM}$		0.1238	-13.49	11.3	0.1315	-8.11	12.2	0.0695	-51.43	33.1
	R^2_l	0.1115	0.1202	7.80	0	0.1197	7.35	0	0.0550	-50.67	27.5
	$R^2_{l,adj,MS}$		0.1037	-7.00	2.4	0.1031	-7.53	2.4	0.0393	-64.75	33.0
	$R^2_{l,adj,LM}$		0.0980	-12.11	12.3	0.1023	-8.25	14.3	0.0397	-64.39	39.5
	$R^2_{l,adj,SAS_{DEV}}$		0.1997	79.10	0	0.2001	79.46	0	0.0957	-14.17	27.5
	$R^2_{l,adj,SAS_{AIC}}$		0.0859	-22.96	6.8	0.0855	-23.32	6.9	0.0227	-79.64	39.1
100	R^2_O	0.1447	0.1456	0.62	0	0.1450	0.21	0	0.0855	-40.91	14.0
	$R^2_{O,adj,MS}$		0.1369	-5.39	0.2	0.1363	-5.81	0.2	0.0762	-47.34	18.0
	$R^2_{O,adj,LM}$		0.1368	-5.46	5.2	0.1316	-9.05	3.9	0.0657	-54.60	29.6
	R^2_l	0.1130	0.1140	0.88	0	0.1145	1.33	0	0.0537	-52.48	21.1
	$R^2_{l,adj,MS}$		0.1057	-6.46	0.2	0.1062	-6.02	0.2	0.0458	-59.47	24.7
	$R^2_{l,adj,LM}$		0.1059	-6.28	6.4	0.1034	-8.50	4.9	0.0375	-66.81	35.7
	$R^2_{l,adj,SAS_{DEV}}$		0.1918	69.73	0	0.1927	70.53	0	0.0937	-17.08	21.1
	$R^2_{l,adj,SAS_{AIC}}$		0.0973	-13.89	0.3	0.0979	-13.36	0.3	0.0377	-66.64	28.6

n	R^2	R^2_{true}	logit link			probit link			complementary link		
			$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$
1,000	R^2_O	0.1462	0.1465	0.21	0	0.1462	0.00	0	0.0900	-38.44	0
	$R^2_{O,adj,MS}$		0.1456	-0.41	0	0.1454	-0.55	0	0.0891	-39.06	0
	$R^2_{O,adj,LM}$		0.1458	-0.27	0	0.1446	-1.09	0	0.0835	-42.89	3.0
	R^2_l	0.1144	0.1143	-0.09	0	0.1141	-0.26	0	0.0610	-46.68	0.4
	$R^2_{l,adj,MS}$		0.1135	-0.79	0	0.1133	-0.96	0	0.0603	-47.29	0.4
	$R^2_{l,adj,LM}$		0.1143	-0.09	0	0.1128	-1.40	0	0.0546	-52.27	5.6
	$R^2_{l,adj,SAS_{DEV}}$		0.1937	69.32	0	0.1933	68.97	0	0.1071	-6.38	0.4
	$R^2_{l,adj,SAS_{AIC}}$		0.1126	-1.57	0	0.1125	-1.66	0	0.0595	-47.99	0.4
2,000	R^2_O	0.1458	0.1453	-0.34	0	0.1452	-0.41	0	0.0876	-39.92	0
	$R^2_{O,adj,MS}$		0.1449	-0.62	0	0.1448	-0.69	0	0.0872	-40.19	0
	$R^2_{O,adj,LM}$		0.1448	-0.69	0	0.1437	-1.44	0	0.0794	-45.54	0.1
	R^2_l	0.1140	0.1138	-0.18	0	0.1139	-0.09	0	0.0590	-48.25	0
	$R^2_{l,adj,MS}$		0.1134	-0.53	0	0.1134	-0.53	0	0.0587	-48.51	0
	$R^2_{l,adj,LM}$		0.1135	-0.44	0	0.1125	-1.32	0	0.0526	-53.86	0.8
	$R^2_{l,adj,SAS_{DEV}}$		0.1926	68.95	0	0.1925	68.86	0	0.1036	-9.12	0
	$R^2_{l,adj,SAS_{AIC}}$		0.1130	-0.88	0	0.1130	-0.88	0	0.0583	-48.86	0.1

เมื่อ $\tilde{R}^2$ คือค่ามัธยฐาน และ $\text{RB.} = \left[\left(\tilde{R}^2 - R^2_{true}\right) \Big/ R^2_{true}\right] \times 100\%$

<u>ตารางที่ 2</u> ค่าสถิติของสัมประสิทธิ์การตัดสินใจแบบต่างๆ ที่ได้จากการทำซ้ำ 1,000 ครั้ง เมื่อใช้ตัวแบบ $logit(\pi_i) = 0.5 - 0.45x_{i1} + 3.525x_{i2} - 0.1x_{i3} - 0.5x_{i4} + 0.2x_{i5}$ ภายใต้ขนาดตัวอย่าง 50, 100, 1000 และ 2000

n	R^2	R^2_{true}	logit link			probit link			complementary link		
			$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$
50	R^2_O	0.3429	0.4303	25.49	0	0.4239	23.62	0	0.3416	-0.38	1.5
	$R^2_{O,adj,MS}$		0.3656	6.62	0.1	0.3584	4.52	0.1	0.2667	-22.22	5.7
	$R^2_{O,adj,LM}$		0.3535	3.09	2.8	0.3444	0.44	1.6	0.3040	-11.34	4.1
	R^2_l	0.2802	0.3641	29.94	0	0.3667	30.87	0	0.1355	-51.64	30.6
	$R^2_{l,adj,MS}$		0.2870	2.43	0.1	0.2892	3.21	0.1	0.0607	-78.34	40.2
	$R^2_{l,adj,LM}$		0.2841	1.39	5.3	0.2824	0.79	4.1	0.0690	-75.37	39.8
	$R^2_{l,adj,SAS_{DEV}}$		0.5261	87.76	0	0.5281	88.47	0	0.2258	-19.41	30.6
	$R^2_{l,adj,SAS_{AIC}}$		0.1923	-31.37	2.1	0.1939	-30.80	2.3	-0.0304	-110.85	55.3
100	R^2_O	0.3451	0.3838	11.21	0	0.3812	10.46	0	0.3058	-11.39	0.2
	$R^2_{O,adj,MS}$		0.3511	1.74	0	0.3483	0.93	0	0.2689	-22.08	0.8
	$R^2_{O,adj,LM}$		0.3463	0.35	0.1	0.3432	-0.55	0.2	0.3325	-3.65	0.5
	R^2_l	0.2824	0.3205	13.49	0	0.3209	13.63	0	0.1726	-38.88	14.2
	$R^2_{l,adj,MS}$		0.2821	-0.11	0	0.2825	0.04	0	0.1346	-52.34	19.4
	$R^2_{l,adj,LM}$		0.2790	-1.20	0.4	0.2767	-2.02	0.5	0.1753	-37.92	14.3
	$R^2_{l,adj,SAS_{DEV}}$		0.4766	68.77	0	0.4768	68.84	0	0.2820	-0.14	14.2
	$R^2_{l,adj,SAS_{AIC}}$		0.2395	-15.19	0	0.2396	-15.16	0	0.0934	-66.93	28.0
1,000	R^2_O	0.3476	0.3527	1.47	0	0.3520	1.27	0	0.2885	-17.00	0
	$R^2_{O,adj,MS}$		0.3494	0.52	0	0.3488	0.35	0	0.2850	-18.01	0
	$R^2_{O,adj,LM}$		0.3472	-0.12	0	0.3464	-0.35	0	0.3735	7.45	0
	R^2_l	0.2847	0.2885	1.33	0	0.2882	1.23	0	0.1782	-37.41	0
	$R^2_{l,adj,MS}$		0.2846	-0.04	0	0.2844	-0.11	0	0.1745	-38.71	0
	$R^2_{l,adj,LM}$		0.2838	-0.32	0	0.2830	-0.60	0	0.2469	-13.28	0
	$R^2_{l,adj,SAS_{DEV}}$		0.4389	54.16	0	0.4386	54.06	0	0.2917	2.46	0
	$R^2_{l,adj,SAS_{AIC}}$		0.2808	-1.37	0	0.2805	-1.48	0	0.1707	-40.04	0
2,000	R^2_O	0.3482	0.3491	0.26	0	0.3486	0.11	0	0.2828	-18.78	0
	$R^2_{O,adj,MS}$		0.3474	-0.23	0	0.3470	-0.34	0	0.2810	-19.30	0

n	R^2	R^2_{true}	logit link			probit link			complementary link		
			$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$
	$R^2_{O,adj,LM}$		0.3469	-0.37	0	0.3464	-0.52	0	0.3745	7.55	0
	R^2_l	0.2850	0.2860	0.35	0	0.2859	0.32	0	0.1771	-37.86	0
	$R^2_{l,adj,MS}$		0.2841	-0.32	0	0.2840	-0.35	0	0.1752	-38.53	0
	$R^2_{l,adj,LM}$		0.2838	-0.42	0	0.2830	-0.70	0	0.2501	-12.25	0
	$R^2_{l,adj,SAS_{DEV}}$		0.4358	52.91	0	0.4356	52.84	0	0.2898	1.68	0
	$R^2_{l,adj,SAS_{AIC}}$		0.2822	-0.98	0	0.2821	-1.02	0	0.1733	-39.19	0

เมื่อ $\tilde{R}^2$ คือค่ามัธยฐาน และ $\text{RB.} = \left[\left(\tilde{R}^2 - R^2_{true} \right) \Big/ R^2_{true} \right] \times 100\%$

<u>ตารางที่ 3</u> ค่าสถิติของสัมประสิทธิ์การตัดสินใจแบบต่างๆ ที่ได้จากการทำซ้ำ 1,000 ครั้ง เมื่อใช้ตัวแบบ $logit(\pi_i) = 0.5 - 0.45x_{i1} + 3.525x_{i2} - 0.1x_{i3} - 0.5x_{i4} + 0.2x_{i5} - 0.4x_{i6} + 0.3x_{i7} - 0.3x_{i8} - 0.05x_{i9} + 0.8x_{i10}$ ภายใต้ขนาดตัวอย่าง 50, 100, 1000 และ 2000

n	R^2	R^2_{true}	logit link			probit link			complementary link		
			$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$
50	R^2_O	0.4875	0.6605	35.49	0	0.6477	32.86	0	0.5976	22.58	0
	$R^2_{O,adj,MS}$		0.5735	17.64	0	0.5574	14.34	0	0.4944	1.42	1.7
	$R^2_{O,adj,LM}$		0.4767	-2.22	1.5	0.4753	-2.50	1.2	0.5229	7.26	1.9
	R^2_l	0.4149	0.6127	47.67	0	0.6140	47.99	0	0.0664	-84.00	45.4
	$R^2_{l,adj,MS}$		0.4588	10.58	0	0.4593	10.70	0	-0.0794	-119.14	56.9
	$R^2_{l,adj,LM}$		0.4049	-2.41	3.3	0.4141	-0.19	2.5	-0.0058	-101.40	51.0
	$R^2_{l,adj,SAS_{DEV}}$		0.7607	83.35	0	0.7609	83.39	0	0.1172	-71.75	45.4
	$R^2_{l,adj,SAS_{AIC}}$		0.2275	-45.17	5.4	0.2287	-44.88	5.5	-0.3047	-173.44	77.8
100	R^2_O	0.4916	0.5678	15.50	0	0.5608	14.08	0	0.5115	4.05	0
	$R^2_{O,adj,MS}$		0.5193	5.63	0	0.5114	4.03	0	0.4566	-7.12	0
	$R^2_{O,adj,LM}$		0.4945	0.59	0	0.4818	-1.99	0	0.4479	-8.89	0
	R^2_l	0.4178	0.5034	20.49	0	0.5040	20.63	0	0.2109	-49.52	26.8
	$R^2_{l,adj,MS}$		0.4268	2.15	0	0.4278	2.39	0	0.1374	-67.11	34.3
	$R^2_{l,adj,LM}$		0.4188	0.24	0	0.4093	-2.03	0	0.1204	-71.18	34.5
	$R^2_{l,adj,SAS_{DEV}}$		0.6664	59.50	0	0.6683	59.96	0	0.3374	-19.24	26.8
	$R^2_{l,adj,SAS_{AIC}}$		0.3327	-20.37	0	0.3349	-19.84	0	0.0467	-88.82	45.1
1,000	R^2_O	0.4942	0.5010	1.38	0	0.5000	1.17	0	0.4563	-7.67	0
	$R^2_{O,adj,MS}$		0.4960	0.36	0	0.4949	0.14	0	0.4508	-8.78	0
	$R^2_{O,adj,LM}$		0.4956	0.28	0	0.4938	-0.08	0	0.5180	4.82	0
	R^2_l	0.4200	0.4270	1.67	0	0.4268	1.62	0	0.3191	-24.02	0
	$R^2_{l,adj,MS}$		0.4195	-0.12	0	0.4192	-0.19	0	0.3116	-25.81	0
	$R^2_{l,adj,LM}$		0.4215	0.36	0	0.4196	-0.10	0	0.3681	-12.36	0
	$R^2_{l,adj,SAS_{DEV}}$		0.5943	41.50	0	0.5938	41.38	0	0.4749	13.07	0
	$R^2_{l,adj,SAS_{AIC}}$		0.4117	-1.98	0	0.4115	-2.02	0	0.3039	-27.64	0
2,000	R^2_O	0.4942	0.4968	0.53	0	0.4962	0.40	0	0.4534	-8.26	0
	$R^2_{O,adj,MS}$		0.4943	0.02	0	0.4937	-0.10	0	0.4506	-8.82	0

n	R^2	R^2_{true}	logit link			probit link			complementary link		
			$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$	$\tilde{R}^2$	RB.	$\% \notin [0,1]$
	$R^2_{O,adj,LM}$		0.4930	-0.24	0	0.4946	0.08	0	0.5222	5.67	0
	R^2_{l}	0.4199	0.4220	0.50	0	0.4216	0.40	0	0.3187	-24.10	0
	$R^2_{l,adj,MS}$		0.4182	-0.40	0	0.4177	-0.52	0	0.3150	-24.98	0
	$R^2_{l,adj,LM}$		0.4186	-0.31	0	0.4195	-0.10	0	0.3773	-10.15	0
	$R^2_{l,adj,SAS_{DEV}}$		0.5893	40.34	0	0.5885	40.15	0	0.4747	13.05	0
	$R^2_{l,adj,SAS_{AIC}}$		0.4143	-1.33	0	0.4139	-1.43	0	0.3112	-25.89	0

เมื่อ $\tilde{R}^2$ คือค่ามัธยฐาน และ $\text{RB.} = \left[\left(\tilde{R}^2 - R^2_{true} \right) \middle/ R^2_{true} \right] \times 100\%$