บทความวิจัย

การแจกแจงของสถิติ Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL) และ Deviance สำหรับตัวแบบการถดถอยลอจิสติค เมื่อตัวอย่างมีขนาดเล็ก

กฤษฎา เหล็กดี และ ลี่ลี อิงศรีสว่าง*

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการแจกแจงของสถิติที่ใช้ในการพิจารณาความเหมาะสมของตัวแบบการถดถอยลอจิสติค (Logistic Regression Model) เมื่อตัวอย่างมีขนาดเล็ก โดยศึกษาการแจกแจงของสถิติ 5 ตัวคือ สถิติ Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL) และ Deviance วิธีการศึกษาใช้การจำลองตัวแบบการถดถอยลอจิสติค ที่ตัวแปรตามมีค่าเป็น 0 และ 1 และมีตัวแปรอิสระ 4 ตัวเป็นตัวแปรต่อเนื่อง 2 ตัว และตัวแปรหุ่น (Dummy Variable) 2 ตัว ทำการจำลองทั้งหมด 4 สถานการณ์ คือที่ตัวอย่างขนาด 10, 30, 50 และ 100 แต่ละสถานการณ์ทำซ้ำ 1,000 รอบ การพิจารณาการแจกแจงของสถิติแต่ละตัวจะเปรียบเทียบกับการแจกแจงแบบ Chi-square โดยพิจารณาเฉพาะบริเวณหางขวาของการแจกแจงเท่านั้น ซึ่งเป็นบริเวณที่สำคัญที่เกี่ยวข้องกับการตัดสินใจยอมรับหรือปฏิเสธสมมุติฐานหลัก (H_0) กำหนดระดับนัยสำคัญของการทดสอบ (α) คือ 0.01, 0.05 และ 0.10 ผลการศึกษาพบว่า เมื่อตัวอย่างมีขนาด 10 สถิติที่มีการแจกแจงใกล้การแจกแจงแบบ Chi-square มากที่สุดที่ $\alpha = 0.01$ คือสถิติ Likelihood Ratio ที่ $\alpha = 0.05$ คือสถิติ Deviance และที่ $\alpha = 0.10$ คือสถิติ Score เมื่อตัวอย่างมีขนาด 30, 50 และ 100 สถิติ HL มีการแจกแจงใกล้การแจกแจงแบบ Chi-square มากที่สุดที่ระดับนัยสำคัญ 0.01, 0.05 และ 0.10 อย่างไรก็ตามในการนำไปใช้ ถ้าสมมุติให้ สถิติ Likelihood Ratio สถิติ Deviance สถิติ Score และ สถิติ HL มีการแจกแจงแบบ Chi-square ระดับนัยสำคัญของการทดสอบจะเปลี่ยนไปจากที่กำหนดไว้

คำสำคัญ: ตัวแบบการถดถอยลอจิสติค, Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL), Deviance

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

*ผู้นิพนธ์ประสานงาน, email: fscilli@ku.ac.th

The Distribution of Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL), and Deviance for a Small Sample Logistic Regression Model

Krisada Lekdee and Lily Ingsrisawang*

ABSTRACT

The objective of this research involved the investigation of the distribution of the statistics that were used to examine whether or not the logistic regression model fit the data when the sample size was small. These statistics were Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL), and Deviance. The simulation study was employed. The data were simulated for the logistic regression models with 2 dummy and 2 continuous variables with samples sizes of 10, 30, 50 and 100. For each sample size 1,000 simulation runs were made. The distributions of those statistics (upper tails only) were compared with the Chi-square distributions. The upper tail of the distribution was an important segment since it was used for hypothesis testing. The levels of significance (α) were set at 0.01, 0.05 and 0.10. The study found that, for the sample size of 10, at $\alpha = 0.01$, the distribution of Likelihood Ratio was closest to the Chi-square distribution; at $\alpha = 0.05$, it was the distribution of Deviance; and at $\alpha = 0.10$, it was the distribution of Score. For the sample sizes of 30, 50 and 100, the distribution of HL was closest to the Chi-square distribution at the significance levels of 0.01, 0.05 and 0.10. However, if the Score, Deviance, Likelihood Ratio, and HL were used and assumed to have the Chi-square distributions, the levels of significance would change from the ones that were set at the beginning.

Keywords: Logistic Regression Model, Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL), Deviance

Department of Statistics, Faculty of Science, Kasetsart University
*Corresponding author, e-mail: fscilli@ku.ac.th

บทนำ

การถดถอยลอจิสติค (Logistic Regression) เป็นการถดถอยที่มีตัวแบบที่ใช้แสดงความสัมพันธ์ระหว่างตัวแปรตามที่มีค่าเป็น 0 หรือ 1 กับตัวแปรอิสระที่เป็นได้ทั้งชนิดต่อเนื่องและชนิดไม่ต่อเนื่องและมีได้มากกว่า 1 ตัว ตัวอย่างในงานวิจัยทางการแพทย์เช่น การศึกษาการเสียชีวิตของผู้ป่วยภายในระยะเวลา 5 ปีหลังเข้ารับการรักษาโรคมะเร็ง กำหนดให้ตัวแปรตามมีค่า 0 เมื่อผู้ป่วยรอดชีพ และ 1 เมื่อผู้ป่วยเสียชีวิต สิ่งที่สนใจคือมีปัจจัยหรือตัวแปรอิสระใดบ้างที่สามารถใช้พยากรณ์ความน่าจะเป็นหรือความเสี่ยงที่ผู้ป่วยจะเสียชีวิต ตัวแบบการถดถอยลอจิสติค มีลักษณะดังนี้

ให้ Y มีการแจกแจงแบบเบอร์นูลี (Bernoulli Distribution) ความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจ $\mathrm{P}(Y = 1) = P$ ความน่าจะเป็นที่จะไม่เกิดเหตุการณ์ $\mathrm{P}(Y = 0) = 1 - P$ $E(Y) = \mathrm{P}$ และ $V(Y) = 1 - P$

นิยาม $\log it(P) = \log(\frac{P}{1 - P}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$

เมื่อ $X_1, X_2, ..., X_p$ คือตัวแปรอิสระ และ $\beta_0, \beta_1, \beta_2, ..., \beta_p$ คือสัมประสิทธิ์การถดถอย ตัวแบบการถดถอยลอจิสติค แสดงความน่าจะเป็นหรือความเสี่ยงที่จะเกิดเหตุการณ์ที่สนใจคือ

$$P(Y = 1 \mid X_1, X_2, ..., X_p) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)}$$

การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติค ใช้วิธีการประมาณแบบ Maximum Likelihood ใช้วิธีการวนซ้ำ (Iterative Methods) และหาคำตอบของสมการที่ไม่เป็นเชิงเส้น (Nonlinear Equations) ซึ่งมีหลายวิธี แต่ที่มีประสิทธิภาพมากที่สุดและใช้กันมากคือวิธีของ Newton-Raphson และ Fisher's Scoring [1]

การตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติค มีสถิติให้พิจารณาเลือกใช้หลายตัว เช่นสถิติ Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL) และ Deviance สถิติเหล่านี้มีการแจกแจงแบบ Asymptotic Chi-square คือเมื่อตัวอย่างมีขนาดใหญ่ จะมีการแจกแจงใกล้การแจกแจงแบบ Chi-square [2-5] แต่ในทางปฏิบัติขนาดตัวอย่างที่ใช้ศึกษาอาจมีขนาดเล็ก ทำให้เกิดปัญหาว่าการใช้สถิติข้างต้นเพื่อตรวจสอบความเหมาะสมของตัวแบบ ยังคงใกล้เคียงกับการแจกแจงแบบ Chi-square หรือไม่ มีการศึกษาเกี่ยวกับตัวแบบการถดถอยลอจิสติค และสถิติที่ใช้ตรวจสอบความเหมาะสมของตัวแบบอย่างกว้างขวาง เช่น Chen และคณะ [6] ศึกษาตัวแบบทางคณิตศาสตร์สำหรับการอนุมานในตัวแบบการถดถอยลอจิสติค เมื่อตัวอย่างมีขนาดเล็ก ประยุกต์ใช้เรื่องคลื่นเดี่ยวใต้น้ำ (Internal Solitary Wave) วิธีการศึกษาใช้การทดลองที่ตัวอย่างขนาด 54 พบว่า เมื่อนำทฤษฎีตัวอย่างขนาดใหญ่ไปใช้กับตัวอย่างขนาดเล็ก สถิติ Likelihood Ratio และ Score ให้ผลสรุปการทดสอบสมมุติฐานตรงกันคือปฏิเสธว่าตัวแบบเหมาะสมกับข้อมูล ตรงกันข้ามกับสถิติ Wald และ Deviance ที่ให้ผลสรุปยอมรับว่าตัวแบบมีความเหมาะสม Pulkstenis และ Robinson [7] ศึกษาสถิติที่ใช้ทดสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติค ที่มีตัวแปรอิสระเป็นตัวแปรต่อเนื่อง พบว่า สถิติ Deviance มีการแจกแจงแตกต่างจากการแจกแจงแบบ Chi-square จึงสร้างสถิติสำหรับการทดสอบความเหมาะสมของตัวแบบดังกล่าวขึ้นมาใหม่

โดยใช้หลักการคล้ายกันกับการสร้างสถิติ HL แล้วจำลองสถานการณ์แบบต่างๆ ขึ้นมา เพื่อเปรียบเทียบตัวสถิติที่สร้างขึ้นมาใหม่กับสถิติ HL และ Kramer [8] ศึกษาการแจกแจงของสถิติ HL โดยใช้การจำลองตัวแบบการถดถอยลอจิสติค ที่มีตัวแปรอิสระเป็นตัวแปรต่อเนื่อง 20 ตัวและตัวแปรหุ่น 3 ตัว ที่ตัวอย่างขนาด 50,000 พบว่าสถิติ HL มีการแจกแจงใกล้การแจกแจงแบบ Chi-square

อย่างไรก็ตามการศึกษาการแจกแจงของสถิติที่กล่าวมาข้างต้นพร้อมกันทั้ง 5 ตัวคือสถิติ Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL) และ Deviance สำหรับตัวแบบการถดถอยลอจิสติค เมื่อตัวอย่างมีขนาดเล็ก เปรียบเทียบกับการแจกแจงแบบ Chi-square ยังไม่มีมาก่อน ผู้วิจัยจึงสนใจที่จะศึกษาการแจกแจงของสถิติดังกล่าวเมื่อตัวอย่างมีขนาดเล็ก เปรียบเทียบกับการแจกแจงแบบ Chi-square เพื่อนำผลที่ได้ไปใช้ในการเลือกสถิติสำหรับตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติค เมื่อตัวอย่างมีขนาดเล็กต่อไป

วิธีการศึกษา

การวิจัยครั้งนี้ใช้การจำลองตัวแบบการถดถอยลอจิสติค ที่ตัวแปรตาม Y มีการแจกแจงแบบเบอร์นูลี (Bernoulli Distribution) มี P เป็นพารามิเตอร์ แสดงความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจ และมีตัวแปรอิสระ 4 ตัว คือ X_1, X_2 เป็นตัวแปรหุ่น (Dummy Variables) และ Z_1, Z_2 เป็นตัวแปรชนิดต่อเนื่องมีการแจกแจงแบบปกติ โดยจำลองทั้งหมด 4 สถานการณ์ คือ ที่ขนาดตัวอย่าง 10, 30, 50 และ 100 กำหนดระดับนัยสำคัญของการทดสอบ (α) คือ 0.01, 0.05 และ 0.10 แต่ละสถานการณ์ใช้เทคนิคการจำลองมอนติคาร์โลทำการวนซ้ำ 1,000 รอบ โปรแกรมที่ใช้ศึกษาคือ โปรแกรม SAS ของภาควิชาสถิติ มหาวิทยาลัยเกษตรศาสตร์ มีขั้นตอนดังนี้

1. ผลิตค่าของตัวแปรอิสระ X_1 และ X_2 ให้มีการแจกแจงแบบเบอร์นูลี โดยกำหนดให้ X_1 มีค่า 1 อยู่ 40% และ X_2 มีค่า 1 อยู่ 50% ฟังก์ชันที่ใช้คือ RAND (‘BERNOULLI’, 0.40) และ RAND (‘BERNOULLI’, 0.50) ตามลำดับ

2. ผลิตค่าของตัวแปรอิสระ Z_1 และ Z_2 เป็นตัวแปรต่อเนื่อง มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากันคือ 0.5 และความแปรปรวนเป็น 0.1 และ 0.5 ตามลำดับ ฟังก์ชันที่ใช้คือ 0.5+RANNOR (-1)*0.1 และ 0.5+RANNOR (-1)*0.5 ตามลำดับ

3. กำหนดค่า $\beta_0 = -1.4$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 1$, $\beta_4 = 1$

4. ให้ xbeta $= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z_1 + \beta_4 Z_2$ นั่นคือ
   xbeta $= -1.4 + X_1 + 0.5\ X_2 + Z_1 + Z_2$

5. คำนวณค่า phat $= P(Y = 1 | X_1, X_2, Z_1, Z_2) = \exp(\text{xbeta})/(1+\exp(\text{xbeta}))$ ตามตัวแบบการถดถอยลอจิสติค ซึ่ง phat จะมีค่าอยู่ระหว่าง 0 ถึง 1

6. ผลิตค่าของตัวแปรตาม Y ให้มีค่าเป็น 0 หรือ 1 โดยวิธีสร้างเลขสุ่มที่มีการแจกแจงแบบ Uniform (0,1) ถ้าตัวเลขสุ่มที่ได้มีค่ามากกว่าค่า phat ให้ $Y = 1$ ถ้าเลขสุ่มที่ได้น้อยกว่าหรือเท่ากับ phat ให้ $Y = 0$ ฟังก์ชันที่ใช้คือ RANUNI (-1)

7. ผลิตค่า X_1, X_2, Z_1, Z_2 และ Y ตามขั้นตอนที่ 1- 6 ให้มีจำนวนเท่ากับขนาดตัวอย่างที่กำหนดในแต่ละสถานการณ์ คือ 10, 30, 50 และ 100

8. ใช้ PROC LOGISTIC และ PROC GENMOD ประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติค บันทึกค่าสถิติ Wald, Score, Likelihood Ratio, HL และ Deviance ของแต่ละสถานการณ์

9. แต่ละสถานการณ์ทำการวนซ้ำ 1,000 รอบ ซึ่งจะทำให้ได้ค่าสถิติแต่ละตัวจำนวน 1,000 ค่า

10. นำค่าสถิติแต่ละตัว มาวิเคราะห์หาลักษณะการแจกแจง ดังนี้

10.1 สร้าง Histogram และ QQ Plot ของสถิติแต่ละตัว เปรียบเทียบกับการแจกแจงแบบ Chi-square ที่มี Degree of Freedom (DF) เท่ากัน โดยที่สถิติ Wald, Score, Likelihood Ratio มี DF = 4 (จำนวนตัวแปรอิสระ) สถิติ HL มี DF = 8 และ สถิติ Deviance มี DF เท่ากับผลต่างของจำนวนค่าสังเกตกับจำนวนพารามิเตอร์

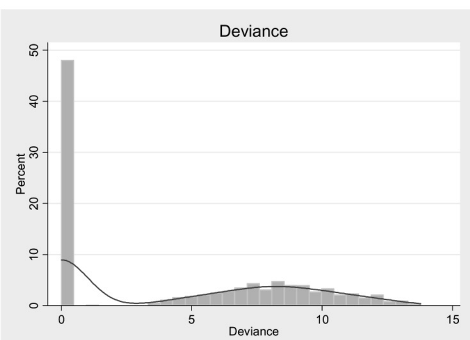
10.2 ที่หางขวาของการแจกแจง หาค่าสถิติแต่ละตัวที่ α = 0.01, 0.05 และ 0.10 จากค่าควอนไทล์ (Quantile) ที่ 0.99, 0.95 และ 0.90 ตามลำดับ เปรียบเทียบกับค่าสถิติ Chi-square ที่มี DF เท่ากัน และที่ α เดียวกัน สำหรับค่าสถิติ Chi-square นั้นสร้างจากฟังก์ชัน CINV(i-i/1000,DF), เมื่อ i = 1, 2, ..., 1,000 ซึ่งจะได้ค่าสถิติ Chi-square 1,000 ค่า ที่ α = 0.001, 0.002, 0.003, ...,0.999 ตามลำดับ

10.3 นำค่าสถิติ Chi-square ที่มี DF = 4 ที่ α = 0.01, 0.05 และ 0.10 ไปพิจารณาว่าตรงกับตำแหน่งควอนไทล์ที่เท่าไรของการแจกแจงของสถิติ Wald, Score, Likelihood Ratio นำค่าสถิติ Chi-square ที่มี DF = 8 ไปพิจารณาว่าตรงกับตำแหน่งควอนไทล์ที่เท่าไรของการแจกแจงของสถิติ HL และนำค่าสถิติ Chi-square ที่มี DF เท่ากับผลต่างของจำนวนค่าสังเกตกับจำนวนพารามิเตอร์ไปพิจารณาว่าตรงกับตำแหน่งควอนไทล์ที่เท่าไรของการแจกแจงของสถิติ Deviance

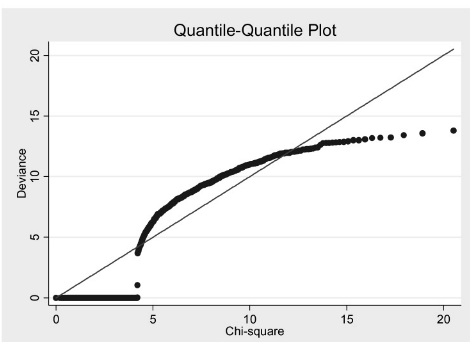
ผลการศึกษา

1. Histogram และ QQ Plot

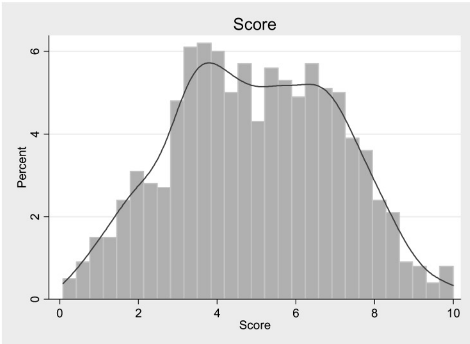
พิจารณาการแจกแจงของสถิติแต่ละตัวจาก Histogram และ QQ Plot พบว่าสถิติที่มีการแจกแจงใกล้การแจกแจงแบบ Chi-square มากกว่าสถิติตัวอื่นๆ เมื่อ n = 10 คือสถิติ Deviance และ Score เมื่อ n = 30, 50 และ 100 คือสถิติ HL แสดงดังรูปที่ 1-10



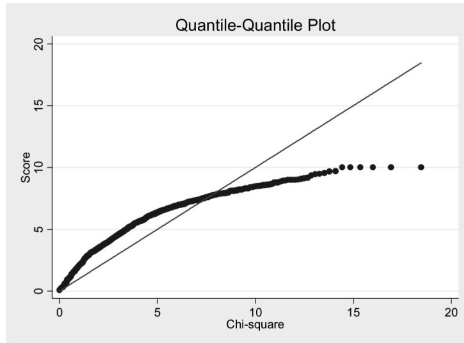
รูปที่ 1 Histogram ของสถิติ Deviance เมื่อ n = 10



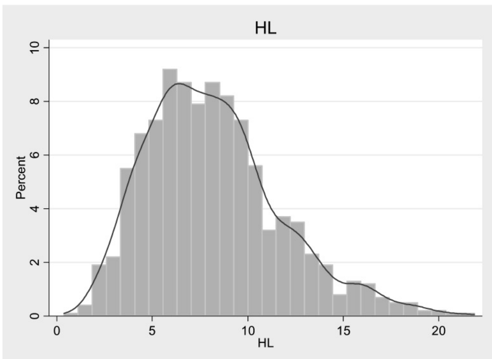
รูปที่ 2 QQ Plot ของสถิติ Deviance เมื่อ n = 10



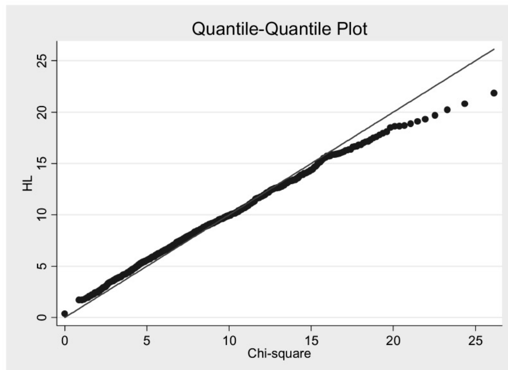
รูปที่ 3 Histogram ของสถิติ Score เมื่อ n = 10



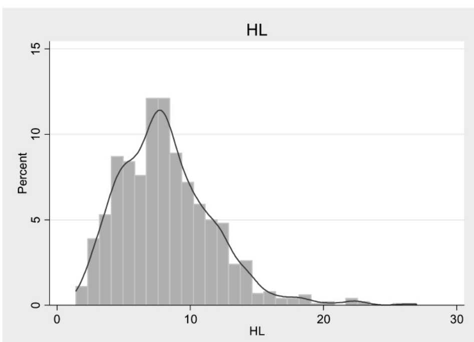
รูปที่ 4 QQ Plot ของสถิติ Score เมื่อ n = 10



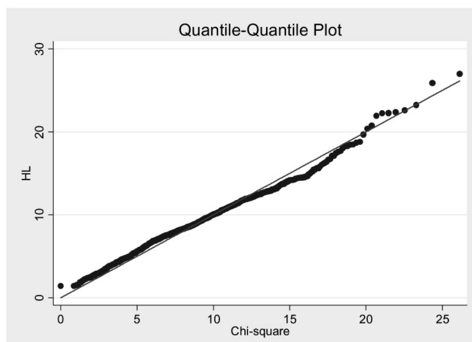
รูปที่ 5 Histogram ของสถิติ HL เมื่อ n = 30



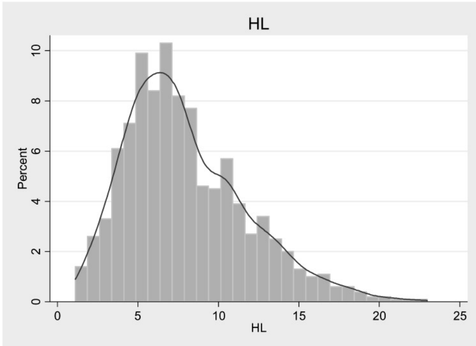
รูปที่ 6 QQ Plot ของสถิติ HL เมื่อ n = 30



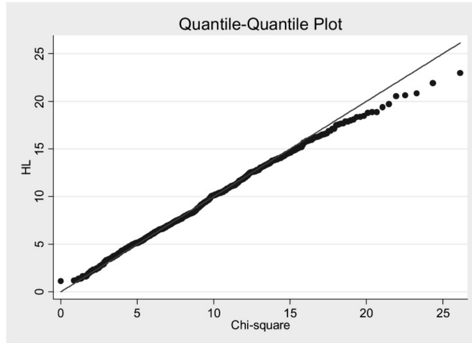
รูปที่ 7 Histogram ของสถิติ HL เมื่อ n = 50



รูปที่ 8 QQ Plot ของสถิติ HL เมื่อ n = 50



รูปที่ 9 Histogram ของสถิติ HL เมื่อ n = 100



รูปที่ 10 QQ Plot ของสถิติ HL เมื่อ n = 100

2. เปรียบเทียบค่าสถิติแต่ละตัวกับค่าสถิติ Chi-square ที่มี DF เท่ากัน

สถิติ Wald, Score, Likelihood Ratio มี DF = 4 (จำนวนตัวแปรอิสระ) สถิติ HL มี DF = 8 และสถิติ Deviance มี DF เท่ากับผลต่างของจำนวนค่าสังเกตกับจำนวนพารามิเตอร์ เปรียบเทียบค่าสถิติแต่ละตัวกับค่าสถิติ Chi-square ที่มี DF เท่ากัน ที่ = 0.01, 0.05 และ 0.10 แสดงในตารางที่ 1

ตารางที่ 1 เปรียบเทียบค่าสถิติแต่ละตัวกับค่าสถิติ Chi-square ที่มี DF เท่ากัน

n	**α**	**Chi-square DF=4**	**Wald DF=4**	**Score DF=4**	**Likelihood Ratio DF=4**	**Chi-square DF=8**	**HL DF=8**	**Chi-square DF=5**	**Deviance DF=5**
10	0.01	13.28	2.78	9.49	13.86	20.09	10.73	15.08	12.90
	0.05	9.49	2.29	8.30	13.85	15.51	9.73	11.07	11.56
	0.10	7.78	2.03	7.73	13.79	13.36	9.16	9.24	10.48
30	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=25	Deviance DF=25
	0.01	13.28	8.52	16.34	23.30	20.09	18.60	44.31	40.23
	0.05	9.49	7.65	13.22	17.17	15.51	15.11	37.65	38.86
	0.10	7.78	7.23	11.79	15.02	13.36	12.90	34.38	38.00

ตารางที่ 1 (ต่อ)

n	α	**Chi-square DF=4**	**Wald DF=4**	**Score DF=4**	**Likelihood Ratio DF=4**	**Chi-square DF=8**	**HL DF=8**	**Chi-square DF=45**	**Deviance DF=45**
50	0.01	13.28	12.95	21.03	26.79	20.09	20.38	69.96	67.40
	0.05	9.49	11.52	16.60	20.24	15.71	14.41	61.66	66.10
	0.10	7.78	10.76	14.70	17.62	13.36	12.72	57.51	65.24
100	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=95	Deviance DF=95
	0.01	13.28	21.84	29.98	35.17	20.09	18.79	129.97	134.88
	0.05	9.49	19.20	24.62	28.35	15.71	15.07	118.75	132.67
	0.10	7.78	17.67	21.97	24.61	13.36	13.26	113.04	130.88

จากตารางที่ 1 เมื่อ n = 10 ที่ $\alpha = 0.01$ สถิติ Likelihood Ratio (DF = 4) มีค่าใกล้ค่าสถิติ Chi-square (DF = 4) มากที่สุด ที่ $\alpha = 0.05$ สถิติ Deviance (DF = 5) มีค่าใกล้ค่าสถิติ Chi-square (DF = 5) มากที่สุด และที่ $\alpha = 0.10$ สถิติ Score (DF = 4) มีค่าใกล้ค่าสถิติ Chi-square (DF = 4) มากที่สุด เมื่อ n = 30, 50 และ 100 สถิติ HL (DF = 8) มีค่าใกล้ค่าสถิติ Chi-square (DF = 8) มากที่สุดที่ระดับนัยสำคัญ 0.01, 0.05 และ 0.10

ในการทดสอบสมมุติฐาน H_0 ที่ว่าตัวแบบการถดถอยลอจิสติคเหมาะสมกับข้อมูล ถ้าสมมุติให้สถิติแต่ละตัวมีการแจกแจงแบบ Chi-square ระดับนัยสำคัญของการทดสอบจะเปลี่ยนไป ตัวอย่างเช่น เมื่อ n = 10 ถ้าใช้การแจกแจงจริงของสถิติ Deviance จะปฏิเสธ H_0 ที่ $\alpha = 0.05$ เมื่อสถิติ Deviance มีค่ามากกว่า 11.56 แต่ถ้าสมมุติให้ สถิติ Deviance มีการแจกแจงแบบ Chi-square จะปฏิเสธ H_0 เมื่อสถิติ Deviance มีค่ามากกว่า 11.07 ซึ่งจะทำให้ระดับนัยสำคัญของการทดสอบเปลี่ยนจาก 0.05 เป็น 0.069 ดังแสดงในตารางที่ 2 และเมื่อ n = 30 ถ้าใช้การแจกแจงจริงของสถิติ HL จะปฏิเสธ H_0 ที่ $\alpha = 0.01$ เมื่อสถิติ HL มีค่ามากกว่า 18.60 แต่ถ้าสมมุติให้สถิติ HL มีการแจกแจงแบบ Chi-square จะปฏิเสธ H_0 เมื่อสถิติ HL มีค่ามากกว่า 20.09 ซึ่งจะทำให้ระดับนัยสำคัญของการทดสอบเปลี่ยนจาก 0.01 เป็น 0.003

ตารางที่ 2 ระดับนัยสำคัญของการทดสอบเมื่อสมมุติให้สถิติแต่ละตัวมีการแจกแจงแบบ Chi-square

n	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=5	Deviance DF=5
10	0.01	0.01	0.000	0.000	0.280	0.01	0.000	0.01	0.000
	0.05	0.05	0.000	0.009	0.468	0.05	0.000	0.05	0.069
	0.10	0.10	0.000	0.095	0.513	0.10	0.003	0.10	0.187
30	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=25	Deviance DF=25
	0.01	0.01	0.000	0.047	0.150	0.01	0.003	0.01	0.000
	0.05	0.05	0.002	0.233	0.351	0.05	0.045	0.05	0.119
	0.10	0.10	0.038	0.391	0.486	0.10	0.084	0.10	0.382
50	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=45	Deviance DF=45
	0.01	0.01	0.004	0.159	0.245	0.01	0.010	0.01	0.000
	0.05	0.05	0.203	0.400	0.470	0.05	0.030	0.05	0.321
	0.10	0.10	0.412	0.555	0.607	0.10	0.075	0.10	0.620
100	α	Chi-square DF=4	Wald DF=4	Score DF=4	Likelihood Ratio DF=4	Chi-square DF=8	HL DF=8	Chi-square DF=95	Deviance DF=95
	0.01	0.01	0.392	0.538	0.591	0.01	0.005	0.01	0.137
	0.05	0.05	0.715	0.763	0.793	0.05	0.042	0.05	0.701
	0.10	0.10	0.827	0.845	0.857	0.10	0.095	0.10	0.898

จากตารางที่ 2 ถ้าสมมุติให้สถิติทุกตัวมีการแจกแจงแบบ Chi-square ในการทดสอบสมมุติฐาน เมื่อ n = 10 ที่ α = 0.05 ถ้าเลือกใช้สถิติ Deviance ระดับนัยสำคัญของการทดสอบจะเปลี่ยนไปน้อยที่สุด และที่ α = 0.10 ถ้าเลือกใช้สถิติ Score ระดับนัยสำคัญของการทดสอบจะเปลี่ยนไปน้อยที่สุด แต่ที่ α = 0.01 ระดับนัยสำคัญของการทดสอบจะเปลี่ยนไปมากไม่ว่าจะใช้สถิติตัวใดก็ตาม เมื่อ n = 30, 50 และ 100 ถ้าเลือกใช้สถิติ HL ระดับนัยสำคัญของการทดสอบจะเปลี่ยนไปน้อยที่สุดที่ระดับนัยสำคัญ 0.01, 0.05 และ 0.10

3. ร้อยละของการปฏิเสธ H_0 เมื่อให้สถิติแต่ละตัวมีการแจกแจงแบบ Chi-square

ค่าร้อยละของการปฏิเสธ H_0 ที่ $\alpha = 0.01$, 0.05 และ 0.10 เมื่อสมมุติให้สถิติแต่ละตัวมีการแจกแจงแบบ Chi-square แสดงในตารางที่ 3

ตารางที่ 3 ค่าร้อยละของการปฏิเสธ H_0 เมื่อสมมุติให้สถิติแต่ละตัวมีการแจกแจงแบบ Chi-square

n	α	Wald (%)	Score (%)	Likelihood Ratio (%)	HL (%)	Deviance (%)
10	0.01	0.0	0.0	28.0	0.0	0.0
	0.05	0.0	0.9	46.8	0.0	**6.9**
	0.10	0.0	**9.5**	51.3	0.3	18.7
30	0.01	0.0	4.7	15.0	**0.3**	0.0
	0.05	0.2	23.3	35.1	**4.5**	11.9
	0.10	3.8	39.1	48.6	**8.4**	38.2
50	0.01	0.4	15.9	24.5	**1.0**	0.0
	0.05	20.3	40.0	47.0	**3.0**	32.1
	0.10	41.2	55.5	60.7	**7.5**	62.0
100	0.01	39.2	53.8	59.1	**0.5**	13.7
	0.05	71.5	76.3	79.3	**4.2**	70.1
	0.10	82.7	84.5	85.7	**9.5**	89.8

จากตารางที่ 3 ถ้าสมมุติให้สถิติทุกตัวมีการแจกแจงแบบ Chi-square ในการทดสอบสมมุติฐาน เมื่อ n = 10 ที่ $\alpha = 0.05$ ถ้าเลือกใช้สถิติ Deviance ค่าร้อยละของการปฏิเสธ H_0 จะใกล้กับค่า α มากที่สุด และที่ $\alpha = 0.10$ ถ้าเลือกใช้สถิติ Score ค่าร้อยละของการปฏิเสธ H_0 จะใกล้กับค่า α มากที่สุด แต่ที่ $\alpha = 0.01$ ค่าร้อยละของการปฏิเสธ H_0 จะต่างจากค่า α มาก ไม่ว่าจะเลือกใช้สถิติตัวใดก็ตาม เมื่อ n = 30, 50 และ 100 ถ้าเลือกใช้สถิติ HL ค่าร้อยละของการปฏิเสธ H_0 จะใกล้กับ α มากที่สุดที่ระดับนัยสำคัญ 0.01, 0.05 และ 0.10

สรุปและวิจารณ์ผลการศึกษา

ผลการศึกษาการแจกแจงของสถิติ Wald, Score, Likelihood Ratio, Hosmer and Lemeshow (HL) และ Deviance ที่ใช้สำหรับตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติค เมื่อตัวอย่างมีขนาดเล็ก โดยใช้การจำลองตัวแบบการถดถอยลอจิสติค ที่มีตัวแปรหุ่น 2 ตัว และตัวแปรต่อเนื่อง 2 ตัว โดยทำการจำลองทั้งหมด 4 สถานการณ์ คือ ที่ตัวอย่างขนาด 10, 30, 50 และ 100 แต่ละสถานการณ์ทำซ้ำ 1,000 รอบ พบว่า เมื่อตัวอย่างมีขนาด 10 สถิติที่มีการแจกแจงใกล้การแจกแจงแบบ Chi-square มากที่สุดที่

α = 0.05 คือ Deviance และที่ α = 0.10 คือ Score การศึกษาครั้งนี้ทำการเปรียบเทียบการแจกแจงของสถิติแต่ละตัวกับการแจกแจงแบบ Chi-square เฉพาะบริเวณหางขวาของการแจกแจงเท่านั้น เนื่องจากเป็นบริเวณที่สำคัญที่เกี่ยวข้องกับการตัดสินใจยอมรับหรือปฏิเสธสมมุติฐานหลัก (H_0) ที่ว่าตัวแบบการถดถอยลอจิสติคเหมาะสมกับข้อมูล ผลที่ได้จึงแตกต่างจากทฤษฎีตัวอย่างขนาดใหญ่ที่ว่า สถิติ Deviance และ Score จะมีการแจกแจงใกล้การแจกแจงแบบ Chi-square ต่อเมื่อตัวอย่างมีขนาดใหญ่ [2, 3, 5] ซึ่งเป็นการกล่าวถึงบริเวณทั้งหมดของการแจกแจง และเมื่อกำหนด α = 0.01 ไม่ควรใช้ตัวแบบการถดถอยลอจิสติคกับตัวอย่างขนาด 10 เนื่องจากไม่มีสถิติตัวใดเหมาะสมที่จะใช้พิจารณาความเหมาะสมของตัวแบบ แม้ว่าที่ α = 0.01 สถิติ Likelihood Ratio จะมีค่าใกล้ค่าสถิติ Chi-square มากที่สุดก็ตาม แต่ถ้าสมมุติให้สถิติ Likelihood Ratio มีการแจกแจงแบบ Chi-square แล้วระดับนัยสำคัญของการทดสอบจะเปลี่ยนไปมากคือจาก 0.01 เป็น 0.28 จึงไม่เหมาะสมที่จะนำไปใช้

เมื่อตัวอย่างมีขนาด 30, 50 และ 100 สถิติ HL มีการแจกแจงใกล้การแจกแจงแบบ Chi-square มากที่สุดที่ระดับนัยสำคัญ 0.01, 0.05 และ 0.10 Hosmer and Lemeshow [4] นำเสนอสถิติ HL สำหรับตรวจสอบความเหมาะสมของตัวแบบการถดถอยลอจิสติค โดยแบ่งค่าสังเกตออกเป็น 10 กลุ่มเท่าๆ กัน โดยใช้ค่าความน่าจะเป็นที่พยากรณ์ได้เป็นตัวแบ่งกลุ่ม แล้วคำนวณค่าสถิติ HL จากจำนวนครั้งของการเกิดเหตุการณ์ในแต่ละกลุ่ม และจำนวนครั้งที่คาดหวังของการเกิดเหตุการณ์ในแต่ละกลุ่ม โดยใช้หลักการเดียวกันกับการคำนวณค่าสถิติ Pearson Chi-square สถิติ HL จะมีการแจกแจงใกล้การแจกแจงแบบ Chi-square ที่มี DF = 8 เมื่อตัวอย่างมีขนาดใหญ่ คือมีขนาดใหญ่กว่า 400 แต่การศึกษาในครั้งนี้พบว่าสถิติ HL มีการแจกแจงใกล้การแจกแจงแบบ Chi-square ที่มี DF = 8 เมื่อตัวอย่างมีขนาดเล็ก เนื่องจากพิจารณาเฉพาะที่หางขวาของการแจกแจงเท่านั้น อย่างไรก็ตามถ้าสมมุติให้ Deviance, Score และ HL มีการแจกแจงแบบ Chi-square เมื่อตัวอย่างมีขนาดเล็ก ระดับนัยสำคัญของการทดสอบสมมุติฐานจะเปลี่ยนไปจากที่กำหนดไว้ ดังนั้นในการนำไปใช้ จึงต้องพิจารณาค่าระดับนัยสำคัญของการทดสอบที่เปลี่ยนไปด้วย ว่าอยู่ในเกณฑ์ที่ยอมรับได้หรือไม่

เอกสารอ้างอิง

1. Smith, T., Smith, B., and Honner, W. 2009. PROC GENMOD for Analysis of Correlated Outcome Data Using the LOGIT Link Function. Available from: URL: http://www.lexjansen.com/wuss/2004/data_analysis/ i_das_proc_genmod_for_analys.pdf. 27 September 2009.
2. Agresti, A. 2002. Categorical Data Analysis. 2nd Edition. New Jersey. John Wiley & Sons.
3. Collett, D. 2003. Modelling Binary Data. 2nd Edition. Florida. Chapman and Hall/CRC.
4. Hosmer, D. W., and Lemeshow, S. 2000. Applied Logistic Regression, 2nd Edition. New York. John Wiley and Sons.
5. Corrente, J. E., Gomes, S. G., and Paula, A. 2002. Correction Factor for the Deviance Distribution of Proportion Data. *Revista de Matematicae Estat\'\i\stica* 20: 175-193.

6. Chen, T-H., Chen, C-Y., Yang, H-C. P., and Chen, C-W. 2008. A Mathematical Tool for Inference in Logistic Regression with Small-Sized Data Sets: A Practical Application on ISW-Ridge Relationships. *Mathematical Problems in Engineering* Article ID 186372: 1-12.
7. Pulkstenis, E., and Robinson, T. J. 2002. Two Goodness-of-Fit Tests for Logistic Regression Models with Continuous Covariates. *Statistics in Medicine* 21(1): 79-93.
8. Kramer, A. A. 2008. Using ODS to Perform Simulations on Statistics for SAS Procedures. Available from: URL: http://analytics.ncsu.edu/sesug/2005/SA05_05.PDF. 30 September 2008.

ได้รับบทความวันที่ 14 กันยายน 2552
ยอมรับตีพิมพ์วันที่ 18 ตุลาคม 2552