NU Science Journal 2007; 4(2): 154 - 164

การวิเคราะห์ตัวแบบโลจิทสะสมเมื่อมีการยุบกลุ่มของตัวแปรตอบสนองเชิงกลุ่มแบบมีลำดับ

ชญาณัฐ โพธิ์นอก จิราภรณ์ พลศรี และ วีรานันท์ พงศาภักดี*

The Analysis of Cumulative Logit Model When Collapsing Categories of Outcome

Chayanat Phonork, Jiraporn Pholsri and Veeranun Pongsapukdee*

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

*Corresponding author. E-mail: veeranun@su.ac.th

บทคัดย่อ

ตัวแบบที่นิยมใช้ในการวิเคราะห์ข้อมูลเมื่อตัวแปรตอบสนองเป็นตัวแปรเชิงกลุ่มแบบมีลำดับ มีหลายตัวแบบ งานวิจัยนี้ศึกษาเฉพาะตัวแบบ proportional odds ซึ่งมีข้อสมมติว่าพารามิเตอร์ที่บอกความชันคงที่เสมอ โดยสนใจศึกษาว่าเมื่อมีการยุบกลุ่มของตัวแปรตอบสนองระดับที่อยู่ติดกันและระดับน้อยที่สุดกับมากที่สุดเข้าด้วยกันแล้วข้อสมมติของตัวแบบ proportional odds ยังเป็นจริงหรือไม่ พร้อมทั้งศึกษาผลลัพธ์อื่น ๆ ของตัวแบบด้วย จากข้อมูลที่นำมาศึกษา ผลการวิเคราะห์พบว่าการยุบกลุ่มของตัวแปรตอบสนองในระดับที่อยู่ติดกันเข้าด้วยกันนั้นความชันคงที่ แต่เมื่อยุบระดับน้อยที่สุดกับมากที่สุดเข้าด้วยกันพบว่าความชันไม่คงที่ ซึ่งไม่สอดคล้องกับข้อสมมติของตัวแบบ proportional odds และยังพบว่าการยุบกลุ่มของตัวแปรตอบสนองระดับที่อยู่ติดกันเข้าด้วยกันค่าประมาณสัมประสิทธิ์โลจิทสะสมไม่เปลี่ยนแปลงมากนัก และค่า P-value มีแนวโน้มเพิ่มขึ้น แต่ยังน้อยกว่านัยสำคัญทางสถิติ 0.05 ของการทดสอบสัมประสิทธิ์โลจิทสะสมจากตัวสถิติวาลด์คงเดิม แต่เมื่อยุบระดับน้อยที่สุดกับมากที่สุดเข้าด้วยกันพบว่าค่าประมาณสัมประสิทธิ์โลจิทสะสมเปลี่ยนแปลงไปมาก ค่า P-value สูงกว่าเดิมและมากกว่าระดับนัยสำคัญทางสถิติ 0.05 ของการทดสอบสัมประสิทธิ์โลจิทสะสมจากตัวสถิติวาลด์และตัวสถิติอัตราส่วนภาวะน่าจะเป็น

คำสำคัญ : ตัวแบบโลจิทสะสม ตัวแบบ proportional odds ตัวแปรตอบสนองเชิงกลุ่มแบบมีลำดับ การยุบกลุ่ม

Abstract

There are many popular models for analyzing data with an ordered categorical outcome. This research considers the proportional odds model that poses the equality of the log-odds ratio or the slope parameters over all the cut-off point as the proportional odds assumption. The objectives are to investigate the validity of this assumption when collapsing of the adjacent categories and the extreme categories of ordinal response and to study results from goodness-of-fit statistics. From the sample data, the research results show that, when collapsing of adjacent categories of ordinal response, the proportional odds assumption is valid whereas the point estimates remain almost unchanged and are significant. However, when collapsing of the extreme categories of ordinal response, the proportional odds assumption is not hold, whereas the point estimates are changed and are not significant, from Wald test and the likelihood ratio test.

Keywords: Cumulative Logit Models, Proportional Odds Model, Ordered Categorical Variable, Collapsing.

บทนำ

ข้อมูลเชิงกลุ่มแบบมีลำดับ พบเห็นอยู่ทั่วไป ทั้งทางด้านการแพทย์ เช่น อาการของโรค (รุนแรง, ปานกลาง, ไม่รุนแรง) ด้านการเกษตร เช่น อาการของแมลงเมื่อได้รับยาฆ่าแมลง (ตาย, สลบ, ไม่แสดงอาการ) ตลอดจนด้านสังคมและพฤติกรรมศาสตร์ เช่น ระดับความพึงพอใจ (มากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด) เป็นต้น

การวิเคราะห์ข้อมูลเชิงกลุ่มแบบมีลำดับ เมื่อตัวอย่างมีขนาดเล็กนั้น เราควรยุบกลุ่มของตัวแปรตอบสนองเข้าด้วยกันเมื่อจำนวนคาดหวัง (Expected number) ในเซลล์มีจำนวนน้อยๆ เพื่อทำให้ตัวสถิติที่ใช้ทดสอบลู่เข้าสู่การแจกแจงแบบไคสแควร์เมื่อใกล้อนันต์ (asymptotic approximation: Altman, 1991) แต่บางครั้งหากความถี่คาดหวังในเซลล์มีจำนวนมากเราก็สามารถยุบกลุ่มของตัวแปรตอบสนองที่อยู่ติดกันเข้าด้วยกันเพื่อทำให้การวิเคราะห์และการสรุปผลง่ายยิ่งขึ้น (Manor *et al.*, 2000)

การวิเคราะห์ข้อมูลโดยใช้ตัวแบบ proportional odds นั้นเป็นที่รู้กันดีว่าพารามิเตอร์ที่บอกความชัน (slope parameter)ในตัวแบบนี้จะไม่เปลี่ยนแปลงเมื่อมีการยุบกลุ่มของตัวแปรตอบสนองเชิงกลุ่มแบบมีลำดับที่อยู่ติดกันเข้าด้วยกัน (Peterson *et al.*, 1990; Greenland, 1994) มีเพียงพารามิเตอร์ที่บอกจุดตัด (intercept parameter) เท่านั้นที่เปลี่ยนแปลง กล่าวอีกนัยหนึ่งตัวแบบ proportional odds มี

ข้อสมมติว่าอิทธิพลของตัวแปรอธิบายคงที่ สำหรับทุกกลุ่มของตัวแปรตอบสนอง (Dobson, 2002) ดังนั้นจึงมีความเหมาะสมที่จะพิจารณาการยุบกลุ่มใดๆ ของตัวแปรตอบสนองในตัวแบบนี้เพื่อทำให้จำนวนค่าสังเกตในแต่ละกลุ่มสูงขึ้นและทำให้การแจกแจงลู่เข้าสู่การแจกแจงแบบไคสแควร์เมื่อใกล้อนันต์ (asymptotic approximation) ได้เร็วขึ้น

ในการวิจัยครั้งนี้ผู้วิจัยต้องการศึกษาต่อไปว่าในการยุบกลุ่มของตัวแปรตอบสนองแบบมีลำดับในระดับน้อยที่สุดกับระดับมากที่สุดเช่นระดับความคิดเห็นต่อการทำงานของรัฐบาล (เห็นด้วยอย่างมาก เห็นด้วย เฉยๆ ไม่เห็นด้วย ไม่เห็นด้วยอย่างมาก) อาจยุบระดับเห็นด้วยอย่างมากและไม่เห็นด้วยอย่างมากเข้าด้วยกันเพื่อเปรียบเทียบกลุ่มประชาชนมีความคิดเห็นที่รุนแรงต่อการทำงานของรัฐบาลกับประชาชนที่มีความคิดเห็นกลางๆ การยุบกลุ่มของตัวแปรตอบสนองแบบมีลำดับระดับน้อยที่สุดกับระดับมากที่สุดนั้นให้ผลเช่นเดียวกับการยุบระดับที่อยู่ติดกันหรือไม่ (Murad *et al.*, 2003) เนื่องจากตัวแบบ proportional odds มีข้อสมมติว่าสัมประสิทธิ์โลจิทสะสมไม่เปลี่ยนแปลง (invariant) ในแต่ละ cut point ของกลุ่มของตัวแปรตอบสนองหรือสัมประสิทธิ์โลจิทสะสมไม่ขึ้นกับกลุ่มของตัวแปรตอบสนองนั่นเอง (Ananth and Kleinbaum, 1997) จึงอาจยุบกลุ่มใดๆ ที่สนใจได้ โดยทดสอบนัยสำคัญของสัมประสิทธิ์โลจิทสะสมด้วยตัวสถิติวาลด์ (Wald test) และตัวสถิติอัตราส่วนภาวะน่าจะเป็นที่ระดับนัยสำคัญ 0.05 และศึกษาความเหมาะสมของตัวแบบที่ได้โดยพิจารณาจากตัวสถิติ -2Log L, AIC (Akaike's Information Criterion) และ SC (Schwarz Criterion)

ตัวแบบโลจิมสะสม (Cumulative Logit Model)

ตัวแบบโลจิทสะสมถูกเสนอโดย Walker และ Duncan ในปี 1967 ต่อมาได้ถูกนำมาปรับใช้และเรียกว่า proportional odds model โดย McCullagh (1980) พิจารณาตัวแปรตอบสนองเชิงกลุ่มแบบมีลำดับ ที่มี k กลุ่ม ซึ่งมีการแจกแจงแบบพหุนาม (multinomial distribution) สมมติว่า Y_i เป็นค่าตอบสนองของ Y เมื่อกำหนด X_i, $i=1,\ldots,n$ ความสัมพันธ์ระหว่างตัวแปรอธิบายกับความน่าจะเป็นสะสมของการแจกแจงของ Y กำหนดโดย

$$\log\left[\frac{P(Y_i \le j|\mathbf{X_i})}{P(Y_i > j|\mathbf{X_i})}\right] = \alpha_j + \boldsymbol{\beta}'\mathbf{X_i}, \quad i=1,...,n, \quad j=1,...,k-1 \tag{1}$$

โดยที่จุดตัด $\alpha_1 \le \alpha_2 \le \ldots \le \alpha_{k-1}$ ($\alpha_j : j=1,\ldots,k-1$) คือ log odds ของ Y ที่น้อยกว่าหรือเท่ากับ j เมื่อ X เป็นศูนย์ นั่นคือ $P(Y \le j) = \frac{e^{\alpha_j}}{1+e^{\alpha_j}}$ เมื่อ X มีค่าเป็นศูนย์ ส่วนเวกเตอร์สัมประสิทธิ์ $\boldsymbol{\beta}$ แทน log odds ratio ของ Y ที่น้อยกว่าหรือเท่ากับ j เมื่อ Xi , l = 1, 2, … , p เพิ่มขึ้น 1 หน่วย ขณะที่ตัวแปรอธิบายตัวอื่น ๆ มีค่าคงที่ ทำให้ได้ว่า $\boldsymbol{\beta}$ ไม่ขึ้นกับ j ดังนั้นตัวแบบนี้มีข้อสมมติว่าความสัมพันธ์ของ X และ Y เป็นอิสระกับ j หรือจุดที่แบ่ง Y ออกเป็นสองกลุ่ม โดย McCullagh (1980) เรียกข้อ

สมมติของความเท่ากันของ log odds ratio ในทุก ๆ จุดตัดว่า proportional odds assumption และเรียกตัวแบบนี้ว่า proportional odds model ประมาณค่าพารามิเตอร์ $\boldsymbol{\beta}$ ในตัวแบบ proportional odds นี้ด้วย $\boldsymbol{\beta}$ โดยใช้วิธีฟิชเชอร์-สกอริง (Fisher scoring method) ในการแก้สมการหาตัวประมาณภาวะน่าจะเป็นสูงสุด (maximum likelihood method) และประมาณค่า odds ratio ได้จาก $e^{\hat{\beta}}$ และใช้ตัวสถิติวาลด์ในการทดสอบสมมติฐาน $H_0 : \beta_l = 0$ การปฏิเสธสมมติฐานว่างหมายความว่าตัวแปรอธิบายนั้นมีความสัมพันธ์กับตัวแปรตอบสนองแบบมีลำดับ

การทดสอบเกี่ยวกับพารามิเตอร์ของตัวแบบ

1. สถิติทดสอบอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio Test Statistic : LRTS)

สถิติทดสอบอัตราส่วนภาวะน่าจะเป็นใช้สำหรับทดสอบว่ามีสัมประสิทธิ์ของตัวแปรอธิบายในตัวแบบอย่างน้อย 1 ตัวที่ไม่เป็นศูนย์ สถิติทดสอบอัตราส่วนภาวะน่าจะเป็นคำนวณได้จากผลต่างของ -2 Log L (null model) + (-2 Log L (fitted model)) ในการทดสอบสมมติฐาน $\beta_l = 0$ ทุกค่า l, l= 1, 2, …, p จะใช้ตัวสถิติทดสอบอัตราส่วนภาวะน่าจะเป็น ซึ่งมีการแจกแจงแบบไคสแควร์ด้วยองศาแห่งความเป็นอิสระที่มีค่าเท่ากับ p

2. การทดสอบด้วยตัวสถิติวาลด์ (Wald test)

โดยการเปรียบเทียบค่า $\hat{\beta}_l$ กับค่าความคลาดเคลื่อนมาตรฐาน (Standard error) ของค่า $\hat{\beta}_l$ นั้นๆ กล่าวคือ $W_l = \frac{\hat{\beta}_l}{s.e.(\hat{\beta}_l)}$ โดยที่ตัวสถิติ W มีการแจกแจงแบบปกติมาตรฐาน หรือ $\left(\frac{\hat{\beta}_l}{s.e.(\hat{\beta}_l)}\right)^2$ มีการแจกแจงแบบไคสแควร์ที่มีองศาความเป็นอิสระเท่ากับ 1 โดยที่ $H_0 : \beta_l = 0$ เป็นสมมติฐานในการทดสอบ

การทดสอบความเหมาะสมของตัวแบบ

ตัวสถิติที่ใช้ในการตรวจสอบความเหมาะสมของตัวแบบที่นิยมใช้มีดังนี้

1. Akaike's Information Criterion : AIC

Akaike's Information Criterion หรือAIC เป็นเกณฑ์ที่มีลักษณะทั่วไปซึ่งสามารถที่จะประยุกต์กับตัวแบบใด ๆ ที่สามารถจะประมาณค่าได้โดยวิธีภาวะน่าจะเป็นสูงสุด (maximum likelihood) และAIC นี้ยังนิยมใช้กันทั่วไปในตัวแบบที่ไม่เป็นเชิงเส้น (non-linear models) สำหรับ AIC ของแต่ละตัวแบบสามารถคำนวณได้จาก $AIC = -2\log L + 2k$ เมื่อ $\log L$ แทนค่าของ Log-likelihood ที่ได้มาจากการประมาณพารามิเตอร์ในตัวแบบด้วยวิธีภาวะน่าจะเป็นสูงสุด (maximum

likelihood) โดยที่ k แทนจำนวนพารามิเตอร์ในตัวแบบ ค่า AIC ที่น้อยที่สุดถือว่าตัวแบบมีความเหมาะสมมากที่สุดสำหรับข้อมูล (Burnham and Anderson, 1992 : 16-30)

2. Schwarz Criterion : SC

Schwarz Criterion หรือ SC ซึ่งนิยามโดย - 2 log L + k log (n) โดยที่ k แทนจำนวนพารามิเตอร์ในตัวแบบ และ n คือขนาดตัวอย่าง ค่า SC ที่น้อยที่สุดถือว่าตัวแบบมีความเหมาะสมสำหรับข้อมูลมากที่สุด (Raftery, 1986 ,1995)

3. -2 Log L

-2 Log L คือ ตัวสถิติใช้ในการตรวจสอบความเหมาะสมของตัวแบบ เมื่อ L คือ ภาวะน่าจะเป็นของตัวแบบโลจิทสะสม

$$L = \prod_{i=1}^{n}\left[\prod_{j=1}^{k}\left(P(Y_i \le j \mid \mathbf{X_i}) - P(Y_i \le j-1 \mid \mathbf{X_i})\right)^{y_{ij}}\right]$$

$$= \prod_{i=1}^{n}\left\{\prod_{j=1}^{k}\left(\frac{\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{X_i})}{1+\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{X_i})} - \frac{\exp(\alpha_{j-1} + \boldsymbol{\beta}'\mathbf{X_i})}{1+\exp(\alpha_{j-1} + \boldsymbol{\beta}'\mathbf{X_i})}\right)^{y_{ij}}\right\}$$

ถ้าตัวแบบที่ต้องการเลือกมีหลายตัวแบบ ตัวแบบที่มีความเหมาะสมกับข้อมูลมากที่สุดคือตัวแบบที่มีค่า -2 Log L น้อยที่สุด

การประยุกต์กับข้อมูลการรักษาโรคมะเร็งปอดด้วยยาเคมีบำบัด

ข้อมูลที่นำมาศึกษาเป็นข้อมูลจริงจาก Holtbrugge และ Schumacher (1991) ซึ่งเก็บข้อมูลจากผู้ป่วยโรคมะเร็งปอดเพื่อเปรียบเทียบวิธีการรักษา 2 วิธี คือให้รับยาทีละชนิดและรับยาชนิดสลับกัน โดยพิจารณาจากขนาดของเซลล์มะเร็ง และใช้ proportional odds model และ stereotype model ในการวิเคราะห์ข้อมูล ความแตกต่างของทั้งสองตัวแบบพิจารณาจากการประมาณค่าพารามิเตอร์และนัยสำคัญของการทดสอบ ผลการวิจัยพบว่าค่าประมาณพารามิเตอร์ที่ได้จาก proportional odds model มีความเอนเอียงน้อยกว่า stereotype model แต่ stereotype model จะให้ค่าประมาณพารามิเตอร์ที่ดีขึ้นเมื่อมีการยุบระดับที่อยู่ติดกันเข้าด้วยกัน

ตาราง 1 ผลของการรักษาโรคมะเร็งปอดด้วยยาเคมีบำบัดจำแนกตามเพศและวิธีการรักษา

วิธีการรักษา	เพศ	ผลของการรักษาด้วยยาเคมี				รวม
		หายดี (1)	ทุเลา (2)	ไม่เปลี่ยนแปลง (3)	รุนแรงขึ้น (4)	(299)
รับยาทีละชนิด	ชาย	26	29	45	28	128
	หญิง	2	5	12	4	23
สลับยา	ชาย	20	20	44	41	125
	หญิง	1	3	7	12	23

ตาราง 2 การทดสอบ proportion odds assumption สำหรับข้อมูลการรักษาโรคมะเร็งปอดด้วยยาเคมีบำบัด

การยุบกลุ่ม	ค่าสถิติไคสแควร์	D.F.	P-value
ไม่มีการยุบกลุ่ม	6.7018	4	0.1525
(1, 2) , 3, 4	1.3106	3	0.5193

หมายเหตุ : ยังมีการยุบกลุ่มอีก 3 แบบ คือ 1. ยุบกลุ่ม 1, 2 และ 3 เข้าด้วยกัน 2. ยุบกลุ่ม 1 และ 2 เข้าด้วยกัน กับกลุ่ม 3 และ 4 เข้าด้วยกัน และ 3. ยุบกลุ่ม 1 และ 4 เข้าด้วยกัน กับกลุ่ม 2 และ 3 เข้าด้วยกัน ซึ่งเป็นการยุบให้เหลือ 2 กลุ่ม จึงมีสมการโลจิทเพียง 1 สมการ ไม่จำเป็นต้องตรวจสอบข้อสมมติของตัวแบบ proportional odds ซึ่งจะทำเฉพาะกรณียุบให้เหลือ 3 กลุ่มขึ้นไปเท่านั้น

จากตาราง 2 พบว่า เมื่อไม่มีการยุบกลุ่มค่า P-value มากกว่าระดับนัยสำคัญทางสถิติ 0.05 พบว่าข้อมูลสอดคล้องกับข้อสมมติของตัวแบบ proportional odds เมื่อทำการยุบกลุ่ม 1 และ 2 เข้าด้วยกัน พบว่าค่า P-value มากกว่า 0.05 พบว่าข้อสมมติของตัวแบบ proportional odds ยังเป็นจริง หรือข้อมูลใหม่ที่ได้จากการยุบกลุ่ม 1 และ 2 เข้าด้วยกันสอดคล้องกับข้อสมมติของตัวแบบ proportional odds

ตาราง 3 ค่าประมาณสัมประสิทธิ์โลจิทสะสมสำหรับข้อมูลการรักษาโรคมะเร็งปอดด้วยยาเคมีบำบัด

การยุบกลุ่ม	พารามิเตอร์	$\hat{\beta}$	S.E$(\hat{\beta})$	Exp$(\hat{\beta})$	Wald Test	D.F.	P-value
ไม่มีการ	วิธีการรักษา	-0.5807	0.2119	0.56	7.5131	1	0.0061
ยุบกลุ่ม	เพศ	-0.5414	0.2953	0.582	3.3619	1	0.0667
(1, 2) ,3 ,4	วิธีการรักษา	-0.6118	0.216	0.542	8.026	1	0.0046
	เพศ	-0.5046	0.2979	0.604	2.8691	1	0.0903
(1, 2, 3), 4	วิธีการรักษา	-0.7313	0.2633	0.481	7.7165	1	0.0055
	เพศ	-0.3578	0.3453	0.699	1.074	1	0.3000
(1, 2) , (3 ,4)	วิธีการรักษา	-0.5022	0.2457	0.605	4.1781	1	0.0409
	เพศ	-0.6543	0.3715	0.52	3.1025	1	0.0782
(1, 4) , (2 ,3)	วิธีการรักษา	0.4174	0.234	0.96	3.1833	1	0.0744
	เพศ	-0.1734	0.3267	0.443	0.2816	1	0.5956

จากตาราง 3 พบว่าเมื่อไม่มีการยุบกลุ่ม ค่าประมาณสัมประสิทธิ์โลจิทสะสมสำหรับตัวแปรวิธีการรักษา เท่ากับ -0.5807 และมีนัยสำคัญทางสถิติที่ระดับ 0.05 นั่นคือวิธีการรักษามีผลต่อผลการรักษา ส่วนเพศไม่มีผลต่อผลการรักษา เมื่อทำการยุบกลุ่มของตัวแปรตอบสนองระดับติดกันเข้าด้วยกัน พบว่าค่าประมาณสัมประสิทธิ์โลจิทสะสมสำหรับตัวแปรวิธีการรักษา ยังใกล้เคียงกับค่าประมาณสัมประสิทธิ์โลจิทสะสมเมื่อไม่มีการยุบกลุ่ม และยังมีนัยสำคัญเช่นเดิม แต่เมื่อยุบกลุ่มระดับน้อยที่สุดกับระดับมากที่สุดเข้าด้วยกัน พบว่าค่าประมาณสัมประสิทธิ์โลจิทสะสมสำหรับตัวแปรวิธีการรักษาเปลี่ยนแปลงไปมาก และไม่มีนัยสำคัญเหมือนก่อน

ตาราง 4 การเปรียบเทียบความเหมาะสมของตัวแบบสำหรับข้อมูลการรักษาโรคมะเร็งปอดด้วยยาเคมีบำบัดด้วยตัวสถิติ -2 Log L, SC และ AIC

ตัวแบบ	-2 Log L	SC	AIC
เมื่อไม่มีการยุบกลุ่ม	789.057	817.559	799.057
(1, 2) ,3 ,4	642.778	665.580	650.778
(1, 2, 3), 4	348.019	365.120	354.019
(1, 2) , (3 ,4)	381.263	398.365	387.263
(1, 4) , (2 ,3)	407.809	424.910	413.809

จากตาราง 4 พบว่าตัวแบบเมื่อยุบกลุ่ม 1 2 และ 3 เข้าด้วยกัน หรือรวมระดับ หายดี ทุเลา และไม่เปลี่ยนแปลงเข้าด้วยกัน ให้ค่าสถิติ -2 Log L, SC และ AIC ต่ำที่สุด จึงเป็นตัวแบบที่เหมาะสมที่สุด ดังนั้นสมการโลจิทที่เหมาะสมคือ $\text{logit}[P(Y \le 1 | X_1, X_2)] = 2.4618 - 0.7313X_1 - 0.3578X_2$ หมายความว่า ผู้ป่วยที่รักษาแบบรับยาทีละชนิดมีโอกาสที่อาการจะไม่รุนแรงขึ้น (หายดี ทุเลาและไม่เปลี่ยนแปลง) เป็น exp (-0.7313) = 0.481 เท่าของผู้ป่วยที่รักษาแบบสลับยา

การประยุกต์กับข้อมูลความคิดเห็นทางการเมือง

ข้อมูลความคิดเห็นทางการเมืองเป็นข้อมูลจริงจาก Senie และคณะ (1981)

ตาราง 5 ความถี่ของประชาชนที่มีความคิดเห็นทางการเมืองแบบต่าง ๆ จำแนกตามการเข้าร่วมพรรคการเมือง

การเข้าร่วมพรรคการเมือง	ความคิดเห็นทางการเมือง (Political Ideology)					
	Very Liberal (1)	Slightly Liberal (2)	Moderate (3)	Slightly Conservative (4)	Very Conservative (5)	รวม (835)
Democratic	80	81	171	41	55	428
Republican	30	46	148	84	99	407

ตาราง 6 การทดสอบ proportion odds assumption สำหรับข้อมูลการความคิดเห็นทางการเมือง

การยุบกลุ่ม	ค่าสถิติไคสแควร์	D.F.	P-value
ไม่มีการยุบกลุ่ม	3.9106	3	0.2713
(1, 2) , 3, 4, 5	3.6374	2	0.1622
(1, 2), 3, (4, 5)	0.4214	2	0.8100
1, 2, 3, (4, 5)	0.1852	1	0.6670
(1, 5) , (2, 4) , 3	19.6378	2	<.0001

จากตาราง 6 เมื่อไม่มีการยุบกลุ่ม ค่า P-value มากกว่าระดับนัยสำคัญทางสถิติ 0.05 พบว่าข้อมูลสอดคล้องกับข้อสมมติของตัวแบบ proportional odds เมื่อทำการยุบกลุ่มระดับติดกันเข้าด้วยกัน

พบว่าค่า P-value มากกว่า 0.05 แสดงว่าข้อสมมติของตัวแบบ proportional odds ยังเป็นจริง แต่เมื่อยุบกลุ่มระดับน้อยที่สุดกับระดับมากที่สุดเข้าด้วยกัน พบว่าค่า P-value น้อยกว่า 0.05 แสดงว่าข้อสมมติของตัวแบบ proportional odds ไม่เป็นจริง หรือข้อมูลใหม่ที่เกิดจากการยุบระดับน้อยที่สุดกับระดับมากที่สุดเข้าด้วยกัน ไม่สอดคล้องกับข้อสมมติของตัวแบบ proportional odds นั่นเอง

ตาราง 7 ค่าประมาณสัมประสิทธิ์โลจิทสะสมสำหรับข้อมูลความคิดเห็นทางการเมือง

การยุบกลุ่ม	พารามิเตอร์	$\hat{\beta}$	S.E$(\hat{\beta})$	Exp$(\hat{\beta})$	Wald Test	D.F.	P-value
ไม่มีการยุบกลุ่ม	พรรคการเมือง	0.9745	0.1291	2.6498	57.0182	1	<.0001
(1, 2) , 3, 4, 5	พรรคการเมือง	0.9661	0.1299	2.6277	55.2784	1	<.0001
(1, 2), 3, (4, 5)	พรรคการเมือง	1.0139	0.1312	2.7563	59.6735	1	<.0001
1, 2, 3, (4, 5)	พรรคการเมือง	1.0059	0.1322	2.7344	57.9280	1	<.0001
(1, 5) , (2, 4) , 3	พรรคการเมือง	-0.0864	0.1275	0.9172	0.4589	1	0.4981

จากตาราง 7 เมื่อไม่มีการยุบกลุ่ม ค่าประมาณสัมประสิทธิ์โลจิทสะสมสำหรับตัวแปรพรรคการเมือง เท่ากับ 0.9745 และมีนัยสำคัญทางสถิติที่ระดับ 0.05 เมื่อทำการยุบกลุ่มของตัวแปรตอบสนองระดับติดกันเข้าด้วยกัน พบว่าค่าประมาณสัมประสิทธิ์โลจิทสะสมสำหรับตัวแปรพรรคการเมือง ยังใกล้เคียงกับค่าประมาณสัมประสิทธิ์โลจิทสะสมเมื่อไม่มีการยุบกลุ่ม และยังมีนัยสำคัญเช่นเดิม แต่เมื่อยุบกลุ่มระดับน้อยที่สุดกับระดับมากที่สุดเข้าด้วยกัน พบว่าค่าประมาณสัมประสิทธิ์โลจิทสะสมสำหรับตัวแปรพรรคการเมืองเปลี่ยนแปลงไปมาก และไม่มีนัยสำคัญทางสถิติเหมือนก่อน

ตาราง 8 การเปรียบเทียบความเหมาะสมของตัวแบบสำหรับข้อมูลความคิดเห็นทางการเมือง ด้วยตัวสถิติ -2 Log L, SC และ AIC

ตัวแบบ	-2 Log L	SC	AIC
ไม่มีการยุบกลุ่ม	2474.985	2508.622	2484.985
(1, 2) , 3, 4, 5	2149.556	2176.466	2157.556
(1, 2), 3, (4, 5)	2088.227	2115.137	2096.227
1, 2, 3, (4, 5)	1762.834	1783.016	1768.834
(1, 5) , (2, 4) , 3	1825.226	1845.409	1831.226

จากตาราง 8 ตัวแบบเมื่อยุบกลุ่ม 4 และ 5 เข้าด้วยกัน หรือรวมกลุ่มอนุรักษ์นิยมและอนุรักษ์นิยมรุนแรงเข้าด้วยกัน ให้ค่าสถิติ -2 Log L, SC และ AIC ต่ำที่สุด จึงเป็นตัวแบบที่เหมาะสมที่สุด ดังนั้นสมการโลจิทที่เหมาะสมมี 2 สมการ สมการแรก คือ $\log it[P(Y \le 1 \mid X)] = -1.4990 + 1.0059X$ หมายความว่าประชาชนที่เข้าร่วมกับพรรคฝ่ายประชาธิปไตยมีโอกาสที่จะมีความคิดเห็นทางการเมืองอยู่ในกลุ่มประชาธิปไตยรุนแรงเป็น exp(1.0059) = 2.7344 เท่าของประชาชนที่เข้าร่วมกับพรรคฝ่ายรีพับริกัน และสมการที่สอง คือ $\log it[P(Y \le 2 \mid X)] = 0.2135 + 1.0059X$ หมายความว่าประชาชนที่เข้าร่วมกับพรรคฝ่ายประชาธิปไตยมีโอกาสที่จะมีความคิดเห็นทางการเมืองอยู่ในกลุ่มประชาธิปไตยหรือประชาธิปไตยรุนแรงเป็น exp(1.0059) = 2.7344 เท่าของประชาชนที่เข้าร่วมกับพรรคฝ่ายรีพับริกัน

สรุปผลการวิจัย

ผลของการวิจัยนี้สรุปจากข้อมูลจริง 2 ชุด พบว่าข้อสมมติของตัวแบบ proportional odds ที่ว่าอิทธิพลของตัวแปรอธิบายคงที่ สำหรับทุกกลุ่มของตัวแปรตอบสนองเป็นจริงเมื่อยุบกลุ่มระดับติดกันเข้าด้วยกัน แต่อาจไม่เป็นจริงเมื่อมีการยุบกลุ่มในระดับน้อยที่สุดกับระดับมากที่สุดเข้าด้วยกัน ส่วนผลลัพธ์อื่น ๆ พบว่าเมื่อยุบกลุ่มระดับติดกันเข้าด้วยกัน ค่าประมาณสัมประสิทธิ์โลจิทสะสมไม่เปลี่ยนแปลงมากนัก และยังมีนัยสำคัญทางสถิติเช่นเดิม แต่ในกรณีที่ยุบกลุ่มในระดับน้อยที่สุดกับระดับมากที่สุดเข้าด้วยกัน พบว่าค่าประมาณสัมประสิทธิ์โลจิทสะสมเปลี่ยนแปลงไปมากและไม่มีนัยสำคัญทางสถิติ

ข้อเสนอแนะ

จากผลสรุปของงานวิจัย ที่ว่าข้อสมมติของตัวแบบ Proportional odds อาจไม่เป็นจริงเมื่อมีการยุบกลุ่มในระดับน้อยที่สุดกับระดับมากที่สุดเข้าด้วยกัน ดังนั้นถ้าต้องการที่จะเปรียบเทียบกรณีนี้ ควรเลือกใช้ตัวแบบอื่น ๆ สำหรับวิเคราะห์ข้อมูลเชิงกลุ่มแบบมีลำดับ

งานวิจัยนี้ใช้ข้อมูลเพียง 2 ชุด ดังนั้นในการวิจัยครั้งต่อไปอาจตรวจสอบข้อสมมติของตัวแบบ proportional odds เมื่อมีการการยุบกลุ่มจากข้อมูลมากขึ้นโดยอาจใช้ข้อมูลตัวอย่างหลากหลายตลอดจนระดับนัยสำคัญทางสถิติระดับต่าง ๆ ด้วย หรืออาจมีการจำลอง แบบข้อมูลเพื่อตรวจสอบข้อสมมติดังกล่าว

เอกสารอ้างอิง

Ananth, C. V. and Kleinbaum, D.G. (1997). Regression Models for Ordinal Responses: A Review of Methods and Applications. International Journal of Epidemiology, *26*, 1323-1333.

Altman, D.G. (1991). Practical Statistics for Medical Research. London: Chapman and Hall.

Burnham, K. P. and D. R. Anderson. (1992). Data-based selection of an appropriate biological model: the key to modern data analysis. Pages 16-30 in Wildlife 2001: Populations, McCullough, D. R., and R. H. Barrett (eds.) Elsevier Applied Science, New York, USA.

Dobson, A. J. (2002). An Introduction to Generalized Linear Models. London: Chapman and Hall.

Greenland, S. (1994). Alternative Models for Ordinal Logistic Regression. Statistics in Medicine, *13*, 1665-1677.

Holtbrugge, W. and Schumacher M. (1991). A Comparison of Regression Models for The Analysis of Ordered Categorical Data , Applied Statistics, *40*, 249-259.

Manor, O., Matthews, S. and Power, C. (2000). Dichotomous or Categorical Response? Analysing Self-Rated Health and Lifetime Social Class. International Journal of Epidemiology, *29*, 149-157.

McCullagh, P. (1980). Regression Models for Ordinal Data (with discussion). Journal of Royal Statistical Society, Ser. B, *42*, 109-142.

Murad, H., Fleischman, A., Sadetzki, S. Geyer O. and Freedman L. S. (2003). Small sample and Ordered Logistic Regression: Does it Help to Collapse categories of outcome?. The American Statistician, *57*, Number 3, 155-160(6).

Peterson, B. and Harrell, F. E. (1990). Partial Proportional odds models for Ordinal Response Variables. Appl. Statist., *30*, 22-31.

Raftery, A. E. (1986). Choosing Models for Cross-classifications, Amer. Sociol. Rev. 51: 145-146.

Raftery, A.E. (1995). Bayesian model selection in social research (with Discussion). Sociological Methodology, *25*, 111-196.

Senie R.T., Rosen P.P., Lesser M.L. and Kinne D.W. (1981). Breast self-examination and medical examination Related to Breast Cancer Stage. Am J Public Health., *71*, 583-590.

Walker, S. H. and Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. Biometrika ,54 : 167-179.