
การวิเคราะห์เป้าหมายทางการศึกษาของนักเรียนโดยใช้ตัวแบบโลจิต ที่มีตัวแปรตอบสนอง 3 กลุ่ม
แบบมีลำดับและไม่มีลำดับ

The Data Analysis of Students' Goals using the Logit Models for Ordinal and
Nominal Response Categories

เมษิยา แยมเจริญกิจ เกษศิริรินทร์ มัชฌิมา และ วีรานันท์ พงศาภักดี*

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

Maysiya Yamjaroenkit, Kessirin Machima and Veeranun Pongsapukdee*

Department of Statistics, Faculty of Science, Silpakorn University.

บทคัดย่อ

งานวิจัยนี้มุ่งศึกษาอิทธิพลต่าง ๆ ที่อาจมีผลกระทบต่อเป้าหมายของนักเรียน (Students' Goals) ซึ่งแบ่งออกเป็น 3 กลุ่มคือนักเรียนมีผลการเรียนดี นักเรียนเป็นที่ยอมรับของบุคคลทั่วไป และนักเรียนที่มีความสามารถทางด้านกีฬา ภายใต้สัดส่วนของแต่ละกลุ่มเท่ากับ π_1 , π_2 , และ π_3 ตามลำดับ ตัวแปรอธิบายที่ศึกษาอิทธิพลต่าง ๆ รวม 10 ตัวแปร ได้แก่ เพศ ระดับชั้นเรียน อายุ เชื้อชาติ ความเป็นเมือง/ชนบท โรงเรียน ผลการเรียน กีฬา รูปร่างหน้าตา และจำนวนเงินที่ได้รับ การทดลองใช้ข้อมูลจริงจากเว็บไซต์ <http://lib.stat.cmu.edu> ซึ่งเป็นข้อมูลตัวอย่างจากประชากรนักเรียนชั้นประถมศึกษาปีที่ 4-6 ใน 3 รัฐของสหรัฐอเมริกาได้แก่ Ingham, Clinton Counties, และ Michigan จำนวน 478 คน ในฐานข้อมูล Popular Kids การวิเคราะห์ข้อมูลใช้โปรแกรม SAS Version 9.0 และ SPSS version 11.5 ภายใต้ตัวแบบโลจิต (logit models) ทั้งกรณีของตัวแบบ Baseline – category logits models เมื่อตัวแปรตอบสนองเป็นแบบไม่มีลำดับ และกรณีของตัวแบบ Proportional odds models ซึ่งเป็นตัวแบบโลจิตสะสม (Cumulative logit model) เมื่อตัวแปรตอบสนองเป็นแบบมีลำดับ ผลการวิจัยพบว่า ตัวแปรอธิบายที่มีอิทธิพลต่อตัวแปรตอบสนองหรือเป้าหมายของนักเรียน เมื่อตัวแปรตอบสนองเป็นแบบไม่มีลำดับคือ ความเป็นเมือง/ชนบท (Urban) ผลการเรียน (CGPA) และกีฬา (Sports) และตัวแปรอธิบายที่มีอิทธิพลต่อตัวแปรตอบสนอง เมื่อตัวแปรตอบสนองเป็นแบบมีลำดับคือ ความเป็นเมือง/ชนบท และผลการเรียน และยังพบว่า ตัวแบบโลจิต เมื่อตัวแปรตอบสนองเป็นแบบไม่มีลำดับ มีความเหมาะสมกับข้อมูลชุดนี้มากกว่าตัวแบบโลจิตสะสมเมื่อตัวแปรตอบสนองเป็นแบบมีลำดับ เนื่องจากตัวแบบโลจิตให้ค่าสถิติอัตราส่วนภาวะน่าจะเป็น (G^2) และค่า SC (Schwarz Criterion) ที่น้อยกว่าของตัวแบบโลจิตสะสม นอกจากนี้จากการตรวจสอบส่วนเหลือ (residuals) ระหว่างค่าสังเกตและค่าพยากรณ์แต่ละค่าจากตัวแบบยืนยันว่า กรณีของตัวแปรตอบสนองเป็นแบบไม่มีลำดับ มีการกระจายของส่วนเหลือที่มีค่าอยู่ระหว่าง (-2,2) และมีการกระจายตัวที่ดีกว่าของกรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ โดยตัวแบบที่เหมาะสมคือ

$$\log(\hat{\pi}_1/\hat{\pi}_3) = -0.6283 + 0.6079 Urban - 0.2514 CGPA + 0.5741 Sports$$

$$\log(\hat{\pi}_2/\hat{\pi}_3) = -2.5866 + 0.3466 Urban + 0.2468 CGPA + 0.8548 Sports$$

$$\log(\hat{\pi}_1/\hat{\pi}_2) = 1.9583 + 0.2613 Urban - 0.4982 CGPA - 0.2807 Sports$$

คำสำคัญ : ตัวแปรตอบสนองแบบไม่มีลำดับ ตัวแปรตอบสนองแบบมีลำดับ ตัวแบบโลจิต และ ตัวแบบโลจิตสะสม

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

This research investigates factors affecting the response variable of students' goals, which are classified into the three categories: student can make good grades, student is popular, and student can be good in sports, with the proportions of these three categories corresponding to π_1 , π_2 , and π_3 respectively. The explanatory variables or factors consist of Gender, Age, Race, Urban, School, CGPA (Cumulative Grade Point Average), Sports, Looks and Money. The analyses of data using SAS version 9.1 and SPSS version 11.5, under both logit models for nominal responses and cumulative logit models for ordinal responses, are performed. The data are obtained from website <http://lib.stat.cmu.edu> on "Popular Kids data base", where a sample of 478 students in grades 4-6 was sampled from the three school districts in USA: Ingham, Clinton Counties, and Michigan. The results reveal that the explanatory variables affecting the students' goals, when responses categories are nominal, include Urban, CGPA and Sports. When the response categories are ordinal, we obtain Urban and CGPA. Moreover, the logit models for nominal response categories are more suitable than the cumulative logit models for ordinal response categories, due to each value of the likelihood ratio (G^2) and the Schwarz Criterion (SC) of the first case is less than that of the second case. Beside these, the residual plots of the first case lie within $(-2, 2)$ and also its scatter plots seem to be more spread out. Therefore, the appropriated logit models from the nominal response categories are concluded as follows:

$$\log(\hat{\pi}_1/\hat{\pi}_3) = -0.6283 + 0.6079\text{Urban} - 0.2514\text{CGPA} + 0.5741\text{Sports}$$

$$\log(\hat{\pi}_2/\hat{\pi}_3) = -2.5866 + 0.3466\text{Urban} + 0.2468\text{CGPA} + 0.8548\text{Sports}$$

$$\log(\hat{\pi}_1/\hat{\pi}_2) = 1.9583 + 0.2613\text{Urban} - 0.4982\text{CGPA} - 0.2807\text{Sports}$$

Keywords : Nominal response, Ordinal response, Logit model, Cumulative logit model, Modeling

ในปัจจุบันเราเริ่มมองเห็นปัญหาเกี่ยวกับ การเลือกโรงเรียนของนักเรียน ซึ่งนักเรียนส่วนใหญ่จะเลือกโรงเรียนที่นักเรียนคิดว่าตรงกับความต้องการของตนและตรงตามเป้าหมายที่กำหนดไว้ โดยทั่วไปนักเรียนอาจคิดว่า วิชาต่าง ๆ ที่เรียนไม่สามารถนำมาใช้ในชีวิตประจำวันได้ ดังนั้นนักเรียนจึงมองหาโรงเรียนที่เหมาะสมสำหรับเป้าหมายของตนเองที่วางไว้ เพราะฉะนั้นโรงเรียนจึงควรตระหนักถึงความสำคัญของปัญหาที่เกิดขึ้นและศึกษาความต้องการที่แท้จริงของนักเรียน รวมถึง เป้าหมายที่นักเรียนวางไว้เพื่อนำมาปรับปรุงและพัฒนาโรงเรียนให้ทันสมัย สามารถทำให้นักเรียนถึงเป้าหมายที่กำหนดไว้ได้ ในการวิจัยครั้งนี้ผู้วิจัยต้องการศึกษาปัจจัยที่ส่งผลกระทบต่อเป้าหมายต่าง ๆ ของนักเรียน (Students' Goals) และตรวจสอบหาตัวแบบที่เหมาะสมกับข้อมูลชุดนี้มากที่สุด ตลอดจนพยากรณ์ความน่าจะเป็นของเป้าหมายแต่ละกลุ่มด้วย

ตัวแบบเชิงสถิติเพื่อการวิเคราะห์ข้อมูลอาศัยตัวแบบ Multinomial logits (McFadden, 1974) ที่สามารถกำหนดกลุ่มฐานในการเปรียบเทียบหรือที่นิยมเรียกต่อมาว่าตัวแบบโลจิสติกแบบกลุ่มฐาน (Baseline-category logits models, Agresti, 2002) ซึ่งใช้สำหรับกรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ และตัวแบบโลจิสติกสะสมหรือตัวแบบ Cumulative logit models ที่มีลักษณะของรูปแบบ Proportional odds models (Walker and Duncan, 1967, McCullagh, 1980) ซึ่งใช้สำหรับกรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ การวางแผนการศึกษาด้านตัวแบบนั้น เนื่องจากลักษณะธรรมชาติของตัวแปรตอบสนองเปรียบเทียบเสมือนเป็นตัวแปรที่ตัดสินใจเป็นไปได้ทั้งแบบไม่มีลำดับหรือแบบมีลำดับ ดังนั้นจึงจำเป็นต้องตรวจสอบความเหมาะสมของตัวแบบของทั้งสองกรณีดังกล่าวแล้วเลือกตัวแบบที่เหมาะสมต่อไป นั่นคือ เมื่อพิจารณาตัวแปรตอบสนองเป็นแบบไม่มีลำดับ เลือกตรวจสอบด้วยตัวแบบ Multinomial logits ของกรณีตัวแบบ Baseline-category logits models และเมื่อพิจารณาตัวแปรตอบสนองเป็นแบบมีลำดับ เลือกตรวจสอบด้วยตัวแบบโลจิสติกสะสมของกรณีตัวแบบ Proportional odds model จากการสร้างตัวแบบจากทั้ง 2 กรณี เราสามารถคำนวณตัวสถิติภาวะสารถูปดี (Goodness-of-Fit Statistics) และตรวจสอบค่าของตัวสถิติต่าง ๆ เช่น Likelihood Ratio, Score และ Wald ตลอดจนการเปรียบเทียบค่าของตัวสถิติอัตราส่วนภาวะน่าจะเป็น (G^2 หรือ -2 Log L), SC (Schwarz

Criterion) การวิเคราะห์ส่วนเหลือ (Residual analysis) ของตัวแบบ เพื่อหาตัวแบบที่เหมาะสมกับข้อมูลชุดนี้มากที่สุด และศึกษาอิทธิพลที่กระทบต่อเป้าหมายของนักเรียนต่อไป

ดังนั้นงานวิจัยนี้จึงสนใจการนำตัวแบบโลจิสติกในกรณีตัวแบบ Baseline - category logit model เมื่อตัวแปรตอบสนองเป็นแบบไม่มีลำดับ และตัวแบบโลจิสติกสะสมในกรณีตัวแบบ Proportional odds model เมื่อตัวแปรตอบสนองเป็นแบบมีลำดับ มาวิเคราะห์เป้าหมายของนักเรียนด้วยข้อมูลจริง ซึ่งเป็นข้อมูลจริงจากเว็บไซต์ <http://lib.stat.cmu.edu> จากฐานข้อมูล Popular Kids ซึ่งเป็นข้อมูลตัวอย่างจากการสำรวจตัวแทนของประชากรนักเรียนชั้นประถมศึกษาปีที่ 4-6 ใน 3 รัฐของสหรัฐอเมริกา ได้แก่ Ingham, Clinton Counties และ Michigan จำนวน 478 คน โดยกำหนดขอบเขตของตัวแปรตอบสนอง คือ เป้าหมายของนักเรียน (Students' Goals) มี 3 กลุ่ม ดังนี้

- 1 = เมื่อเป้าหมายของนักเรียนคือ นักเรียนมีผลการเรียนดี
- 2 = เมื่อเป้าหมายของนักเรียนคือ นักเรียนเป็นที่ยอมรับของบุคคลทั่วไป
- 3 = เมื่อเป้าหมายของนักเรียนคือ นักเรียนมีความสามารถทางด้านกีฬา

ส่วนตัวแปรอธิบายประกอบด้วยตัวแปรที่อาจมีอิทธิพลต่อตัวแปรตอบสนองหรือเป้าหมายของนักเรียนรวมทั้งหมด 10 ตัวแปรได้แก่ เพศ (Gender) ระดับชั้นเรียน (Grade) อายุ (Age) เชื้อชาติ (Race) ความเป็นเมือง/ชนบท (Urban) โรงเรียน (School) ผลการเรียน (CGPA) กีฬา (Sports) รูปร่างหน้าตา (Looks) และจำนวนเงินที่นักเรียนได้รับ (Money) การประมวลผลและการวิเคราะห์ข้อมูลใช้โปรแกรมสำเร็จรูป SAS (Statistical Algorithm Systems) version 9.0 และ SPSS (Statistical Package for Social Sciences) version 11.5 ในการวิเคราะห์ทั้ง 2 กรณี คือ ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ และตัวแปรตอบสนองเป็นแบบมีลำดับ

ระเบียบวิธีการวิจัย

เพื่อให้งานวิจัยนี้เป็นไปตามวัตถุประสงค์และมีความชัดเจน ผู้วิจัยได้แบ่งวิธีการวิจัยเป็น 3 ตอน ดังนี้

ตอนที่ 1 การวิเคราะห์ข้อมูล กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ

1.1 กำหนดความสัมพันธ์ของตัวแปรตอบสนอง และตัวแปรอธิบายในรูปตัวแบบโลจิสต์ ในกรณีตัวแบบ baseline - category logit model เมื่อตัวแปรตอบสนองเป็นแบบไม่มีลำดับ

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \alpha_j + \beta_j'x, \quad j=1, \dots, J-1.$$

โดยที่ α_j, β_j แทน เวกเตอร์ของพารามิเตอร์ในตัวแบบ เมื่อ $j = 1, 2, \dots, J-1$ แทนกลุ่มของตัวแปรตอบสนองที่มี $J=3$ กลุ่ม

1.2 ทำการคัดเลือกตัวแปรอธิบายเข้าตัวแบบโลจิสต์ โดยใช้วิธีการเลือกตัวแปรแบบขั้นตอนโดยวิธีอัตราส่วนภาวะน่าจะเป็น (Stepwise (likelihood ratio : LR))

1.3 นำตัวแปรอธิบายที่ถูกคัดเลือกเข้าตัวแบบในข้อ 1.2 มาประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood or ML) และทดสอบนัยสำคัญทางสถิติของค่าพารามิเตอร์ด้วยตัวสถิติ Wald (Wald test)

1.4 ตรวจสอบความเหมาะสมของตัวแบบ (Model Fit statistics) โดยใช้ตัวสถิติ Deviance และ Pearson Goodness-of-Fit Statistics

1.5 นำตัวแบบโลจิสต์ (logit models) ที่ได้มาใช้ในการพยากรณ์ ความน่าจะเป็นที่นักเรียนจะอยู่กลุ่มใด

ตอนที่ 2 การวิเคราะห์ข้อมูล กรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ

2.1 กำหนดความสัมพันธ์ของตัวแปรตอบสนอง และตัวแปรอธิบายในรูปตัวแบบโลจิสต์สะสม (cumulative logit model) เมื่อตัวแปรตอบสนองเป็นแบบมีลำดับ 1, 2, 3 ตามเป้าหมายทั่วไปคือ ผลการเรียนรู้ การยอมรับ และกีฬาตามลำดับ

$$\text{logit} [P(Y \leq j | x)] = \alpha_j + \beta x, \quad j=1, \dots, J-1.$$

ตอนที่ 1 ผลการวิเคราะห์ข้อมูล กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ

ตารางที่ 1.1 จำนวนและร้อยละของนักเรียนจำแนกตามกลุ่มของตัวแปรตอบสนอง คือ เป้าหมายของนักเรียน

เป้าหมายของนักเรียน (Goals)	ค่าของตัวแปร	ความถี่ (Total Frequency)	เปอร์เซ็นต์ (Percent)
มีผลการเรียนดี เป็นที่ยอมรับของบุคคลทั่วไป มีความสามารถทางด้านกีฬา	1	247	51.67
	2	141	29.50
	3	90	18.83
รวม		478	100.00

2.2 ตรวจสอบข้อสมมติของตัวแบบ Proportional Odds (Proportional assumptions) ด้วยการทดสอบสกอร์ (Score test)

2.3 ทำการคัดเลือกตัวแปรอธิบายเข้าตัวแบบโลจิสต์สะสม โดยใช้วิธีการเลือกตัวแปรแบบขั้นตอนโดยวิธีอัตราส่วนภาวะน่าจะเป็น (Stepwise (likelihood ratio : LR))

2.4 ประมาณค่าพารามิเตอร์ในตัวแบบโลจิสต์สะสม (Cumulative logit model) ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood or ML) และทดสอบนัยสำคัญทางสถิติของค่าพารามิเตอร์ด้วยตัวสถิติ Wald (Wald test)

2.5 ทำการวิเคราะห์ความสัมพันธ์ของ ความน่าจะเป็นของการพยากรณ์กับค่าสังเกตของตัวแปรตอบสนอง โดยใช้ตัวสถิติ Somers'D, Gamma, Tau-a และ c และตรวจสอบความเหมาะสมของตัวแบบด้วยตัวสถิติ Deviance และ Pearson Goodness-of-Fit Statistics

2.6 นำตัวแบบโลจิสต์สะสม (Cumulative logit model) ที่ได้ มาใช้ในการพยากรณ์ความน่าจะเป็นที่นักเรียนจะอยู่กลุ่มใด

ตอนที่ 3 การเลือกตัวแบบที่เหมาะสมกับข้อมูล

เลือกตัวแบบที่เหมาะสมมากที่สุด ระหว่างตัวแบบในตอนที่ 1 และตัวแบบในตอนที่ 2 เมื่อตัวแปรตอบสนองเป็นแบบไม่มีลำดับ และเมื่อตัวแปรตอบสนองเป็นแบบมีลำดับ โดยเปรียบเทียบตัวสถิติต่างๆ เช่น อัตราส่วนภาวะน่าจะเป็น (G^2), SC (Schwarz Criterion) และการวิเคราะห์ส่วนเหลือ (Residual analysis) ของทั้งสองตัวแบบ ค่าสถิติที่มีค่าต่ำกว่าชี้ให้เห็นถึงความคลาดเคลื่อนของตัวแบบที่น้อยกว่า (วีรพันธ์ พงศาภักดี, 2544)

ผลการวิจัย

การวิเคราะห์ตัวแปรตอบสนอง (เป้าหมายของนักเรียน) ภายใต้วัตถุประสงค์เพื่อจะหาคำตอบเกี่ยวกับ ปัจจัยที่ส่งผลต่อเป้าหมายของนักเรียน และการเลือกตัวแบบที่เหมาะสมกับข้อมูล แบ่งการวิเคราะห์ข้อมูลออกเป็น 3 ตอน ดังนี้

จากตารางที่ 1.1 พบว่า นักเรียน 478 คน เป็นนักเรียน

ที่มีเป้าหมาย (Students' Goals) คือ มีผลการเรียนดีมากที่สุด จำนวน 247 คน คิดเป็นร้อยละ 51.67 รองลงมาเป็นนักเรียนที่มีเป้าหมาย คือ เป็นที่ยอมรับของบุคคลทั่วไป จำนวน 141 คน คิดเป็นร้อยละ 29.50 น้อยที่สุดเป็นนักเรียนที่มีเป้าหมาย คือ มีความสามารถทางด้านกีฬา จำนวน 90 คน คิดเป็นร้อยละ 18.83

ตารางที่ 1.2 ค่าสถิติ Wald และ P - value สำหรับผลกระทบของตัวแปรอธิบายที่มีผลต่อตัวแปรตอบสนอง

ปัจจัย (Effect)	องศาอิสระ (DF)	ค่าสถิติ Wald	P - value
Urban	2	15.1705	0.0005
Grades	2	23.1051	< .0001
Sports	2	27.6274	< .0001

จากตารางที่ 1.2 พบว่า ตัวแปรอธิบายที่ถูกคัดเลือกเข้า ภาวะน่าจะเป็น (Stepwise (Likelihood ratio: LR)) (Agresti, 2002) ได้แก่ Urban, CGPA และ Sports

ตารางที่ 1.3 ค่าประมาณสัมประสิทธิ์ของตัวแบบโลจิสต์ กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ

พารามิเตอร์	เป้าหมาย (Goals)	องศาอิสระ (DF)	β	$S.E.(\beta)$	ค่าสถิติ Wald	P - value
Intercept	1	1	-0.6283	0.5873	1.1443	0.2847
Intercept	2	1	-2.5866	0.6620	15.2689	< .0001
Urban	1	1	0.6079	0.1598	14.4714	0.0001
Urban	2	1	0.3466	0.1741	3.9654	0.0464
CGPA	1	1	-0.2514	0.1256	4.0088	0.0453
CGPA	2	1	0.2468	0.1373	3.2341	0.0721
Sports	1	1	0.5741	0.1530	14.0814	0.0002
Sports	2	1	0.8548	0.1629	27.5360	< .0001

จากตารางที่ 1.3 ค่าประมาณสัมประสิทธิ์ของตัวแบบโลจิสต์ที่ได้จากการประมาณค่าด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood or ML) สามารถเขียนสมการโลจิสต์ได้ดังนี้

$$\log(\hat{\pi}_1/\hat{\pi}_3) = -0.6283 + 0.6079Urban - 0.2514CGPA + 0.5741Sports \quad (1.1)$$

$$\log(\hat{\pi}_2/\hat{\pi}_3) = -2.5866 + 0.3466Urban + 0.2468CGPA + 0.8548Sports \quad (1.2)$$

$$\log(\hat{\pi}_1/\hat{\pi}_2) = 1.9583 + 0.2613Urban - 0.4982CGPA - 0.2807Sports \quad (1.3)$$

การประมาณค่าความน่าจะเป็นของตัวแปรตอบสนอง (Estimating Response Probabilities)

ค่าความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ นักเรียนมีผลการเรียนดี สามารถเขียนสมการได้ดังนี้

$$\hat{\pi}_1 = \frac{e^{-0.6283+0.6079Urban-0.2514CGPA+0.5741Sports}}{1 + e^{-0.6283+0.6079Urban-0.2514CGPA+0.5741Sports} + e^{-2.5866+0.3466Urban+0.2468CGPA+0.8548Sports}} \quad (1.4)$$

ค่าความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ นักเรียนเป็นที่ยอมรับของบุคคลทั่วไป สามารถเขียนสมการ ได้ดังนี้

$$\hat{\pi}_2 = \frac{e^{-2.5866+0.3466 \text{ Urban}+0.2468 \text{ CGPA}+0.8548 \text{ Sports}}}{1 + e^{-0.6283+0.6079 \text{ Urban}-0.2514 \text{ CGPA}+0.5741 \text{ Sports}} + e^{-2.5866+0.3466 \text{ Urban}+0.2468 \text{ CGPA}+0.8548 \text{ Sports}}} \quad (1.5)$$

และค่าความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมายคือ นักเรียนมีความสามารถทางด้านกีฬา สามารถเขียนสมการ ได้ดังนี้

$$\hat{\pi}_3 = \frac{1}{1 + e^{-0.6283+0.6079 \text{ Urban}-0.2514 \text{ CGPA}+0.5741 \text{ Sports}} + e^{-2.5866+0.3466 \text{ Urban}+0.2468 \text{ CGPA}+0.8548 \text{ Sports}}} \quad (1.6)$$

ตัวอย่างผลของการวิเคราะห์ข้อมูลและการตีความหมายของผลลัพธ์

ถ้าต้องการพยากรณ์ นักเรียนคนหนึ่งที่อาศัยอยู่ในเมือง และให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก และให้ความสำคัญกับกีฬา (Sports) เป็นอันดับที่สอง ว่าจะมีค่าความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ นักเรียนมีผลการเรียนดี สามารถทำได้ดังนี้

$$\hat{\pi}_1 = \frac{e^{-0.6283+0.6079(1)-0.2514(1)+0.5741(2)}}{1 + e^{-0.6283+0.6079(1)-0.2514(1)+0.5741(2)} + e^{-2.5866+0.3466(1)+0.2468(1)+0.8548(2)}} = 0.5781$$

นั่นคือ ถ้านักเรียนอาศัยอยู่ในเมือง ให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก และให้ความสำคัญกับกีฬา (Sports) เป็นอันดับที่สอง ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ มีผลการเรียนดี เท่ากับ 0.5781

ในทำนองเดียวกัน ถ้านักเรียนอาศัยอยู่ในเมือง ให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก และให้ความสำคัญกับกีฬา (Sports) เป็นอันดับที่สอง ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ เป็นที่ยอมรับของบุคคลทั่วไป สามารถทำได้ดังนี้

$$\hat{\pi}_2 = \frac{e^{-2.5866+0.3466(1)+0.2468(1)+0.8548(2)}}{1 + e^{-0.6283+0.6079(1)-0.2514(1)+0.5741(2)} + e^{-2.5866+0.3466(1)+0.2468(1)+0.8548(2)}} = 0.1812$$

นั่นคือ ถ้านักเรียนอาศัยอยู่ในเมือง ให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก และให้ความสำคัญกับกีฬา (Sports) เป็นอันดับที่สอง ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ เป็นที่ยอมรับของบุคคลทั่วไป เท่ากับ 0.1812

และถ้านักเรียนอาศัยอยู่ในเมือง ให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก และให้ความสำคัญกับกีฬา (Sports) เป็นอันดับที่สอง ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ มีความสามารถทางด้านกีฬา สามารถทำได้ดังนี้

$$\hat{\pi}_3 = \frac{1}{1 + e^{-0.6283+0.6079(1)-0.2514(1)+0.5741(2)} + e^{-2.5866+0.3466(1)+0.2468(1)+0.8548(2)}} = 0.2407$$

นั่นคือ ถ้านักเรียนอาศัยอยู่ในเมือง ให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก และให้ความสำคัญกับกีฬา (Sports) เป็นอันดับที่สอง ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ มีความสามารถทางด้านกีฬา เท่ากับ 0.2407

ตารางที่ 1.4 ค่า Deviance และ Pearson Goodness-of-Fit Statistics กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ

Criterion	Value	DF	Value / DF	P – value
Deviance	692.2033	664	1.0425	0.2173
Pearson	723.1856	664	1.0891	0.0552

จากตารางที่ 1.4 พบว่า ตัวสถิติ Deviance เท่ากับ 692.2033 และ ตัวสถิติ Pearson เท่ากับ 723.1856 ซึ่งให้ค่า P-value 0.2173 และ 0.0552 ตามลำดับ เนื่องจาก P-values

มีค่ามากกว่า 0.01 ทั้งสองค่า ดังนั้น ตัวแบบมีความเหมาะสมที่จะใช้ในการพยากรณ์ (Lawal, 2003). ภายใต้ระดับนัยสำคัญทางสถิติ 0.01

ตารางที่ 1.5 ค่าประมาณสำหรับ Odds Ratio กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ

Effect	เป้าหมาย (Goals)	Point Estimate	99% Wald Confidence Limits
Urban	1	1.837	(1.217 , 2.772)
Urban	2	1.414	(0.903 , 2.214)
CGPA	1	0.778	(0.563 , 1.075)
CGPA	2	1.280	(0.899 , 1.823)
Sports	1	1.776	(1.197 , 2.633)
Sports	2	2.351	(1.545 , 3.576)

จากตารางที่ 1.5 พบว่า นักเรียนที่อาศัยอยู่ในเมืองมีเป้าหมายคือ มีผลการเรียนดีมากว่า มีความสามารถทางด้านกีฬา มากกว่านักเรียนที่อาศัยอยู่ในชนบท เป็น 1.837 เท่า และช่วงความเชื่อมั่น 99% ของ Wald คือ (1.217 , 2.772) และนักเรียนที่อาศัยอยู่ในเมือง มีเป้าหมายคือ เป็นที่ยอมรับของบุคคลทั่วไป มากกว่ามีความสามารถทางด้านกีฬา มากกว่านักเรียนที่อาศัยอยู่ในชนบท เป็น 1.414 เท่า และช่วงความเชื่อมั่น 99% ของ Wald คือ (0.903 , 2.214)

นักเรียนที่ให้ความสำคัญกับการเรียนเป็นอันดับหนึ่ง มีเป้าหมายคือมีผลการเรียน มากกว่า มีความสามารถทางด้านกีฬาน้อยกว่านักเรียนที่ให้ความสำคัญกับการเรียนเป็นอันดับสุดท้าย เป็น 0.778 เท่า และช่วงความเชื่อมั่น 99% ของ Wald คือ (0.563 , 1.075) และนักเรียนที่ให้ความสำคัญกับการเรียนเป็น

อันดับหนึ่ง มีเป้าหมายคือ เป็นที่ยอมรับของบุคคลทั่วไป มากกว่านักเรียนที่ให้ความสำคัญกับการเรียนเป็นอันดับสุดท้าย เป็น 1.280 เท่า และช่วงความเชื่อมั่น 99% ของ Wald คือ (0.899 , 1.823)

นักเรียนที่ให้ความสำคัญกับกีฬาเป็นอันดับหนึ่ง มีเป้าหมายคือ มีผลการเรียนมากกว่า มีความสามารถทางด้านกีฬา มากกว่านักเรียนที่ให้ความสำคัญกับกีฬาเป็นอันดับสุดท้าย เป็น 1.776 เท่า และช่วงความเชื่อมั่น 99% ของ Wald คือ (1.197 , 2.633) ส่วนนักเรียนที่ให้ความสำคัญกับกีฬาเป็นอันดับหนึ่ง มีเป้าหมายคือ เป็นที่ยอมรับของบุคคลทั่วไป มากกว่า นักเรียนที่ให้ความสำคัญกับกีฬาเป็นอันดับสุดท้ายเป็น 2.351 เท่า และช่วงความเชื่อมั่น 99% ของ Wald คือ (1.545 , 3.576)

ตอนที่ 2 ผลการวิเคราะห์ข้อมูล กรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ

ตารางที่ 2.1 การทดสอบข้อสมมติของตัวแบบ Proportional Odds ด้วย Score Test

ค่าสถิติไคสแควร์	องศาอิสระ (DF)	P – value
6.4434	2	0.0399

จากตารางที่ 2.1 พบว่า ค่าสถิติไคสแควร์ เท่ากับ 6.4434 องศาอิสระเท่ากับ 2 ค่า P-value เท่ากับ 0.0399 เนื่องจาก ค่า P-value มากกว่า 0.01 ดังนั้น ข้อมูลชุดนี้เป็นไปตามข้อสมมติของ

ตัวแบบ Proportional Odds (Agresti, 2007) ที่ระดับนัยสำคัญ 0.01

ตารางที่ 2.2 ค่าสถิติ Score และ P - value สำหรับการคัดเลือกตัวแปรอธิบายเข้าตัวแบบโลจิสติกส์

ปัจจัย (Effect)	องศาอิสระ (DF)	ค่าสถิติ Score	P – value
CGPA	1	17.2207	< .0001
Urban	1	13.9818	0.0002

จากตารางที่ 2.2 พบว่า ตัวแปรอธิบายที่ถูกคัดเลือกเข้าตัวแบบโลจิสติกส์ ด้วยวิธีการเลือกตัวแปรแบบขั้นตอนโดยวิธี

อัตราส่วนภาวะน่าจะเป็น (Stepwise (Likelihood Ratio : LR)) (Agresti, 2002) ได้แก่ Urban และ CGPA

ตารางที่ 2.3 ค่าประมาณสัมประสิทธิ์ของตัวแบบโลจิตสะสม กรณีที่ตัวแปรตอบสนองเป็นแบบ มีลำดับ

พารามิเตอร์	องศาอิสระ (DF)	$\hat{\beta}$	$S.E.(\hat{\beta})$	ค่าสถิติ Wald	P - value
Intercept 1	1	0.1524	0.3242	0.2210	0.6383
Intercept 2	1	1.6186	0.3336	23.5472	< .0001
Urban	1	0.4019	0.1082	13.7890	0.0002
CGPA	1	-0.3433	0.0836	16.8424	< .0001

จากตารางที่ 2.3 พบว่า ค่าประมาณสัมประสิทธิ์โลจิตสะสม สำหรับตัวแปรความเป็นเมือง/ชนบท (Urban) และตัวแปรการเรียน (CGPA) เท่ากับ 0.4019 และ -0.3433 ค่าสถิติ Wald เท่ากับ 13.7890 และ 16.8424 ตามลำดับ เนื่องจาก ค่า P - value น้อยกว่า 0.01 แสดงว่า สัมประสิทธิ์โลจิตสะสมสำหรับ ตัวแปรความเป็นเมือง/ชนบท (Urban) และตัวแปรการเรียน (CGPA) มีนัยสำคัญที่ระดับ 0.01 ดังนั้น สามารถเขียนสมการโลจิตสะสมได้ 2 สมการ ได้ดังนี้

$$\logit[P(Y \leq 1 | X_1, X_2)] = 0.1524 + 0.4019 \text{Urban} - 0.3433 \text{CGPA} \quad (2.1)$$

$$\logit[P(Y \leq 2 | X_1, X_2)] = 1.6186 + 0.4019 \text{Urban} - 0.3433 \text{CGPA} \quad (2.2)$$

การประมาณค่าความน่าจะเป็นของตัวแปรตอบสนอง (Estimating Response Probabilities)

ค่าความน่าจะเป็นสะสม สามารถเขียนได้ดังสมการต่อไปนี้

$$P(Y \leq 1) = \frac{e^{0.1524 + 0.4019 \text{Urban} - 0.3433 \text{CGPA}}}{1 + e^{0.1524 + 0.4019 \text{Urban} - 0.3433 \text{CGPA}}}$$

$$P(Y \leq 2) = \frac{e^{1.6186 + 0.4019 \text{Urban} - 0.3433 \text{CGPA}}}{1 + e^{1.6186 + 0.4019 \text{Urban} - 0.3433 \text{CGPA}}}$$

ค่าความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ นักเรียนมีผลการเรียนดี สามารถเขียนสมการ ได้ดังนี้

$$P(Y=1) = P(Y \leq 1)$$

ค่าความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ นักเรียนเป็นที่ยอมรับของบุคคลทั่วไป สามารถเขียนสมการ ได้ดังนี้

$$P(Y=2) = P(Y \leq 2) - P(Y \leq 1)$$

และค่าความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมายคือ นักเรียนมีความสามารถทางด้านกีฬา สามารถเขียนสมการ ได้ดังนี้

$$P(Y=3) = 1 - P(Y \leq 2)$$

ตัวอย่างผลของการวิเคราะห์ข้อมูลและการตีความหมายของผลลัพธ์

ถ้าต้องการพยากรณ์ นักเรียนคนหนึ่งที่อาศัยอยู่ในเมือง และ ให้ความสำคัญกับการเรียน (Grades) เป็นอันดับแรกว่าจะมีค่าความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ นักเรียนมีผลการเรียนดี สามารถทำได้ดังนี้

$$\begin{aligned} P(Y=1) &= P(Y \leq 1) \\ &= \frac{e^{0.1524 + 0.4019(1) - 0.3433(1)}}{1 + e^{0.1524 + 0.4019(1) - 0.3433(1)}} = 0.5526 \end{aligned}$$

นั่นคือ ถ้านักเรียนอาศัยอยู่ในเมือง และให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ มีผลการเรียนดี เท่ากับ 0.5526

ถ้านักเรียนอาศัยอยู่ในเมือง และให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ เป็นที่ยอมรับของบุคคลทั่วไป สามารถทำได้ดังนี้

$$\begin{aligned} P(Y=2) &= P(Y \leq 2) - P(Y \leq 1) \\ &= \frac{e^{1.6186 + 0.4019(1) - 0.3433(1)}}{1 + e^{1.6186 + 0.4019(1) - 0.3433(1)}} - 0.5526 \\ &= 0.2899 \end{aligned}$$

นั่นคือ ถ้านักเรียนอาศัยอยู่ในเมือง และให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ เป็นที่ยอมรับของบุคคลทั่วไป เท่ากับ 0.2899

และถ้านักเรียนอาศัยอยู่ในเมือง และให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ มีความสามารถทางด้านกีฬาสามารถทำได้ดังนี้

$$P(Y=3) = 1 - P(Y \leq 2) = 1 - 0.8425 = 0.1575$$

นั่นคือ ถ้านักเรียนอาศัยอยู่ในเมือง และให้ความสำคัญกับการเรียน (CGPA) เป็นอันดับแรก ความน่าจะเป็นที่นักเรียนจะอยู่ในกลุ่มที่นักเรียนมีเป้าหมาย คือ มีความสามารถทางด้านกีฬา เท่ากับ 0.1575

ตารางที่ 2.4 ค่า Deviance และ Pearson Goodness-of-Fit Statistics กรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ

Criterion	Value	DF	Value / DF	P – value
Deviance	729.7318	668	1.0924	0.0487
Pearson	719.6322	668	1.0773	0.0814

จากตารางที่ 2.4 พบว่า ตัวสถิติ Deviance เท่ากับ 729.7318 และ ตัวสถิติ Pearson เท่ากับ 719.6322 ซึ่งให้ค่า P-values 0.0487 และ 0.0814 ตามลำดับ เนื่องจาก P-values มีค่ามากกว่า 0.01 ทั้งสองค่า ดังนั้น ตัวแบบมีความเหมาะสมที่จะใช้ในการพยากรณ์ (Lawal, 2003). ภายใต้ระดับนัยสำคัญทางสถิติ 0.01

ตารางที่ 2.5 ค่าประมาณสำหรับ Odds Ratio กรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ

Effect	Point Estimate	99% Wald Confidence Limits
Urban	1.495	(1.131 , 1.975)
CGPA	0.709	(0.572 , 0.880)

จากตารางที่ 2.5 พบว่า นักเรียนที่อาศัยอยู่ในเมือง มีเป้าหมายคือ มีผลการเรียนดี มากกว่า มีเป้าหมายคือ เป็นที่ยอมรับของบุคคลทั่วไปและมีความสามารถทางด้านกีฬา มากกว่านักเรียนที่อาศัยอยู่ชานเมืองและในชนบท เป็น 1.495 เท่า และช่วงความเชื่อมั่น 99% ของ Wald คือ (1.131 , 1.975) แต่นักเรียนที่ให้ความสำคัญกับการเรียนเป็นอันดับหนึ่ง

มีเป้าหมายคือ มีผลการเรียนดี มากกว่า มีเป้าหมายคือ เป็นที่ยอมรับของบุคคลทั่วไปและมีความสามารถทางด้านกีฬาน้อยกว่านักเรียนที่ให้ความสำคัญกับการเรียนเป็นอันดับรองลงมา เป็น 0.709 เท่า และช่วงความ-เชื่อมั่น 99% ของ Wald คือ (0.572 , 0.880)

ตารางที่ 2.6 ความน่าจะเป็นของการพยากรณ์กับค่าสังเกตและค่าความสัมพันธ์ กรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ

ตัวสถิติ	ค่าความสัมพันธ์
Percent Concordant	58.1
Percent Disconcordant	33.8
Somers'D	0.242
Gamma	0.264
Tau - a	0.148
c	0.621

จากตารางที่ 2.6 พบว่า จำนวนคู่ของค่าสังเกตที่สอดคล้องกัน (Percent Concordant) เท่ากับ 58.1 เปอร์เซ็นต์ และจำนวนคู่ของค่าสังเกตที่ไม่สอดคล้องกัน (Percent Disconcordant) เท่ากับ 33.8 เปอร์เซ็นต์ ซึ่งแสดงว่า จำนวนคู่ของค่าสังเกตที่สอดคล้องกัน มากกว่าจำนวนคู่ของค่าสังเกตที่ไม่สอดคล้องกัน

และจากตารางพบว่า ค่าสถิติ Somers'D เท่ากับ 0.242, Gamma เท่ากับ 0.264, Tau-a เท่ากับ 0.148 และ c เท่ากับ 0.621 จะเห็นว่า ค่าสถิติ Somers'D, Gamma, Tau - a และ c ให้ความสัมพันธ์เชิงบวก แสดงว่า ตัวแปรที่ได้มีความสัมพันธ์กับตัวแบบ

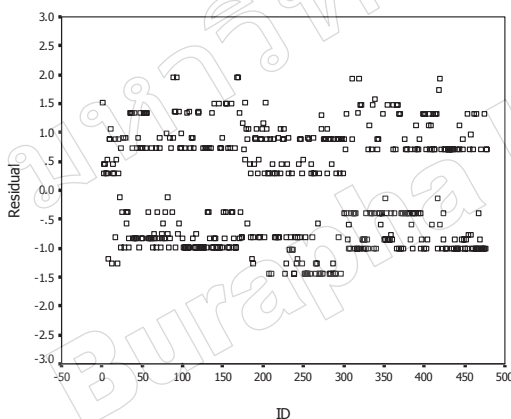
ตอนที่ 3 การเลือกตัวแบบที่เหมาะสมกับข้อมูล

ตารางที่ 3.1 การเปรียบเทียบความเหมาะสมของตัวแบบ ด้วยตัวสถิติ -2 Log L และ SC

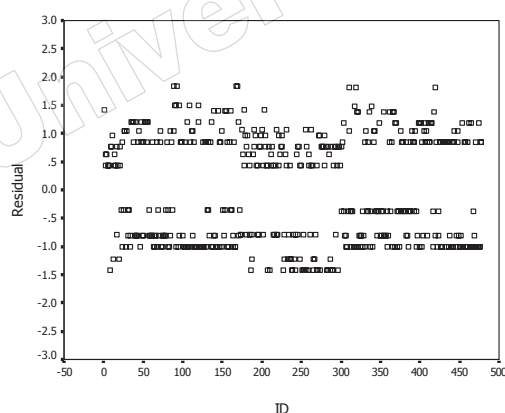
ตัวแบบ	-2 Log L	SC
ตัวแบบโลจิส	901.950	951.307
ตัวแบบโลจิสสะสม	939.478	964.157

จากตารางที่ 3.1 พบว่า ตัวแบบโลจิส ให้ค่าสถิติ -2 Log L เท่ากับ 901.950 ค่าสถิติ SC เท่ากับ 951.307 และตัวแบบโลจิสสะสม ให้ค่าสถิติ -2 Log L เท่ากับ 939.478 ค่าสถิติ SC เท่ากับ 964.157 จะเห็นว่า ตัวแบบโลจิสให้ค่าสถิติ -2 Log L และ ค่าสถิติ SC ต่ำกว่าตัวแบบโลจิสสะสม จึงน่าจะมีความคลาดเคลื่อนน้อยกว่า

การตรวจสอบส่วนเหลือ



ภาพที่ 1.1 กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ



ภาพที่ 1.2 กรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ

จากภาพที่ 1.1 และภาพที่ 1.2 พบว่า กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับมีการกระจายของส่วนเหลือ (Residual) ของแต่ละค่าสังเกตมีค่าอยู่ระหว่าง (-1.5, 2) และมีการกระจายตัวดีกว่า กรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ เนื่องจากตัวแบบ โลจิส ให้ค่าสถิติ -2 Log L เท่ากับ 901.950 ค่าสถิติ SC เท่ากับ 951.307 และตัวแบบโลจิสสะสม ให้ค่าสถิติ -2 Log L เท่ากับ 939.478 ค่าสถิติ SC เท่ากับ 964.157 จะเห็นว่า ตัวแบบโลจิสให้ค่าสถิติ -2 Log L และค่าสถิติ SC ต่ำกว่า

ตัวแบบโลจิสสะสม และการตรวจสอบส่วนเหลือ กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ ยังพบว่า การกระจายของส่วนเหลือ (Residual) ของแต่ละค่าสังเกต นอกจากมีการกระจายตัวดีกว่าแล้ว ค่าส่วนเหลือยังมีช่วงแคบ (-1.5, 2) กว่าช่วงส่วนเหลือของกรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ (-2, 2) ด้วย ดังนั้น สมการโลจิส กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ จึงมีความเหมาะสมกับข้อมูลชุดนี้มากที่สุด

สรุปผลการวิจัย

ตัวแบบในกรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ มีตัวแปรอธิบายที่ถูกคัดเลือกเข้าตัวแบบโลจิส คือ Urban, CGPA และ Sports และตัวแบบในกรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ มีตัวแปรอธิบายที่ถูกคัดเลือกเข้าตัวแบบโลจิสสะสม คือ Urban และ CGPA เนื่องจากค่าสถิติอัตราส่วนภาวะน่าจะเป็น (G^2) หรือ $-2 \log L$ และ SC (Schwarz Criterion) ของตัวแบบโลจิสที่มีค่าน้อยกว่าของตัวแบบโลจิสสะสม นอกจากนี้ การตรวจสอบส่วนเหลือ กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ พบว่า การกระจายของส่วนเหลือของแต่ละค่าสังเกต มีค่าอยู่ระหว่าง $(-1.5, 2)$ และมีการกระจายตัวดีกว่า กรณีที่ตัวแปรตอบสนองเป็นแบบมีลำดับ ซึ่งมีค่าอยู่ระหว่าง $(-2, 2)$ ดังนั้น สมการโลจิส กรณีที่ตัวแปรตอบสนองเป็นแบบไม่มีลำดับ ได้แก่

$$\log(\hat{\pi}_1/\hat{\pi}_3) = -0.6283 + 0.6079Urban - 0.2514CGPA + 0.5741Sports$$

$$\log(\hat{\pi}_2/\hat{\pi}_3) = -2.5866 + 0.3466Urban + 0.2468CGPA + 0.8548Sports$$

$$\log(\hat{\pi}_1/\hat{\pi}_2) = 1.9583 + 0.2613Urban - 0.4982CGPA - 0.2807Sports$$

จึงมีความเหมาะสมกับข้อมูลชุดนี้มากที่สุด และสามารถนำไปวิเคราะห์ข้อมูลและพยากรณ์ภาวะน่าจะเป็นของตัวแปรตอบสนอง หรือเป้าหมายของนักเรียนต่อไป

อภิปรายผลและข้อเสนอแนะการวิจัย

1. จากผลของการวิจัย พบว่าตัวแบบโลจิส เมื่อตัวแปรตอบสนองเป็นแบบไม่มีลำดับ มีความเหมาะสมมากกว่าตัวแบบโลจิสสะสม เมื่อตัวแปรตอบสนองเป็นแบบมีลำดับนั้นอาจเป็นเพราะตัวแปรตอบสนองไม่สามารถจัดลำดับได้ชัดเจน การจัดลำดับของตัวแปรตอบสนองจึงไม่มีผลลัพธ์ดีขึ้น

2. การสลับลำดับของตัวแปรตอบสนองหรือเป้าหมายของนักเรียน สำหรับตัวแบบโลจิส เมื่อตัวแปรตอบสนองเป็นแบบไม่มีลำดับหรือแบบมีลำดับ พบว่า การสลับลำดับตัวแปรตอบสนองหรือเป้าหมายของนักเรียนไม่ทำให้ผลลัพธ์ดีขึ้น ซึ่งสอดคล้องกับผลลัพธ์ทางทฤษฎี (Lawal, 2003, Agresti, 2007)

3. การวิจัยครั้งนี้เลือกใช้ตัวแบบโลจิสเมื่อตัวแปรตอบสนองแบบไม่มีลำดับ ในกรณีตัวแบบ Baseline - category logits model ส่วนการเลือกใช้ตัวแบบโลจิสสะสม เมื่อตัวแปรตอบสนองแบบมีลำดับ ยังมีตัวแบบอื่น ๆ ที่อาจสนใจศึกษาเพิ่มเติม เช่น ตัวแบบ Adjacent-categories logit model และตัวแบบ Generalized logit model with different slopes

4. อาจนำตัวแบบนาร่องที่เหมาะสมจากผลของการวิจัยนี้ไปศึกษาและประยุกต์สำหรับข้อมูลที่เป็นตัวแทนที่ดีสำหรับประชากรนักเรียนในประเทศไทยต่อไป โดยอาจเพิ่มตัวแปรที่ผู้เชี่ยวชาญเสนอแนะได้

เอกสารอ้างอิง

- วีรพันธ์ พงศาภักดี. (2544). การวิเคราะห์ข้อมูลเชิงกลุ่ม: ทฤษฎีและการประยุกต์ (กับ GLIM และ SPSS/FW). นครปฐม: โรงพิมพ์มหาวิทยาลัยศิลปากร.
- Agresti, A. (2002). *Categorical Data Analysis*. Second Edition, New York : John Willey & Sons, Inc.
- (2007). *An Introduction to Categorical Data Analysis*. Second Edition, New York : John Willey & Sons, Inc.
- Lawal, H. B. (2003). *Categorical Data Analysis with SAS and SPSS Applications*. London : Lawrence Erlbaum Associates. Inc.
- McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical society B* 42, 109-142.
- <http://lib.stat.cmu.edu>, December, 2008.
- McFadden, D.(1974). *Conditional Logit Analysis of Qualitative Choice Behavior*. Frontiers in Econometrics, ed. by P. Zarembka. Academic Press, New York, 105-142.
- Walker, S. H. and Duncan, D. B. (1967). Estimation of Probability of An Event as A Function of Several Independent Variables. *Biometrika*. 54, 167-179.