

ตัวแบบการถดถอยลอจิสติกในการพยากรณ์ความน่าจะเป็นของการชำระหนี้ได้ของครัวเรือน : กรณีศึกษา
จังหวัดปทุมธานี

Logistic Regression Model for Predicting the Probability of Household Debt Settlement:
A Case Study of Pathum Thani Province

ภาคสุภาวงศ์ มาปรีดา¹, แสงหล้า ชัยมงคล²

¹นักศึกษาระดับปริญญาโท สาขาสถิติประยุกต์ มหาวิทยาลัยธรรมศาสตร์

Email: paksupang_p5796@hotmail.com

²คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบการถดถอยลอจิสติกที่เหมาะสมในการพยากรณ์ความน่าจะเป็นของความสามารถในการชำระหนี้ได้ของครัวเรือนโดยพิจารณาจากตัวชี้วัดมาตรฐานครอบครัวเข้มแข็ง โดยนำข้อมูลมาตรฐานครอบครัวเข้มแข็งในระดับครอบครัว จังหวัดปทุมธานี ปี พ.ศ. 2559 ที่เก็บรวบรวมโดยสำนักงานกิจการสตรีและสถาบันครอบครัวมาสร้างตัวแบบการถดถอยลอจิสติกที่มีการแก้ปัญหาความสัมพันธ์เชิงเส้นระหว่างตัวชี้วัดย่อย โดยใช้วิธีการแลสโซ และวิธีการสร้างองค์ประกอบหลักของตัวชี้วัดย่อย นอกจากนี้ในงานวิจัยนี้ยังได้มีการปรับค่าเอนเอียงของค่าประมาณพารามิเตอร์ ในกรณีที่เกิดสัดส่วนของเหตุการณ์หนึ่งน้อยกว่าเมื่อเทียบกับการเกิดอีกเหตุการณ์หนึ่ง ผลการวิจัยพบว่าตัวแบบการถดถอยลอจิสติกแบบแลสโซเป็นตัวแบบที่เหมาะสมที่สุดในการพยากรณ์ความสามารถในการชำระหนี้ได้ของครัวเรือน

คำสำคัญ: การคัดเลือกตัวแบบ การถดถอยแลสโซ การวิเคราะห์องค์ประกอบหลัก

Abstract

The purpose of this research is to establish a suitable logistic regression model for predicting the probability of household debt repayment, based on robust family standard indicators. Strong family standard Pathumthani Province, collected in the year 2016 by the Office of Women's Affairs and Family Institute, was used to create a logistic regression model that solved the linear relationship problem between the sub-indicators. The ways to solve this problem were used the LASSO method and construct the principal components of the sub-indicators. In addition, the bias of the parameter estimation was adjusted in the case of rare events in data. The results show that the Lasso logistics regression model is the most suitable model for predicting the household's ability to pay debts.

Keywords: Model Selection Lasso Principle Component Analysis Rare Events

บทนำ

สถานการณ์การเปลี่ยนแปลงทางสังคมและเศรษฐกิจในช่วงที่ผ่านมาส่งผลกระทบต่อคนในสังคมเป็นอย่างมาก ประชาชนภายในประเทศต้องเผชิญกับปัญหาความยากจนและหนี้สิน จากการสำรวจสถานภาพหนี้ครัวเรือนไทยโดยพบว่าในปัจจุบันครัวเรือนไทยส่วนใหญ่ 91.1 เปอร์เซ็นต์มีหนี้สิน มหาวิทยาลัยหอการค้าไทยเผยหนี้ครัวเรือนของไทยปี พ.ศ. 2560 เพิ่มสูงสุดในรอบ 10 ปี (prachatai.com สืบค้นเมื่อ 21 ต.ค. 2560) จากงานวิจัยที่เกี่ยวข้องเกี่ยวกับความสามารถในการชำระหนี้ของครัวเรือนที่ผ่านมาส่วนมากจะเน้นศึกษาปัจจัยทางเศรษฐกิจและสังคม เช่นงานวิจัยของ โศภนิต หนูทิศ และประเสริฐ จรรยาสุภาพ (2554) ได้ทำการศึกษาปัจจัยที่มีผลต่อการค้างชำระหนี้ของลูกหนี้กองทุนหมุนเวียนเพื่อการเกษตร ผลที่ได้จากการศึกษาพบว่า ปัจจัยทางที่มีผลต่อการค้างชำระหนี้ ได้แก่ จำนวนแรงงานในครอบครัวและค่าใช้จ่ายภาคการเกษตร แต่งานวิจัยนี้ผู้วิจัยเน้น ศึกษาเกี่ยวกับปัจจัยตัวชี้วัดระดับความเข้มแข็งของครอบครัวที่มีผลต่อความสามารถในการชำระหนี้ได้ของครัวเรือนโดยใช้ครอบครัวในจังหวัดปทุมธานีเป็นกรณีศึกษา

ดัชนีชี้วัดมาตรฐานครอบครัวเข้มแข็งที่นำมาวิจัยครั้งนี้ได้มาจากการพัฒนาของ สำนักงานกิจการสตรีและสถาบันครอบครัวที่ร่วมมือกับสถาบันแห่งชาติเพื่อการพัฒนาเด็กและครอบครัว มหาวิทยาลัยมหิดลและสถาบันวิจัยและให้คำปรึกษาแห่งมหาวิทยาลัยธรรมศาสตร์ ได้พัฒนามาตรฐานครอบครัวเข้มแข็งระดับครอบครัวที่ประกอบด้วย 5 ด้าน 10 ตัวชี้วัด 47 ตัวชี้วัดย่อย ซึ่งตัวชี้วัดย่อยของตัวชี้วัดแต่ละตัวมีความสัมพันธ์ (ธนัญญา เกิดผล, 2559) และสอดคล้องกับแนวทางการพัฒนาของตัวชี้วัดในแต่ละด้านที่สามารถนำไปใช้วัดความเข้มแข็งของครอบครัวในระดับครอบครัว (วรวิมล โรมรัตน์พันธ์, 2556)

ในงานวิจัยนี้ต้องการหาตัวแบบที่เหมาะสมต่อการพยากรณ์ความสามารถในการชำระหนี้ได้ของครัวเรือนและแก้ปัญหาความสัมพันธ์ระหว่างตัวชี้วัดย่อย โดยได้นำข้อมูลดัชนีชี้วัดมาตรฐานครอบครัวเข้มแข็งที่เก็บข้อมูลเมื่อปี พ.ศ. 2559 ของจังหวัดปทุมธานีมาทำการวิเคราะห์ด้วยการใช้ตัวแบบการถดถอยลอจิสติก โดยกำหนดให้ตัวแปรตอบสนอง Y มีค่า 2 ค่าคือ $Y = 1$ เมื่อครอบครัวมีความสามารถในการชำระหนี้ได้ และ $Y = 0$ เมื่อครอบครัวไม่มีความสามารถในการชำระหนี้ และให้ตัวชี้วัดย่อยทั้ง 47 ตัวชี้วัดมาตรฐานครอบครัวเข้มแข็งเป็นตัวแปรอธิบาย แนวทางในการแก้ปัญหาความสัมพันธ์เชิงเส้นระหว่างตัวแปรอธิบายผู้วิจัยได้ศึกษา 2 วิธีคือใช้วิธีการประมาณค่าและคัดเลือกตัวแปรอธิบายตามวิธีการของ LASSO ที่ถูกเสนอโดย Tibshirani (1996) และอีกวิธีหนึ่ง คือใช้วิธีการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) โดยวิธีนี้จะทำการสร้างองค์ประกอบหลัก โดยทำการสร้างเซตของตัวแปรใหม่หรือองค์ประกอบหลักของแต่ละตัวชี้วัดในแต่ละด้านให้เป็นฟังก์ชันเชิงเส้นของตัวชี้วัดย่อยและนำองค์ประกอบหลักที่ได้เป็นตัวแปรอธิบายของการวิเคราะห์การถดถอยลอจิสติก

แต่เนื่องจากข้อมูลมาตรฐานครอบครัวเข้มแข็ง ปี พ.ศ. 2559 ที่ได้เก็บมาจากกลุ่มตัวอย่างพบว่าจำนวนครอบครัวที่ไม่สามารถชำระหนี้ได้มีจำนวนน้อยมากเมื่อเทียบกับจำนวนครอบครัวที่มีความสามารถในการชำระหนี้ได้ หรือเรียกเหตุการณ์นี้ว่า Rare Event ดังนั้นผู้วิจัยจึงได้ปรับสัมประสิทธิ์การถดถอยที่ได้จากการใช้องค์ประกอบหลักเป็นตัวแปรอธิบายตามวิธีการของ Rare Event

วัตถุประสงค์ของการวิจัย

เพื่อหาตัวแบบการถดถอยลอจิสติกที่เหมาะสมในการพยากรณ์ความน่าจะเป็นของความสามารถในการชำระหนี้ได้ของครัวเรือนจากตัวแปรดัชนีชี้วัดมาตรฐานครอบครัวเข้มแข็ง

วิธีการวิจัย

การดำเนินการวิจัยแบ่งออกเป็น 4 ขั้นตอน ดังนี้

1. การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยครั้งนี้ เป็นข้อมูลทุติยภูมิที่ได้มาจากสำนักงานกิจการสตรีและสถาบันครอบครัว ที่ได้ทำการเก็บรวบรวมข้อมูลครอบครัวเข้มแข็งของจังหวัดปทุมธานีในปี พ.ศ. 2559 โดยการสำรวจตัวอย่างครอบครัวจากครอบครัวไทยในจังหวัดปทุมธานีจำนวน 1,087 ครอบครัว

2. การจัดการข้อมูล

2.1 การกำหนดตัวแปรในการวิเคราะห์การถดถอยลอจิสติก

การกำหนดตัวแปรทำได้โดยกำหนดให้ตัวชี้วัดของความสามารถในการชำระหนี้สินของครัวเรือนเป็นตัวแปรตอบสนองหรือตัวแปร (Y) โดยแบ่งออกเป็น 2 ลักษณะคือ ครอบครัวที่สามารถชำระหนี้สินได้ ($Y = 1$) และครอบครัวที่ไม่สามารถชำระหนี้สินได้ ($Y = 0$) สำหรับตัวชี้วัดย่อยอื่น ๆ อีก 49 ตัว กำหนดให้เป็นตัวแปรอธิบายหรือตัวแปร X_{ijk} เมื่อ i แทนด้าน โดยที่ $i = 1, 2, 3, 4, 5$, j แทนตัวชี้วัดในแต่ละด้าน โดยที่ $j = 1, 2, \dots, r_i$ และ k แทนตัวชี้วัดย่อยในแต่ละตัวชี้วัด โดยที่ $k = 1, 2, \dots, r_{ij}$ โดยตัวชี้วัดย่อยที่มีลักษณะเป็นข้อมูลเชิงกลุ่ม โดยจะกำหนดให้มี 2 กลุ่ม คือ ใช่ และไม่ใช่ และตัวชี้วัดย่อยที่แสดงระดับความคิดเห็นให้มีการวัดที่ให้เป็นระดับคะแนน 4 ระดับตามระดับความคิดเห็น คือ 1, 2, 3 และ 4

2.2 ตรวจสอบความสมบูรณ์ของข้อมูลที่ใช้ในการวิเคราะห์

ข้อมูลที่จะนำมาใช้ในการวิเคราะห์จะต้องมีข้อมูลครบทุกตัวแปร ดังนั้นครอบครัวที่มีข้อมูลในแต่ละตัวแปรไม่สมบูรณ์จะไม่ถูกนำมาวิเคราะห์ หลังจากการจัดการข้อมูลจะมีข้อมูลของจำนวนครอบครัวเหลือทั้งหมด 919 ครอบครัวที่จะนำมาใช้ในการวิเคราะห์ข้อมูล โดยเป็นครอบครัวที่มีความสามารถในการชำระหนี้ได้จำนวน 917 ครอบครัว และมีเพียง 12 ครอบครัวที่ไม่สามารถชำระหนี้ได้

3. การวิเคราะห์ข้อมูล

3.1 สร้างตัวแบบการถดถอยลอจิสติก โดยใช้วิธี LASSO

LASSO เป็นวิธีที่นิยมใช้ในการวิเคราะห์การถดถอยกรณีที่มีตัวแปรอธิบายมีความสัมพันธ์เชิง

เส้นกัน ซึ่งวิธี LASSO มี Penalty Function คือ $P_\lambda(\beta) = \lambda \sum_{i=1}^p |\beta_i|$ โดยที่ β และ λ คือพารามิเตอร์ปรับค่า ในการหาค่าพารามิเตอร์ปรับค่าทั่วไปนิยมใช้วิธี Cross-Validation ในการหาค่าพารามิเตอร์ปรับค่าที่เหมาะสม ในที่นี้การคัดเลือกตัวแปรแบบ LASSO จะใช้โปรแกรม RStudio โดยใช้ชุดคำสั่ง “glmnet” และเลือกค่าพารามิเตอร์ปรับค่า ($\lambda = 0.0008242$) ซึ่งเป็นค่าที่ทำให้อัตราพยากรณ์ถูกต้องสูงสุด

3.2 สร้างองค์ประกอบหลักของตัวชี้วัดย่อยในแต่ละตัวชี้วัดโดยใช้วิธีการ PCA

PCA เป็นการสร้างเซตของตัวแปรใหม่หรือองค์ประกอบหลักเป็นฟังก์ชันเชิงเส้นของตัวแปรเดิม โดยที่องค์ประกอบเหล่านี้ไม่มีความสัมพันธ์กันและเรียงลำดับตามขนาดของความแปรปรวน ในการวิเคราะห์องค์ประกอบหลักมีแนวคิดที่ว่าความแปรผันทั้งหมดที่เกิดขึ้นของค่าสังเกตใดๆจากเดิมที่มี p ตัวนั้นจะสามารถอธิบายได้ด้วยองค์ประกอบหลักเพียง 1 หรือ 2 องค์ประกอบหลัก

3.3 สร้างตัวแบบการถดถอยลอจิสติก โดยใช้คะแนนองค์ประกอบหลักเป็นตัวแปรอธิบาย สำหรับการวิเคราะห์การถดถอยลอจิสติกเป็นสมการที่แสดงความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

$$P(Y = 1 | x) = \pi_i = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

ในการวิเคราะห์การถดถอยลอจิสติก ความสัมพันธ์ระหว่างตัวแปรอธิบายกับตัวแปรตอบสนองไม่เป็นความสัมพันธ์เชิงเส้น ซึ่งในการวิเคราะห์จะต้องมีการปรับให้ความสัมพันธ์อยู่ในรูปเชิงเส้น ในรูปของ odds และในการเขียนตัวแบบลอจิสติกจะเขียนให้อยู่ในรูป log ของ odds เรียกว่า logit ดังต่อไปนี้

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

การประมาณค่าพารามิเตอร์ใช้วิธีการภาวะน่าจะเป็นสูงสุด (Maximum likelihood) โดยที่พยายามหาค่า $\hat{\beta}$ ที่ทำให้

$$\begin{aligned} \ln L(\beta | y) &= \sum_{\{Y=1\}_i} \ln(\pi_i) + \sum_{\{Y=0\}_i} \ln(1 - \pi_i) \\ &= -\sum_{i=1}^n \ln(1 + e^{(1-2Y)_i \beta_i}) \end{aligned}$$

มีค่าสูงสุดและความแปรปรวนของตัวประมาณค่าพารามิเตอร์คือ

$$V(\hat{\beta}) = \left[\sum_{i=1}^n \pi_i (1 - \pi_i) x_i' x_i \right]^{-1}$$

โดยตัวประมาณที่ได้ เป็นตัวประมาณที่ไม่เอนเอียง (unbiased) แต่ในกรณีของ rare event $P(Y = 1 | x) = \pi_i$ มีค่าน้อยมากๆ ทำให้ $\hat{\beta}$ ขาดคุณสมบัติที่ดีของตัวประมาณค่า จึงต้องทำการปรับค่าที่อาศัยความรู้ล่วงหน้าเกี่ยวกับสัดส่วนของ $Y=1$ ในประชากร (τ) และสัดส่วนของ $Y=1$ ในตัวอย่าง (\bar{y}) วิธีการปรับอาจทำได้โดยการปรับค่า $\hat{\beta}_0$ หรือใช้วิธีการถ่วงน้ำหนักฟังก์ชันภาวะน่าจะเป็น ซึ่งวิธีการเหล่านี้มีให้เลือกใช้ในชุดคำสั่งของ “relogit” ของโปรแกรม R (Kosuke and Gary, 2007) โดยในงานวิจัยนี้เลือกใช้วิธีการปรับค่าเอนเอียงของค่าประมาณสัมประสิทธิ์การถดถอย

3.4 เปรียบเทียบความสามารถในการพยากรณ์ได้ถูกต้องของตัวแบบ

การเลือกตัวแบบการถดถอยลอจิสติกที่เหมาะสมในการพยากรณ์ความสามารถในการชำระหนี้ได้ของครัวเรือนจะพิจารณาจากความสามารถในการพยากรณ์ได้ถูกต้องของตัวแบบแต่ละแบบโดยเฉพาะการพยากรณ์ถูกต้องในกลุ่มที่ครอบครัวไม่มีความสามารถในการชำระหนี้ได้

4. ผลการวิเคราะห์ข้อมูล

ตารางที่ 1: สัมประสิทธิ์น้ำหนัก (loading) และความหมายขององค์ประกอบหลักของตัวชี้วัดแต่ละตัว

ด้าน	ตัวชี้วัด	ดัชนีตัวชี้วัดย่อย	องค์ประกอบหลัก	
1	ตัวชี้วัดที่ 1	สมาชิกในครอบครัวมีการแสดงออกที่สื่อถึงความรัก (x_{111})	$PC_{111} = 0.575 x_{111} + 0.576 x_{112} + 0.580 x_{113}$ โดยที่ PC_{111} หมายถึงสมาชิกในครอบครัวแสดงออกถึงความรักและความเอาใจใส่ระหว่างกัน	
		สมาชิกในครอบครัวมีการช่วยเหลือแบ่งเบาภาระงาน (x_{112})		
		สมาชิกในครอบครัวมีการทำกิจกรรมประจำวันร่วมกัน (x_{113})		
	ตัวชี้วัดที่ 2	สมาชิกในครอบครัวมีการพูดคุยหรือรับฟังกันและกัน (x_{121})	$PC_{121} = 0.577 x_{121} + 0.577 x_{122} + 0.576 x_{123}$ โดยที่ PC_{121} หมายถึงสมาชิกในครอบครัวยอมรับและเคารพความคิดเห็นซึ่งกันและกัน	
		สมาชิกในครอบครัวมีการยอมรับความเห็นที่แตกต่างกัน (x_{122})		
		สมาชิกในครอบครัวเคารพความเป็นส่วนตัวของแต่ละคน (x_{123})		
	ตัวชี้วัดที่ 3	สมาชิกในครอบครัวมีการพูดจาต่อกันด้วยดีหรือใช้เหตุผล (x_{131})	$PC_{131} = 0.410 x_{131} + 0.413 x_{132} + 0.413 x_{133} + 0.409 x_{134} + 0.396 x_{135} + 0.029 x_{136} + 0.407 x_{137} - 0.013 x_{138}$ โดยที่ PC_{131} หมายถึงสมาชิกในครอบครัวสื่อสารกันอย่างมีคุณภาพ	
		สมาชิกในครอบครัวมีโอกาสท้วงติง แสดงความคิดเห็น (x_{132})		
		สมาชิกในครอบครัวมีการแสดงความชื่นชมเห็นคุณค่าซึ่งกัน (x_{133})		
		สมาชิกในครอบครัวต่างดูแลเอาใจใส่ซึ่งกันและกัน (x_{134})		
		เมื่อประสบปัญหาสมาชิกในครอบครัวสามารถหาทางออก (x_{135})		
		สมาชิกในครอบครัวไม่ทำร้ายจิตใจเมื่อมีความขัดแย้ง (x_{136})		
		สมาชิกในครอบครัวไว้วางใจที่จะเล่าเรื่องต่าง ๆ ที่เกิดขึ้น (x_{137})		
		สมาชิกในครอบครัวไม่ใช้อำนาจเหนือกันและกัน (x_{138})		
	ตัวชี้วัดที่ 4	สมาชิกในครอบครัวร่วมตัดสินใจเรื่องสำคัญด้วยกัน (x_{141})	$PC_{141} = -0.706 x_{141} - 0.706 x_{142} + 0.034 x_{143}$ โดยที่ PC_{141} หมายถึงสมาชิกในครอบครัวมีส่วนร่วมแก้ปัญหาข้อขัดแย้งด้วยเหตุผล	
		สมาชิกในครอบครัวร่วมกันแก้ปัญหาด้วยเหตุผล (x_{142})		
		สมาชิกในครอบครัวไม่ทำร้ายร่างกายเมื่อมีความขัดแย้ง (x_{143})		
				$PC_{142} = -0.020 x_{141} - 0.027 x_{142} + 0.999 x_{143}$ และ PC_{142} หมายถึงสมาชิกในครอบครัวไม่ทำร้าย

ด้าน	ตัวชี้วัด	ดัชนีตัวชี้วัดย่อย	องค์ประกอบหลัก
		ขัดแย้ง (x_{143})	ร่างกายเมื่อมีความขัดแย้ง
2	ตัวชี้วัดที่ 5	สมาชิกในครอบครัวทำหน้าที่ของตนอย่างเหมาะสม (x_{251})	$PC_{251} = 0.500x_{251} + 0.500x_{252} + 0.500x_{253} + 0.499x_{254}$ โดยที่ PC_{251} หมายถึงครอบครัวทำบทบาทหน้าที่ของครอบครัวต่อสมาชิก
		ครอบครัวมีการดูแล ความปลอดภัยของสมาชิกทุกคน (x_{252})	
		ครอบครัวมีการใช้คุณธรรมสำคัญในการดำเนินชีวิต (x_{253})	
		ครอบครัวปฏิบัติตนเป็นแบบอย่างต่อสมาชิกในครอบครัว (x_{254})	
3	ตัวชี้วัดที่ 6	ครอบครัวมีสมาชิกที่สามารถหารายได้มาเลี้ยงดูตนเอง (x_{361})	$PC_{361} = 0.446x_{361} - 0.508x_{362} + 0.677x_{363} + 0.061x_{364}$ โดยที่ PC_{361} หมายถึงครอบครัวสามารถพึ่งตนเองด้านเศรษฐกิจ
		ครอบครัวมีการทำบัญชีรายรับ-รายจ่าย (x_{362})	
		ครอบครัวมีการออมในรูปแบบต่าง ๆ (x_{363})	
		ครอบครัวมีที่อยู่อาศัยเป็นของตนเอง (x_{364})	
	ตัวชี้วัดที่ 7	สมาชิกในครอบครัวมีสิทธิได้รับหลักประกันสุขภาพ (x_{371})	$PC_{371} = 0.003x_{371} + 0.707x_{374} + 0.707x_{376}$ $PC_{372} = 0.999x_{371} + 0.001x_{374} + 0.003x_{376}$ โดยที่ PC_{371} หมายถึงครอบครัวสามารถเลี่ยงพฤติกรรมเสี่ยงและดูแลสมาชิกที่มีภาวะพึ่งพิงกันได้ และ PC_{372} หมายถึงครอบครัวมีสิทธิได้รับหลักประกันสุขภาพ
		ครอบครัวสามารถเลี่ยงพฤติกรรมเสี่ยงของสมาชิกภายในครอบครัว (x_{374})	
		ครอบครัวสามารถดูแลสมาชิกที่มีภาวะพึ่งพิงสูง (x_{376})	

ด้าน	ตัวชี้วัด	ดัชนีตัวชี้วัดย่อย	องค์ประกอบหลัก
	ตัวชี้วัดที่ 8	สมาชิกในครอบครัวมีการแลกเปลี่ยนข้อมูลข่าวสาร (x_{381})	$PC_{381} = 0.529x_{381} + 0.529x_{382} + 0.505x_{383} + 0.426x_{384}$ โดยที่ PC_{381} หมายถึงสมาชิกในครอบครัวสามารถ พึ่งตนเองด้านข้อมูลข่าวสารและการเรียนรู้
		สมาชิกในครอบครัวมีการนำข้อมูลข่าวสารไปใช้ ประโยชน์ในชีวิตประจำวัน (x_{382})	
		สมาชิกในครอบครัวได้รับรู้เรื่องเกี่ยวกับครอบครัว (x_{383})	
		ระดับการศึกษาในระบบของสมาชิกในครอบครัว (x_{384})	
4	ตัวชี้วัดที่ 9	สมาชิกในครอบครัวมีความสัมพันธ์ที่ดี (x_{491})	$PC_{491} = 0.274x_{491} + 0.375x_{492} + 0.333x_{493}$ $+ 0.519x_{494} + 0.543x_{495} + 0.327x_{496}$ $PC_{492} = 0.682x_{491} + 0.549x_{492} - 0.261x_{493}$ $- 0.291x_{494} + 0.139x_{495} - 0.242x_{496}$ โดยที่ PC_{491} หมายถึงครอบครัวมีทุนทางสังคมใน การสร้างบรรยากาศความเป็นครอบครัว และ PC_{492} หมายถึงสมาชิกในครอบครัวมี ความสัมพันธ์ที่ดีและมีการช่วยเหลือเกื้อกูลกัน ระหว่างเครือญาติ
		สมาชิกในครอบครัวมีการเกื้อกูลกันระหว่างเครือ ญาติ (x_{492})	
		สมาชิกในครอบครัวมีความรู้สึกปลอดภัยและ ไว้วางใจกัน (x_{493})	
		สมาชิกในครอบครัวเป็นสมาชิกกลุ่มทางสังคม (x_{494})	
		สมาชิกในครอบครัวมีส่วนร่วมกิจกรรมพัฒนา คุณภาพชีวิต (x_{495})	
		ครอบครัวมีการอบรมให้สมาชิกในครอบครัวใช้ บริการพื้นที่สาธารณะ (x_{496})	
5	ตัวชี้วัดที่ 10	ครอบครัวมีการเตรียมพร้อมการรองรับความ ยากลำบาก ($x_{5,10,1}$)	$PC_{5,10,1} = 0.569x_{5,10,1} + 0.502x_{5,10,2} + 0.393x_{5,10,3}$ $+ 0.512x_{5,10,4} + 0.072x_{5,10,5}$ $PC_{5,10,2} = 0.112x_{5,10,1} + 0.035x_{5,10,2} + 0.142x_{5,10,3}$ $- 0.089x_{5,10,4} + 0.978x_{5,10,5}$ โดยที่ $PC_{5,10,1}$ หมายถึงครอบครัวมีการหลีกเลี่ยง ภาวะเสี่ยงและการปรับตัวได้ในภาวะยากลำบาก และ $PC_{5,10,2}$ หมายถึงสมาชิกในครอบครัวสามารถ เข้าร่วมกิจกรรมเพื่อเตรียมความพร้อมรับมือกับปัญหาภัย พิบัติ
		ครอบครัวร่วมมือกันแก้ไขปัญหาเมื่อเกิดภาวะวิกฤติ ($x_{5,10,2}$)	
		ครอบครัวสามารถปรับตัวเมื่อเกิดภาวะวิกฤติ ($x_{5,10,3}$)	
		ครอบครัวเข้าร่วมกิจกรรมในการป้องกันปัญหา สังคม ($x_{5,10,4}$)	
		ครอบครัวเข้าร่วมกิจกรรมเพื่อเตรียมพร้อมการ ป้องกันปัญหาภัยพิบัติ ($x_{5,10,5}$)	

หมายเหตุ กำหนดให้ตัวประกอบหลัก PC_{ijk} เมื่อ i แทนด้าน โดยที่ $i = 1,2,3,4,5$, j แทนตัวชี้วัด
ในแต่ละด้าน โดยที่ $j = 1,2,\dots,10$ และ k แทนตัวประกอบหลัก โดยที่ $k = 1,2$

จากการสร้างองค์ประกอบหลักของตัวชี้วัดแต่ละตัวในแต่ละด้านจะเห็นว่าความแปรผันทั้งหมดของ
ตัวชี้วัดย่อยของตัวชี้วัดที่ 1, 2, 5, 6, และ 8 จะสามารถอธิบายได้ด้วยองค์ประกอบหลักเพียง 1 องค์ประกอบ
โดยจะแสดงเป็นภาพรวมของคุณลักษณะของตัวชี้วัดย่อยของตัวชี้วัดนั้นๆ ส่วนตัวชี้วัดที่เหลือ (ตัวชี้วัดที่ 3, 4,

7, 9, และ 10) สามารถทำการจัดหมวดหมู่ตัวชี้วัดย่อยออกเป็น 2 กลุ่มอย่างชัดเจน โดยดูได้จากสัมประสิทธิ์น้ำหนักของตัวชี้วัดย่อยที่มีค่ามากในองค์ประกอบหลักใดองค์ประกอบเดียวเท่านั้น และความหมายขององค์ประกอบหลักแต่ละองค์ประกอบได้แสดงไว้ในตารางที่ 1

ตารางที่ 2: ค่าประมาณ ความคลาดเคลื่อนมาตรฐานและ ค่าอัตราส่วนออกดส์ ของพารามิเตอร์ที่มีนัยสำคัญทางสถิติของตัวแบบการถดถอยลอจิสติกแบบLASSO

ตัวชี้วัด	ตัวชี้วัดย่อย	$\hat{\beta}$	S.E.	$Exp(\hat{\beta})$
ตัวชี้วัดที่ 6	ครอบครัวมีการออมในรูปแบบต่าง ๆ (x_{363})	2.795	0.765	16.365
ตัวชี้วัดที่ 9	ครอบครัวมีการช่วยเหลือเกื้อกูลกันระหว่างเครือญาติ (x_{492})	2.952	1.379	19.145
ตัวชี้วัดที่ 9	ครอบครัวมีการอบรมให้สมาชิกในครอบครัวใช้พื้นที่สาธารณะ (x_{496})	2.259	1.364	9.569
ตัวชี้วัดที่ 10	ครอบครัวร่วมมือกันแก้ไขปัญหาเมื่อเกิดภาวะวิกฤต ($x_{5,10,2}$)	2.978	1.121	19.64
	Constant	-5.762	1.983	

ตัวแบบพยากรณ์ความน่าจะเป็นที่ครอบครัวมีความสามารถชำระหนี้ได้ที่ได้จากการวิเคราะห์การถดถอยลอจิสติกแบบ LASSO สามารถเขียนได้ดังนี้

$$P(Y_i = 1 | x_i) = \hat{\pi}(x_i) = \frac{\exp(-5.762 + 2.795x_{363} + 2.952x_{492} + 2.259x_{496} + 2.978x_{5,10,2})}{1 + \exp(-5.762 + 2.795x_{363} + 2.952x_{492} + 2.259x_{496} + 2.978x_{5,10,2})}$$

หรือเขียนตัวแบบการถดถอยลอจิสติกแบบ LASSO ในรูปสมการเชิงเส้นได้ดังนี้

$$\text{logit}(\hat{\pi}(x)) = \log\left(\frac{\hat{\pi}(x_i)}{1 - \hat{\pi}(x_i)}\right) = -5.762 + 2.795x_{363} + 2.952x_{492} + 2.259x_{496} + 2.978x_{5,10,2}$$

จากตัวแบบที่ได้จะเห็นว่าดัชนีชี้วัดความเข้มแข็งของครอบครัวที่มีผลต่อการพยากรณ์ความน่าจะเป็นของครอบครัวที่สามารถชำระหนี้ได้ มี 3 ด้าน ได้แก่ ด้านการพึ่งตนเอง ด้านทุนทางสังคม และด้านการหลีกเลี่ยงความเสี่ยงและการปรับตัวได้ในภาวะยากลำบาก โดยมี 3 ตัวชี้วัด ได้แก่ตัวชี้วัดที่ 6 ตัวชี้วัดที่ 9 และตัวชี้วัดที่ 10 โดยที่ตัวชี้วัดที่ 9 มีตัวชี้วัดย่อย 2 ตัวที่มีนัยสำคัญทางสถิติ ซึ่งตัวชี้วัดย่อยเหล่านี้มีอิทธิพลทางบวกกับความน่าจะเป็นที่ครอบครัวจะสามารถชำระหนี้ได้ โดยค่าที่เพิ่มขึ้น 1 คะแนนจะมีผลทำให้โอกาสที่

ครอบครัวจะสามารถชำระหนี้ได้เป็น 16, 19, 9.5, และ 19.6 เท่าของโอกาสที่ครอบครัวจะไม่สามารถชำระหนี้ได้ สำหรับตัวชี้วัดย่อย (x_{363} , x_{492} , x_{496} และ $x_{5,10,2}$) ตามลำดับ

ตารางที่ 3: ค่าประมาณ ความคลาดเคลื่อนมาตรฐานและ ค่าอัตราส่วนออกดส์ ของพารามิเตอร์ที่มีนัยสำคัญทางสถิติของตัวแบบการถดถอยลอจิสติกที่ใช้อังก์ประกอบหลักเป็นตัวแปรอธิบาย

ตัวชี้วัด	องค์ประกอบหลัก	$\hat{\beta}$	S.E.	$Exp(\hat{\beta})$
ตัวชี้วัดที่ 1	สมาชิกในครอบครัวแสดงออกถึงความรักและความเอาใจใส่ระหว่างกัน (PC_{111})	1.763	0.560	5.832
ตัวชี้วัดที่ 2	สมาชิกในครอบครัวยอมรับและเคารพความคิดเห็นซึ่งกันและกัน (PC_{121})	1.870	0.538	6.489
ตัวชี้วัดที่ 3	สมาชิกในครอบครัวสื่อสารกันอย่างมีคุณภาพ (PC_{132})	0.158	0.082	1.171
ตัวชี้วัดที่ 6	การพึ่งตนเองด้านเศรษฐกิจ (PC_{361})	0.599	0.163	1.802
ตัวชี้วัดที่ 10	การหลีกเลี่ยงภาวะเสี่ยงและการปรับตัวได้ในภาวะยากลำบาก ($PC_{5,10,2}$)	0.620	0.191	1.859
	Constant	5.423	0.506	

ตัวแบบพยากรณ์ความน่าจะเป็นที่ครอบครัวมีความสามารถชำระหนี้ได้ ที่ได้จากการวิเคราะห์การถดถอยลอจิสติกที่ใช้อังก์ประกอบหลักของตัวชี้วัดเป็นตัวแปรอธิบาย สามารถเขียนได้ ดังต่อไปนี้

$$\hat{\pi}(PC_i) = \frac{\exp(5.423 + 1.763PC_{111} + 1.870PC_{121} + 0.158PC_{132} + 0.599PC_{361} + 0.620PC_{5,10,2})}{1 + \exp(5.423 + 1.763PC_{111} + 1.870PC_{121} + 0.158PC_{132} + 0.599PC_{361} + 0.620PC_{5,10,2})}$$

และสามารถเขียนตัวแบบในรูปตัวแบบเชิงเส้น ได้ดังนี้

$$\text{logit}(\hat{\pi}(x)) = \log\left(\frac{\hat{\pi}(x_i)}{1 - \hat{\pi}(x_i)}\right) = 5.423 + 1.763PC_{111} + 1.870PC_{121} + 0.158PC_{132} + 0.599PC_{361} + 0.620PC_{5,10,2}$$

จากตัวแบบที่ได้จะเห็นว่าดัชนีชี้วัดความเข้มแข็งของครอบครัวที่มีผลต่อการพยากรณ์ความน่าจะเป็นของครอบครัวที่สามารถชำระหนี้ได้ มี 3 ด้าน (ด้านสัมพันธภาพ ด้านการพึ่งตนเอง และด้านการหลีกเลี่ยงภาวะเสี่ยงและการปรับตัวได้ในภาวะยากลำบาก) ประกอบด้วย 5 ตัวชี้วัด โดยที่ด้านที่ 1 (ด้านสัมพันธภาพ) มี 3 ตัวชี้วัด ซึ่งองค์ประกอบหลักของตัวชี้วัดเหล่านี้มีอิทธิพลทางบวกกับความน่าจะเป็นที่ครอบครัวจะสามารถชำระหนี้ได้ โดยค่าคะแนนองค์ประกอบหลักที่เพิ่มขึ้น 1 คะแนนจะมีผลทำให้โอกาสที่ครอบครัวจะสามารถชำระหนี้ได้

ชำระหนี้ได้เป็น 5.8, 6.4, 1.17, 1.8 และ 1.86 เท่าของโอกาสที่ครอบครัวจะไม่สามารถชำระหนี้ได้ สำหรับองค์ประกอบหลัก (สมาชิกในครอบครัวแสดงออกถึงความรักและความเอาใจใส่ระหว่างกัน ,สมาชิกในครอบครัวยอมรับและเคารพความคิดเห็นซึ่งกันและกัน ,สมาชิกในครอบครัวสื่อสารกันอย่างมีคุณภาพ ,การพึ่งตนเองด้านเศรษฐกิจ และการหลีกเลี่ยงภาวะเสี่ยงและการปรับตัวได้ในภาวะยากลำบาก) ตามลำดับ

ตารางที่ 4: ค่าประมาณ ความคลาดเคลื่อนมาตรฐานและ ค่าอัตราส่วนออดส์ ของพารามิเตอร์ที่มีนัยสำคัญทางสถิติของตัวแบบการถดถอยลอจิสติกที่ใช้อันดับองค์ประกอบหลักเป็นตัวแปรอธิบาย ที่มีการปรับค่าเอนเอียงด้วยวิธีการของ Rare Events

ตัวชี้วัด	องค์ประกอบหลัก	$\hat{\beta}$	S.E.	$Exp(\hat{\beta})$
ตัวชี้วัดที่ 1	สมาชิกในครอบครัวแสดงออกถึงความรักและความเอาใจใส่ระหว่างกัน (PC_{111})	2.656	0.857	14.235
ตัวชี้วัดที่ 2	สมาชิกในครอบครัวยอมรับและเคารพความคิดเห็นซึ่งกันและกัน (PC_{121})	2.641	0.817	14.027
ตัวชี้วัดที่ 3	สมาชิกในครอบครัวสื่อสารกันอย่างมีคุณภาพ (PC_{132})	2.278	0.702	9.757
ตัวชี้วัดที่ 6	การพึ่งตนเองด้านเศรษฐกิจ (PC_{361})	1.752	0.634	5.766
ตัวชี้วัดที่ 10	การหลีกเลี่ยงภาวะเสี่ยงและการปรับตัวได้ในภาวะยากลำบาก ($PC_{5,10,2}$)	2.985	1.393	19.786
	Constant	4.235	2.590	

ตัวแบบพยากรณ์ความน่าจะเป็นที่ครอบครัวมีความสามารถชำระหนี้ได้ ที่ได้จากการวิเคราะห์การถดถอยลอจิสติกที่ใช้อันดับองค์ประกอบหลักของตัวชี้วัดเป็นตัวแปรอธิบาย และมีการปรับค่าเอนเอียงด้วยวิธีการของ Rare Events สามารถเขียนได้ ดังต่อไปนี้

$$\hat{\pi}(PC_i) = \frac{\exp(4.235 + 2.656PC_{111} + 2.641PC_{121} + 2.278PC_{132} + 1.752PC_{361} + 2.985PC_{5,10,2})}{1 + \exp(4.235 + 2.656PC_{111} + 2.641PC_{121} + 2.278PC_{132} + 1.752PC_{361} + 2.985PC_{5,10,2})}$$

และสามารถเขียนตัวแบบในรูปแบบเชิงเส้น ได้ดังนี้

$$\text{logit}(\hat{\pi}(x)) = \log\left(\frac{\hat{\pi}(x_i)}{1 - \hat{\pi}(x_i)}\right) = 4.235 + 2.656PC_{111} + 2.641PC_{121} + 2.278PC_{132} + 1.752PC_{361} + 2.985PC_{5,10,2}$$

จากตัวแบบที่ได้จะเห็นว่า การปรับค่าเอนเอียงของค่าประมาณพารามิเตอร์จะทำให้เห็นผลขององค์ประกอบหลักแต่ละองค์ประกอบได้ชัดเจนมากยิ่งขึ้น โดยค่าคะแนนองค์ประกอบหลักที่เพิ่มขึ้น 1 คะแนนจะมีผลทำให้โอกาสที่ครอบครัวจะสามารถชำระหนี้ได้เป็น 14.2, 14, 9.8, 5.8 และ 19.8 เท่าของโอกาสที่ครอบครัวจะไม่สามารถชำระหนี้ได้ สำหรับองค์ประกอบหลัก (สมาชิกในครอบครัวแสดงออกถึงความรักและความเอาใจใส่ระหว่างกัน ,สมาชิกในครอบครัวยอมรับและเคารพความคิดเห็นซึ่งกันและกัน ,สมาชิกในครอบครัวสื่อสารกันอย่างมีคุณภาพ ,การพึ่งตนเองด้านเศรษฐกิจ และการหลีกเลี่ยงภาวะเสี่ยงและการปรับตัวได้ในภาวะยากลำบาก) ตามลำดับ

ตารางที่ 5: ร้อยละความถูกต้องของการพยากรณ์ของตัวแบบการถดถอยลอจิสติกโดยใช้วิธี LASSO วิธีองค์ประกอบหลัก (PC) และ วิธี Rare Events

กลุ่ม	LASSO	PC	Rare Events
ครอบครัวที่สามารถชำระหนี้ได้ ($Y = 1$)	96.7%	99.9%	99.9%
ครอบครัวที่ไม่สามารถชำระหนี้ได้ ($Y = 0$)	41.7%	25.0%	33.3%
ความถูกต้องโดยรวมของข้อมูล	96.0%	98.9%	99.6%

จากตารางที่ 5 จะพบว่าตัวแบบการถดถอยลอจิสติก โดยใช้วิธี LASSO วิธีองค์ประกอบหลัก และ วิธี Rare Events มีความสามารถในการพยากรณ์กลุ่มครอบครัวที่มีความสามารถชำระหนี้ได้ ถูกต้องสูงกว่า ร้อยละ 96 โดยเฉพาะวิธีองค์ประกอบหลัก และ วิธี Rare Events ที่พยากรณ์กลุ่มได้ถูกต้องร้อยละ 100 แต่การพยากรณ์กลุ่มครอบครัวที่ไม่สามารถชำระหนี้ได้นั้น พบว่า ทั้ง 3 วิธีพยากรณ์ได้ถูกต้องน้อยมาก โดยวิธีการ LASSO พยากรณ์ถูกต้องมากที่สุดคือร้อยละ 41.7 และวิธีการ PC พยากรณ์ถูกน้อยที่สุดเพียงร้อยละ 25 และเมื่อมีการปรับค่าเอนเอียงของค่าประมาณพารามิเตอร์ด้วยวิธี Rare Events ทำให้การพยากรณ์ถูกเพิ่มขึ้นเป็นร้อยละ 33.3

สรุปผลการวิจัย

ตัวแบบการถดถอยลอจิสติกที่เหมาะสมที่ใช้ในการพยากรณ์ความน่าจะเป็นของความสามารถในการชำระหนี้ได้ของครัวเรือนเมื่อพิจารณาจากดัชนีชี้วัดความเข้มแข็งของครอบครัวคือตัวแบบการถดถอยลอจิสติกที่ใช้วิธีการของ LASSO มาแก้ปัญหาความสัมพันธ์เชิงเส้นระหว่างตัวชี้วัดย่อยของมาตรฐานครอบครัวเข้มแข็ง เนื่องจากตัวแบบการถดถอยลอจิสติก LASSO มีร้อยละของการพยากรณ์ถูกในการพยากรณ์กลุ่มครอบครัวที่ไม่สามารถชำระหนี้ได้ซึ่งมีสัดส่วนในตัวอย่างน้อยมากเมื่อเทียบกับกลุ่มครอบครัวที่สามารถชำระหนี้ได้สูงกว่าวิธีการวิเคราะห์ที่ใช้วิธีองค์ประกอบหลักทั้งที่มีและไม่มีการปรับค่าเอนเอียงของค่าประมาณสัมประสิทธิ์การถดถอย นอกจากนั้นเมื่อพิจารณาตัวแบบจะพบว่าตัวแบบการถดถอยลอจิสติก LASSO เป็นตัวแบบที่ง่ายกว่า

ตัวแบบที่ใช้อธิบายประกอบหลักเพราะสามารถใช้ตัวชี้วัดย่อยได้โดยตรง ซึ่งจากผลการวิจัยครั้งนี้พบว่า ตัวชี้วัดย่อยที่มีผลต่อการพยากรณ์ความน่าจะเป็นของความสามารถในการชำระหนี้ได้ของครัวเรือน ได้แก่ ครอบครัวมีการออมในรูปแบบต่าง ๆ สมาชิกในครัวเรือนมีการช่วยเหลือเกื้อกูลกันระหว่างเครือญาติ ครอบครัวมีการอบรมให้สมาชิกในครอบครัวใช้บริการพื้นที่สาธารณะ และครอบครัวร่วมมือกันแก้ไขปัญหาเมื่อเกิดภาวะวิกฤติ ซึ่งตัวชี้วัดย่อยเหล่านี้มีอิทธิพลทางบวกกับความน่าจะเป็นที่ครอบครัวจะสามารถชำระหนี้ได้โดยค่าที่เพิ่มขึ้น 1 คะแนนจะมีผลทำให้โอกาสที่ครอบครัวจะสามารถชำระหนี้ได้สูงกว่าเกิน 10 เท่าของโอกาสที่ครอบครัวจะไม่สามารถชำระหนี้ได้

ข้อเสนอแนะ

ในงานวิจัยนี้ ผู้วิจัยศึกษาการแก้ปัญหาตัวแบบการถดถอยลอจิสติกโดยใช้วิธีการ LASSO โดยไม่ได้พิจารณาในกรณีของ Rare Events ในการปรับค่าเอนเอียงของค่าประมาณ ในอนาคต ก่อนการสร้างตัวแบบการถดถอยลอจิสติกโดยใช้วิธีการ LASSO จะต้องพิจารณาการปรับค่าเอนเอียงของค่าประมาณจากวิธี Rare Events ก่อน

รายการอ้างอิง

- กอง บก. ข่าวเศรษฐกิจ [ออนไลน์] ม.หอการค้า เผยสำรวจพบหนี้ครัวเรือน [อ้างเมื่อ 17 ตุลาคม 2560]. จาก <https://prachatai.com/journal/2017/10/73684>
- สำนักงานจังหวัดปทุมธานี [ออนไลน์] บรรยายสรุปจังหวัดปทุมธานี. [อ้างเมื่อ 15 กุมภาพันธ์ 2561]. จาก <http://103.28.101.10/briefprovince/filedoc/13000000.pdf>
- กัลยา วาณิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 4 กรุงเทพฯ: ธรรมสาร; 2552.
- โสภณิศ หนูทิศ และประเสริฐ จรรยาสุภาพ. (2554). ปัจจัยที่มีความสัมพันธ์ต่อการค้างชำระหนี้ของลูกหนี้กองทุน หมุนเวียน เพื่อการกู้ยืมแก่เกษตรกรและผู้ยากจน กรณีศึกษา : จังหวัดชัยภูมิ. วารุณี โรมรัตน์พันธ์. (2556:1-283). โครงการรวบรวมและวิเคราะห์ตัวชี้วัดด้านครอบครัวตามมาตรฐานครอบครัวเข้มแข็งสำนักงานกิจการสตรีและสถาบันครอบครัว. พิมพ์ครั้งที่ 1. กรุงเทพฯ:สำนักงานกิจการสตรีและสถาบันครอบครัว กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์.
- วิฐรา พิงพาพงศ์. (2558). บทวิเคราะห์วิธีวิเคราะห์การถดถอยเชิงเส้นสำหรับข้อมูลที่มีมิติสูง. วารสารวิทยาศาสตร์และเทคโนโลยี, 23(2),213-223.
- Kosuke Imai, Gary King, and Olivia Lau. 2007. “relogit: Rare Events Logistic Regression for Dichotomous Dependent Variables” in Kosuke