บทที่ 3
ระเบียบวิธีวิจัย

3.1 วิธีการวิจัย

ในการศึกษาถึงการก่อหนี้สินของบุคคลที่ทำงานอยู่ในจังหวัดสมุทรปราการ ในครั้งนี้มีระเบียบและวิธีวิจัยดังนี้

3.1.1 กลุ่มตัวอย่างที่ใช้ในการศึกษา

ในการศึกษาครั้งนี้ ทำการเก็บรวบรวมข้อมูลจากบุคคลที่ทำงานอยู่ในเขตจังหวัดสมุทรปราการ โดยทำการสุ่มตัวอย่างโดยใช้สูตรของทาโร่ยามาเน่ จากขนาดของประชากรมากกว่าหนึ่งแสนคน และกำหนดให้ความคลาดเคลื่อนของการสุ่มตัวอย่างที่ยอมรับได้ไม่เกิน 5 เปอร์เซ็นต์ เพราะฉะนั้นจะทำการสุ่มตัวอย่างจำนวน 400 ตัวอย่าง (ตารางที่ 3.1)

ตารางที่ 3.1 ขนาดของกลุ่มตัวอย่างของทาโร ยามาเน่ ที่ระดับความเชื่อมั่น 95 % และความคลาดเคลื่อนต่างๆ

ขนาดประชากร	ขนาดของกลุ่มตัวอย่างที่ระดับความคลาดเคลื่อน (e)					
	± 1%	± 2%	± 3%	± 4%	± 5%	± 10%
500	*	*	*	*	222	83
1,000	*	*	*	385	286	91
1,500	*	*	638	441	316	94
2,000	*	*	714	476	333	95
2,500	*	1,250	769	500	345	96
3,000	*	1,364	811	517	353	97
3,500	*	1,458	843	530	359	97
4,000	*	1,538	870	541	364	98
4,500	*	1,607	891	549	367	98
5,000	*	1,667	909	556	370	98
6,000	*	1,765	938	566	375	98
7,000	*	1,842	959	574	378	99
8,000	*	1,905	976	580	381	99
9,000	*	1,957	989	584	383	99
10,000	5,000	2,000	1,000	588	385	99
15,000	6,000	2,143	1,034	600	390	99
20,000	6,667	2,222	1,053	606	392	100
25,000	7,143	2,273	1,064	610	394	100
50,000	8,333	2,381	1,087	617	397	100
100,000	9,091	2,439	1,099	621	398	100
∞	10,000	2,500	1,111	625	400	100

หมายเหตุ. * หมายถึงขนาดตัวอย่างไม่เหมาะสมที่จะสมมติให้เป็นการกระจายแบบปกติ จึงไม่สามารถใช้สูตรคำนวณขนาดของกลุ่มตัวอย่างได้

ที่มา: จักรกฤษณ์ สำราญใจ (2544)

การเลือกกลุ่มตัวอย่างจะใช้วิธีการสุ่มตัวอย่างแบบเท่ากัน โดยจะแบ่งการสุ่มตัวอย่างเท่ากันในทุกๆ อำเภอของจังหวัดสมุทรปราการ จาก 6 อำเภอ วิธีการเก็บข้อมูลแบบสอบถามจะใช้วิธีการสุ่มแบบบังเอิญ (Accidental Random Sampling) กำหนดจำนวนตัวอย่าง 400 ตัวอย่าง

3.1.2 วิธีการเก็บรวบรวมข้อมูล

ในการศึกษาครั้งนี้ได้มีการเก็บรวบรวมข้อมูลซึ่งแบ่งออกเป็น 2 ประเภท ดังนี้

1) ข้อมูลปฐมภูมิ (Primary data) เป็นการเก็บรวบรวมข้อมูลที่ได้จากการออกแบบสอบถามที่สร้างขึ้น โดยสอบถามจากกลุ่มตัวอย่างของบุคคลที่ทำงานอยู่ในจังหวัดสมุทรปราการ

2) ข้อมูลทุติยภูมิ (secondary data) เป็นการเก็บรวบรวมข้อมูลภาวะหนี้สินที่ได้จากเอกสารข้อมูลทางสถิติที่หน่วยงานต่างๆ ได้จัดเก็บรวบรวมไว้ เช่น สำนักงานสถิติแห่งชาติ เอกสารและเว็บไซต์ต่างๆ ที่เกี่ยวกับจังหวัดสมุทรปราการ

3.2 แบบจำลอง Logit Model

พระคมสัน เจริญวงค์ (2554) ได้กล่าวถึง George Judge, et al (1988) เป็นปรมาจารย์ที่ได้คิดค้น Logit Model ขึ้นและต่อมา William Greene (2003) ได้อธิบายที่มาของแบบจำลองโลจิต โดยเริ่มจากลักษณะของข้อมูล แล้วโยงไปจนถึงฟังก์ชั่นโลจิต ซึ่งผลลัพธ์สุดท้ายออกมาเป็นฟังก์ชั่นเส้นตรงของ log odds โดยใช้แนวคิดเรื่อง repeated observations ซึ่งสอดคล้องกับการอธิบายของ George Judge, et al (1988) ต่อมาได้มีนักเศรษฐศาสตร์รางวัลโนเบล ชื่อ Daniel McFadden (1975) (รางวัลโบเลบ ปี พ.ศ. 2543) ได้เขียนถึง “ปัญหาทางเลือก” ซึ่งเป็นปัญหาใหญ่ทางเศรษฐศาสตร์คู่กับปัญหาความขาดแคลน และแสดงความสัมพันธ์ของแบบจำลองโลจิตกับ Utility ไว้อย่างมีหลักการในหนังสือชื่อว่า “Urban Tavel Demand : A Behavioral Analysis ร่วมกับ Thomas Domas Domencich ตั้งแต่ปี พ.ศ. 2518 ซึ่งนักวิชาการทั่วโลกได้ยึดถือเป็นแบบแผนของการวิเคราะห์ทางเลือก (Choices) ด้วยแบบจำลองโลจิตตั้งแต่บัดนั้นเป็นต้นมา

Logit Model หรือการวิเคราะห์ถดถอยโลจิต (logit Regression Analysis) คือ เทคนิคการวิเคราะห์ตัวแปรหลายตัว เป็นเทคนิคทางสถิติที่สามารถควบคุมตัวแปรได้หลายตัวพร้อมๆ กัน เพื่อศึกษาอิทธิพลของปัจจัยที่มีผลต่อตัวแปรเป้าหมายที่กำหนด หากสามารถความคุมปัจจัยที่มีผลต่อตัวแปรตามได้มากเท่าใด การสรุปเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรต้น และตัวแปรตามจะถูกต้องมากขึ้นเท่านั้น

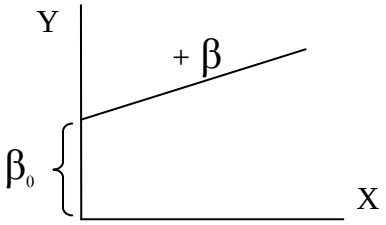
เทคนิคการวิเคราะห์ตัวแปรหลายตัว นอกจากการวิเคราะห์ถดถอยโลจิต ได้แก่ การวิเคราะห์ถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression Analysis : MLR) ใช้วิธีประมาณ

ค่าพารามิเตอร์ คือ OLS (Ordinary Least Squares Method) แต่การวิเคราะห์ถดถอยเชิงเส้นแบบพหุแตกต่างจากการวิเคราะห์ถดถอยโลจิต คือ การวิเคราะห์ถดถอยโลจิต ตัวแปรตามมีค่าเป็น 1 กับ 0 ส่วนตัวแปรต้นอาจมีค่าเป็น 1 กับ 0 หรือตัวแปรเชิงปริมาณก็ได้ ส่วนการวิเคราะห์ถดถอยเชิงเส้นแบบพหุ ทั้งตัวแปรตาม และตัวแปรต้นเป็นตัวแปรเชิงปริมาณ คือ เป็นข้อมูลช่วง หรืออัตราส่วน

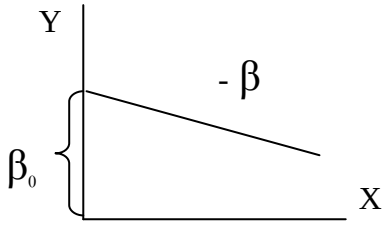
การวิเคราะห์ถดถอยโลจิต จะมีการนำเสนอในรูปแบบของสมการ ดังนี้

$$\text{Log}\left(\frac{P(y=1)}{1-P(y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots$$

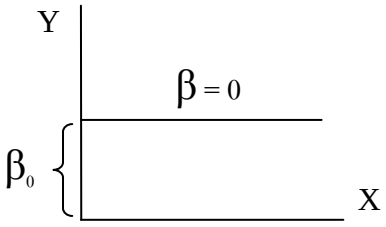
การวิเคราะห์ถดถอยเชิงเส้นแบบพหุ ตัวแปรตามอยู่ด้านซ้ายมือ และตัวแปรต้นอยู่ด้านขวามือ ตัวแปรต้นจะมีค่าสัมประสิทธิ (β) แสดงลักษณะความสัมพันธ์ระหว่างตัวแปรต้น กับตัวแปรตามว่าเป็นไปในทิศทางใด + หรือ – และค่าสัมประสิทธิ (โดยไม่คำนึงเครื่องหมาย) ยังแสดงปริมาณความสัมพันธ์ระหว่างตัวแปรต้น กับตัวแปรตามว่ามีค่ามากน้อยเพียงใด ถ้ามีค่ามากก็แสดงว่าตัวแปรนั้นมีความสัมพันธ์กับตัวแปรตามมาก เมื่อพิจารณากราฟจะพบว่าค่าสัมประสิทธิ คือ ค่าความลาดชัน (slope) นั่นคือ ความสัมพันธ์ระหว่างตัวแปรต้น กับตัวแปรตามยิ่งมีค่ามากความลาดชันก็จะมาก หมายความว่า เมื่อค่าของตัวแปรต้นเปลี่ยนแปลง ค่าของตัวแปรตามก็จะเปลี่ยนแปลงมาก ส่วนค่า β_0 คือ ค่าคงที่ (constant) หรือจุดตัด (intercept) หรือค่าจุดเริ่มต้น (origin) ดังรูป



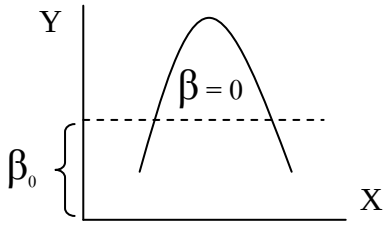
ภาพที่ 3.1 ความสัมพันธ์เชิงบวก



ภาพที่ 3.2 ความสัมพันธ์เชิงลบ



ภาพที่ 3.3 ความสัมพันธ์เท่ากับศูนย์



ภาพที่ 3.4 ความสัมพันธ์เส้นโค้ง

ค่าสัมประสิทธิเป็นบวก หมายถึง การผันแปรของตัวแปรตาม และตัวแปรต้นเป็นไปในทิศทางเดียวกัน คือ ถ้าค่าของตัวแปรต้นเพิ่มขึ้น ค่าของตัวแปรตามจะเพิ่มขึ้นตามไปด้วย

ค่าสัมประสิทธิเป็นลบ หมายถึง การผันแปรของตัวแปรตาม และตัวแปรต้นเป็นไปในทิศทางตรงกันข้ามกัน คือ ถ้าค่าของตัวแปรต้นเพิ่มขึ้น ค่าของตัวแปรตามจะลดลง

ความสัมพันธ์ = 0 หมายถึง ตัวแปรต้นไม่มีความสัมพันธ์กับตัวแปรตาม ไม่ว่า X จะมีค่าเพิ่มขึ้นเท่าใด Y มีค่าเท่าเดิม

ความสัมพันธ์เชิงเส้นโค้ง หมายถึง X มีความสัมพันธ์เชิงเส้นโค้งกับ Y ทำให้ถ้าใช้เทคนิคการวิเคราะห์ถดถอยมาใช้ X จะไม่มีความสัมพันธ์กับ Y เพราะค่าสัมประสิทธิ์ถดถอยมีค่าเป็น 0 ซึ่งในความเป็นจริง X กับ Y มีความสัมพันธ์กัน แต่สัมพันธ์เชิงเส้นโค้ง

ความสัมพันธ์มาก – น้อย ระหว่างตัวแปรต้นและตัวแปรตาม พิจารณาจากค่าตัวเลขส่วนมากเป็นจุดทศนิยม ยิ่งค่าสัมประสิทธิมีค่ามากเท่าใด (โดยไม่คำนึงถึงเครื่องหมายบวกหรือลบ) ก็แสดงว่าตัวแปรต้นและตัวแปรตามมีความสัมพันธ์กันมากเท่านั้น

แนวคิดหลักของการวิเคราะห์ถดถอยโลจิต คือ ค่าล็อกธรรมชาติ (natural logarithm) ของอัตราส่วนของโอกาส (odd ratio) โดยการหาค่าโลจิต (logit) ของอัตราส่วนโอกาสและเปลี่ยนให้เป็นค่าล็อกธรรมชาติ (ln) คือ $\exp\beta$ หรือค่า e^{β}

ความหมายของ $\exp\beta$ หรือ e^{β} สรุปได้ 3 ลักษณะ ดังนี้

1. ถ้า $\beta > 0$ จะทำให้ $e^{\beta} > 1$ หมายความว่าค่า Odds เพิ่มขึ้น หรือโอกาสที่จะเกิดเหตุการณ์ที่สนใจเพิ่มมากขึ้น

2. ถ้า $\beta < 0$ จะทำให้ $e^{\beta} < 1$ หมายความว่าค่า Odds ลดลง หรือโอกาสที่จะเกิดเหตุการณ์ที่สนใจลดลง

3. ถ้า $\beta = 0$ จะทำให้ $e^{\beta} = 1$ หมายความว่าค่า Odds ไม่เพิ่มขึ้นหรือลดลง

Odd Ratio ของตัวแปรกลุ่มจะแสดงถึงโอกาสที่จะเกิดเหตุการณ์ของกลุ่มนี้เป็นกี่เท่าของกลุ่มอ้างอิง

$\exp\beta$ หรือ e^{β} เป็นค่าที่แสดงถึงการเปลี่ยนแปลงของ Odd Ratio เมื่อตัวแปรทำนายเปลี่ยนไป 1 หน่วย ถ้าค่า $\exp\beta$ มีค่ามากกว่า 1 แสดงว่าเมื่อตัวแปรทำนายมีค่าเพิ่มขึ้น จะช่วยเพิ่มโอกาสของการเกิดเหตุการณ์ ($Y \rightarrow 1$) สำหรับตัวแปรทำนายที่เป็นทวิ เช่น เพศ จะช่วยเพิ่มโอกาสของการเกิดเหตุการณ์ ($Y = 1$) น้อยกว่าเป็นกี่เท่าเมื่อเทียบกับกลุ่มอ้างอิง ($Y = 0$)

กรณีที่ตัวแปรต้นเป็นตัวแปรเชิงปริมาณ เมื่อตัวแปรทำนายเพิ่มขึ้น 1 หน่วย จะทำให้โอกาสของการเกิดเหตุการณ์เพิ่มขึ้นเป็น $(e^{\beta} - 1) \times 100\%$

นอกจากสิ่งที่ต้องพิจารณาหลังจากได้ผลการวิเคราะห์การถดถอยแบบโลจิตแล้วคือการพิจารณาว่าแบบจำลองเข้ากับข้อมูลได้ดี (goodness of fit) หรือไม่มากน้อยเพียงใด มีประเด็นที่ต้องพิจารณา ดังนี้

1) การทดสอบทางสถิติเกี่ยวกับตัวแปรอิสระแต่ละตัว ได้แก่ การพิจารณาค่าทางสถิติพื้นฐาน เช่น ค่าเฉลี่ย (Mean) ค่ามัธยฐาน (Median) ค่าสูงสุด (Maximum) ค่าต่ำสุด (Minimum) ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) และเพื่อทดสอบว่าการวิเคราะห์ด้วยแบบจำลอง Logit Model ในการศึกษาเป็นไปตามสมมติฐานของสมการโลจิต ต้องพิจารณาเงื่อนไข ดังนี้

1.1) ตัวแปรตาม (Y) ต้องเป็น Binary Response ตัวแปรต้น (X) เป็น Dummy Variable / Interval / Ratio Scale ก็ได้

1.2) ค่าเฉลี่ยของความคลาดเคลื่อน = 0 [E(€) = 0] ถ้าตัว variance ของตัว error term มีค่าไม่คงที่ แสดงว่ามีปัญหา Heteroscedasticity ซึ่งสามารถทดสอบได้ด้วยวิธี White's test

1.3) ค่าความคลาดเคลื่อนไม่มีความสัมพันธ์กันเอง [Cov(€i€j) = 0] พิจารณาจากตัวทดสอบของ Durbin Watson

1.4) ตัวแปรอิสระกับค่าความคลาดเคลื่อนต้องเป็นอิสระแก่กัน คือ ถ้าค่าความคลาดเคลื่อนทำให้ตัวแปรอิสระมีผลต่อตัวแปรตามน้อย แสดงว่ามีปัญหา ซึ่งสามารถทดสอบได้โดยการพิจารณาค่าการกระจายตัวแบบปกติ (Jarque-Bera : JB) เป็นการทดสอบการแจกแจงปกติของตัวคลาดเคลื่อน (Normality Test) ใช้กับการทดสอบตัวอย่างขนาดใหญ่ที่มีตัวอย่างตั้งแต่ 30 ตัวอย่างขึ้นไป

1.5) ตัวแปรอิสระจะต้องไม่มีความสัมพันธ์กันเอง นั่นก็คือ การทดสอบปัญหา Multicolinearity

1.6) จำนวนตัวอย่างต้องมีอย่างน้อย หรือเท่ากับ 30 x P [$n \geq 30 \times P$] [P = Parameter]

2) แบบจำลองในภาพรวม พิจารณาค่า -2LL (log likelihood) ถ้าแบบจำลองเหมาะสมดีมากกับข้อมูล ค่า -2LL จะเท่ากับ 0 (ค่า -2LL เป็นค่าคำนวณแบบจำลองที่มีค่าคงที่ ซึ่งมีจุดตัดเพียงจุดเดียว ใช้เป็นฐานในการเปรียบเทียบกับแบบจำลองที่มีตัวแปรต้นเพิ่มขึ้น) ค่าของ -2LL ลดลง แสดงว่าตัวแปรต้นนั้นทำให้แบบจำลองเข้ากับข้อมูลได้มากขึ้น

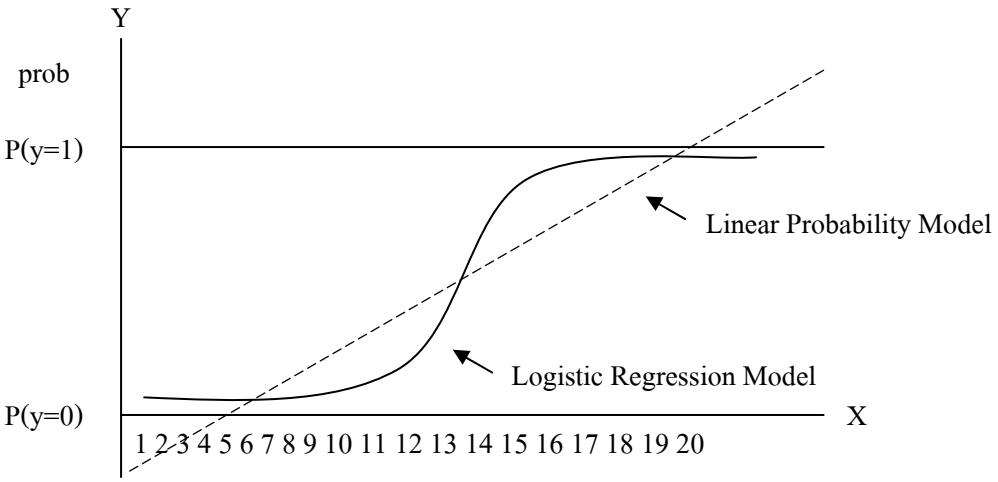
นอกจากค่า -2LL สามารถพิจารณาแบบจำลองในภาพรวมด้วยสถิติทดสอบ 4 ตัว ได้แก่ (1) อัตราส่วนความเป็นไปได้ (likelihood ratio) (2) สถิติทดสอบ (score) (3) สถิติทดสอบวอลด์ (Wald tests) และ (4) สถิติทดสอบการเข้ากับข้อมูลได้ดีฮอสเมอร์แอนด์เลมโชว์ (Hosmer & Lemeshow) โดยทั่วไปสถิติทั้ง 4 ตัวจะให้ผลคล้ายคลึงกัน แต่ถ้าให้ผลแตกต่างกันควรใช้อัตราส่วนความเป็นไปได้ (likelihood ratio) และการทดสอบคะแนน (score) เท่านั้น หรือพิจารณาจากค่า McFadden R-Square ซึ่งโดยปกติจะมีค่าไม่เกิน 0.40

3) ความเหมาะสมใช้ได้ดีของแบบจำลอง (goodness of fit statistics) สำหรับแบบจำลอง Logit Model ให้มีความสำคัญค่อนข้างน้อยกับ goodness of fit statistics แต่ให้

ความสำคัญมากกับการวิเคราะห์ คือ เครื่องหมายที่แสดงความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตาม(β) เป็น + หรือ – หรือ 0

4) การประเมินความถูกต้องของโอกาส / ความน่าจะเป็นที่ได้จากการพยากรณ์ (predicated probabilities) พิจารณาจาก Classification Table หรือ Overall Correct Classification

แบบจำลอง Logit Model ตัวแปรตามมีค่าเพียงสองค่าคือ 1 กับ 0 เท่านั้น ดังนั้นถ้าตัวแปรต้น X มีค่าน้อยๆ ก็ทำให้โอกาสที่ตัวแปรตาม Y มีค่าเท่ากับ 0 ถ้าตัวแปรต้น X มีค่ามากๆ ก็ทำให้โอกาสที่ตัวแปรตาม Y มีค่าเท่ากับ 1



ภาพที่ 3.5 ฟังก์ชั่นเส้นโค้งของ Logistic Regression

ในทางคณิตศาสตร์สามารถเขียนฟังก์ชั่น Logistics ได้ดังนี้

$$P(Y=1) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots}}$$

$$= \frac{e^{\beta x}}{1 + e^{\beta x}}$$

$$= \frac{1}{1 + e^{-\beta x}}$$

ฟังก์ชั่นนี้มีความหมายว่า เมื่อใส่ค่า X (ตัวแปรต้น) จะทำให้ค่า Y (ตัวแปรตาม) ที่ได้มีค่าระหว่าง 0 กับ 1 โดยไม่รวมค่า Y = 0 และ Y = 1

การวิเคราะห์ถดถอยโลจิสติกส์ จะมีการแปลงค่าที่ประมาณได้ให้อยู่ช่วง 1 และ 0 ดังนั้นถ้าตัวแปรตามมีค่าเป็น 1, 2, 3 ซึ่งเป็นลักษณะของตัวแปรแบบ Nominal Scale ทำให้ไม่

สามารถนำมาเข้าสมการวิเคราะห์ถดถอยได้ เพราะไม่สามารถบวกลบคูณหารได้เหมือนตัวเลขปกติ ดังนั้น ต้องทำให้เป็นตัวแปรหุ่นก่อน โดยที่จำนวนตัวแปรหุ่นของตัวแปรแบบ Nominal Scale แต่ละตัวจะน้อยกว่าจำนวนกลุ่มย่อยของตัวแปรนั้นอยู่เสมอ

สมการวิเคราะห์ถดถอยจะมีตัวแปรหุ่นกี่ตัวก็ได้ ตราบใดที่ยังมีอัตราความเป็นอิสระมากพอสมควร คือจำนวนตัวแปรตามยังน้อยกว่าจำนวนวิเคราะห์ (กลุ่มตัวอย่าง)

1) ความเป็นเส้นตรง (linearity) ของความสัมพันธ์ระหว่างตัวแปรต้น และตัวแปรตาม ดังแสดงในภาพที่ 3.1, 3.2, 3.3 ดังนั้นหาก X กับ Y มีความสัมพันธ์กันแบบเส้นโค้ง การวิเคราะห์ความถดถอย โลจิต จะทำให้ได้ค่าสัมประสิทธิ์ = 0 ซึ่งแสดงว่า X กับ Y ไม่มีความสัมพันธ์กัน ทั้งที่ในความเป็นจริงแล้วมีความสัมพันธ์กัน ดังนั้นต้องปรับสมการให้เป็นสมการเส้นตรงก่อนจึงใช้การวิเคราะห์ความถดถอยโลจิตได้

2) ความสัมพันธ์กันมากระหว่างตัวแปรอิสระด้วยกัน (Multicollinearity) วัตถุประสงค์การวิเคราะห์ถดถอยโลจิต คือ ศึกษาอิทธิพลตัวแปรต้นแต่ละตัวที่มีต่อตัวแปรตาม และอิทธิพลของตัวแปรต้นทั้งหมดต่อตัวแปรตาม และเมื่อได้แบบแผนความสัมพันธ์ระหว่างตัวแปรแล้วก็นำไปประมาณค่าของโอกาส หรือความน่าจะเป็น (probabilities) ที่หน่วยวิเคราะห์แต่ละหน่วยจะเป็นสมาชิกของกลุ่มใดกลุ่มหนึ่ง ดังนั้น หากตัวแปรต้นแต่ละตัวสัมพันธ์กันก็ไม่สามารถแยกอธิบายอิทธิพลของตัวแปรต้นแต่ละตัวต่อตัวแปรตามได้ ซึ่งหากสมการมีปัญหา Multicollinearity สามารถแก้ปัญหาได้ 3 แบบ คือ

2.1) ตัดตัวแปรใดตัวแปรหนึ่งที่สัมพันธ์กันมากออกไป

2.2) สร้างตัวแปรใหม่จาก 2 ตัวแปรตาม แล้วใช้ตัวแปรใหม่นั้นแทน

2.3) การขจัดผลของตัวแปรตัวหนึ่งออกจากตัวแปรอีกตัวหนึ่ง

การพิจารณาค่าเปอร์เซ็นต์การทำนาย เป็นการเปรียบเทียบจำนวนกรณีที่เป็นค่าสังเกต (observed cases) กับจำนวนกรณีที่เป็นค่าทำนาย หรือจำแนกได้โดยใช้แบบจำลองว่าถูกต้องร้อยละเท่าใด ยิ่งถูกต้องมาก แบบจำลองยิ่งใช้ได้ดีมาก ค่าของโอกาสที่กำหนดไว้ล่วงหน้า (pior probabilities) คือค่าของโอกาสของแต่ละกรณีที่จะถูกจำแนกได้อย่างถูกต้องในกลุ่มใดกลุ่มหนึ่ง ค่าดังกล่าวนี้ คืออัตราส่วนของขนาดของประชากรของแต่ละกลุ่ม ส่วนค่าของโอกาสที่ได้ภายหลัง (posterior probabilities) คือค่าของโอกาสที่แต่ละกรณีจะถูกกำหนดโดยแบบจำลองที่ใช้ เมื่อนำแบบจำลองไปใช้จะถูกต้องได้ร้อยละเท่าใดของตัวอย่างทั้งหมด

3.3 วิธีการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้จะทำการศึกษาเกี่ยวกับปัจจัยทางเศรษฐกิจและสังคมที่มีผลต่อภาระหนี้สิน เช่น เพศ อายุ สถานภาพ ระดับการศึกษา อาชีพ สมาชิกที่เป็นภาระพึ่งพิง ลักษณะของรายได้ที่ได้รับ รายได้เฉลี่ยต่อเดือน เงินออม รายจ่ายเฉลี่ยต่อเดือน หนี้สินนอกระบบ

ในการวิเคราะห์ข้อมูลของปัจจัยที่มีผลกระทบต่อภาระหนี้ จะใช้การวิเคราะห์ข้อมูลด้วยวิธีการ Logit Model และทำการเก็บข้อมูลจากกลุ่มตัวอย่างจำนวน 400 ตัวอย่าง ซึ่งแสดงโดยสมการแบบจำลองดังนี้

$$Y = f \{X1, X2, X3A, X3B, X3C, X4, X5A, X5B, X5C, X6, X7, X8, X9, X10, X11\}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{3A} X_{3A} + \beta_{3B} X_{3B} + \beta_{3C} X_{3C} + \beta_4 X_4 + \beta_{5A} X_{5A} + \beta_{5B} X_{5B} + \beta_{5C} X_{5C} + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + €$$

ตัวแปรตาม

Y = หนี้สิน[1]ของบุคคลที่ทำงานอยู่ในเขตจังหวัดสมุทรปราการ
= 1 มีหนี้สินในระบบ
= 0 ไม่มีหนี้สินในระบบ

ตัวแปรต้น

X1 = เพศ
X2 = อายุ (ปี)
X3 = สถานภาพ
X4 = ระดับการศึกษา
X5 = อาชีพ
X6 = สมาชิกที่เป็นภาระพึ่งพิง
X7 = ลักษณะของรายได้ที่ได้รับ
X8 = รายได้เฉลี่ยต่อเดือน

---

[1] หนี้สิน หมายรวมถึง หนี้ที่เกิดจากการกู้ยืม อันได้แก่ ธนาคารพาณิชย์ เครือญาติ ไฟแนนซ์ ลิซซิ่ง สหกรณ์ กลุ่มออมทรัพย์ กองทุนหมู่บ้าน บุคคลภายนอก ผู้ปล่อยกู้

X9 = เงินออม[2]

X10 = รายจ่ายเฉลี่ยต่อเดือน

X11 = หนี้สินนอกระบบ

Y เป็นตัวแปรหุ่น

กำหนดให้ 1 = มีหนี้สิน

0 = ไม่มีหนี้สิน

X1 เป็นตัวแปรหุ่น

กำหนดให้ 1 = เพศชาย

0 = เพศหญิง

X2 เป็นตัวแปรเชิงปริมาณ คือ อายุ

X3A เป็นตัวแปรหุ่น

กำหนดให้ 1 = สมรส

0 = โสด

X3B เป็นตัวแปรหุ่น

กำหนดให้ 1 = หย่าร้าง

0 = โสด

X3C เป็นตัวแปรหุ่น

กำหนดให้ 1 = แยกกันอยู่

0 = โสด

โดยให้สถานภาพโสดเป็นตัวแปรฐาน

X4 เป็นตัวแปรหุ่น

กำหนดให้ 1 = ปริญญาตรี/สูงกว่าปริญญาตรี

0 = ต่ำกว่าปริญญาตรี

X5A เป็นตัวแปรหุ่น

กำหนดให้ 1 = ประกอบธุกิจส่วนตัว / ค้าขาย

0 = พนักงานบริษัทเอกชน

---

[2] กรณีถ้าบุคคลนั้นมีเงินเหลือในบัญชี ถือว่าบุคคลนั้นมีเงินออม

X5B เป็นตัวแปรหุ่น

กำหนดให้ 1 = รับจ้าง / แรงงาน

0 = พนักงานบริษัทเอกชน

X5C เป็นตัวแปรหุ่น

กำหนดให้ 1 = ข้าราชการ / รัฐวิสาหกิจ

0 = พนักงานบริษัทเอกชน

โดยให้พนักงานบริษัทเอกชนเป็นตัวแปรฐาน

X6 เป็นตัวแปรหุ่น

กำหนดให้ 1 = มีสมาชิกที่เป็นภาระพึ่งพิง

0 = ไม่มีสมาชิกที่เป็นภาระพึ่งพิง

X7 เป็นตัวแปรหุ่น

กำหนดให้ 1 = รับเป็นเงินเดือนประจำ

0 = ไม่ได้รับเป็นเงินเดือนประจำ

X8 เป็นตัวแปรเชิงปริมาณ คือ รายได้เฉลี่ยต่อเดือน

X9 เป็นตัวแปรหุ่น

กำหนดให้ 1 = มีเงินออม

0 = ไม่มีเงินออม

X10 เป็นตัวแปรเชิงปริมาณ คือ รายจ่ายเฉลี่ยต่อเดือน

X11 เป็นตัวแปรหุ่น

กำหนดให้ 1 = มีหนี้สินนอกระบบ

0 = ไม่มีหนี้สินนอกระบบ

β_0 คือ ค่าคงที่

$\beta_1, \beta_2, \beta_3, \ldots \beta_{11}$ คือ ค่าพารามิเตอร์

ϵ คือ ค่า Error