การวิเคราะห์ความถดถอยโลจิสติค

Logistic Regression Analysis

ความหมายของการวิเคราะห์ความถดถอยโลจิสติค

- การวิเคราะห์ความถดถอยโลจิสติคแบบ 2 กลุ่ม (Binary Logistic Regression Analysis) เป็นเทคนิคการวิเคราะห์ความถดถอยแบบหนึ่งที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพมีค่าได้เพียง 2 ค่า ส่วนตัวแปรอิสระอาจจะเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพหรืออาจจะมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพก็ได้
- การวิเคราะห์แบบโลจิสติคไม่มีเงื่อนไขเกี่ยวกับการแจกแจงของตัวแปรอิสระ และเงื่อนไขเกี่ยวกับเมทริกซ์ค่าแปรปรวนและแปรปรวนร่วมของแต่ละกลุ่ม
- การวิเคราะห์ความถดถอยโลจิสติกเป็นการพยากรณ์โอกาสที่แต่ละหน่วยจะอยู่ในกลุ่มใดกลุ่มหนึ่ง

ประเภทของการวิเคราะห์ความถดถอยโลจิสติค แบ่งตามจำนวนตัวแปรอิสระ

1. มีตัวแปรอิสระเพียง 1 ตัว การวิเคราะห์ความถดถอยโลจิสติคอย่างง่าย (Simple Logistic Regression)

2. มีตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป การวิเคราะห์ความถดถอยโลจิสติกเชิงพหุ (Multiple Logistic Regression)

ประเภทของการวิเคราะห์ความถดถอยโลจิสติค แบ่งตามจำนวนกลุ่มของตัวแปรตาม

1. เมื่อตัวแปรตามมีค่าได้เพียง 2 ค่า หรือแบ่งหน่วยข้อมูลออกเป็น 2 กลุ่มย่อย เช่น

   Y = สถานภาพของผู้ที่แต่งงานแล้ว

$$Y = \begin{cases} 1 & \text{ถ้าหย่า/แยกกันอยู่} \\ 0 & \text{ถ้าอยู่ด้วยกัน} \end{cases}$$

กรณีจะเรียกว่า การวิเคราะห์ความถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression)

2. เมื่อตัวแปรตามมีค่ามากกว่า 2 ค่า หรือแบ่งหน่วยข้อมูลออกเป็นกลุ่มย่อยมากกว่า 2 กลุ่ม เช่น การศึกษาสถานภาพของลูกค้าสินเชื่อ

Y =	1	ถ้าเป็นลูกหนี้ชั้นดี
	2	ถ้าเป็นลูกหนี้ที่มีปัญหาบ้าง
	3	ถ้าเป็นลูกหนี้ (NPL)

กรณีจะเรียกว่า การวิเคราะห์ความถดถอยโลจิสติกแบบหลายกลุ่ม (Multinomial Logistic Regression Analysis)

ตัวอย่างที่ 1

- การศึกษาการเป็นโรคหัวใจ กับอายุ น้ำหนัก โคเลสเตอรอล ฯลฯ

$$Y = \begin{cases} 1 & \text{ถ้าคนไข้เป็นโรคหัวใจ} \\ 0 & \text{ถ้าคนไข้ไม่เป็นโรคหัวใจ} \end{cases}$$

ส่วนตัวแปรอิสระคือ อายุ น้ำหนัก โคเลสเตอรอล

สำหรับการวิเคราะห์ความถดถอยโลจิสติก จะคำนวณหาความน่าจะเป็นหรือโอกาสที่เหตุการณ์จะเกิด เช่น

P{Y=1} = P{คนไข้เป็นโรคหัวใจ}

P{Y=0} = P{คนไข้ไม่เป็นโรคหัวใจ}

ตัวอย่างที่ 2

- การศึกษาการเสพยาเสพติดของเด็กวัยรุ่น กับเพศ อายุ สภาพครอบครัว รายได้ครอบครัว การศึกษาของบิดามารดา

$$Y = \begin{cases} 1 & \text{ถ้าเด็กเสพยาเสพติด} \\ 0 & \text{ถ้าเด็กไม่ได้เสพยาเสพติด} \end{cases}$$

ส่วนตัวแปรอิสระคือ เพศ อายุ สภาพครอบครัว รายได้ครอบครัว การศึกษาของบิดามารดา

สำหรับการวิเคราะห์ความถดถอยโลจิสติก จะคำนวณหาความน่าจะเป็นหรือโอกาสที่เหตุการณ์จะเกิด เช่น

P{Y=1} = P{เด็กเสพยาเสพติด}

P{Y=0} = P{เด็กไม่ได้เสพยาเสพติด}

วัตถุประสงค์

1. เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ พร้อมทั้งศึกษาระดับความสัมพันธ์ระหว่าง ตัวแปรอิสระแต่ละตัวกับตัวแปรตาม หรือ ศึกษาว่าตัวแปรอิสระตัวใดบ้างที่มีอิทธิพลหรือมีผลกระทบต่อตัวแปรตาม และตัวแปรอิสระตัวใดที่มีผลกระทบต่อตัวแปรตามมาก เช่นศึกษาว่าปัจจัยใดบ้างเป็นปัจจัยสำคัญที่ทำให้มีโอกาสเป็นโรคหัวใจ หรือปัจจัยใดเป็นปัจจัยที่สำคัญที่ทำให้เด็กวัยรุ่นมีโอกาสติดยาเสพติด
2. เพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ โดยการใช้สมการที่สร้างขึ้นด้วยปัจจัยหรือตัวแปรอิสระที่มีผลต่อตัวแปรตามที่ได้จากการศึกษาในวัตถุประสงค์ข้อที่ 1 เมื่อทราบค่าตัวแปรอิสระจะทำให้สามารถพยากรณ์โอกาสที่จะเกิดเหตุการณ์ได้

- แนวคิดที่เป็นแกนหลักของการถดถอยโลจิสติค คือ Logit ซึ่งหมายถึงค่า Natural Logatithm ของอัตราส่วนของโอกาส (Odd ratio) เพื่อให้ผู้ศึกษาเข้าใจถึงที่มาของอัตราส่วนของโอกาส
- ตัวอย่าง

การเที่ยวกลางคืน	เพศ		รวม
	ชาย	หญิง	
เที่ยว	73	15	88
ไม่เที่ยว	23	11	34
รวม	96	26	122

ค่าโอกาส (Odd) ของผู้ชายที่จะเที่ยวกลางคืนคือ $\frac{\mathbf{73}}{\mathbf{23}} = \mathbf{3.17}$

ค่าโอกาส (Odd) ของผู้หญิงที่จะเที่ยวกลางคืนคือ $\frac{\mathbf{15}}{\mathbf{11}} = \mathbf{1.26}$

อัตราส่วนโอกาส (Odds ratio) ที่เด็กผู้ชายจะเที่ยวกลางคืนเปรียบเทียบกับ

เด็กผู้หญิง คำนวณได้จาก $\frac{\mathbf{3.17}}{\mathbf{1.36}} = \mathbf{2.33}$

ซึ่งหมายความว่าเด็กผู้ชายวัยรุ่นจะเที่ยวกลางคืนมากกว่าผู้หญิงถึง 2.33 เท่า

- การหาค่า Logit ของอัตราส่วนโอกาสดังกล่าว กระทำได้โดยการเปลี่ยน 2.33 เป็น natural logarithm $\boldsymbol{ln2.33}$ ซึ่งจะได้เท่ากับ 0.85 ซึ่ง ค่า 0.85 นี้คือค่าสัมประสิทธิ์ถดถอยยกกำลัง $\boldsymbol{Exp\ \beta}$ ของเพศ ถ้าเราใช้การวิเคราะห์ถดถอยโลจิสติคที่ผลลัพธ์ (ตัวแปรตาม) มีสองค่าว่าขึ้นอยู่กับเพศมากน้อยเพียงใด
- สมมติว่า p เป็นโอกาสที่เด็กจะเที่ยว และ (1-p) เป็นโอกาสที่เด็กจะไม่เที่ยว ซึ่งโอกาสจะเที่ยวหรือไม่ขึ้นอยู่กับเบี้ยเลี้ยงรายวัน (stipend)

$$\boldsymbol{ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X}$$

ซึ่ง

$$\boldsymbol{ln\left(\frac{p}{1-p}\right) = logitY}$$

$$\boldsymbol{logitY = \beta_0 + \beta_1 X}$$

หากเราแก้สมการ Log (Antilog) เราจะได้

$$\boldsymbol{p} = \text{โอกาส}(\boldsymbol{Y} = \text{ผลลัพธ์}(\text{การเที่ยวกลางคืน}) | \boldsymbol{X} = \boldsymbol{x}) = \frac{\boldsymbol{e}^{\beta_0 + \beta_1 \boldsymbol{X}}}{\boldsymbol{1} + \boldsymbol{e}^{\beta_0 + \beta_1 \boldsymbol{X}}}$$

การแปลงค่าของโอกาส(odds) ให้เป็น Natural logarithm

การวิเคราะห์ความถดถอยโลจิสติคอย่างง่าย

Simple Logistic Regression

การวิเคราะห์ความถดถอยโลจิสติคอย่างง่าย

(Simple Logistic Regression)

- เป็นการวิเคราะห์ความถดถอยโลจิสติคที่มีตัวแปรอิสระ 1 ตัว (X)
- การวิเคราะห์ ความถดถอยโลจิสติคอย่างง่ายแบบมี 2 กลุ่ม

ในกรณีนี้ตัวแปรตาม Y มีค่าได้เพียง 2 ค่า คือ 0 กับ 1 ดังนั้นตัวแปรตาม Y จะมีการแจกแจงแบบ Bernoulli

$$\mathbf{P\{Y = y\} = p^{y} (1 - p)^{1-y}; y = 0, 1} \qquad (1.1)$$

สำหรับตัวอย่างหน่วยที่ i จะได้ว่า

$$\mathbf{P\{Y_i = y_i\} = p^{y_i}(1 - p)^{1-y_i}; y_i = 0, 1} \qquad (1.2)$$

จากสมการ (1.2) เมื่อ $\boldsymbol{y_i = 0}$ จะได้

$$\mathbf{P\{Y_i = 0\} = p^{0}(1 - p)^{1-0} = 1 - p}$$

จากสมการ (1.2) เมื่อ $\boldsymbol{y_i = 1}$ จะได้

$$\mathbf{P\{Y_i = 1\} = p^1(1-p)^{1-1} = p}$$

$$\mathbf{E\{Y_i\}} = \sum \boldsymbol{y_i}\,\mathbf{P\{Y_i = y_i\}}$$

$$\mathbf{E\{Y_i\} = 0P\{Y_i = 0\} + 1P\{Y_i = 1\}}$$

$$\therefore\ \mathbf{E}\{\boldsymbol{Y}\} = \mathbf{0(1-p) + 1(p) = p} \qquad (1.3)$$

ซึ่งทำให้ $\mathbf{0} \le \mathbf{E}\{\boldsymbol{Y}\} \le \mathbf{1}$

เนื่องจาก Y มีค่าได้เพียง 2 ค่า คือ 0 และ 1 จึงทำให้ความสัมพันธ์ระหว่าง X และ Y ไม่ได้อยู่ในรูปเชิงเส้น แต่จะอยู่ในรูป

$$\mathrm{E}\{Y\} = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}} \qquad (1.4)$$

โดยเรียกสมการ (1.4) ว่า Logistic Response Function โดยที่

$$\boldsymbol{0 \le E\{Y\} \le 1}$$

จากสมการที่ (1.3) และ (1.4) จะได้

$$\boldsymbol{E\{Y\} = p = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}} \qquad (1.5)$$

โดยที่ $\boldsymbol{p = P\{Y = 1\} = P\{}$เกิดเหตุการณ์ที่สนใจ$\boldsymbol{\} = E\{Y\}}$

$\therefore \boldsymbol{P\{}$ไม่เกิดเหตุการณ์ที่สนใจ$\boldsymbol{\} = P\{Y = 0\} = 1 - p}$

ตัวอย่าง

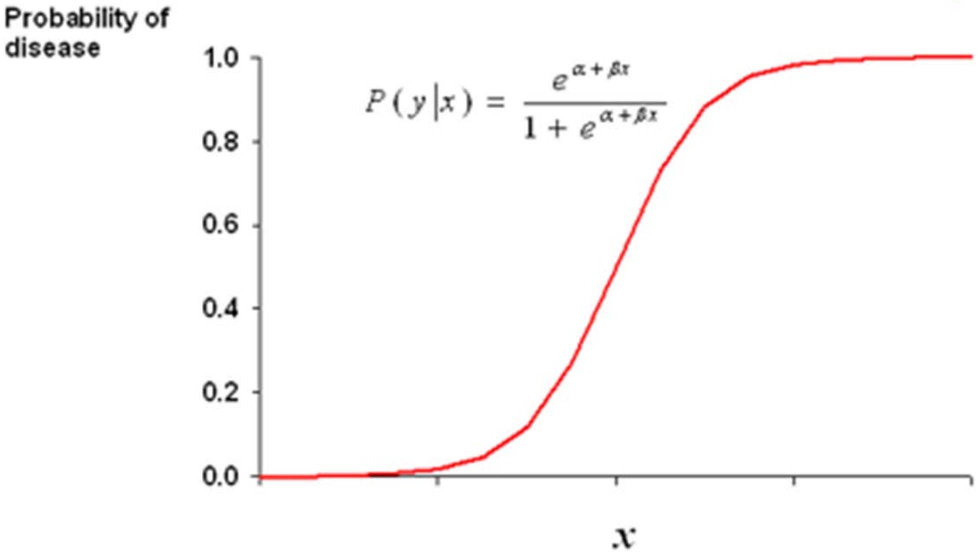
$$P\{\text{เกิดเหตุการณ์}\} = P\{\text{ลูกค้าซื้อสินค้า}\} = P\{Y = 1\}$$

$$P\{\text{ไม่เกิดเหตุการณ์}\} = P\{\text{ลูกค้าไม่ซื้อสินค้า}\} = P\{Y = 0\}$$

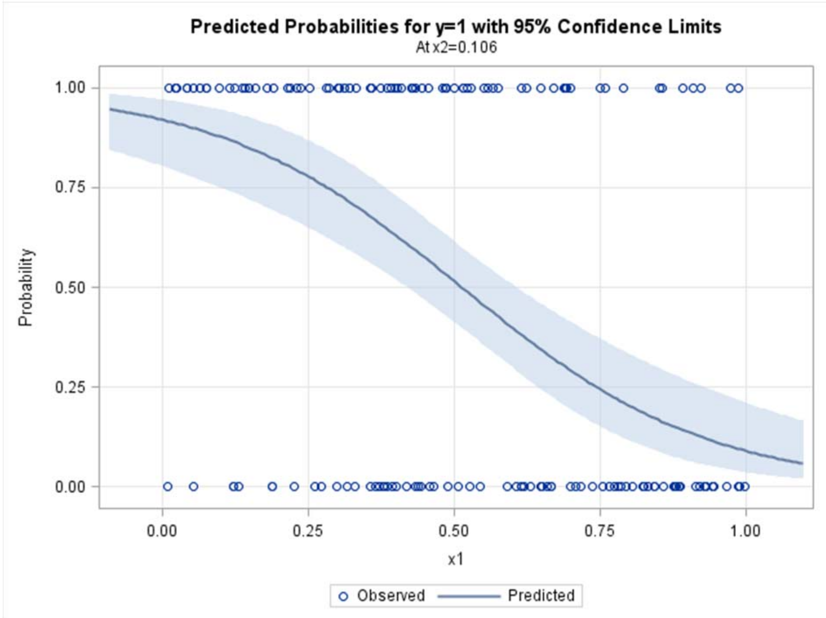
$$= 1 - P\{\text{ลูกค้าซื้อสินค้า}\}$$

กราฟของ $P\{$เกิดเหตุการณ์$\}$

เมื่อ $\beta_1 > 0$



เมื่อ $\beta_1 < 0$



Source: https://onlinecourses.science.psu.edu/stat507/sites/onlinecourses.science.psu.edu.stat507/files/lesson09/graph_function.gif
http://blogs.sas.com/content/iml/files/2014/06/LogisticSim.png

การประมาณค่าสัมประสิทธิ์ความถดถอยโลจิสติค

กรณีใช้ข้อมูลตัวอย่าง เป้าหมายของการวิเคราะห์คือการประมาณค่า $\boldsymbol{\beta_0}$และ $\boldsymbol{\beta_1}$ เนื่องจาก $\mathbf{Y_i}$ มีการแจกแจงแบบ Bernoulli ซึ่งมีฟังก์ชันความน่าจะเป็นดังแสดงในสมการที่ (1.1) และ (1.2) และข้อมูลตัวอย่าง n หน่วยเป็นอิสระกัน ฟังก์ชันความควรจะเป็น (likelihood function) คือ

$$\boldsymbol{L = \prod_{i=1}^{n} p^{Y_i}(1-p)^{1-Y_i}}$$

$$\boldsymbol{L = \prod_{i=1}^{n} \left\{\frac{exp(\beta_0+\beta_1 X)}{1+exp(\beta_0+\beta_1 X)}\right\}^{Y_i} \left\{1-\frac{exp(\beta_0+\beta_1 X)}{1+exp(\beta_0+\beta_1 X)}\right\}^{1-Y_i}}$$

(11.7)

$$P\{\text{ไม่เกิดเหตุการณ์}\} = P\{Y=0\} = 1 - p$$

$$= 1 - \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} = \frac{1+e^{w}-e^{w}}{1+e^{w}} = \frac{1}{1+e^{w}}$$

โดยที่ $w = \beta_0 + \beta_1 X$

สมการที่ (11.7) จะเป็น

$$L = \prod_{i=1}^{n} \left\{\frac{e^{w}}{1+e^{w}}\right\}^{Y_i} \left\{\frac{1}{1+e^{w}}\right\}^{1-Y_i} \qquad (11.8)$$

หาค่า log ของสมการที่ (11.8)

$$log_e(L) = lnL = \sum_{i=1}^{n} (Y_i ln[P(Y_i)] + (1-Y_i) ln[1-P(Y_i)])$$

(11.9)

การประมาณค่า $\boldsymbol{\beta_0}$และ $\boldsymbol{\beta_1}$จะใช้หลักการของความควรจะเป็นสูงสุด (Maximum Likelihood) หรือการประมาณค่า $\boldsymbol{\beta_0}$และ $\boldsymbol{\beta_1}$ที่ทำให้ $\boldsymbol{lnL}$ ในสมการ (11.8) เทียบกับ $\boldsymbol{\beta_0}$และ $\boldsymbol{\beta_1}$แล้วให้เท่ากับศูนย์ อย่างไรก็ตามมาสามารถหาค่า $\boldsymbol{\beta_0}$และ $\boldsymbol{\beta_1}$ได้โดยตรง เนื่องจากสมการไม่ได้อยู่ในรูปเชิงเส้น จึงใช้เทคนิคการทำซ้ำ (Iteration techniques)

Haberman (1978) ได้ศึกษาโดยใช้วิธีการของ Newton-Raphson และพบว่าเป็นวิธีที่มีประสิทธิภาพ ดังนั้นส่วนใหญ่จึงมักใช้โปรแกรมสำเร็จรูปทางสถิติเช่น STATA SAS SPSS ในการประมาณค่า $\boldsymbol{\beta_0}$และ $\boldsymbol{\beta_1}$ โดยที่ค่าประมาณของ $\boldsymbol{\beta_0}$ คือ $\widehat{\boldsymbol{\beta_0}}$และ ค่าประมาณของ $\boldsymbol{\beta_1}$ คือ $\widehat{\boldsymbol{\beta_1}}$และเรียกว่า$\widehat{\boldsymbol{\beta_0}}$ และ $\widehat{\boldsymbol{\beta_1}}$ว่าตัวประมาณความควรจะเป็นสูงสุด (Maximum likelihood Estimator: MLE)

ตัวอย่าง

จากการศึกษาความแตกต่างของการจ่ายเงินปันผลของบริษัทในตลาดหลักทรัพย์ โดยสุ่มตัวอย่างบริษัทในตลาดหลักทรัพย์ที่จ่ายเงินปันผลในปีที่ผ่านมา 20 แห่ง และบริษัทที่ไม่จ่ายเงินปันผลมา 20 แห่งเช่นกัน เนื่องจากขนาดของบริษัทต่างกัน โดยกำหนดว่าถ้าทุนจดทะเบียนมากกว่า 1,000 ล้านบาท คือเป็นบริษัทขนาดใหญ่ ไม่เกิน 1,000 ล้านบาทคือเป็นขนาดเล็ก

ตัวแปรตามคือ $Y = \begin{cases} 1 & \text{ถ้าปีที่แล้วจ่ายเงินปันผล} \\ 0 & \text{ถ้าปีที่แล้วไม่จ่ายเงินปันผล} \end{cases}$

$Size = \begin{cases} 1 & \text{ถ้าเป็นบริษัทขนาดใหญ่} \\ 0 & \text{ถ้าเป็นบริษัทขนาดเล็ก} \end{cases}$

สร้างฟังก์ชันความควรจะเป็น L ดังในสมการ (11.7) และ $\boldsymbol{lnL}$ ดังในสมการ (11.9) ประมาณค่า $\boldsymbol{\beta_0}$และ $\boldsymbol{\beta_1}$ที่ทำให้ $\boldsymbol{lnL}$ ในสมการ (11.9) มีค่ามากที่สุด ได้ค่า $\widehat{\boldsymbol{\beta_0}}$ และ $\widehat{\boldsymbol{\beta_1}}$ ในการประมาณซ้ำหลายๆรอบ

$$\widehat{\boldsymbol{\beta_0}} = -\mathbf{1.163}$$

$$\widehat{\boldsymbol{\beta_1}} = \mathbf{1.322}$$

$$\boldsymbol{E(Y) = P\{Y = 1\} = P}\{\text{บริษัทจะจ่ายเงินปันผล}\}$$

$$= \frac{\boldsymbol{e^{-1.163+1.322Size}}}{\boldsymbol{1+e^{-1.163+1.322Size}}}$$

การปรับรูปแบบความสัมพันธ์ให้อยู่ในรูปเชิงเส้น

จากสมการของการวิเคราะห์ความถดถอยแบบโลจิสติคในสมการที่ 11.6 หรืออยู่ในรูป $\boldsymbol{lnL}$ ในสมการที่ 11.9 จะพบว่า ความสัมพันธ์ไม่ได้อยู่ในรูปเชิงเส้น จึงมีการปรับให้อยู่ในรูปเชิงเส้นได้ดังนี้

กำหนดให้ Odd Ratio (OR) เป็นอัตราส่วนระหว่างโอกาสที่เหตุการณ์จะเกิดกับโอกาสที่เหตุการณ์จะไม่เกิด

$$\therefore \boldsymbol{Odds\ Ratio = OR = \frac{P\{\text{เกิดเหตุการณ์}\}}{P\{\text{ไม่เกิดเหตุการณ์}\}} = \frac{P\{Y=1\}}{P\{Y=0\}} = \frac{p}{1-p}}$$

$$\boldsymbol{= \frac{\dfrac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{\dfrac{1}{1+e^{\beta_0+\beta_1 X}}}}$$

$$\therefore \boldsymbol{OR = e^{\beta_0+\beta_1 X}} \qquad (1.11)$$

- OR จึงเป็นค่าที่แสดงถึงโอกาสที่จะเกิดเหตุการณ์ว่าเป็นกี่เท่าของโอกาสที่จะไม่เกิดเหตุการณ์เช่น
    - OR ของตัวอย่างการศึกษาความแตกต่างของการจ่ายเงินปันผลของบริษัทในตลาดหลักทรัพย์ ถ้าได้ OR ของบริษัท A =2.5 แสดงว่า โอกาสที่บริษัท A จะจ่ายเงินปันผลเป็น 2.5 เท่าของโอกาสที่จะไม่จ่ายเงินปันผล
    - OR ของตัวอย่าง การติดยาเสพติดของวัยรุ่น ถ้าได้ OR ของนายไพรัช =3 แสดงว่า โอกาสที่นายไพรัชจะเสพยาเป็น 3 เท่าของโอกาสที่นายไพรัชจะไม่เสพยา

ความหมายของค่า Odd Ratio

- ถ้าค่า OR มากกว่า 1 แสดงว่า โอกาสที่จะเกิดเหตุการณ์มากกว่าโอกาสที่จะไม่เกิดเหตุการณ์

- ถ้าค่า OR = 1 แสดงว่า โอกาสที่จะเกิดเหตุการณ์และไม่เกิดเหตุการณ์เท่ากัน

- ถ้าค่า OR น้อยกว่า 1 แสดงว่า โอกาสที่จะเกิดเหตุการณ์ต่ำกว่าโอกาสที่จะไม่เกิดเหตุการณ์

นำสมการที่ (1.10) มาหาค่า $\boldsymbol{log_e\left(e^{\beta_0+\beta_1 X}\right)}$

$$\boldsymbol{log_e(OR) \quad = log_e\left(e^{\beta_0+\beta_1 X}\right)}$$

$$\boldsymbol{= \beta_0 + \beta_1 X}$$

ด้านขวาของสมการที่ (1.11) อยู่ในรูปเชิงเส้น โดยเรียกสมการที่ (1.11) ว่าฟังก์ชันตอบสนองโลจิท (Logit Response function) หรือกรณีที่ใช้ข้อมูลตัวอย่าง สมการ (1.11) จะกลายเป็น

$$\boldsymbol{log_e(OR) \quad = ln(\widehat{OR}) \;= \hat{\beta}_0 + \hat{\beta}_1 X} \qquad (1.12)$$

จากตัวอย่าง การศึกษาความแตกต่างของการจ่ายเงินปันผลของบริษัทในตลาดหลักทรัพย์

$$\widehat{\beta_0} = -1.163$$

$$\widehat{\beta_1} = 1.322$$

ค่าประมาณของ Odd-ratio คือ

$$\widehat{OR} = e^{\hat{\beta}_0+\hat{\beta}_1 X}$$

$$\widehat{OR} = e^{-1.163+1.322Size} \qquad (1.13)$$

ฟังก์ชันตอบสนองโลจิท คือ

$$ln(\widehat{OR}) = -1.163 + 1.322Size \qquad (1.14)$$

ตัวอย่าง การศึกษาความแตกต่างของการจ่ายเงินปันผลของบริษัทในตลาดหลักทรัพย์

กรณีที่มีตัวแปรอิสระเพียง 1 ตัว และเป็นเชิงคุณภาพ ตัวแปรอิสระคือขนาดของบริษัท หรือ ตัวแปรSIZE มีค่าได้เพียง 2 ค่า ดังนั้นการประมาณค่า $\boldsymbol{\beta_0}$และ$\boldsymbol{\beta_1}$ อาจศึกษาได้โดยการสร้างตารางจำแนก 2 ทาง

การจ่ายเงินปันผล Y	ขนาด (Size)		รวม
	ใหญ่ (1)	เล็ก (0)	
จ่ายเงินปันผล	15	5	20
ไม่จ่ายเงินปันผล	4	16	20
รวม	19	21	40

การคำนวณความน่าจะเป็นทำได้ดังนี้

1. $\boldsymbol{P}\{\text{จ่ายเงินปันผล}\} = \mathbf{P}\{\boldsymbol{Y} = \mathbf{1}\} = \dfrac{\mathbf{20}}{\mathbf{40}} = \mathbf{0.5}$
2. $\boldsymbol{P}\{\text{บริษัทใดๆในตลาดหลักทรัพย์จะจ่ายเงินปันผลโดยกำหนดว่าเป็นบริษัทใหญ่ๆ}\} =$ $\mathbf{P}\{\boldsymbol{Y} = \mathbf{1} | \boldsymbol{SIZE} = \mathbf{1}\} = \dfrac{\mathbf{15}}{\mathbf{19}} = \mathbf{0.79}$
3. $\boldsymbol{P}\{\text{บริษัทในตลาดหลักทรัพย์จะจ่ายเงินปันผลโดยกำหนดว่าเป็นบริษทเล็ก}\} =$ $\mathbf{P}\{\boldsymbol{Y} = \mathbf{1} | \boldsymbol{SIZE} = \mathbf{0}\} = \dfrac{\mathbf{5}}{\mathbf{21}} = \mathbf{0.24}$
4. $\boldsymbol{Odd\ Ratio}$ เมื่อกำหนดว่าเป็นบริษัทขนาดใหญ่ $= \dfrac{\mathbf{P}\{\boldsymbol{Y}=\mathbf{1}|\boldsymbol{SIZE}=\mathbf{1}\}}{\mathbf{P}\{\boldsymbol{Y}=\mathbf{0}|\boldsymbol{SIZE}=\mathbf{1}\}}$ **หรือ**

$$\boldsymbol{OR}\{\boldsymbol{Y} = \mathbf{1} | \boldsymbol{SIZE} = \mathbf{1}\} = \frac{\mathbf{15/19}}{\mathbf{4/19}} = \mathbf{15/4} = \mathbf{3.75} \tag{1.15}$$

5. $\boldsymbol{Odd\ Ratio}$ เมื่อกำหนดว่าเป็นบริษัทขนาดเล็ก $= \dfrac{P\{Y=1|SIZE=0\}}{P\{Y=0|SIZE=0\}}$

หรือ $\boldsymbol{OR\{Y = 1|SIZE = 0\} = \frac{5/21}{16/21} = 0.3125}$ (1.16)

หมายถึงโอกาสที่บริษัทขนาดเล็กจะจ่ายเงินปันผลน้อยกว่าโอกาสที่จะไม่จ่าย

หาค่า ln Odd ของสมการที่ (1.15) และ (1.16) จะได้

$$\boldsymbol{ln\big(OR(Y = 1|Size = 1)\big) = ln(3.75) = 1.322} \quad (1.17)$$

$$\boldsymbol{ln\big(OR(Y = 1|Size = 0)\big) = ln(0.3125) = -1.163} \quad (1.18)$$

จัดสมการ (1.14)-(1.15) ให้อยู่ในรูปฟังก์ชันของขนาดบริษัท

$$\boldsymbol{ln\big(OR(Y = 1|Size)\big) = -1.163 + 1.322Size} \quad (1.19)$$

- **สัมประสิทธิ์ของตัวแปรอิสระ Size ในสมการที่ (1.19) $= +1.322$ หมายถึง ถ้าขนาดของบริษัทใหญ่ขึ้น log ของ OR ของบริษัทขนาดใหญ่จะจ่ายเงินปันผลมากกว่าของบริษัทขนาดเล็ก**
- **นอกจากนี้ เมื่อเปรียบเทียบสมการที่ (1.19) กับฟังก์ชันตอบสนองโลจิทในสมการที่ (1.14) จะพบว่า $\widehat{\beta_0} = -1.163$ และ $\widehat{\beta_1} = 1.322$** หรือกล่าวได้ว่า กรณีที่มีตัวแปรอิสระเพียงตัวเดียวและเป็นตัวแปรเชิงคุณภาพ การประมาณค่าพารามิเตอร์ $\boldsymbol{\beta_0}$**และ** $\boldsymbol{\beta_1}$ อาจใช้การวิเคราะห์ความถดถอยโลจิสติคและการหาความน่าจะเป็นจากตารางแจกแจงความถี่แบบ 2 ทาง จะให้ผลลัพธ์เหมือนกัน

ความหมายของสัมประสิทธิ์ความถดถอยโลจิสติค

เนื่องจาก $\boldsymbol{Odd\ Ratio = OR = \frac{p}{1-p} = e^{\beta_0+\beta_1 X}}$

หรือ $\boldsymbol{\widehat{OR} = \frac{\hat{p}}{1-\hat{p}} = e^{\hat{\beta}_0+\hat{\beta}_1 X}}$

จากตัวอย่าง $\boldsymbol{\widehat{\beta_0} = -1.163}$ และ $\boldsymbol{\widehat{\beta_1} = 1.322}$

$$\boldsymbol{\therefore \widehat{OR} = \frac{\hat{p}}{1-\hat{p}} = e^{-1.163+1.322 Size}}$$

$\boldsymbol{e^{1.322} = 3.75}$ ถ้าขนาดของบริษัทใหญ่ขึ้น (จากบริษัทขนาดเล็กเป็นบริษัทขนาดใหญ่) จะได้ว่า Odds ของการจ่ายเงินปันผลจะเพิ่มขึ้น 3.75 เท่า นั่นคือโอกาสที่บริษัทจะจ่ายเงินปันผลของบริษัทขนาดใหญ่เป็น 3.75 เท่าของบริษัทขนาดเล็ก

$$\boldsymbol{\therefore \hat{P}\{Y=1\} = \hat{p} = \frac{1}{1+e^{-(-1.163+1.322 Size)}}}$$

- โอกาสที่บริษัทขนาดเล็ก (Size=0) จะจ่ายเงินปันผลเท่ากับ

$$\widehat{P}\{Y = 1|Size = 0\} = \widehat{p} = \frac{1}{1 + e^{-(-1.163+1.322Size)}}$$

$$= \frac{1}{1 + e^{1.163}} = 0.24$$

- โอกาสที่บริษัทขนาดใหญ่ (Size=1) จะจ่ายเงินปันผลเท่ากับ

$$\widehat{P}\{Y = 1|Size = 1\} = \widehat{p} = \frac{1}{1 + e^{-(-1.163+1.322Size)}}$$

$$= \frac{1}{1 + e^{-0.159}} = 0.54$$

การวัดความสัมพันธ์โดยใช้ Odd Ratio

ตัวอย่าง

<u>กรณีที่มีตัวแปรอิสระ 1 ตัว</u>

$$\boldsymbol{ln(Odd\ Ratio) = ln(OR) = ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X}$$

<u>กรณีที่ตัวแปรอิสระ X เป็นตัวแปรเชิงคุณภาพที่มีค่าได้ 2 ค่า</u>

การศึกษาการเป็นโรคหัวใจ กับอายุ น้ำหนัก โคเลสเตอรอล ฯลฯ

$$Y = \begin{cases} 1 & \text{ถ้าคนไข้เป็นโรคหัวใจ} \\ 0 & \text{ถ้าคนไข้ไม่เป็นโรคหัวใจ} \end{cases}$$

$$X = \begin{cases} 1 & \text{ถ้าพ่อหรือแม่เป็นโรคหัวใจ} \\ 0 & \text{ถ้าพ่อและแม่ไม่เป็นโรคหัวใจ} \end{cases}$$

$$\boldsymbol{P\{Y = 1\} = \mathbf{p} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}}$$

$$\boldsymbol{ln(OR) = \beta_0 + \beta_1 X}$$

- ถ้าพ่อและแม่ไม่เป็นโรคหัวใจ (X=0)

$$\therefore \boldsymbol{ln(OR: X = 0) = \beta_0 + \beta_1(0) = \beta_0}$$

- ถ้าพ่อหรือแม่เป็นโรคหัวใจ (X=1)

$$\therefore \boldsymbol{ln(OR: X = 1) = \beta_0 + \beta_1(1) = \beta_0 + \beta_1}$$

$$\boldsymbol{\frac{OR(X = 1)}{OR(X = 0)} = e^{\beta_1}}$$

กรณีที่ตัวแปรอิสระ X เป็นตัวแปรเชิงปริมาณ

X = โคเลสเตอรอล

$$P\{\text{เป็นโรคหัวใจ}\} = p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

ถ้าค่าโคเลสเตอรอล $X = \mathrm{x}$

$$ln(OR: X = x) = \beta_0 + \beta_1 x$$

ถ้าค่าโคเลสเตอรอล $X = \mathrm{x} + 1$

$$ln(OR: X = x + 1) = \beta_0 + \beta_1 (x + 1)$$

$$\frac{OR(X = x + 1)}{OR(X = x)} = e^{\beta_1}$$

- สำหรับกรณีที่ตัวแปรอิสระเป็นตัวแปรเชิงปริมาณ การเปรียบเทียบจะไม่สนใจค่าที่ต่างกันเพียง 1 หน่วย นั้นคือจะไม่เทียบโอกาสที่จะเป็นโรคหัวใจเมื่อโคเลสตรอรอลเพิ่มขึ้น 1 เช่นเพิ่มจาก 200 เป็น 220 หรือในรูปทั่วไป $\boldsymbol{X}$ จะเพิ่มขึ้น m หน่วย

$$\frac{\boldsymbol{OR(X = x + m)}}{\boldsymbol{OR(X = x)}} = \boldsymbol{e^{m\beta_1}}$$

ดังนั้น $\boldsymbol{e^{m\beta_1}}$ = โอกาสที่จะเกิดเหตุการณ์เทียบกับโอกาสที่ไม่เกิดเหตุการณ์เมื่อตัวแปรอิสระเปลี่ยนไป m หน่วย

- ถ้า $\boldsymbol{\beta_1 = 0}$ จะทำให้ $\boldsymbol{e^{\beta_1} = 1}$ หมายถึง โอกาสที่จะเกิดเหตุการณ์เท่ากับโอกาสที่จะไม่เกิดเหตุการณ์

- ถ้า $\boldsymbol{\beta_1 > 0}$ จะทำให้ $\boldsymbol{e^{\beta_1} > 1}$ หมายถึง $\boldsymbol{OR > 1}$ หรือโอกาสที่จะเกิดเหตุการณ์มากกว่าโอกาสที่จะไม่เกิดเหตุการณ์

- ถ้า $\boldsymbol{\beta_1 < 0}$ จะทำให้ $\boldsymbol{e^{\beta_1} < 1}$ หมายถึง $\boldsymbol{OR < 1}$ หรือโอกาสที่จะเกิดเหตุการณ์น้อยกว่าโอกาสที่จะไม่เกิดเหตุการณ์

การวิเคราะห์ความถดถอยโลจิสติกเชิงพหุ

Multiple Logistic Regression

- การวิเคราะห์ความถดถอยโลจิสติกที่มีตัวแปรอิสระมากกว่า 1 ตัว ในรูปทั่วไปกำหนดให้ มี p ตัว คือ$\boldsymbol{X_1, X_2, X_3, \ldots, X_p}$ ส่วนตัวแปรตาม $\boldsymbol{Y}$ ยังคงมีค่าได้ 2 ค่า คือ

$$\boldsymbol{Y} = \begin{cases} 1 & \text{ถ้าเกิดเหตุการณ์} \\ 0 & \text{ถ้าไม่เกิดเหตุการณ์} \end{cases}$$

และ $\boldsymbol{Y}$ ยังคงมีการแจกแจงแบบ Bernoulli ที่มีความน่าจะเป็น p

$$p = E\{Y\} = P\{Y = 1\} = \frac{e^{\beta_0 + \beta_1 X_1 + \cdots \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \cdots \beta_p X_p}}$$

หรือ $p\{\text{เกิดเหตุการณ์}\} = P\{Y = 1\} = \frac{1}{1+e^{-(\beta_0 + \beta_1 X_1 + \cdots \beta_p X_p)}}$

และ $p\{\text{ไม่เกิดเหตุการณ์}\} = P\{Y = 0\} = \frac{1}{1+e^{\beta_0 + \beta_1 X_1 + \cdots \beta_p X_p}}$

$$= 1 - P\{Y = 1\} = 1 - \mathrm{p}$$

$$Odd\ Ratio = \mathrm{OR} = \frac{p}{1 - p} = e^{\beta_0 + \beta_1 X_1 + \cdots \beta_p X_p}$$

ฟังก์ชันตอบสนองโลจิท คือ

$$\boldsymbol{log_e(OR)} \quad = \mathbf{ln(OR)} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} X_1 + \cdots \boldsymbol{\beta_p} X_p$$

หรือ กรณีที่ใช้ข้อมูลตัวอย่าง

$$\boldsymbol{log_e\left(\widehat{OR}\right)} = \widehat{\boldsymbol{\beta}}_{\mathbf{0}} + \widehat{\boldsymbol{\beta}}_{\mathbf{1}} X_1 + \cdots + \widehat{\boldsymbol{\beta}}_{\boldsymbol{p}} X_p$$

การประเมินผลแบบจำลองถดถอย Logistic

- การประเมินค่าแบบจำลองที่ใช้ในการวิเคราะห์ว่าดีหรือไม่อย่างไร ซึ่งทำได้หลังจากที่ได้ผลจากการวิเคราะห์ถดถอยแบบ Logistic นักวิจัยต้องชี้ให้เห็นว่าแบบจำลองที่ตนใช้เข้ากับข้อมูลได้ดี (Goodness of fit) หรือไม่มากน้อยเพียงใด
  1. แบบจำลองภาพรวมเป็นอย่างไร
  2. การทดสอบสถิติเกี่ยวกับตัวแปลอิสระแต่ละตัว
  3. ความเหมาะสมใช้ได้ดีของแบบจำลอง (goodness of fit statistics)
  4. การประเมินความถูกต้องของโอกาส/ ความน่าจะเป็นได้จากการพยากรณ์ (predicted probability)

การประเมินแบบจำลองภาพรวม

(Overall model evaluation)

- แบบจำลองจะใช้กับข้อมูลได้อย่างเหมาะสม ถ้าหากว่ามันให้ผลดีกว่าแบบจำลองที่มีจุดตัดแต่เพียงอย่างเดียว
- การประเมินความเหมาะสมใช้ได้ดีของแบบจำลอง Logistic ค่าของ -2LL (Log likelihood) ถ้าแบบจำลองสอดคล้องหรือเข้าได้ดีกับข้อมูล ค่าของ -2LL จะเป็น 0 โดยทั่วไปการหาค่าดังกล่าวนี้จะเริ่มจากการใช้แบบจำลองที่มีแค่ค่าจุดตัดเพียงอย่างเดียวซึ่งจะเป็นฐานที่ใช้เปรียบเทียบกับแบบจำลองที่มีตัวแปรอิสระเพิ่มขึ้น ถ้าค่าของ -2LL ลดลง แสดงว่าตัวแปรอิสระตัวนั้นทำให้แบบจำลองเข้ากับข้อมูลได้ดีมากขึ้น (Mazzarot, 1998 ; Windsperger, 2003)

- สถิติทดสอบ
  - อัตราส่วนความเป็นไปได้ (Likelihood ratio)
  - คะแนน (Score)
  - สถิติ Wald tests
  - สถิติทดสอบการเข้ากับข้อมูลได้ดี Hosmer &Lemeshow

โดยทั่วไปสถิติทั้งสามตัวนี้จะให้ผลคล้ายคลึงกัน แต่ถ้าให้ผลแตกต่างกันควรใช้อัตราส่วนความเป็นไปได้ และการทดสอบคะแนนเท่านั้น (Menard, 1995)

ตัวอย่าง

- การใช้ข้อมูล Mroz.dta จาก T.A. Mroz (1987), “The Sensitivity of an Empirical Model of Married Women’s Hours of Work to Economic and Statistical Assumptions,” *Econometrica* 55, 765-799. เป็นข้อมูลเกี่ยวกับการทำงาน โดยมีตัวแปรสำคัญคือ inlf คืออยู่ในกำลังแรงงาน หรือ การมีงานทำที่เป็น Binary Choice และมีปัจจัยทาง Demographic ที่จะเป็นตัวกำหนดการตัดสินใจในการเข้าสู่ตลาดแรงงาน หรืออาจกล่าวได้ว่าแรงงานจะตัดสินใจทำงานเมื่อ Market Offered Wage สูงกว่าReservation wage โดยในข้อมูลมีตัวแปรที่น่าสนใจ คือ

Estimates of Married Women's Labor Supply Responses from Previous Studies

Evaluated at:

Wife's Wage:	$4.50	Husband's Wage:	$7.00
Wife's Hours:	1500	Husband's Hours:	2000

Household nonlabor income:	$1000
Labor income marginal tax rate:	0.339
Nonlabor income marginal tax rate:	0.280

		Wage Effect	Income Effect (per $1000)
1. Boskin (1973)	Instrumental Variables	29	−16.9
2. Cogan (1980a)	Tobit	865	−32.3
3. Cogan (1980a)	Instrumental Variables	349	−11.7
4. Cogan (1980b)	Tobit	632	−22.8
5. Cogan (1980b)	Fixed Costs	196	−8.5
6. Cogan (1981)	Fixed Costs	269	−22.4
7. Greenhalgh (1980)	Instrumental Variables	213	−65.6
8. Hausman (1981)	Convex Budget Set	328	−125.0
9. Hausman (1981)	Non-convex Budget Set	335	−118.0
10. Hausman (1981)	Fixed Costs	305	−113.0
11. Heckman (1976)	Tobit	1462	−73.4
12. Heckman (1976)	Generalized Tobit	−499	51.0
13. Heckman (1980)	Generalized Tobit	1401	−18.7
14. Layard et al. (1980)	Tobit	128	−118.2
15. Layard et al. (1980)	Instrumental Variables	22	−11.8
16. Leuthold (1978)	1967 Estimates	14	−3.0
17. Leuthold (1978)	1969 Estimates	45	−7.1
18. Leuthold (1978)	1971 Estimates	33	5.8
19. Nakamura and Nakamura (1981)		−16	−15.0
20. Schultz (1980)	Tobit	123	−67.0
21. Schultz (1980)	Instrumental Variables	−26	−1.9

Note: See Mroz (1984) for information on how these wage and income effects are calculated. All effects evaluated in terms of 1975 dollars.

TABLE II

METHODOLOGICAL CLASSIFICATION OF PREVIOUS STUDIES

	Data Set	Single Worker Model	Exogenous Wage	Exogenous Experience	No Sample Selection Controls	Tobit Restrictions	No Taxes	Exogenous Taxes
1. Boskin (1973)	SEO, 1967				×			×
2. Cogan (1980a)	NLS, 1966	×		×		×	×	
3. Cogan (1980a)	NLS, 1966	×		×	×		×	
4. Cogan (1980b)	PSID, 1975	×		×		×	×	
5. Cogan (1980b)	PSID, 1975	×		×			×	
6. Cogan (1981)	PSID, 1975	×		×			×	
7. Greenhalgh (1980)	GHS, 1971	×		×	×			×
8. Hausman (1981)	PSID, 1975	×	×	?		×		
9. Hausman (1981)	PSID, 1975	×	×	?		×		
10. Hausman (1981)	PSID, 1975	×	×	?				
11. Heckman (1976)	NLS, 1967			×		×	×	
12. Heckman (1976)	NLS, 1967			×			×	
13. Heckman (1980)	NLS, 1967						×	
14. Layard, et al. (1980)	GHS, 1974			×		×		×
15. Layard, et al. (1980)	GHS, 1974			×				×
16. Leuthold (1978)	NLS, 1967		×		×			×
17. Leuthold (1978)	NLS, 1969		×		×			×
18. Leuthold (1978)	NLS, 1971		×		×			×
19. Nakumura and Nakumura (1981)	1970 U.S. Census	×						
20. Schultz (1980)	SEO, 1966					×		
21. Schultz (1980)	SEO, 1966				×			

TABLE III

MEANS OF THE DATA

(Standard Deviation in Parentheses.)

Variable name	Full Sample	Working Women
Wife's age	42.5	42.0
	(8.1)	(7.7)
Wife's education	12.3	12.7
	(2.3)	(2.3)
Children less than 6	0.24	0.14
	(0.52)	(0.39)
Children between 6 and 18	1.35	1.35
	(1.32)	(1.32)
Husband's age	45.1	44.6
	(8.1)	(8.0)
Husband's education	12.5	12.6
	(3.0)	(3.0)
Wife's wage (in dollars)	—	4.18
(average hourly earnings)		(3.31)
Husband's wage	7.48	7.23
(average hourly earnings)	(4.23)	(3.57)
Household nonlabor income ($1000)	3.76	3.39
	(5.90)	(6.07)
Wife's hours of work	740.6	1302.9
	(871.3)	(776.3)
Husband's hours of work	2267.3	2233.5
	(595.6)	(582.9)
Number of Observations	753	428

Construction of the Sample

The data used in this analysis come from the twelve year Michigan Panel Study of Income Dynamics tape (Wave XII tape). The following sequential criteria were used to select this sample.

Selection Criteria	Number of Observations Deleted
Total number of records: 6373	
1. Nonrandom low income sample	2876
2. Not living in the U.S. in both 1975 and 1976	13
3. Change in head or wife in household from 1975 to 1976	331
4. Duplicate records due to later splitoffs	379
5. Female head of household in 1975	590
6. Not classified as married in both 1975 and 1976	212
7. Head retired, permanently disabled or did not work in 1975	362
8. No wife interview in 1976	15
9. Incomplete information on birthdates of children	5
10. Nonwhite head of household	146
11. Wife less than 30 years old in 1975	613
12. Wife older than 60 in 1975	31
13. Husband less than 30 years old in 1975	9
14. Husband older than 60 in 1975	28
15. Husband works more than 5200 hrs. in 1975	3
16. Husband's education not reported	2
17. Family income in upper bracket in 1975 (Fam. income > $99,999, not known exactly)	3
18. Unknown taxable income in 1975	1
19. Husband reported no labor earnings in 1975	1

Remaining sample: 753 observations.

Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
inlf	753	.5683931	.4956295	0	1
kidslt6	753	.2377158	.523959	0	3
kidsge6	753	1.353254	1.319874	0	8
age	753	42.53785	8.072574	30	60
educ	753	12.28685	2.280246	5	17
husage	753	45.12085	8.058793	30	60
huseduc	753	12.49137	3.020804	3	17
huswage	753	7.482179	4.230559	.4121	40.509
unem	753	8.623506	3.114934	3	14
city	753	.6427623	.4795042	0	1

```
inlf:  =1 if in lab frce, 1975                     Number of Observations
_______________________________              ______________________________
                                             Total    Integers   Nonintegers
|      #                    Negative             -           -             -
|      #                    Zero               325         325             -
|  #   #                    Positive           428         428             -
|  #   #                                     _____       _____         _____
|  #   #                    Total              753         753             -
|  #   #                    Missing              -
+----------------------                      _____
0                      1                       753
   (2 unique values)
```

OLS

Source	SS	df	MS
Model	25.9631084	9	2.88478983
Residual	158.764647	743	.213680548
Total	184.727756	752	.245648611

Number of obs	=	753
F(9, 743)	=	13.50
Prob > F	=	0.0000
R-squared	=	0.1405
Adj R-squared	=	0.1301
Root MSE	=	.46226

inlf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
kidslt6	-.3061292	.0363436	-8.42	0.000	-.3774776	-.2347807
kidsge6	-.0168673	.0140314	-1.20	0.230	-.0444132	.0106786
age	-.0086999	.0046712	-1.86	0.063	-.0178702	.0004704
educ	.0576548	.0094287	6.11	0.000	.0391447	.0761649
husage	-.0050784	.0046313	-1.10	0.273	-.0141704	.0040136
huseduc	-.0119508	.0075389	-1.59	0.113	-.0267509	.0028492
huswage	-.0119324	.0045379	-2.63	0.009	-.0208411	-.0030237
unem	-.0025946	.0055647	-0.47	0.641	-.0135189	.0083297
city	.0086857	.0380814	0.23	0.820	-.0660743	.0834457
_cons	.8101666	.1745997	4.64	0.000	.467399	1.152934

Logistic Regression

```
Iteration 0:   log likelihood =  -514.8732
Iteration 1:   log likelihood = -457.55568
Iteration 2:   log likelihood = -457.33231
Iteration 3:   log likelihood = -457.33218
Iteration 4:   log likelihood = -457.33218

Logistic regression                               Number of obs   =        753
                                                  LR chi2(9)      =     115.08
                                                  Prob > chi2     =     0.0000
Log likelihood = -457.33218                       Pseudo R2       =     0.1118
```

inlf	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
kidslt6	-1.504129	.1989	-7.56	0.000	-1.893966	-1.114292
kidsge6	-.0898004	.0675764	-1.33	0.184	-.2222478	.042647
age	-.041185	.022117	-1.86	0.063	-.0845336	.0021636
educ	.2758894	.0472609	5.84	0.000	.1832597	.3685191
husage	-.0265899	.0220151	-1.21	0.227	-.0697387	.0165588
huseduc	-.05648	.0356272	-1.59	0.113	-.126308	.013348
huswage	-.0577641	.0219868	-2.63	0.009	-.1008574	-.0146709
unem	-.0122302	.0262631	-0.47	0.641	-.0637049	.0392445
city	.0537199	.1808348	0.30	0.766	-.3007098	.4081496
_cons	1.553326	.8510901	1.83	0.068	-.1147802	3.221432

ต้องหาค่า marginal fixed effect เพื่อระบุการเปลี่ยนแปลงของความน่าจะเป็นจากการที่ตัวแปรมีการเปลี่ยนแปลง

```
Marginal effects after logit
      y  = Pr(inlf) (predict)
         =  .57535332
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
kidslt6	-.3674917	.04876	-7.54	0.000	-.463052	-.271932	.237716
kidsge6	-.0219402	.01651	-1.33	0.184	-.054298	.010418	1.35325
age	-.0100624	.0054	-1.86	0.063	-.020653	.000528	42.5378
educ	.0674058	.01153	5.85	0.000	.044806	.090006	12.2869
husage	-.0064965	.00538	-1.21	0.227	-.017036	.004043	45.1208
huseduc	-.0137993	.0087	-1.59	0.113	-.030856	.003258	12.4914
huswage	-.014113	.00537	-2.63	0.009	-.024642	-.003584	7.48218
unem	-.0029881	.00642	-0.47	0.641	-.015565	.009588	8.62351
city*	.0131392	.04428	0.30	0.767	-.073641	.099919	.642762

(*) dy/dx is for discrete change of dummy variable from 0 to 1

การทดสอบสมมติฐาน

ตัวอย่างที่ 1

```
. test educ

 ( 1)  [inlf]educ = 0

           chi2(  1) =   28.08
         Prob > chi2 =    0.0000
```

ตัวอย่างที่ 2

```
. test huseduc

 ( 1)  [inlf]huseduc = 0

           chi2(  1) =    6.95
         Prob > chi2 =    0.0084
```

การทดสอบสมมติฐาน

ตัวอย่างที่ 3

```
. test kidslt6 kidsge6

 ( 1)  [inlf]kidslt6 = 0
 ( 2)  [inlf]kidsge6 = 0

           chi2(  2) =   55.21
         Prob > chi2 =    0.0000
```

ตัวอย่างที่ 4

```
. test  kidslt6= kidsge6

 ( 1)  [inlf]kidslt6 - [inlf]kidsge6 = 0

           chi2(  1) =   47.39
         Prob > chi2 =    0.0000
```

การทดสอบความแม่นยำ

```
Logistic model for inlf

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       343           165  |        508
     -     |        85           160  |        245
-----------+--------------------------+-----------
   Total   |       428           325  |        753

Classified + if predicted Pr(D) >= .5
True D defined as inlf != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   80.14%
Specificity                     Pr( -|~D)   49.23%
Positive predictive value       Pr( D| +)   67.52%
Negative predictive value       Pr(~D| -)   65.31%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   50.77%
False - rate for true D         Pr( -| D)   19.86%
False + rate for classified +   Pr(~D| +)   32.48%
False - rate for classified -   Pr( D| -)   34.69%
--------------------------------------------------
Correctly classified                        66.80%
--------------------------------------------------
```

Goodness-Of-Fit Test

```
Logistic model for inlf, goodness-of-fit test

       number of observations =         753
 number of covariate patterns =         753
            Pearson chi2(746) =         750.17
                  Prob > chi2 =           0.4503
```

Measure of Fit for Logit

Measures of Fit for logit of inlf

Log-Lik Intercept Only:	-514.873	Log-Lik Full Model:	-461.088
D(746):	922.176	LR(6):	107.570
		Prob > LR:	0.000
McFadden's R2:	0.104	McFadden's Adj R2:	0.091
Maximum Likelihood R2:	0.133	Cragg & Uhler's R2:	0.179
McKelvey and Zavoina's R2:	0.184	Efron's R2:	0.134
Variance of y*:	4.033	Variance of error:	3.290
Count R2:	0.668	Adj Count R2:	0.231
AIC:	1.243	AIC*n:	936.176
BIC:	-4019.376	BIC':	-67.826

หนังสืออ้างอิง

- กัลยา วาณิชย์บัญชา (2552) การวิเคราะห์ข้อมูลหลายตัวแปร. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย
- สุชาต ประสิทธิ์รัฐสินธุ์ (2548) เทคนิคการวิเคราะห์ตัวแปรหลายตัวสำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์. หจก. สามลดา
- Thomas A. Mroz (1987). The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions, Econometrica 55, 765-799.