บทที่ 2
แนวคิดและทฤษฎี

การศึกษาเกณฑ์การคัดเลือกตัวแบบในการวิเคราะห์ความถดถอยพหุนามแบบติดกลุ่มจะเป็นการพิจารณาหาเกณฑ์การคัดเลือกตัวแบบที่ให้ค่าพยากรณ์มีความถูกต้องและแม่นยำมากที่สุด ดังนั้นงานวิจัยครั้งนี้จะเป็นการศึกษาเกณฑ์การคัดเลือกตัวแบบและเปรียบเทียบความถูกต้องของการพยากรณ์ ซึ่งประกอบด้วย เกณฑ์การคัดเลือกตัวแบบโดยสถิติทดสอบเอฟบางส่วน (The partial F-test statistic) ด้วยวิธีการกำจัดตัวแปรแบบถอยหลัง(Backward Elimination (BW)) ซึ่งวิธีการนี้เป็นวิธีการที่เริ่มต้นด้วยการพิจารณาตัวแปรอิสระทุกตัวแปร แล้วค่อยตัดค่าสัมประสิทธิ์ความถดถอยของตัวแปรที่ไม่มีระดับนัยสำคัญออก

เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Akaike's Information Criterion (AIC)) เป็นเกณฑ์ที่ใช้ในการคัดเลือกตัวแบบ โดยใช้แนวคิดของ คูลล์แบล็ค – ไลท์เบอร์ (Kulback-Leibler(1951)) มาใช้ในการพิจารณาหาตัวแบบที่เหมาะสมเพื่อนำไปใช้ในการพยากรณ์ โดยตัวแบบที่ให้ค่า AIC ต่ำสุดจะเป็นตัวแบบที่ดีที่สุด

เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์(Bayesian Information Criterion(BIC)) เป็นเกณฑ์ที่นำแนวคิดของเบส์(Bayesian Approach) มาใช้ในการพิจารณา สำหรับกรณีนี้จะกำหนดการแจกแจงก่อน (Prior Distribution) เป็นแบบไม่มีข้อสนเทศ(Non-informative Prior) เนื่องจากเราไม่ทราบลักษณะพื้นฐานของข้อมูล ดังนั้นการแจกแจงก่อนจะเป็นการแจกแจงสม่ำเสมอ(Uniform Distribution) และนำมารวมเข้ากับฟังก์ชันภาวะน่าจะเป็นของแต่ละตัวแบบ ซึ่งจะได้ความน่าจะเป็นภายหลัง(Posterior Distribution) มาคำนวณหาค่า $\mathrm{BIC} = -2\log(\text{Bayes Factor})$ โดยตัวแบบที่ให้ค่า BIC ต่ำสุดจะเป็นตัวแบบที่ดีที่สุด ทั้ง 3 เกณฑ์นี้จะพิจารณาภายใต้การทดสอบสมมติฐานแบบติดกลุ่ม(Nested Hypothesis) ซึ่งมีแนวคิด และทฤษฎี ดังนี้

2.1 การวิเคราะห์ความถดถอยพหุนาม(Polynomial Regression Analysis)

ตัวแบบและข้อตกลงเบื้องต้นที่นำมาใช้ในการทำวิจัยครั้งนี้ คือ ตัวแบบความถดถอยพหุนามซึ่งมีรูปแบบทั่วไปดังนี้

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots\dots\dots + \beta_p X_i^p + \varepsilon_i \qquad ; i = 1,2,3,\dots\dots,n$$

โดย β_1 คือ สัมประสิทธิ์ผลกระทบเชิงเส้น(Linear Effect Coefficient)

β_2 คือ สัมประสิทธิ์ผลกระทบกำลังสอง(Quadratic Effect Coefficient)

. .

. .

. .

$\beta_{p'}$ คือ สัมประสิทธิ์ผลกระทบกำลัง p'

สำหรับการวิจัยครั้งนี้ได้ศึกษาถึงตัวแบบที่ประกอบด้วยตัวแปรอิสระ 2 ตัวแปร มีกำลังสูงสุดไม่เกิน 6 และมีพจน์อันตรกิริยา สำหรับตัวอย่างกรณีที่ตัวแบบประกอบด้วยตัวแปรอิสระ 2 ตัวแปร ตัวแบบสมบูรณ์(Complete Model)มีกำลังสูงสุดไม่เกิน 2 และมีพจน์อันตรกิริยา มีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_{10}x_{i1} + \beta_{20}x_{i1}^2 + \beta_{01}x_{i2} + \beta_{02}x_{i2}^2 + \beta_{03}x_{i1}x_{i2} + \varepsilon_i \qquad (2.1)$$

โดย β_{03} คือ สัมประสิทธิ์ผลกระทบอันตรกิริยา(Interaction Effect Coefficient)
สำหรับตัวแบบเชิงเส้นทั่วไปจะมีรูปแบบดังสมการที่ (1.1)(หน้า 4)

ตัวแบบความถดถอยพหุนามในกรณีนี้จะเป็นตัวแบบเชิงเส้น นั่นคือ เป็นการเชิงเส้นในพารามิเตอร์ ซึ่งหมายความว่า ในแต่ละเทอมของตัวแบบจะมีพารามิเตอร์เพียง 1 ตัว คูณด้วยค่าคงที่ของตัวแปรอิสระ และตัวแปรอิสระที่มีอิทธิพลต่อการเปลี่ยนแปลงของ Y จะกำหนดให้เป็นตัวแปรที่ถูกควบคุมให้คงที่

เนื่องจากตัวแบบที่ใช้ในการพิจารณาเป็นตัวแบบเชิงเส้น ดังนั้นในการประมาณค่าพารามิเตอร์เราสามารถใช้วิธีการกำลังสองน้อยสุดและวิธีการภาวะน่าจะเป็นสูงสุด ในการประมาณค่าของพารามิเตอร์สัมประสิทธิ์ความถดถอย $\underset{\sim}{\beta}$ สำหรับตัวแบบความถดถอยเชิงเส้นข้างต้นมีรูปแบบของตัวประมาณ ดังนี้

ตัวประมาณค่ากำลังสองน้อยสุด(Least Square Estimator: $\underset{\sim}{\mathbf{b}}$) คือ

$$\underset{\sim}{b} = (X'X)^{-1}X'\underset{\sim}{Y}$$

สำหรับเกณฑ์การคัดเลือกตัวแบบโดยสถิติทดสอบเอฟบางส่วน(The partial F-test statistic) ด้วยวิธีการกำจัดตัวแปรแบบถอยหลัง(Backward Elimination(BW))

ตัวประมาณค่าภาวะน่าจะเป็นสูงสุด(Maximum Likelihood Estimator: $\mathbf{b}_{\sim M}$) คือ

$$b_{\sim M} = (X'X)^{-1}X'\underset{\sim}{Y}$$

สำหรับเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Akaike's Information Criterion (AIC)) และ เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์(Bayesian Information Criterion (BIC))

การเพิ่มระดับขั้นของ p' จะเพิ่มจาก 1 จนถึง n-1 โดยระดับขั้นของพหุนามในแต่ละพจน์เกิดจากการรวมกันในพจน์ของตัวแปรอิสระ กำลังในแต่ละพจน์ของพหุนามจะกำหนดภายใต้พจน์ที่มีกำลังสูงสุด เช่นตัวแบบที่มีกำลังสูงสุดเป็น 2 จะได้ตัวแบบที่สมบูรณ์ดังสมการที่(2.1) (หน้า 12) ซึ่งจะสังเกตได้ว่าในแต่ละพจน์มีกำลังไม่เกิน 2

ในกรณีที่ค่าของ x มีค่าเพิ่มสูงขึ้น ดังนั้น ถ้า x มีค่ามาก การหาเมทริกซ์ผกผันของ $(X'X)$จะเกิดความคลาดเคลื่อนสูง เช่น x จะมีค่ามากเมื่อ $x > 1$ และจะมีค่าน้อยเมื่อ $0 < x < 1$ ซึ่งการแก้ปัญหาในการวิจัยในครั้งนี้ ได้ใช้วิธีการพหุนามเชิงตั้งฉาก(Orthogonal Polynomial) และการแปลงเมทริกซ์ X เข้าสู่ศูนย์กลาง ซึ่งจะกล่าวถึงในหัวข้อ 2.2

2.2 **การแก้ไขปัญหาการเกิดความสัมพันธ์ร่วมเชิงพหุ**(Multicollinearity)

เนื่องจากตัวแบบความถดถอยพหุนามเมทริกซ์ X จะเกิดความความสัมพันธ์ร่วมเชิงพหุ และ $\det(X'X) = 0$ ดังนั้นจึงจำเป็นที่จะต้องปรับปรุงแก้ไขด้วยวิธีการพหุนามเชิงตั้งฉาก ซึ่งมีรายละเอียดดังนี้

2.2.1 **วิธีการพหุนามเชิงตั้งฉาก**(Orthogonal Polynomial)[1]

การใช้วิธีการพหุนามเชิงตั้งฉาก(Orthogonal Polynomial) เป็นวิธีการที่ทำให้$(X'X)$เป็นเมทริกซ์เส้นทแยงมุม (Diagonal Matrix) ซึ่งจะทำให้การหา $(X'X)^{-1}$ ทำได้ง่ายขึ้น นอกจากนี้วิธีการพหุนามเชิงตั้งฉาก(Orthogonal Polynomial) สามารถช่วยในการแก้ไขปัญหาการเกิดความสัมพันธ์ร่วมเชิงพหุ (Multicollinearity) ระหว่าง $x_{i2}, x_{i2}^2, \ldots$ เป็นต้น ดังนั้น ในการพิจารณาพหุนามเชิงตั้งฉาก(Orthogonal Polynomial) อาจเขียนพหุนามเชิงตั้งฉากในรูปของ

$$y_i = \gamma_0\psi_0(x_i) + \gamma_1\psi_1(x_i) + \ldots\ldots\ldots + \gamma_{p'}\psi_{p'}(x_i) + \varepsilon_i \quad ; i = 1,2,\ldots\ldots\ldots\ldots,n$$

เมื่อ $\psi_j(x_i)$ เป็นพหุนามระดับขั้นที่ j ใน x_i (j=0,1,2,.......,p)

และ $\psi_j(x_i)$ เป็นพหุนามเชิงตั้งฉาก กล่าวคือ $\sum_{i=1}^{n}\psi_j(x_i)\psi_k(x_i) = 0 \quad (\forall_j, k, j \neq k)$

ซึ่งสามารถเขียนสมการในตัวแบบเชิงเส้นได้ดังนี้

$$\underset{\sim}{Y} = X\underset{\sim}{\gamma} + \underset{\sim}{\varepsilon}$$

---

[1] ธีระพร วีระถาวร. <u>ตัวแบบเชิงเส้น ทฤษฎีและการประยุกต์</u>. กรุงเทพมหานคร: บริษัท พิมพ์ดี จำกัด,2541 หน้า 269-275.

เมื่อ
$$X = \begin{bmatrix} \psi_0(x_1) & \psi_1(x_1) & \cdots & \psi_{p'}(x_1) \\ \psi_0(x_2) & \psi_1(x_2) & \cdots & \psi_{p'}(x_2) \\ \vdots & \vdots & & \vdots \\ \psi_0(x_n) & \psi_1(x_n) & \cdots & \psi_{p'}(x_n) \end{bmatrix}$$

$$X' X = \begin{bmatrix} \sum_i \psi_0^2(x_i) & 0 & \cdots & 0 \\ 0 & \sum_i \psi_1^2(x_i) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sum_i \psi_k^2(x_i) \end{bmatrix}$$

ในกรณีที่เมทริกซ์ X มีค่าลำดับชั้น p' จะได้ว่า

$$\hat{\underset{\sim}{\gamma}} = (X' X)^{-1} X' \underset{\sim}{Y} \qquad \text{ซึ่ง} \quad \hat{\gamma}_j = \frac{\sum_{i=1}^{n} \psi_j(x_i)(y_i)}{\sum_{i=1}^{n} \psi_j^2(x_i)} \quad ; \ j = 0,1,2,\ldots\ldots\ldots,p'$$

การวิจัยครั้งนี้จะกำหนดให้เมทริกซ์ X เป็นผลกระทบคงที่ และค่าของ x มีปริภูมิห่างเท่ากัน (Equally Spaced x - Values)

สมมติว่าค่าของ $x_i = i - \frac{1}{2}(n+1) \qquad i = 1,2,\ldots\ldots\ldots,n$

เช่น $n = 4$ จะได้ว่า $x_1 = \frac{-3}{2}, x_2 = \frac{-1}{2}, x_3 = \frac{1}{2}, x_4 = \frac{3}{2}$

$n = 5$ จะได้ว่า $x_1 = -2, x_2 = -1, x_3 = 0, x_4 = 1, x_5 = 2$

$n = 6$ จะได้ว่า $x_1 = \frac{-5}{2}, x_2 = \frac{-3}{2}, x_3 = \frac{-1}{2}, x_4 = \frac{1}{2}, x_5 = \frac{3}{2},$

$x_6 = \frac{5}{2}$

2.2.2 การแปลงเมทริกซ์ X เพื่อให้เป็นแนวตั้งเชิงตั้งฉาก ด้วยกระบวนการของกราม-ชมิดต์(Gram-Schmidt Orthogonalization Process)

ในการประมาณอิทธิพลบางตัวแปรในตัวแบบเชิงเส้นเราจำเป็นต้องตรวจสอบค่าลำดับชั้นของ X ถ้า X มีค่าลำดับชั้นเท่ากับจำนวนแนวตั้งของมันก็จะทำให้การหาค่าประมาณอิทธิพลของตัวแปรอิสระง่ายขึ้นเพราะทำให้สามารถหา $(X'X)^{-1}$ ได้ ซึ่งในวิทยานิพนธ์เล่มนี้ได้ใช้กระบวนการของกราม-ชมิดต์รวมทั้งการสร้างฐานหลักเชิงตั้งฉากปรกติ(orthonormal basis) สำหรับ n มิติ ซึ่งเป็นผลคูณภายในปริภูมิ V ถ้าเราเริ่มจากการพิจารณาฐานหลัก(basis) $x_1,\dots\dots,x_n$ โดยมีกระบวนการดังนี้

เริ่มจากการกำหนดให้ $$\underset{\sim 1}{u} = \left(\frac{1}{\|\underset{\sim}{x}\|}\right)\underset{\sim 1}{x}$$

โดย $\underset{\sim 1}{u}$ เป็นเวกเตอร์หนึ่งหน่วยในทิศทางของ $\underset{\sim 1}{x}$

และกำหนด $\underset{\sim 2}{u},\dots,\underset{\sim k}{u}$ โดย

$$\underset{\sim k+1}{u} = \left(\frac{1}{\|\underset{\sim k+1}{x} - \underset{\sim k}{p}\|}\right)(\underset{\sim k+1}{x} - \underset{\sim k}{p})$$ สำหรับ k=1,…,n-1

ให้ $\underset{\sim k}{p}$ แทนภาพฉายเชิงตั้งฉากของ $\underset{\sim k+1}{x}$

$$\underset{\sim k}{p} = \langle \underset{\sim k+1}{x}, \underset{\sim 1}{u} \rangle \underset{\sim 1}{u} + \langle \underset{\sim k+1}{x}, \underset{\sim 2}{u} \rangle \underset{\sim 2}{u} + \dots + \langle \underset{\sim k+1}{x}, \underset{\sim k}{u} \rangle \underset{\sim k}{u}$$

ดังนั้น $\underset{\sim k+1}{u}$ เป็นเวกเตอร์หนึ่งหน่วยที่ตั้งฉากกับ $\underset{\sim 1}{u}$ และ เป็นอิสระเชิงเส้น(linearly independence) ทั้งนี้ $(\underset{\sim 1}{u},\underset{\sim 2}{u},\dots,\underset{\sim k}{u})$ เป็นฐานหลักเชิงตั้งฉากปรกติ(orthonormal basis) ด้วย

2.2.3 การแปลงเมทริกซ์ X เข้าสู่ศูนย์กลางและการปรับมาตรา(Centering and Scaling)

การแปลงเมทริกซ์เข้าสู่ศูนย์กลาง เป็นวิธีการหนึ่งในการแก้ไขปัญหาการเกิดความสัมพันธ์ร่วมเชิงพหุของตัวแปรอิสระ ซึ่งเป็นวิธีที่ใช้เมทริกซ์สหสัมพันธ์ มาแทนที่เมทริกซ์ $X'X$ โดยพิจารณาจากการกำหนดให้ข้อมูลมีรูปแบบดังตารางนั่นคือ

x_0	x_1	x_2	…	$x_{p'}$	Y
1	x_{11}	x_{21}	…	$x_{p',1}$	y_1
1	x_{12}	x_{22}	…	$x_{p',2}$	y_2
…					
1	x_{1n}	x_{2n}	…	$x_{p',n}$	y_n
ผลบวกของคอลัมน์	$\sum x_{1i}$	$\sum x_{2i}$	…	$\sum x_{p',i}$	$\sum y_i$
ค่าเฉลี่ยของคอลัมน์	$\overline{x}_1$	$\overline{x}_2$	…	$\overline{x}_{p'}$	$\overline{y}$

จากตัวแบบ

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{p'} x_{p'} + \varepsilon$$

สามารถเขียนในรูปแบบของ

$$y = \{\beta_0 + \beta_1 \overline{x}_1 + \beta_2 \overline{x}_2 + \ldots + \beta_{p'} \overline{x}_{p'}\} + \beta_1 (x_1 - \overline{x}_1) + \beta_2 (x_2 - \overline{x}_2) + \ldots + \beta_{p'} (x_{p'} - \overline{x}_{p'}) + \varepsilon$$

เมื่อ $\overline{y}, \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{p'}$ เป็นค่าจำนวนจริงที่ได้มาจากข้อมูล ถ้าเราเขียน $z_j = x_j - \overline{x}_j$ $j=1,2,3,\ldots,p'$, $\beta_0' = \{\beta_0 + \beta_1 \overline{x}_1 + \beta_2 \overline{x}_2 + \ldots + \beta_{p'} \overline{x}_{p'}\}$ ดังนั้นสามารถเขียนตัวแบบได้ว่า

$$y = \beta_0' + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{p'} z_{p'} + \varepsilon$$

เราสามารถแสดงการแปลงข้อมูลโดยการพิจารณาจากตัวแปรด้านบน เมื่อ $z_j = x_j - \overline{x}_j$ $j=1,2,3,\ldots,p'$ สำหรับ $i=1,2,\ldots,n$ β_0' สามารถทำให้อยู่ในรูปแบบอย่างง่ายคือ

$$b_0' + b_1 \overline{z}_1 + b_2 \overline{z}_2 + \ldots + b_{p'} \overline{z}_{p'} = \overline{y}$$

ซึ่งหมายถึง

$$b_0' = \overline{y}$$

จากค่ากลางข้างต้นเราจะนำมาใช้ในการหาเมทริกซ์สหสัมพันธ์ของตัวแปร กรณีที่เมทริกซ์มีขนาดใหญ่เช่น 5x5 จะทำให้เกิดความคลาดเคลื่อนสูง ดังนั้นเราจะทำการแปลงค่ากลางโดย

$$z_{ji} = \frac{(x_{ji} - \bar{x}_j)}{S_{jj}^{1/2}}$$

$$y_i = \frac{(y_i - \bar{y}\ )}{S_{yy}^{1/2}}$$

เมื่อ $S_{yy} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$

$$S_{jj} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

โดย $X = \begin{bmatrix} z_{11} & z_{12} & . & . & . & z_{1,p'} \\ z_{21} & z_{22} & . & . & . & z_{2,p'} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ z_{n1} & z_{n2} & . & . & . & z_{n,p'} \end{bmatrix}$

$$X' X = \begin{bmatrix} 1 & r_{12} & . & . & . & r_{1,p'} \\ r_{21} & 1 & . & . & . & r_{2,p'} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ r_{p'-1,1} & r_{p'-1,2} & . & . & . & 1 \end{bmatrix}$$

ดังนั้น ตัวแบบจะอยู่ในรูปของ $yS_{yy}^{1/2} = b_1 S_{11}^{1/2} z_1 + ... + b_{p'} S_{p'p'}^{1/2} z_{p'} + \varepsilon$

$$b_j = \left(\frac{S_{jj}}{S_{yy}}\right) b_j' \qquad j=1,2,\ldots,p'$$

2.3 เกณฑ์ที่นำมาใช้ในการคัดเลือกตัวแบบ

เกณฑ์ที่นำมาใช้ในการคัดเลือกตัวแบบความถดถอยพหุนามแบบติดกลุ่ม ประกอบด้วย 3 เกณฑ์ นั่นคือ

2.3.1 เกณฑ์การคัดเลือกตัวแบบโดยสถิติทดสอบเอฟบางส่วน(The partial F-test statistic) ด้วยวิธีการกำจัดตัวแปรแบบถอยหลัง(Backward Elimination(BW))

2.3.2 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Akaike's Information Criterion(AIC))

2.3.3 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์ (Bayesian Information Criterion(BIC))

โดยมีรายละเอียดดังนี้

2.3.1.เกณฑ์การคัดเลือกตัวแบบโดยสถิติทดสอบเอฟบางส่วน (The partial F-test statistic)ด้วยวิธีการกำจัดตัวแปรแบบถอยหลัง(Backward Elimination(BW)) [2]

วิธีการนี้เริ่มต้นด้วยสมการความถดถอยด้วยตัวแบบที่สมบูรณ์ คือ ประกอบด้วยตัวแปรอิสระทุกตัวที่ใช้พิจารณา แล้วจะคัดตัวแปรอิสระออกจากสมการครั้งละ 1 ตัวแปร โดยเลือกตัวแปรที่มีความสัมพันธ์กับ y น้อยที่สุด (เมื่อตัวแปรอื่น ๆ คงที่) และค่าสัมประสิทธิ์ความถดถอยของตัวแปรนั้นไม่มีนัยสำคัญ จากนั้นคำนวณหาสมการความถดถอยสำหรับตัวแปรอิสระที่เหลือ และคัดตัวแปรอิสระที่มีนัยสำคัญน้อยที่สุดออก จนกระทั่งไม่มีตัวแปรอิสระใดถูกคัดออกแล้วตัวแปรอิสระที่เหลืออยู่ จะอยู่ในสมการความถดถอยทุกตัวเป็นอันสิ้นสุดของวิธีการนี้ สามารถแสดงขั้นตอนได้ดังนี้

1) สร้างสมการความถดถอยตัวแบบสมบูรณ์ (Complete Model) ที่ประกอบด้วยตัวแปรอิสระทุกตัว

2) คำนวณหาค่าเอฟบางส่วน(Partial F) ของตัวแปรอิสระทุกตัว เสมือนว่าตัวแปรอิสระนั้นเข้าสู่สมการเป็นตัวสุดท้าย

3) เลือกตัวแปรอิสระที่มีค่าเอฟบางส่วน(Partial F) น้อยที่สุด (F_L) นำค่า (F_L) เปรียบเทียบกับค่า F ณ ระดับนัยสำคัญที่กำหนด ถ้าพบว่า $F_L < F_0$ จะตัดตัวแปรอิสระที่มีค่าเอฟบางส่วน (Partial F) น้อยที่สุดนั้นออกจากสมการและคำนวณหาค่าสัมประสิทธิ์ของสมการความถดถอยเมื่อตัดตัวแปรอิสระนั้นออกแล้ว จากนั้นกลับไปทำในขั้นตอนที่ 2 ถ้า $F_L > F_0$ จะหยุดกระบวนการคัดเลือกตัวแปรอิสระ และได้สมการถดถอยที่เหมาะสม

---

[2] นพมาศ อัครจันทรโชติ."การเปรียบเทียบวิธีการสร้างตัวแบบในการวิเคราะห์ความถดถอยพหุนามกรณีที่มี 2 ตัวแปรอิสระซึ่งเกิดอันตรกิริยา".วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.

2.3.2 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Akaike's Information Criterion(AIC))

ในปี ค.ศ. 1973 อาไคเคะได้เสนอเกณฑ์ในการคัดเลือกตัวแบบเพื่อใช้เป็นเครื่องมือในการหาตัวแบบที่ให้ค่าพยากรณ์แม่นยำที่สุด นั่นคือ เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ(Akaike's Information Criterion(AIC)) เป็นเกณฑ์ที่พิจารณาจากการประมาณความคลาดเคลื่อนรวมเข้ากับข้อสนเทศ(Information)ของค่าสังเกต และใช้แนวคิดจากค่าต่ำสุดของข้อสนเทศของคูลส์แบล็ค - ไลท์เบอร์ (Kulback - Leibler(1951)) เพื่อนำมาใช้ในการปรับค่าประมาณของการพยากรณ์ให้มีความแม่นยำมากขึ้น โดยมีข้อกำหนดเบื้องต้น คือ ตัวประมาณได้มาจากวิธีการประมาณค่าภาวะน่าจะเป็นสูงสุด

เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ(Akaike's Information Criterion (AIC)) คือ

$$AIC = -2\sum_{i=1}^{n} \log f(X_i \mid \hat{\underset{\sim}{\beta}}_M) + 2p$$

หรือ

$$AIC_k = -2\log(ML_k) + 2p_k \tag{2.2}$$

เมื่อ ML_k คือ ฟังก์ชันภาวะน่าจะเป็นสูงสุด(Maximum likelihood Function) ของตัวแบบที่ k

p_k คือ จำนวนพารามิเตอร์ของตัวแบบที่ k

ซึ่งตัวแบบที่ให้ค่า AIC ต่ำสุดจะเป็นตัวแบบที่ให้ค่าพยากรณ์แม่นยำและถูกต้อง

2.3.2.1 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Akaike's Information Criterion(AIC))สำหรับการแจกแจงแบบปกติ

ตัวแบบและฟังก์ชันภาวะน่าจะเป็นที่นำมาใช้ในการพิจารณามีรูปแบบดังสมการ (1.1) (หน้า 4)ซึ่งจะได้ฟังก์ชันภาวะน่าจะเป็นดังนี้

$$ML = L(\underset{\sim}{\beta}_M, \sigma^2; \underset{\sim}{Y}) = f(y_1 \mid \beta_1, \sigma^2) \times f(y_2 \mid \beta_2, \sigma^2) \times \ldots\ldots \times f(y_n \mid \beta_n, \sigma^2)$$

$$= f(\underset{\sim}{Y} \mid \underset{\sim}{\beta}_M, \sigma^2 I_n)$$

$$= (2\pi\sigma^2)^{\frac{-n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim}{\beta}_M)^T(\underset{\sim}{Y} - X\underset{\sim}{\beta}_M)\right\}$$

$$\therefore \log L(\underset{\sim}{\beta}_M, \sigma^2; \underset{\sim}{Y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim}{\beta}_M)^T(\underset{\sim}{Y} - X\underset{\sim}{\beta}_M)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim M}{\beta})$$
$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{n}{2}$$
$$= -\frac{n}{2}(\log(2\pi) + 1) - \frac{n}{2}\log(\hat{\sigma}^2)$$

ซึ่ง $\underset{\sim M}{\beta} = (X'X)^{-1}X'\underset{\sim}{Y}$

และ $\hat{\sigma}^2 = \frac{1}{n}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim M}{\beta})$

ดังนั้น AIC สำหรับการแจกแจงแบบปกติ คือ

$$\begin{aligned}
AIC &= -2\log L(\underset{\sim M}{\beta}, \sigma^2; \underset{\sim}{Y}) + 2p \\
&= -2\left(-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim M}{\beta})\right) + 2p \\
&= -2\left(-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim M}{\beta})\right) + 2p \\
&= -2\left(-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{n}{2}\right) + 2p \\
&= -2\left(-\frac{n}{2}(\log(2\pi) + 1) + \frac{n}{2}\log(\hat{\sigma}^2)\right) + 2p \\
&= n(\log(2\pi) + 1) + n\log(\hat{\sigma}^2) + 2p
\end{aligned}$$

2.3.2.2 AIC **สำหรับการเปรียบเทียบตัวแบบ**

การเปรียบเทียบตัวแบบเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ(Akaike's Information Criterion) จะอาศัยหลักการทดสอบด้วยอัตราส่วนภาวะน่าจะเป็น(Likelihood Ratio Test) และลำดับของการเปรียบเทียบตัวแบบจะอาศัยหลักการของออกส์แคม เรเซอร์ (occam's razor) นั่นคือเมื่อนำตัวแบบ 2 ตัวแบบมาแข่งขันกัน ตัวแบบที่มีรูปแบบอย่างง่ายจะมีแนวโน้มใน การยอมรับมาก และเมื่อตัวแบบอย่างง่ายถูกปฏิเสธ ตัวแบบที่ติดกลุ่มกับตัวแบบอย่างง่ายจะถูกปฏิเสธไปด้วย ซึ่งมีขั้นตอนการทดสอบดังนี้

(1) กำหนดให้ตัวแบบ

$$M_1 : \underset{\sim}{Y} = X\underset{\sim M}{\beta} + \underset{\sim}{\varepsilon} \quad \text{เมื่อ } \underset{\sim}{\varepsilon} \sim N_n(0,\sigma_0^2 I_n)$$

$$M_2 : \underset{\sim}{Y} = X\underset{\sim M}{\beta} + \underset{\sim}{\varepsilon} \quad \text{เมื่อ } \underset{\sim}{\varepsilon} \sim N_n(0,\sigma_1^2 I_n)$$

โดยตัวแบบ $M_1 \subset M_2$ และทำการทดสอบสมมติฐานในรูปแบบของการทดสอบสมมติฐานแบบติดกลุ่มนั่นคือ

$$H_0 : M_1 : \underset{\sim}{Y} = X\underset{\sim M}{\beta} + \underset{\sim}{\varepsilon}$$

$$H_1 : M_2 : \underset{\sim}{Y} = X\underset{\sim M}{\beta} + \underset{\sim}{\varepsilon} \quad \text{เมื่อ } H_0 \subset H_1$$

(2)คำนวณค่า AIC ของแต่ละตัวแบบที่นำมาเปรียบเทียบจากสูตรในสมการที่ (2.2)

(3)เปรียบเทียบค่า AIC โดยใช้หลักการของการทดสอบอัตราส่วนภาวะน่าจะเป็น(Likelihood Ratio Test) จากสูตร

$$AIC = -2\log\left(\frac{L(\underset{\sim M}{\beta},\sigma_1^2 : \underset{\sim}{Y})}{L(\underset{\sim M}{\beta},\sigma_0^2 : \underset{\sim}{Y})}\right) + 2(p_1 - p_0)$$

หรือ $AIC_1 - AIC_0 = -2\log(ML_1) + 2p_1 + 2\log(ML_0) - 2p_0$

โดย AIC_1 แทน AIC สำหรับตัวแบบ H_1, AIC_0 = AIC สำหรับตัวแบบ H_0

$$= -2\log\left(\frac{ML_1}{ML_0}\right) + 2(p_1 - p_0)$$

โดย $-2\log\left(\frac{ML_1}{ML_0}\right)$ มีการแจกแจงเป็นแบบไคสแควร์ $\chi^2_{df=(p_1-p_0)}$

(4)เปรียบเทียบค่า AIC กับค่า $\chi^2_{df=(p_1-p_0)}$ จากตารางที่กำหนดระดับนัยสำคัญ α โดยมีองศาความเป็นอิสระเท่ากับ $(p_1 - p_0)$

การตัดสินใจ $AIC_{10} < \chi^2_{\alpha, df=(p_1-p_0)}$ ยอมรับ H_0

$AIC_{10} \geq \chi^2_{\alpha, df=(p_1-p_0)}$ ยอมรับ H_1

โดย AIC_{10} แทน AIC สำหรับตัวแบบ H_1 เทียบกับตัวแบบ H_0

(5)พิจารณาตามขั้นตอนที่ (1) - (4) จนกระทั่งครบทุกตัวแบบ

(6)นำตัวแบบที่มีระดับนัยสำคัญทางสถิติมาพิจารณาหาตัวแบบ ซึ่งตัวแบบที่ให้ค่า AIC ต่ำสุดจะเป็นตัวแบบที่ดีที่สุด

2.3.3 **เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์** (Bayesian Information Criterion (BIC))

ในปี ค.ศ. 1978 ชวาร์ซได้เสนอเกณฑ์ในการคัดเลือกตัวแบบโดยข้อสนเทศ นั่นคือ เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของชวาร์ซ (Schwarz's Information Criterion (SIC, BIC, SBC)) ซึ่งเป็นเกณฑ์ที่รู้จักกันอย่างกว้างขวางและใช้เป็นเครื่องมือในการคัดเลือกตัวแบบทางด้านสถิติ

เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของชวาร์ซ สามารถนำไปพัฒนาโดยใช้แนวคิดของเบส์ หรือเรียกว่าเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์ (Bayesian Information Criterion(BIC))เพื่อนำมาใช้ในการแก้ไขปัญหาการคัดเลือกตัวแบบ โดยกำหนดให้ $p(M_k)_{(1\le k\le L)}$ แทน ความหนาแน่นก่อน(Prior Density) บนตัวแบบ $M_1, M_2, \ldots, M_L$ และกำหนดให้ $p(\underset{\sim}{\beta}_k \mid M_k)$ แทน ความน่าจะเป็นก่อนบน $\underset{\sim}{\beta}_k$ ภายใต้ตัวแบบ M_k การประยุกต์ใช้ทฤษฎีของเบส์สำหรับความหนาแน่นร่วมก่อนของ M_k บน $\underset{\sim}{\beta}_k$ สามารถเขียนได้ดังนี้

$$p(M_k, \underset{\sim}{\beta}_k \mid X) = \frac{p(M_k)p(\underset{\sim}{\beta}_k \mid M_k)p(X \mid M_k, \underset{\sim}{\beta}_k)}{p(X)}$$

เมื่อ $p(X)$ แทน การแจกแจงขอบของ X

$p(X \mid M_k, \underset{\sim}{\beta}_k)$ แทน ความน่าจะเป็นของ X เมื่อกำหนด β_k หรือ ภาวะน่าจะเป็นของ β_k

งานวิจัยครั้งนี้ จะทำการประมาณเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์ Bayesian Information Criterion (BIC) ด้วยตัวประกอบของเบส์(Bayes Factor) เสนอโดย โรเบิร์ต อี เคสส์ (Robert E. Kass) และ เอเดร์น อี ราฟร์เทอรี (Adrian E. Raftery) ซึ่งเป็นการพิจารณาอัตราส่วนภาวะน่าจะเป็นของการแจกแจงภายหลัง ซึ่งการคัดเลือกตัวแบบด้วยแนวคิดของเบส์มีรูปแบบของสมการดังนี้

$$\mathrm{BIC} = -2\log(\text{Bayes Factor})$$

โดยตัวประกอบของเบส์ (Bayes Factor) คือ อัตราส่วนออดส์ของความน่าจะเป็นภายหลังที่ใช้วัดความชัดเจนในการยอมรับ H_0 กับสมมติฐานทางเลือกที่นำมาทำการทดสอบ สำหรับรายละเอียดจะกล่าวถึงในหัวข้อ 2.6

2.3.3.1 **เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์** (Bayesian Information Criterion (BIC))**สำหรับการวิเคราะห์ความถดถอยพหุนามแบบติดกลุ่ม**

2.3.3.1.1 **ตัวแบบและฟังก์ชันภาวะน่าจะเป็น**

กำหนดให้ $\underset{\sim}{Y} = X\underset{\sim M}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\varepsilon} \sim N_n(0,\sigma^2 I_n)$ และ $\underset{\sim}{Y}$ แทน เวกเตอร์ของตัวแปรตามขนาด n x 1 X แทน เมทริกซ์ของค่าสังเกตขนาด n x (p′+1) และ $\underset{\sim M}{\beta}$ แทน เวกเตอร์พารามิเตอร์ขนาด (p′+1) x 1 ซึ่งมีฟังก์ชันภาวะน่าจะเป็นดังนี้

$$ML = L(\underset{\sim M}{\beta},\sigma^2;\underset{\sim}{Y}) = f(y_1 \mid \beta_1,\sigma^2)\times f(y_2 \mid \beta_2,\sigma^2)\times \ldots\ldots\ldots \times f(y_n \mid \beta_n,\sigma^2)$$

$$= f(\underset{\sim}{Y} \mid \underset{\sim M}{\beta},\sigma^2 I_n)$$

$$= (2\pi\sigma^2)^{\frac{-n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim M}{\beta})\right\}$$

ซึ่ง $\underset{\sim M}{\beta} = (X'X)^{-1}X'\underset{\sim}{Y}$

และ $\hat{\sigma}^2 = \frac{1}{n}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim M}{\beta})$

2.3.3.1.2 **การแจกแจงก่อนและการแจกแจงภายหลังขอบ**

การแจกแจงก่อน(Prior Distribution)

การแจกแจงก่อน(Prior Distribution) ของพารามิเตอร์ได้ใช้แนวคิดของ Maximal Data Informative Prior (Zellner(1971)) ซึ่งเป็นแบบไม่มีข้อสนเทศ(Noninformative Prior) กล่าวคือ เราไม่ทราบลักษณะพื้นฐานของข้อมูล และข้อมูลมีลักษณะคลุมเครือ ไม่ชัดเจนซึ่งเหตุผลในการใช้ Non–Informative Prior Distribution เนื่องจาก

1.เป็นการใช้ข้อมูลที่มีอยู่อธิบายลักษณะของข้อมูลเอง

2.บ่อยครั้งเราไม่สามารถขจัดการแจกแจงก่อน ที่เกิดจากความเชื่อของแต่ละบุคคลหรือที่เรียกว่า จิตวิสัย(Subjective Prior)ได้ เนื่องจากข้อจำกัดของเวลาและค่าใช้จ่าย หรือเนื่องจากความไม่ชำนาญของผู้กำหนด ดังนั้น การแจกแจงก่อนของ $\underset{\sim}{\beta}$ และ $\log(\sigma)$ จะเป็นการแจกแจงแบบสม่ำเสมอ(Uniform Distribution) และมีความเป็นอิสระซึ่งกันและกัน โดยมีรูปแบบดังนี้

$$p(\underset{\sim}{\beta},\sigma) \propto \frac{1}{\sigma} \qquad ;-\infty < \underset{\sim}{\beta} < \infty, 0 < \sigma < \infty$$

การแจกแจงภายหลังขอบ

ฟังก์ชันความหนาแน่นน่าจะเป็นร่วมภายหลังสำหรับตัวแบบ คือ

$$p(X \mid M) \propto \text{likelihood} \times \text{prior}$$
$$p(X \mid M) \propto l(\underset{\sim M}{\beta}, \sigma \mid \underset{\sim}{Y}) \times p(\underset{\sim M}{\beta}, \sigma)$$
$$p(X \mid M) \propto \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})'(\underset{\sim}{Y} - X\underset{\sim M}{\beta})\right]$$
$$p(X \mid M) \propto \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}[\nu s^2 + (\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}})'X'X(\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}})]\right]$$

โดย $\underset{\sim M}{\hat{\beta}} = (X'X)^{-1}X'\underset{\sim}{Y}$ และ $s^2 = \dfrac{(\underset{\sim}{Y} - X\underset{\sim M}{\beta})'(\underset{\sim}{Y} - X\underset{\sim M}{\beta})}{\nu}$

M = ตัวแบบที่นำมาพิจารณา

เมื่อ $\nu = n - p$

ν แทน ระดับขั้นความเสรี

n แทน ขนาดตัวอย่าง

p แทน จำนวนพารามิเตอร์ในตัวแบบ

การแจกแจงภายหลังขอบของตัวแบบความถดถอย

$$p(X \mid M) \propto \int_0^\infty \text{likelihood} \times \text{prior}\, d\sigma$$
$$\propto \int_0^\infty l(\beta \mid_{\sim M} \sigma, M) \times p(\underset{\sim M}{\beta}, \sigma) d\sigma$$
$$\propto \int_0^\infty \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})'(\underset{\sim}{Y} - X\underset{\sim M}{\beta})\right] \bullet \left(\frac{1}{\sigma}\right) d\sigma$$
$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\beta})'(\underset{\sim}{Y} - X\underset{\sim M}{\beta})\right] d\sigma$$
$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\hat{\beta}} - X(\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}}))'(\underset{\sim}{Y} - X\underset{\sim M}{\hat{\beta}} - X(\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}}))\right] d\sigma$$
$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}(\underset{\sim}{Y} - X\underset{\sim M}{\hat{\beta}})'(\underset{\sim}{Y} - X\underset{\sim M}{\hat{\beta}}) + (\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}})'X'X(\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}}))\right] d\sigma$$

ส่วนพจน์ $(\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}})'X'(\underset{\sim}{Y} - X\underset{\sim M}{\hat{\beta}}) = (\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}})'[X'\underset{\sim}{Y} - X'X(X'X)^{-1}X'\underset{\sim}{Y}] = 0$

$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}(\nu s^2 + (\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}})'X'X(\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}}))\right] d\sigma$$

ใช้เทคนิคการอินทิเกรตของแกมมาจะได้ว่า

$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp\left[-\frac{a}{2\sigma^2}\right] d\sigma \qquad ; a = (\nu s^2 + (\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}})' X' X (\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}}))$$

$$\propto \int_0^\infty \frac{1}{\sigma^{n+1}} \exp[-x] d\sigma \qquad ; x = \frac{a}{2\sigma^2}$$

$$\propto \int_0^\infty \left(\frac{a}{2x}\right)^{-\frac{1}{2}(n+1)} \exp[-x] \left(\frac{a}{2}\right)^{\frac{1}{2}} x^{-\frac{3}{2}} dx$$

$$\propto \int_0^\infty \left(\frac{a}{2}\right)^{-\frac{1}{2}(n+1)} \bullet x^{\frac{n}{2}+\frac{1}{2}} \exp[-x] \left(\frac{a}{2}\right)^{\frac{1}{2}} \frac{x^{-\frac{3}{2}}}{2} dx$$

$$\propto \int_0^\infty \left(\frac{a}{2}\right)^{-\frac{n}{2}} \bullet x^{\frac{n}{2}-1} \exp[-x] \bullet \frac{1}{2} \, dx$$

$$\propto \frac{2^{\left(\frac{n-2}{2}\right)}}{a^{\frac{n}{2}}} \int_0^\infty x^{\left(\frac{n-2}{2}\right)} e^{-x} dx$$

$$\propto \frac{2^{\left(\frac{n-2}{2}\right)} \Gamma\left(\frac{n}{2}\right)}{a^{\frac{n}{2}}}$$

$$\propto 2^{\left(\frac{n-2}{2}\right)} \Gamma\left(\frac{n}{2}\right) (\nu s^2 + (\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}})' X' X (\underset{\sim M}{\beta} - \underset{\sim M}{\hat{\beta}}))^{-\frac{n}{2}}$$

ดังนั้น BIC สำหรับการวิเคราะห์ความถดถอยคือ

$$\begin{aligned} \mathrm{BIC} &= -2\log(\text{Bayes Factor}) \\ &= -2\log\left(\frac{p(X \mid M_0)}{p(X \mid M_1)}\right) \\ &= -2\log\left(\frac{(\nu_0 s_0{}^2 + (\underset{\sim 0M}{\beta} - \underset{\sim 0M}{\hat{\beta}})' X_0' X_0 (\underset{\sim 0M}{\beta} - \underset{\sim 0M}{\hat{\beta}}))}{(\nu_1 s_1{}^2 + (\underset{\sim 1M}{\beta} - \underset{\sim 1M}{\hat{\beta}})' X_1' X_1 (\underset{\sim 1M}{\beta} - \underset{\sim 1M}{\hat{\beta}}))}\right)^{-\frac{n}{2}} \end{aligned}$$

2.4 **การทดสอบสมมติฐานแบบติดกลุ่ม**(Nested Hypothesis)[3]

สิ่งที่จำเป็นในการสร้างตัวแบบ นั่นคือเราต้องการหาวิธีการทางด้านสถิติในการกำหนดตัว

---

[3] Terry Sincich. "Business Statistics By Example(Fifth Edition)" .Prentice Hall International Editions. pp.728-733

แบบที่เหมาะสมที่สุดให้กับข้อมูลที่ได้จากการสำรวจ ซึ่งการวิจัยครั้งนี้ได้ใช้การทดสอบสมมติฐานแบบติดกลุ่ม(Nested Hypothesis) ในการทดสอบตัวแบบติดกลุ่ม (Nested Model)

นิยามที่ 1 ตัวแบบ 2 ตัวแบบจะติดกัน ถ้าในแต่ละพจน์ของตัวแบบแรกเป็นส่วนหนึ่งของตัวแบบที่สอง ซึ่งตัวแบบที่สองจะมีพจน์มากกว่าตัวแบบแรกอย่างน้อย 1 พจน์ ตัวแบบที่สองที่มีความซับซ้อนมากกว่าตัวแบบแรกจะเรียกว่าตัวแบบที่สมบูรณ์(Complete Model) และตัวแบบแรกที่เป็นตัวแบบอย่างง่ายของตัวแบบสอง เรียกว่า ตัวแบบลดรูป(Reduced Model)

ตัวอย่าง ความเหมาะสมของตัวแบบอันตรกิริยาเชิงเส้นตรงสำหรับค่าเฉลี่ยข้อมูล A ซึ่งประกอบด้วย ตัวแปรเชิงปริมาณ 2 ตัวแปร ดังนั้นตัวแบบอันตรกิริยา (Interaction Model)คือ

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

ในกรณีที่กำหนดให้ความสัมพันธ์ระหว่าง y และ x เป็นตัวแบบพหุนามกำลังสองจะได้ตัวแบบดังนี้

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2$$

ซึ่งจะเห็นได้ว่า 2 ตัวแบบนี้ติดกัน(Nested Model)

ตัวแบบเส้นโค้งที่ประกอบไปด้วยพจน์กำลังสองของ x_1 และ x_2 ซึ่งจะเหมือนกับพจน์ของอันตรกิริยา ดังนั้น 2 ตัวแบบนี้ติดกลุ่มกัน ในกรณีนี้ตัวแบบอันตรกิริยา (Interaction)จะติดกลุ่มกับตัวแบบที่สองซึ่งจะมีความซับซ้อน โดยอยู่ในรูปของตัวแบบพหุนามกำลังสอง ดังนั้นตัวแบบพหุนามกำลังสองจะเป็นตัวแบบที่สมบูรณ์(Complete Model) และ ตัวแบบอันตรกิริยา (Interaction) จะเป็นตัวแบบลดรูป(Reduced Model)

ในการทดสอบสมมติฐานแบบติดกลุ่มจะเป็นการเปรียบเทียบตัวแบบที่ติดกัน(Nested model) นั่นคือ ถ้าตัวแบบเส้นโค้งไม่สามารถที่จะให้ข้อมูลในการพยากรณ์ค่า Y ได้ นั่นแสดงว่าพจน์กำลังสองของ 2 ตัวแปรอาจจะไม่จำเป็นต้องนำมาพิจารณา ดังนั้นในการทดสอบสมมติฐานเราจะทดสอบสมมติฐานพารามิเตอร์ของเทอม x_1^2 และ x_2^2 ว่าเท่ากับ 0 หรือไม่

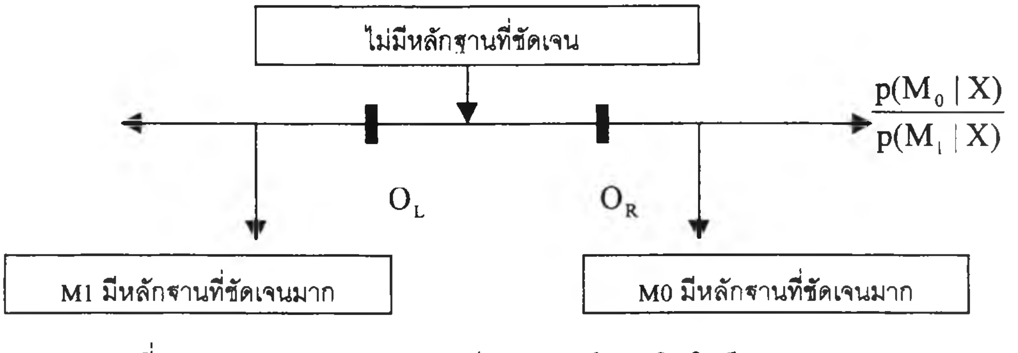
$$H_0 : \beta_4 = \beta_5 = 0$$
$$H_1 : \beta_i \neq 0 \text{ i=4,5}$$

2.5 การค้นหาปริภูมิตัวแบบของเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์ (Bayesian Information Criterion(BIC))[4]

การคัดเลือกตัวแบบในแนวคิดของเบส์จะพิจารณาหาค่าความน่าจะเป็นภายหลังสำหรับทุกๆ ตัวแบบที่เราสนใจ ซึ่งอาจจะมีปริภูมิตัวแบบที่เป็นไปได้ทั้งหมดมีขนาดใหญ่ และยากที่จะพิจารณาตัวแบบที่เป็นไปได้ทั้งหมด ตัวอย่างเช่น ในกรณีที่มีตัวแปรอิสระ 27 ตัวแปรจะมีตัวแบบที่เป็นไปได้ทั้งหมด $2^{27} = 134{,}217{,}728$ ตัวแบบ ดังนั้นเราจึงต้องการวิธีการในการค้นหาปริภูมิตัวแบบ เพื่อความง่ายและรวดเร็วในการหาความน่าจะเป็นภายหลังสำหรับแต่ละตัวแบบ ดังนั้นวิธีการที่นำมาใช้ในการหาปริภูมิตัวแบบ คือ ออกส์แคม วินโดว์ (Occam's Window) ซึ่งเป็นวิธีการพิจารณาอัตราส่วนภาวะน่าจะเป็นภายหลังสำหรับการเปรียบเทียบตัวแบบ 2 ตัวแบบ โดยมีหลักในการพิจารณาดังนี้

- ถ้าล็อกออดส์ภายหลัง(Log Posterior Odds) มีค่าเป็นบวก นั่นแสดงว่า ข้อมูลมีความชัดเจนในการสนับสนุนตัวแบบอย่างง่าย ดังนั้นเราจะปฏิเสธ M_1 และจะมาพิจารณา M_0 เราสามารถกำหนดล็อกออดส์ภายหลัง(Log Posterior Odds) เป็นค่าคงที่บวก O_R ก่อนที่จะปฏิเสธ M_1
- ถ้าล็อกออดส์ภายหลัง(Log Posterior Odds) มีค่าน้อยและเป็นลบ นั่นแสดงว่า ข้อมูลที่มีอยู่ไม่มีความชัดเจนในการสนับสนุนตัวแบบอย่างง่าย ดังนั้นเรายังเก็บตัวแบบทั้งสองนี้เอาไว้ก่อน
- ถ้าล็อกออดส์ภายหลัง(Log Posterior Odds) มีค่ามากและเป็นบวก นั่นคือ มีค่าน้อยกว่า $O_L = \log(c)$ เมื่อ c กำหนดจากสมการที่ (2.3) ดังนั้นเราจะปฏิเสธ M_0 และจะพิจารณา M_1

การเปรียบเทียบตัวแบบจากหลักการข้างต้นสามารถพิจารณาได้จากรูปที่ 1



รูปที่ 2.1 แสดงการแปลผลของวิธีการออกส์แคม วินโดว์ (Occam's Window)

---

[4] David Madigan and Adrian E. Raftery. "Model Selection and Accounting for Model Uncertainly in Graphical Models Using Occam's Window". Technical Report. July 14,1992. Revised; June 22,1993 ;October 10,1993. University of Washington. pp. 5-10

ออกส์แคม วินโดว์ (Occam's Window) เป็นวิธีการหนึ่งในการค้นหาปริภูมิตัวแบบ เสนอโดย Madigan และ Raftery(1993) ซึ่งถูกนำมาใช้ในตัวแบบความถดถอย วิธีการนี้ถ้าตัวแบบมีความน่าจะเป็นภายหลังห่างจากตัวแบบที่ดีที่สุด ตัวแบบนั้นจะไม่น่าเชื่อถือและไม่นำมาพิจารณาอีกต่อไป

$$A' = \left\{ M_1 : \frac{\max_1 \{p(M_0 \mid X)\}}{p(M_1 \mid X)} \le c \right\} \qquad (2.3)$$

ดังนั้น ตัวแบบที่เป็นไปตามสมการที่(2.3) ก็จะถูกตัดออกจากปริภูมิตัวแบบ ในสมการที่(2.3) c เป็นค่าที่มีขนาดใหญ่ และเป็นค่าที่ถูกเลือกโดยผู้วิจัย สำหรับงานวิจัยครั้งนี้ผู้วิจัยได้กำหนด c=20 สำหรับการเปรียบเทียบในกรณีที่ระดับนัยสำคัญเท่ากับ 0.05 และ c=100 สำหรับการเปรียบเทียบในกรณีที่ระดับนัยสำคัญเท่ากับ 0.01(Jeffreys (1961) Appendix B.) สำหรับตัวแบบที่ยอมรับจะสอดคล้องกับ

$$B = \left\{ H_0 : \exists H_1 \in A, H_1 \subset H_0, \frac{p(M_0 \mid X)}{p(M_1 \mid X)} > 1 \right\} \qquad (2.4)$$

สมการที่ 2.4 จะเป็นการแยกตัวแบบที่ถูกยอมรับออกจากสมการที่ 2.3

หลักการของออกส์แคม วินโดว์ (Occam's Window) จะเป็นการยืนยันการแปลผลอัตราส่วนความน่าจะเป็นภายหลังของตัวแบบ $\frac{p(M_0 \mid X)}{p(M_1 \mid X)}$ เมื่อ M_0 มีตัวแปรอิสระน้อยกว่า M_1 อย่างน้อย 1 ตัว ซึ่งมีกระบวนการในการพิจารณาดังนี้

กระบวนการ Up – Down

ขั้นตอนแรกของการเลือกตัวแบบเริ่มต้น ในทางปฏิบัติเราจะเลือกตัวแบบเริ่มต้นจากตัวแบบที่เป็นไปได้ทั้งหมด ซึ่งวิธีการในการค้นหาตัวแบบเริ่มต้นจะมี 2 วิธีนั่นคือ

"Up" เป็นกระบวนการในการค้นหาตัวแบบเริ่มต้นโดยการเพิ่มตัวแปร และ "Down" เป็นกระบวนการในการค้นหาตัวแบบเริ่มต้นโดยการลดตัวแปร เมื่อเริ่มจากตัวแบบที่ได้จากการกำหนดตัวแปร ขั้นตอนแรกจะใช้ "Up" และจะทำการกำหนดตัวแบบเริ่มต้นจาก "Down"

กำหนดให้ A และ C เป็นสมาชิกของปริภูมิตัวแบบ M เมื่อ A แทนตัวแบบที่ได้รับการยอมรับ และ C แทนตัวแบบภายใต้การพิจารณา สำหรับทั้งสองกระบวนการ จะเริ่มจากการกำหนด A=ϕ และ C เท่ากับตัวแบบเริ่มต้นที่ได้จากการกำหนด กำหนดให้ $O_L=1$ และ $O_R=c$ แทนขอบเขตทางซ้ายและขวาสำหรับการแสดงออกส์แคม วินโดว์ (Occam's Window) ดังรูปที่ 2.1(หน้า 27)

กระบวนการ Down สำหรับการคัดเลือกตัวแบบ

1.เลือกตัวแบบ M จาก C

2. $C \leftarrow C - M$ และ $A \leftarrow A + M$

3. เลือกตัวแบบย่อย M_0 จาก M โดยการลดตัวแปรออกจาก M

4.คำนวณ ค่า $B = \log \dfrac{p(M_0 \mid X)}{p(M \mid X)}$

5.ถ้า $B > O_R$ ดังนั้น $A \leftarrow A - M$ และถ้า $M_0 \notin C$ $C \leftarrow C + M_0$

6.ถ้า $O_L \le B \le O_R$ ดังนั้นถ้า $M_0 \notin C$ $C \leftarrow C + M_0$

7.ถ้ามีตัวแบบย่อย M ให้กลับไปทำในขั้นตอนที่ 3

8.ถ้า $C \neq \phi$ กลับไปทำในขั้นตอนที่ 1

กระบวนการ Up สำหรับการคัดเลือกตัวแบบ

1.เลือกตัวแบบ M จาก C

2. $C \leftarrow C - M$ และ $A \leftarrow A + M$

3. เลือกตัวแบบย่อย M_1 จาก M โดยการเพิ่มตัวแปรเข้าไปใน M

4.คำนวณ ค่า $B = \log \dfrac{p(M \mid X)}{p(M_1 \mid X)}$

5.ถ้า $B < O_L$ ดังนั้น $A \leftarrow A - M$ และถ้า $M_1 \notin C$ $C \leftarrow C + M_1$

6.ถ้า $O_L \le B \le O_R$ ดังนั้นถ้า $M_1 \notin C$ $C \leftarrow C + M_1$

7.ถ้ามีตัวแบบย่อย M ให้กลับไปทำในขั้นตอนที่ 3

8.ถ้า $C \neq \phi$ กลับไปทำในขั้นตอนที่ 1

การกำหนดตัวแบบในการเปรียบเทียบ จะประกอบด้วยลำดับของการเปรียบเทียบตัวแบบติดกลุ่ม (Nested Model) เป็นคู่ๆ ถ้าตัวแบบ H_0 ถูกปฏิเสธ ดังนั้นทุกตัวแบบที่ติดกลุ่ม กับตัวแบบข้างต้นจะถูกปฏิเสธไปด้วย โดยเฉพาะถ้ามีหลักฐานสนับสนุน H_0 แล้ว H_1 จะถูกปฏิเสธ แต่หากต้องการปฏิเสธ H_0 เราต้องมีหลักฐานที่มากพอ ถ้าหลักฐานที่มียังไม่เพียงพอ จะไม่มีตัวแบบใดถูกปฏิเสธออกไป

2.6 **ตัวประกอบของเบส์** (Bayes Factor) [5]

ตัวประกอบของเบส์ (Bayes factor) คือ อัตราส่วนออดส์ภายหลัง(Posterior Odds) ของสองสมมติฐานเมื่อความน่าจะเป็นก่อน(Prior probability) ของสองสมมติฐาน คือ $\frac{1}{2}$ ซึ่งจะเป็นตัวที่ใช้วัดความชัดเจนในการยอมรับสมมติฐาน H_0

นิยามที่ 2

กำหนดข้อมูล X ภายใต้การทดสอบสองสมมติฐาน H_0 และ H_1 มีความหนาแน่นน่าจะเป็น $p(X \mid H_0)$ หรือ $p(X \mid H_1)$ และกำหนดการแจกแจงก่อน $p(H_0)$ และ $p(H_1) = 1 - p(H_0)$ จะได้ความน่าจะเป็นภายหลัง คือ $p(X \mid H_0)$ และ $p(X \mid H_1) = 1 - p(X \mid H_0)$ ดังนั้นแนวคิดของการแจกแจงก่อนที่จะแปลงไปเป็น ความน่าจะเป็นภายหลังจากการพิจารณาข้อมูลที่มีอยู่ เราจะทำการแปลงให้อยู่ในรูปแบบที่ง่าย โดยอาศัยอัตราส่วนออดส์ (odds= probability/(1-probability) นั่นคือ

$$p(H_k \mid X) = \frac{p(X \mid H_k)p(H_k)}{p(X \mid H_0)p(H_0) + p(X \mid H_1)p(H_1)} \quad \text{(k=0,1)}$$

ดังนั้นจะได้ว่า

$$\underset{(1)}{\frac{p(H_0 \mid X)}{p(H_1 \mid X)}} = \underset{(2)}{\frac{p(X \mid H_0)}{p(X \mid H_1)}} \bullet \underset{(3)}{\frac{p(H_0)}{p(H_1)}}$$

โดย (1) แทน Posterior odds (2) แทน Bayes factor และ (3) แทน prior odds

จากสมการข้างต้น เราจะทำการแปลงให้อยู่ในรูปผลคูณอย่างง่าย นั่นคือ

$$B_{01} = \frac{p(H_0 \mid X)/p(H_1 \mid X)}{p(H_0)/p(H_1)}$$

$$= \frac{\frac{p(X \mid H_0)p(H_0)}{p(X)} \Big/ \frac{p(X \mid H_1)p(H_1)}{p(X)}}{p(H_0)/p(H_1)}$$

$$= \frac{p(X \mid H_0)}{p(X \mid H_1)}$$

โดย B_{01} แทน อัตราส่วนออดส์ภายหลัง(Posterior Odds)ของ H_0 และ H_1

---

[5] Kass. And Raftery, A.E. "Baye Factor" Journal Of The American Statistical Association. Vol. 90 ,377-395.

ดังนั้น ตัวประกอบของเบส์ (Bayes factor)สำหรับการวิเคราะห์ความถดถอยมีรูปแบบดังนี้

$$B_{01} = \frac{\int_0^\infty p(\underset{\sim}{\beta}_0 \mid \sigma_0, H_0) p(\underset{\sim}{\beta}_0, \sigma_0)\ d\underset{\sim}{\sigma}_0}{\int_0^\infty p(\underset{\sim}{\beta}_1 \mid \sigma_1, H_1) p(\underset{\sim}{\beta}_1, \sigma_1)\ d\underset{\sim}{\sigma}_1}$$

โดยที่

$\underset{\sim}{\beta}_0, \underset{\sim}{\beta}_1$ แทน เวกเตอร์ของพารามิเตอร์ภายใต้ H_0, H_1

$p(\underset{\sim}{\beta}_0, \sigma_0), p(\underset{\sim}{\beta}_1, \sigma_1)$ แทน การแจกแจงก่อนของ $\underset{\sim}{\beta}_0, \underset{\sim}{\beta}_1$

$p(\underset{\sim}{\beta}_0 \mid \sigma_0, H_0), p(\underset{\sim}{\beta}_1 \mid \sigma_1, H_1)$ แทน ความน่าจะเป็นของ x เมื่อกำหนด $\underset{\sim}{\beta}_0, \underset{\sim}{\beta}_1$ หรือภาวะน่าจะเป็นของ $\underset{\sim}{\beta}_0, \underset{\sim}{\beta}_1$