

การเปรียบเทียบตัวแบบกำลังสองน้อยสุด กับตัวประมาณริคจ์ในตัวแบบถดถอยเชิงเส้นพหุคูณ เมื่อข้อมูลมีปัญหาพหุสัมพันธ์

*ปัทมวดี นันทนานันตร์

ในงานวิจัยด้านต่างๆ ไม่ว่าจะเป็นด้านวิทยาศาสตร์ สังคมศาสตร์ หรือในด้านธุรกิจนั้น หากต้องการคาดคะเนหรือพยากรณ์ค่าของตัวแปรตาม(dependent variable) ที่เราสนใจ ว่าขึ้นอยู่กับกลุ่มของตัวแปรอิสระ(independent variables) อย่างไร โดยทั่วไปจะใช้เทคนิคที่เรียกว่าการวิเคราะห์การถดถอยพหุคูณ(Multiple Regression Analysis) และนิยมประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยของตัวแบบ ด้วยตัวประมาณกำลังสองน้อยสุด(Least Squares Estimator) เนื่องจากตัวประมาณดังกล่าวจะมีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียง และในบรรดาตัวประมาณที่ไม่เอนเอียงเชิงเส้นทั้งหมด ตัวประมาณกำลังสองน้อยสุดจะให้ค่าความแปรปรวนต่ำสุด ซึ่งเรียกว่ามีคุณสมบัติ BLUE (Best Linear Unbiased Estimator) สำหรับตัวแบบการถดถอยพหุคูณเชิงเส้นอยู่ในรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \varepsilon_i \quad ; i = 1, 2, \dots, n \quad , j = 1, 2, \dots, k$$

หรือเขียนในรูปเมทริกซ์ ได้ดังนี้

$$\underline{y} = \underline{X} \underline{\beta} + \underline{\varepsilon}$$

เมื่อ	$\underline{y}$	แทนเวกเตอร์ของตัวแปรตามที่มีขนาด $(n \times 1)$
	$\underline{X}$	แทนเมทริกซ์ของตัวแปรอิสระที่มีขนาด $(n \times q)$
	$\underline{\beta}$	แทนเวกเตอร์ของสัมประสิทธิ์ถดถอยที่มีขนาด $(q \times 1)$
	$\underline{\varepsilon}$	แทนเวกเตอร์ของความคลาดเคลื่อนที่มีขนาด $(n \times 1)$

n แทนขนาดตัวอย่าง

k แทนจำนวนตัวแปรอิสระในตัวแบบ

และ q แทนจำนวนพารามิเตอร์ที่ต้องการประมาณค่าในตัวแบบ $= k+1$

ซึ่งตัวประมาณกำลังสองน้อยสุดที่กล่าวถึงข้างต้นอยู่ในรูปแบบดังนี้

$$\hat{\beta}_{LS} = (X'X)^{-1} X'y \quad (1.1)$$

แต่ในการประมาณค่าพารามิเตอร์ด้วยตัวประมาณกำลังสองน้อยสื่อนั้น ลักษณะของข้อมูลจะต้องเป็นไปตามข้อสมมติ (Assumptions) คือ ความคลาดเคลื่อนสุ่ม (random error) $\varepsilon_i; i = 1, \dots, n$ มีค่าเฉลี่ยเท่ากับศูนย์ ความแปรปรวนเท่ากับ σ^2 คงที่ทุกค่า i และความแปรปรวนร่วมของ $(\varepsilon_i, \varepsilon_j); i \neq j$ ทุกคู่เท่ากับศูนย์หรือไม่มีสหสัมพันธ์กัน (uncorrelation) ดังนั้น ในกรณีที่ข้อมูลไม่เป็นไปตามข้อสมมติดังกล่าว การใช้ตัวประมาณกำลังสองน้อยสุดจะทำให้ตัวประมาณพารามิเตอร์ขาดคุณสมบัติ BLUE นอกจากนี้ ตัวประมาณกำลังสองน้อยสุดอาจขาดคุณสมบัติ BLUE อันเนื่องจากสาเหตุอื่นได้เช่นกัน เช่น กรณีที่ตัวแปรอิสระมีความสัมพันธ์กันที่เรียกว่า มีพหุสัมพันธ์ (Multicollinearity) ดังนั้น จึงได้มีการคิดค้นหาตัวประมาณอื่นแทนตัวประมาณกำลังสองน้อยสุดเมื่อเกิดกรณีที่ไม่เหมาะสมที่จะใช้ตัวประมาณกำลังสองน้อยสุด เช่น ตัวประมาณการถดถอยริดจ์ (Ridge Regression Estimator) เป็นต้น

ได้กล่าวไปแล้วว่า ตัวประมาณกำลังสองน้อยสุดจากสมการ (1.1) มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียง และให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Squares Error) ต่ำสุดในบรรดาตัวประมาณที่ไม่เอนเอียงเชิงเส้นทั้งหมด แต่ในการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณด้วยตัวประมาณกำลังสองน้อยสื่อนั้น สมมติฐานที่สำคัญข้อหนึ่ง คือ ตัวแปรอิสระแต่ละตัวต้องไม่มีความสัมพันธ์กับตัวแปรอิสระตัวอื่นซึ่งในทางปฏิบัติเกิดขึ้นได้น้อยมาก และเมื่อตัวแปรอิสระเกิดพหุสัมพันธ์กันสูงจะทำให้เมทริกซ์ $X'X$ เกิดเงื่อนไขที่ไม่ดี (ill-condition) นั่นคือ ทำให้ $|X'X|$ มีค่าเล็กลงเข้าใกล้ศูนย์เนื่องจาก เมทริกซ์ความแปรปรวนร่วมของค่าประมาณสัมประสิทธิ์การถดถอยอยู่ในรูป $\text{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ จึงส่งผลให้ความแปรปรวนของค่าประมาณสัมประสิทธิ์การถดถอยมีค่ามาก และมีผลทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณมีค่ามาก ดังนั้น ถ้าตัวแปรอิสระมีความสัมพันธ์กันสูง เราอาจแก้ไขได้โดยการตัดตัวแปรอิสระบางตัวออกจากตัวแบบ แต่ในบางกรณี ความสัมพันธ์ระหว่างตัวแปรอิสระไม่ชัดเจน ทำให้การตัดตัวแปรอิสระตัวใดตัวหนึ่งออกจากตัวแบบทำได้ยาก หรืออาจไม่ต้องการตัดตัวแปรอิสระตัวใดออกจากตัวแบบ เพราะถือว่า ตัวแปรอิสระทุกตัวมีผลในการอธิบายตัวแปรตามได้มากพอ ๆ กัน

Hoerl และ Kennard(1970) ได้ศึกษาหาตัวประมาณค่าพารามิเตอร์ในแบบการถดถอยพหุคูณเชิงเส้นที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าตัวประมาณกำลังสองน้อยสุด เมื่อข้อมูลเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยตัวประมาณนี้สร้างจากหลักการนำค่าคงที่ค่าหนึ่งที่มากกว่าศูนย์(k) มาบวกกับสมาชิกทุกตัวในแนวทแยงมุมของเมทริกซ์ $X'X$ เนื่องจาก เมทริกซ์ความแปรปรวนร่วมของตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณอยู่ในรูป $Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$ ดังนั้น การนำค่าคงที่ที่มากกว่าศูนย์ มาบวกกับสมาชิกทุกตัวในแนวทแยงมุมของเมทริกซ์ $X'X$ จะทำให้ $(X'X)^{-1}$ มีค่าลดลง ซึ่งจะทำให้ได้ตัวประมาณที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าตัวประมาณกำลังสองน้อยสุด โดยตัวประมาณดังกล่าวเรียกว่า ตัวประมาณการถดถอยริดจ์(Ridge Regression Estimator) ซึ่งมีรูปแบบของตัวประมาณดังนี้

$$\hat{\beta}_{\sim RID} = (X'X + kI)^{-1} \cdot X' y \quad ; k > 0 \quad (1.2)$$

เมื่อ k เป็น ค่าคงที่ หรือเรียกว่าเป็น พารามิเตอร์ที่เอนเอียง(Biasing Parameter)

I เป็น เมทริกซ์เอกลักษณ์(Identity Matrix) ที่มีขนาด (q x q)

ซึ่งตัวประมาณการถดถอยริดจ์เป็นตัวประมาณที่เอนเอียง แต่จะมีค่าความแปรปรวนจากตัวอย่างน้อยกว่าตัวประมาณกำลังสองน้อยสุด ซึ่งจะทำให้ตัวประมาณมีความแม่นยำสูงกว่า

ในการวิเคราะห์การถดถอยพหุคูณเมื่อข้อมูลเกิดพหุสัมพันธ์นั้น การเลือกค่า k ที่เหมาะสม จะสามารถทำให้ตัวประมาณริดจ์มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าตัวประมาณกำลังสองน้อยสุดได้ แต่ในทางปฏิบัติเราไม่ทราบว่าค่า k ที่เหมาะสมควรเป็นเท่าไร ซึ่งมีนักวิจัยได้เสนอวิธีการหาค่า k ที่เหมาะสมจำนวนหลายวิธีด้วยกัน โดยเริ่มแรก Hoerl และ Kennard (1970) ใช้เทคนิคการหาค่า k โดยพิจารณาจากกราฟ(Graphical Technique) หรือเรียกว่า Ridge Trace แต่เป็นวิธีการที่ยุ่งยาก จึงมีผู้คิดค้นวิธีการหาค่า k ที่เหมาะสมโดยไม่ใช้กราฟ (Nongraphical Technique) ไว้มากมาย ซึ่ง Gibbons (1981) ได้ทำการวิจัยเพื่อเปรียบเทียบวิธีการต่างๆ ในการหาค่า k ที่เหมาะสม ผลการวิจัยพบว่า โดยภาพรวม วิธีการเลือกค่า k ที่ดีที่สุด คือ วิธีการประมาณค่า k ที่คิดค้นโดย Hoerl , Kennard และ Bladwin(1975) ซึ่งแทนด้วย k_{HKB} และเขียนได้ในรูป

$$k_{HKB} = \frac{qs^2}{\hat{\beta}_{\sim LS}' \hat{\beta}_{\sim}} \quad (1.3)$$

เมื่อ
$$s^2 = \frac{\left(\underset{\sim}{y} - X \underset{\sim}{\hat{\beta}}_{LS} \right)' \left(\underset{\sim}{y} - X \underset{\sim}{\hat{\beta}}_{LS} \right)}{n - q}$$
 เป็นตัวประมาณพารามิเตอร์ σ^2

ซึ่งในการศึกษาครั้งนี้จะศึกษาในกรณีที่ข้อมูลมีปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ โดยทำการศึกษาตัวประมาณกำลังสองน้อยสุดซึ่งอยู่ในรูปแบบสมการ (1.1) และตัวประมาณริดจ์ซึ่งอยู่ในรูปแบบสมการ (1.2) โดยใช้ค่า k_{HKB} ซึ่งอยู่ในรูปแบบสมการ (1.3) และเปรียบเทียบประสิทธิภาพของตัวประมาณโดยใช้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Squares Error : RMSE) ซึ่งตัวประมาณใดที่ให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำสุด จะถือว่าเป็นตัวประมาณที่มีประสิทธิภาพสูงสุดในแต่ละสถานการณ์ และข้อมูลที่ใช้ในการศึกษาครั้งนี้ได้จากการจำลองด้วยเทคนิคการจำลองมอนติคาร์โล (Monte Carlo Simulation Technique) โดยทำการจำลองซ้ำจำนวน 1,000 ครั้ง ในแต่ละสถานการณ์

ขอบเขตของการวิจัย

1. ตัวแบบการถดถอยพหุคูณเชิงเส้นในการศึกษาครั้งนี้ อยู่ในรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \quad ; i = 1, 2, \dots, n$$

หรือเขียนในรูปเมทริกซ์ ได้ดังนี้

$$\underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ แทนเวกเตอร์ของตัวแปรตามที่มีขนาด $(n \times 1)$

X แทนเมทริกซ์ของตัวแปรอิสระที่มีขนาด $(n \times q)$

$\underset{\sim}{\beta}$ แทนเวกเตอร์ของสัมประสิทธิ์ถดถอยที่มีขนาด $(q \times 1)$

$\underset{\sim}{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อนที่มีขนาด $(n \times 1)$

n แทนขนาดตัวอย่าง

และ q แทนจำนวนพารามิเตอร์ที่ต้องการประมาณค่าในตัวแบบ = 4

โดยมีข้อกำหนด(Assumptions) ว่า อันดับ (rank) ของเมทริกซ์ X เท่ากับ q (full rank) ; ($q < n$) ความคลาดเคลื่อนเป็นตัวแปรสุ่มที่เป็นอิสระกันและมีการแจกแจงเดียวกัน ที่มีค่าเฉลี่ย $E(\varepsilon) = 0$ และ เมทริกซ์ความแปรปรวน - ความแปรปรวนร่วม $E(\varepsilon \varepsilon') = \sigma^2 I$ ซึ่งศึกษาในกรณีที่ $\beta = 1$

2. การแจกแจงของความคลาดเคลื่อนสุ่มมีการแจกแจงปกติ(Normal Distribution)

$$f(\varepsilon) = \frac{1}{\sigma^2 \sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon - \mu)^2\right\} \quad ; -\infty < \mu < \infty, \sigma^2 > 0$$

เมื่อ μ แทน ค่าเฉลี่ยของความคลาดเคลื่อนสุ่ม
 σ^2 แทน ความแปรปรวนของความคลาดเคลื่อนสุ่ม

ซึ่งศึกษาในกรณีที่ $\mu = 0$ และ $\sigma^2 = 3$

3. จำนวนตัวแปรอิสระที่ใช้ในการศึกษาเท่ากับ 3 ตัวแปร และศึกษาในกรณีมีพหุสัมพันธ์ระหว่างตัวแปรอิสระเพียง 1 คู่ (x_1 กับ x_2) โดยมีระดับสหสัมพันธ์ (ρ) ระหว่างตัวแปรอิสระคู่ดังกล่าว 7 ระดับ คือ 0.1, 0.3, 0.5, 0.7, 0.9, 0.95 และ 0.99 ส่วนตัวแปรอิสระอีก 1 ตัวที่เหลือ (x_3) ไม่มีสหสัมพันธ์กับตัวแปรอิสระตัวอื่นๆ ซึ่งศึกษาในกรณีที่ตัวแปรอิสระ x_3 มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 5 ส่วนในการสร้างตัวแปรอิสระ x_1 กับ x_2 ให้มีสหสัมพันธ์กัน ใช้วิธีการจำลองของ Wilchern และ Churchill(1978) ซึ่งมีสูตรในการคำนวณดังนี้

$$X_{ij} = (1 - \rho^2)^{1/2} \cdot Z_{ij} + \rho Z_{i4} \quad ; i = 1, 2, \dots, n, j = 1, 2, 3$$

เมื่อ $Z_{i1}, \dots, Z_{i4}$ แทน ตัวเลขสุ่มที่มีการแจกแจงปกติมาตรฐานที่เป็นอิสระต่อกัน (Independent Standard Normal Random Number)

ρ แทน ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ x_1 กับ x_2

4. ขนาดตัวอย่าง(n) มี 6 ระดับ คือ 20, 30, 35, 40, 50 และ 60

5. เกณฑ์ที่ใช้ในการพิจารณา

ทำการเปรียบเทียบประสิทธิภาพของตัวประมาณพารามิเตอร์แต่ละตัว โดยการเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณการถดถอยพหุคูณเชิงเส้น (Root Mean Squares Error : RMSE) โดยมีสูตรในการคำนวณดังต่อไปนี้

$$RMSE = \sqrt{\sum_{j=1}^{1,000} \frac{(\beta_{ij} - \hat{\beta}_{ij})^2}{1,000}}$$

เมื่อ β_{ij} แทน ค่าจริงของพารามิเตอร์ในสมการถดถอยตัวที่ i ในการจำลองรอบที่ j

เมื่อ $i = 1, 2, 3, 4$ และ $j = 1, 2, \dots, 1,000$

$\hat{\beta}_{ij}$ แทน ค่าประมาณของพารามิเตอร์ในสมการถดถอยตัวที่ i ในการจำลองรอบที่ j

ผลการทดลอง

ตารางที่ 1 แสดงค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง(RMSE) จากการใช้ตัวประมาณกำลังสองน้อยสุด(LS)กับตัวประมาณริตจ์(RID)ในการประมาณค่าพารามิเตอร์ โดยจำแนกตามระดับขนาดตัวอย่าง(n)และระดับสหสัมพันธ์ของตัวแปรอิสระ x_1 กับ $x_2 (\rho)$

ขนาดตัวอย่าง(n)	ตัวประมาณ	ระดับสหสัมพันธ์ของตัวแปรอิสระ x_1 กับ $x_2 (\rho)$						
		0.1	0.3	0.5	0.7	0.9	0.95	0.99
20	LS	0.007302*	0.008040*	0.009631	0.011750	0.017943	0.023882	0.049483
	RID	0.007782	0.008257	0.009214*	0.010136*	0.012366*	0.014221*	0.022753*
30	LS	0.005696*	0.006135*	0.007055	0.008261	0.012111	0.015478	0.030785
	RID	0.006070	0.006282	0.006729*	0.007267*	0.008509*	0.009326*	0.013971*
35	LS	0.005184*	0.005443*	0.005988*	0.006766	0.009811	0.012459	0.024632
	RID	0.005524	0.005690	0.006031	0.006423*	0.007399*	0.008065*	0.011896*
40	LS	0.004742*	0.004927*	0.005370*	0.006015	0.008661	0.010671	0.020466
	RID	0.005160	0.005263	0.005474	0.005830*	0.006412*	0.006925*	0.009931*
50	LS	0.004358*	0.004511*	0.004871*	0.005456	0.007742	0.009399	0.017754
	RID	0.004744	0.004839	0.005008	0.005309*	0.005840*	0.006307*	0.009019*
60	LS	0.004009*	0.004145*	0.004435*	0.004888	0.006867	0.008255	0.015436
	RID	0.004352	0.004439	0.004594	0.004861*	0.005328*	0.005754*	0.008055*

หมายเหตุ * แทนค่า RMSE ที่มีค่าต่ำสุดในแต่ละสถานการณ์

ตารางที่ 2 แสดงตัวประมาณค่าพารามิเตอร์ที่ควรเลือกใช้ในแต่ละสถานการณ์ของระดับขนาดตัวอย่าง (n) และระดับสหสัมพันธ์ของตัวแปรอิสระ x_1 กับ $x_2 (\rho)$

ขนาดตัวอย่าง(n)	ระดับสหสัมพันธ์ของตัวแปรอิสระ x_1 กับ $x_2 (\rho)$						
	0.1	0.3	0.5	0.7	0.9	0.95	0.99
20	LS	LS	RID	RID	RID	RID	RID
30	LS	LS	RID	RID	RID	RID	RID
35	LS	LS	LS	RID	RID	RID	RID
40	LS	LS	LS	RID	RID	RID	RID
50	LS	LS	LS	RID	RID	RID	RID
60	LS	LS	LS	RID	RID	RID	RID

ผลการทดลองดังแสดงได้จากตารางที่ 1 และ 2 ซึ่งแสดงค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง(RMSE) จากการใช้ตัวประมาณกำลังสองน้อยสุด (LS) กับตัวประมาณริดจ์ (RID) ในการประมาณค่าพารามิเตอร์ โดยจำแนกตามระดับขนาดตัวอย่าง(n)และระดับสหสัมพันธ์ของตัวแปรอิสระ x_1 กับ $x_2 (\rho)$ และแสดงตัวประมาณค่าพารามิเตอร์ที่ควรเลือกใช้ในแต่ละสถานการณ์ นำมาสรุปผลการทดลองได้ดังนี้

1. เมื่อระดับสหสัมพันธ์ในตัวแปรอิสระมีค่าน้อยกว่าหรือเท่ากับ 0.5

ตัวประมาณกำลังสองน้อยสุดจะให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าตัวประมาณริดจ์ ในทุกระดับขนาดตัวอย่าง ยกเว้นขนาดตัวอย่างเท่ากับ 20 และ 30

2. เมื่อระดับสหสัมพันธ์ในตัวแปรอิสระมีค่ามากกว่า 0.5

ตัวประมาณริดจ์จะให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง(RMSE) ต่ำกว่าตัวประมาณกำลังสองน้อยสุดในทุกระดับขนาดตัวอย่าง

3. ผลกระทบจากปัจจัยต่างๆ ที่มีต่อค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) พบว่า ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง(RMSE)จะแปรผันตามกับระดับสหสัมพันธ์ในตัวแปรอิสระ และแปรผกผันกับระดับขนาดตัวอย่าง เมื่อกำหนดให้ปัจจัยอื่นๆ คงที่

การประยุกต์ใช้กับข้อมูลจริง

การที่จะนำผลการศึกษาที่ได้ในครั้งนี้นำไปประยุกต์ใช้กับข้อมูลจริงในการประมาณค่าหรือพยากรณ์ค่าของตัวแปรตามที่ต้องการได้ มีหลักในการปฏิบัติดังนี้

1. ข้อมูลจริงจะต้องมีตัวแปรอิสระ 3 ตัวแปร และมีสหสัมพันธ์ระหว่างตัวแปรอิสระเพียง 1 คู่ เท่านั้น โดยตัวแปรอิสระอีกตัวหนึ่งจะต้องไม่มีสหสัมพันธ์กับตัวแปรอิสระตัวอื่น ๆ ที่เหลือ จึงจะสามารถนำผลการศึกษาค้างนี้ไปประยุกต์ใช้กับข้อมูลจริงได้ เช่น หากต้องการพยากรณ์ตัวแปรตามคือเงินเดือนแรกเข้าของพนักงาน โดยตัวแปรอิสระที่จะใช้ในการพยากรณ์ตัวแปรตามมี 3 ตัวแปร ได้แก่ สาขาที่จบอายุ และประสบการณ์ทำงาน ซึ่งจะพบว่าอายุและประสบการณ์ทำงานน่าจะมีสหสัมพันธ์กัน ส่วนสาขาที่จบไม่น่าจะมีสหสัมพันธ์กับอายุและประสบการณ์ทำงาน ดังนั้น ในกรณีเช่นนี้ ควรใช้ตัวประมาณจริงในการประมาณค่าพารามิเตอร์ในตัวแบบ เพราะจะทำให้ผลการพยากรณ์มีความคลาดเคลื่อนต่ำกว่าการใช้ตัวประมาณกำลังสองน้อยสุดในการประมาณค่าพารามิเตอร์ในตัวแบบ

2. ผลการวิจัยครั้งนี้จะนำไปใช้กับข้อมูลจริงได้ในกรณีที่ความคลาดเคลื่อนสุ่มมีการแจกแจงปกติ เท่านั้น ดังนั้น จึงควรทดสอบการแจกแจงของความคลาดเคลื่อนสุ่มก่อนว่ามีการแจกแจงปกติหรือไม่ โดยสามารถทดสอบการแจกแจงของตัวแปรตามแทนได้ เนื่องจากหากความคลาดเคลื่อนสุ่มมีการแจกแจงปกติ ตัวแปรตามก็จะมีแจกแจงปกติเช่นกัน ในการทดสอบว่าตัวแปรตามมีการแจกแจงปกติหรือไม่ นั้นทำได้โดยใช้สถิติทดสอบ Kolmogorov - Smirnov Test (K-S Test) หรือ Shapiro - Wilk Test

3. ในการที่จะทดสอบว่าระดับพหุสัมพันธ์ของตัวแปรอิสระอยู่ในระดับใดนั้น ทดสอบโดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) โดยที่

ถ้าพบว่าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าน้อยกว่า 0.40 (ตัวแปรอิสระมีพหุสัมพันธ์กันในระดับต่ำ) ควรใช้ตัวประมาณกำลังสองน้อยสุดในการประมาณค่าพารามิเตอร์ในตัวแบบ

ถ้าพบว่าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ระหว่าง 0.40 - 0.60 (ตัวแปรอิสระมีพหุสัมพันธ์กันในระดับปานกลาง) และขนาดตัวอย่างที่ใช้น้อยกว่า 35 ควรใช้ตัวประมาณจริงในการประมาณค่าพารามิเตอร์ แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ระหว่าง 0.41 - 0.60 และขนาดตัวอย่างที่ใช้อยู่ระหว่าง 35 - 60 ควรใช้ตัวประมาณตัวประมาณกำลังสองน้อยสุดในการประมาณค่าพารามิเตอร์ในตัวแบบ

ถ้าพบว่าค่าสัมประสิทธิ์สหสัมพันธ์มีค่ามากกว่า 0.60 (ตัวแปรอิสระมีพหุสัมพันธ์กันในระดับค่อนข้างสูงถึงสูงมาก) ควรใช้ตัวประมาณจริงในการประมาณค่าพารามิเตอร์ ในตัวแบบ

ซึ่งในการเลือกใช้ตัวประมาณค่าพารามิเตอร์ที่เหมาะสม จะสามารถทำให้ได้ผลการพยากรณ์ที่มีความถูกต้องมากยิ่งขึ้น และสามารถเขียนตารางสรุปการเลือกใช้ตัวประมาณพารามิเตอร์ในสถานการณ์ต่าง ๆ ได้ดังนี้

ขนาดตัวอย่าง	ระดับสหสัมพันธ์ของ ตัวแปรอิสระ	ตัวประมาณค่าพารามิเตอร์ ที่ควรเลือกใช้
น้อยกว่า 35	น้อยกว่า 0.4	ตัวประมาณกำลังสองน้อยสุด(LS)
	0.4 ถึง 0.6	ตัวประมาณริดจ์(RID)
	มากกว่า 0.6	ตัวประมาณริดจ์(RID)
35 ถึง 60	น้อยกว่า 0.4	ตัวประมาณกำลังสองน้อยสุด(LS)
	0.4 ถึง 0.6	ตัวประมาณกำลังสองน้อยสุด(LS)
	มากกว่า 0.6	ตัวประมาณริดจ์(RID)

บรรณานุกรม

- Askin, R. G., and Montgomery, D. C. (1980). Augmented robust estimators. *Technometrics*, 22, pp.333-341.
- Chatterjee, Samprit and Price, Bertram. (1977). *Regression Analysis by Example*. New York : John Wiley & Sons.
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression : Applications to nonorthogonal problems. *Technometrics*, 12, pp.69-82.
- Hoerl, A. E., Kennard, R. W., and Bladwin, K. F. (1975). Ridge Regression : Some simulations. *Communication Statistics*, 4, pp.105-123.
- Koenker, R. W., and Bassett, G. W. (1978). Regression quantiles. *Econometrica*, 46, pp.33-50.
- Neter, John and Wasserman, William. (1990). *Applied Linear Statistical Models*. Homewood, Illinois : Richard D. Irwin.