บทที่ 2

ทฤษฎีที่เกี่ยวข้องกับการวิจัย

ตัวแบบการถดถอยพหุนามหลักเกณฑ์ดี (Well-formulated Polynomial Regression Models)

ในปี ค.ศ. 1987 เพโซโต (Peixoto) ได้เสนอทฤษฎีบทที่เกี่ยวกับตัวแบบการถดถอยพหุนามหลักเกณฑ์ดี โดยอาศัยความรู้ทางด้านทฤษฎีเซต (Set theory) รายละเอียดของนิยามและทฤษฎีที่เกี่ยวข้องมีดังนี้

ให้ตัวแปรอิสระทั้งหมดของตัวแบบการถดถอยพหุนามเป็นเซต S ซึ่งเป็นเซตย่อยของ $\{ x_1^{i1} x_2^{i2} \ldots x_m^{im} : i_1, i_2, \ldots, i_m = 0, 1, \ldots, q \}$ โดยที่ $m \geq 1$ และ $q \geq 2$

นิยามที่ 1

ตัวแปรอิสระที่อยู่ในรูป $x_1^{h1-i1} x_2^{h2-i2} \ldots x_m^{hm-im}$ โดยที่ $i_j = 0, 1, \ldots, h_j$ สำหรับทุก j (j=1,...,m) จะเรียกว่าอยู่ต่ำกว่าหรือต่ำกว่าโดยระดับชั้น (inferior or hierarchically inferior) ของตัวแปรอิสระ $x_1^{h1} x_2^{h2} \ldots x_m^{hm}$ ก็ต่อเมื่อมี i_j อย่างน้อยที่สุด 1 ตัวที่มากกว่า 0 เราใช้สัญลักษณ์ $z_1 < z_2$ แทนความหมายว่าตัวแปรอิสระ z_1 อยู่ต่ำกว่า z_2

นิยามที่ 2

ตัวแปรอิสระ z_2 จะเรียกว่าอยู่เหนือกว่าหรือเหนือกว่าโดยระดับชั้น (superior or hierarchically superior) ของตัวแปรอิสระ z_1 ถ้า z_1 อยู่ต่ำกว่า z_2

นิยามที่ 3

ตัวแปรอิสระ z_1 และ z_2 จะเปรียบเทียบได้โดยระดับชั้น (hierarchically comparable) ถ้า z_1 อยู่ต่ำกว่าหรือเหนือกว่า อย่างใดอย่างหนึ่งกับ z_2 และจะเปรียบเทียบไม่ได้โดยระดับชั้น (hierarchically incomparable) ถ้าไม่มีตัวใดต่ำกว่าอีกตัว เราใช้สัญลักษณ์ $z_1 \| z_2$ แทนความหมายว่าตัวแปรอิสระ z_1 เปรียบเทียบไม่ได้โดยระดับชั้นกับ z_2

นิยามที่ 4

ตัวแปรอิสระ z ถูกเรียกว่าใหญ่สุดเฉพาะกลุ่มโดยระดับชั้น (hierarchically maximal) ในเซต S ถ้าไม่มีตัวแปรอิสระใดในเซต S อยู่เหนือกว่า z และตัวแปรอิสระ z ที่ใหญ่สุดเฉพาะกลุ่มโดยระดับชั้นจะเป็นตัวแปรอิสระที่มากสุดโดยระดับชั้น (hierarchically greatest) ในเซต S ถ้าตัวแปรอิสระทุกตัวในเซต S นั้นอยู่ต่ำกว่า z

นิยามที่ 5

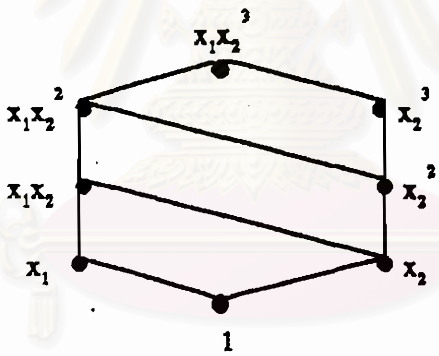
z_2 จะปกคลุม (cover) z_1 ก็ต่อเมื่อ $z_1 < z_2$ และไม่มีตัวแปรอิสระ z ใด ๆ ที่จะทำให้ $z_1 < z < z_2$

เราสามารถเขียนโครงสร้างโดยลำดับชั้น (hierarchically structure) ของตัวแปรอิสระในรูปของ Hasse Diagram ได้ โดยแต่ละตัวแปรอิสระจะแทนด้วยวงกลม ถ้า z_2 ปกคลุม z_1 แล้ว วงกลมที่แทน z_2 จะอยู่สูงกว่าวงกลมที่แทน z_1 และวงกลม 2 วงนี้จะถูกเชื่อมโดยเส้นตรง ส่วนตัวแปรอิสระที่มีกำลังรวม[1] (total exponent) เท่ากัน จะมีวงกลมอยู่ในระดับเดียวกัน

ตัวอย่างที่ 1 จากตัวแบบการถดถอยพหุนาม

$$y_i = \beta_{00} + \beta_{10}x_{1i} + \beta_{01}x_{2i} + \beta_{11}x_{1i}x_{2i} + \beta_{02}x_{2i}^2 + \beta_{12}x_{1i}x_{2i}^2 + \beta_{03}x_{2i}^3 + \beta_{13}x_{1i}x_{2i}^3 + \varepsilon_i$$

ทำให้เราสามารถเขียน Hasse Diagram ได้ดังนี้



นิยามที่ 6

ตัวแบบการถดถอยพหุนามใด ๆ จะเรียกว่าเป็นตัวแบบหลักเกณฑ์ดีก็ต่อเมื่อ ถ้ามีตัวแปรอิสระ z อยู่ในตัวแบบแล้ว ตัวแปรอิสระทุกตัวที่อยู่ต่ำกว่า z จะต้องอยู่ในตัวแบบด้วย

หมายเหตุ ในปี ค.ศ. 1979 เบิร์นฮาร์ดทและจุง (Bernhardt and Jung) เรียกตัวแบบหลักเกณฑ์ดีว่า ตัวแบบสมบูรณ์ (complete model)

---

[1] กำลังรวมของตัวแปรอิสระที่อยู่ในรูป $x_1^{h1}\ x_2^{h2}\ \ldots\ x_m^{hm}$ คือ $h_1+h_2+\ldots+h_m$

<u>ทฤษฎีบทที่ 1</u>

ให้ $\mathbf{X}$[1] $= [(x_{ij})]$ $i = 1,2,\ldots,n$; $j = 1,2,\ldots,m$ และ $\mathbf{W}$ เป็นการแปลงเชิงเส้นของ $\mathbf{X}$ อยู่ในรูป

(1) $$\mathbf{W} = \mathbf{XA} + \mathbf{1}_n\mathbf{b}'$$

โดยที่ $\mathbf{A} = \text{diag}(a_1,a_2,\ldots,a_m)$ เป็นเมทริกซ์ไม่ใช่เอกฐาน

$\mathbf{1}_n$ เป็นเวกเตอร์ $n\times 1$ ที่มีสมาชิกเป็น 1 ทั้งหมด

และ $\mathbf{b}' = (b_1,b_2,\ldots,b_m)$ เป็นเวกเตอร์ขนาด $1\times m$

ให้ $Z(X)$ เป็นเมทริกซ์ตัวแบบของตัวแบบการถดถอยพหุนามดังที่กำหนดในบทที่ 1 และให้ $Z^*(X)$ เป็นเมทริกซ์ตัวแบบของตัวแบบหลักเกณฑ์ดีซึ่งได้จากการเพิ่มพจน์ที่อยู่ต่ำกว่าที่ขาดหายใน $Z(X)$

<u>ทฤษฎีบทที่ 2</u>

ให้ $\aleph[\,Z(X)\,]$ แทนปริภูมิการประมาณ (estimation space) ของตัวแบบ หรือบางทีเรียกว่าเป็นปริภูมิแนวตั้ง (column space) ของ $Z(X)$ จะได้ว่า

(1) $\aleph[\,Z^*(X)\,] = \aleph[\,Z^*(W)\,]$ สำหรับทุก ๆ W ที่อยู่ในรูปสมการ (1)

(2) ถ้าตัวแบบใน $Z(X)$ เป็นตัวแบบหลักเกณฑ์ดีแล้ว $\aleph[\,Z(X)\,] = \aleph[\,Z(W)\,]$ สำหรับทุก ๆ W ที่อยู่ในรูปของสมการ (1)

(3) ถ้า $Z^*(X)$ มีค่าลำดับชั้นแนวตั้งเต็มอัตรา (full column rank) และตัวแบบไม่เป็นตัวแบบหลักเกณฑ์ดีแล้ว $\aleph[\,Z(X)\,] \neq \aleph[\,Z(W)\,]$ สำหรับทุก ๆ W ที่อยู่ในรูปสมการ (1) ซึ่งทุก ๆ $b \neq 0$

พจน์ของตัวแบบการถดถอยพหุนามจะยืนยงภายใต้การแปลง (invariant under transformation) ก็ต่อเมื่อตัวแบบการถดถอยพหุนามนั้นเป็นตัวแบบหลักเกณฑ์ดี ยกเว้นถ้าการแปลงเชิงเส้นนั้นเป็นการแปลงมาตราส่วน[2] (scale transformation) การแปลงมาตราส่วนไม่สามารถเปลี่ยนปริภูมิการประมาณของตัวแบบการถดถอยพหุนามได้ แต่การแปลงเลื่อนขนาน[3]

---

[1] อักษรตัวใหญ่เข้มจะหมายถึงเมทริกซ์ ส่วนอักษรตัวเล็กเข้มหมายถึงเวกเตอร์

[2] คือการแปลงในรูปแบบสมการ (1) โดยที่ $b = 0$

[3] คือการแปลงในรูปแบบสมการ (1) โดยที่ $b \neq 0$

(translation transformation) จะเลื่อนจุดเริ่มต้นของตัวแปรอิสระเริ่มต้น $x_1, x_2, ..., x_m$ และอาจจะมีผล กระทบกับปริภูมิการประมาณของตัวแบบที่ไม่มีหลักเกณฑ์ดี (Non Well-formulated Models) ได้

ริชาร์ด บี ดาร์ลิงตัน (Richard B. Darlington ,1990) ได้กล่าวถึงผลกระทบของการแปลงค่าตัวแปรอิสระ หรือการดึงค่าเข้าศูนย์กลาง (centering) ในการถดถอยพหุนามที่จะมีต่อค่าต่าง ๆ ดังนี้

1. การวัดความสัมพันธ์เชิงเดียว (Measure of Simple Relationship) เช่น สหสัมพันธ์เชิงเดียว (simple correlation) สัมประสิทธิ์การถดถอยเชิงเดียว (simple regression coefficient) ซึ่งพจน์หรือตัวแปรอิสระที่จะได้รับผลกระทบคือทุกพจน์ยกเว้นพจน์ที่มีกำลัง 1 (first-power) หรือพจน์เชิงเส้น (linear term)

2. การวัดบทบาทส่วนหนึ่งอย่างเดียว (Measure of Unique Contribution) เช่น สัมประสิทธิ์การถดถอยย่อยหรือบางส่วน (partial regression coefficient) สหสัมพันธ์บางส่วน (partial correlation) และค่าทดสอบทีหรือเอฟที่ใช้ทดสอบนัยสำคัญ ซึ่งจะกระทบต่อทุกพจน์ยกเว้นพจน์ที่มีกำลังสูงสุด

สำหรับผลกระทบจากการแปลงตัวแปรในการถดถอยพหุนามที่มีต่อค่าทดสอบทีนั้นก็ได้มีการกล่าวไว้ในบทความของกรีเพนทรอก ไรอัน และสมิธ (Griepentrog, Ryan, and Smith) ในปี ค.ศ.1982

การเลือกสมการการถดถอยที่ดีที่สุด (Selecting the Best Regression Equation)

การสร้างตัวแบบจากวิธีการเลือกสมการการถดถอยที่ดีที่สุดซึ่งกล่าวไว้ในงานวิจัยมี 3 วิธีดังนี้

<u>1. การสร้างตัวแบบด้วยวิธีกำจัดตัวแปรอิสระย้อนหลัง</u> (Model Building by Backward Elimination Method)

วิธีการนี้เริ่มต้นด้วยสมการการถดถอยที่ประกอบด้วยตัวแปรอิสระทุกตัวที่ใช้พิจารณา แล้วจะคัดตัวแปรอิสระออกครั้งละ 1 ตัว ตัวแปรอิสระตัวแรกที่จะถูกคัดออก คือตัวแปรอิสระที่มีความสัมพันธ์กับ y น้อยที่สุด (โดยกำหนดให้ตัวแปรอื่น ๆ คงที่) และค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระนั้นไม่มีนัยสำคัญ แต่ถ้าค่าสัมประสิทธิ์การถดถอยของตัวแปรนั้นมีนัยสำคัญ แสดงว่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระที่เหลือก็จะมีนัยสำคัญด้วย ซึ่งทำให้สมการการถดถอยประกอบด้วยตัวแปรอิสระทุกตัว ขั้นตอนต่อไปหลังการคัดออกแล้วจะคำนวณหาสมการการถดถอยสำหรับตัวแปรอิสระที่เหลือ และคัดตัวแปรอิสระที่มีนัยสำคัญน้อยที่สุดออก

ถ้าไม่มีตัวแปรอิสระใดถูกคัดออกตัวแปรอิสระทุกตัวก็จะอยู่ในสมการ และจะเป็นการจบของวิธีการนี้ เราสามารถแสดงขั้นตอนของวิธีดังกล่าวดังนี้

1.1 สมการการถดถอยประกอบด้วยทุก ๆ ตัวแปรอิสระ

1.2 คำนวณค่าสถิติเอฟบางส่วน (partial F) ของทุก ๆ ตัวแปรอิสระโดยทำเสมือนว่าตัวแปรอิสระนั้นเข้าสู่สมการเป็นตัวสุดท้าย

1.3 ค่าสถิติเอฟบางส่วนที่น้อยที่สุด (F_L) จะถูกเปรียบเทียบกับค่า F ณ ระดับนัยสำคัญที่กำหนด (F_O)

ก) ถ้า $F_L < F_O$ จะตัดตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนน้อยสุดนั้นออกจากสมการ และคำนวณค่าสัมประสิทธิ์ของสมการถดถอยเมื่อตัดตัวแปรอิสระแล้วโดยวิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Squares Method) จากนั้นกลับไปที่ 1.2

ข) ถ้า $F_L > F_O$ จะได้สมการการถดถอยที่เหมาะสม

<u>2. การสร้างตัวแบบด้วยวิธีการถดถอยขั้นบันได</u> (Model Building by Stepwise Regression Method)

วิธีการนี้เป็นวิธีการเพิ่มเติมของวิธีการคัดเลือกตัวแปรอิสระข้างหน้า (Forward Selection) โดยเพิ่มเติมในแต่ละขั้นตอนของการคัดเลือกตัวแปรอิสระแต่ละตัวเข้าสมการการถดถอยด้วยการกำจัดตัวแปรอิสระอื่น ๆ ที่เข้ามาอยู่ในสมการการถดถอยก่อนตัวแปรอิสระตัวล่าสุดที่เพิ่มจะเข้าไปในสมการตามหลักเกณฑ์ของวิธีการกำจัดตัวแปรอิสระย้อนหลัง เราสามารถแสดงขั้นตอนต่าง ๆ ของวิธีดังกล่าวได้ดังนี้

2.1 สมการการถดถอยเริ่มจากไม่มีตัวแปรอิสระอยู่ในสมการ

2.2 เลือกตัวแปรอิสระที่มีสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) กับ y สูงสุดเข้าสมการเป็นตัวแรก และทดสอบว่าตัวแปรนั้นมีนัยสำคัญ มิฉะนั้นจะได้ว่า $y = \bar{y}$ เป็นสมการการถดถอยที่เหมาะสม

2.3 คำนวณค่าสถิติเอฟบางส่วนสำหรับทุกตัวแปรอิสระที่ไม่อยู่ในสมการ

2.4 ค่าสถิติเอฟบางส่วนที่มากที่สุด (F_U) จะถูกเปรียบเทียบกับค่า F ณ ระดับนัยสำคัญที่กำหนด (F_O)

ก) ถ้า $F_U < F_O$ จะไม่นำตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนมากสุดนั้นเข้าสมการ

ข) ถ้า $F_U > F_O$ จะนำตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนมากสุดนั้นเข้าสมการ และคำนวณค่าสัมประสิทธิ์ของสมการการถดถอยเมื่อนำตัวแปรอิสระนั้นเข้าสมการโดยวิธีกำลังสองน้อยสุดสามัญ

2.5 คำนวณค่าสถิติเอฟบางส่วนของทุกตัวแปรอิสระที่อยู่ในสมการถดถอย

2.6 ค่าสถิติเอฟบางส่วนที่ต่ำที่สุด (F_L) จะถูกเปรียบเทียบกับค่าเอฟ ณ ระดับนัยสำคัญที่กำหนด (F_O)

ก) ถ้า $F_L < F_O$ จะตัดตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนต่ำสุดนั้นออกจากสมการ และคำนวณค่าสัมประสิทธิ์ของสมการการถดถอยเมื่อตัดตัวแปรอิสระแล้ว

ข) ถ้า $F_L > F_O$ จะไม่ตัดตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนนั้นออกจากสมการ

2.7 ถ้าไม่มีตัวแปรอิสระเข้าและออกจากสมการแล้วจะได้สมการที่เหมาะสม มิฉะนั้นจะกลับไปที่ 2.3

3. <u>การสร้างตัวแบบด้วยวิธีตัวแบบหลักเกณฑ์ดี</u> (Model Building by Well-formulated Model)

การสร้างตัวแบบจากวิธีตัวแบบหลักเกณฑ์ดีนี้สร้างโดยการเพิ่มตัวแปรอิสระทุกตัวที่อยู่ต่ำกว่าสำหรับทุกตัวแปรอิสระที่อยู่ในตัวแบบแล้วโดยอาศัยนิยามที่ 1

การประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณโดยวิธีกำลังสองน้อยสุดสามัญ[1]

วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณนี้มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้นที่คิดขึ้นโดยคาร์ล เฟรดริก เกาส์ (Carl Friedrich Gauss) ในปี ค.ศ. 1777-1855 และอังเดร แอนดรีวิช มาร์โคฟ (Andrei Andreevich Markov) ในปี ค.ศ. 1855-1922 โดยมีหลักในการประมาณค่าสัมประสิทธิ์คือ ทำให้ผลรวมกำลังสองของความคลาดเคลื่อน (Sum of Squares Error (SSE)) มีค่าน้อยที่สุด ซึ่งแสดงรายละเอียดดังนี้

---

[1] ปราณี รัตนัง, "การประมาณสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาดมีการแจกแจงแบบเบ้ และมีการแจงแจงแบบหางยาวกว่าการแจกแจงปกติ, " (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย,2531)หน้า 16-17.

<u>นิยามที่ 7</u>

จากสมการ $y = X\beta + \varepsilon$ จะได้ว่าตัวประมาณกำลังสองน้อยสุดของ β คือ $\hat{\beta}$ ที่ทำให้ผลรวมกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด

จากนิยามที่ 7 เราจะทำการหาตัวประมาณกำลังสองน้อยสุดได้ดังนี้

กำหนด $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)'$ จะได้ว่า $\hat{y} = X\hat{\beta}$ ดังนั้น ε เป็นเวกเตอร์ของความคลาดเคลื่อน ซึ่งผลบวกกำลังสองของความคลาดเคลื่อนคือ

$$
\begin{aligned}
SSE &= (\hat{\varepsilon}'\hat{\varepsilon}) \\
&= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= (y' - \hat{\beta}'X')(y - X\hat{\beta}) \\
&= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \\
&= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}
\end{aligned}
$$

การหาค่าน้อยที่สุดของผลรวมกำลังสองของความคลาดเคลื่อนทำได้โดยการหาอนุพันธ์ (differentiate) เทียบกับ β_i ; $i = 0,1,\ldots,p$ (p เป็นจำนวนตัวแปร) แล้วกำหนดให้เท่ากับ 0

$$
\frac{\partial(\hat{\varepsilon}'\hat{\varepsilon})}{\partial\beta_i} = \frac{\partial(y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta})}{\partial\beta_i} = 0
$$

จะได้ว่า $-2X'y + 2X'X\hat{\beta} = 0$

$$(X'X)\hat{\beta} = X'y$$

ดังนั้น $\hat{\beta} = (X'X)^{-1}X'y$

และเมทริกซ์ความแปรปรวนร่วมของตัวประมาณ β คือ

$$
\begin{aligned}
cov(\hat{\beta}) &= cov[(X'X)^{-1}X'y] \\
&= (X'X)^{-1}X'\,cov(y)\,X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

การทดสอบเอฟบางส่วน (Partial F-test)

การทดสอบเอฟบางส่วนเป็นการทดสอบที่ใช้ตรวจสอบนัยสำคัญของ β_j เพื่อตัดสินใจว่าตัวแปรอิสระใดควรอยู่ในสมการหรือตัวแปรอิสระใดควรตัดออกจากสมการ โดยที่ β_j จะปรากฏอยู่ ณ ตำแหน่งใดในแบบจำลองก็ได้ แต่ในทางปฏิบัติจะทำโดยถือว่าตัวแปรอิสระนั้นเข้าสู่สมการเป็นตัวสุดท้าย

ให้ z_j คือฟังก์ชันของตัวแปรอิสระ x_j และ

(2) $$y = \beta_0 + \beta_1 z_1 + \dots + \beta_q z_q + \varepsilon$$

เป็นสมการแสดงความสัมพันธ์ของ y กับ z จากสมการ (2) เราสามารถหาค่าตัวประมาณ ${}_1\hat{\beta}' = ({}_1\hat{\beta}_1 , {}_1\hat{\beta}_2 , \dots , {}_1\hat{\beta}_q)$ โดยวิธีกำลังสองน้อยสุดและค่าผลบวกกำลังสองที่เกี่ยวข้องได้ดังนี้

ก) ${}_1\hat{\beta} = ({}_1Z'{}_1Z)^{-1}{}_1Z'y$ เมื่อ ${}_1Z$ คือเมทริกซ์ขนาด n×q

ข) $SSR_1 = {}_1\hat{\beta}'{}_1Z'y$

ค) $SSE_1 = y'y - {}_1\hat{\beta}'{}_1Z'y$

และ $MSE_1 = \sigma_1^2$

$$= \frac{1}{n-q}(y'y - {}_1\hat{\beta}'{}_1Z'y)$$

โดยที่ SSR หมายถึง Sum Squares of Regression

(3) กำหนด $y = \beta_0 + \beta_1 z_1 + \dots + \beta_q z_q + \beta_{q+1} z_{q+1} + \dots + \beta_p z_p + \varepsilon$

เป็นสมการแสดงความสัมพันธ์ของ y กับ z โดยที่ p>q จากสมการ (3) สามารถหาค่าตัวประมาณ ${}_2\hat{\beta}' = ({}_2\hat{\beta}_1 , {}_2\hat{\beta}_2 , \dots, {}_2\hat{\beta}_q , {}_2\hat{\beta}_{q+1} , \dots, {}_2\hat{\beta}_p)$ โดยวิธีกำลังสองน้อยสุดและค่าผลบวกกำลังสอง (Sum of Squares) ที่เกี่ยวข้องได้ดังนี้

ก) ${}_2\hat{\beta} = ({}_2Z'{}_2Z)^{-1}{}_2Z'y$ เมื่อ ${}_2Z$ คือเมทริกซ์ขนาด n×p

ข) $SSR_2 = {}_2\hat{\beta}'{}_2Z'y$

ค) $SSE_2 = y'y - {}_2\hat{\beta}'{}_2Z'y$

และ $MSE_2 = \sigma_2^2$

$$= \frac{1}{(n-p)} (y'y - {}_2\hat{\beta}'{}_2Z'y)$$

จากผลลัพธ์ข้างต้นจะพบว่า Extra Sum of Squares Regression คือ ESSR โดยที่

$$\begin{aligned} ESSR &= SSR_2 - SSR_1 \\ &= {}_2\hat{\beta}'_2 Z'y - {}_1\hat{\beta}'_1 Z'y \end{aligned}$$

ซึ่ง Extra Sum of Squares นี้เป็นค่าผลบวกกำลังสองของตัวแปรอิสระ z_{q+1} , z_{q+2} , z_p ที่เพิ่มขึ้นมาจากสมการ (2)

โดยอาศัยความรู้เรื่อง Distribution of Quadratic Form เราสามารถพิสูจน์ได้ว่า[1]

$$\frac{ESSR}{\sigma^2} \sim \chi^2_{(p-q)} \text{ และ } \frac{SSE_2}{\sigma^2} \sim \chi^2_{(n-p)}$$

$$\text{ดังนั้นค่าทดสอบเอฟบางส่วน} = \frac{ESSR/(p-q)\sigma^2}{SSE_2/(n-p)\sigma^2} = \frac{ESSR/(p-q)}{\hat{\sigma}_2^2}$$

จะมีการแจกแจงแบบเอฟ[2] ณ ระดับขั้นความเสรี (p-q,n-p) และเราจะปฏิเสธสมมติฐาน $H_0 : \beta_{q+1} = \beta_{q+2} = \ldots = \beta_p = 0$ ณ ระดับนัยสำคัญ α เมื่อ ค่าทดสอบเอฟบางส่วน > $F_{1-\alpha,p-q,n-p}$

จากความรู้ในเรื่องการทดสอบเอฟบางส่วนนี้ เราสามารถนำมาประยุกต์ใช้กับวิธีการหาสมการการถดถอยที่ดีที่สุดได้เช่น วิธีกำจัดตัวแปรย้อนหลัง วิธีการถดถอยขั้นบันได เป็นต้น โดยการนำไปใช้ทำได้ดังนี้

จากสมการ $y = \beta_0 + \beta_1 z_1 + \ldots + \beta_p z_p + \varepsilon$ เราจะหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบ (Mean Squares Error) และ ผลบวกกำลังสองของความถดถอย (Sum of Squares Regression) ของเฉพาะ z_j จากสมการต่อไปนี้

$$MSE = \sigma^2 = \frac{1}{(n-p)} (y'y - \hat{\beta}'Z'y)$$

$$SS(z_j/z_1, z_2, \ldots, z_{j-1}, z_{j+1}, \ldots, z_p) = SS(z_1, z_2, \ldots, z_p) - SS(z_1, z_2, \ldots, z_{j-1}, z_{j+1}, \ldots, z_p)$$

ดังนั้น ค่าสถิติเอฟบางส่วน คือ

$$F_C = \frac{SS(z_j/z_1, z_2, \ldots, z_{j-1}, z_{j+1}, \ldots, z_p)}{\hat{\sigma}^2}$$

---

[1] จาก S. R. Searle

[2] $U = \dfrac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2} \sim F_{(\nu_1, \nu_2)}$

โดยจะปฏิเสธ $H_0: \beta_j = 0$ ณ ระดับนัยสำคัญ α เมื่อ

$$F_C > F_{1-\alpha,1,n-p}$$

สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

สัมประสิทธิ์สหสัมพันธ์เป็นตัวสถิติที่ใช้วัดความสัมพันธ์เชิงเส้นของตัวแปร ว่ามีความสัมพันธ์กันมากน้อยเพียงใด ซึ่งค่าสัมประสิทธิ์นั้นจะมีค่าอยู่ระหว่าง -1 และ 1 ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ -1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นกันมากในทางตรงกันข้าม ถ้าเป็น 1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นกันมากในทิศทางเดียวกัน ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นศูนย์หรือเข้าใกล้ศูนย์ ถือว่าตัวแปรทั้งสองไม่มีความสัมพันธ์เชิงเส้นกันหรือมีความสัมพันธ์เชิงเส้นกันน้อย

1. สัมประสิทธิ์สหสัมพันธ์เชิงเดียว (Simple Correlation Coefficient) เป็นตัวสถิติที่ใช้วัดความสัมพันธ์ของตัวแปร 2 ตัวใด ๆ ซึ่งมีสูตรดังนี้

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

โดยที่ n = ขนาดตัวอย่าง

y_i = ตัวแปรตามที่ i

และ x_i = ตัวแปรอิสระที่ i

2. สัมประสิทธิ์สหสัมพันธ์บางส่วน (Coefficient of Partial Correlation) เป็นตัวสถิติที่ใช้เป็นค่าวัดระดับและทิศทางความสัมพันธ์เชิงเส้นระหว่างตัวแปรสองตัว โดยที่ควบคุมให้ตัวแปรอื่น ๆ ที่เหลือไม่ให้มีการเปลี่ยนแปลงหรือมีค่าคงที่ เช่น ถ้าตัวแปรอิสระคือ x_1,x_2,x_3 และตัวแปร y และต้องการหาสัมประสิทธิ์สหสัมพันธ์บางส่วนของตัวแปรตามกับตัวแปรอิสระ สัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง y กับ x_1 โดยควบคุมให้ x_2 และ x_3 ไม่มีการเปลี่ยนแปลง จะแทนด้วยสัญลักษณ์ $r_{y1.23}$ และสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง y กับ x_2 และ x_3 คือ $r_{y2.13}$ และ $r_{y3.12}$ ตามลำดับ สำหรับการคำนวณหาค่าของสัมประสิทธิ์สหสัมพันธ์บางส่วนคู่ใด ๆ นั้น ทำได้ดังนี้

$$r^2_{ij.1,2,3,\ldots,i-1,i+1,\ldots,j-1,j+1,\ldots,k} = \left(\frac{a_{ij}^2}{a_{ii}a_{jj}}\right)$$

และ $r_{ij.\,1,2,3,\ldots,i-1,i+1,\ldots,j-1,j+1,\ldots,k} = \text{sign}\sqrt{r^2}$

เมื่อ $r_{ij.1,2,3,\ldots,i-1,i+1,\ldots,j-1,j+1,\ldots,k}$ คือค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรที่ i และ j โดยที่ตัวแปรอื่นคงที่

r_{ij} คือค่าสัมประสิทธิ์สหสัมพันธ์เชิงเดียวระหว่างตัวแปรที่ i และ j

k เป็นจำนวนตัวแปรทั้งหมด

และ a_{ij} เป็นสมาชิกแถวที่ i แนวตั้งที่ j ของเมทริกซ์ผกผันของเมทริกซ์สหสัมพันธ์ (R)

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & & r_{2k} \\ \vdots & \vdots & & \vdots \\ r_{k1} & & \cdots & 1 \end{bmatrix}$$

$$A = R^{-1} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & & \cdots & a_{kk} \end{bmatrix}$$

โดยที่ R คือเมทริกซ์สหสัมพันธ์

และ A คือเมทริกซ์ผกผันของเมทริกซ์สหสัมพันธ์

ส่วนเครื่องหมายกำหนดตามเครื่องหมายของสัมประสิทธิ์การถดถอยในสมการถดถอย