บทที่ 3
แบบจำลองเชิงเส้นรูปทั่วไป
(General Linear Model)

3.1 ข้อตกลงสมการถดถอย

ในบทที่ 2 นักศึกษาได้ศึกษาสมการถดถอยในลักษณะของ Simple Regression ซึ่งเป็นสมการถดถอยในรูปเฉพาะที่มีตัวแปรอิสระเพียงตัวเดียว ซึ่งโดยอาศัยหลักเกณฑ์ดังที่กล่าวมาในบทดังกล่าวเราย่อมสามารถขยายความสู่กรณีทั่วไปได้ อย่างไรก็ตามเพื่อความสะดวกรวดเร็วในการศึกษาเราควรจะได้ศึกษาถึงรูปทั่วไปของสมการถดถอยพหุคูณ (Multiple Regression) เป็นการเฉพาะซึ่งเราสามารถลดรูปสู่กรณีเฉพาะได้ทุกกรณี การศึกษาในส่วนนี้จะนำเสนอโดยใช้ความรู้ทางทฤษฎีเมตริกซ์และเวคเตอร์เป็นพื้นฐาน นักศึกษาจึงจำเป็นต้องมีความรู้พื้นฐานในเรื่องเหล่านี้ดีพอ

ให้ความสัมพันธ์เชิงเส้น[1] ระหว่างตัวแปร Y และตัวแปรอิสระ $X_1, X_2, \ldots X_k$ ปรากฏดังนี้

$$Y_i = \beta_1+\beta_2X_{2i}+\beta_3X_{3i}+ \ldots +\beta_kX_{ki} + u_i \quad ; \quad i = 1, 2, \ldots, n \quad \ldots\ldots 1$$

$$= \textbf{Systematic} \text{ part + Random Part}$$

โดยที่ $X_{1i} = 1$ เสมอ ; $i = 1, 2, \ldots, n$ หรือ $X_1 = (1, 1, \ldots, 1)'$

หรือสามารถเสนอในรูปย่อ ๆ ได้ดังนี้

$$Y_i = \beta_1+\sum_{i=2}^{k} \beta_jX_{ji}+u_i \; ; \quad i = 1, 2, \ldots, n$$

---

[1] คำว่าสมการเชิงเส้น (Linear Equation หรือ Linear Model) ในที่นี้หมายถึง Linear in Parameter β ขอให้เข้าใจไว้ในชั้นนี้ว่าในแบบจำลอง $Y = \beta_1+\sum_{j=2}^{k}\beta_jX_j+u$ นั้น β_j ทำหน้าที่เป็นตัวไม่ทราบค่า (Unknown Parameter) ขณะที่ X_j เป็นตัวแปรอิสระที่ทราบค่าได้จากข้อมูลที่มีอยู่ ด้วยเหตุนี้แบบจำลอง $Y = \beta_1+\sum_{j=2}^{k}\beta_jX_j+u$ ที่ β_j ยกกำลัง 1 เท่านั้นที่เป็น Linear Model หรือ Linear Equation ถ้า β_j ยกกำลังอื่น ๆ หรืออยู่ในรูปอื่น เช่น อยู่ในรูปผลคูณ $\beta_s\beta_t$ อยู่ในรูปผลหาร β_s/β_t อยู่ในรูป log เช่น $\log\beta_j$, e^{β_j} หรือรูปกำลังของ X_j เช่น $\beta_sX_j^{\beta_j}$ เป็นต้น จะมิใช่ Linear Model แต่เรียกว่า Nonlinear Model ซึ่งจะต้องใช้วิธีประมาณค่า β_j ที่แตกต่างกันออกไป ส่วนกรณีที่ตัวแปร X_j ยกกำลังอื่นที่มิใช่ 1 ขณะที่ β_j ยกกำลัง 1 เช่น $Y = \beta_1+\Sigma\beta_jX_j^{c_j}+u$ จะไม่ถือว่าเป็น Nonlinear Model แต่เรียกชื่อเฉพาะว่า Curvilinear Model ซึ่งยังคงใช้วิธีประมาณค่า β_j เช่นเดียวกับกรณีของ Linear Model ทุกประการ เพียงแต่แปลงรูป $X_j^{c_j}$ ให้เป็นตัวแปรใหม่ คือ W_j เพื่อให้มีรูปร่างเป็น Linear Model ได้อย่างสมบูรณ์

จากสมการ 1 เมื่อเรากำหนดค่าให้ $i = 1, 2, \ldots, n$ จะปรากฏระบบสมการวิวิธพันธ์เชิงเส้น (Linear Non-Homogeneous Equation) ทั้งสิ้น n สมการดังนี้

$$Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \ldots + \beta_k X_{k1} + u_1 \qquad \ldots\ldots (1)$$

$$Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \ldots + \beta_k X_{k2} + u_2 \qquad \ldots\ldots (2)$$

$$Y_3 = \beta_1 + \beta_2 X_{23} + \beta_3 X_{33} + \ldots + \beta_k X_{k3} + u_3 \qquad \ldots\ldots (3)$$

$$\vdots$$

$$Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \ldots + \beta_k X_{kn} + u_n \qquad \ldots\ldots (n)$$

จัดเป็นรูปเมตริกซ์ได้ดังนี้

$$Y = X\beta + U$$

โดยที่ Y คือเวคเตอร์ขนาด $n \times 1$ ของตัวแปรตาม Y คือ $Y = (Y_1, Y_2, \ldots, Y_n)'$

U คือ เวคเตอร์ขนาด $n \times 1$ ของ Stochastic Term (หรือ เรียกอีกอย่างหนึ่งว่า Error Term หรือ Disturbance Term) คือ $U = (u_1, u_2, \ldots, u_n)'$

β คือเวคเตอร์ขนาด $k \times 1$ ของพารามิเตอร์ β คือ $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ และ

X คือเมตริกซ์ขนาด $n \times k$ ของสัมประสิทธิของพารามิเตอร์ β หรือนัยหนึ่ง X คือเมตริกซ์ของเวคเตอร์ $X_1, X_2, \ldots, X_k$ กล่าวคือ

$$X = \begin{bmatrix} 1 & X_{21} & X_{31} & \ldots & X_{k1} \\ 1 & X_{22} & X_{32} & \ldots & X_{k2} \\ 1 & X_{23} & X_{33} & \ldots & X_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{2n} & X_{3n} & \ldots & X_{kn} \end{bmatrix} = (X_1, X_2, \ldots, X_k)$$

จากแบบจำลอง $Y = X\beta + U$ จะเห็นว่าเรามีความจำเป็นต้องควบคุมความเคลื่อนไหวทั้ง U และ X ไว้หลายประการเพื่อประโยชน์ในการประมาณค่าเวคเตอร์ β หรือนัยหนึ่ง β_j ; $j = 1, 2, \ldots, k$ เหตุที่ต้องควบคุมตัวแปร U ก็เพราะ U เป็นตัวแปรที่เป็นแหล่งข้อผิดพลาดทั้งปวงจากการกำหนดความสัมพันธ์ทางคณิตศาสตร์ระหว่างตัวแปร Y กับตัวแปร X's เช่น ความผิดพลาดจากการวัดค่า Y ความผิดพลาดจากพฤติกรรมของมนุษย์ที่ผันผวนอยู่เสมอ ความผิดพลาดจากการปรับข้อมูล ความผิดพลาดจากการกำหนดตัวแปร X's ที่คาดว่าจะสามารถอธิบายความเคลื่อนไหวของตัวแปร Y ไว้ไม่ครบถ้วน ซึ่งความผิดพลาดเหล่านี้จะส่งผลให้ค่า $\hat{Y}$ คลาดเคลื่อนไปจากค่า Y จริงอยู่ตลอดเวลา และด้วยเหตุที่ U เป็นแหล่งรับความผิดพลาด

ทั้งมวลเอาไว้ ค่าของ U จึงไม่แน่นอนและไม่อาจวัดหรือสังเกตได้ เราจึงเห็นได้ว่า U เป็น Unobservable Variable ด้วยเหตุนี้ U จึงมีความเคลื่อนไหวได้ตลอดเวลา และเป็นตัวแปรที่ Sensitive มาก ผลเสียในงานพยากรณ์ (Forcasting) และงานสร้างฟังก์ชัน (Model Formulation) จึงได้รับความกระทบกระเทือนเพราะสาเหตุจากความเคลื่อนไหวของ U เป็นส่วนใหญ่ และเพื่อให้งานประมาณค่าดำเนินไปได้เราจึงจำเป็นต้องควบคุมความเคลื่อนไหวของ U ไว้ในกรอบที่พิจารณาเห็นว่าจำเป็นด้วยการกำหนดเป็นข้อตกลงขึ้นดังนี้

1. u เป็นเวคเตอร์ของตัวแปรสุ่ม
2. $E(U) = 0$ เมื่อ 0 คือ Zero vector หรือ $0 = (0, 0, \ldots, 0)'$
3. $E(UU') = \sigma^2 I_n$
4. $U \sim N(0, \sigma^2 I_n)$

สำหรับตัวแปร X's ก็เป็นไปในทำนองเดียวกัน คือ ในงานคาดหมายหรือพยากรณ์เกี่ยวกับค่าของ Y นั้นเราจะเห็นได้ว่า Y ย่อมมีความแปรเปลี่ยนไปได้จากสาเหตุหลายประการ โดยเราจะใช้ X_j แทนสาเหตุที่ j^{th} ที่ผลักดันให้ Y เปลี่ยนแปลงค่าโดยที่ X_j และ X_l ต้องมิใช่สาเหตุเดียวกัน หรือสืบเนื่องถึงกัน ขณะเดียวกัน X's เหล่านี้จะต้องทำหน้าที่เป็นเพียงเงื่อนไขที่มีผลให้ Y มีค่าแปรเปลี่ยนไปเท่านั้น ดังนั้นเราจึงต้องเพิ่มข้อตกลงไว้อีกหลายประการดังนี้

5. $X = (X_1, X_2, \ldots, X_k)$ คือเซทของค่าคงที่ หรือ X เป็น Fixed Matrix
6. X ต้องมี Full Rank หรือ $r(X) = k \leqslant n$

สำหรับข้อตกลงเหล่านี้อาจเป็นจริงหรือไม่เป็นจริงก็ได้ในงานทางปฏิบัติ หน้าที่ของนักวิจัยก็คือ ต้องหาทางตรวจสอบดูว่างานที่ตนกำลังดำเนินการอยู่นั้นขัดแย้งกับข้อตกลงเหล่านี้หรือไม่ ถ้าขัดแย้งก็จำเป็นต้องหาทางแก้ไขให้ถูกต้องหรือสอดคล้องกับข้อตกลงต่อไป เรื่องการตรวจสอบความถูกต้องเหล่านี้จะได้กล่าวถึงต่อไปในส่วนของ Second Order Test

สำหรับความหมายของข้อตกลงทั้ง 6 ประการนี้ ผู้เขียนขอถือโอกาสอธิบายเพิ่มเติมดังต่อไปนี้ ส่วนรายละเอียดอื่น ๆ ได้กล่าวไว้แล้วในบทที่ 2

1. U เป็นเวคเตอร์ของตัวแปรสุ่ม หมายความว่า u_i ; $i = 1, 2, \ldots, n$ เป็นตัวแปรสุ่ม การกำหนดให้ตัวแปร u_i เป็นตัวแปรสุ่มมีจุดมุ่งหมายเพื่อให้สามารถควบคุมความเคลื่อนไหวของ u_i ได้โดยอาศัยทฤษฎีความน่าจะเป็นซึ่งเราสามารถกำหนด pdf ของ u_i เพื่อใช้ประโยชน์ในการหา Sampling Distribution ของ $\hat{\beta}$ และ Y ได้ ความจริงถ้ามองในแง่ของความเหมาะสม

ของข้อตกลงเราจะพบว่าเป็นสิ่งที่สมควรกำหนดเช่นนี้ เพราะ u มีความเคลื่อนไหวในลักษณะที่เป็นไปโดยสุ่มอยู่โดยธรรมชาติซึ่งเราไม่อาจคาดหมายล่วงหน้าได้ ข้อตกลงนี้จะส่งผลให้ Y เป็นตัวแปรสุ่มด้วยเพราะ Y เป็นฟังก์ชันของ U

2. $E(U) = 0$ หรือ $E(u_i) = 0 ; i = 1, 2, \ldots, n$ การกำหนดเช่นนี้หมายถึง การกำหนดว่าโดยเฉลี่ยแล้วค่าของ u_i จะหักลบกันหมดจนไม่ส่งผลกระทบต่อความสัมพันธ์ระหว่าง Y กับ X' s ขณะเดียวกันลองพิจารณาสมการ $Y = X\beta + U$ เราจะพบว่า $E(Y) = X\beta$ แสดงว่าเมื่อมีการทดลองซ้ำหลาย ๆ ครั้ง ณ ค่า $X = X_0$ ค่าเฉลี่ยของ Y จะปรากฏอยู่บน Surface $Y = X\beta$ ดังนั้นข้อตกลงว่า $E(U) = 0$ จึงถือเสมือนกับเรามีข้อตกลงว่า $E(Y) = X\beta$ และชี้ให้เห็นว่าในการพยากรณ์นั้นเราจะพยากรณ์ค่า $E(Y_i)$ มิใช่ Y_i

3. $E(UU') = \sigma^2 I_n$ เป็นข้อตกลงที่ผนวกข้อตกลงสำคัญ 2 ประการไว้ด้วยกันคือ

$$E(u_i u_j) = 0 ; \; i \neq j \quad \text{และ} \quad E(u_i^2) = \sigma^2 ; \; i = j$$

พิจารณา $E(UU')$ จะพบว่า

$$E(UU') = V(U) = E \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} (u_1, u_2, \ldots, u_n)$$

$$= \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \ldots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \ldots & E(u_2u_n) \\ \vdots & \vdots & & \vdots \\ E(u_nu_1) & E(u_nu_2) & \ldots & E(u_n^2) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix} = \sigma^2 I_n$$

ขณะเดียวกันข้อตกลงนี้ส่งผลสะท้อนสู่ตัวแปร Y ด้วยดังนี้

$\because \quad Y = X\beta + U$ ดังนั้น $V(Y) = V(U) = \sigma^2 I_n$

จากข้อตกลงนี้ เราสามารถจำแนกการพิจารณาในรายละเอียดเป็นข้อตกลงย่อยได้ดังนี้

ก. $E(u_i^2) = E(u_iu_i) = \sigma^2$; $i = j$ แสดงว่าในทุกจุดของค่า X' s ที่คงที่ ค่าของ u_i จะมีการแจกแจงค่าเบี่ยงไปจากค่าเฉลี่ย (คือ $E(u_i) = 0$) เท่ากันเสมอ หรือนัยหนึ่งความกว้างของฐานโค้งของ Y จะเท่ากันเสมอในทุกค่าเฉพาะของ X หรือฐานโค้งของ u จะเท่ากันเสมอในทุกค่าเฉพาะของ X[1]

ข. $E(u_iu_j) = 0$ หรือ $E(Y_iY_j) = 0$; $i \neq j$ หมายความว่าค่าของ u หรือค่าของ Y ในต่างช่วงต่างวาระ (period) กันจะไม่สัมพันธ์หรือกระทบสืบเนื่องถึงกัน

4. $U \sim N(0, \sigma^2 I_n)$ หรือ $u_i \sim N(0, \sigma^2)$; $i = 1, 2, ..., n$ ข้อตกลงนี้มีผลสะท้อนให้เสมือนตกลงว่า $Y \sim N(X\beta, \sigma^2 I_n)$[2]

จากข้อตกลงที่ 3 เราตกลงว่าฐานโค้งของตัวแปรสุ่ม u หรือ Y จะต้องกว้างเท่ากันเสมอ ณ ทุกค่าเฉพาะของ X แต่ข้อตกลงนี้มิได้จำกัดว่าโค้งของ u จะมีรูปร่างอย่างใด ด้วยเหตุนี้จึงต้องเพิ่มข้อจำกัดว่าโค้งดังกล่าวจะต้องเป็นโค้งปกติ ทั้งนี้เพื่อประโยชน์ในการวิเคราะห์โดยเฉพาะในแง่ของการสร้าง Sampling Distribution ของ $\hat{\beta}$ การหา Confidence Region ของ β การหา Sampling Distribution ของ Y ตลอดจนการสร้าง test สำหรับทดสอบสมมุติฐานเกี่ยวกับ β และ $C\beta$[3]

5. X เป็น Fixed Matrix หรือ X เป็นเซทของค่าคงที่โดยที่ $X = (X_1, X_2, ..., X_k)$ เมื่อ $X_1 = (1, 1, 1, ..., 1)'$ ข้อตกลงนี้หมายความว่า X_j จะต้องเป็นเวคเตอร์ของค่าคงที่[4] กล่าวคือ ณ ค่า $X_j = X_{j0}$ และทำการสำรวจและทดลองซ้ำ ๆ ณ ค่า $X_j = X_{j0}$ นี้ค่าของ Y ที่ปรากฏขึ้นอาจซ้ำค่ากันหรือมีค่าแตกต่างกันก็ได้ การผันแปรในค่าของ Y ณ ค่า $X_j = X_{j0}$ นี้มิใช่เพราะอิทธิพลของ X_j หากเป็นไปเพราะความผันแปรในค่าของ Y และอิทธิพลของ u

---

[1] ขอให้ย้อนไปดูรายละเอียดกรณี Simple Regression ในบทที่ 2

[2] สามารถพิสูจน์ได้โดยง่ายโดยอาศัย mgf-technique สำหรับ $Y \sim N(X\beta, \sigma^2 In)$ ก็แสดงว่า $Y_i \sim N(\beta_1 + \sum_{j=2}^{k} \beta_j X_{ji}\ ,\ \sigma^2)$ $i = 1, 2, ..., n$

[3] $C\beta$ โดยที่ C คือเมตริกซ์ของค่าคงที่ใด ๆ คือกรณีทั่วไป ถ้า $C = I_k$ ก็แปลว่าเรากำลังสร้าง test สำหรับทดสอบสมมุติฐานเกี่ยวกับ β

[4] ถือว่า X_j เป็น Mathematical Variable มิใช่ Random Variable ดังนั้นค่าของ X_j จึงแปรเปลี่ยนไปสู่ค่าต่าง ๆ ได้ตามความต้องการของผู้ทดสอบมิใช่แปรเปลี่ยนไปในลักษณะของ Random Variable ที่เป็นไปโดยเราไม่อาจควบคุมได้

6. X เป็นเมตริกซ์ที่มี Full Rank หมายความว่า สดมภ์ต่าง ๆ ของเมตริกซ์ X คือ เวคเตอร์ $X_1, X_2, \ldots, X_k$ จะต้องเป็นอิสระเชิงเส้นต่อกัน (ไม่มี Linear Dependent) หรือนัยหนึ่งตัวแปรอิสระ X_j และ X_s ; $j \neq s$ ต้องไม่เกี่ยวข้องกัน มิใช่ตัวแปรตัวเดียวกันหรือไม่มีสหสัมพันธ์ที่สมบูรณ์ต่อกัน

ในทางทฤษฎีเมตริกซ์เราสามารถพิสูจน์ได้โดยง่ายว่า ถ้า X มี Full Rank แล้ว เมตริกซ์ $(X'X)^{-1}$ จะมีค่าปรากฏ (หรือ $(X'X)^{-1}$ exist) และเนื่องจาก $\hat{\beta} = (X'X)^{-1}X'Y$ จึงเห็นได้ว่าเรามีความจำเป็นต้องมีเมตริกซ์ $(X'X)^{-1}$ จึงจะสามารถประมาณค่า β ได้ ดังนั้น การตั้งข้อตกลงว่าเมตริกซ์ X ต้องมี Full Rank จึงเป็นข้อตกลงที่มีผลสืบเนื่องโดยตรงต่อการปรากฏค่าของ $\hat{\beta}$

3.2 การประมาณค่าพารามิเตอร์ β

3.2.1 หนทางเลือกในการประมาณค่า β

ในที่นี้จะเสนอวิธีประมาณค่าพารามิเตอร์ β เป็น 2 วิธี วิธีแรกจะประมาณค่าโดยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square, OLS) ซึ่งเป็นวิธีที่ไม่จำเป็นต้องอาศัยข้อตกลงเกี่ยวกับ pdf ของตัวแปรสุ่ม u (หรือนัยหนึ่งคือ pdf ของ Y) กับอีกวิธีหนึ่งคือวิธี MLE (Maximum Likelihood Estimation) ซึ่งจำเป็นต้องอาศัย pdf ของ Y ดังนี้

วิธีที่ 1 Ordinary Least Square (OLS)

จากสมการ $Y = X\beta + U$

ให้ $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)'$ เป็นเวคเตอร์ของตัวประมาณค่าของ β ซึ่งเมื่อแทนที่ $\hat{\beta}$ ลงในสมการ $Y = X\beta + U$ จะได้ $Y = X\hat{\beta} + e$ เมื่อ $e = \hat{U} = Y - X\hat{\beta}$

พิจารณา Sum Square of Residual จะพบว่า

$$\Sigma e_i^2 = (e_1, e_2, \ldots, e_n)\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = e'e$$

$$
\begin{aligned}
e'e &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\
&= (Y' - \hat{\beta}'X')(Y - X\hat{\beta})^{1} \\
&= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}^{\;2} \\
&= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}
\end{aligned}
$$

ดังนั้น $\frac{\partial}{\partial\hat{\beta}} e'e = \frac{\partial}{\partial\hat{\beta}}(Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}) = 0$

=> $0 - 2X'Y + 2X'X\hat{\beta} = 0$ [3] ทำให้ได้ **Normal Equation ดังนี้**

$(X'X)\hat{\beta} = X'Y$ ---------- 2

และเนื่องจากเรามีข้อตกลงว่าเมตริกซ์ X ต้องมี full rank $(X'X)^{-1}$ จึงมีค่าปรากฏ

ดังนั้น $\hat{\beta} = (X'X)^{-1}X'Y$ คือค่าประมาณของ β ---------- 3

จากสมการ 3 แสดงว่า $\hat{\beta}$ คือ OLS-Estimator ของ β และเนื่องจาก $(X'X)^{-1}X'$ เป็นเมตริกซ์ของค่าคงที่ ดังนั้น $\hat{\beta}$ จึงเป็นฟังก์ชันเชิงเส้นของ Y เราจึงเรียก $\hat{\beta}$ ว่า Linear Estimator

---

[1] $(A \pm B)' = A' \pm B'$, $(ABC)' = C'B'A'$

[2] $Y'X\hat{\beta}$ และ $\hat{\beta}'X'Y$ ต่างก็เป็น scalar เพราะมีขนาด 1×1 ดังนั้นจึงรวมกันได้

[3] การดิฟเฟอเรนชิเอต scalar f เทียบต่อเวคเตอร์ v สามารถกระทำได้โดยง่าย เพียงแต่จัดผลของการดิฟเฟอเรนชิเอตให้อยู่ในรูปเวคเตอร์ ดังนี้คือ $\frac{\partial f}{\partial v} = \left(\frac{\partial f}{\partial v_i}\right)_{n\times1}$ เช่น

$$
f = a'x = (a_1, a_2, \ldots, a_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + a_2x_2 + \ldots + a_nx_n \;;\; \frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial X_1} \\ \frac{\partial f}{\partial X_2} \\ \vdots \\ \frac{\partial f}{\partial X_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}
$$

$$
f = X'AX = \sum_i^n a_{ii}X_i^2 + 2\sum_{i \neq j}^n a_ia_jX_iX_j = (X_1, X_2, \ldots, X_n)\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \;;\; \frac{\partial f}{\partial X} = \begin{bmatrix} 2a_{11}X_1 + 2\sum_{j\neq1}^n a_1a_jX_j \\ 2a_{22}X_2 + 2\sum_{j\neq2}^n a_2a_jX_j \\ \vdots \\ 2a_{nn}X_n + 2\sum_{j\neq n}^n a_na_jX_j \end{bmatrix}
$$

นั่นคือ $\frac{\partial}{\partial X}a'X = a$ และ $\frac{\partial}{\partial X}X'AX = 2AX$

วิธีที่ 2 Maximum Likelihood Estimation Method (MLE)

$$L = \prod_i^n f_{U_i}(u_i) = \prod_i^n \{\frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{u_i^2}{2\sigma^2}\}\}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{\sum_1^n u_i^2}{2\sigma^2}\}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\frac{U'U}{2\sigma^2}\}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{\frac{-(Y-X\beta)'(Y-X\beta)}{2\sigma^2}\}$$

$$\therefore \quad \ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{(Y-X\beta)'(Y-X\beta)}{2\sigma^2}$$

ดังนั้น $$\frac{\partial}{\partial\beta}\ln L = 0 - \frac{1}{2\sigma^2}\{0 - 2X'Y + 2X'X\beta\} = 0$$

=> $$\hat{\beta} = (X'X)^{-1}X'Y \qquad \ldots\ldots. 4$$

จะเห็นว่า $\hat{\beta}$ ในสมการ 4 ซึ่งเป็น MLE ของ β ก็คือตัวประมาณค่าตัวเดียวกันกับ $\hat{\beta}$ ตามวิธี OLS ที่ได้ตามสมการ 3 และขอให้สังเกตว่า $\hat{\beta}$ ก็คือ Linear Function ของ Y

3.2.2 คุณสมบัติของตัวประมาณค่า $\hat{\beta}$ และทฤษฎีที่เกี่ยวข้อง

ในการศึกษาคุณสมบัติของตัวประมาณค่า $\hat{\beta}$ เราจะต้องพิจารณาถึง All Possible Set ของ $\hat{\beta}$ ซึ่งมีค่าแตกต่างกันไปทุกครั้งที่มีการทดลองซ้ำ ๆ เมื่อกำหนดให้ X คงที่อยู่ ณ $X = X_0$ กล่าวคือ เมื่อเรากำหนดให้ $X = X_0$ แล้วทำการบันทึกข้อมูลหรือทดลองซ้ำ ๆ อยู่ ณ จุดนั้น ผลที่ได้รับก็คือทำให้เราได้รับเวคเตอร์ Y และเวคเตอร์ U ที่แตกต่างกันไป ซึ่งมีผลให้ $\hat{\beta}$ มีค่าผันแปรไปได้มากมาย ปัญหาก็คือ $\hat{\beta}$ จะเป็นตัวประมาณค่าที่ปราศจากอคติของ β หรือไม่ $V(\hat{\beta})$ มีค่าเท่าไร และ $\hat{\beta}$ มีการแจกแจง (Sampling Distribution) อย่างใด? เราจะได้พิสูจน์ให้เห็นคุณสมบัติเหล่านี้ได้ดังต่อไปนี้

3.2.2.1 $E(\hat{\beta}) = \beta$

จาก $\hat{\beta} = (X'X)^{-1}X'Y$ และจากข้อตกลงที่ 2 ที่ว่า $E(U) = 0$ หรือ $E(Y) = X\beta$ ดังนั้น $E(\hat{\beta}) = E\{(X'X)^{-1}X'Y\} = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'X\beta = \beta$ หรือพิสูจน์อีกวิธีหนึ่งโดยการจัดรูป $\hat{\beta}$ ให้เป็นฟังก์ชันของ β และ U ได้ดังนี้

จาก $\hat{\beta} = (X'X)^{-1}X'Y$ และจากแบบจำลอง $Y = X\beta+U$

ดังนั้น $\hat{\beta} = (X'X)^{-1}X'(X\beta+U)$

$= (X'X)^{-1}(X'X)\beta+(X'X)^{-1}X'U$

$$\Rightarrow \quad \hat{\beta} = \beta+(X'X)^{-1}X'U \qquad \text{....... 5}$$

แสดงว่า $\hat{\beta}$ เป็นฟังก์ชันของ β และ U

ดังนั้น $E(\hat{\beta}) = \beta+(X'X)^{-1}X'E(U) = \beta$

3.2.2.2 $V(\hat{\beta}) = ?$

$$V(\hat{\beta}) = E\{\hat{\beta}-E(\hat{\beta})\}\{\hat{\beta}-E(\hat{\beta})\}' = E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$$

$$= E\begin{bmatrix}\hat{\beta}_1-\beta_1\\ \hat{\beta}_2-\beta_2\\ \vdots\\ \hat{\beta}_k-\beta_k\end{bmatrix}(\hat{\beta}_1-\beta_1, \hat{\beta}_2-\beta_2, \ldots, \hat{\beta}_k-\beta_k)$$

$$= \begin{bmatrix} E(\hat{\beta}_1-\beta_1)^2 & E(\hat{\beta}_1-\beta_1)(\hat{\beta}_2-\beta_2) & \ldots & E(\hat{\beta}_1-\beta_1)(\hat{\beta}_k-\beta_k) \\ E(\hat{\beta}_2-\beta_2)(\hat{\beta}_1-\beta_1) & E(\hat{\beta}_2-\beta_2)^2 & \ldots & E(\hat{\beta}_2-\beta_2)(\hat{\beta}_k-\beta_k) \\ \vdots & \vdots & & \vdots \\ E(\hat{\beta}_k-\beta_k)(\hat{\beta}_1-\beta_1) & E(\hat{\beta}_k-\beta_k)(\hat{\beta}_2-\beta_2) & \ldots & E(\hat{\beta}_k-\beta_k)^2 \end{bmatrix}$$

แต่จากสมการที่ 5 เราทราบว่า $\hat{\beta}$ เป็นฟังก์ชันของ β และ U คือ $\hat{\beta} = \beta+(X'X)^{-1}X'U$

ดังนั้น $V(\hat{\beta}) = E\{\beta+(X'X)^{-1}X'U-\beta\}\{\beta+(X'X)^{-1}X'U-\beta\}'$

$= E\{(X'X)^{-1}X'U\}\{U'X(X'X)^{-1}\}$

$= (X'X)^{-1}X'E(UU')X(X'X)^{-1}$

$= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}$

$= \sigma^2(X'X)^{-1}(X'X)(X'X)^{-1}$

$= \sigma^2(X'X)^{-1}$

หรือพิสูจน์อีกวิธีหนึ่งโดยอาศัย Linear Transformation ดังนี้

จากความรู้เดิมที่ว่า ถ้า $Z = m_1Y_1+m_2Y_2 = (m_1, m_2)\begin{bmatrix}Y_1\\Y_2\end{bmatrix}$ ดังนั้น $E(Z) = m_1\mu_1+m_2\mu_2$ เมื่อ $E(Y_1) = \mu_1$ และ $E(Y_2) = \mu_2$ และสามารถพิสูจน์ได้โดยง่ายว่า $V(Z) = E(Z-E(Z))^2 = m_1^2\sigma_1^2+m_2^2\sigma_2^2+2m_1m_2\sigma_{12}$ เมื่อ $V(Y_1) = \sigma_1^2$, $V(Y_2) = \sigma_2^2$ และ $Cov(Y_1, Y_2) = \sigma_{12}$

จาก $V(Z) = m_1^2\sigma_1^2+m_2^2\sigma_2^2+2m_1m_2\sigma_{12}$ จะเห็นว่าเป็น Quadratic Form ซึ่งเราสามารถเสนอได้ในรูปเมตริกซ์ดังนี้

$$V(Z) = m_1^2\sigma_1^2+m_2^2\sigma_2^2+2m_1m_2\sigma_{12} = (m_1, m_2)\begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{bmatrix}\begin{bmatrix}m_1\\m_2\end{bmatrix}$$

และโดยอาศัยความรู้เรื่องนี้เราสามารถขยายความสู่กรณีทั่วไปคือ $Z = MY$ เมื่อ M เป็นเมตริกซ์ขนาด $p\times q$ Y เป็นเวคเตอร์ขนาด $q\times 1$ และ Z เป็นเวคเตอร์ขนาด $p\times 1$ กล่าวคือ

$$\begin{bmatrix}Z_1\\Z_2\\Z_3\\ \vdots\\ Z_p\end{bmatrix} = \begin{bmatrix}m_{11} & m_{12} & \dots & m_{1q}\\ m_{21} & m_{22} & \dots & m_{2q}\\ m_{31} & m_{32} & \dots & m_{3q}\\ \vdots & \vdots & & \vdots\\ m_{p1} & m_{p2} & \dots & m_{pq}\end{bmatrix}\begin{bmatrix}Y_1\\Y_2\\Y_3\\ \vdots\\ Y_q\end{bmatrix}$$

เราสามารถพิสูจน์ได้โดยง่ายได้ว่า $E(Z) = ME(Y) = M\mu$ เมื่อ $\mu = \begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \vdots\\ \mu_q\end{bmatrix}$

$V(Z) = MV(Y)M'$ เมื่อ $V(Y) = \begin{bmatrix}\sigma_1^2 & \sigma_{12} \dots \sigma_{1q}\\ \sigma_{21} & \sigma_2^2 \dots \sigma_{2q}\\ \vdots & \vdots \quad\quad \vdots\\ \sigma_{q1} & \sigma_{q2} \dots \sigma_q^2\end{bmatrix}$ = Variance - Covariance Matrix ของ Y

ดังนั้น จาก $\hat{\beta} = (X'X)^{-1}X'Y$ ให้ $M = (X'X)^{-1}X'$ และเราทราบว่า $V(Y) = \sigma^2 I_n$

ดังนั้น $V(\hat{\beta}) = MV(Y)M' = (X'X)^{-1}X' \sigma^2 I_n \{(X'X)^{-1}X'\}'$

$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$

3.2.2.3 $\hat{\beta}$ เป็น Best Linear Unbiased Estimator (BLUE) ของ β

การพิสูจน์ว่า $\hat{\beta}$ เป็น BLUE นี้ได้เคยพิสูจน์มาแล้วครั้งหนึ่งในตอน 2.3.2 ซึ่งเป็นกรณีของ Simple Linear Regression ในที่นี้จะพิสูจน์สำหรับกรณีทั่วไปดังนี้

ให้ $\hat{\beta}^*$ เป็น Unbiased Estimator อีกตัวหนึ่งของ β นอกเหนือจากที่มีอยู่เดิมคือ $\hat{\beta} = (X'X)^{-1}X'Y$ โดยที่ $\hat{\beta}^* \neq \hat{\beta}$

กำหนดให้ $\hat{\beta}^* = \{ (X'X)^{-1}X' + C \} Y = (X'X)^{-1}X'Y + CY = \hat{\beta} + CY$ โดยที่ C คือเมตริกซ์ขนาด $k \times n$ ของค่าคงที่ กล่าวคือ $C = (c_{ij})_{k\times n}$ หรือ C คือเมตริกซ์ขนาดเดียวกันกับ $(X'X)^{-1}X'$

การพิสูจน์จะพิสูจน์เป็น 3 กรณีคือ

(1) $\hat{\beta}$ เป็น Linear Function ของตัวแปรสุ่ม Y_i ; $i = 1, 2, \ldots, n$

(2) $\hat{\beta}^*$ เป็น Unbiased Estimator ของ β ภายใต้เงื่อนไข

(3) $V(\hat{\beta}^*) \geqslant V(\hat{\beta})$

1. จาก $\hat{\beta} = (X'X)^{-1}X'Y$ เมื่อกำหนดให้ $M = (X'X)^{-1}X'$ จะพบว่า $\hat{\beta} = MY$ หรือ

$\hat{\beta}_j = \sum_i^n m_{ji}Y_i$; $j = 1,2,\ldots,k$ ซึ่งแสดงให้เห็นว่า $\hat{\beta}$ เป็น Linear Function ของตัวแปรสุ่ม Y_i ; $i = 1,2,\ldots,n$

ในทำนองเดียวกัน $\hat{\beta}^* = \hat{\beta} + CY = (X'X)^{-1}X'Y + CY = (M+C)Y$ จะพบว่า $\hat{\beta}_j^* = \sum_i^n (m_{ji} + c_{ji})Y_i$; $j = 1,2,\ldots,k$ ซึ่งแสดงว่า $\hat{\beta}^*$ ก็เป็น Linear Function ของตัวแปรสุ่ม Y_i ; $i = 1, 2,\ldots,n$ เช่นกัน

2. พิจารณา $\hat{\beta}^* = \hat{\beta} + CY$ โดยที่ $\hat{\beta} = (X'X)^{-1}X'Y$ และ $Y = X\beta + U$ ดังนั้น

$$
\begin{aligned}
\hat{\beta}^* &= (X'X)^{-1}X'(X\beta + U) + C(X\beta + U) \\
&= \beta + (X'X)^{-1}X'U + CX\beta + CU \\
&= \beta + CX\beta + \{ (X'X)^{-1}X' + C \}U
\end{aligned}
$$

ดังนั้น $E(\hat{\beta}^*) = \beta + CX\beta + 0 = \beta + CX\beta \neq \beta$ ซึ่งแสดงว่า $\hat{\beta}^*$ จะเป็น Unbiased Estimator ของ β ได้ก็ต่อเมื่อ $CX = 0$[1] หรือนัยหนึ่ง $CX = 0$ คือเงื่อนไขที่จะทำให้ $\hat{\beta}^*$ เป็น Unbiased Estimator ของ β

---

[1] $\beta + CX\beta$ จะเท่ากับ β ได้เฉพาะเมื่อ $CX\beta = 0$ แต่ $\beta \neq 0$ เพราะเป็นพารามิเตอร์ที่เรามุ่งประมาณค่า ดังนั้นจึงเป็นไปได้เพียงประการเดียวคือ $CX = 0$ ขอให้สังเกตว่า 0 ในที่นี้คือ Zero Matrix ขนาด $k \times k$

3. พิจารณา $V(\hat{\beta}^*)$ จะพบว่า

$$V(\hat{\beta}^*) = E(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)'$$

แต่ $\hat{\beta}^* - \beta = \beta + CX\beta + \{(X'X)^{-1}X' + C\}U - \beta$ และเนื่องจาก $CX = 0$

ดังนั้น $\hat{\beta}^* - \beta = \{(X'X)^{-1}X' + C\}U$

นั่นคือ
$$\begin{aligned} V(\hat{\beta}^*) &= E(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)' = \{(X'X)^{-1}X' + C\}\, E(UU')\, \{(X'X)^{-1}X' + C\}' \\ &= \sigma^2 \{(X'X)^{-1}X' + C\}\{X(X'X)^{-1} + C'\} \\ &= \sigma^2 \{(X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'C' + CX(X'X)^{-1} + CC'\} \end{aligned}$$

แต่เนื่องจาก $CX = 0$ และ $(CX)' = X'C' = 0$ ดังนั้น

$$V(\hat{\beta}^*) = \sigma^2 \{(X'X)^{-1} + CC'\} = \sigma^2(X'X)^{-1} + \sigma^2 CC'$$

นั่นคือ $V(\hat{\beta}^*) = V(\hat{\beta}) + \sigma^2 CC'$ หรือ $V(\hat{\beta}_j^*) = V(\hat{\beta}_j) + \sigma^2 C_j C_j'$[1] ; $j = 1,2,\ldots,k$

เมื่อ C_j คือ Row ที่ j ของเมตริกซ์ C

พิจารณา $V(\hat{\beta}_j^*) = V(\hat{\beta}_j) + \sigma^2 C_j C_j'$ ไม่เป็นจำนวนติดลบ หรือเมตริกซ์ CC' เป็น Semi-Positive Definite Matrix แสดงว่า $V(\hat{\beta}_j^*) - V(\hat{\beta}_j) \geq 0$ หรือ $V(\hat{\beta}_j^*) \geq V(\hat{\beta}_j)$; $j = 1,2,\ldots,k$ หรือ $V(\hat{\beta}^*) \geq V(\hat{\beta})$

แสดงว่า $\hat{\beta}$ มี Variance ต่ำที่สุด (Minimum Variance) หรือเป็น Best Estimator นั่นคือ ผลการพิสูจน์ข้อ 1. - 3. ยืนยันให้เห็นว่า $\hat{\beta}$ เป็น BLUE ของ β

3.2.2.4 การประมาณค่า $V(\hat{\beta})$

จากสูตร $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$ จะพบว่าเราสามารถหาค่า $(X'X)^{-1}$ ได้เพราะ X เป็น Fixed Matrix แต่ยังไม่อาจทราบค่าของ $\sigma^2(X'X)^{-1}$ ได้เพราะ σ^2 เป็นพารามิเตอร์ที่ไม่ทราบค่า ด้วยเหตุนี้จึงยังไม่อาจทราบค่า $V(\hat{\beta})$ ได้ หนทางหนึ่งที่พอจะทำให้ทราบค่า $V(\hat{\beta})$ ได้ก็คือการประมาณค่า σ^2 ขึ้นใช้แทน σ^2

วิธีการที่จะประมาณค่า σ^2 ขึ้นนั้นโดยปกติเราจะประมาณโดยอาศัย Expected Mean of Square of Residual (EMS) หรือโดยอาศัย MLE ดังนี้

วิธีที่ 1 โดยอาศัย EMS

จาก Sum of Square Residual $\Sigma e_i^2 = e'e$ จะพบว่า

$$e'e = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

---

[1] C_j คือแถวที่ j ของแมตริกซ์ C

แต่ $Y = X\beta + U$ และ $\hat{\beta} = \beta + (X'X)^{-1}X'U$

$$
\begin{aligned}
\Rightarrow \quad e'e &= [X\beta + U - X\{\beta + (X'X)^{-1}X'U\}]'[X\beta + U - X\{\beta + (X'X)^{-1}X'U\}] \\
&= \{U - X(X'X)^{-1}X'U\}'\{U - X(X'X)^{-1}X'U\} \\
&= [\{I_n - X(X'X)^{-1}X'\}U]'[\{I_n - X(X'X)^{-1}X'\}U] \\
&= U'\{I_n - X(X'X)^{-1}X'\}\{I_n - X(X'X)^{-1}X'\}U \\
&= U'\{I_n - X(X'X)^{-1}X'\}U \quad [1]
\end{aligned}
$$

$$
\begin{aligned}
E(e'e) &= \sigma^2 \mathrm{tr}\left(I_n - X(X'X)^{-1}X'\right) \quad [2] \\
&= \sigma^2\left(\mathrm{tr}\, I_n - \mathrm{tr}\, X(X'X)^{-1}X'\right) \quad [3] \\
&= \sigma^2\left(n - \mathrm{tr}(X'X)^{-1}X'X\right) \quad [4] \\
&= \sigma^2(n - \mathrm{tr}\, I_k) \\
&= \sigma^2(n - k)
\end{aligned}
$$

นั่นคือ $E(\Sigma e_i^2) = \sigma^2(n-k)$ หรือ $E\left(\frac{\Sigma e_i^2}{n-k}\right) = \sigma^2$ [5]

ดังนั้น $\hat{\sigma}^2 = \frac{\Sigma e_i^2}{n-k}$ คือตัวประมาณค่าที่ปราศจากอคติของ σ^2 หรือเสนอ $\hat{\sigma}^2$ ในรูปเมตริกซ์ได้ดังนี้

$$\hat{\sigma}^2 = \frac{e'e}{n-k} = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n-k}$$

---

[1] เมตริกซ์ $M = \left(I_n - X(X'X)^{-1}X'\right)$ เป็น Idempotent Symmetric Matrix เราสามารถพิสูจน์ได้ว่า $M^2 = MM = M$ และ $M' = M$ นอกจากนี้ $\mathrm{tr}\, M = r(M)$

[2] ถ้า $E(u_iu_j) = 0$; $i \neq j$ และ $E(u_i^2) = \sigma^2$ เราสามารถพิสูจน์ได้โดยง่ายว่า $E(U'AU) = \sigma^2 \mathrm{tr}\, A$ ดังนี้

$$U'AU = (u_1, u_2, \ldots, u_n)\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \sum_i^n a_{ii}u_i^2 + 2\sum_{i<j} a_{ij}u_iu_j$$

ดังนั้น
$$
\begin{aligned}
E(U'AU) &= \sum_i^n a_{ii}E(u_i^2) + 2\sum_{i<j} a_{ij}E(u_iu_j) = \sum_i^n a_{ii}\sigma^2 + 0 \\
&= \sigma^2(a_{11} + a_{22} + \ldots + a_{nn}) = \sigma^2\, \mathrm{tr}\, A
\end{aligned}
$$

[3] $\mathrm{tr}(A \pm B) = \mathrm{tr}\, A \pm \mathrm{tr}\, B$

[4] $\mathrm{tr}(ABC) = \mathrm{tr}\, BCA = \mathrm{tr}(CAB)$

[5] $E\left(\frac{\Sigma e_i^2}{n-k}\right)$ คือ Expected Mean Square of Residual.

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'Y^{1}$$

$$= Y'Y - \hat{\beta}'X'Y$$

ดังนั้น $\hat{\sigma}^2 = \dfrac{Y'Y - \hat{\beta}'X'Y}{n-k} = \dfrac{\Sigma Y_i^2 - (\hat{\beta}_1 \Sigma Y_i + \hat{\beta}_2 \Sigma X_{2i} Y_i + \ldots + \hat{\beta}_k \Sigma X_{ki} Y_i)}{n-k}$

วิธีที่ 2 วิธี MLE

จาก $L = \dfrac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\dfrac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}\right\}$

แต่เนื่องจาก $E(\hat{\beta}) = \beta$ หรือ $\hat{\beta}$ เป็นตัวประมาณค่าที่ปราศจากอคติของ β

ดังนั้น $L = \dfrac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\dfrac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{2\sigma^2}\right\}$

$$\frac{\partial}{\partial\sigma} \ln L = \frac{\partial}{\partial\sigma}\left\{-\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{2\sigma^2}\right\} = 0$$

$$\Rightarrow \quad -\frac{n}{\sigma} + \frac{1}{\sigma^3}(Y - X\hat{\beta})'(Y - X\hat{\beta}) = 0$$

ดังนั้น $\hat{\sigma}^2 = \dfrac{1}{n}(Y - X\hat{\beta})'(Y - X\hat{\beta})$ คือ MLE ของ σ^2

แต่เมื่อพิจารณา $\hat{\sigma}^2$ จะพบว่า $\hat{\sigma}^2$ เป็น Biased Estimator ของ σ^2 กล่าวคือ

$$E(\hat{\sigma}^2) = \frac{1}{n} E(Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$= \frac{1}{n} E\{U'(I_n - X(X'X)^{-1}X')U\}$$

$$= \frac{(n-k)}{n}\sigma^2 \neq \sigma^2$$

แต่ $E\left(\dfrac{n}{n-k}\hat{\sigma}^2\right) = \sigma^2$ แสดงว่า $\dfrac{n\hat{\sigma}^2}{n-k}$ เป็น Unbiased Estimator ของ σ^2

ดังนั้น $\dfrac{n}{n-k}\left\{\dfrac{1}{n}(Y - X\hat{\beta})'(Y - X\hat{\beta})\right\} = \dfrac{1}{n-k}(Y - X\hat{\beta})'(Y - X\hat{\beta})$ คือ Unbiased Estimator ของ σ^2

---

[1] จาก Normal Equation (สมการ 2) คือ $(X'X)\hat{\beta} = X'Y$ ดังนั้น $\hat{\beta}'X'X\hat{\beta} = \hat{\beta}'X'Y$

3.2.2.5 Sampling Distribution ของ $\hat{\beta}$

การหาฟังก์ชันการแจกแจงของ $\hat{\beta}$ เมื่อ $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)'$ เป็นเรื่องที่เกี่ยวข้องอยู่กับ Multivariate Distribution โดยเฉพาะอย่างยิ่งก็คือ Multivariate Normal Distribution ดังนั้น ก่อนที่จะทำการศึกษาเรื่องการแจกแจงของ $\hat{\beta}$ ผู้เขียนขอแนะนำความรู้เกี่ยวกับ Multivariate Normal Distribution เสียก่อน ทั้งนี้จะไม่กล่าวถึงการพิสูจน์ นักศึกษาที่สนใจสามารถศึกษาเพิ่มเติมได้จากเอกสารอ้างอิง[1] และจะได้แนะนำการแจกแจงของ Quadratic Form ในโอกาสต่อ ๆ ไปเท่าที่เห็นว่าจำเป็นต้องอ้างอิง

ถ้า $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$ เป็นเวคเตอร์ของตัวแปรสุ่ม $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$ เป็นเวคเตอร์ของค่าเฉลี่ยของ

ตัวแปรสุ่ม $Y_1, Y_2, \ldots, Y_n$, $V = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$

เป็น Positive Definite Variance - Covariance Matrix ของตัวแปรสุ่ม $Y_1, Y_2, \ldots, Y_n$ และให้

$t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$ เป็นเวคเตอร์ของค่าคงที่ใด ๆ ดังนั้นเราจึงสามารถเสนอ Multivariate Normal Distribution ได้ดังนี้

$$f_{(x_1, x_2, \ldots, x_n)} = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left\{-\frac{(Y-\mu)'V^{-1}(Y-\mu)}{2}\right\};\ -\infty < Y_i < \infty;$$

$$i = 1, 2, \ldots, n$$

ตัวแปรสุ่ม Y มี mgf ดังนี้คือ[2] $M_Y(t_1, t_2, \ldots, t_n) = \exp\left\{t'\mu + \frac{t'Vt}{2}\right\}$

---

[1] อ่าน Hogg, R.W. and Craig, A.T., Introduction of Mathematical Statistic (The Macmilland Company, London, 1970) p. 379 - 394 และ Graybill, F.A., An Introduction to Linear Statistical Model, Vol I (McGraw-Hill Book Company, Inc., N.Y., 1961) P. 48 - 92.

[2] แสดงว่า $E(Y) = \mu$, $V(Y) = V$ ขอให้สังเกตว่า $\exp\left\{t'\mu + \frac{t'Vt}{2}\right\}$ สามารถกระจายออกมาในรูป Linear Form ของ t และ Quadratic Form ของ t ได้ดังนี้

$$\exp\left\{t'\mu + \frac{t'Vt}{2}\right\} = \exp\left\{\Sigma t_i\mu_i + (\Sigma t_i^2\sigma_{ii}^2 + \Sigma\Sigma t_it_j\sigma_{ij})/2\right\}$$

ในขณะเดียวกัน เราสามารถหาการแจกแจงของฟังก์ชันของตัวแปรสุ่ม Y คือ W = GY

โดยที่ $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$ และ G เป็นเมตริกซ์ขนาด m x n ได้ดังนี้

เนื่องจาก $M_w(t_1, t_2, \ldots, t_m) = E(e^{t'w})$ และเนื่องจาก W = GY ดังนั้น

$\Rightarrow M_W(t_1, t_2, \ldots, t_m) = E(e^{t'GY})$ แสดงว่าในการหา $M_W(t_1, t_2, \ldots, t_m)$ นั้นเราเพียงแต่แทนที่ t' ของ $M_Y(t_1, t_2, \ldots, t_n)$ ด้วย $t'G$ เท่านั้นก็จะสามารถทราบ mgf ของ W ได้[1]

$$M_W(t_1, t_2, \ldots, t_m) = M_{GY}(t_1, t_2, \ldots, t_m)$$

$$= \exp\left\{ t'G\mu + \frac{(t'G)V(t'G)'}{2} \right\}$$

$$= \exp\left\{ t'(G\mu) + \frac{t'(GVG')t}{2} \right\}$$

แสดงว่า $W \sim N(G\mu, GVG')$

และเราสามารถหาฟังก์ชันการแจกแจงต่าง ๆ อันเป็นกรณีเฉพาะของ w ได้ดังนี้

กำหนดให้ $c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ และ $w = c\,Y = c_1Y_1 + c_2Y_2 + \ldots + c_nY_n$

หรือ W คือ Linear Function ของ Y เราจะพบว่า $M_W(t) = E(e^{tW}) = E(e^{tC'Y}) = M_Y(tC')$ แสดงว่าในการหา $M_W(t)$ นั้นเราเพียงแต่แทนที่ t' ใน $M_Y(t_1, t_2, \ldots, t_n) = \exp\left\{ t'\mu + \frac{t'Vt}{2} \right\}$ ด้วย tC'

ดังนั้น $M_W(t) = \exp\left\{ tC'\mu + \frac{(tC')V(tC')'}{2} \right\} = \exp\left\{ tC'\mu + \frac{t(C'VC)t}{2} \right\}$[2]

$$= \exp\left\{ t(C'\mu) + \frac{(C'VC)t^2}{2} \right\}$$

แสดงว่า $W \sim N(C'\mu, C'VC)$

---

[1]ในกรณี Univariate เช่น Normal Variate เราพบว่า $M_Y(t) = e^{\mu t + \sigma^2 t^2/2}$ ถ้าเราต้องการทราบ mgf ของ W = aY เราจะพบว่า $M_W(t) = M_{aY}(t) = M_Y(at)$ ซึ่งแสดงว่าเพียงแต่แทนที่ t ใน $M_Y(t)$ ด้วย at เราก็สามารถหา mgf ของ W ได้

[2]ในกรณีนี้ t เป็นตัวคงที่ (หรือเวคเตอร์ขนาด 1×1) ทั้งนี้เพราะ $W = C'Y = \Sigma c_iY_i$ เป็น Scalar

โดยอาศัยความรู้ดังที่กล่าวมานี้ เราสามารถหาการแจกแจงของ $\hat{\beta}$ ได้โดยง่ายดังนี้

เนื่องจาก $\hat{\beta} = (X'X)^{-1}X'Y$ ถ้าเรากำหนดให้ $(X'X)^{-1}X' = c$ จะพบว่า $\hat{\beta} = CY$ โดยที่ C คือเมตริกซ์ขนาด k x n

ดังนั้น $M_{\hat{\beta}}(t_1, t_2, \ldots, t_k) = \exp\left\{ t'(C\mu) + \frac{t'(CVC')t}{2} \right\}$ โดยที่ $\mu = E(Y) = X\beta$ และ

$$V = V(Y) = \sigma^2 I_n$$

ดังนั้น $M_{\hat{\beta}}(t) = \exp\left\{ t'(CX\beta) + \frac{t'(C\sigma^2 I_n C')t}{2} \right\} \Rightarrow \hat{\beta} \sim N(CX\beta, \sigma^2 C I_n C')$

พิจารณา $CX\beta$ จะพบว่า $CX\beta = [(X'X)^{-1}X']X\beta = \beta$ และพบว่า $\sigma^2 C I_n C' = \sigma^2[(X'X)^{-1}X'][(X'X)^{-1}X']' = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$

นั่นคือ $M_{\hat{\beta}}(t_1, t_2, \ldots, t_k) = \exp\left\{ t'\beta + \frac{t'\sigma^2(X'X)^{-1}t}{2} \right\}$

แสดงว่า $\hat{\beta} \sim N[\beta, \sigma^2(X'X)^{-1}]$

3.2.3 สัมประสิทธิแห่งการตัดสินใจ (Coefficient of Determination) : R^2

สัมประสิทธิการตัดสินใจ คือ ดัชนีที่ใช้วัดความแม่นยำ (Precision) ของสมการประมาณค่า โดยถือว่า ดัชนีดังกล่าวคืออัตราส่วนเปรียบเทียบของความผันแปรที่อธิบายได้ด้วยแบบจำลอง (Explained Variation) ต่อความผันแปรรวมทั้งหมดของตัวแปรสุ่ม Y หมายความว่า โดยปกติของงานวิเคราะห์ความถดถอยนั้นเรามีความสนใจในธรรมชาติของตัวแปร Y และมีความปรารถนาที่จะทราบพฤติการณ์ของตัวแปรสุ่ม Y ในอนาคต แต่เนื่องจากตัวแปร Y เป็นตัวแปรสุ่ม ความผันผวนในพฤติการณ์ของ Y ย่อมมีสาเหตุมาจากแหล่งต่าง ๆ หลายแหล่งผสมผสานกัน วิธีในทางปฏิบัติของเราในกรณีเช่นนี้ก็คือให้พยายามแยกหรือดึงเอาสาเหตุที่ผสมผสานกันดังกล่าวแล้วออกมาแล้วนำเสนอไว้ในรูปของตัวแปรอิสระ X's ให้มากที่สุดเท่าที่จะมากได้ ด้วยเหตุนี้ส่วนหนึ่งของสาเหตุที่ทำให้ค่าของ Y ผันแปรไปได้จึงปรากฏให้เห็นได้เด่นชัดขึ้นและมองเห็นได้ว่าคืออะไรด้วยการเสนอสมการความสัมพันธ์นี้ไว้ในรูปของ Y = f(X's) เรียกตัวแปร X's เหล่านี้ว่า Explained Variable หรือIndependent Variable หรือ Regressor ดังนี้ส่วนหนึ่งของความผันแปรของ Y (Total Variation) จึงระบุได้ว่ามาจากสาเหตุแห่งความผันแปรที่มองเห็นได้ อธิบายได้ซึ่งก็คือความผันแปรอันเนื่องมาจากตัวแปรอิสระ X's เหล่านี้นี่เอง เรียกความผันแปรส่วนนี้ว่า

Explained Variation หรือ Variation due to Regression แต่เนื่องจากสาเหตุที่ผสมผสานกันแล้วมีผลให้ตัวแปรสุ่ม Y มีค่าผันผวนไปนั้นมีจำนวนสาเหตุที่มากมายยากที่จะระบุได้ครบถ้วน จึงเป็นสิ่งที่เป็นไปไม่ได้ที่เราจะดึงสาเหตุเหล่านั้นออกมาให้ปรากฏตัวได้ครบ อาจเพราะคาดคิดไปไม่ถึงหรือเข้าใจผิดคิดว่ามิใช่สาเหตุ หรือไม่รู้ไม่เข้าใจ รวมตลอดไปถึงความบกพร่องทั้งมวลในการบันทึกข้อมูล รวมถึงสาเหตุภายในอื่น ๆ จึงทำให้สาเหตุแห่งความผันแปรของ Y บางส่วนตกค้างหรือหลงเหลืออยู่ สาเหตุเหล่านี้จะแอบส่งอิทธิพลต่อความผันผวนในค่าของ Y อย่างซ่อนเร้น โดยที่เราเองไม่อาจบ่งบอกระบุหรืออธิบายได้ว่าเป็นใครบ้าง ความผันแปรส่วนที่ซ่อนเร้นนี้เรียกว่า Unexplained Variation หรือ Error หรือ Residual ด้วยเหตุผลดังกล่าวเราจึงสามารถสรุปได้ว่า

ความผันแปรรวมของ Y = (ความผันแปรในส่วนที่เนื่องจาก X's) + (ความผันแปรที่แฝงเร้น)

หรือ Total Variation of Y = (Explained Variation of Y) + (Unexplained Variation of Y)

ซึ่งความจริงข้อนี้สามารถพิสูจน์ให้เห็นได้ ซึ่งจะได้แสดงการพิสูจน์ให้เห็นในลำดับต่อไป

จากข้อเท็จจริงดังกล่าวจะเห็นได้ว่า ความผันแปรของตัวแปรสุ่ม Y เกิดขึ้นจากผลรวมของความผันแปร 2 ส่วนคือความผันแปรจากแหล่งที่ระบุหรือมองเห็นได้กับความผันแปรจากแหล่งที่ไม่อาจระบุได้หรือมองเห็นไม่ได้หรืออธิบายไม่ได้ว่ามาจากสาเหตุใด และเนื่องจากการประกอบกันของแหล่งความผันแปรทั้งสองนี้มีผลให้ได้ความผันแปรรวมซึ่งเป็นปริมาณคงที่ สิ่งที่จะเป็นเครื่องบ่งชี้ถึงคุณภาพของงาน และความแม่นยำของงานก็คือความสามารถในการดึงเอาปัจจัยที่อธิบายความผันผวนของ Y มาแสดงไว้อย่างครบถ้วน หมายความว่า ถ้างานใดผู้วิจัยสามารถดึงเอาปัจจัยดังกล่าวมาแสดงไว้ครบถ้วนมากเพียงใด งานนั้นก็จะมีคุณภาพสูงมากเพียงนั้น นั่นคือ ในทางปฏิบัตินักวิจัยพึงลดจำนวนความผันแปรที่แฝงเร้นให้มากที่สุดเท่าที่จะมากได้ โดยการพยายามเพิ่มปริมาณความผันแปรในส่วนที่อธิบายได้ให้สูงขึ้นซึ่งมีวิธีปฏิบัติหลายวิธี เช่น พยายามระบุตัวแปรอิสระให้ครบถ้วน ระบุตัวแปรอิสระที่มีความสำคัญและมีอิทธิพลต่อ Y อย่างแท้จริง ควบคุมขั้นตอนการปฏิบัติงานวิจัยตั้งแต่วางแผนเตรียมการ การบันทึกข้อมูลเรื่อยมาจนถึงการเสนอผลให้รัดกุม ฯลฯ ซึ่งล้วนมีผลให้เราสามารถเพิ่มความผันแปรในส่วนที่อธิบายได้ให้สูงขึ้น และลดปริมาณของความผันแปรในส่วนที่ซ่อนเร้นได้ขณะเดียวกัน

แต่เนื่องจากการวัดค่าความแม่นยำหรือคุณภาพของงานโดยเสนอผลในรูปของปริมาณเดี่ยว (Absolute Quantity) เป็นสิ่งที่เข้าใจได้ยาก วิธีที่จะทำให้เข้าใจได้ง่ายและเข้าใจได้ทันที

ผู้วิจัยควรใช้ปริมาณเปรียบเทียบ (Relative Quantity) โดยเสนอคุณภาพของงานในลักษณะต่อไปนี้คือ

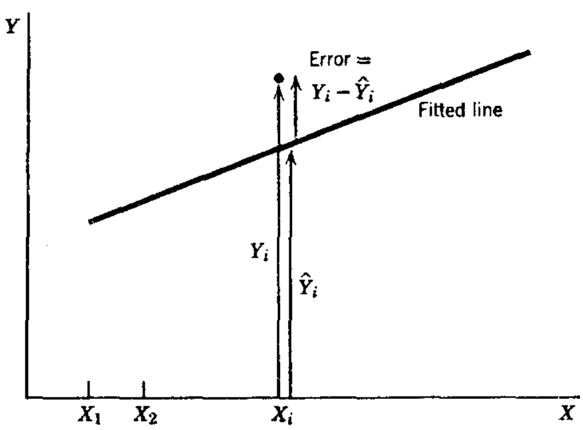
$$\text{ส.ป.ส.การตัดสินใจ} = \frac{\text{Explained Variation}}{\text{Total Variation}} \text{ หรือ } \frac{\text{Explained Variation}}{\text{Total Variation}} \times 100\%$$

$$\text{หรือ} \quad = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}} \text{ หรือ } (1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}) \times 100\%$$

จะเห็นได้ว่าถ้า Explained Variation มีค่าสูงขึ้น อัตราส่วนเปรียบเทียบจะมีค่าสูงขึ้น และสูงที่สุดเท่ากับ 1 ถ้า Explained Variation มีค่าเท่ากับ Total Variation ซึ่งมีความหมายว่าแบบจำลองที่เรากำหนดตัวแปรอิสระขึ้นมาเพื่ออธิบายความผันผวนในค่าของ Y นั้น สามารถอธิบายความผันผวนดังกล่าวได้ครบถ้วน และแสดงให้เห็นได้ในขณะเดียวกันได้ว่า Unexplained Variation มีค่าเท่ากับ 0 ในทำนองเดียวกันถ้า Explained Variation มีค่าต่ำลง อัตราส่วนเปรียบเทียบ Explained Variation/Total Variation จะมีค่าต่ำลง มีผลให้ Unexplained Variation/Total Variation มีค่าสูงขึ้น และ ส.ป.ส. การตัดสินใจจะมีค่าต่ำที่สุดถ้า Unexplained Variation มีค่าเท่ากับ Total Variation ซึ่งแสดงให้เห็นว่า งานประมาณค่าแบบจำลองดังกล่าวล้มเหลวโดยสิ้นเชิง และจากที่กล่าวมาแล้วทั้งหมดนี้แสดงให้เห็นว่า $0 \leqslant$ ส.ป.ส. การตัดสินใจ $\leqslant 1$

ต่อไปนี้จะแสดงการพิสูจน์ให้เห็นความเป็นจริงข้างต้นโดยหลักเกณฑ์ทางคณิตศาสตร์ ทั้งนี้จะเริ่มจากกรณี Simple Regression คือ กรณีของสมการ $Y = \alpha + \beta X + u$ เพราะเข้าใจได้ง่าย จากนั้นจึงค่อยขยายสู่กรณีทั่วไปในภายหลัง

พิจารณาสมการ $Y = \hat{\alpha} + \hat{\beta}X$ ซึ่งเป็นสมการที่ใช้ประมาณค่าสมการ $E(Y) = \alpha + \beta X$



จากภาพเมื่อกำหนดให้ $P_i(X_i, Y_i)$ คือค่าสังเกตชุดใด ๆ ที่วัดค่า Y_i ณ ค่า $X = X_i$ จะเห็นได้ว่าจุด P_i มิได้วางอยู่บน Regression Line ดังนั้นค่าของ Y_i ซึ่งคาดหมายได้ด้วย $\hat{Y}_i$ จึงประมาณได้จากสมการถดถอย $\hat{Y} = \hat{\alpha}+\hat{\beta}X$ ระยะที่ $\hat{Y}_i$ ประมาณค่า Y_i คลาดเคลื่อนไปจากความจริงคือ ระยะ e_i ด้วยเหตุนี้เราจึงพบว่า ณ ค่า $X = X_i$ นั้นจะส่งผลให้

$Y_i = \hat{Y}_i + e$ (เทียบต่อแกนเดิมที่มีจุดกำเนิดเป็น (0, 0))

หรือ $y_i = \hat{y}_i + e$ (เทียบต่อแกนใหม่ที่มีจุดกำเนิดเป็น $(\bar{X}, \bar{Y})$)

จากสมการ $y_i = \hat{y}_i + e_i$ เมื่อเรายกกำลังสองโดยตลอดและรวมตลอดในทุกค่าของ i

จะพบว่า $\Sigma y_i^2 = \Sigma \hat{y}_i^2 + \Sigma e_i^2 + 2\Sigma e_i\hat{y}_i$

พิจารณา $\Sigma e_i\hat{y}_i$ จะพบว่า

$$\Sigma e_i\hat{y}_i = \Sigma(y_i-\hat{y}_i)\hat{y}_i = \Sigma y_i\hat{y}_i - \Sigma\hat{y}_i^2 = \Sigma y_i(\hat{\beta}x_i) - \Sigma(\hat{\beta}x_i)^2$$

$$= \hat{\beta}\Sigma x_iy_i - \hat{\beta}^2\Sigma x_i^2 = \hat{\beta}\Sigma x_iy_i - \hat{\beta}\frac{\Sigma x_iy_i}{\Sigma x_i^2}\cdot\Sigma x_i^2$$

ดังนั้น $\Sigma e_i\hat{y}_i = \hat{\beta}\Sigma x_iy_i - \hat{\beta}\Sigma x_iy_i = 0$

ดังนั้น $\Sigma y_i^2 = \Sigma\hat{y}_i^2 + \Sigma e_i^2$ หรือ (Total Variation) = (Explained Variation) + (Unexplained Variation) หรือ (Total SS) = (Explained SS) + (Unexplained SS)[1]

จากสมการ $\Sigma y_i^2 = \Sigma\hat{y}_i^2 + \Sigma e_i^2$ เมื่อนำ Σy_i^2 หารตลอดจะพบว่า

$$1 = \frac{\Sigma\hat{y}_i^2}{\Sigma y_i^2} + \frac{\Sigma e_i^2}{\Sigma y_i^2}$$

$$=> \quad \frac{\Sigma\hat{y}_i^2}{\Sigma y_i^2} = 1 - \frac{\Sigma e_i^2}{\Sigma y_i^2}$$

ดังนั้น ส.ป.ส. การตัดสินใจก็คือ $\frac{\Sigma\hat{y}_i^2}{\Sigma y_i^2}$ หรือ $1 - \frac{\Sigma e_i^2}{\Sigma y_i^2}$

พิจารณา $\frac{\Sigma\hat{y}_i^2}{\Sigma y_i^2}$ จะพบว่า $\frac{\Sigma\hat{y}_i^2}{\Sigma y_i^2} = \frac{\hat{\beta}^2\Sigma x_i^2}{\Sigma y_i^2} = \left(\frac{\Sigma x_iy_i}{\Sigma x_i^2}\right)^2\frac{\Sigma x_i^2}{\Sigma y_i^2}$

$$= \frac{(\Sigma x_iy_i)^2}{\Sigma x_i^2\Sigma y_i^2} = \left(\frac{\Sigma x_iy_i}{\sqrt{\Sigma x_i^2\Sigma y_i^2}}\right)^2 = r^2$$

---

[1] SS ย่อมาจาก Sum of Square ขอให้สังเกตว่า $\Sigma y_i^2 = \Sigma(Y_i-\bar{Y})^2$ และ $\Sigma\hat{y}_i^2 = \Sigma(\hat{Y}_i-\bar{Y})^2$ สำหรับ Σe_i^2 เราอาจเรียกได้หลายชื่อ เช่น Residual SS, Error SS หรือ Unexplained SS

แสดงว่า ส.ป.ส. การตัดสินใจก็คือ (ส.ป.ส. สหสัมพันธ์)^2 นั่นเอง ด้วยเหตุนี้เราจึงถือว่า r^2 เป็นดัชนีช่วยตัดสินใจว่าสมการถดถอยมีความแม่นยำหรือมีคุณภาพสูงเพียงใด และเมื่อพิจารณาสมการ $r^2 = 1 - \frac{\Sigma e_i^2}{\Sigma y_i^2}$ จะพบว่าในกรณีที่สมการถดถอยมีคุณภาพต่ำนั้น Unexplained Variation จะมีค่าสูงมากและจะมีคุณภาพต่ำที่สุดถ้า $\Sigma e_i^2 = \Sigma y_i^2$ ซึ่งมีผลให้ $r^2 = 1-1 = 0$ ขณะเดียวกัน ถ้าสมการถดถอยมีคุณภาพสูง Σe_i^2 จะมีค่าต่ำและจะมีคุณภาพสูงที่สุดถ้า Unexplained Variation หรือ Σe_i^2 มีค่าเท่ากับ 0 ซึ่งมีผลให้ $r^2 = 1-0 = 1$ ด้วยเหตุนี้สัมประสิทธิการตัดสินใจจึงมีค่าอยู่ระหว่าง 0 ถึง 1 หรือ $0 \leqslant r^2 \leqslant 1$

สำหรับในกรณีทั่วไปคือ กรณี $Y = X\beta + U$ ซึ่งประมาณได้ด้วยความสัมพันธ์ $Y = X\hat{\beta} + e$ จะพบว่า[1]

ให้ $\hat{Y} = X\hat{\beta}$ คือ fitted vector ของ Y นั่นคือ $e = Y - \hat{Y}$

$$\text{ดังนั้น } \Sigma Y_i^2 = Y'Y = (\hat{Y} + Y - \hat{Y})'(\hat{Y} + Y - \hat{Y}) = (\hat{Y} + e)'(\hat{Y} + e)$$

$$= \hat{Y}'\hat{Y} + 2\hat{Y}'e + e'e$$

$$\text{พิจารณา } \hat{Y}'e \quad \text{พบว่า} \quad \hat{Y}'e = (X\hat{\beta})'e = \hat{\beta}'(X'e)$$

$$\text{แต่ } X'e = X'[Y - \hat{Y}] = X'[Y - X\hat{\beta}] = X'[Y - X(X'X)^{-1}X'Y]$$

$$= X'[I - X(X'X)^{-1}X']Y = X'Y - X'Y = 0$$

$$\text{ดังนั้น } Y'Y = \hat{Y}'\hat{Y} + e'e = (X\hat{\beta})'(X\hat{\beta}) + e'e = \hat{\beta}'X'Y + e'e$$

ดังนั้น เมื่อนำ Correction Factor คือ nY^2 หักลบตลอดทั้งด้านซ้ายและด้านขวา จะพบว่า

$$Y'Y - nY^2 = (\hat{\beta}'X'Y - nY^2) + e'e$$

นั่นคือ (Total Sum Square) = (Explained Sum square) + (Unexplained Sum Square)

---

[1] ในทางทฤษฎีนั้นเราจะต้องถือว่า $Y = X\beta + U$ โดยที่ U เป็นตัวแปรสุ่มที่ผนวกเอาความผิดพลาดทั้งมวลไว้ เรียกความสัมพันธ์นี้ว่า Functional Form และถ้าจะนำไปใช้ในทางปฏิบัติเราจะใช้เฉพาะ True Surface คือ $Y = X\beta$ โดยที่ Y ในที่นี้ก็คือ E(Y) นั่นเอง เรียกความสัมพันธ์นี้ว่า Basic Form ขณะเดียวกัน เมื่อเราประมาณค่า β ได้ด้วย $\hat{\beta}$ รูปสมการทางทฤษฎีก็คือ $Y = X\hat{\beta} + e$ โดยที่ e คือค่าประมาณของ U เมื่อ $e = Y - \hat{Y} = Y - X\hat{\beta}$ และในทางปฏิบัติเราจะใช้เฉพาะ $Y = X\hat{\beta}$ หรือ $\hat{Y} = X\hat{\beta}$

และพบว่า $R^2 = \frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} = \frac{\text{Regression SS}}{\text{Total SS}}$

พิจารณา $R^2 = \frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2}$

จะพบว่า $R^2 = \frac{\hat{\beta}_2 \Sigma X_{2i}Y_i + \hat{\beta}_3 \Sigma X_{3i}Y_i + \ldots + \hat{\beta}_k \Sigma X_{ki}Y_i}{\Sigma y_i^2}$

จะเห็นได้ว่าเศษจะมีค่าเพิ่มขึ้นเสมอในทุกครั้งที่เราเพิ่มจำนวนตัวแปรอิสระ X เข้าสู่แบบจำลอง หรือนัยหนึ่ง ถ้า k มีค่าสูงขึ้นเศษของ R^2 จะสูงขึ้นขณะที่ส่วนคือ Σy_i^2 หรือ $Y'Y - n\bar{Y}^2$ มีค่าคงที่ นั่นหมายความว่าค่าของ R^2 จะเพิ่มขึ้นเสมอในทุกครั้งที่เราเพิ่มตัวแปรอิสระ X ที่คาดคิดน่าจะเป็นปัจจัยที่ช่วยอธิบายความผันผวนในค่าของ Y ลงในแบบจำลอง เรื่องนี้เป็นสิ่งที่ต้องคิดเพราะความจริงแล้วตัวแปรอิสระที่เราคาดคิดว่าจะอธิบายความผันผวนของ Y ได้นั้นอาจไม่มีผลหรืออิทธิพลต่อ Y เลย และแม้แต่เรานำตัวแปร X ที่ไม่มีความหมายต่อ Y เลยเพิ่มเข้าไปในแบบจำลอง ค่าของ R^2 ก็จะยังคงสูงขึ้นเพราะโครงสร้างของสูตรเอื้ออำนวยให้เป็นเช่นนั้น ด้วยเหตุนี้สูตร $R^2 = \frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2}$ จึงยังไม่เหมาะเพราะมีช่องโหว่ให้เกิดความเข้าใจผิดและมีผลเสียหายถึงการคัดเลือกตัวแปรอิสระ X's ด้วย วิธีแก้ไขให้ R^2 มีความหมายที่ชัดเจนและเป็นดัชนีที่บ่งบอกอิทธิพลร่วมที่ X's มีต่อ Y อย่างแท้จริงก็คือ การถ่วงน้ำหนักทั้งเศษและส่วนด้วย degree of freedom โดยใช้สูตร R^2 ดังนี้

$$\bar{R}^2 = 1 - \frac{e'e/n-k}{\Sigma y_i^2/n-1}$$

$$= 1 - \frac{(Y'Y - \hat{\beta}'X'Y)/n-k}{(Y'Y - n\bar{Y}^2)/n-1}$$

$\bar{R}^2$ หรือ adjusted R^2 นี้จะมีค่าใกล้เคียงความจริงขึ้น เพราะในทุกครั้งที่เราเพิ่มตัวแปรอิสระ X เข้าสู่แบบจำลอง ค่าของ $(Y'Y - \hat{\beta}'X'Y)$ จะถูกถ่วงน้ำหนักด้วย $n-k$ เมื่อ k – เคือจำนวนตัวแปรอิสระ X's ค่าของ R^2 จึงไม่เพิ่มขึ้นอย่าง "พรวดพราด" แต่จะเพิ่มขึ้นอย่างสมเหตุสมผล ตัวแปร X ตัวใดที่ไม่มีอิทธิพลต่อความผันผวนในค่าของ Y เลยก็จะไม่มีผลให้ R^2 เพิ่มขึ้นจากเดิม หรือถ้าเพิ่มขึ้นก็จะเพิ่มขึ้นในอัตราส่วนค่อนข้างต่ำ ดังนั้นในทางปฏิบัติเราพึงเลือกใช้ $\bar{R}^2$ เป็นดัชนีการตัดสินใจเพราะเหมาะสมกว่า

อนึ่งเราสามารถพิสูจน์ได้ว่า R^2 ก็คือกำลังสองของสหสัมพันธ์เชิงเส้นระหว่าง Y กับ $\hat{Y}$ ขอให้ระลึกว่า multiple correlation คือสหสัมพันธ์ระหว่างตัวแปรสุ่มตัวหนึ่งกับตัวแปรสุ่มอีกตัวหนึ่ง ในที่นี้ก็คือระหว่าง Y กับ $\hat{Y}$ โดย $\hat{Y}$ คือตัวแทนของตัวแปรอิสระ X's ทั้งกลุ่ม การพิสูจน์ปรากฏดังนี้

พิสูจน์

$$\because r_{Y\hat{Y}} = \Sigma(Y_i - \hat{Y})(\hat{Y}_i - \hat{Y})/\sqrt{\Sigma(Y_i - \hat{Y})^2\Sigma(\hat{Y}_i - \hat{Y})^2}$$

แต่เนื่องจาก $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2X_{2i} + \hat{\beta}_3X_{3i} + \ldots + \hat{\beta}_kX_{ki}; i = 1,2,\ldots,n$ ....(1)

โดยที่ $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_1\bar{X}_1 - \hat{\beta}_2\bar{X} - \ldots - \hat{\beta}_j\bar{X}_k$ ....(2)

จากสมการ (1) เมื่อ sum และเฉลี่ยด้วยการหารด้วย n แล้วแทนค่า $\hat{\beta}_1$ ด้วยสมการที่ (2) จะพบว่า $\hat{Y} = \bar{Y}$

ดังนั้น

$$\begin{aligned}
\Sigma(Y_i - \bar{Y})(\hat{Y}_i - \hat{Y}) &= \Sigma(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y}) \\
&= \Sigma Y_i\hat{Y}_i - n\bar{Y}^2 \\
&= \hat{Y}'Y - n\bar{Y}^2 \\
&= \hat{\beta}'X'Y - n\bar{Y}^2 \\
\Sigma(Y_i - \bar{Y})^2 &= \Sigma Y_i^2 - n\bar{Y}^2 \\
&= Y'Y - n\bar{Y}^2 \\
\Sigma(\hat{Y}_i - \bar{Y})^2 &= \Sigma\hat{Y}_i^2 - n\bar{Y}^2 \\
&= \hat{\beta}'X'Y - n\bar{Y}^2
\end{aligned}$$

ดังนั้น

$$\begin{aligned}
r^2_{Y\hat{Y}} &= (\hat{\beta}'X'Y \quad n\bar{Y}^2)^2/(Y'Y - n\bar{Y}^2)(\hat{\beta}'X'Y - n\bar{Y}^2) \\
&\frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} \\
&= R^2
\end{aligned}$$

แสดงว่า R^2 สามารถคำนวณหาได้จากสหสัมพันธ์เชิงเส้นระหว่างเวคเตอร์ Y กับ $\hat{Y}$

หมายเหตุ R เรียกว่า Multiple Correlation Coefficient เขียนให้เต็มรูปได้เป็น $R_{y.x_2x_3...x_k}$ ใช้เป็นดัชนีวัดสหสัมพันธ์ร่วมระหว่าง ($X_2, X_3,... X_k$) กับ Y เรื่องของ Multiple Correlation นี้จะได้กล่าวถึงอีกครั้งหนึ่งในตอน 3.2.6

เกี่ยวกับความผันแปร $Y'Y = \hat{Y}'\hat{Y} + e'e$ สามารถอธิบายให้เห็นความจริงในรูปทรงทางเรขาคณิตได้ดังนี้

จาก $E(Y) = E(X\beta + u) = X\beta$

หรือ $E(Y) = \beta_1 1 + \beta_2 X_2 + \ ...... \ + \beta_k X_k$

โดยที่ $1 = (1,1,..., 1)'$ และ $X_j = (X_{j1}, X_{j2},..., X_{jn})'$; $j = 2,3,..., k$ เป็นเวคเตอร์ขนาด $n \times 1$

แสดงว่า E(Y) คือ linear combination ของสดมภ์ของแมตริกซ์ X ซึ่งในทุกครั้งที่สัมประสิทธิ์ $\beta_1, \beta_2, \ ...... \ , \beta_k$ เปลี่ยนค่าไปย่อมมีผลให้เกิดเวคเตอร์ E(Y) ชุดใหม่ซึ่งผลจากการเปลี่ยนค่า $\beta_1, \beta_2, \ ...... \ , \beta_k$ ทำให้เกิด regression plane ขึ้น

และด้วยเหตุที่ $Y = X\beta + u$ แสดงว่าเวคเตอร์ Y นั้นโดยปกติจะไม่ทาบอยู่บนระนาบ $E(Y) = X\beta$

ทุกครั้งเมื่อเรามีค่าสังเกต เราจะเริ่มประมาณค่า $\beta_1, \beta_2, \ ...... \ , \beta_k$ โดยวิธี least square เพื่อให้ได้ $\hat{\beta}_1, \hat{\beta}_2, \ ...... \ , \hat{\beta}_k$ ที่มีผลให้ระยะทางกำลังสองของ e (Square length) สั้นที่สุด

ระยะทางกำลังสองของ e คือ $\sum_{1}^{n} e_i^2 = e'e = (Y - X\hat{\beta})' \ (Y - X\hat{\beta})$

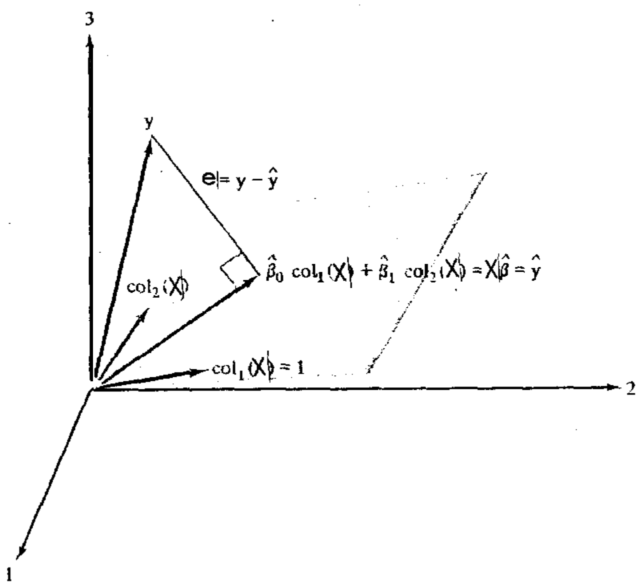
และเมื่อเราประมาณค่า $\beta_1, \beta_2, \ ...... \ , \beta_k$ ได้คือ $\hat{\beta}_1, \hat{\beta}_2, \ ...... \ , \hat{\beta}_k$ ค่าเหล่านี้คือสัมประสิทธิ์ของเวคเตอร์ 1, X2, X3,..., Xk ซึ่ง linear combination ของเวคเตอร์เหล่านี้จะสร้างให้เกิดเวคเตอร์ $\hat{Y}$ ขึ้น เวคเตอร์ $\hat{Y}$ จึงเป็นเวคเตอร์ในระนาบ เรียกว่า estimated plane

$$\hat{Y} = \hat{\beta}_1 1 + \hat{\beta}_2 X_2 + \ ...... \ + \hat{\beta}_k X_k$$

ขอให้พิจารณาภาพในหน้าต่อไปซึ่งแสดงให้เห็นเฉพาะกรณีของ Simple linear regression ระนาบที่แลเงาในภาพต่อไปนี้คือระนาบ $\hat{Y} = \hat{\beta}_1 1 + \hat{\beta}_2 X_2$

ขอให้สังเกตว่าเวคเตอร์ $\hat{Y}$ จะอยู่ในระนาบเพราะเวคเตอร์ $\hat{Y}$ เป็น linear combination ของเวคเตอร์ 1 และ X (ในภาพเขียนนั้น col1(X) และ col2(X)) แต่เวคเตอร์ Y ไม่แน่ว่าจะอยู่ในระนาบหรือไม่ ถ้าเวคเตอร์ Y อยู่ในระนาบ (คือระนาบพุ่งไปทาบเวคเตอร์ Y) ก็แปลว่า $\hat{Y}$ =

$\hat{Y}$ คือระยะ e ระยะ e จะตั้งฉากกับ $\hat{Y}$ เสมอเพราะ $\hat{Y}'e = 0$ ซึ่งในทางทฤษฎีเราจะพยายามประมาณค่า $\hat{\beta}_1$ และ $\hat{\beta}_2$ เพื่อให้ได้ค่าที่เหมาะสมที่สุด ค่าที่เหมาะสมที่สุดคือ $\hat{\beta}_1$ และ $\hat{\beta}_2$ ที่ทำให้ระยะ e สั้นที่สุด



โดยอาศัยทฤษฎีสามเหลี่ยมมุมฉากของปิธากอรัสเราทราบว่า $Y'Y = \hat{Y}'\hat{Y} + e'e$

หรือ (ความยาว Y)^2 = (ความยาว $\hat{Y}$)^2 + (ความยาว e)^2

แต่(ความยาว e)^2 หรือ Σe_i^2 (นิยามความยาวของเวคเตอร์ เช่น e คือ $\sqrt{e'e}$ ดังนั้น e'e ก็คือ (ความยาว e)^2)จะสั้นที่สุดได้เมื่อเราประมาณค่า $\hat{\beta}_1$ และ $\hat{\beta}_2$ ได้อย่างเหมาะสมจนทำให้ระนาบ $\hat{Y} = \hat{\beta}_1 1 + \hat{\beta}_2 X$ ลอยใกล้เวคเตอร์ Y และ (ระยะทาง e)^2 จะสั้นที่สุดเมื่อระนาบ $\hat{Y} = \hat{\beta}_1 1 + \hat{\beta}_2 X$ ทาบเวคเตอร์ Y พอดี

อนึ่ง R^2 ซึ่งเราสามารถพิสูจน์ได้ว่ามีค่าเท่ากับ $(\hat{\beta}' X' Y - n\bar{Y}^2) \;/\; (Y' Y - n\bar{Y}^2)$ นั้นความจริงแล้วก็คือ (Simple correlation ระหว่างตัวแปร Y และ $\hat{Y}$)^2 นั่นเอง เรื่องนี้สามารถพิสูจน์ได้ดังนี้

$$r_{Y\hat{Y}} = \frac{\Sigma(Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{\sqrt{\Sigma(Y_i - \bar{Y})^2 \; \Sigma(\hat{Y}_i - \hat{Y})^2}}$$

โดยที่ $Y = (Y_1, Y_2, \dots, Y_n)'$ และ $\hat{Y} = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)'$ หรือ

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_k \;;\; i = 1, 2, \dots, n$$

โดยที่ $\hat{\beta}_1 = \bar{Y} - \beta_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3 - \dots - \hat{\beta}_k\bar{X}_k$

ดังนั้น $\bar{\hat{Y}} = (\bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3 - \dots - \hat{\beta}_k\bar{X}_k + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \dots + \hat{\beta}_k X_k = \bar{Y}$

นั่นคือ $r_{Y\hat{Y}} = \dfrac{\Sigma(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{\Sigma(Y_i - \bar{Y})^2 \, \Sigma(\hat{Y}_i - \bar{Y})^2}} = \dfrac{a}{bc}$

ซึ่งสามารถพิสูจน์ได้ว่า

$$a = \Sigma\hat{Y}_i Y_i - n\bar{Y}^2 = \hat{Y}'Y - n\bar{Y}^2 = \hat{\beta}'X'Y - n\bar{Y}^2$$

$$b = \sqrt{\Sigma Y_i^2 - n\bar{Y}^2} = \sqrt{Y'Y - n\bar{Y}^2}$$

$$c = \sqrt{\Sigma\hat{Y}_i^2 - n\bar{Y}^2} = \sqrt{Y'Y - n\bar{Y}^2} = \sqrt{\hat{\beta}'X'Y - n\bar{Y}^2}$$

ดังนั้น $r_{Y\hat{Y}} = \dfrac{a}{bc} = \dfrac{\hat{\beta}'X'Y - n\bar{Y}^2}{\sqrt{Y'Y - n\bar{Y}^2}}$ และ R^2 ก็คือ $r^2_{Y\hat{Y}}$

3.2.4 กรณีเฉพาะเมื่อ $k = 2$ หรือกรณีเฉพาะเมื่อ **General Linear Regression** ลดรูปลงสู่รูปของ **Simple Linear Regression**

3.2.4.1 การประมาณค่าพารามิเตอร์

จากสมการ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$; $i = 1, 2, \ldots, n$ เมื่อ $k = 2$ สมการรูปทั่วไปนี้จะลดรูปเป็น

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \ ; \ i = 1, 2, \ldots, n$$

และเพื่อให้สอดคล้องกับรูปที่แสดงไว้ในบทที่ 2 จึงกำหนดให้ $\beta_1 = \alpha$ และ $\beta_2 = \beta$ ดังนั้น สมการถดถอยเชิงเส้นกรณี $k = 2$ จึงเสนอได้ดังนี้คือ

$$Y_i = \alpha + \beta X_i + u_i \ ; \ i = 1, 2, \ldots, n$$

ดังนั้น $X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$, $X'X = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ X_1 & X_2 & X_3 & \ldots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}$

$$(X'X)^{-1} = \frac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ X_1 & X_2 & X_3 & \ldots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}$$

ดังนั้นเวคเตอร์ $\hat{\beta} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (X'X)^{-1}X'Y$

$$= \frac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix} \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}$$

=> $$\hat{\alpha} = \frac{\Sigma X_i^2 \Sigma Y_i - \Sigma X_i \Sigma X_i Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2}$$

$$\hat{\beta} = \frac{n\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2}$$

จัดรูปเสียใหม่เพื่อให้เป็นรูปง่ายจะพบว่า

$$\hat{\beta} = \frac{n\Sigma X_i Y_i - \Sigma X_i \Sigma Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} = \frac{\Sigma X_i Y_i - \frac{\Sigma X_i \Sigma Y_i}{n}}{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}}$$

$$= \frac{\Sigma (X_i - \overline{X})(Y_i - \overline{Y})^1}{\Sigma (X_i - \overline{X})^2} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

เมื่อ $x_i = (X_i - \overline{X})$, $y_i = (Y_i - \overline{Y})$

และ $$\hat{\alpha} = \frac{\Sigma X_i^2 \Sigma Y_i - \Sigma X_i \Sigma X_i Y_i}{n\, \Sigma Xf\text{-} (\Sigma X_i)^2}$$

$$= \frac{\Sigma X_i^2 \Sigma Y_i - \frac{(\Sigma X_i)^2 \Sigma Y_i}{n} - \Sigma X_i \Sigma X_i Y_i + \frac{(\Sigma X_i)^2 \Sigma Y_i}{n}\ ^2}{n\left\{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right\}}$$

$$= \frac{\Sigma Y_i \left\{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right\} - \Sigma X_i \left\{\Sigma X_i Y_i - \frac{\Sigma X_i \Sigma Y_i}{n}\right\}}{n\left\{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right\}}$$

$$= \frac{\Sigma Y_i (\Sigma x_i^2) - \Sigma X_i (\Sigma x_i y_i)}{n\Sigma x_i^2}$$

---

[1] $\Sigma(X_i - \overline{X})(Y_i - \overline{Y}) = \Sigma(X_i Y_i - \overline{Y}X_i - \overline{X}Y_i + \overline{X}\overline{Y})$

$= \Sigma X_i Y_i - n\overline{X}\overline{Y} - n\overline{X}\overline{Y} + n\overline{X}\overline{Y}$

$= \Sigma X_i Y_i - \frac{\Sigma X_i \Sigma Y_i}{n}$

[2] บวกเข้าและลบออกด้วย $\frac{(\Sigma X_i)^2 \Sigma Y_i}{n}$

$$= \bar{Y} - \hat{\beta}\bar{X}$$

นั่นคือ ในกรณีที่ k = 2 เราจะพบว่า $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$, $\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$

3.2.4.2 การประมาณค่า $V(\hat{\beta})$

จากสูตร ๓ $\hat{V}(\hat{\beta}) = \hat{\sigma}^2 \frac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 & \Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}$

$$= \frac{\hat{\sigma}^2}{n\Sigma x_i^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\hat{\sigma}^2 \Sigma X_i^2}{n\Sigma x_i^2} & \frac{-\hat{\sigma}^2 \Sigma X_i}{n\Sigma x_i^2} \\ \frac{-\hat{\sigma}^2 \Sigma X_i}{n\Sigma x_i^2} & \frac{n\hat{\sigma}^2}{n\Sigma x_i^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\hat{\sigma}^2 \Sigma X_i^2}{n\Sigma x_i^2} & \frac{-\hat{\sigma}^2 \bar{X}}{\Sigma x_i^2} \\ \frac{-\hat{\sigma}^2 \bar{X}}{\Sigma x_i^2} & \frac{\hat{\sigma}^2}{\Sigma x_i^2} \end{bmatrix}$$

นั่นคือ $\hat{V}(\hat{\alpha}) = \frac{\hat{\sigma}^2 \Sigma X_i^2}{n\Sigma x_i^2}$, $\hat{V}(\hat{\beta}) = \frac{\hat{\sigma}^2}{\Sigma x_i^2}$ และ $Cov(\hat{\alpha}, \hat{\beta}) = \frac{-\hat{\sigma}^2 \bar{X}^2}{\Sigma x_i^2}$

โดยที่ $\hat{\sigma}^2 = \frac{\Sigma e_i^2}{n-2}$

3.2.4.3 Sampling Distribution ของ $\hat{\alpha}$ และ $\hat{\beta}$

การมีความรู้ความเข้าใจเกี่ยวกับ Sampling Distribution ของ $\hat{\alpha}$ และ $\hat{\beta}$ จะมีส่วนอย่างสำคัญที่จะช่วยให้นักศึกษาสามารถสร้างช่วงเชื่อมั่นของ α และ β รวมถึงการสร้างเขตเชื่อมั่นของ (α, β) และการทดสอบสมมุติฐานของ α และ β ในที่นี้จะกล่าวถึงเฉพาะการนำประโยชน์ไปใช้ในประเด็นดังกล่าวโดยไม่เสนอวิธีการสร้างเขตเชื่อมั่น และวิธีพัฒนา Test Statistics นักศึกษาจะสามารถสร้างเขตเชื่อมั่นและพัฒนา Test Statistics ได้เองเมื่อเรียนวิชาทฤษฎีสถิติ 1 - 2

ก. Sampling Distribution ของ $\hat{\beta}$

จาก $$\hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$

ให้ $$\frac{x_i}{\Sigma x_i^2} = w_i$$

ดังนั้น $$\hat{\beta} = \Sigma w_i y_i = \Sigma w_i(Y_i - \bar{Y}) = \Sigma w_i Y_i - \bar{Y}\Sigma w_i$$
$$= \Sigma w_i Y_i{}^1$$

แต่เนื่องจาก $Y_i = \alpha + \beta X_i + u_i$ และ $u_i \sim N(0, \sigma^2)$

$$\Rightarrow \quad M_{Y_i}(t) = e^{(\alpha+\beta X_i)t} M_{u_i}(t)$$
$$= e^{(\alpha+\beta X_i)t} e^{\sigma^2 t^2/2}$$
$$= e^{(\alpha+\beta X_i)t+\sigma^2 t^2/2}$$

แสดงว่า $Y_i \sim N(\alpha + \beta X_i, \sigma^2)$

ดังนั้น $$M_{\hat{\beta}}(t) = M_{\Sigma w_i Y_i}(t) = \prod_i^n M_{Y_i}(w_i t)$$
$$= \prod_i^n \{ e^{(\alpha+\beta X_i)w_i t+\sigma^2 (w_i t)^2/2} \}$$
$$= e^{t\Sigma(\alpha+\beta X_i)w_i + (\Sigma \sigma^2 w_i^2)t^2/2}$$

พิจารณา $\Sigma(\alpha + \beta X_i)w_i$ จะพบว่า

$$\Sigma(\alpha + \beta X_i)w_i = \alpha\Sigma w_i + \beta\Sigma X_i w_i{}^2$$
$$= 0 + \beta$$
$$= \beta$$

---

[1] $\bar{Y}\Sigma w_i = \bar{Y}\Sigma\left(\frac{x_i}{\Sigma x_i^2}\right) = \frac{\bar{Y}}{\Sigma x_i^2}(\Sigma x_i) = \frac{\bar{Y}}{\Sigma x_i^2}\{\Sigma(X_i - \bar{X})\} = 0$ ทั้งนี้ขอให้ทำความเข้าใจไว้ว่า Σx_i^2 เป็นค่าคงที่ ในขณะที่ x_i^2 มิใช่ค่าคงที่เพราะค่าของ x_i^2 จะแปรค่าไปเรื่อย ๆ ตาม running index (i) ซึ่งเมื่อรวมตลอดค่า x_i^2 ทุกค่าตั้งแต่ 1 ถึง n ผลรวมจึงเป็นตัวคงที่

[2] $\Sigma w_i = \Sigma\left(\frac{x_i}{\Sigma x_i^2}\right) = \frac{1}{\Sigma x_i^2}\cdot\Sigma x_i = \frac{1}{\Sigma x_i^2}\cdot\Sigma(X_i - \bar{X}) = 0$

$\Sigma X_i w_i = \Sigma(X_i - \bar{X})w_i = \Sigma x_i w_i = \Sigma x_i\left(\frac{x_i}{\Sigma x_i^2}\right) = \left(\frac{1}{\Sigma x_i^2}\right)\Sigma x_i x_i = 1$

และเมื่อพิจารณา $\Sigma\sigma^2 w_i^2$ จะพบว่า

$$\Sigma\sigma^2 w_i^2 = \sigma^2\Sigma\left(\frac{x_i}{\Sigma x_i^2}\right)^2 = \frac{\sigma^2}{(\Sigma x_i^2)^2}\cdot\Sigma x_i^2 = \frac{\sigma^2}{\Sigma x_i^2}$$

ดังนั้น $M_{\hat{\beta}}(t) = e^{\beta t+(\sigma^2/\Sigma x_i^2)t^2/2}$ ซึ่งแสดงว่า $\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\Sigma x_i^2})$ หรือนัยหนึ่ง $\hat{\beta}$ เป็นตัวแปรสุ่ม[1] ที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ β มีค่าความแปรปรวนเท่ากับ $\frac{\sigma^2}{\Sigma x_i^2}$

เมื่อประมาณค่า σ^2 ด้วย $\hat{\sigma}^2 = \frac{\Sigma e_i^2}{n-2}$ เราจะพบว่า $\frac{\hat{\beta}-\beta}{\hat{\sigma}\sqrt{1/\Sigma x_i^2}} \sim t_{n-2}$ ซึ่งเราสามารถสามารถนำตัวสถิติ $t = \frac{\hat{\beta}-\beta}{\hat{\sigma}\sqrt{1/\Sigma x_i^2}}$ นี้ไปใช้ประโยชน์ดังนี้

1) การสร้างช่วงเชื่อมั่นของ β

โดยอาศัยทฤษฎีการประมาณค่าด้วยช่วงเชื่อมั่น (Interval Estimation Theory) จะพบว่า

$$\Pr\left\{-t_{n-2,\,1-\frac{\alpha}{2}} < \frac{\hat{\beta}-\beta}{\hat{\sigma}\sqrt{1/\Sigma x_i^2}} < t_{n-2,\,1-\frac{\alpha}{2}}\right\} = 1-\alpha$$

นั่นคือช่วงเชื่อมั่น $(1-\alpha)$ 100% ที่เชื่อว่าจะครอบคลุมเอาค่าจริงของ β ไว้คือช่วง

$$\left(\hat{\beta} - t_{n-2,\,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{1}{\Sigma x_i^2}},\ \hat{\beta} + t_{n-2,\,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{1}{\Sigma x_i^2}}\right)$$

2) การทดสอบสมมุติฐานเกี่ยวกับ β

โดยอาศัย Maximum Likelihood Ratio Test (MLRT)[2] เราสามารถพัฒนาตัวทดสอบ (Test Statistics) สำหรับ $H_0 : \beta = 0$ VS $H_1 : \beta \neq 0$ ได้ดังนี้[3]

---

[1] ทั้งนี้เพราะค่าประมาณของ β จะมีค่าเปลี่ยนแปลงไปเรื่อย ๆ ตามกลุ่มตัวอย่าง

[2] อ่านมนตรี พิริยะกุล **ทฤษฎีสถิติ 2** (โรงพิมพ์วิคตอรี, กรุงเทพมหานคร พ.ศ. 2525) บทที่ 5-6

[3] ถ้าปฏิเสธสมมุติฐานหลักก็แสดงว่า $\beta \neq 0$ หรือตัวแปร X ยังคงปรากฏอยู่ในแบบจำลอง และมีอิทธิพลต่อ Y ถ้าไม่อาจปฏิเสธ H_o แสดงว่า $\beta = 0$ แสดงว่า X ไม่ควรปรากฏอยู่ในแบบจำลองหรือ X ไม่มีอิทธิพลต่อ Y

ปฏิเสธสมมุติฐานหลัก ณ ระดับนัยสำคัญ α ถ้า $\left| \dfrac{\hat{\beta}}{\hat{\sigma}\sqrt{1/\Sigma x_i^2}} \right| > t_{n-2,\,1-\frac{\alpha}{2}}$

และไม่อาจปฏิเสธสมมุติฐานหลัก ณ ระดับนัยสำคัญ α ถ้า $\left| \dfrac{\hat{\beta}}{\hat{\sigma}\sqrt{1/\Sigma x_i^2}} \right| < t_{n-2,\,1-\frac{\alpha}{2}}$

ข. Sampling Distribution ของ $\hat{\alpha}$

จาก $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$

เราทราบว่า $Y_i \sim N(\alpha+\beta X_i, \sigma^2)$ ดังนั้น $\bar{Y} \sim N\left(\alpha+\beta\bar{X}, \frac{\sigma^2}{n}\right)$ [1] และเราทราบจากข้อ ก. ว่า $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\Sigma x_i^2}\right)$ ดังนั้น

$$M_{\hat{\alpha}}(t) = M_{\bar{Y}}(t) \cdot M_{\hat{\beta}}(-\bar{X}t)$$

$$= e^{(\alpha+\beta\bar{X})t+(\sigma^2/n)t^2/2} \cdot e^{\beta(-\bar{X}t)+(\sigma^2/\Sigma x_i^2)(-\bar{X}t/2)^2}$$

$$= e^{(\alpha+\beta(\bar{X}-\bar{X}))t+(\sigma^2/n+(\sigma^2\bar{X}^2/\Sigma x_i^2))t^2/2}$$

$$= e^{\alpha t+\sigma^2(1/n+\bar{X}^2/\Sigma x^2)t^2/2}$$

$$= e^{\alpha t+\sigma^2(\Sigma X_i^2/n\Sigma x_i^2)t^2/2}$$ [2]

นั่นคือ $\hat{\alpha} \sim N\left[\alpha, \sigma^2\left(\frac{\Sigma X_i^2}{n\Sigma x_i^2}\right)\right]$

(1) การสร้างช่วงเชื่อมั่นสำหรับ α

จาก $\Pr\left\{ -t_{n-2,\,1-\frac{\alpha}{2}} \leqslant \dfrac{\hat{\alpha}-\alpha}{\hat{\sigma}\sqrt{\dfrac{\Sigma X_i^2}{n\Sigma x_i^2}}} \leqslant t_{n-2,\,1-\frac{\alpha}{2}} \right\} = 1-\alpha$

$$\Rightarrow \Pr\left\{ \hat{\alpha} - t_{n-2,\,1-\frac{\alpha}{2}}\hat{\sigma}\sqrt{\frac{\Sigma X_i^2}{n\Sigma x_i^2}} \leqslant \alpha \leqslant \hat{\alpha} + t_{n-2,\,1-\frac{\alpha}{2}}\hat{\sigma}\sqrt{\frac{\Sigma X_i^2}{n\Sigma x_i^2}} \right\} = 1-\alpha$$

---

[1] พิสูจน์ได้โดยอาศัย mgf • technique

[2] $$\frac{1}{n}+\frac{\bar{X}^2}{\Sigma x_i^2} = \frac{\Sigma x_i^2+n\bar{X}^2}{n\,\Sigma x_i^2} = \frac{\Sigma(X_i-\bar{X})^2+n\bar{X}^2}{n\Sigma x_i^2} = \frac{(\Sigma X_i^2 - n\bar{X}^2)+n\bar{X}^2}{n\Sigma x_i^2} = \frac{\Sigma X_i^2}{n\Sigma x_i^2}$$

แสด งว่าช่ว งเชื่อมั่น (Random Interval)[1] ต่อไปนี้คือ

$\left(\hat{\alpha} - t_{n-2\,,\,1-\frac{\alpha}{2}}\, \hat{\sigma} \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}\,,\ \hat{\alpha} + t_{n-2\,,\,1-\frac{\alpha}{2}}\, \hat{\sigma} \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}} \right)$ จะเป็นช่วงที่ครอบคลุมเอาค่าจริงของ α ไว้

(2) การทดสอบสมมุติฐานเกี่ยวกับ α

แม้ว่า α จะเป็นสัมประสิทธิความถดถอยซึ่งเราไม่ค่อยให้ความสนใจมากก็ตาม แต่ในสถานการณ์ทางปฏิบัติบางสถานการณ์เราอาจจำเป็นต้องทดสอบเพื่อย้ำว่า $\alpha = 0$ อย่างแน่นอน หรือ $\alpha \neq 0$ อย่างแน่ชัด เช่นใน Consumption Function (C = f(Y)) หรือเหตุการณ์ทางวิทยาศาสตร์บางเรื่องเช่น เรื่องการดูดซึมตัวยาตามผนังลำไส้ ซึ่งจำเป็นต้องทดสอบเพื่อย้ำว่า $\alpha = 0$ เสมอ ด้วยเหตุนี้เราจึงมีความจำเป็นต้องทดสอบสมมุติฐานเกี่ยวกับ α อยู่บ้างแม้จะไม่ทุกครั้งทุกสถานการณ์ก็ตาม

สำหรับการทดสอบ $H_0 : \alpha = 0$ VS $H_1 : \alpha \neq 0$ นั้น โดยอาศัย MLRT จะพบว่า เราจะปฏิเสธสมมุติฐานหลักถ้า $|t_c| > t_{n-2\,,\,1-\frac{\alpha}{2}}$ และไม่อาจปฏิเสธสมมุติฐานหลัก ถ้า $|t_c| \leq t_{n-2\,,\,1-\frac{\alpha}{2}}$

โดยที่ $t_c = \dfrac{\hat{\alpha}}{\hat{\sigma}\sqrt{\dfrac{\sum X_i^2}{n\sum x_i^2}}}$

ตัวอย่าง 3.1 ในการทดลองครั้งหนึ่ง นักวิจัยสงสัยว่าผลผลิตข้าวโพด (ถัง/ไร่) น่าจะผันผวนไปเพราะสาเหตุจากการกำหนดปริมาณปุ๋ยและความชื้นหรือไม่ (ควบคุมสภาพของดินและอุณหภูมิในแปลงเพาะไว้แล้ว) ด้วยเหตุนี้ผู้วิจัยจึงกำหนดแบบจำลองไว้ดังนี้

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \ ; \ i = 1,2,\ldots,n$$

โดยที่ Y_i = ปริมาณผลผลิตข้าวโพด (ถัง/ไร่) จากแปลงทดลองที่ i X_{2i} = ปริมาณปุ๋ย (กิโลกรัม/ไร่) ที่ใส่ลงในแปลงทดลองที่ i และ X_{3i} = ปริมาณน้ำฝน (นิ้ว) ที่ตกในบริเวณแปลงทดลองที่ i

จากการบันทึกข้อมูลในศูนย์ทดลองต่าง ๆ 7 แห่ง ปรากฏข้อมูลดังนี้

---

[1] โดยทางทฤษฎี ช่วงเชื่อมั่นจะเป็นช่วงเชื่อมั่นสุ่มเสมอทั้งนี้เพราะช่วงเชื่อมั่นจะเปลี่ยนช่วงไปได้ตามตัวอย่างสุ่ม (Random Sample) ซึ่งหมายถึงการเปลี่ยนช่วงไปได้ตามตัวสถิติซึ่งเป็นตัวแปรสุ่มที่ผันแปรค่าตามตัวอย่างสุ่ม

Y ปริมาณผลผลิตข้าวโพด (ถัง/ไร่)	X_2 ปริมาณปุ๋ยที่ใช้ (กิโลกรัม/ไร่)	X_3 ปริมาณน้ำฝน (นิ้ว)
40	100	36
45	200	33
50	300	37
65	400	37
70	500	34
70	600	32
80	700	36

จงวิเคราะห์สมการถดถอยดังกล่าว พร้อมทั้งอภิปรายผล

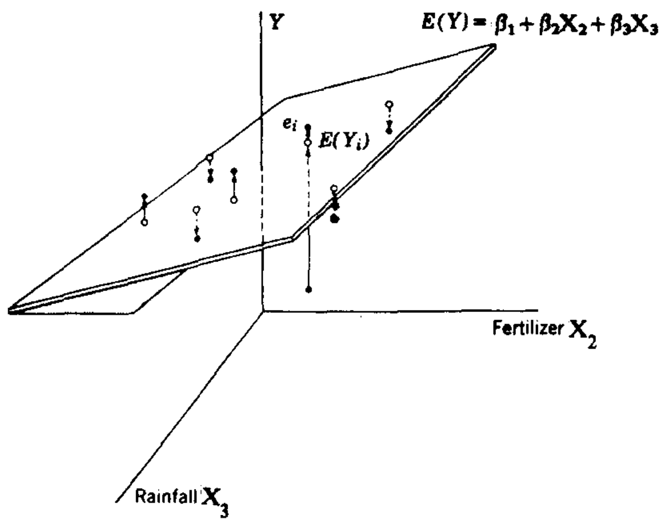
วิธีทำ จากสมการ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \; ; \; i = 1,2,\ldots,n$

แสดงว่า $E(Y_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \quad ; \; i = 1,2,\ldots,n$ คือสมการระนาบ (plane) ที่ค่าเฉลี่ยของ Y_i ปรากฏอยู่ หรือนัยหนึ่ง $E(Y_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \; ; \; i = 1,2,\ldots,n$ คือสมการของระนาบที่พุ่งผ่านไปในกลุ่มค่าเฉลี่ยของ Y_i ขอให้สังเกตว่า β_1 คือระยะจุดตัดบนแกน Y, β_2 คือความชันของระนาบเทียบต่อระนาบ YX_2[1] และ β_3 คือความชันของระนาบเทียบต่อระนาบ YX_3[2]

สมมุติว่าระนาบ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$ คือระนาบจริง (True Plane) มีลักษณะปรากฏดังภาพ และค่าสังเกต $Z_i = (Y_i, X_{2i}, X_{3i})$ คือจุดโคออร์ดิเนตที่เห็นเป็นจุดดำซึ่งอาจอยู่บนผิวระนาบ อยู่เหนือระนาบ หรืออยู่ใต้ระนาบ ดังภาพ[3]

[1,2] จากสมการ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$ ถ้าเรากำหนดให้ $X_3 = 0$ จะพบว่า $E(Y) = \beta_1 + \beta_2 X_2$ และถ้าเรากำหนดให้ $X_2 = 0$ จะพบว่า $E(Y) = \beta_1 + \beta_3 X_3$ แสดงว่า β_1 คือจุดตัดบนแกน Y, β_2 คือความชันของเส้นตรงที่เกิดจากรอยตัดกันระหว่างระนาบ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$ กับระนาบ YX_2 และ β_3 คือความชันของเส้นตรงที่เกิดจากรอยตัดกันระหว่างระนาบ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$ กับระนาบ YX_3 ขอให้ระลึกไว้ด้วยว่าในทางคณิตศาสตร์นั้นสมการเส้น (Curve) เกิดจากการตัดกันของระนาบ

[3] จุดดำ คือ โคออร์ดิเนตของค่าสังเกต $Z_i = (Y_i, X_{2i}, X_{3i})$ จุด o คือ จุดที่แนบอยู่บนระนาบลูกศร ↑ แสดงว่าค่าสังเกตอยู่ใต้ระนาบ เครื่องหมาย ↓ แสดงว่าค่าสังเกตอยู่เหนือระนาบ



ภารกิจของเราก็คือการอาศัยข้อสนเทศ (Information) จากค่าสังเกต $Z_i = (Y_i, X_{2i}, X_{3i})$; $i = 1,2,\ldots,n$ มาเป็นเครื่องมือในการประมาณสมการ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$ ซึ่งเป็น true plane การประมาณดังกล่าวก็คือ การประมาณค่าจุดตัดบนแกน Y (ประมาณ β_1) และความชันของระนาบ (ประมาณ β_2 และ β_3) ถ้าค่าประมาณของจุดตัดบนแกน Y[1] และความชันมีค่าใกล้เคียงกับค่าจริงระนาบที่ประมาณขึ้นมาก็จะใกล้เคียงกับระนาบจริง เครื่องมือที่ใช้วัดว่าค่าประมาณดังกล่าวใกล้เคียงกับความจริงเพียงใดก็คือ $V(\hat{\beta})$ ถ้า $V(\hat{\beta})$ มีค่าใกล้ 0 หรือเท่ากับ 0 (Zero Matrix) ระนาบที่ประมาณขึ้นก็จะทับไปบนระนาบจริง ถ้า $V(\hat{\beta})$ มีค่ามาก ระนาบที่ประมาณขึ้นก็จะเบี่ยงไปจากระนาบจริง จะเบี่ยงไปในทิศใดมากน้อยเพียงใดขึ้นอยู่กับว่า $V(\hat{\beta}_j)$; $j = 1,2,3$ มีค่าสูงต่ำแตกต่างกันเพียงใด ถ้า $V(\hat{\beta}_1)$ มีค่าสูงแสดงว่าระนาบที่ประมาณขึ้นมาอาจลอยอยู่สูงหรือต่ำกว่าระนาบจริงมาก ถ้า $V(\hat{\beta}_2)$ และ $V(\hat{\beta}_3)$ มีค่าสูงระนาบที่ประมาณขึ้นมานั้นอาจเอียงซ้ายขวาหรือก้มผิดความจริง ดังนี้เป็นต้น

---

[1] ในสมการ $E(Y) = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k$ นั้น ในความหมายเชิงคณิตศาสตร์เราถือว่า β_1 คือค่าเริ่มต้นของ E(Y) หรือจุดตัดบนแกน Y

อนึ่ง จากสมการ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$ แสดงให้เห็นว่าในการพยากรณ์นั้นเรามุ่งพยากรณ์ค่าของ $E(Y)$ มิใช่พยากรณ์ค่า Y ที่เป็นเช่นนี้เพราะ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ เป็นสมการที่เป็น Linear Function ของตัวแปรสุ่ม u แต่เนื่องจาก u เป็น Unobservable Random Variable การพยากรณ์ค่าของ Y จึงเป็นสิ่งที่กระทำมิได้ อย่างไรก็ตามในทางปฏิบัติเราไม่นิยมเขียนสมการพยากรณ์ว่า $E(Y) = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$ แต่นิยมเขียนเป็น $Y = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$ ดังนั้นจึงขอย้ำให้นักศึกษาเข้าใจไว้ในที่นี้ว่าในสมการพยากรณ์ $Y = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$ นั้นค่า Y ในที่นี้ก็คือ $E(Y)$ ซึ่งก็คือรูปย่อ ของ $E(Y|X_2, X_3)$ นั่นเอง[1]

จากข้อมูลเราสามารถประมาณสมการ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$[2] ได้ดังนี้

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = (X'X)^{-1}X'Y \text{ และ } \hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} \text{ เมื่อ } \hat{\sigma}^2 = \frac{e'e}{n-k}$$

โดยที่ $e = Y - X\hat{\beta}$ หรือ $\hat{\sigma}^2 = \frac{1}{n-k}(Y'Y - \hat{\beta}'X'Y)$ ซึ่งในที่นี้ $n = 7,\ k = 3$

จากข้อมูลพบว่า

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 100 & 200 & 300 & 400 & 500 & 600 & 700 \\ 36 & 33 & 37 & 37 & 34 & 32 & 36 \end{bmatrix} \begin{bmatrix} 1 & 100 & 36 \\ 1 & 200 & 33 \\ 1 & 300 & 37 \\ 1 & 400 & 37 \\ 1 & 500 & 34 \\ 1 & 600 & 32 \\ 1 & 700 & 36 \end{bmatrix} = \begin{bmatrix} 7 & 2{,}800 & 245 \\ 2{,}800 & 1{,}400{,}000 & 97{,}500 \\ 245 & 97{,}500 & 8{,}599 \end{bmatrix}$$

---

[1] $E(Y | X_2, X_3)$ สำหรับตัวอย่างนี้อ่านว่า "ค่าคาดหมายหรือค่าเฉลี่ยของตัวแปรสุ่ม Y ณ สถานีทดลอง (หรือ ณ วาระสังเกต) ที่กำหนดให้ $X_2 = X_{20}$ และ $X_3 = X_{30}$" ตามตัวอย่าง 3.1 สมมุติเราต้องการพยากรณ์ประมาณผลผลิตถั่วเฉลี่ยเมื่อใช้ปุ๋ย 750 ถัง/ไร่ และมีปริมาณน้ำฝน 29 นิ้ว ค่าพยากรณ์ก็คือ $E(Y)$ ซึ่งก็คือ $E(Y | X_2 = 750, X_3 = 29)$

[2] ต่อไปจะเขียนว่า $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3$ หรือในกรณีทั่วไปก็คือ $Y = X\beta$ และเขียนสมการคาดหมายว่า $Y = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$ หรือในกรณีทั่วไปคือ $Y = X\hat{\beta}$ โดยขอให้เข้าใจร่วมกันไว้ว่า Y ในสมการเหล่านี้ก็คือ $E(Y)$ หรือ $E(Y|X)$ นั่นเอง และบ่อยครั้งที่เราเสนอสมการคาดหมายว่า $\hat{Y} = X\hat{\beta}$ ซึ่งก็ขอให้เข้าใจว่า $\hat{Y}$ ก็คือ $E(Y)$ ในสมการ $E(Y) = X\hat{\beta}$ นั่นเอง

$$(X'X)^{-1} = \begin{bmatrix} 55.91410907607 & -4.18856259(10^{-3}) & -1.54559505413 \\ -4.18856259(10^{-3}) & 3.70942812(10^{-6}) & 7.72797827(10^{-5}) \\ -1\ 54559505413 & 7.72797827(10^{-5}) & 4.32766615(10^{-2}) \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 100 & 200 & 300 & 400 & 500 & 600 & 700 \\ 36 & 33 & 37 & 37 & 34 & 32 & 36 \end{bmatrix} \begin{bmatrix} 40 \\ 45 \\ 50 \\ 65 \\ 70 \\ 70 \\ 80 \end{bmatrix} = \begin{bmatrix} 420 \\ 187{,}000 \\ 14{,}680 \end{bmatrix}$$

ดังนั้น $\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 11.3292 \\ 00\ .068934 \\ 00\ .602782 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$

ดังนั้นสมการถดถอยที่ใช้ประมาณระนาบ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$ ก็คือ

$$E(Y) = 11.3292 + .068934X_2 + .602782X_3$$[1]

ปัญหาก็คือสมการถดถอยนี้สามารถกะประมาณที่ตั้งและลักษณะของระนาบจริงได้ดีเพียงใด เครื่องมือที่ใช้ก็คือ $\hat{V}(\hat{\beta})$

จาก $\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$ โดยที่ $\hat{\sigma}^2 = \frac{1}{n-k}(Y'Y - \hat{\beta}'X'Y)$

จากข้อมูลพบว่า

---

[1] ในทางปฏิบัติเราจะเขียนดังนี้คือ $Y = 11.3292 + 0.68934\,X_2 + .602782\,X_3$ โดยที่ Y ในที่นี้ก็คือ E(Y) หรือ $E(Y \mid X_2, X_3)$

$$Y'Y = (40, 45, 50, 65, 70, 70, 80)\begin{bmatrix}40\\45\\50\\65\\70\\70\\80\end{bmatrix} = 26,550$$

$$\hat{\beta}'X'Y = (11.3292, \quad .068934, \quad .602782)\begin{bmatrix}420\\187,000\\14,680\end{bmatrix} = 26,497.68160732$$

ดังนั้น $\hat{\sigma}^2 = \frac{1}{7-3}(26,550 - 26,497.68160732) = 13.07959817$

นั่นคือ $\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} = \begin{bmatrix}731.3340787485 & -5.47847156(10^{-2}) & -20.21576224156\\ -5.47847156(10^{-2}) & 4.85178293(10^{-5}) & 1.01078811(10^{-3})\\ -20.21576224156 & 1.01078811(10^{-3}) & .56604143276\end{bmatrix}$

$\Rightarrow \hat{V}(\hat{\beta}_1) = 731.3340787485,\ \hat{V}(\hat{\beta}_2) = .000048518,\ \hat{V}(\hat{\beta}_3) = .56604143276$

เราจึงสามารถเสนอสมการถดถอยได้ดังนี้

$$Y = \underset{[27.043]}{11.3292} + \underset{[.00697]}{.068934X_2} + \underset{[.7524]}{.602782X_3}$$

หมายเหตุ ค่าในวงเล็บคือ Standard Error

เมื่อพิจารณาสมการถดถอยเราจะพบว่า $\hat{V}(\hat{\beta}_2)$ และ $\hat{V}(\hat{\beta}_3)$ มีค่าค่อนข้างต่ำ แสดงว่า Estimated Plane มีลักษณะใกล้เคียงกับ True Plane เพราะค่าประมาณความแปรปรวนของความชันมีค่าต่ำซึ่งแสดงว่า Estimated Plane มีลักษณะการเอียงซ้ายขวา $(\hat{\beta}_2)$ ใกล้เคียงกับ True Plane และมีลักษณะการก้มเงย $(\hat{\beta}_3)$ ไม่ต่างจาก True Plane แต่เมื่อพิจารณา $\hat{V}(\hat{\beta}_1)$ พบว่า $\hat{V}(\hat{\beta}_1)$ มีค่าสูง แสดงว่า Estimated Plane อาจลอยอยู่เหนือหรือใต้ True Plane มากสักหน่อย ซึ่งแสดงให้

เห็นว่าถ้าหากเรานำ Estimated Plane ไปคาดหมายค่า E(Y) ค่าประมาณของ E(Y) อาจคลาดเคลื่อนไปจากค่าที่ควรจะเป็นในอัตราค่อนข้างสูง ซึ่งสาเหตุของความคลาดเคลื่อนนี้จะมาจากความบกพร่องในค่าประมาณของ β_1 เป็นสำคัญ

อนึ่งผู้เขียนขออธิบายความหมายของ $\hat{\beta}$ เพิ่มเติมเพื่อประโยชน์ในการอภิปรายผลดังนี้

โดยปกติ $\hat{\beta}$ มีความหมายเชิงคณิตศาสตร์ในแง่ของการบอกรูปร่างลักษณะของ Estimated Plane หรือ Estimated Surface แล้วแต่กรณีโดยมี $\hat{V}(\hat{\beta})$ เป็นดัชนีที่ใช้บ่งบอกว่า Estimated Plane หรือ Estimated Surface นั้นมีลักษณะ (Regression Coefficient β_j) และที่ตั้ง (β_1) ใกล้เคียงกับ True Surface เพียงใด แต่ในแง่ของการประยุกต์เราพึงมองความหมายของ $\hat{\beta}$ ในอีกมุมหนึ่งเพิ่มขึ้นจากการมองในเชิงคณิตศาสตร์ดังนี้

เนื่องจาก $\hat{\beta}_j = \frac{\Delta Y}{\Delta X_j}$; j = 2,3,...,k คือค่าความชันของ Surface (เมื่อพิจารณาจากสมการเส้นตรงในระนาบ X_jY) ดังนั้น $\hat{\beta}_j = \frac{\Delta Y}{\Delta X_j}$ จึงแสดงถึงอัตราการเปลี่ยนค่าของ Y อันเนื่องมาจากการเปลี่ยนค่าของตัวแปร X_j เมื่อ X_s คงที่ ; $s \neq j$[1]

ตัวอย่างเช่น $\hat{\beta}_2 = \frac{\Delta Y}{\Delta X_2} = .5 = \frac{1}{2}$ แสดงว่า ถ้า X_2 เปลี่ยนค่าไป 2 **หน่วยเมื่อใดจะมีผล**ให้ Y เปลี่ยนค่าไป 1 หน่วยเสมอ ขณะเดียวกัน ถ้า $\hat{\beta}_2 = 5$ แสดงว่า $\hat{\beta}_2 = \frac{5}{1} = \frac{\Delta Y}{\Delta X_2}$ แสดงว่าถ้า X_2 เปลี่ยนค่าไป 1 หน่วยจะมีผลผลักดันให้ Y เปลี่ยนแปลงค่าไปถึง 5 หน่วย หรือ $\hat{\beta}_2 = 20$ แสดงว่า $\hat{\beta}_2 = \frac{20}{1} = \frac{\Delta Y}{\Delta X_2}$ แสดงว่าถ้า X_2 เปลี่ยนค่าไป 1 หน่วยจะมีผลให้ Y เปลี่ยนแปลงค่าไปถึง 20 หน่วย ตามตัวอย่างที่ผ่านมา $\hat{\beta}_2 = .0689 = \frac{.0689}{1}$ หรือ $\frac{689}{10,000}$ แสดงว่า X_2 (ปริมาณปุ๋ย) เปลี่ยนแปลงไป 10,000 หน่วย จะมีผลให้ปริมาณผลผลิตข้าวโพด (Y) เปลี่ยนแปลงไป (เพิ่มขึ้น) เพียง 689 หน่วย ขณะเดียวกัน $\hat{\beta}_3 = .6028 = \frac{6,028}{10,000}$ แสดงว่าปริมาณน้ำฝน (X_3) เปลี่ยนแปลงไป 10,000 หน่วยจะมีผลให้ปริมาณผลผลิตข้าวโพดเปลี่ยนแปลง (เพิ่มขึ้น) ถึง 6,028 หน่วย ดังนี้ย่อมสรุปได้ว่าค่า $\hat{\beta}_j$ ก็คืออิทธิพลของตัวแปรอิสระ X_j ที่มีต่อตัวแปร Y ถ้า $\hat{\beta}_j$ มีค่าสูงแสดงว่า X_j มีอิทธิพลต่อ Y มากขณะเดียวกันถ้า $\hat{\beta}_j$ มีค่าต่ำแสดงว่า X_j มีอิทธิพลต่อตัวแปร Y น้อย นอกจากนี้เครื่องหมายของ $\hat{\beta}_j$ มีความหมายถึงการเปลี่ยนแปลงในทางเพิ่มขึ้น

[1] ด้วยเหตุนี้เราจึงเรียก β_j ว่า Partial Regression Coefficient (อ่านตอน 3.2.6.1)

หรือลดลงของตัวแปร Y อันเนื่องมาจากตัวแปร X_j กล่าวคือ ถ้า $\hat{\beta}_j$ มีค่าติดลบแสดงว่า ถ้า X_j เปลี่ยนแปลงค่าเพิ่มขึ้น (ลดลง) จะมีผลให้ตัวแปร Y มีค่าลดลง (เพิ่มขึ้น) ขณะเดียวกัน ถ้า $\hat{\beta}_j$ มีค่าบวก แสดงว่าถ้า X_j เปลี่ยนแปลงค่าเพิ่มขึ้น (ลดลง) จะมีผลให้ตัวแปร Y มีค่าเพิ่มขึ้น (ลดลง) ดัชนีที่ใช้วัดดูว่าตัวแปร $X_2, X_3, \ldots, X_k$ มีอิทธิพลร่วมกันในการผลักดันให้ตัวแปร Y เปลี่ยนแปลงค่าไปมากน้อยเพียงใด คือ Coefficient of Determination, R^2 (หรือเขียนให้ถูกต้องตามหลักทางทฤษฎีก็คือ $R^2_{Y \cdot x_2 x_3 \cdots x_k}$) ความรู้และข้อจำกัดรวมทั้งข้อขัดแย้งต่าง ๆ เกี่ยวกับการนำ Multiple Correlation และ Simple Correlation มาใช้ในงานวิเคราะห์ความถดถอยนี้จะกล่าวถึงในบทที่ 4

3.2.5 การสร้างเขตความเชื่อมั่น (Confidence Region)

การศึกษาเกี่ยวกับการสร้างเขตความเชื่อมั่นนี้เป็นเรื่องที่เกี่ยวข้องอยู่กับ Multivariate Normal Distribution โดยเฉพาะอย่างยิ่งก็คือการแจกแจงของ Quadratic Form ดังนั้นก่อนที่จะศึกษาเรื่องนี้ผู้เขียนขอแนะนำเกี่ยวกับเรื่องราวโดยสรุปของการแจกแจง Quadratic Form ตลอดจนการนำการแจกแจงดังกล่าวไปใช้ประโยชน์เสียก่อน

3.2.5.1 การแจกแจงของ Quadratic Form

ให้ $X_j \sim N(\mu_j, \sigma_j^2)$; $j = 1,2,\ldots,n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน โดยอาศัยความรู้เกี่ยวกับ mgf เราสามารถพิสูจน์ได้ว่า[1] ตัวแปรสุ่ม Q คือ

$$Q = \sum_i^n (X_j - \mu_j)^2/\sigma_j^2 \sim \chi^2_{(n)}$$

ขอให้สังเกตว่า $Q = \sum_i^n \left(\frac{X_j - \mu_j}{\sigma_j}\right)^2 = \frac{1}{\sigma_1^2}(X_1 - \mu_1)^2 + \frac{1}{\sigma_2^2}(X_2 - \mu_2)^2 + \ldots + \frac{1}{\sigma_n^2}(X_n - \mu_n)^2$

$$= (X_1 - \mu_1, X_2 - \mu_2, \ldots, X_n - \mu_n) \begin{bmatrix} 1/\sigma_1^2 & & 0 \\ & 1/\sigma_2^2 & \\ & & \ddots \\ 0 & & & 1/\sigma_n^2 \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_n - \mu_n \end{bmatrix}$$

$$= (X - \mu)' V^{-1} (X - \mu)$$

---

[1] อ่านมนตรี พิริยะกุล **ทฤษฎีสถิติ 2** (โรงพิมพ์วิคตอรี กรุงเทพมหานคร, 2525) หน้า 69

เมื่อ $X' = (X_1, X_2, \ldots, X_n)$ และ $\mu' = (\mu_1, \mu_2, \ldots, \mu_n)$ หรือ X และ μ เป็นเวคเตอร์ขนาด $n \times 1$ และ V คือ Variance - Covariance Matrix ของตัวแปรสุ่ม $X_1, X_2, \ldots, X_n$

ขอให้สังเกตว่าในที่นี้เราถือว่า $X_1, X_2, \ldots, X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน V จึงเป็น diagonal matrix กล่าวคือ $V = \text{diag}.(\sigma_1^2, \sigma_1^2, \ldots, \sigma_n^2)$

ต่อไปนี้จะแสดงให้เห็นว่า ถ้า $X_1, X_2, \ldots, X_n$ มิได้เป็นอิสระต่อกัน[1] ซึ่งในกรณีเช่นนี้เรา ย่อมเลียนแบบตัวแปรสุ่ม Q ได้ดังนี้คือ

$$Q = (X - \mu)'V^{-1}(X - \mu)$$

เพียงแต่ในกรณีนี้ V มิใช่ $\text{diag}.(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$ แต่ V เป็น Positive Definite Matrix ใด ๆ ปัญหาก็คือ Q มีการแจกแจงแบบใด .

การหาฟังก์ชันการแจกแจงของตัวแปรสุ่ม Q เราจะเริ่มต้นด้วยการหา mgf ของ Q คือ $M_Q(t)$ ดังนี้

$$M_Q(t) = E(e^{Qt}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{Qt} f(x_1, x_2, \ldots, x_n)dx_1dx_2\ldots dx_n$$

แต่ $f(x_1, x_2, \ldots, x_n)$ คือ Joint pdf ของตัวแปรสุ่มปกติ $X_1, X_2, \ldots, X_n$ โดยที่ $X_j \sim N(\mu_j, \sigma_j^2)$ ดังนั้น $f(x_1, x_2, \ldots, x_n)$ คือ Multivariate Normal Distribution

$$\Rightarrow M_Q(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left\{ t(X - \mu)'V^{-1}(X - \mu) \right\} \times \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left\{ \frac{-(X - \mu)'V^{-1}(X - \mu)}{2} \right\} dx_1dx_2\ldots dx_n$$

$$= \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left\{ \frac{-(X - \mu)'V^{-1}(X - \mu)(1 - 2t)}{2} \right\} dx_1dx_2\ldots dx_n$$

จาก Integrand คือ $\frac{1}{(2\pi)^{n/2}|V|^{1/2}} e^{-\frac{(X-\mu)'V^{-1}(X-\mu)(1-2t)}{2}}$ เราจะพบว่า Integrand นี้มีรูปใกล้เคียงกับ Multivariate Normal ส่วนที่เกินเข้ามาคือ $(1 - 2t)$ ใน Exponent Term เราจึงจัดรูป Integrand

---

[1] ต่อไปเราจะได้นำความรู้ส่วนนี้ไปใช้ในการหาฟังก์ชั่นการแจกแจงของ $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ ซึ่งโดยปกติแล้วจะไม่เป็นอิสระต่อกัน

เสียใหม่โดยให้ $(1-2t)V^{-1} = S^{-1}$ ซึ่งจะพบว่า $|S^{-1}| = |(1-2t)V^{-1}| = (1-2t)^n|V^{-1}|$ และเนื่องจากในเมตริกซ์ D ใด ๆ นั้น $|D^{-1}| = 1/|D|$ ถ้า $|D| \neq 0$ ดังนั้น

$|S^{-1}| = (1-2t)^n|V^{-1}|$ จึงกลายเป็น $\frac{1}{|S|} = \frac{(1-2t)^n}{|V|}$ หรือ $|V| = (1-2t)^n|S|$ ดังนั้น integrand จึงเปลี่ยนรูปจากเดิมเป็น

$$\frac{1}{(2\pi)^{n/2}\sqrt{(1-2t)^n|S|}} e^{-\frac{(X-\mu)'S^{-1}(X-\mu)}{2}}$$

ดังนั้น
$$M_Q(t) = \frac{1}{(1-2t)^{n/2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}|S|^{1/2}} \exp\left\{\frac{-(X-\mu)^{-1}S^{-1}(X-\mu)}{2}\right\} dx_1dx_2\ldots dx$$

$= \frac{1}{(1-2t)^{n/2}}$; $t < \frac{1}{2}$ ซึ่งก็คือ mgf ของตัวแปรสุ่ม $\chi^2_{(n)}$

=> $M_Q(t) = \frac{1}{(1-2t)^{n/2}}$ ซึ่งย่อมแสดงว่า ถ้า V เป็น Positive Definite ใด ๆ แล้ว Quadratic Form $(X-\mu)'V^{-1}(X-\mu)$ จะมีการแจกแจงแบบ $\chi^2_{(n)}$

3.2.5.2 เขตเชื่อมั่นของ β ที่ประมาณด้วย $\hat{\beta}$

ในที่นี้จะเสนอในรูปทั่วไปก่อน ส่วนกรณีเฉพาะต่าง ๆ จะได้แนะนำไว้เป็นบางจุดเท่าที่จำเป็นทั้งนี้เพื่อเปิดโอกาสให้นักศึกษาได้ศึกษาด้วยตนเองในเรื่องดังกล่าว

ให้ $C = (c_{ij})_{r\times k}$ เป็นเมตริกซ์ของค่าคงที่ใด ๆ โดยที่ $r(C) = r$ หรือ C มี **Full Rank** และให้ $\hat{\gamma} = C\hat{\beta}$ คือ ฟังก์ชันของ $\hat{\beta}$ กล่าวคือ

$$\begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_r \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \ldots & c_{1k} \\ c_{21} & c_{22} & \ldots & c_{2k} \\ \vdots & \vdots & & \vdots \\ c_{r1} & c_{r2} & \ldots & c_{rk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$ [1]

หรือนัยหนึ่ง

$$\hat{\gamma}_i = c_{i1}\hat{\beta}_1 + c_{i2}\hat{\beta}_2 + \ldots + c_{ik}\hat{\beta}_k = \sum_{j}^{k} c_{ij}\hat{\beta}_j \ ; \ i = 1,2,\ldots,r$$

---

[1] ศึกษาเพิ่มเติมได้จาก **Wonnacott, R.J. and Wonnacott. T.H. Econometrics (John Wiley and Sons, Inc., NY, 1970), p. 249**

เมื่อ $\hat{\beta} \sim N[\beta, \sigma^2(X'X)^{-1}]$ พิจารณา $C\hat{\beta}$ จะพบว่า $E(\hat{\gamma}) = E(C\hat{\beta}) = C\beta$ และ $V(\hat{\gamma}) = V(C\hat{\beta}) = C\sigma^2(X'X)^{-1}C'$ [1]

ดังนั้น $\hat{\gamma} = C\hat{\beta} \sim N[C\beta, \sigma^2 C(X'X)^{-1}C']$

โดยอาศัยความรู้ในตอน 3.2.5.2 เราทราบว่า ตัวแปรสุ่ม $Q \sim \chi^2_{(r)}$

ดังนั้น $Q = (C\hat{\beta} - C\beta)'[\sigma^2 C(X'X)^{-1}C']^{-1}(C\hat{\beta} - C\beta)$ [2] จึงมีการแจกแจงแบบ $\chi^2_{(r)}$ ทั้งนี้เพราะ rank ของ C มีค่าเท่ากับ r

หมายเหตุ การกำหนดให้ $\hat{\gamma} = C\hat{\beta}$ เป็นการศึกษาเกี่ยวกับการสร้างเขตเชื่อมั่น [3] ของ β ในลักษณะทั่วไป (Generalized Approach) เพื่อให้มองเห็นภาพที่ใกล้ตัวนักศึกษามากขึ้น ผู้เขียนขอเสนอกรณีเฉพาะไว้หลาย ๆ กรณีดังนี้

กรณีเฉพาะที่ 1 $C = I_k$ กล่าวคือ

$$C = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{k \times k}$$

ดังนั้น $\hat{\gamma} = C\hat{\beta} = I_k\hat{\beta} = \hat{\beta}$

$$\Rightarrow \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

นั่นคือ ถ้าเรากำหนดให้ $C = I_k$ แสดงว่าเรากำลังศึกษาเพื่อสร้างเขตเชื่อมั่นของ β ทั้งเวคเตอร์ หรือนัยหนึ่งก็คือ เรากำลังจะสร้างเขตเชื่อมั่นร่วม (Joint Confidence Region) ของพารามิเตอร์ β

---

[1] ดูตอน 3.2.2.4

[2] $X_j \sim N(\mu_j, \sigma^2_j)$ ดังนั้น $Z_j = \frac{X_j - \mu_j}{\sigma_j}$ และ $\sum_j^r Z_j^2 \sim \chi^2_{(r)}$

[3] หมายถึงทั้ง Point Estimation และ Interval Estimation ทั้งนี้เพราะ Interval Estimation ต้องอาศัยข้อสนเทศจาก Point Estimation

ทุกตัวพร้อม ๆ กัน

กรณีเฉพาะที่ 2 $C = (0,0,\ldots,0,1,0\ldots,0)$

$C\hat{\beta} = \hat{\beta}_j$ นั่นคือ $\hat{\gamma} = \hat{\beta}_j$ แสดงว่าเรากำลังศึกษาเพื่อสร้างเขตเชื่อมั่นของ β_j

กรณีเฉพาะที่ 3 $C = (1, X_{20}, X_{30},\ldots,X_{k0})$

เมื่อ $C = (1, X_{20}, X_{30},\ldots,X_{k0})$ แสดงว่า $\hat{\gamma} = C\hat{\beta} = \hat{\beta}_1+\hat{\beta}_2X_{20}+\ldots+\hat{\beta}_KX_{k0}$ ดังนั้น การกำหนดให้ $C = (1, X_{20}, X_{30},\ldots,X_{k0})$ ก็แสดงว่า เรากำลังศึกษาเกี่ยวกับการสร้างเขตเชื่อมั่นของค่าพยากรณ์ของ E(Y) หรือเขียนให้เต็มรูปก็คือ $E(Y|X_2 = X_{20}, X_3 = X_{30},\ldots,X_k = X_{k0})$ ทั้งนี้เพราะเมื่อเราต้องการพยากรณ์ค่า E(Y) ณ ค่า $X_1 = X_{10}, X_2 = X_{20},\ldots,X_k = X_{k0}$ เราจะพบว่า $E(Y) = \hat{\beta}_1+\hat{\beta}_2X_{20}+\ldots+\hat{\beta}_kX_{k0}$

$$\Rightarrow E(Y) = (1, X_{20}, X_{30},\ldots,X_{k0})\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_k\end{bmatrix} = r'\hat{\beta} = C\hat{\beta} \text{ เมื่อ } C = r'$$

กรณีเฉพาะที่ 4 เมื่อต้องการสร้างเขตเชื่อมั่นของ $\beta_2, \beta_3,\ldots\beta_k$ (ไม่รวม β_1)

ในกรณีนี้เราจะพบว่า

$$C = \left[\begin{array}{c|cccc}0 & 1 & 0 & \ldots & 0\\ 0 & 0 & 1 & \ldots & 0\\ 0 & 0 & 0 & \ldots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & \ldots & 1\end{array}\right]_{k\times k} \text{ และ } \hat{\beta} = \begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \vdots\\ \hat{\beta}_k\end{bmatrix}$$

$$\text{ดังนั้น } \hat{\gamma} = C\hat{\beta} = \left[\begin{array}{c|cccc}0 & 1 & 0 & \ldots & 0\\ 0 & 0 & 1 & \ldots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & \ldots & 1\end{array}\right]\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_k\end{bmatrix} = \begin{bmatrix}\hat{\beta}_2\\ \hat{\beta}_3\\ \vdots\\ \hat{\beta}_k\end{bmatrix}$$

และถ้าเราต้องการสร้างเขตเชื่อมั่นของเฉพาะ $\beta_r, \beta_{r+1},\ldots,\beta_k$ เราก็สามารถกำหนดเมตริกซ์ C ได้ดังนี้

$$C = \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

r - 1col k - r+1col

กรณีเฉพาะที่ 5 $C = (a_1, a_2, \ldots, a_k)$[1]

ในกรณีนี้เราจะพบว่า $\hat{\gamma} = C\hat{\beta} = a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + \ldots + a_k\hat{\beta}_k$ แสดงว่า เรากำลังศึกษาเพื่อสร้างเขตเชื่อมั่นของ Linear Combination ของ β_j ; j = 1,2,...,k เช่น C = (0, 2, -1, 0, 0,...,0) แสดงว่า $\hat{\gamma} = 2\hat{\beta}_2 - \hat{\beta}_3$

สำหรับการสร้างเขตเชื่อมั่นของ $C\beta$ นั้น เราพบว่า $\hat{\gamma} = C\hat{\beta}$ และทราบว่าตัวแปรสุ่ม Q มีการแจกแจงแบบ $\chi^2_{(r)}$ เมื่อ r(C) = r โดยที่

$$Q = (C\hat{\beta} - C\beta)'[C\sigma^2(X'X)^{-1}C']^{-1}(C\hat{\beta} - C\beta)$$

ขณะเดียวกัน เราทราบว่า $\frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-k)}$[2]

ดังนั้นโดยอาศัยนิยามของตัวแปรสุ่ม F คือ $F_{\nu_1, \nu_2} = \frac{\chi^2_1/\nu_1}{\chi^2_2/\nu_2}$ เมื่อกำหนดให้ $Q = \chi^2_1$ และ $\frac{(n-k)\hat{\sigma}^2}{\sigma^2} = \chi^2_2$

ดังนั้น
$$\frac{(C\hat{\beta} - C\beta)'[C\sigma^2(X'X)^{-1}C']^{-1}(C\hat{\beta} - C\beta)/r}{\frac{(n-k)\hat{\sigma}^2}{\sigma^2}/n-k} \sim F_{r,\,n-k}$$

$$\Rightarrow \quad \frac{(C\hat{\beta} - C\beta)'[C(X'X)^{-1}C']^{-1}(C\hat{\beta} - C\beta)}{r\hat{\sigma}^2} \sim F_{r,n-k}$$

และโดยอาศัยสมการความน่าจะเป็น

---

[1] กรณีเฉพาะ 2, 3, 5 คือกรณีที่แมตริกซ์ C ลดรูปลงเป็นแมตริกซ์ขนาด $1 \times k$ หรือ Row Vector ขนาด k แต่โดยปกติแล้วเรานิยมเสนอเวคเตอร์ในรูป Column Vector ดังนั้น ถ้าเราจะกำหนดให้ C เป็น Column Vector. รูปของสูตรจะเปลี่ยนไปเล็กน้อยแต่ผลลัพธ์ตรงกัน

[2] เราทราบจากทฤษฎีสถิติว่า $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ เมื่อนิยามว่า $S^2 = \frac{1}{n-1}\sum_i^n (X_i - \bar{X})^2$

$$\Pr \left\{ \frac{(C\hat{\beta} - C\beta)'[C(X'X)^{-1}C'](C\hat{\beta} - C\beta)}{r\hat{\sigma}^2} \leqslant \in \right\} = 1 - \alpha$$ [1]

เราสามารถพิสูจน์โดยอาศัย Quantile เป็นเครื่องมือได้โดยง่ายว่า $\in = F_{r,n-k,\,1-\alpha}$ ดังนั้นเขตเชื่อมั่น $(1 - \alpha)$ 100% ที่เชื่อว่า Random Region จะครอบคลุมค่าจริงของ $C\beta$ ไว้ก็คือ

$$(C\beta - C\hat{\beta})'[C(X'X)^{-1}C']^{-1}(C\beta - C\hat{\beta}) \leqslant r\hat{\sigma}^2F_{r,n-k,\,1-\alpha}$$

สำหรับกรณีเฉพาะที่น่าสนใจสำหรับการสร้างเขตเชื่อมั่นปรากฏดังนี้

กรณีเฉพาะที่ 1 เมื่อ $C = I_k$ ซึ่งเป็นกรณีของการสร้างเขตเชื่อมั่นของพารามิเตอร์ β_j ทุกตัวพร้อม ๆ กัน (Joint Confidence Region) ในกรณีนี้เราจะพบว่าเขตเชื่อมั่น $(1 - \alpha)$ 100% คือ

$$(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta}) \leqslant r\hat{\sigma}^2F_{r,n-k,\,1-\alpha}$$

กรณีนี้เป็นกรณีเฉพาะที่สำคัญ เพราะในงานวิเคราะห์ความถดถอยเราจะใช้กรณีนี้เป็นปกติ ผู้เขียนจึงขอถือโอกาสนี้กล่าวถึงเรื่องนี้ค่อยข้างละเอียดเพื่อชี้ให้เห็นถึงแนวทางของการนำไปใช้ประโยชน์โดยเฉพาะอย่างยิ่งก็คือการนำเสนอเขตเชื่อมั่นในรูปทรงทางเรขาคณิต สำหรับกรณีเขตเชื่อมั่น $(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta}) \leqslant r\hat{\sigma}^2F_{r,(n-k),\,1-\alpha}$ เราสามารถแสดงให้เห็นได้ว่าเขตเชื่อมั่นดังกล่าวคือ k - Dimensional Ellipsoid สำหรับการแสดงดังกล่าว ผู้เขียนขอเสนอจากกรณีเฉพาะคือกรณี Simple Linear Regression แล้วขยายความสู่กรณีทั่วไป

[1] เช่นจากสมการ $P_r\left\{ \left| \frac{\bar{x}-\mu}{s/\sqrt{n}} \right| < \in \right\} = 1 - \alpha$ เราสามารถพิสูจน์ได้โดยง่ายได้ว่า $\in = t_{n-1,1-\alpha/2}$

ดังนั้นช่วงเชื่อมั่น $(1-\alpha)$ 100% หรือ $(1-\alpha)$ 100% Random Interval ที่คาดว่าจะครอบคลุมค่า μ เอาไว้ก็คือ

$\left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| \leqslant t_{n-1,\,1-\alpha/2}$ หรือ

$$\bar{X} - t_{n-1,1-\alpha/2}\, s/\sqrt{n} \leqslant \mu \leqslant \bar{X} + t_{n-1,1-\alpha/2}\, s/\sqrt{n}$$

สำหรับกรณีของ F เราพบว่าค่า F ทางด้านหางซ้ายมีค่าใกล้ 0 ซึ่งเมื่อจัดออกเป็นช่วงแล้วค่าทางซ้ายก็จะไม่ต่างจาก 0 ซึ่งเป็น lower bound ของ F ดังนั้นเราจึงไม่นิยามเขตเชื่อมั่นด้วย absolute form แต่กำหนดไว้เฉพาะด้านขวาของเขตถือว่าด้านซ้ายมีค่าไม่ต่างจาก 0 ขอให้สังเกตว่าเราจะไม่แยก α ออกเป็น $\frac{\alpha}{2}$

จากสมการ Simple Linear Regression $Y_i = \beta_1 + \beta_2 X_i + u_i$; $i = 1,2,\ldots,n$ จะพบว่า[1]

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \text{และ} \quad (X'X) = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ X_1 & X_2 & X_3 & \ldots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}, \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

ดังนั้นเขตเชื่อมั่นคือ $(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta}) < 2\hat{\sigma}^2 F_{2,\, n-2,\, 1-\alpha} = K^*$ [2]

$$\Rightarrow (\beta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2) \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix} \begin{bmatrix} \beta_1 - \hat{\beta}_1 \\ \beta_2 - \hat{\beta}_2 \end{bmatrix} < K^*$$

หรือ $n(\beta_1 - \hat{\beta}_1)^2 + 2\Sigma X_i(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + \Sigma X_i^2(\beta_2 - \hat{\beta}_2)^2 - K^* < 0$

โดยอาศัยความรู้เกี่ยวกับสมการกำลังสองรูปทั่วไป จะพบว่า $A = n$, $B = 2\Sigma X_i$, $C = \Sigma X_i^2$, $D = E = 0$, $F = -K^*$ และพบว่า $B^2 - 4AC = 4(\Sigma X_i)^2 - 4n\,\Sigma X_i^2 = 4\{(\Sigma X_i)^2 - n\Sigma X_i^2\}$

---

[1] เพื่อความเข้าใจที่ถูกต้องผู้เขียนขอทบทวนความรู้เกี่ยวกับหลักทางเรขาคณิตวิเคราะห์ในส่วนที่เกี่ยวกับ Ellipse เสียก่อนดังนี้

จากสมการกำลังสองรูปทั่วไป $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$

ถ้า $B^2 - 4AC < 0$ สมการกำลังสองดังกล่าวนี้จะเป็นสมการ Ellipse และเราสามารถแปลงรูปสมการนี้สู่รูปมาตรฐานของ Ellipse คือ $x^2/a^2 + y^2/b^2 = 1$ ซึ่งเป็นสมการ Ellipse ที่มีศูนย์กลางที่จุดกำเนิดหรือ$(x-g)^2/a^2 + (y-h)^2/b^2$ ซึ่งเป็นสมการ Ellipse ที่มีศูนย์กลางที่จุด (g, h) ได้โดยอาศัยการหมุนแกนให้เปลี่ยนไปจากตำแหน่งเดิม ϕ องศา สู่แกนใหม่คือแกน x' และ y' โดยที่เราสามารถคำนวณหาค่า ϕ ได้จากสมการ $\cot 2\phi = \frac{A-C}{B}$ และแทนค่า x และ y ในสมการรูปทั่วไปด้วย

$x = x'\cos\phi - y'\sin\phi$

$y = x'\sin\phi + y'\cos\phi$

ตัวอย่างเช่น สมการ $x^2 + xy + y^2 - 6 = 0$ เราจะพบว่า $B^2 - 4AC = 1 - 4(1).(1) = -3 < 0$ แสดงว่าสมการกำลังสองนี้คือสมการ Ellipse

จาก $\cot 2\phi = \frac{A-C}{B} = \frac{1-1}{1} = 0$ แสดงว่า $2\phi = 90°$ หรือ $\phi = 45°$ แสดงว่าถ้าเราหมุนแกน xy ไปจากตำแหน่งเดิม $45°$ สมการกำลังสองจะลดรูปสู่รูปมาตรฐานของ Ellipse เมื่อเทียบต่อแกนใหม่ $x'y'$

ดังนั้นจากสมการ $x^2+xy+y^2 = 0$ ถ้าเราแทนที่ x ด้วย $x'\cos 45° - y'\sin 45° = 1/\sqrt{2}(x'-y')$และแทนที่ y ด้วย $x'\sin 45° + y'\cos 45° = 1/\sqrt{2}(x'+y')$ จะพบว่า $x^2+xy+y^2-6 = \frac{3}{2}x'^2 + \frac{1}{2}y'^2 - 6 = 0$ หรือ $\frac{x'^2}{2^2} + \frac{y'^2}{(2\sqrt{3})^2} = 1$ ซึ่งก็คือสมการ Ellipse ที่มีแกน y' เป็น major axes โดยมี semi - major axis $= 2\sqrt{3}$ และ semi - minor axis $= 2$

[2] ในที่นี้ $C = I_2$ ดังนั้น $r = 2$ และเนื่องจาก $k = 2$ ดังนั้น $F_{2,\, n-k,\, 1-\alpha} = K^*$

$$\Rightarrow \quad B^2 - 4AC \quad = -4n \left\{ \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n} \right\} = -4n\Sigma x_i^2 < 0$$

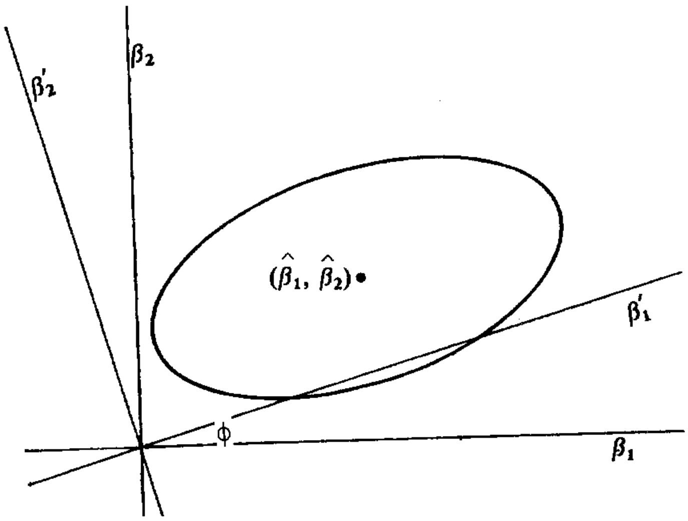
ซึ่งแสดงว่าสมการ

$$n(\beta_1 - \hat{\beta}_1)^2 + 2\Sigma X_i(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + \Sigma X_i^2(\beta_2 - \hat{\beta}_2)^2 - K^* \leqslant 0$$ คือสมการ Ellipse

ดังนั้นเมื่อทราบค่าของตัวแปร X และทราบ $\hat{\beta}_1$ และ $\hat{\beta}_2$ เราย่อมสามารถคำนวณหาค่า ϕ ได้จากสมการ $\cot 2\phi = \frac{A - C}{B}$ และโดยการเปลี่ยนรูปตัวแปรสู่แกน β_1' และ β_2' เราย่อมสามารถเสนอสมการเขตมั่นใจได้ดังนี้คือ

$$\frac{(\beta_1' - \hat{\beta}_1)^2}{K_1^2} + \frac{(\beta_2' - \hat{\beta}_2)^2}{K_2^2} \leqslant 1$$

ซึ่งก็คือสมการ Ellipse ที่มีจุดศูนย์กลางที่จุด $(\hat{\beta}_1, \hat{\beta}_2)$ เมื่อเทียบต่อแกนเดิมคือแกน β_1 และ β_2 โดยที่ K_1, K_2 คือค่าคงที่ที่เกิดขึ้นจากการแปลงรูปสมการ และเราสามารถเสนอเขตเชื่อมั่นได้ดังภาพ



จะเห็นได้ว่าเส้นรอบรูป (Contour) มีศูนย์กลางที่ $(\hat{\beta}_1, \hat{\beta}_2)$ มีลักษณะของ Ellipse ซึ่งแสดงถึง Joint Confidence Region ของ β_1 และ β_2 โดยเราเชื่อมั่นถึง $(1 - \alpha)$ 100% ว่า Contour นี้จะครอบคลุมค่าจริงคือ (β_1, β_2) เอาไว้ จากภาพนี้เรายังสามารถหา Confidence Limit ของเฉพาะ β_1 และ β_2 ได้โดยพิจารณาได้จากเฉพาะแกน β_1' และแกน β_2' ตามลำดับ

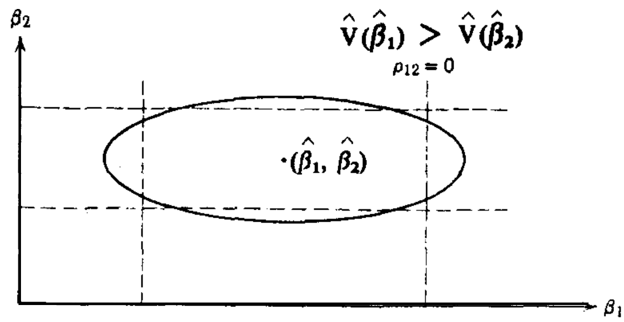
ถ้า $Cov(\hat{\beta}_1, \hat{\beta}_2) = 0$ ซึ่งแสดงว่าไม่มีเทอมไขว้ $(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)$ ในสมการกำลัง 2 กล่าวคือ ถ้า $Cov\ (\hat{\beta}_1, \hat{\beta}_2) = 0$ สมการจะลดรูปลงเหลือ

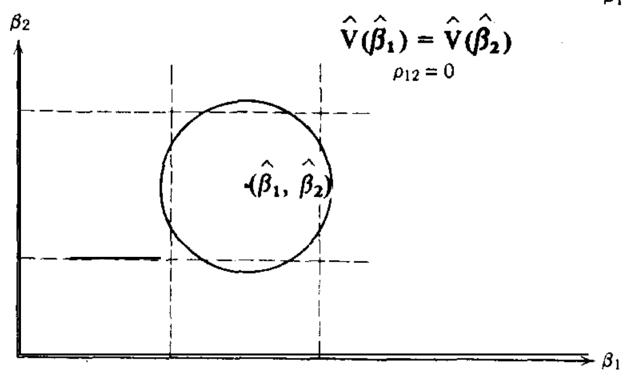
$$n(\beta_1 - \hat{\beta}_1)^2 + \Sigma X_i^2(\beta_2 - \hat{\beta}_2)^2 \leqslant K^*$$

$$\text{หรือ} \quad \frac{(\beta_1 - \hat{\beta}_1)^2}{K_1^2} + \frac{(\beta_2 - \hat{\beta}_2)^2}{K_2^2} \leqslant 1$$

เมื่อ $K_1 = \sqrt{\frac{K^*}{n}}$, $K_2 = \sqrt{\frac{K^*}{\Sigma X_i^2}}$

ในกรณีนี้เราจะไม่มีความจำเป็นต้องหมุนแกน contour จะปรากฏในแกนเดิมดังภาพ





ขอให้สังเกตว่า ปัจจัยที่มีผลต่อรูปทรงของ Ellipse ก็คือสัมประสิทธิของ $(\beta_1-\hat{\beta}_1)^2$ และ $(\beta_2-\hat{\beta}_2)^2$ ซึ่งก็คือสมาชิกในแนว Main Diagonal ของเมตริกซ์ $X'X$ นั่นเอง และถ้าเราพิจารณาให้ลึกลงไปโดยเปลี่ยนแปลงรูปของสมการเพียงเล็กน้อยก็จะพบว่าปัจจัยดังกล่าวก็คือ $\hat{V}(\hat{\beta}_1)$ และ $\hat{V}(\hat{\beta}_2)$ ขอให้สังเกตว่า ถ้า $\hat{V}(\hat{\beta}_1) = \hat{V}(\hat{\beta}_2)$ Contour นี้จะกลายเป็นวงกลมที่มีจุดศูนย์กลางที่ $(\hat{\beta}_1, \hat{\beta}_2)$ ถ้า $\hat{V}(\hat{\beta}_1) \neq \hat{V}(\hat{\beta}_2)$ Contour จะเป็นรูป Ellipse โดยที่ถ้า $\hat{V}(\hat{\beta}_1) > \hat{V}(\hat{\beta}_2)$ Contour จะมี Major Axes บนแกน β_1 ถ้า $\hat{V}(\hat{\beta}_1) < \hat{V}(\hat{\beta}_2)$ contour จะมี Major Axes บนแกน β_2

สำหรับกรณีของสมการ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i ; i = 1, 2, \ldots, n$ เขตเชื่อมั่นร่วมจะปรากฏเป็นรูป Ellipsoid[1] โดยมีจุดศูนย์กลางที่ $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ เมื่อเทียบต่อแกนเดิม

กรณีเฉพาะที่ 2

$$C = \underbrace{\begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ & & \\ & & \\ & & \\ 0 & \cdots & 0 \end{bmatrix}}_{k-r \text{ col}} \!\!\!\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \end{matrix} \left.\underbrace{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}}_{r \text{ col}}\right\} r \text{ row}$$

กรณีนี้คือกรณีที่เรามุ่งสนใจสร้างเขตเชื่อมั่นร่วมของพารามิเตอร์ r ตัว คือ $\beta^{r} = (\beta_r, \beta_{r+1}, \ldots, \beta_k)'$ ในกรณีนี้เราจะพบว่า

$$C\beta = \left[\begin{array}{ccc:cccc} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{array}\right] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ \vdots \\ \\ \beta_k \end{bmatrix} = \begin{bmatrix} \beta_r \\ \beta_{r+1} \\ \\ \vdots \\ \\ \beta_k \end{bmatrix}_{r\times 1} \text{ และ } C\hat{\beta} = \begin{bmatrix} \hat{\beta}_r \\ \hat{\beta}_{r+1} \\ \\ \vdots \\ \\ \hat{\beta}_k \end{bmatrix}_{r\times 1}$$

เขียนย่อเป็น $\hat{\beta}^{r}$

---

[1] จากรูปทั่วไปของ Quadric Surface คือ

$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + K = 0$$

เมื่อหมุนแกนไป α, θ และ γ องศาตามลำดับโดยแทนที่ x, y และ z ดังนี้คือ

$$x = x'\cos\alpha_1 + y'\cos\alpha_2 + z'\cos\alpha_3$$
$$y = x'\cos\theta_1 + y'\cos\theta_2 + z'\cos\theta_3$$
$$z = x'\cos\gamma_1 + y'\cos\gamma_2 + z'\cos\gamma_3$$

เราสามารถจัดรูปสมการนี้ซึ่งเป็นรูปทั่วไปเป็นรูปมาตรฐานคือ

$$\frac{x'^2}{a^2} + \frac{y'^2}{b^2} + \frac{z'^2}{c^2} = 1 \quad \text{หรือ} \quad \frac{(x'-h)^2}{a^2} + \frac{(y'-g)^2}{b^2} + \frac{(z'-l)^2}{c^2} = 1$$

เพื่อความสะดวกขอเปลี่ยนแปลงสัญญลักษณ์ดังนี้คือให้ $V = (X'X)^{-1}$ ดังนั้นสำหรับกรณีเฉพาะนี้จะเห็นว่า Partition เมตริกซ์ V ได้ดังนี้ คือ

$$V = \left[\begin{array}{c|c} V_{11} & V_{12} \\ \hline V_{21} & V_{22} \end{array}\right]_{r \times k}$$

เมื่อ V_{11} เป็น Submatrix ขนาด $(k-r)\times(k-r)$, V_{12} เป็น Submatrix ขนาด $(k-r)\times r$, V_{21} เป็น Sub matrix ขนาด $r\times(k-r)$ และ V_{22} เป็น Submatrix ขนาด $r\times r$

ดังนั้น $CVC' = V_{22}$ แสดงว่า Joint Confident Region ของ $\beta_r, \beta_{r+1}, \ldots, \beta_k$ คือ

$$(\beta^r - \hat{\beta}^r)' V_{22}^{-1} (\beta^r - \hat{\beta}^r) \leq r\hat{\sigma}^2 F_{r, n-k, 1-\alpha}$$

สำหรับกรณีนี้อาจเปลี่ยนแปลงไปได้มากมายตามความต้องการ ทั้งนี้ขึ้นอยู่กับวิธี Partition เมตริกซ์ C เป็นสำคัญ ขอให้พิจารณาตัวอย่างต่อไปนี้ ซึ่งจะแสดงวิธีการใช้ประโยชน์จากกรณีนี้

ตัวอย่าง 3.2 จากผลลัพธ์ในตัวอย่าง 3.1 จงสร้าง Joint Confidence Region ของ β_2 และ β_3 และ หา Confidence Limit Wı ง β_1

วิธีทำ ก. การสร้าง Confidence Region ของ β_2 และ β_3

$$C = \left[\begin{array}{c|cc} 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right], \quad V = (X'X)^{-1} = \left[\begin{array}{c|cc} 55.914 & -.0042 & -1.546 \\ \hline -.0042 & .0000037 & .0000773 \\ -1.546 & .0000773 & .0433 \end{array}\right]$$

$$C\beta = \left[\begin{array}{c|cc} 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} \quad \text{และ } C\hat{\beta} = \begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

$$V_{22} = CVC' = C(X'X)^{-1}C'$$

$$\Rightarrow V_{22} = \left[\begin{array}{c|cc} 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c|cc} -55.914 & -.0042 & 1.546 \\ \hline -.0042 & .0000037 & .0000773 \\ -1.546 & .0000773 & .0433 \end{array}\right]$$

$$= \begin{bmatrix} .0000037 & .0000773 \\ .0000773 & .0433 \end{bmatrix}$$

$$\Rightarrow \quad V_{22}^{-1} = \begin{bmatrix} 280{,}000 & -500 \\ -500 & 24 \end{bmatrix}$$

จึงพบว่าเขตเชื่อมั่นร่วมของ β_2 และ β_3 คือ

$$(\beta_2 - .07\,,\, \beta_3 - .60) \begin{bmatrix} 280{,}000 & -500 \\ -500 & 24 \end{bmatrix} \begin{bmatrix} \beta_2 - .07 \\ \beta_3 - .06 \end{bmatrix} \leqslant \quad 2(13.08)(6.94)$$

$$280{,}000\,(\beta_2 - .07)^2 - 1{,}000\,(\beta_2 - .07)(\beta_3 - .60) + 24\,(\beta_3 - .60)^2 \leqslant 181.55$$

จะเห็นว่า $A = 280{,}000$, $B = -1{,}000$, $C = 24$

ในการหมุนแกน β_2 และ β_3 ไป ϕ องศา เมื่อ $\cot 2\phi = \frac{A - C}{B}$ และแทนที่ β_2 และ β_3 โดยที่ $\beta_2 = \beta_2' \cos\phi - \beta_3' \sin\phi$ และ $\beta_3 = \beta_2' \sin\phi + \beta_3' \cos\phi$ สมการของเขตเชื่อมั่นจะเปลี่ยนรูปเป็นรูปมาตรฐานของสมการ Ellipse ที่มีจุดศูนย์กลางที่ (.07, .60) สำหรับการวาดรูปขอให้เป็นภาระของนักศึกษา

ข. การสร้าง Confidence Limit ของ β_1

ให้ $C = (1, 0, 0)$ ดังนั้น

$$C\beta = (1, 0, 0) \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \beta_1 \text{ หรือ } C\hat{\beta} = \hat{\beta}_1$$

$$CVC' = (1, 0, 0) \begin{bmatrix} 55.914 & -.0042 & -1.546 \\ -.0042 & .0000037 & .0000037 \\ -1.546 & .0000773 & .0433 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 55.914$$

ดังนั้น $(CVC')^{-1} = \dfrac{1}{55.914}$

ดังนั้น Confidence Limit ของ β_1 คือ $(\beta_1 - 11.33)\left(\frac{1}{55.914}\right)(\beta_1 - 11.33) \leqslant 1(13.08)(7.71)$

$$\text{หรือ } \frac{1}{55.914}\ (\beta_1 - 11.33)^2 \leqslant 100.85$$

$$(\beta_1 - 11.33)^2 \leqslant 5{,}638.93$$

$$\Rightarrow \quad |\beta_1 - 11.33| \leqslant 75.09$$

นั่นคือ $-63.76 \leqslant \beta_1 \leqslant 86.42$ หรือนัยหนึ่งเราสามารถเชื่อมั่นได้ถึง 95% ว่า Random Interval $(-63.76, 86.42)$ จะเป็นช่วงที่ครอบคลุมค่าจริงของ β_1 เอาไว้

หมายเหตุ โดยทั่วไปเราจะพบว่า Confidence Limit เสนอไว้ดังนี้ คือ

$$\hat{\beta}_j - t_{n-k,\,1-\frac{\alpha}{2}}\hat{\sigma}_{\hat{\beta}_j} < \beta_j < \hat{\beta}_j + t_{n-k,\,1-\alpha}\hat{\sigma}_{\hat{\beta}_j}$$

ความจริงรูปนี้ก็คือรูปเดียวกันกับที่เสนอไว้ข้างบน กล่าวคือ จากเขตเชื่อมั่นรูป $(\beta-\hat{\beta})(CVC')^{-1}(\beta-\hat{\beta}) \leqslant r\,\hat{\sigma}^2 F_{r,\,n-k,\,1-\alpha}$ เมื่อกำหนดให้ $C = (0, \dots 0, 1, 0, \dots, 0)$ รูปทั่วไปจะลดรูปเป็น $\frac{1}{v_j}(\beta_j - \hat{\beta}_j)^2 \leqslant 1\cdot\hat{\sigma}^2\cdot F_{1,\,n-k,\,1-\alpha}$ หรือ $(\beta_j - \hat{\beta}_j)^2 \leqslant v_j\,\hat{\sigma}^2\ F_{1,\,n-k,\,1-\alpha}$

เมื่อ v_j คือสมาชิกตัวที่ (j, j) ของเมตริกซ์ $(X'X)^{-1}$ และเนื่องจาก $F_{1,\,n-k} = t^2_{n-k}$ และ $V(\hat{\beta}_j) = v_j\,\hat{\sigma}^2$ ดังนั้น เขตเชื่อมั่นจึงกลายเป็น $(\beta_j - \hat{\beta}_j)^2 \leqslant \hat{V}(\hat{\beta}_j)\,t^2_{n-k}$ หรือ
$|\beta_j - \hat{\beta}_j| \leqslant t_{n-k,1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{\beta}_j)}$; $j = 1, 2, \dots, k$

ตัวอย่าง 3.3 ในกรณีของแบบจำลอง $Y = X\beta + U$ ซึ่งเราสามารถประมาณค่า β ได้ด้วย $\hat{\beta} = (X'X)^{-1}X'Y$ ทำให้ได้สมการคาดหมาย $E(Y) = X\hat{\beta}$ หรือที่นิยมเขียนกันโดยทั่วไปเป็น $\hat{Y} = X\hat{\beta}$ นั้น

ก. ถ้ากำหนดให้ค่าของ X's ในอนาคต ในรูปของเวคเตอร์ C คือ $C' = (1, X_{2,\,n+i}, X_{3,\,n+i}, \dots, X_{k,\,n+i})$ จงพิสูจน์ว่า ค่าพยากรณ์ของ Y ในอนาคตคือ $\hat{Y}_{n+i} = C'\hat{\beta}$ และ $\hat{V}(\hat{Y}_{n+i}) = C'\,\hat{V}(\hat{\beta})\,C$

ข. จากผลในข้อ ก. จงหาค่าช่วงพยากรณ์ (Prediction Interval) ของ Y_{n+i} สำหรับสมการ $Y = \alpha + \beta X + u$

วิธีทำ โดยอาศัยทฤษฎีเกี่ยวกับ **Multivariate Normal** Density เราย่อมทราบได้ทันทีว่า $C'\hat{\beta} \sim N(C'\beta, C'\hat{V}(\hat{\beta})C)$ โดยที่ $\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$ โดยอาศัยวิธีการที่กล่าวมาแล้วในตอน 3.2.2.4 แต่ในที่นี้จะแสดงวิธีทำอีกลักษณะหนึ่งโดยไม่อาศัยความรู้เรื่อง Multivariate Normal Density เป็นพื้นฐาน ทั้งนี้เพื่อให้นักศึกษามีโอกาสเรียนรู้วิธีพัฒนาเขตเชื่อมั่นโดยวิธีอื่น

ก. เมื่อ $C' = (1, X_{2,n+i}, X_{3,n+i}, \ldots, X_{k,n+i})$[1]; $i = 1, 2, 3 \ldots$

ดังนั้น

$$Y_{n+i} = (\hat{\beta}_1 + \hat{\beta}_2 X_{2,n+i} + \ldots + \hat{\beta}_k X_{k,n+i})$$

$$= (1, X_{2,n+i}, X_{3,n+i}, \ldots, X_{k,n+i})\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_k \end{bmatrix} = C'\hat{\beta}$$

จาก $\hat{Y}_{n+i} = C'\hat{\beta}$ เราจะพบว่า

1) $E(\hat{Y}_{n+i}) = E(C'\hat{\beta}) = C'E(\hat{\beta}) = C'\beta$ ซึ่งแสดงว่า

$\hat{Y}_{n+i} = \hat{\beta}_1 + \hat{\beta}_2 X_{2,n+i} + \hat{\beta}_3 X_{3,n+i} + \ldots + \hat{\beta}_k X_{k,n+i}$ เป็นค่าประมาณที่ใช้พยากรณ์

$Y_{n+i} = \beta_1 + \beta_2 X_{2,n+i} + \beta_3 X_{3,n+i} + \ldots + \beta_k X_{k,n+i}$ ได้ดีโดยไม่มีอคติ (Unbias)

2)
$$\begin{aligned} V(\hat{Y}_{n+i}) &= E\{(C'\hat{\beta} - C'\beta)(C'\hat{\beta} - C'\beta)'\} \\ &= E\{C'(\hat{\beta} - \beta)(\hat{\beta}'C - \beta'C)\} \\ &= E\{C'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'C\} \end{aligned}$$

---

[1] $C' = (1, X_{2,n+i}, X_{3,n+i}, \ldots, X_{k,n+i})$ คือค่าของ X's ณ period ที่ i นอก Sample period โดยที่ $i = 1, 2, 3, \ldots$ ตัวอย่างเช่นถ้าเราบันทึกข้อมูลอนุกรมเวลาไว้ 11 ปี ตั้งแต่ ปี พ.ศ. 2514-2524 ในที่นี้จะพบว่า n = 11 = พ.ศ. 2524 ดังนั้นค่าในอนาคตของ X_j คือ $X_{j,n+i}$ ถ้า i = 1 ค่า $X_{j,n+1}$ คือค่าของ X_j ใน พ.ศ. 2525 ถ้า i = 5 ค่า $X_{j,n+5}$ คือค่าของ X_j ในปี พ.ศ.2529 ซึ่งถ้าเราทราบค่าของ $X_{j,n+i}$ ในทุกค่าของ j และ i เราย่อมพยากรณ์ค่าของ Y_{n+i} ได้ โดยปกติเราสามารถทราบค่า $X_{j,n+i}$ ได้จากนโยบาย การคาดหวัง แนวโน้ม ตลอดจนถึงการกำหนดข้อตกลงร่วมบางประการ เช่น ถือว่าอัตราการเปลี่ยนแปลงคงที่เสมอ

$$= C'\{E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\}C$$
$$= C'V(\hat{\beta})C$$
$$= C'\sigma^2(X'X)^{-1}C$$

นั่นคือ $V(\hat{Y}_{n+l}) = \sigma^2 C'(X'X)^{-1}C$ และเมื่อประมาณค่า σ^2 ด้วย $\hat{\sigma}^2 = \dfrac{Y'Y - \hat{\beta}'X'Y}{n-k}$

จะพบว่า $\hat{V}(\hat{Y}_{n+l}) = \hat{\sigma}^2 C'(X'X)^{-1}C$[1]

และเนื่องจาก $C'\hat{\beta} \sim N(C'\beta, \hat{\sigma}^2 C'(X'X)^{-1}C)$ ดังนั้นตัวแปรสุ่ม $\dfrac{C'\hat{\beta}-C'\beta}{\hat{\sigma}\sqrt{C'(X'X)^{-1}C}}$ จึงมีการแจกแจงแบบ t มี $df = n-k$

และโดยอาศัยความรู้เกี่ยวกับช่วงเชื่อมั่น (Interval Estimation Theory) เราทราบว่า

$$Pr\left\{\left|\frac{C'\hat{\beta}-C'\beta}{\hat{\sigma}\sqrt{C'(X'X)^{-1}C}}\right| \leqslant t_{n-k,1-\alpha/2}\right\} = 1-\alpha$$

หรือ $$Pr\left\{-t_{n-k,1-\alpha/2} \leqslant \frac{C'\hat{\beta}-C'\beta}{\hat{\sigma}\sqrt{C'(X'X)^{-1}C}} \leqslant t_{n-k,1-\alpha/2}\right\} = 1-\alpha$$

$$Pr\left\{C'\hat{\beta} - t_{n-k,1-\alpha/2}\,\hat{\sigma}\sqrt{C'(X'X)^{-1}C} \leqslant C'\beta \leqslant C'\hat{\beta} + t_{n-k,1-\alpha/2}\,\hat{\sigma}\sqrt{C'(X'X)^{-1}C}\right\} = 1-\alpha$$

นั่นคือ ช่วงพยากรณ์ (หรือช่วงที่เราเชื่อมั่นได้ถึง $(1-\alpha)$ 100% ว่า เป็นช่วงที่จะครอบคลุมเอาค่าจริงของ Y_{n+l} ไว้) ของ $E(Y_{n+1})$ คือ

$$\left(C'\hat{\beta} - t_{n-k,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{C'(X'X)^{-1}C},\ C'\hat{\beta} + t_{n-k,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{C'(X'X)^{-1}C}\right)$$

ข. จากกรณีทั่วไปในข้อ ก. เราสามารถนำมาลดรูปใช้กับกรณีเฉพาะ เมื่อ $k = 2$ คือ กรณี Simple Regression สำหรับสมการ $Y = \alpha+\beta x+u$ ได้ดังนี้

ให้ $C' = (1, X_0)$ คือค่าในอนาคตของ X ณ จังหวะเวลาใด ๆ ที่ $X_{n+l} = X_0$ ดังนั้น

$$\hat{Y}_{n+l} = \hat{\alpha} + \hat{\beta}X_{n+l} = \hat{\alpha} + \hat{\beta}X_0 = (1, X_0)\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix} = C'\hat{\beta}$$

---

[1] ถ้ากำหนดให้ C เป็น Row Vector คือ $C = (1, X_{2,n+l}, \ldots, X_{k,n+l})$ เช่นที่นิยามในตอน 3.2.5 จะพบว่า $\hat{Y}_{n+l} = C\hat{\beta}$ และ $V(\hat{Y}_{n+l}) = \sigma^2 C(X'X)^{-1}C'$

และพบว่า $V(\hat{Y}_{n+i}) = \sigma^2 C'(X'X)^{-1} C = \sigma^2 (1, X_0) \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ X_0 \end{bmatrix}$

แต่ $\begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \overset{n}{\Sigma} X_i^2 \end{bmatrix}^{-1} = \dfrac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix} = \dfrac{1}{n\Sigma x_i^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ \Sigma X_i & n \end{bmatrix}$

เมื่อกำหนดให้ $x_i = X_i - \bar{X}$

ดังนั้น $V(\hat{Y}_{n+i}) = \dfrac{\sigma^2}{n\Sigma x_i^2} (1, X_0) \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix} \begin{bmatrix} 1 \\ X_0 \end{bmatrix}$

$$= \frac{\sigma^2}{n\Sigma x_i^2} \{ \sum_i^n X_i^2 + nX_0^2 - 2X_0 \sum_i^n X_i \}$$

$$= \frac{\sigma^2}{n\Sigma x_i^2} \sum_i^n (X_i - X_0)^2$$

$$= \frac{\sigma^2}{n\Sigma x_i^2} \sum_i^n (X_i - \bar{X} + \bar{X} - X_0)^2$$

$$= \frac{\sigma^2}{n\Sigma x_i^2} \left\{ \sum_i^n (X_i - \bar{X})^2 + n(\bar{X} - X_0)^2 \right\}^{1}$$

$$= \frac{\sigma^2}{n\Sigma x_i^2} \{ \sum_i^n x_i^2 + n(\bar{X} - X_0)^2 \}$$

$$= \sigma^2 \{ \frac{1}{n} + \frac{(\bar{X} - X_0)^2}{\Sigma x_i^2} \}$$

นั่นคือ $V(\hat{Y}_{n+i}) = \sigma^2 \{ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\Sigma x_i^2} \}$ และ $\hat{V}(\hat{Y}_{n+i}) = \hat{\sigma}^2 \{ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\Sigma x_i^2} \}$

โดยที่ $\hat{\sigma}^2 = \frac{1}{n-k} \sum_i^n e_i^2$ หรือ $\hat{\sigma}^2 = \frac{1}{n-2}(Y'Y - \hat{\beta}'X'Y) = \frac{1}{n-2}(\sum_i^n Y_i^2 - \hat{\alpha}\sum_i^n Y_i - \hat{\beta}\sum_i^n X_i Y_i)$

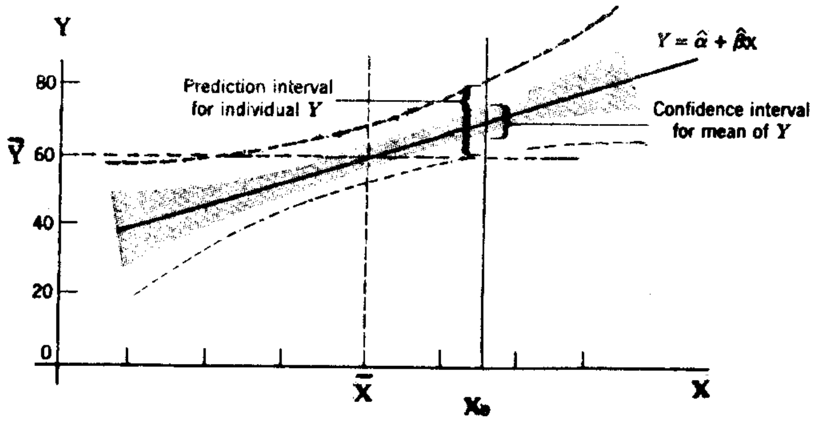
---

[1] เทอมไขว้คือ $2(X_i - X_0)\sum_i^n (X_i - \bar{X}) = 0$ เพราะ $\sum_i^n (X_i - \bar{X}) = 0$

ดังนั้นช่วงพยากรณ์ ณ ระดับความเชื่อมั่น $(1-\alpha)100\%$ของ Y_{n+i} หรือ $E(Y_{n+i})$ คือ

$$\left(\hat{\alpha}+\hat{\beta}X_{n+i}-t_{n-2,\,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}}\;,\;\hat{\alpha}+\hat{\beta}X_{n+i}+t_{n-2,\,1-\frac{\alpha}{2}}\,\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}}\right)$$

ข้อสังเกต จาก $\hat{V}(Y_{n+i}) = \hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}}$ จะพบว่า ถ้าค่าของ X_{n+i} หรือ X_0 ใกล้เคียงกับ $\bar{X}$ (ค่าเฉลี่ยของ X ใน Sample Period) $\hat{V}(Y_{n+i})$ จะมีค่าต่ำที่สุดทำให้ช่วงพยากรณ์ ณ จุด $X_0 = \bar{X}$ เป็นช่วงที่แคบที่สุด และเมื่อใดก็ตามที่ค่าของ X_0 แตกต่างไปจาก $\bar{X}$ มากขึ้น $\hat{V}(Y_{n+i})$ ก็จะมากขึ้น ช่วงพยากรณ์ก็จะกว้างขึ้น ซึ่งมีผลให้ความแม่นยำลดน้อยลง ขอให้สังเกตความเป็นจริงได้จากภาพข้างล่างนี้



จากภาพที่เห็นจึงมีข้อเตือนใจนักศึกษาข้อหนึ่ง ซึ่งนับว่าสำคัญมากคือ เราไม่ควรใช้สมการถดถอยพยากรณ์ค่า Y ในอนาคตให้ห่างไกลปัจจุบัน (ข้อมูลใน Sample Period) มากนัก เพราะช่วงพยากรณ์ของ Y ในอนาคตที่ไกลปัจจุบันมากจะเป็นช่วงที่กว้างมาก ค่าพยากรณ์ที่ได้จะขาดความแม่นยำ ทั้ง $\hat{\alpha}$ และ $\hat{\beta}$ เองก็ขาดความแม่นยำเพราะมีเขตเชื่อมั่นกว้างขวางมาก อีกประการหนึ่งเราไม่อาจคาดหมายได้ว่าในอนาคตที่ห่างไกลปัจจุบันนั้น แบบแผนความสัมพันธ์ $Y = f(X's)$ จะเปลี่ยนแปลงโครงสร้างไปจากเดิมที่พบใน Sample Period หรือไม่ เช่น แบบแผนความสัมพันธ์อาจเป็นรูปเส้นโค้งขณะที่ความสัมพันธ์ที่พบจากข้อมูลใน Sample Period เป็นรูปสมการเส้นตรง ซึ่งในกรณีเช่นนี้ถ้าเรานำสมการถดถอยไปพยากรณ์ค่า Y ในอนาคต

ค่าพยากรณ์ที่ได้จะคลาดเคลื่อนผิดความจริงไปมาก

หมายเหตุ ตัวอย่างนี้นิยาม C ในรูป Column Vector แต่ถ้านิยาม C ในรูปเมตริกซ์ขนาด $r \times k$ แล้วลดรูปให้ $r = 1$ จะได้ $C = (1, X_{2\,n+i}, X_{3\,n+i}, \ldots, X_{k\,n+i})$ เช่นที่ปฏิบัติมาในตอนต้น ซึ่งถ้านิยามในรูปนี้จะพบว่า $\hat{Y}_{n+i} = C\hat{\beta}$ และ $V(\hat{Y}_{n+i}) = \sigma^2 C(X'X)^{-1}C'$

3.2.5.3 การทดสอบสมมุติฐาน

การทดสอบสมมุติฐานในที่นี้หมายถึงการทดสอบดูว่าพารามิเตอร์ β มีค่าเท่ากับค่าใดค่าหนึ่งที่สนใจหรือไม่ ในที่นี้จะอธิบายแยกเป็น 2 กรณี คือ กรณีทดสอบว่า $\beta = 0$ กับ $\beta = \beta^*$ ดังนี้

พิจารณา $H_0 : \beta = 0$ หรือนัยหนึ่ง $H_0 : \beta_j = 0 ; j = 2, 3, \ldots, k$

การทดสอบสมมุติฐานรูปนี้มีจุดมุ่งหมายเพื่อศึกษาดูว่าตัวแปรอิสระ X_j , $j = 2, 3, \ldots, k$ มีอิทธิพลต่อความผันผวนหรือความเปลี่ยนแปลงในค่าของตัวแปรสุ่ม Y หรือไม่

พิจารณาแบบจำลองต่อไปนี้

$$E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k \beta_k$$

ขอให้สังเกตว่าตัวแปรอิสระ X_j จะเกาะติดอยู่กับพารามิเตอร์ β_j ถ้าหาก $\beta_j = 0$ ก็แสดงว่าเทอม $\beta_j X_j$ ไม่ปรากฏอยู่ในแบบจำลอง ซึ่งหมายความถึงว่า X_j มิได้มีอิทธิพลต่อความเปลี่ยนแปลงในค่าของตัวแปรสุ่ม Y ถ้า $\beta_j \neq 0$ ก็แสดงว่าเทอม $\beta_j X_j$ จะยังคงปรากฏอยู่ในแบบจำลองหรือ X_j มีอิทธิพลต่อความเปลี่ยนแปลงในค่าของตัวแปรอิสระ Y ดังนั้นโดยทั่วไปเราจึงมักสนใจที่จะตรวจสอบดูเสียก่อนว่า X_j มีอิทธิพลต่อ Y หรือไม่ซึ่งสามารถเสนอในรูปสมมุติฐานคือ $H_0 : \beta_j = 0$ VS $H_1 : \beta_j \neq 0 ; j = 1, 2, \ldots, k$[1] ถ้าปฏิเสธสมมุติฐาน (หรือมีนัยสำคัญทางสถิติ) ก็แสดงว่ายอมรับ $H_1 : \beta_j \neq 0$ หรือ X_j มีอิทธิพลต่อ Y ถ้ายอมรับสมมุติฐาน (ความจริงในทางทฤษฎีเราใช้คำว่า "ไม่อาจปฏิเสธ" ไม่นิยมใช้คำว่า "ยอมรับ" เพราะเกี่ยวข้องอยู่กับ β-error) ก็แสดงว่ายอมรับ $\beta_j = 0$ หรือ X_j ไม่มีอิทธิพลต่อ Y หรือ X_j ไม่ควรปรากฏอยู่ในแบบจำลอง

---

[1] โดยปกติเราจะไม่สนใจทดสอบสมมุติฐานเกี่ยวกับ β_1 เพราะมิได้ให้ประโยชน์ต่อการตีความแต่ประการใด แต่จะทดสอบด้วยก็ไม่มีอะไรเสียหาย

ปัญหาที่สืบเนื่องจากการปฏิเสธ $H_0 : \beta = 0$ ก็คือเมื่อเรายอมรับว่า $\beta \neq 0$ ซึ่งมีความหมายว่า $X_1, X_2, \ldots, X_k$ มีอิทธิพลต่อ Y[1] ก็คือ X_j มีอิทธิพลต่อ Y มากน้อยเพียงใด อย่าลืมว่าขนาด (magnitude) ของ β_j คือค่าที่แสดงระดับอิทธิพลของ X_j ที่มีต่อ Y ดังนั้นถ้าเราต้องการปรับปรุงแบบจำลองให้เหมาะสม การทดสอบว่า $H_0 : \beta = 0$ VS $H_1 : \beta \neq 0$ จึงไม่เพียงพอ สมควรทดสอบต่อไปว่าถ้า $\beta \neq 0$ แล้ว β เท่ากับเท่าไรกันแน่ ทางออกของปัญหานี้ก็คือทดสอบว่า $H_0 : \beta = \beta^*$ VS $H_1 : \beta = \beta^{**}$ ($\beta^{**} \neq \beta^*$ หรือ $\beta^{**} > \beta^*$ หรือ $\beta^{**} < \beta^*$) โดยที่ β^* และ β^{**} คือค่าของ β ที่พึงได้รับจาก Priori Information เช่น อาจได้รับจากงานที่มีผู้เคยทำการศึกษาวิจัยเอาไว้ หรือจากงานสำรวจ

การศึกษาเรื่องการทดสอบสมมุติฐานในที่นี้ผู้เขียนจะเสนอไว้เป็น 3 วิธี โดยจะเริ่มจากกรณีทั่วไปเสมอ คือ $H_0 : C\beta = \gamma_0$ VS $H_1 : C\beta \neq \gamma_0$ ส่วนกรณีเฉพาะให้นักศึกษาทดลองกระทำเองโดยวิธีเปลี่ยนรูปเมตริกซ์ C

การทดสอบสมมุติฐาน $H_0 : C\beta = \gamma_0$ VS $H_1 : C\beta \neq \gamma_0$ โดยที่ γ_0 คือเวคเตอร์ของค่าคงที่ที่สงสัยว่าค่าจริงของ β จะเท่ากับค่าดังกล่าว และ C เป็นเมตริกซ์ใด ๆ ขนาด $r \times k$, $r \leq k$ ที่มี Full Rank สามารถกระทำได้หลายวิธีดังนี้

วิธีที่ 1 สร้างเขตเชื่อมั่นสำหรับ $C\beta$ โดยอาศัยสมการ

$$(C\beta - C\hat{\beta})' [C(X'X)^{-1}C']^{-1} (C\beta - C\hat{\beta}) \leq \hat{\sigma}^2 r F_{r, n-k, 1-\alpha}$$

แล้วตรวจดูว่าเขตเชื่อมั่นดังกล่าวนี้ครอบคลุมเวคเตอร์ γ_0 ไว้หรือไม่ ถ้า γ_0 เป็นค่าที่ปรากฏในเขตเชื่อมั่นเราจะยอมรับสมมุติฐาน ถ้า γ_0 ไม่ปรากฏในเขตเชื่อมั่นเราจะปฏิเสธสมมุติฐาน

ตัวอย่างเช่นในกรณีของแบบจำลอง $E(Y) = \beta_0 + \beta_1 X$ โดยมีสมมุติฐานว่า

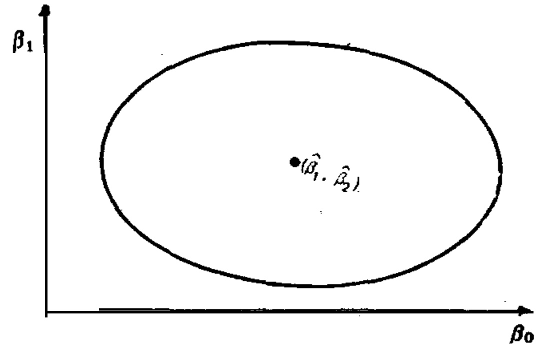
$$H_0 : \beta_0 = 2, \beta_1 = 3$$

$$\text{หรือ } H_0 : \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

สมมุติว่า $\hat{\beta}_0 = 1.8$, $\hat{\beta}_1 = 2$ และเขตเชื่อมั่นร่วมของ $C\beta$ ปรากฏดังภาพ

---

[1] $X_1 = (1, 1, 1, \ldots, 1)'$ การยอมรับหรือปฏิเสธเกี่ยวกับ β_1 จึงไม่มีประโยชน์ในการอภิปรายผล ในทางปฏิบัติเราจึงนิยมแยก β_1 ออกจากทดสอบ หรือขจัด β_1 ทิ้งไปโดยใช้ข้อมูลในรูป Deviated Form



จะเห็นว่าเมื่อเราพล็อตเวคเตอร์ $\gamma_0 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ ลงบนระนาบ $\beta_0\beta_1$ ปรากฏว่า γ_0 ปรากฏอยู่ในเขตเชื่อมั่น ในกรณีเช่นนี้ก็เป็นการยืนยันว่า $C\beta = \gamma_0$ จริง (หรือยอมรับ H_0)

วิธีนี้อาจไม่สะดวกนักโดยเฉพาะอย่างยิ่งเมื่อมีตัวแปรอิสระมากกว่า 2 ตัว เพราะจะมีปัญหาเรื่องการวาดภาพ Joint Confidence Region

วิธีที่ 2 อาศัยสมการ $\Pr\left\{\frac{(C\beta - C\hat{\beta})'[C(X'X)^{-1}C']^{-1}(C\beta - C\hat{\beta})}{r\hat{\sigma}^2} \leqslant F_{r,n-k,1-\alpha}\right\} = 1-\alpha$

จากสมการ $\Pr\left\{\frac{(C\beta - C\hat{\beta})'[C(X'X)^{-1}C']^{-1}(C\beta - C\hat{\beta})}{r\hat{\sigma}^2} \leqslant F_{r,n-k,1-\alpha}\right\} = 1-\alpha$ เมื่อกำหนดสมมุติฐานว่า $H_0 : C\beta = \gamma_0$ ดังนั้นเมื่อแทนที่ $C\beta$ ในสมการดังกล่าวด้วย γ_0 เราจึงปฏิเสธสมมุติฐาน $H_0 : C\beta = \gamma_0$ เมื่อ

$$(\gamma_0 - C\hat{\beta})'(C(X'X)^{-1}C')^{-1}(\gamma_0 - C\hat{\beta}) \leqslant r\hat{\sigma}^2 F_{r,n-k,1-\alpha}$$

ตัวอย่าง 3.3 ระยะทางที่อนุภาคเคลื่อนผ่านสื่อชนิดหนึ่งจะผันแปรไปตามระยะเวลาที่ใช้และปริมาณความชื้นของสื่อชนิดนั้นปรากฏดังสมการต่อไปนี้

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

เมื่อ Y = ระยะทางที่อนุภาคเคลื่อนที่ X_2 = เวลาที่อนุภาคเคลื่อนที่ X_3 = ความชื้นของสื่อ

ผลการบันทึกข้อมูลปรากฏดังนี้

Y	6.0	13.0	13.0	29.2	33.1	32.0	46.2	117.5
X_2	1	2	3	4	5	6	8	20
X_3	10	10	12	11	14	15	18	30

ก. จงกะประมาณสมการ $E(Y) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$

ข. จงทดสอบสมมุติฐาน $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ VS $H_1 : \beta_j \neq 0 ;\ j = 1, 2, 3$

$H_0 : \beta_1 = 25, \beta_2 = 5, \beta_3 = 0$ VS $H_1 : \beta_1 \neq 25, \beta_2 \neq 5, \beta_3 \neq 0$ และ

$H_0 : \beta_1 - \beta_2 + 2\beta_3 = 10$ VS $H_1 : \beta_1 - \beta_2 + 2\beta_3 \neq 10$

วิธีทำ ก. จากข้อมูลพบว่า

$$X'X = \begin{bmatrix} 8 & 49 & 120 \\ 49 & 555 & 1014 \\ 120 & 1014 & 2110 \end{bmatrix}, \quad (X'X)^{-1} = \begin{bmatrix} 15.2589 & 1.9536 & -1.8066 \\ 1.9536 & .2649 & -.23841 \\ -1.8066 & -.23841 & .21779 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 290 \\ 3,264.9 \\ 5,967.2 \end{bmatrix}, \quad Y'Y = 19,286.9 \quad , \bar{Y} = \frac{290}{8}$$

$$\text{ดังนั้น } \hat{\beta} = (X'X)^{-1} X'Y = \begin{bmatrix} 22.79 \\ 8.79 \\ -2.69 \end{bmatrix} \quad , \hat{\beta}' X'Y = 19,240.5$$

แสดงว่า $E(Y) = 22.79 + 8.79 X_2 - 2.69 X_3$

โดยที่ $\hat{\sigma}^2 = \frac{1}{8-3}(19,286.9 - 19,240.5) = \frac{46.4}{5} = 9.28$

$$R^2 = \frac{\hat{\beta}' X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} = \frac{19,240.5 - 10,512.5}{19,286.9 - 10,512.5} = .9947 = 99.47\%$$

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = 9.28 \begin{bmatrix} 15.2589 & 1.9536 & -1.8066 \\ 1.9536 & .2649 & -.23841 \\ -1.8066 & -.23841 & .21779 \end{bmatrix}$$

ดังนั้น $\hat{V}(\hat{\beta}_1) = 141.6$, $\hat{V}(\hat{\beta}_2) = 2.46$; $\hat{V}(\hat{\beta}_3) = 2.02$, $s_{\hat{\beta}_1} = 11.899$, $s_{\hat{\beta}_2} = 1.57$, $s_{\hat{\beta}_3} = 1.42$

และเนื่องจาก $t_c = \hat{\beta}_j / s_{\hat{\beta}_j}$; $j = 1, 2, 3$ สำหรับ $H_0 : \beta_j = 0$ VS $H_1 : \beta_j \neq 0$; $j = 1, 2, 3$

โดยที่ $t_{n-k, 1-\frac{\alpha}{2}} = t_{5, .975} = 2.571$ และ $t_{5, .995} = 4.032$

ดังนั้นเราจึงสามารถเสนอรูปที่สมบูรณ์ของสมการถดถอยได้ดังนี้

$$\begin{array}{llll} E(Y) = 22.79 + & 8.79 X_2 & - 2.69 X_3 & : R^2 = 99.47\% \\ (11.899) & (1.57) & (1.42) & \\ (1.92) & (5.59^{**}) & (1.89) & \end{array}$$

หมายเหตุ ค่าในวงเล็บแถวที่ 1 หมายถึง Standard Error คือ $s_{\hat{\beta}_1}$, $s_{\hat{\beta}_2}$ และ $s_{\hat{\beta}_3}$ ตามลำดับ ส่วนในแถวที่ 2 หมายถึง t_c

ข. การทดสอบสมมุติฐานกระทำดังนี้

1. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ VS $H_1 : \beta_j \neq 0$; $j = 1, 2, 3$ หรือ

$$H_0 : \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{VS } H_1 : \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

แสดงว่า $C = I_3$ และ $r_0 = 0$ และจากรูปทั่วไปคือจะปฏิเสธ $H_0 : C\beta = r_0$ VS $H_1 : C\beta \neq r_0$ เมื่อ $(r_0 - C\hat{\beta})' [C(X'X)^{-1} C']^{-1} (r_0 - C\hat{\beta}) \leqslant r\hat{\sigma}^2 F_{r, n-k, 1-\alpha}$ เราจึงปฏิเสธสมมุติฐานนี้เมื่อ $\hat{\beta}'(X'X)\beta \leqslant r\hat{\sigma}^2 F_{3, 5, 1-\alpha}$ เมื่อ r คือ Rank ของเมตริกซ์ C

$$\hat{\beta}'(X'X)\hat{\beta} = (22.79, 8.79, -2.69) \begin{bmatrix} 8 & 49 & 120 \\ 49 & 555 & 1,014 \\ 120 & 1,014 & 2,110 \end{bmatrix} \begin{bmatrix} 22.79 \\ 8.79 \\ -2.69 \end{bmatrix} = 19,271.09$$

$$r\hat{\sigma}^2 F_{3,5,.95} = 3(9.28)(5.4095) = 150.6$$

จะเห็นว่า $\hat{\beta}'(X'X)\hat{\beta}$ มีค่ามากกว่า $r\hat{\sigma}^2 F_{3,5,95}$ เราจึงไม่อาจปฏิเสธสมมุติฐานหลัก

2. $H_0 : \beta_1 = 25, \beta_2 = 5, \beta_3 = 0$ VS $H_1 : \beta_1 \neq 25, \beta_2 \neq 5, \beta_3 \neq 0$

$$\text{หรือ } H_0 : \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 25 \\ 5 \\ 0 \end{bmatrix}$$

$$\text{VS } H_1 : \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \neq \begin{bmatrix} 25 \\ 5 \\ 0 \end{bmatrix}$$

จะเห็นว่า $C = I_3$ และ $\gamma_0 = \begin{bmatrix} 25 \\ 5 \\ 0 \end{bmatrix}$ ดังนั้น $\gamma_0 - C\hat{\beta} = \begin{bmatrix} 25 \\ 5 \\ 0 \end{bmatrix} - \begin{bmatrix} 22.79 \\ 8.79 \\ -2.69 \end{bmatrix} = \begin{bmatrix} 2.21 \\ -3.79 \\ 2.69 \end{bmatrix}$

จะพบว่า

$$(\gamma_0 - C\hat{\beta})'(X'X)(\gamma_0 - C\hat{\beta}) = (2.21, -3.79, 2.69) \begin{bmatrix} 8 & 49 & 120 \\ 49 & 555 & 1,014 \\ 120 & 1,014 & 2,110 \end{bmatrix} \begin{bmatrix} 2.21 \\ -3.79 \\ 2.69 \end{bmatrix} = 3,209.594$$

และ $r\hat{\sigma}^2 F_{3,5,.95} = 3(9.28)(5.4095) = 150.6$

จะเห็นว่า $(\gamma_0 - C\hat{\beta})'(X'X)(\gamma_0 - C\hat{\beta})$ มีค่ามากกว่า $r\hat{\sigma}^2 F_{3,5,.95}$ ดังนั้นเราจึงไม่อาจปฏิเสธสมมุติฐานหลัก

3. $H_0 : \beta_1 - \beta_2 + 2\beta_3 =$ VS $H_1 : \beta_1 - \beta_2 + 2\beta_3 \neq 0$

$$\text{หรือ } (1, -1, 2) \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = 0 \quad \text{VS } H_1 : (1, -1, 2) \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \neq 0$$

จะเห็นว่า $C = (1, -1, 2)$, $\gamma_0 = 0$ ดังนั้น $\gamma_0 - C\hat{\beta} = 0 - (1, -1, 2) \begin{bmatrix} 22.79 \\ 8.79 \\ -2.69 \end{bmatrix} = -8.62$

จะเห็นว่า $(\gamma_0 - C\hat{\beta})' [C(X'X)^{-1}C']^{-1} (\gamma_0 - C\hat{\beta}) =$

$$(-8.62) \left[(1, -1, 2) \begin{bmatrix} 15.2589 & 1.9536 & -1.8066 \\ 1.9536 & .2649 & -.23841 \\ -1.8066 & -.23841 & .21779 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \right]^{-1} (-8.62)$$

$$= \frac{(8.62)^2}{.1609} = 461.805$$

และ $r\hat{\sigma}^2 F_{r, n-k, 1-\alpha} = 1(9.28) F_{1, 5, .95} = 1(9.28)(6.6079) = 61.32$

จะเห็นว่า $(\gamma_0 - C\hat{\beta})' [C(X'X)^{-1}C']^{-1} (\gamma_0 - C\hat{\beta})$ มีค่าสูงกว่า $r\hat{\sigma}^2 F_{1, 5, .95}$ เราจึงไม่อาจปฏิเสธสมมุติฐานหลัก

วิธีที่ 3 ทดสอบสมมุติฐานโดยอาศัยตารางวิเคราะห์ความแปรปรวน (ANOVA)

การใช้วิธี ANOVA ในที่นี้ผู้เขียนขอเสนอเป็น 2 วิธี คือ $H_0 : \beta = \beta^*$ VS $H_1 : \beta \neq \beta^*$ กรณีหนึ่ง กับการทดสอบ Subhypothesis อีกกรณีหนึ่ง สำหรับกรณีแรกคือ

$$H_0 : \beta = \beta^* \text{ VS } H_1 : \beta \neq \beta^*$$

เป็นสมมุติฐานทั่วไปเช่นเดียวกันกับ $H_0 : C\beta = \gamma_0$ เมื่อ $C = I_k$ กรณีเฉพาะของวิธีนี้คือกรณีที่ $\beta^* = 0$ กล่าวคือ $H_0 : \beta = 0$ VS $H_1 : \beta \neq 0$ ซึ่งสอดคล้องกับสมมุติฐาน $H_0 : C\beta = \gamma_0$ VS $H_1 : C\beta \neq \gamma_0$ เมื่อ $C = I_k$ และ $\gamma_0 = 0$ ส่วนกรณีที่ 2 คือกรณีของ Subhypothesis หมายถึงกรณีที่เรามุ่งสนใจทดสอบเฉพาะบางส่วนของ β โดยไม่สนใจว่าส่วนที่เหลือจะมีความหมายความสำคัญเพียงใด กรณีเฉพาะก็คือการทดสอบว่า $\beta_2 = \beta_3 = \cdots = \beta_k = 0$ โดยไม่สนใจ β_1

กรณีที่ 1 $H_0 : \beta = \beta^*$ VS $H_1 : \beta \neq \beta^*$

ตาราง ANOVA ปรากฏดังนี้

ANOVA[1]

$H_0 : \beta = \beta^* \text{ VS } H_1 : \beta \neq \beta^*$

SOV	df	SS	MS
Total	n	$Q = (Y-X\beta^*)'(Y-X\beta^*)$	
Due to β (Regression)	k	$Q_2 = (Y-X\beta^*)' X(X'X)^{-1} X' (Y-X\beta^*)$	Q_2/k
Error (Residual)	n-k	$Q_1 = Y'Y-\hat{\beta}'X'Y = Q-Q_2$	$Q_1/n-k = \hat{\sigma}^2$ หรือ s^2

ถ้าสมมุติฐานหลักเป็นจริง $\dfrac{Q_2/k}{Q_1/n-k} = F_c$ จะมีการแจกแจงแบบ $F_{k,n-k}$ สำหรับกรณีเฉพาะ คือ เมื่อ $\beta^* = 0$ หรือ $\beta = 0$ VS $H_1 : \beta \neq 0$ ตาราง ANOVA จะลดรูปลงดังนี้

ANOVA

$H_0 : \beta = 0 \text{ VS } H_1 : \beta \neq 0$

SOV	df	SS	MS
Total	n	$Q = Y'Y$	
Due to β (Regression)	k	$Q_2 = \hat{\beta}'X'Y$	Q_2/k
Error (Residual)	n-k	$Q_1 = Y'Y-\hat{\beta}'X'Y = Q-Q_2$	$Q_1/n-k = \sigma^2$ หรือ s^2

สมการ $E(Y) = \beta_1 + \beta_2 X_2 + ... + \beta_k X_k$ จะสามารถนำไปใช้พยากรณ์ค่า E(Y) ได้อย่างน่าพึงพอใจ ถ้าหากว่า $F_c > F_{k,n-k}$ อย่างน้อย 4 เท่า[2] ทั้งนี้มิได้หมายความว่า ถ้า $F_c > F_{k,n-k}$ ไม่ถึง

[1] การพิสูจน์ความเป็นมากระทำโดยอาศัย Maximum Likelihood Ratio Test ผู้สนใจสามารถศึกษา รายละเอียดได้จาก Graybill. F.A., Introduction to Linear Statistical Model V.1, (Mc Graw - Hill Book Company, Inc., NY), p.128-139

[2] Drapper, N.R. and Smith, H., Applied Regression Analysis, (John Wiley and Sons, Inc., NY), p.64

4 เท่าแล้วสมการนี้จะใช้พยากรณ์ไม่ได้เพียงแต่น่าพึงพอใจหรือไม่เท่านั้นที่ยังน่าสงสัยอยู่

ตัวอย่าง 3.4 โดยอาศัยข้อมูลในตัวอย่าง 3.2 จงทดสอบสมมุติฐานต่อไปนี้

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \ \text{VS}\ H_1 : \beta_j \neq 0 ; j = 1, 2, 3$$

วิธีทำ จากสมมุติฐานแสดงว่า $\beta^* = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

ดังนั้น $Q = Y'Y = 19{,}286.9$, $Q_2 = \hat{\beta}'X'Y = 19{,}240.5$, $Q_1 = Q - Q_2 = 46.4$ ตาราง ANOVA ปรากฏดังนี้

ANOVA

$H_0 : \beta = 0$ VS $H_1\ \beta \neq 0$

SOV	df	SS	MS	F
Total	8	19,286.9		
Due to β	3	19,240.5	6,413.5	691.1'
Error	5	46.4	9.28	

$F_{3, 5, .95} = 5.4095$

จะเห็นว่า $F_c > 5.4095$ เราจึงปฏิเสธสมมุติฐานหลักและเชื่อว่าตัวแปรอิสระ X_1, X_2, X_3 ต่างก็มีอิทธิพลร่วมต่อความผันแปรในค่าของตัวแปรตาม Y ขอให้สังเกตว่า F_c มีค่ามากกว่า F_{tab} เกินกว่า 4 เท่า ทำให้เชื่อว่าแบบจำลองนี้น่าจะมีประโยชน์ในเชิงพยากรณ์ได้ดีถ้าไม่มีปัญหาใน Second Order Test

ผลจากการวิเคราะห์มีข้อที่พึงสังเกตประการหนึ่งคือ β_1 เข้าไปมีส่วนร่วมในการอภิปรายผลด้วยทั้ง ๆ ที่ X_1 มิได้มีความหมายในเชิงอภิปรายมากนักในทางปฏิบัติแล้วเราควรขจัด β_1 (ซึ่งก็คือสัมประสิทธิ์ของ X_1) ออกไปจากการวิเคราะห์เสียก่อน บางครั้งเราอาจมีความจำเป็นต้องผลักดัน β หลายตัวออกไปจากการมีส่วนร่วมในการวิเคราะห์ เช่น กรณี Sequential F – test และ

Partial F - test ด้วยเหตุผลประการดังกล่าวจึงมีผลนำไปสู่การพัฒนาวิธีทดสอบ **Subhypothesis** ซึ่งจะกล่าวถึงในลำดับต่อไปนี้

กรณีที่ 2 การทดสอบ Subhypothesis[1]

จากสมการ $Y = X\beta + U$ เรา Partition เมตริกซ์ X และเวคเตอร์ β ออกได้ดังนี้คือ

$$X = (X_1, X_2) \text{ และ } \beta = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$$

โดยที่ X_1 เป็น Submatrix ขนาด $n \times r$ และ X_2 เป็น Submatrix ขนาด $n \times (k-r)$ ขณะเดียวกันเราก็ Partition β ให้มีลักษณะที่ทำให้การคูณเมตริกซ์ $X\beta$ เป็นไปตามนิยามของการคูณ กล่าวคือ γ_1 เป็นเวคเตอร์ขนาด $r \times 1$ และ γ_2 เป็นเวคเตอร์ขนาด $(k-r) \times 1$ ดังไดอะแกรม

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \left[\begin{array}{cccc:ccc} 1 & X_{11} & \cdots & X_{r1} & X_{r+1,1} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{r2} & X_{r+1,2} & \cdots & X_{k2} \\ \vdots & & & & \vdots & & \\ 1 & X_{1n} & \cdots & X_{rn} & X_{r+1,n} & \cdots & X_{kn} \end{array}\right] \left[\begin{array}{c} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \\ \hdashline \beta_{r+1} \\ \vdots \\ \beta_k \end{array}\right]$$

r col　　　　　　k - r col

ดังนั้น $Y = X\beta + U = (X_1, X_2)(\gamma_1, \gamma_2)' + U = X_1\gamma_1 + X_2\gamma_2 + U$

ในที่นี้ผู้เขียนจะเสนอเฉพาะกรณีของ $H_0 : \gamma_1 = 0$ เท่านั้น จะไม่กล่าวถึง $H_0 : \gamma_1 = \gamma_1^*$ แต่ประการใด

จากสมการ $Y = X\beta + U$ เราสามารถคำนวณหา $\hat{\beta}$ ได้จากสมการนอร์มอล $(X'X)\hat{\beta} = X'Y$ และพบว่า $\hat{\beta} = (X'X)^{-1}X'Y$ และความผันแปรของ Y อันเนื่องมาจาก X คือ $\hat{\beta}'X'Y$

จากสมการ $Y = X_1\gamma_1 + X_2\gamma_2 + U$ เมื่อสมมุติฐานหลักเป็นจริงคือ $\gamma_1 = 0$ จะพบว่า $Y = X_2\gamma_2 + U$ และเราสามารถคำนวณหา $\hat{\gamma}_2$ ได้จากสมการนอร์มอล $(X_2' X_2)\hat{\gamma}_2 = X_2' Y$ และพบว่า $\hat{\gamma}_2 = (X_2' X_2)^{-1} X_2' Y$ และความผันแปรอันเนื่องมาจาก X_2 คือ $\hat{\gamma}_2' X_2' Y$

---

[1] ศึกษารายละเอียดเพิ่มเติมจาก Graybill, F.A., Ibid., p.133-140

ดังนั้น ความผันแปรอันเนื่องมาจากเฉพาะอิทธิพลของกลุ่ม X_1 เมื่อขจัดอิทธิพลของกลุ่ม X_2 ออกแล้ว คือ $\hat{\beta}'X'Y - \hat{\gamma}_2'X_2'Y$

ดังนั้นตารางวิเคราะห์ความแปรปรวนของ Subhypothesis $\gamma_1 = 0$ คือ

ANOVA

$H_0 : \gamma_1 = 0$ VS $H_1 : \gamma_1 \neq 0$

SOV	df	SS	MS	F
Total	n	$Y'Y$		
Due to β	k	$\hat{\beta}'X'Y$		
Due to γ_2	k - r	$\hat{\gamma}_2'X_2'Y$		
Due to γ_1	r	$\hat{\beta}'X'Y - \hat{\gamma}_2'X_2'Y = Q_1$	Q_1/r	$\frac{n-k}{r}\frac{Q_1}{Q_0}$
Error	n - k	$Y'Y - \hat{\beta}'X'Y = Q_0$	$Q_0/n-k$	

ตัวอย่าง 3.5 โดยอาศัยข้อมูลจากตัวอย่าง 3.2 จงทดสอบสมมุติฐานว่า $\beta_2 = \beta_3 = 0$

วิธีทำ จากข้อมูลตามตัวอย่าง 3.2 สำหรับแบบจำลอง $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$; $i = 1, 2, \ldots, 8$ จะพบว่า

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \hdashline \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$$

$$=> \quad Y = X_1\gamma_1 + X_2\gamma_2 + U$$

เนื่องจากสมมุติฐานหลักระบุว่า $\beta_2 = \beta_3 = 0$ หรือ $\begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ แสดงว่า $\gamma_2 = 0$

ดังนั้นแบบจำลอง $Y = X\beta + U$ จึงลดรูปลงเป็น $Y = X_1\gamma_1 + U$ เมื่อ H_0 จริง

ซึ่งในกรณีของสมการลดรูป $Y = X_1\gamma_1 + U$ หรือ $Y_i = \beta_1 X_{1i} + u_i$; $i = 1, 2, \ldots, n$ เราจะพบ

ว่าสมการนอร์มอลคือ $(X_1' X_1)\hat{\beta}_1 = X_1' Y$ เมื่อ $X_1 = (1, 1, 1, 1, 1, 1, 1, 1)'$ ดังนั้น $X_1' X_1 = 8$ และ $X_1' Y = \sum_i^8 Y_i = 290$

$$= > \quad 8\hat{\beta}_1 = 290 \text{ หรือ } \hat{\beta}_1 = \frac{290}{8} = 36.3 = \sum^n Y_i / n = \bar{Y}$$

ดังนั้น $\hat{\gamma}_1' X_1' Y = (36.3)(290) = (\frac{290}{8})\, 290 = 8\, (\frac{290}{8})^2 = 10{,}512.5 = n\bar{Y}^2$

และจากแบบจำลอง $Y = X\beta + U$ เราสามารถคำนวณหา $\hat{\beta}$ ได้จากสมการนอร์มอล $(X'X)\hat{\beta} = X'Y$ หรือ $\hat{\beta} = (X'X)^{-1}X'Y$ และพบว่า $\hat{\beta}'X'Y = 19{,}240.5$

ดังนั้นอิทธิพลของเฉพาะ X_2 และ X_3 คือ $\hat{\beta}'X'Y - \hat{\gamma}_1' X_1' Y$ = 19,2405 - 10,512.5 = 8,728

ตาราง ANOVA สำหรับสมมุติฐาน $\gamma_2 = 0$ จึงปรากฏดังนี้

ANOVA

$H_0 : \beta_2 = \beta_3 = 0$ VS $H_1 : \beta_j \neq 0$; j = 2, 3

sov	df	SS	MS	F
Total	8	19,286.9		
Due to β	3	19,240.5		
Due to γ_1	1	10,512.5		
Due to γ_2 (หรือ $\gamma_2 \mid \gamma_1$)	2	8,728.0	4,364	470.3 *
Error	8-3 = 5	46.4	9.28	

$F_{2,5,.95} = 5.786$

จะเห็นว่า $F_c > 5.786$ เราจึงปฏิเสธสมมุติฐานหลักและเชื่อว่าระยะเวลาที่อนุภาคใช้ในการเคลื่อนที่และปริมาณความชื้นของสื่อมีอิทธิพลต่อความผันแปรของระยะทางที่อนุภาคพึงเคลื่อนที่ได้

ผลจากตัวอย่างที่ 3.4 นำไปสู่ตาราง ANOVA ที่นิยมใช้กันอย่างกว้างขวางในทางปฏิบัติ ซึ่งผู้วิจัยต้องแยกอิทธิพลของ X_1 ซึ่งไม่มีความหมายในเชิงอภิปรายผลออกเสียก่อนดังนี้

ANOVA

$H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0 \text{ VS } H_1 : \beta_j \neq 0 ; j = 2, 3, \ldots, k$

SOV	df	ss	MS	F
Total	n	$Y'Y$		
β_1 (mean)	1	$n\bar{Y}^2$		
$\beta_1, \beta_2, \beta_3, \ldots, \beta_k$	k	$\hat{\beta}'X'Y$		
$\beta_2, \beta_3, \ldots, \beta_k$	k-1	$\hat{\beta}'X'Y - n\bar{Y}^2 = Q_1$	$Q_1/k-1$	$\frac{n-k}{k-1}\frac{Q_1}{Q_0}$
Error (หรือ Residual)	n-k	$Y'Y - \hat{\beta}'X'Y = Q_0$	$Q_0/k-1$	

$$R^2 = \frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2}$$

3.2.6 การปรับข้อมูลให้อยู่รูปค่าเบี่ยงเบนจากค่าเฉลี่ย (Deviation From Mean)

การวิเคราะห์สมการถดถอยโดยใช้ข้อมูลเดิม (Initial form หรือ Deviation from Zero) มีปัญหาสำคัญอยู่ประการหนึ่งคือข้อมูลเดิมมักเป็นปริมาณค่อนข้างสูง ซึ่งมีผลให้ผลรวมกำลังสอง (Sum Square) และผลรวมเทอมไขว้ (Sum of Cross Product Term) มีค่าสูงมาก ทำให้การ**วิเคราะห์ข้อมูลดำเนินไปได้โดยยาก** แม้จะใช้คอมพิวเตอร์เข้าช่วยบางครั้งก็มีปัญหา เพราะข้อมูลที่เกี่ยวข้องมีจำนวนหลักทศนิยมเกินกว่าจำนวนหลักที่เครื่องจะรับได้โดยเฉพาะในไมโครคอมพิวเตอร์ ทางเลือกสำหรับปัญหานี้ก็คือการวิเคราะห์ข้อมูลโดยอาศัยข้อมูลที่แปลงรูปแล้วอาศัย**วิธีย้ายแกนสู่แกนที่มีจุดกำเนิดเป็น** $\bar{Z} = (\bar{Y}, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_k)$ หรือในทางปฏิบัติการคือการเปลี่ยน**ข้อมูล** Y_i **และ** X_{ji} **เป็น** $y_i = Y_i - \bar{Y}$ และ $x_{ji} = X_{ji} - \bar{X}_j$; $j = 2, 3, \ldots, k$ ตามลำดับ วิธีนี้นอกจาก**ทำให้ขนาด (Magnitude) ของข้อมูลเล็กลง** ซึ่งมีผลให้เราวิเคราะห์ข้อมูลง่ายกว่าเดิมแล้วยังมีผลดี**ที่สำคัญ 2 ประการดังนี้**

1. การปรับค่าข้อมูลมีผลให้ขนาดของเมตริกซ์ $X'X$ ลดขนาดลงจาก $n \times n$ เป็น $(n-1) \times (n-1)$ ทำให้การคำนวณหาส่วนกลับคือ $(X'X)^{-1}$ กระทำได้ง่ายขึ้น ที่กล่าวว่าง่ายเพราะ

การปรับข้อมูลทำให้สมาชิกของเมตริกซ์ $(X'X)$ มีขนาดเล็กลงทั้งขนาดของเมตริกซ์ก็ลดลงไป 1 มิติ การหา $(X'X)^{-1}$ จึงกระทำได้ง่ายกว่าเดิมเป็นอย่างมาก

2. การปรับข้อมูลทำให้สมาชิกของเมตริกซ์ $(X'X)$ มีลักษณะใกล้เคียงกับ Correlation Matrix เพียงแต่เรานำรากที่สองของสมาชิกใน Main Diagonal ของ $(X'X)$ ที่สอดคล้องกันไปปรับค่าตามสูตร

$$r_{pq} = \frac{\sum_{i}^{n} x_{pi} x_{qi}}{\sqrt{\sum_{i}^{n} x_{pi}^2 \sum_{i}^{n} x_{qi}^2}} \quad ; p, q = 2, 3, \ldots, k$$

เราก็จะได้ Correlation Matrix หรือ Correlation Table ตามต้องการ รายละเอียดเรื่องนี้จะได้กล่าวถึงอีกครั้งในบทที่ 4

สำหรับการพัฒนาวิธีประมาณค่าตามนัยแห่งการปรับข้อมูลในที่นี้จะขอกล่าวถึงเฉพาะในส่วนที่สัมพันธ์สืบเนื่องกับ ANOVA ที่ต่อเนื่องมาจากตอน 3.2.5 รายละเอียดอื่น ๆ จะกล่าวถึงอีกครั้ง

จากระบบสมการ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \; ; i = 1, 2, \ldots, n$ ... (1)

เมื่อรวมตลอดในทุกค่าของ i แล้วหารตลอดด้วย n จะได้

$$\bar{Y} = \beta_1 + \beta_2 \bar{X}_2 + \beta_3 \bar{X}_3 + \ldots + \beta_k \bar{X}_k + \bar{u} \quad \ldots (2)$$

(1) - (2) จะได้ $Y_i - \bar{Y} = \beta_2 (X_{2i} - \bar{X}_2) + \beta_3 (X_{3i} - \bar{X}_3) + \ldots + \beta_k (X_{ki} - \bar{X}_k) + (u_i - \bar{u})$

หรือ $y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + (u_i - \bar{u}) \; ; i = 1, 2, \ldots, n$ ... (3)

จากสมการที่ (3) จะเห็นว่าเป็นระบบสมการ ซึ่งสามารถเสนอในรูปเมตริกซ์ได้ดังนี้

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_{21} & x_{31} & x_{41} & \cdots & x_{k1} \\ x_{22} & x_{32} & x_{42} & \cdots & x_{k2} \\ \cdot & & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & & \cdot & & \\ x_{2n} & x_{3n} & x_{4n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 - \bar{u} \\ u_2 - \bar{u} \\ \cdot \\ \cdot \\ \cdot \\ u_n - \bar{u} \end{bmatrix}$$

หรือ $Y = X\beta + (U - \bar{U})$

เมื่อให้ $\hat{\beta}$ คือ OLS ของ β แล้วแทนที่ β ด้วย $\hat{\beta}$ จะได้สมการใหม่คือ $Y = X\hat{\beta} + e$ โดยที่ e คือค่าประมาณของ $(U - \bar{U})$

จากวิธี OLS เราทราบว่า $\sum_{i}^{n} e_i^2 = e'e = (Y - X\hat{\beta})'(Y - X\hat{\beta})$ ซึ่งเราสามารถ Minimize $\sum_{i}^{n} e_i^2$ โดยการดิฟเฟอร์เรนซิเอท เทียบต่อ $\hat{\beta}$ และพบว่า $\hat{\beta} = (X'X)^{-1}X'Y$

ในทำนองเดียวกันกับตอน 3.2 เราสามารถพิสูจน์ได้ว่า

1. $E(\hat{\beta}) = \beta$ และ $\hat{\beta}$ เป็น BLUE ของ β

2. $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$

3. $\hat{\sigma}^2 = \frac{1}{n-k} e'e = \frac{1}{n-k}(Y'Y - \hat{\beta}'X'Y) = \frac{1}{n-k}\left(\sum_{i}^{n} y_i^2 - \sum_{j=2}^{k}\sum_{i=1}^{n}\hat{\beta}_j x_{ji} y_i\right)$

โดยที่ $X'X = \begin{bmatrix} x_{21} & x_{22} & \cdots & x_{2n} \\ x_{31} & x_{32} & \cdots & x_{3n} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} x_{21} & x_{31} & \cdots & k_{k1} \\ x_{22} & x_{32} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{2n} & x_{3n} & \cdots & x_{kn} \end{bmatrix}$

$$= \begin{bmatrix} \sum x_{2i}^2 & \sum x_{2i}x_{3i} & \cdots & \sum x_{2i}x_{ki} \\ \sum x_{2i}x_{3i} & \sum x_{3i}^2 & \cdots & \sum x_{3i}x_{ki} \\ \vdots & \vdots & & \vdots \\ \sum x_{2i}x_{ki} & \sum x_{3i}x_{ki} & \cdots & \sum x_{ki}^2 \end{bmatrix}_{(k-1)\times(k-1)}$$

นอกจากนี้ เรายังสามารถพิสูจน์ได้โดยง่ายว่า

Total SS = Explained SS (หรือ Regression SS) + Unexplained SS (หรือ Residual SS)

หรือ $Y'Y = \hat{\beta}'X'Y + e'e$

หรือ $\sum y_i^2 = (\hat{\beta}_2 \sum x_{2i} y_i + \hat{\beta}_3 \sum x_{3i} y_i + \ldots + \hat{\beta}_k \sum x_{ki} y_i) + e'e$

ขอให้สังเกตว่าค่าของ Sum of Square ทุกแหล่งได้เสนอในรูป Deviation Form ด้วยเหตุนี้เราจึงสามารถนำความสัมพันธ์ดังกล่าวไปสร้างเป็นตาราง ANOVA ได้ทันทีดังนี้

ANOVA

$H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$ VS $H_1 : \beta_j$ ไม่เป็น 0 ทั้งหมด

SOV	df	ss	MS	F_c
Regression	k - 1	$\hat{\beta}'X'Y$ $(= \hat{\beta}_2 \Sigma x_{2i} y_i + \ldots + \hat{\beta}_k \Sigma x_{ki} y_i)$	$\frac{\hat{\beta}'X'Y}{k-1}$	$\frac{(n-k)\,\hat{\beta}'X'Y}{(k-1)(Y'Y - \hat{\beta}'X'Y)}$
Residual	n - k	$Y'Y - \hat{\beta}'X'Y$ $(= \Sigma y_i^2 - \hat{\beta}_2 \Sigma x_{2i} y_i - \ldots - \hat{\beta}_k \Sigma x_{ki} y_i)$	$\frac{Y'Y - \hat{\beta}'X'Y}{n-k}$	
Total	n - 1	$Y'Y = \Sigma y_i^2$		

จะเห็นได้ว่าแหล่งความผันแปรต่าง ๆ มีลักษณะเดิมแต่ Sum of Square เปลี่ยนแปลงสู่รูปของ Deviation Form เรียบร้อยอยู่ในตัวโดยไม่จำเป็นต้องนำ $CF = n\bar{Y}^2$ ไปหักลบออกเช่นที่เคยปฏิบัติในตารางก่อน

$$\text{ดังนั้น } R^2 = \frac{\text{Regression SS}}{\text{Total SS}} = \frac{\hat{\beta}'X'Y}{Y'Y} = \frac{\hat{\beta}_2 \Sigma x_{2i} y_i + \hat{\beta}_3 \Sigma x_{3i} y_i + \ldots + \hat{\beta}_k \Sigma x_{ki} y_i}{\Sigma y_i^2}$$

$$\text{และ } \bar{R}^2 = 1 - \frac{(Y'Y - \hat{\beta}'X'Y)/n-k}{Y'Y/n-1}$$

จะเห็นว่าการประมาณค่าสมการถดถอยวิธีนี้ไม่มี $\hat{\beta}_1$ ปรากฏอยู่ในสมการ แต่ความจริงแล้ว $\hat{\beta}_1$ ยังคงปรากฏอยู่เพียงแต่แอบแฝงอยู่ทำให้เรามองไม่เห็นตัวอย่างเด่นชัดเท่านั้น เรื่องนี้สามารถแสดงให้เห็นได้โดยง่ายดังนี้

จากสมการประมาณค่า $y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \ldots + \hat{\beta}_k x_{ki}$ ------- (1)

ถ้าย้อนคืนสู่รูปเดิมคือ เสนอตัวแปรในรูปของ Deviation From Zero จะพบว่าสมการที่ (1) สามารถคืนสู่รูปเดิมโดยอาศัยความจริงที่ว่า $y_i = Y_i - \bar{Y}$ และ $x_{ji} = X_{ji} - \bar{X}_j$; j = 2, 3, ... , k ดังนี้

$$Y_i - \bar{Y} = \hat{\beta}_2(X_{2i} - \bar{X}_2) + \hat{\beta}_3(X_{3i} - \bar{X}_3) + \ldots + \hat{\beta}_k(X_{ki} - \bar{X}_k)$$

$$\Rightarrow \quad Y_i = (\bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3 - \ldots - \hat{\beta}_k\bar{X}_k) + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki}$$

แต่ $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3 - \ldots - \hat{\beta}_k\bar{X}_k$ ดังนั้นสมการที่ (1) จึงปรับรูปย้อนสู่รูปเดิมได้เป็น

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki} \qquad \ldots\ldots (2)$$

ความจริงข้อนี้ชี้ให้เห็นอย่างเด่นชัดว่า การปรับข้อมูลสู่ Deviation from Mean นั้นมีประโยชน์หลายประการดังที่กล่าวมาทั้งยังมิได้ทำให้ข้อเท็จจริงเดิมเปลี่ยนแปลงไปแต่ประการใด และขอย้ำไว้อีกครั้งหนึ่งว่า ถ้านักศึกษาวิเคราะห์สมการถดถอยโดยนัยแห่งสมการที่ (1) แล้วนักศึกษาอย่าได้ประมาณค่า β_1 อีกเพราะจะทำให้เกิดการประมาณค่าซ้ำซ้อนและผิดความจริง ทั้งนี้เพราะสมการที่ (1) และสมการที่ (2) ก็คือสมการเดียวกันเพียงแต่เทียบต่อแกน Coordinate คนละชุดเท่านั้น $\hat{\beta}_1$ ยังคงปรากฏอยู่ในสมการที่ (1) ไม่เปลี่ยนแปลงเพียงแต่ **แฝงตัวอยู่จน**เรามองไม่เห็นเท่านั้น และจากความจริงข้อนี้เราจึงมีสิทธิจะปรับข้อมูลเดิมโดยการย้ายแกนไปสู่แกนใด ๆ ที่มีจุดกำเนิดใด ๆ ได้เสมอโดยมิได้ทำให้สูญเสียข้อเท็จจริงแต่ประการใด แต่ประโยชน์ที่พึงบังเกิดจากการย้ายแกนสู่แกนที่มีจุดกำเนิดใด ๆ ที่มิใช่ $\bar{Z} = (\bar{Y}, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_k)$ นั้นมีประโยชน์ข้างเคียงน้อยกว่าจึงยังไม่เป็นที่นิยม

ก่อนที่จะผ่านตอนนี้ไปผู้เขียนขอแสดงการพิสูจน์เพื่อแสดงให้เห็นว่าในกรณีของการย้ายแกนสู่แกนชุดใหม่ที่มีจุดกำเนิดเป็น $\bar{Z}$ นั้น $\hat{\sigma}^2$ ยังคงมีค่าเท่ากับ $\frac{e'e}{n-k}$ เหมือนในกรณียังไม่ย้ายแกน เหตุที่จำเป็นต้องพิสูจน์เพราะนักศึกษามักเคลือบแคลงและมีเหตุชวนให้เชื่อว่า $\hat{\sigma}^2$ มีค่าเท่ากับ $\frac{e'e}{n-k-1}$ มากกว่าที่จะเป็น $\frac{e'e}{n-k}$

พิสูจน์ เพื่อประโยชน์ในเชิงทฤษฎี ผู้เขียนจะพิสูจน์ให้เห็นเป็น 2 ประเด็นคือ

$$\frac{Y'Y - \hat{\beta}'X'Y}{\sigma^2} \sim \chi^2_{(n-k)} \text{ และ } \frac{\hat{\beta}'X'Y/k-1}{(Y'Y - \hat{\beta}'X'Y)/n-k} \sim F_{k-1,\,n-k}$$

1. จาก Residual Sum Square $\sum_{i}^{n} y_i^2 - \hat{\beta}_2 \sum_{i}^{n} x_{2i} y_i - \hat{\beta}_3 \sum_{i}^{n} x_{3i} y_i - \ldots - \hat{\beta}_k \sum_{i}^{n} x_{ki} y_i$

จะพบว่า

$$Q = \sum_{i}^{n} y_i^2 - \hat{\beta}_2 \sum_{i}^{n} x_{2i} y_i - \hat{\beta}_3 \sum_{i}^{n} x_{3i} y_i - \ldots - \hat{\beta}_k \sum_{i}^{n} x_{ki} y_i$$

$$= \sum_{i}^{n} (Y_i - \bar{Y})^2 - \hat{\beta}_2^2 \sum_{i}^{n} x_{2i}^2 - \hat{\beta}_3^2 \sum_{i}^{n} x_{3i}^2 - \ldots - \hat{\beta}_k^2 \sum_{i}^{n} x_{ki}^2 \quad [1]$$

$$\therefore \frac{Q}{\sigma^2} = \sum_{i}^{n} \frac{(Y_i - \bar{Y})^2}{\sigma^2} - \frac{\hat{\beta}_2^2}{\sigma^2/\sum x_{2i}^2} - \frac{\hat{\beta}_3^2}{\sigma^2/\sum x_{3i}^2} - \ldots - \frac{\hat{\beta}_k^2}{\sigma^2/\sum x_{ki}^2}$$

แต่ตามสมมุติฐานหลักระบุว่า $\beta_j = 0 \,;\, j = 2, 3, \ldots, k$ ดังนั้น

$$\frac{Q}{\sigma^2} = \sum_{i}^{n} \left(\frac{Y_i - \bar{Y}}{\sigma}\right)^2 - \frac{(\hat{\beta}_2 - \beta_2)^2}{\sigma^2/\sum x_{2i}^2} - \frac{(\hat{\beta}_3 - \beta_3)^2}{\sigma^2/\sum x_{3i}^2} - \ldots - \frac{(\hat{\beta}_k - \beta_k)^2}{\sigma^2/\sum x_{ki}^2}$$

และเนื่องจาก $\hat{\beta}_j \sim N\left(\beta_j, \frac{\sigma^2}{\sum x_{ji}^2}\right)$ หรือ $Z_j = \frac{\hat{\beta}_j - \beta_j}{\sigma/\sqrt{\sum x_{ji}^2}} \sim N(0, 1)$

$$\Rightarrow \quad \frac{Q}{\sigma^2} + \sum_{j=2}^{k} Z_j^2 = \frac{(n-1) S_y^2}{\sigma^2}$$

แต่ $\frac{(n-1) S_y^2}{\sigma^2} \sim \chi^2_{(n-1)}$ และ $\sum_{j=2}^{k} Z_j^2 \sim \chi^2_{(k-1)}$ ดังนั้น $\frac{Q}{\sigma^2} \sim \chi^2_{(n-1-k+1)}$

นั่นคือ $\frac{Q}{\sigma^2} \sim \chi^2_{(n-k)}$

2. $Q_1 = \hat{\beta}' X' Y = \sum_{j=2}^{k} (\hat{\beta}_j \sum_{i}^{n} x_{ji} y_i) = \sum_{j=2}^{k} (\hat{\beta}_j^2 \sum_{i}^{n} x_{ji}^2)$

ดังนั้น $\frac{Q_1}{\sigma^2} = \sum_{j=2}^{k} \frac{\hat{\beta}_j^2}{\sigma^2/\sum x_{ji}^2}$

$$= \sum_{j=2}^{k} \left(\frac{\hat{\beta}_j - \beta_j}{\sigma/\sqrt{\sum x_{ji}^2}}\right)^2 = \sum_{j=2}^{k} Z_j^2 \quad \text{นั่นคือ} \quad \frac{Q}{\sigma^2} \sim \chi^2_{(k-1)}$$

3. $\lambda = \frac{\hat{\beta}' X' Y / k-1}{(Y'Y - \hat{\beta}' X' Y)/n-k} = \frac{Q_1/(k-1)\sigma^2}{Q/(n-k)\sigma^2}$ นั่นคือ $\lambda \sim F_{k-1, n-k}$

---

[1] $\because \quad \hat{\beta}_j = \frac{\sum x_{ji} y_i}{\sum x_{ji}^2}$ ดังนั้น $\sum_{i}^{n} x_{ji} y_i = \hat{\beta}_j \sum_{i}^{n} x_{ji}^2 \,;\, j = 2, 3, \ldots, k$

3.2.7 การปรับข้อมูลให้อยู่ในรูปมาตรฐาน (standardized from)

ความจริงแล้วเรื่องนี้เป็นเรื่องที่กระทำต่อเนื่องกับตอน 3.2.6 นั่นเอง การปรับข้อมูลให้เป็นรูปมาตรฐานนั้นเรามีเหตุผล 2 ประการ ตามขั้นดำเนินการคือ ประการแรกเพื่อลดขนาดของข้อมูลของ Y และ X's ลงด้วยการหักออกด้วยค่ามัชฌิมเลขคณิตคือ จาก Y เป็น y และจาก X_{ji} เป็น x_{ji}; $j = 2, 3, \ldots, k$ การลดขนาด (magnitude) ของข้อมูลลงทำให้เราบริหารข้อมูลได้ง่าย แต่ยังไม่อาจแก้ปัญหาได้ทั้งหมด กล่าวคือ ตราบใดก็ตามที่ตัวแปรมีมาตราการชั่งตวงวัด (measurement scale) ต่างกัน มูลค่าของตัวแปรก็จะต่างกัน ค่าประมาณของพารามิเตอร์ β_j ก็จะยังคงเปรียบเทียบกันไม่ได้ จึงให้ดำเนินการประการที่สองคือหาร (ถ่วงน้ำหนัก) ด้วย s_j ผลลัพธ์ก็คือ X_j กลายเป็น $\frac{X_{ji} - \bar{X}_j}{s_j} = z_{ji}$ เรียกว่าตัวแปรมาตรฐาน ค่าของ z_j จะปรากฏในช่วง -3 ถึง $+3$ ด้วยความน่าจะเป็น .99 เมื่อเป็นเช่นนี้คือตัวแปรทุกตัวมีค่าอยู่ในช่วงเดียวกัน ค่าประมาณของ β_j จึงเทียบกันได้ในระหว่าง j ในผลลัพธ์จาก SPSS จะพบสดมภ์ชื่อ B สำหรับแสดงว่า $\hat{\beta}_j$ กรณีข้อมูลเดิมและสดมภ์ชื่อ beta สำหรับแสดงว่า $\hat{\beta}_j$ กรณีข้อมูลรูปมาตรฐาน การพัฒนาสมการสำหรับข้อมูลลักษณะนี้ปรากฏดังนี้

จากระบบสมการ $Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i;\ i = 1, 2, \ldots, n$ ....(1)

ให้จัดเป็น deviated form ตามตอน 3.2.6 ได้

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + (u_i - \bar{u});\ i = 1, 2, \ldots, n \qquad \ldots\ldots\ldots(2)$$

จากสมการ (2) ให้หารตลอดด้วย σ_y และนำ σ_j/σ_j คูณด้านหน้า x_{ji}; $j = 2, 3, \ldots, k$ ได้สมการ (3) ดังนี้

$$\frac{y_i}{\sigma_y} = \left(\beta_2 \frac{\sigma_2}{\sigma_y}\right)\frac{x_{2i}}{\sigma_2} + \left(\beta_3 \frac{\sigma_3}{\sigma_y}\right)\frac{x_{3i}}{\sigma_3} + \ldots.$$

$$+ \left(\beta_k \frac{\sigma_k}{\sigma_y}\right)\frac{x_{ki}}{\sigma_k} + \frac{u_i - \bar{u}}{\sigma_n};\ i = 1, 2, \ldots, n \qquad \ldots\ldots\ldots(3)$$

แต่ $\frac{x_{ji}}{\sigma_j} = z_{ji}$; $j = 2, 3, \ldots, k$

$$\text{และ } \rho_{jy} = \frac{\Sigma x_{ji} y_i}{\sqrt{\Sigma x_{ji}^2\, \Sigma y_i^2}} = \frac{\Sigma x_{ji} y_i}{\sqrt{\Sigma x_{ji}^2\, \Sigma y_i^2}} \frac{\sqrt{\Sigma x_{ji}^2}}{\sqrt{\Sigma x_{ji}^2}}$$

$$= \frac{\Sigma x_{ji} y_i}{\Sigma x_{ji}^2} \frac{\sqrt{\Sigma x_{ji}^2}}{\sqrt{\Sigma y_i^2}} = \beta_j \frac{\sigma_j}{\sigma_y};\ j = 2, 3, \ldots, k$$

ด้วยเหตุนี้สมการ (3) จึงเปลี่ยนรูปเป็น

$$z_{yi} = \rho_2 z_{2i} + \rho_3 z_{3i} + \dots + \rho_k z_{ki} + z_{ui} \; ; \; i = 1, 2, \dots, n \qquad \text{..........(4)}$$

พึงสังเกตว่า จากสมการ (4) นั้น ตัวแปรเดิมทุกตัวถูกแปลงเป็นรูปมาตรฐาน ซึ่งกรณีนี้ค่าของตัวแปรจะวิ่งอยู่ในช่อง −3 ถึง +3 ด้วยความน่าจะเป็นไม่น้อยกว่า .99 และสัมประสิทธิ์ β_j' เปลี่ยนเป็น ρ_j ซึ่งเราทราบว่า ρ_j นั้น มีค่าอยู่ระหว่าง −1 ถึง +1 ด้วยเหตุนี้ค่าประมาณสัมประสิทธิ์ของสมการที่ (4) จึงเป็นค่าที่เปรียบเทียบกันได้ ทำให้เราอ่านได้จากสมการถดถอยโดยทันทีว่า ตัวแปรใดมีอิทธิพลต่อ Y มากกว่ากัน ข้อที่ขอเตือนให้ระวังในเรื่องนี้ก็คือ การเปรียบเทียบให้เปรียบเทียบเฉพาะขนาดอย่าสนใจเครื่องหมาย (คือให้มองในรูป absolute) ที่เป็นเช่นนี้เพราะ ρ_{jy} ที่ติดลบแปลว่า X_j กับ Y สัมพันธ์กันในทางขัดกัน ขณะที่ ρ_{jy} เป็นบวก แปลว่า X_j กับ Y สัมพันธ์กันในทางส่งเสริมกัน มิได้แปลว่าค่า ρ ที่เป็นลบ มีค่าน้อยกว่าที่เป็นบวก เช่น $\rho_{2y} = -.84$ และ $\rho_{3y} = +.49$ ให้แปลว่า X_2 มีอิทธิพลต่อ Y สูงกว่า X_3 ประมาณ 2 เท่า เป็นต้น

อนึ่ง จากสมการ (4) เราสามารถจัดเป็นรูปเมตริกซ์ได้ดังนี้ คือ

$$\begin{bmatrix} z_{y1} \\ z_{y2} \\ \vdots \\ z_{yn} \end{bmatrix} = \begin{bmatrix} z_{21} & z_{31} & \cdots & z_{k1} \\ z_{22} & z_{32} & \cdots & z_{k2} \\ \vdots & \vdots & & \vdots \\ z_{2n} & z_{3n} & \cdots & z_{kn} \end{bmatrix} \begin{bmatrix} \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix} + \begin{bmatrix} z_{u1} \\ z_{u2} \\ \vdots \\ z_{un} \end{bmatrix}$$

หรือ
$$\underset{(n \times 1)}{Z_y} = \underset{(n \times k-1)}{Z_x} \; \underset{(k-1 \times 1)}{\rho} + \underset{(n \times 1)}{Z_u} \qquad \text{.......(5)}$$

จะพบว่า $Z'_x Z_x = \underset{(k-1 \times k-1)}{nR_x}$ เป็น correlation matrix ของตัวแปรอิสระ

$Z'_x Z_y = \underset{(k-1 \times 1)}{nR_{xy}}$ เป็น correlation vector ของตัวแปรอิสระกับตัวแปรตาม

$Z'_y Z_y = n$

ด้วยเหตุนี้เราจึงพบว่า

$$\hat{\rho} = R_x^{-1} R_{xy}$$

$$s^2 = \frac{1}{n-k+1}[Z'_y Z_y - \hat{\rho}' Z'_x Z_y]$$

$$= \frac{n}{n-k+1}[1 - R'_{xy} R_x^{-1} R_{xy}]$$

$$\hat{V}(\hat{\rho}) = \frac{s^2}{n} R_x^{-1}$$

หมายความว่า การวิเคราะห์สมการถดถอยโดยอาศัยสมการ (5) จะทำได้ง่ายกว่าสมการที่ (1) มาก เพราะเพียงแต่เราสามารถทราบ correlation matrix เราก็สามารถวิเคราะห์สมการถดถอยได้ตั้งแต่ต้นจนจบ เรื่องนี้ขอให้นักศึกษาขยายผลการศึกษาและพิสูจน์สูตรต่าง ๆ ไว้ในรูปของ R_x และ R_{xy} ต่อไปเป็นแบบฝึกหัด

3.3 Generalize Least Square (GLS)[1]

ในสถานการณ์ทางปฏิบัติเราอาจพบว่า $E(u_i u_j) \neq 0$ ในทุกค่าของ $i \neq j$ และ $E(u_i u_j) \neq \sigma^2$ ในทุกค่าของ $i = j$ เช่นใน Dependent Lagged Model, Autoregressive **Model หรือแม้แต่ในกรณี**ที่มีปัญหา **Heteroscedasticity ซึ่งในสถานการณ์เช่นนั้น $E(UU')$ จะมิใช่ $\sigma^2 I_n$ ซึ่งเป็น Scalar Matrix** แต่กลับพบว่า $E(UU') = \sigma^2\Omega$ โดยที่ Ω เป็น Symmetric Positive Definite **Matrix**[2] **ขนาด** $n \times n$ และพบว่า $U \sim N(0, \sigma^2\Omega)$ ดังนั้น

$$L = f_u(u) = \frac{1}{\sigma^n (2\pi)^{n/2}} \frac{1}{|\Omega|^{\frac{1}{2}}} \exp\{-\tfrac{1}{2}\frac{1}{\sigma^2} U'\Omega^{-1}U\}$$

การประมาณค่าพารามิเตอร์ β สำหรับสถานการณ์นี้ ถ้าเราจะยังคงประมาณค่า β ด้วย $\hat{\beta} = (X'X)^{-1} X'Y$ ก็ยังคงใช้ได้เพราะ $\hat{\beta}$ จะยังเป็น **Unbiased Estimator แต่ $\hat{\beta}$ จะไม่เป็น Minimum Variance Estimator**

การประมาณค่า β สำหรับสถานการณ์ที่ **$E(UU') = \sigma^2\Omega$ ในที่นี้ผู้เขียนจะเสนอไว้เป็น** 2 วิธี ดังนี้

---

[1] เรียกอีกอย่างหนึ่งว่า Aitken's GLS ตามนามของ A.C. Aitken ซึ่งเสนอเรื่องนี้ไว้เป็นคนแรกในปีคริสต์ศักราช 1934

[2] *Positive Definite Matrix* คือเมตริกซ์ที่ให้ค่า *Characteristic Value* เป็นบวกทุกค่า โดยปกติใน *GLS* เราถือว่า σ^2 เป็นพารามิเตอร์ที่ไม่ทราบค่าขณะที่ Ω เป็นเมตริกซ์ที่เราต้องทราบค่าของสมาชิกทุกตัว มิเช่นนั้น $\hat{\beta} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y$ จะไม่เป็นตัวสถิติ แต่ในทางปฏิบัติเรามักไม่ทราบค่าสมาชิกของ Ω ซึ่งจะทราบได้ก็โดยการประมาณค่าสมาชิกของ Ω ด้วยวิธีที่เหมาะสม เช่น วิธีของเซลเนอร์-ทีเซล และวิธีของคาแกน การประมวลผลโดยใช้ค่าประมาณ $\hat{\Omega}$ เรียกว่า Estimated GLS (EGLS)

วิธีที่ 1 อาศัย Transformation Technique

เนื่องจาก $E(UU') = \sigma^2\Omega$ โดยที่ Ω เป็น Symmetric Positive Definite Matrix เราจึงสามารถหาเมตริกซ์ P ที่ทำให้ $PP' = \Omega$ ได้[1]

จากสมการ $PP' = \Omega$ จะพบว่า $P^{-1}\Omega(P^{-1})' = I_n$ และ $\Omega^{-1} = (P')^{-1}P^{-1}$ ซึ่งจากความจริงข้อนี้ทำให้เราสามารถแปลงรูปสมการ $Y = X\beta + U$ ที่ $E(UU') = \sigma^2\Omega$ ได้ดังนี้

จากสมการ $Y = X\beta + U$

คูณด้านหน้า (Premultiply) ตลอดด้วย P^{-1} จะได้ $P^{-1}Y = P^{-1}X\beta + P^{-1}U$ หรือ $Y^* = X^*\beta + U^*$ โดยที่ $Y^* = P^{-1}Y$, $X^* = P^{-1}X$ และ $U^* = P^{-1}U$[2]

จากสมการที่แปลงรูปแล้วคือ $Y^* = X^*\beta + U^*$ จะพบว่า

$$\begin{aligned} E(U^*U^{*\prime}) &= E\{P^{-1}UU'(P')^{-1}\} \\ &= P^{-1}E(UU')(P')^{-1} \\ &= \sigma^2 P^{-1}\Omega(P')^{-1} \\ &= \sigma^2 I'' \end{aligned}$$

ซึ่งแสดงให้เห็นว่าสมการ $Y^* = X^*\beta + U^*$ มีคุณสมบัติสอดคล้องกับข้อตกลงว่า $E(u_iu_j) = 0;\ i \neq j$ และ $E(u_iu_j) = \sigma^2;\ i = j$ เราจึงสามารถประมาณค่าพารามิเตอร์ β ได้โดยวิธี OLS ดังที่กล่าวมาแล้วในตอนต้นดังนี้

---

[1] จากหลักเกณฑ์ของพีชคณิตเชิงเส้น

ให้ A เป็น Positive Definite Matrix ขนาด $n \times n$ เราย่อมสามารถหา Characteristive Root ของ A ได้จากสมการ $|A-\lambda I_n| = 0$ ได้ $\lambda_i > 0;\ i = 1,2, \ldots\ldots, n$ และสามารถหา Characteristic Vector ได้จากสมการ $(A-\lambda_i I_n)X_i = 0;\ i = 1, 2, \ldots, n$ และ $X_i = (x_1, x_2, \ldots\ldots, x_n)'$ ซึ่งเมื่อนำ Characteristic Vector มาประกอบเป็นเมตริกซ์ จะได้ Orthogonal Modal Matrix Q กล่าวคือ $QQ' = In$ หรือ $Q' = Q^{-1}$

ดังนั้น $Q'AQ = \text{diag.}(\lambda_1, \lambda_2, \ldots\ldots, \lambda_n) = \Lambda$

ให้ $D = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \ldots, \frac{1}{\sqrt{\lambda_n}}\right)$ และนำ D คูณด้านหน้า (Premultiply) และคูณด้านหลัง (Post-multiply) สมการ $Q'AQ = \Lambda$ จะได้ $(QD)'A(QD) = I_n$ หรือ $A = \{(QD)'\}^{-1}(QD)^{-1} = \{(QD)^{-1}\}'(QD)^{-1}$ และเมื่อให้ $\{(QD))^{-1}\}' = P$ จึงพบว่า $A = PP'$

[2] วิธีปฏิบัติก็คือนำเมตริกซ์ P^{-1} คูณด้านหน้าสมการ $Y = X\beta$ แล้ววิเคราะห์ข้อมูลตามวิธี OLS

ให้ $\hat{\beta}$ เป็น OLS-estimator ของ β และเมื่อแทนที่ $\hat{\beta}$ ลงในสมการ $Y^* = X^*\beta + U^*$ จะได้

$$Y^* = X^*\hat{\beta} + e \text{ และ } e = Y^* - X^*\hat{\beta}$$

โดยอาศัยเทคนิคการ Minimize Σe^2 จะพบว่า

$$
\begin{aligned}
(1) \quad \hat{\beta} &= (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* \\
&= \{(P^{-1}X)'(P^{-1}X)\}^{-1}\{(P^{-1}X)'(P^{-1}Y)\} \\
&= \{X'(P')^{-1}P^{-1}X\}^{-1}\{X'(P')^{-1}P^{-1}Y\} \\
&= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)
\end{aligned}
$$

$$
\begin{aligned}
(2) \quad V(\hat{\beta}) &= \sigma^2(X^{*\prime}X^*)^{-1} \\
&= \sigma^2\{(P^{-1}X)'(P^{-1}X)\}^{-1} \\
&= \sigma^2\{X'(P')^{-1}P^{-1}X\}^{-1} \\
&= \sigma^2(X'\Omega^{-1}X)^{-1}
\end{aligned}
$$

$$
\begin{aligned}
(3) \quad \hat{\sigma}^2 &= \frac{1}{n-k}(Y^{*\prime}Y^* - \hat{\beta}'X^{*\prime}Y^*) \\
&= \frac{1}{n-k}\{(P^{-1}Y)'(P^{-1}Y) - \hat{\beta}'(P^{-1}X)'(P^{-1}Y)\} \\
&= \frac{1}{n-k}(Y'\Omega^{-1}Y - \hat{\beta}'X'\Omega^{-1}Y)
\end{aligned}
$$

(4) ตาราง ANOVA ปรากฏดังนี้

SOV	df	SS	MS	F
β_1	1	$n\bar{Y}^{*2}$		
$\beta_2, \beta_3, \ldots, \beta_k$	k-1	$\hat{\beta}'X'\Omega^{-1}Y - n\bar{Y}^{*2}$	$\frac{1}{k-1}(\hat{\beta}'X'\Omega^{-1}Y - n\bar{Y}^{*2})$	F_c
Residual	n-k	$Y'\Omega^{-1}Y - \hat{\beta}'X'\Omega^{-1}Y$	$\frac{1}{n-k}(Y'\Omega^{-1}Y - \hat{\beta}'X'\Omega^{-1}Y)$	
Total (Corrected)	n-1	$Y'\Omega^{-1}Y - n\bar{Y}^{*2}$	$\frac{1}{n-1}(Y'\Omega^{-1}Y - nY$ [illegible] 3	
Total (Uncorrected)	n	$Y'\Omega^{-1}Y$		

หมายเหตุ

ให้ $1' = [1, 1, \ldots, 1]$ คือ 1 – vector ขนาด $n \times 1$ ซึ่งจะพบว่า

$$11' = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} [1, 1, \ldots, 1] = \begin{bmatrix} 1 & 1 & \cdots \\ 1 & 1 & \cdots \\ \vdots & & \\ 1 & & \end{bmatrix} = \underset{(n \times n)}{J}$$

ดังนั้น $n\bar{Y}^2$ และ $n\bar{Y}^{*2}$ จึงเสนอได้ในรูปเมตริกซ์ดังนี้

$$n\bar{Y}^2 = \frac{1}{n}\left(\sum_i^n Y_i\right)^2$$

$$= \frac{1}{n}\left(\sum_i^n Y_i^2 + \sum_{i \neq j}^n Y_i Y_j\right)$$

$$= \frac{1}{n} Y'JY$$

ดังนั้น $SSR = \hat{\beta}'X'Y - \frac{1}{n} Y'JY$

$$SST = Y'Y - \frac{1}{n} Y'JY$$

และ $n\bar{Y}^{*2} = \frac{1}{n}\left(\sum_i^n Y_i^*\right)^2$

$$= \frac{1}{n}\left(\sum_i^n Y_i^{*2} + \sum_{i \neq j}^n Y_i^* Y_j^*\right)$$

$$= \frac{1}{n} Y^{*\prime}JY^*$$

$$= \frac{1}{n} (P^{-1}Y)'11'(P^{-1}Y)$$

$$= \frac{1}{n} Y'(P')^{-1}11'P^{-1}Y$$

ดังนั้น $SSR = \hat{\beta}'X'\Omega^{-1}Y - \frac{1}{n} Y'(P')^{-1}JP^{-1}Y$

$$SST = Y'\Omega^{-1}Y - \frac{1}{n} Y'(P')^{-1}JP^{-1}Y$$

จะเห็นว่าถึงอย่างไรเราก็ต้องคำนวณหาเมตริกซ์ P

วิธีที่ 2 อาศัย Maximum Likelihood Estimation

เนื่องจาก $U \sim N(0, \sigma^2\Omega)$ ดังนั้น Joint pdf ของ $u_1, u_2, \ldots, u_n$ คือ

$$L = \frac{1}{\sigma^n(2\pi)^{n/2}} \frac{1}{|\Omega|^{1/2}} \exp\{-\frac{1}{2\sigma^2} U'\Omega^{-1}U\} \quad ; -\infty < u_i < \infty; i = 1, 2, \ldots, n$$

$$\text{ดังนั้น } \ln L = -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{1}{2\sigma^2} U'\Omega^{-1}U$$

$$= -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{1}{2\sigma^2}\{Y'\Omega^{-1}Y - 2\beta'X'\Omega^{-1}Y + \beta'X'\Omega^{-1}X\beta\}$$

MLE ของ β และ σ^2 สามารถหาได้จากสมการ $\frac{\partial}{\partial\beta}\ln L = 0$ และ $\frac{\partial}{\partial\sigma}\ln L = 0$ ดังนี้

(1) $\frac{\partial}{\partial\beta}\ln L = 0 \Rightarrow -2X'\Omega^{-1}Y + 2X'\Omega^{-1}X\beta = 0$

ดังนั้น $\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$

(2) $\frac{\partial}{\partial\sigma}\ln L = 0$

$$\Rightarrow -\frac{n}{\sigma} + \frac{1}{2\sigma^3}(Y'\Omega^{-1}Y - 2\beta'X'\Omega^{-1}Y + \beta'X'\Omega^{-1}X\beta) = 0$$

ดังนั้น $\sigma^2 = \frac{1}{n}(Y - X\beta)'\Omega^{-1}(Y - X\beta)$

แทนค่า β ด้วย $\hat{\beta}$ $\Rightarrow$ $\hat{\sigma}^2 = \frac{1}{n}(Y - X\hat{\beta})'\Omega^{-1}(Y - X\hat{\beta}) = \frac{1}{n}e'\Omega^{-1}e$

พิจารณา $\hat{\sigma}^2$ จะพบว่า $Y - X\hat{\beta} = U - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U$ ทั้งนี้เพราะ $\hat{\beta} = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U$

และ $Y = X\beta + U$

ดังนั้น $\hat{\sigma}^2 = \frac{1}{n}\{[U - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U]'\Omega^{-1}[U - X(X\Omega^{-1}X)^{-1}X'\Omega^{-1}U]\}$

$$= \frac{1}{n}\{U'[I_n - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}]'\Omega^{-1}[I_n - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}]U\}$$

ให้ $M = I_n - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$ จะพบว่า $M^2 = MM = M$ แสดงว่า M เป็น Idempotent Matrix

แต่ $M' \neq M$

ดังนั้น $\hat{\sigma}^2 = \frac{1}{n}U'(M'\Omega^{-1}M)U$

พิจารณาแมตริกซ์ M ซึ่งเป็น Idempotent Matrix จะพบว่า

$r(M) = tr(M) = tr\{I_n - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\} = n - tr\{(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X)\} = n-k'$

พิจารณา $M'\Omega^{-1}M$ เนื่องจาก Q เป็น Positive Definite Matrix[2] ดังนั้น $r(M'\Omega^{-1}M) = r(M) = n-k$[3]

แสดงว่า $B = M'\Omega^{-1}M$ เป็น Positive Definite Matrix ที่ $r(B) = r(M) = n - k$

ดังนั้น $E(\hat{\sigma}^2) = \frac{1}{n}E(U'BU) = \frac{\sigma^2}{n}\,tr(B) = \frac{n-k}{n}\sigma^2$

นั่นคือ $E\left(\frac{n}{n-k}\hat{\sigma}^2\right) = \sigma^2$ หรือ $\frac{1}{n-k}(Y - X\hat{\beta})'\Omega^{-1}(Y - X\hat{\beta})$ เป็นตัวประมาณค่าที่ปราศจากอคติของ σ^2

ตัวอย่างที่ 3.6 จงยกตัวอย่างสถานการณ์ที่จำเป็นต้องวิเคราะห์สมการถดถอยโดยอาศัย GLS

วิธีทำ

1. กรณีที่ค่าสังเกตคือค่าเฉลี่ย

สมมุติว่าแบบจำลองเดิมคือ $Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it}$

โดยที่ Y_{it} = ผลผลิตข้าวโพดในแปลงทดลองที่ i ของศูนย์ทดลองที่ t

$i = 1, 2, \ldots, n_t\,;\; t = 1, 2, \ldots, T$

---

[1] ถ้า A เป็น Idempotent Matrix ที่ $r(A) = r$ แล้ว $tr(A) = r$; $r(A)$ หมายถึง Rank ของ A, $tr(A)$ หมายถึง Trace ของ A

[2] ถ้า A เป็น Positive Definite Matrix แล้ว A' และ A^{-1} จะเป็น Positive Definite Matrix ด้วย

[3] ถ้า A ขนาด $n \times n$ เป็น Positive Definite Matrix และ B เป็นเมตริกซ์ขนาด $n \times k$ ที่ $r(B) = k < n$ แล้ว B'AB จะเป็น Positive Definite Matrix

X_{2it} = จำนวนแรงงานที่ใช้ในแปลงทดลองที่ i ของศูนย์ทดลองที่ t

X_{3it} = จำนวนต้นทุนที่ใช้ในแปลงทดลองที่ i ของศูนย์ทดลองที่ t

ในกรณีงานทดลองครั้งนี้ เราจะมีศูนย์ทดลองหลายศูนย์ ศูนย์หนึ่ง ๆ จะทดลองเพาะข้าวโพด n_t แปลง วิธีหนึ่งที่ผู้วิจัยอาจใช้ก็คือ การถัวเฉลี่ยผลผลิต กำลังแรงงาน และต้นทุนที่ใช้ในศูนย์ทดลองที่ t ; t = 1, 2, ..., T เป็นค่าเฉลี่ยต่อ 1 แปลงทดลองเฉพาะศูนย์ แล้วนำข้อมูลดังกล่าวมาวิเคราะห์ กล่าวคือ

$$\bar{Y}_t = \frac{1}{n_t}\Sigma Y_{it}\ ,\ \bar{X}_{jt} = \frac{1}{n_t}\Sigma X_{jit}\ ;\ j = 2, 3$$

แบบจำลองใหม่จึงมีลักษณะดังนี้คือ

$$\bar{Y}_t = \beta_1 + \beta_2\bar{X}_{2t} + \beta_3\bar{X}_{3t} + \bar{u}_t \quad ;\ t = 1, 2, \ldots, T\ (T = \text{จำนวนศูนย์ทดลอง})$$

จากแบบจำลองใหม่นี้นักศึกษาจะพบว่า

$$E(\bar{u}_t) = \frac{1}{n_t}E(\Sigma u_{it}) = \frac{1}{n_t}\Sigma E(u_{it}) = 0$$

$$E(\bar{u}_t^2) = \frac{1}{n_t^2}E(\Sigma u_{it})^2 = \frac{1}{n_t^2}E\{\Sigma u_{it}^2 + \Sigma u_{it}u_{mt}\}$$

$$= \frac{n_t\sigma^2}{n_t^2} = \frac{\sigma^2}{n_t} \quad ;\ t = 1, 2, \ldots, T$$

$$\text{ดังนั้น} \quad E(UU') = \sigma^2\begin{bmatrix} 1/n_1 & & & & \\ & 1/n_2 & & & \\ & & 1/n_3 & & \\ & & & \ddots & \\ & & & & 1/n_T \end{bmatrix}$$

$$= \sigma^2\Omega \text{ ซึ่ง } \Omega^{-1} = \begin{bmatrix} n_1 & & & & \\ & n_2 & & & \\ & & n_3 & & \\ & & & \ddots & \\ & & & & n_T \end{bmatrix}$$

ในกรณีนี้ การประมาณค่าพารามิเตอร์ β ของสมการ $\bar{Y} = \bar{X}\beta + \bar{U}$ ต้องใช้ GLS กล่าวคือ

$$\hat{\beta} = (\bar{X}'\Omega^{-1}\bar{X})^{-1}\bar{X}'\Omega^{-1}\bar{X}\ ,\ \hat{V}(\hat{\beta}) = \hat{\sigma}^2(\bar{X}'\Omega^{-1}\bar{X})^{-1}\ ,\ \hat{\sigma}^2 = \frac{1}{n-3}e'\Omega^{-1}e$$

GLS ในกรณีเฉพาะเช่นนี้เรียกว่า Weighted Least Square (WLS)

สำหรับตัวอย่างนี้ นักศึกษาจะพบว่า Ω เป็นเมตริกซ์ที่เราทราบค่าสมาชิกได้ทุกตัว ซึ่งมีผลให้ $\hat{\beta}$ เป็นตัวสถิติ และพบว่า $P^{-1} = \text{diag}(\sqrt{n}, \sqrt{n_2}, \ldots, \sqrt{n_T})$ กล่าวคือ

$$P^{-1} = \begin{bmatrix} \sqrt{n_1} & & & & \\ & \sqrt{n_2} & & & \\ & & \sqrt{n_3} & & \\ & & & \ddots & \\ & & & & \sqrt{n_T} \end{bmatrix} \quad \text{ที่ } (P')^{-1}P^{-1} = \Omega^{-1}$$

ซึ่งเราสามารถหา $SS(\beta_1)$ ได้โดยง่าย

2. กรณีของ Autocorrelation : AR(1)

จากแบบจำลอง $Y = X\beta + U$ โดยที่ $E(u_i u_j) \neq 0$ และเราตั้งเป็นข้อตกลงว่า $u_t = \rho u_{t-1} + v_t$ โดยถือว่า $E(v_t) = 0$, $E(v_t v_s) = 0$; $t \neq s$ และ $E(v_t^2) = \sigma_V^2$ เมื่อ ρ คือ Serial Correlation Coefficient ที่ $|\rho| < 1$ [1]

ในกรณีเช่นนี้ เมื่อเราอาศัยการแทนค่า Lagged u_t ซ้ำ ๆ กัน ตามหลักของ Markov Process เราจะสามารถพิสูจน์ได้ว่า

$$\rho_s = \rho^s \text{ และ } \sigma_u^2 = \frac{\sigma_V^2}{1-\rho^2} \quad ; s = 1, 2, \ldots, n$$

$$\text{ดังนั้น} \quad E(UU') = \sigma_u^2\Omega = \sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 & . & \rho^{n-1} \\ p & 1 & \rho & . \; . \; . & \rho^{n-2} \\ \rho^2 & p & 1 & . \; . \; . & \rho^{n-3} \\ & & & & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & . \; . \; . & 1 \end{bmatrix}_{n \times n}$$

---

[1] ข้อตกลงของ Stationary คือถือว่า Probability ของ u มิได้อาศัยหรือพึ่งพิงจังหวะเวลาแห่งการบันทึกค่าสังเกต ในที่นี้ก็คือ $|\rho| <1$

นอกจากนี้ยังพบว่า $\Omega^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho & 0 \ldots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \cdots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 \cdots & 0 & -\rho & 1 \end{bmatrix}_{n\times n}$

$$\text{และ} \quad P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 \ldots & 0 & 0 & 0 \\ -\rho & 1 & 0 \ldots & 0 & 0 & 0 \\ 0 & -\rho & 1 \ldots & 0 & 0 & 0 \\ & & & \cdot & \vdots & \vdots \\ 0 & 0 & 0 \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 \ldots & 0 & -\rho & 1 \end{bmatrix}_{n\times n} \quad \text{ที่ } PP' = Q$$

ด้วยเหตุนี้เราจึงต้องประมาณค่า β ด้วยวิธี GLS กล่าวคือ

$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$, $\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'\Omega^{-1}X)^{-1}$; $\hat{\sigma}^2 = \frac{1}{n-k}e'\Omega^{-1}e$ และ $SS(\beta_1) = n\bar{Y}^{*2}$ โดยที่ $Y^* = P^{-1}Y$

แต่ตามตัวอย่างนี้ นักศึกษาจะพบว่า Ω เป็น Unknown Matrix สิ่งที่จะต้องกระทำก่อนการวิเคราะห์เพื่อให้ $\hat{\beta}$, $\hat{V}(\hat{\beta})$ และ $\hat{\sigma}^2$ เป็นตัวสถิติก็คือ การประมาณค่าเมตริกซ์ Ω ซึ่งสามารถกระทำได้หลายวิธี เมื่อแทนที่ Ω ด้วย $\hat{\Omega}$ เราย่อมประมาณค่า β, $V(\hat{\beta})$ และ σ^2 ได้ เทคนิคดังกล่าวนี้เรียกว่า EGLS

3. กรณีที่ $V(u_i)$ เป็นฟังก์ชันของตัวแปรอิสระ X_j [1]

จากแบบจำลอง $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$; $i = 1, 2, \ldots, n$

สมมุติว่า $V(u_i) = \sigma^2(a + bX_{ji}^h)^2$; สำหรับบางค่าของ h เช่น $h = -\frac{1}{2}, -1, \frac{1}{2}, 1$ ซึ่งในกรณีเช่นนี้จะพบว่า

$$E(UU') = \sigma^2 \begin{bmatrix} (a+bX_{j1}^h)^2 & 0 & 0 & \ldots & 0 \\ 0 & (a+bX_{j2}^h)^2 & 0 & \ldots & 0 \\ 0 & 0 & (a+bX_{j3}^h)^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & (a+bX_{jn}^h)^2 \end{bmatrix}_{n \times n} = \sigma^2 \Omega$$

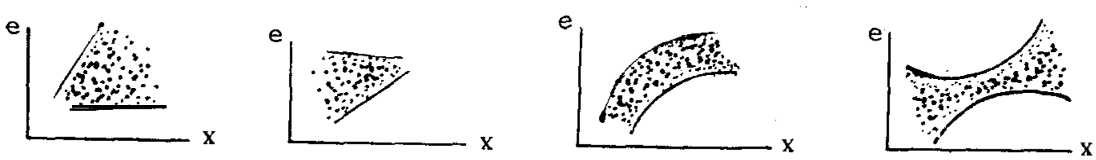
ดังนั้น

$$\Omega^{-1} = \begin{bmatrix} 1/(a+bX_{j1}^h)^2 & 0 & 0 & \ldots & 0 \\ 0 & 1/(a+bX_{j2}^h)^2 & 0 & \ldots & 0 \\ 0 & 0 & 1/(a+bX_{j3}^h)^2 & \cdot & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1/(a+bX_{jn}^h)^2 \end{bmatrix}_{n \times n}$$

และพบว่า

---

[1] นำคู่ลำดับ (X_{ji}, e_i) มาพล็อตลงในระนาบ 2 มิติ จะพบว่า Scatter Diagram ที่แสดงกลุ่มของคู่ลำดับอาจปรากฏลักษณะดังนี้



การทดสอบที่ง่ายที่สุดคือ การทดลองนำค่าของตัวแปรอิสระทุกตัวมาทดลองพล็อตร่วมกับค่าของ e ตัวแปรอิสระใดเป็นตัวการทำให้เกิดภาพดังกล่าว ก็แสดงว่า u เป็นฟังก์ชันของตัวแปรอิสระนั้น วิธีแก้ปัญหาก็คือ ให้ถ่วงน้ำหนักสมการถดถอยทั้งสมการด้วยตัวแปรอิสระ (หรือฟังก์ชันของตัวแปรอิสระ) ดังกล่าว เราเรียกตัวแปรอิสระนี้ว่า Deflator

$$
P^{-1} = \begin{bmatrix} 1/(a+bX_{j1}^h) & 0 & 0 & \cdots & 0 \\ 0 & 1/(a+bX_{j2}^h) & 0 & \cdots & 0 \\ 0 & 0 & 1/(a+bX_{j3}^h) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1/(a+bX_{jn}^h) \end{bmatrix}_{n \times n}
$$

โดยที่ $(P')^{-1}P^{-1} = \Omega^{-1}$

ค่าประมาณของ β, $V(\beta)$ และ σ^2 ตลอดจนตาราง ANOVA สามารถหาได้โดยนัยของข้อ 1 และ 2 ข้างต้น ส่วนค่าของ a และ b ในฟังก์ชั่น $u_i = a+bX_{ji}^h$; $i = 1, 2, \ldots, n$; $h = -\frac{1}{2}, -1, \frac{1}{2}, 1$ สามารถหาได้จากการวิเคราะห์สมการถดถอย $|e_i| = a+bX_{ji}^h$ ซึ่งมีผลให้สามารถประมาณค่า Ω ได้ ด้วย $\hat{\Omega}$ วิธีนี้เรียกว่า EGLS ตัวอย่างนี้เป็นวิธีหนึ่งที่ใช้แก้ปัญหา Heteroscedasticity of Variance[1] ซึ่งเสนอโดย เกลจเซอร์ (Glejser, H.1969) นอกจากนี้การตรวจสอบนัยสำคัญ $H_0 : a = 0$ และ $H_0 : b = 0$ ยังใช้เป็นเกณฑ์ในการทดสอบว่า Heteroscedasticity เกิดขึ้นจริงหรือไม่อีกด้วย

3.4 ตัวแปรดัมมี่ (Dummy Variable)

3.4.1 หลักเกณฑ์ทั่วไป

ตัวแปรดัมมี่ คือตัวแปรอิสระที่เป็น Categorical Variable[2] ที่กลุ่มย่อย (Category) หนึ่งของตัวแปรดังกล่าวจะได้รับระหัส (Coding) หนึ่ง ขณะที่กลุ่มย่อยอื่นจะได้รับระหัสอื่นที่แตกต่างกันไป ซึ่งโดยทั่วไปเรานิยมใช้ระหัส 1 แก่กลุ่มย่อยหนึ่ง และให้ระหัส 0 แก่กลุ่มย่อยอื่น ๆ และโดยทั่วไปเราจะไม่มีข้อจำกัดเกี่ยวกับจำนวนกลุ่มย่อย (Category) ของตัวแปรอิสระ ทั้งนี้เพราะไม่ว่านักวิจัยจะจำแนกตัวแปรอิสระออกเป็นกลุ่มย่อยมากกลุ่มเพียงใด นักวิจัยก็สามารถ

---

[1] การแก้ปัญหา Heteroscedasticity นอกจากจะแก้โดยอาศัย WLS หรือ GLS แล้ว เรายังสามารถใช้เทคนิคการแปลงรูปตัวแปรเพื่อให้ $V(u_i)$ คงที่ (Variance Stabilized Transformation) แก้ปัญหาได้ นักศึกษาจะได้พบเรื่องราวเหล่านี้ในบทต่อไป

[2] ตัวแปรที่มีลักษณะตรงข้ามกับ Categorical Variable เรียกว่า Contineous Variable

สร้างตัวแปรดัมมี่เข้ารับค่าตัวแปรอิสระได้มากเพียงนั้น เช่น นักวิจัยจำแนกตัวแปร ระดับการศึกษาของหัวหน้าครอบครัวออกเป็น 4 กลุ่ม คือ (1) มีการศึกษาระดับประถมศึกษา (2) มีการศึกษาระดับมัธยมศึกษา (3) มีการศึกษาสูงกว่าระดับมัธยมศึกษาแต่ไม่ถึงระดับปริญญาตรี (4) มีการศึกษาตั้งแต่ระดับปริญญาตรีขึ้นไป นักวิจัยก็สามารถจัดตัวแปรดัมมี่ซึ่งมีค่า 0, 1[1] เข้ารับค่าของตัวแปรได้ 4 ตัว หรือ (4 – 1) ตัว แล้วแต่กรณีได้ดังนี้

กรณีที่ 1 ถ้าแบบจำลองมีเทอมคงที่ β_1 (Intercept Term) ปรากฏอยู่จะจัดตัวแปรดัมมี่เข้ารับค่าของตัวแปรได้ (4 – 1) = 3 ตัว ดังนี้

D_1 = 1 ถ้าหัวหน้าครอบครัวมีการศึกษาระดับประถมศึกษา
= 0 ถ้ามีการศึกษาระดับอื่น

D_2 = 1 ถ้าหัวหน้าครอบครัวมีการศึกษาระดับมัธยมศึกษา
= 0 ถ้ามีการศึกษาระดับอื่น

D_3 = 1 ถ้าหัวหน้าครอบครัวมีการศึกษาสูงกว่าระดับมัธยมศึกษาแต่ไม่ถึงระดับปริญญาตรี
= 0 ถ้ามีการศึกษาระดับอื่น

ในกรณีนี้ถ้า $D_1 = D_2 = D_3 = 0$ แสดงว่าหัวหน้าครอบครัวมีการศึกษาระดับปริญญาตรีขึ้นไป

เหตุที่ต้องจัดจำนวนตัวแปรดัมมี่เข้ารับค่าของตัวแปรน้อยกว่าจำนวนกลุ่มย่อยของตัวแปรอยู่ 1 ตัวเสมอก็เพื่อมิให้เกิดปัญหา Multicolinearity ซึ่งถือว่าเป็นกับดัก (Dummy Variable Trap) สำหรับนักวิจัยที่ไม่ระมัดระวัง

---

[1] อาจกำหนดค่าเป็นอย่างอื่นเช่น –1, 0, 1 หรือ 0, 1, 2 ได้แต่ไม่เป็นที่นิยมเพราะมีข้อโต้แย้งเรื่องลำดับคะแนนระหัสเหล่านี้ถูกนำไปใช้ในเรื่องของ Effect Coding และ Orthogonal Coding การให้ระหัสสามารถกระทำได้กับทั้งตัวแปรอิสระและตัวแปรตาม แต่การใช้ตัวแปรดัมมี่เป็นตัวแปรตามอาจมีปัญหาทางทฤษฎีอยู่บ้าง เช่น การอภิปรายผลและปัญหาเกี่ยวกับข้อตกลงของ u การแก้ปัญหาอาจกระทำโดยอาศัย Logit Model อ่าน Maddala, G.S., Op. Cit., p.162-170 สำหรับการกำหนดระหัสให้แก่ตัวแปรดัมมี่เป็นอย่างอื่นที่มิได้มีจุดมุ่งหมายในแง่ของ Orthogonal Coding เช่น กำหนดให้มีค่าเป็น 0, 1, 2 ซึ่งเรียกว่า Polytomous จะมีปัญหาเรื่องลำดับคะแนนที่จะให้แก่ Category เช่น อาจต้องตอบคำถามว่าเพราะเหตุใด Category นั้นจึงให้คะแนน (ระหัส) สูงกว่า ซึ่งนักวิจัยยากที่จะอธิบายได้ นอกจากนี้การออกระหัสในลักษณะ Polytomous มักก่อให้เกิดปัญหา Difterential Effect อีกด้วย อ่าน Weisberg, S. Op. Cit., p.162 และ Kerlinger, F.N. and Pedhazur E.J., Op. Cit., p.116-151

กรณีที่ 2 ถ้าแบบจำลองไม่มีเทอมคงที่ (β_1) ปรากฏอยู่ให้จัดตัวแปรดัมมี่รับค่าของตัวแปรอิสระเท่ากับจำนวนกลุ่มย่อยของตัวแปร

ตัวอย่างเช่น สมมุติว่าเราต้องการวิเคราะห์ความสัมพันธ์ระหว่างรายได้กับระดับการศึกษาของหัวหน้าครอบครัว และอายุราชการจากสมการ

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

โดยที่ Y = รายได้ของหัวหน้าครอบครัว

X_2 = ระดับการศึกษาของหัวหน้าครอบครัว

X_3 = อายุราชการ

โดยการศึกษาครั้งนี้ใช้ข้อมูลสำรวจเป็นหลัก และจำแนกตัวแปรการศึกษาของหัวหน้าครอบครัวเป็น 4 ระดับคือ มีการศึกษาระดับประถมศึกษา มีการศึกษาระดับมัธยมศึกษา มีการศึกษาสูงกว่าระดับมัธยมศึกษาแต่ไม่ถึงระดับปริญญาตรี และมีการศึกษาระดับปริญญาตรีขึ้นไป

สมมุติว่านักวิจัยทำการสำรวจจากครัวเรือนตัวอย่าง (Sampled Houshold) 10 ครัวเรือน ผลปรากฏดังนี้

Y	5,000	6,000	10,000	9,000	4,000	3,000	15,000
X_2	ประถม	มัธยม	ปริญญา	มัธยม	มัธยม-ปริญญา	ประถม	ปริญญา
X_3	10	7	9	8	2	5	12

Y_1	4,500	8,000	7,000
Y_2	มัธยม	ประถม	มัธยม-ปริญญา
X_3	8	30	9

จะเห็นว่าเราสามารถจัดตัวแปรดัมมี่เข้ารับค่าของตัวแปร X_2 ได้ดังนี้

D_1	I	0	0	0	0	1	0	0	1	0
D_2	0		1		0		1		000100	
D_3	0	0	0	0	1	0	0	0	0	1
D_4	0	0	1	0	0	0	1	0	0	0

แบบจำลองเดิมคือ $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$ จึงเปลี่ยนรูปเป็น

$$Y = \beta_1 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta_3 X_3 + u$$

เมตริกซ์ X จะมีค่าปรากฏดังนี้

$$X = \begin{vmatrix} 1 & 1 & 0 & 0 & 10 \\ 1 & 0 & 1 & 0 & 7 \\ 1 & 0 & 0 & 0 & 9 \\ 1 & 0 & 1 & 0 & 8 \\ 1 & 0 & 0 & 1 & 2 \\ 1 & 1 & 0 & 0 & 5 \\ 1 & 0 & 0 & 0 & 12 \\ 1 & 0 & 1 & 0 & 8 \\ 1 & 1 & 0 & 0 & 30 \\ 1 & 0 & 0 & 1 & 9 \end{vmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_3 \end{bmatrix}$$

ขณะเดียวกัน ถ้าเราเพิ่มตัวแปรดัมมี่ D_4 เข้าไปรับค่าของตัวแปร X_2 อีกตัวหนึ่งโดยที่

$D_4 = 1$ ถ้ามีการศึกษาระดับปริญญาตรีขึ้นไป

$= 0$ มีการศึกษาระดับอื่น ๆ

แล้ว แบบจำลองเดิมคือ $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$ จะเปลี่ยนรูปเป็น

$$Y = \beta_1 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \beta_3 X_3 + u$$

ซึ่งมีผลให้เมตริกซ์ X เปลี่ยนลักษณะไปสู่รูปใหม่ดังนี้

$$x = \begin{vmatrix} 1 & 1 & 0 & 0 & 0 & 10 \\ 1 & 0 & 1 & 0 & 0 & 7 \\ 1 & 0 & 0 & 0 & 1 & 9 \\ 1 & 0 & 1 & 0 & 0 & 8 \\ 1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 1 & 0 & 0 & 0 & 5 \\ 1 & 0 & 0 & 0 & 1 & 12 \\ 1 & 0 & 1 & 0 & 0 & 8 \\ 1 & 1 & 0 & 0 & 0 & 30 \\ 1 & 0 & 0 & 1 & 0 & 9 \end{vmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \beta_3 \end{bmatrix}$$

ในกรณีนี้จะพบว่าสดมภ์ต่าง ๆ ของเมตริกซ์ X เกิดจากการประกอบกันเชิงเส้นของกันและกัน กล่าวคือ

สดมภ์ที่ 1 = สดมภ์ที่ 2 + สดมภ์ที่ 3 + สดมภ์ที่ 4 + สดมภ์ที่ 5

สดมภ์ที่ 5 = สดมภ์ที่ 1 – สดมภ์ที่ 2 – สดมภ์ที่ 3 – สดมภ์ที่ 4

เป็นต้น

ซึ่งแสดงว่าเวคเตอร์ในเมตริกซ์ X มีลักษณะของ Linearly Dependent หรือมี Colinearity ผลสะท้อนที่ติดตามมาก็คือเมตริกซ์ $(X'X)^{-1}$ ไม่มีค่าปรากฏ กรณีของการสร้างตัวแปรดัมมี่ในลักษณะเช่นนี้เป็นเครื่องชี้ให้เห็นว่านักวิจัยได้ติดกับดัก (Dummy Variable Trap) เข้าแล้ว

ในขณะเดียวกัน หากนักวิจัยกำหนดแบบจำลองในรูปของ

$$Y = \beta_2 X_2 + \beta_3 X_3 + u$$

ซึ่งไม่มีเทอมคงที่ β_1 ปรากฏอยู่ สถานการณ์เช่นนี้นักวิจัยสามารถสร้างตัวแปรดัมมี่ได้มากเท่ากับจำนวนกลุ่ม (จำนวน Category) ของตัวแปรอิสระ โดยไม่มีปัญหาเรื่องกับดักแต่ประการใด จากข้อมูลในตัวอย่างที่ผ่านมาจะพบว่า เราสามารถเปลี่ยนรูปสมการ $Y = \beta_2 X_2 + \beta_3 X_3 + u$ เป็น $Y = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \beta_3 X_3 + u$ และพบว่า เมตริกซ์ X ต่อไปนี้ไม่มีเวคเตอร์ใดเกิดขึ้น จากการประกอบกันของเวคเตอร์อื่น ๆ

$$x = \begin{vmatrix} 1 & 0 & 0 & 0 & 10 \\ 0 & 1 & 0 & 0 & 7 \\ 0 & 0 & 0 & 1 & 9 \\ 0 & 1 & 0 & 0 & 8 \\ 0 & 0 & 1 & 0 & 2 \\ 1 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 1 & 12 \\ 0 & 1 & 0 & 0 & 8 \\ 1 & 0 & 1 & 0 & 30 \\ 0 & 0 & 0 & 0 & 9 \end{vmatrix}, \quad \beta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \beta_3 \end{bmatrix}$$

ด้วยเหตุผลดังกล่าวเราจึงสามารถสรุปหลักเกณฑ์การสร้างตัวแปรดัมมี่ได้ดังนี้

1. ถ้าแบบจำลองมีเทอมคงที่ β_1 ปรากฏอยู่ให้สร้างตัวแปรดัมมี่เข้ารับค่าตัวแปรอิสระที่สนใจได้น้อยกว่าจำนวนกลุ่มย่อย (Catergory) ของตัวแปรนั้น 1 ตัวเสมอ

เหตุที่เป็นเช่นนี้เพราะในกรณีดังกล่าวถ้าสร้างตัวแปรดัมมี่เท่ากับจำนวนกลุ่มย่อยของตัวแปร จะมีผลให้เกิดปัญหา Colinearity ตามตัวอย่างที่ผ่านมา ถ้ามองเฉพาะสดมภ์ที่ 5 (คือ D_4) จะพบว่า D_4 มิได้เพิ่มข้อสนเทศแก่งานแต่ประการใดเพราะ D_1, D_2 และ D_3 เมื่อ $D_1 = D_2 = D_3 = 0$ ย่อมหมายถึง D_4 อยู่แล้ว

2. ถ้าแบบจำลองไม่มีเทอมคงที่ β_1 ปรากฏอยู่ ให้สร้างตัวแปรดัมมี่เข้ารับค่าของตัวแปรอิสระที่สนใจเท่ากับจำนวนกลุ่มย่อยของตัวแปรนั้น

โดยปกติเรานิยมนำตัวแปรดัมมี่มาใช้ในงานวิเคราะห์ความถดถอยเมื่อตัวแปรอิสระมีธรรมชาติลักษณะใดลักษณะหนึ่งดังต่อไปนี้

1. ตัวแปรอิสระที่ถูกจำแนกออกเป็นกลุ่มย่อยตามกาลเวลา (Temporal Effect) เช่น ดัชนีราคาผู้บริโภค แยกเป็น 2 กลุ่ม ตามปีฐานที่ใช้ในการคำนวณ คือ ปีฐานเดิมกับปีฐานใหม่ รายได้ของผู้บริโภคจำแนกเป็น 2 กลุ่ม **คือ** รายได้ระหว่างปีสงครามกับรายได้ในระหว่างปีที่มีสันติภาพ ดัชนีราคาสินค้าที่สนใจจำแนกเป็น 4 กลุ่มตามไตรมาส (Seasonal Effect) ระยะเวลา (จำนวนปี) ที่คู่สมรสอยู่ร่วมกันก่อนการหย่าร้างจำแนกเป็น k กลุ่ม เป็นต้น

2. ตัวแปรอิสระที่ถูกจำแนกออกเป็นกลุ่มย่อยตามท้องที่สำรวจ หรือ ท้องที่อ้างอิง (Spatial Effect) เช่น ปริมาณการบริโภคสินค้าและบริการชนิดหนึ่งจำแนกเป็น 5 กลุ่มตามภูมิภาค คือ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง ภาคตะวันออก และภาคใต้ รสนิยมของผู้บริโภคจำแนกเป็น 2 กลุ่มย่อยตามประเภทของสินค้าและบริการที่บริโภคอยู่เป็นปกติ ความต้องการซื้อสินค้าและบริการชนิดที่สนใจจำแนกเป็น 2 กลุ่มตามระดับความเจริญของท้องถิ่น คือ ในเขตเทศบาล และนอกเขตเทศบาล เป็นต้น

3. ตัวแปรอิสระเป็นตัวแปรเชิงคุณภาพซึ่งเป็น Categorical Variable อยู่แล้วตามธรรมชาติ เช่น เพศ สถานภาพสมรส ถิ่นที่อยู่ปกติ เชื้อชาติ สัญชาติ ฐานะทางสังคม (Class Effect) ชั้นทางเศรษฐกิจ กลุ่มอายุ ระดับการศึกษา เป็นต้น

4. ตัวแปรเชิงปริมาณที่ได้รับการจัดหมวดหมู่เป็นกลุ่มย่อย (Broad Grouping of Quantitative Variable) เช่น การจัดกลุ่มรายได้ การจัดกลุ่มความชื้นในอากาศ การจัดกลุ่มระดับ IQ เป็นต้น

5. ตัวแปรอิสระที่ถูกจำแนกตามสิ่งทดลอง (Treatment Effect) เช่น การจำแนกสัตว์ทดลองเป็น 2 กลุ่มคือ กลุ่มควบคุม (ไม่รับ Treatment = 0) กับกลุ่มทดลอง (รับ Treatment = 1) เป็นต้น

โดยปกติสมการถดถอยที่มีตัวแปรดัมมี่ปรากฏร่วมอยู่จะมิใช่สมการ Hyperplane เพียงสมการเดียวดังที่เคยพบเห็นในตอนที่ผ่านมา แต่จะเป็นกลุ่มของสมการถดถอยที่วางขนานกันหลายสมการ กล่าวคือ ถ้ามีตัวแปรดัมมี่ m ตัว แต่ละตัวได้รับระหัส 0 และ 1 ก็จะปรากฏสมการที่ขนานกันแต่มีจุดตัดบนแกน Y ต่าง ๆ กันถึง 2^m สมการ

3.4.2 การตรวจสอบนัยสำคัญและการอภิปรายผล

การตรวจสอบนัยสำคัญ และอภิปรายผลเกี่ยวกับพารามิเตอร์ในแบบจำลองที่มีตัวแปรดัมมี่ปะปนอยู่นั้น โดยหลักการแล้วถือว่าเป็นเรื่องเดียวกันกับกรณีปกติที่ไม่มีตัวแปรดัมมี่ แต่ในรายละเอียดแล้วยังมีสิ่งที่ต้องระมัดระวังเพราะโดยปกติแบบจำลองที่มีตัวแปรดัมมี่ปะปนอยู่นั้นมิใช่เป็นแบบจำลองเดียว แต่กลับเป็นแบบจำลองหลาย ๆ แบบที่ผสมผสานกันอยู่และสามารถแยกตัวออกมาเป็นกรณีเฉพาะได้เมื่อแทนค่าของตัวแปรดัมมี่ด้วยระหัส 0 และ 1 ซึ่งสามารถจำแนกแบบจำลองเดิมออกได้ถึง 2^m ชุดที่มีความหมายแตกต่างกัน โดยปกติแล้วเราไม่จำเป็นต้องระมัดระวังในกรณีที่จะทำการตรวจสอบ และอภิปรายผลเกี่ยวกับสัมประสิทธิ์ความถดถอยเพราะแบบจำลองทั้ง 2^m ชุดจะมีสัมประสิทธิ์ความถดถอย (หรือความชัน) เดียวกัน แต่จะต้องระมัดระวังเมื่อต้องทำการตรวจสอบนัยสำคัญและอภิปรายผลเกี่ยวกับ Intercept Term ของแบบจำลองแต่ละชุด เพราะ Intercept Term มักเกิดจากการผสมผสานระหว่างพารามิเตอร์หลายตัว

ตัวอย่างเช่น ในการพยากรณ์ปริมาณการบริโภคสินค้าและบริการชนิดหนึ่ง โดยนักวิจัยคาดว่าปริมาณการบริโภคจะผันแปรไปตามกาลเวลา ระดับชั้นทางสังคมของผู้บริโภค และปัจจัยอื่น ๆ เช่น รายได้ รสนิยม สภาพคล่อง และอื่น ๆ ดังนี้

$$C = f \text{ (ตัวแปรแสดงกาลเวลา, ระดับชั้นทางสังคม, ปัจจัยอื่น ๆ)}$$

กำหนดให้ตัวแปรกาลเวลาจำแนกเป็น 4 ระดับคือ ไตรมาสที่ 1, 2, 3, 4 และกำหนดให้ตัวแปรระดับชั้นทางสังคมจำแนกเป็น 3 ระดับคือ สังคมชั้นสูง สังคมชั้นกลาง และสังคมชั้นต่ำ (ใช้ดัชนีการจำแนกที่เหมาะสม)

แบบจำลองปรากฏดังนี้

$$C = \beta_1 + \alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3 + \gamma_1 S_1 + \gamma_2 S_2 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

โดยที่ $Q_1 = 1$ ถ้าค่าสังเกตที่ได้รับการบันทึก ณ ไตรมาสที่ 1

$= 0$ ณ ไตรมาสอื่น

$Q_2 = 1$ ถ้าค่าสังเกตที่ได้รับการบันทึก ณ ไตรมาสที่ 2

$\quad = 0$ ณ ไตรมาสอื่น

$Q_3 = 1$ ถ้าค่าสังเกตที่ได้รับการบันทึก ณ ไตรมาสที่ 3

$\quad = 0$ ณ ไตรมาสอื่น

$S_1 = 1$ ถ้าค่าสังเกตที่ได้รับการบันทึกจากหน่วยสำรวจในสังคมชั้นสูง

$\quad = 0$ ในสังคมชั้นอื่น

$S_2 = 1$ ถ้าค่าสังเกตที่ได้รับการบันทึกจากหน่วยสำรวจในสังคมชั้นกลาง

$\quad = 0$ ในสังคมชั้นอื่น

หมายเหตุ ถ้า $Q_1 = Q_2 = Q_3 = 0$ แสดงว่าค่าสังเกตได้รับการบันทึก ณ ไตรมาสที่ 4 (Q_4) และถ้า $S_1 = S_2 = 0$ แสดงว่าค่าสังเกตได้รับการบันทึกจากหน่วยสำรวจในสังคมชั้นต่ำ (S_3)

สมมุติว่าผลจากการวิเคราะห์ข้อมูลปรากฏดังนี้

$$C = \hat{\beta}_1 + \hat{\alpha}_1 Q_1 + \hat{\alpha}_2 Q_2 + \hat{\alpha}_3 Q_3 + \hat{\gamma}_1 S_1 + \hat{\gamma}_2 S_2 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k$$

$$\hat{\beta} = (\hat{\beta}_1, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\beta}_2, \ldots, \hat{\beta}_k)' = (X'X)^{-1}X'Y$$

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} = V$$

$$\hat{\sigma}^2 = \frac{1}{n-(k+5)}(Y'Y - \hat{\beta}'X'Y)$$

ทั้งนี้เมตริกซ์ X มีขนาด $n \times (k+5)$ และ $X'X$ มีขนาด $(k+5) \times (k+5)$

จะเห็นว่าแบบจำลองนี้สามารถจำแนกเป็นแบบจำลองเฉพาะสถานการณ์ได้ถึง 2^5 ชุด (32 ชุด) ซึ่งเป็น Hyperplane ที่วางขนานกันใน Space ระบบ k มิติ

การทดสอบสมมุติฐานและอภิปรายผลสำหรับแบบจำลองแต่ละชุด นักวิจัยจะต้องระมัดระวังเนื่องจากค่าจุดตัดบนแกน Y (ในที่นี้คือแกน C) เกิดขึ้นจากการผสมผสานระหว่าง β_1 กับสัมประสิทธิ์ของตัวแปรดัมมี่ ตัวอย่างเช่น แบบจำลองสำหรับไตรมาสที่ 1 $(Q_1 = 1)$ สำหรับสังคมชั้นสูง $(S_1 = 1)$ คือ

$$C = (\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\gamma}_1) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k$$

กรณีนี้จะเห็นว่า $H_0 : (\beta_1 + \alpha_1 + \gamma_1) = 0$ ก็คือสมมุติฐานที่มุ่งตรวจสอบดูว่าปริมาณการบริโภคของคนในสังคมชั้นสูงสำหรับไตรมาสที่ 1 มีอัตราถัวเฉลี่ย (Mean Response) เท่ากับ 0 หรือไม่ และพบว่าตัวสถิติ t สำหรับกรณีเช่นนี้คือ

$$t_c = \frac{(\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\gamma}_1)}{\sqrt{\hat{V}(\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\gamma}_1)}}$$

โดยที่ $\hat{V}(\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\gamma}_1) = \hat{V}(\hat{\beta}_1) + \hat{V}(\hat{\alpha}_1) + \hat{V}(\hat{\gamma}_1) + 2Cov(\hat{\beta}_1, \hat{\alpha}_1) + 2Cov(\hat{\beta}_1, \hat{\gamma}_1) + 2Cov(\hat{\alpha}_1, \hat{\gamma}_1)$ ซึ่งค่า Variance และ Covariance เหล่านี้ปรากฏอยู่ในเมตริกซ์ $V = \hat{\sigma}^2(X'X)^{-1}$

หรือในกรณีแบบจำลองสำหรับไตรมาสที่ 3 ($Q_3 = 1$) สำหรับสังคมชั้นสูง ($S_1 = 1$) คือ

$$C = (\hat{\beta}_1 + \hat{\alpha}_3 + \hat{\gamma}_1) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k$$

จะเห็นว่า $H_0 : (\beta_1 + \alpha_3 + \gamma_1) = 0$ คือสมมุติฐานที่มุ่งทดสอบดูว่าปริมาณการบริโภคของคนในสังคมชั้นสูง ณ ไตรมาสที่ 1 มีอัตราถัวเฉลี่ยเท่ากับ 0 หรือไม่ และพบว่า

$$\hat{V}(\hat{\beta}_1 + \hat{\alpha}_3 + \hat{\gamma}_1) = \hat{V}(\hat{\beta}_1) + \hat{V}(\hat{\alpha}_3) + \hat{V}(\hat{\gamma}_1) + 2Cov(\hat{\beta}_1, \hat{\alpha}_3) + 2Cov(\hat{\beta}_1, \hat{\gamma}_1) + 2Cov(\hat{\alpha}_3, \hat{\gamma}_1)$$

เป็นต้น

สิ่งที่น่าสังเกตก็คือ สัมประสิทธิ์ความถดถอยของตัวแปรดัมมี่ในสมการเดิม ในที่นี้คือ

$$C = \hat{\beta}_1 + \hat{\alpha}_1 Q_1 + \hat{\alpha}_2 Q_2 + \hat{\alpha}_3 Q_3 + \hat{\gamma}_1 S_1 + \hat{\gamma}_2 S_2 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k$$

นั้นมิได้มีความหมายตรงตัวเหมือน $\beta_2, \beta_3, \beta_4, \ldots, \beta_k$ ซึ่งเป็นความชันของ Hyperplane และหมายถึง Relative Change $(\frac{\Delta Y}{\Delta X})$ แต่กลับหมายถึง Differential Effect หรือ Relative Difference เช่น $\alpha_1, \alpha_2, \alpha_3$ หมายถึง Seasonal Effect และ γ_1, γ_2 หมายถึง Class Effect เป็นต้น[1] การทดสอบนัยสำคัญของ Differential Effect สามารถกระทำได้ 2 นัย คือ

1. ถ้าต้องการทดสอบนัยสำคัญของ Differential Effect ตัวใดตัวหนึ่งเดี่ยว ๆ เพียงตัวเดียวให้ใช้วิธีทดสอบตามปกติที่เคยกระทำมาในตอนก่อน เช่น

$$H_0 : \alpha_1 = 0 \quad VS \quad H_1 : \alpha_1 \neq 0$$

ให้ทดสอบโดยใช้ t-test หรือ F-test ก็ได้ ถ้าใช้ t-test จะพบว่า $t_c = \frac{\hat{\alpha}_1}{\sqrt{\hat{V}(\hat{\alpha}_1)}}$ ที่ df เท่ากับ $n-(k+5)$ หรือคำนวณหาค่า t จากผลต่างระหว่าง Intercept Term ของแบบจำลองที่มี α_1

---

[1] ถ้าพิจารณาในแง่ของงานทดลอง $\alpha_1, \alpha_2, \alpha_3, \gamma_1, \gamma_2$ หมายถึง Treatment Effect

ปรากฏอยู่ และแบบจำลองที่ไม่มี α_1 ปรากฏอยู่ (แต่ต้องมีพารามิเตอร์อื่นร่วมกัน) เช่น คำนวณโดยอาศัยแบบจำลอง

$$(\hat{\beta}_1 + \hat{\alpha}_1) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นต่ำ : ไตรมาสที่ 1)}$$

และ
$$\hat{\beta}_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นต่ำ : ไตรมาสที่ 4)}$$

จะพบว่า $H_0 : \alpha_1 = 0$ ก็คือ $H_0 : (\beta_1 + \alpha_1) - \beta_1 = 0$ ซึ่งในกรณีนี้การคำนวณหาค่า t_c แม้จะใช้วิธีเดียวกัน กล่าวคือ

$$t_c = \frac{(\hat{\beta}_1 + \hat{\alpha}_1) - \hat{\beta}_1}{\sqrt{\hat{V}\{(\hat{\beta}_1 + \hat{\alpha}_1) - \hat{\beta}_1\}}} = \frac{\hat{\alpha}_1}{\sqrt{\hat{V}(\hat{\alpha}_1)}} \qquad df = n - (k+5)$$

แต่ให้ความหมายของ α_1 ชัดเจนกว่า ในที่นี้หมายถึงปริมาณความแตกต่างในการบริโภคของคนในสังคมชั้นต่ำเปรียบเทียบระหว่างไตรมาสที่ 1 และไตรมาสที่ 4 (Seasonal Effect)

2. ต้องการเปรียบเทียบระหว่าง Different Effect ให้ทดสอบโดยคำนวณหาค่า t หรือ F จากผลต่างระหว่าง Intercept Term ต่าง ๆ เช่น ต้องการทดสอบ $H_0 : \gamma_1 = \gamma_2$ (ความแตกต่างในปริมาณการบริโภคระหว่างกลุ่มคนในสังคมชั้นสูงและชั้นกลาง : Class Difference) เราสามารถคำนวณหาค่า t_c ได้จากผลต่างระหว่าง Intercept Term ของแบบจำลองคู่ใดคู่หนึ่งต่อไปนี้

$$C = (\hat{\beta}_1 + \hat{\gamma}_1) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นสูง : ไตรมาสที่ 4)}$$
$$C = (\hat{\beta}_1 + \hat{\gamma}_2) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นกลาง : ไตรมาสที่ 4)} \quad \ldots\ldots \text{ ก.}$$

$$C = (\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\gamma}_1) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นสูง : ไตรมาสที่ 1)}$$
$$C = (\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\gamma}_2) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นกลาง : ไตรมาสที่ 1)} \quad \ldots\ldots \text{ ข}$$

$$C = (\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\gamma}_1) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นสูง : ไตรมาสที่ 1, 2)} \quad \ldots\ldots \text{ ค}$$
$$C = (\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\gamma}_2) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นกลาง : ไตรมาสที่ 1, 2)}$$

$$C = (\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 + \hat{\gamma}_1) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นสูง : ไตรมาสที่ 1, 2, 3)}$$
$$C = (\hat{\beta}_1 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 + \hat{\gamma}_2) + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k \quad \text{(สังคมชั้นกลาง : ไตรมาสที่ 1, 2, 3)}$$

ฯลฯ ..... ง

ผลต่างระหว่าง Intercept Term จากแบบจำลองทุกคู่ล้วนส่งผลให้เกิดผลต่าง $(\hat{\gamma}_1 - \hat{\gamma}_2)$ ซึ่งใช้เป็นสถิติสำหรับทดสอบสมมุติฐาน $H_0 : \gamma_1 - \gamma_2 = 0$ เช่นเดียวกัน และพบว่า

$$t_c = \frac{\hat{\gamma}_1 - \hat{\gamma}_2}{\sqrt{\hat{V}(\hat{\gamma}_1 - \hat{\gamma}_2)}}$$

โดยที่ $\hat{V}(\hat{y}_1 - \hat{y}_2) = \hat{V}(\hat{y}_1) + \hat{V}(\hat{y}_2) + 2Cov(\hat{y}_1, \hat{y}_2)$, $df = n-(k+3)$

3.4.3 สรุป

ตัวแปรดัมมี่ถูกนำเข้ามาใช้ในงานวิเคราะห์ความถดถอยก็ด้วยวัตถุประสงค์สำคัญคือ เพื่อใช้เป็นตัวทดแทน (Proxy) หรือตัวแปรทดแทน (Proxy Variable)[1] ของตัวแปรที่ไม่อาจวัดค่าเป็นตัวเลขเชิงประมาณในลักษณะของ Contineous Variable ได้ทั้งนี้เพื่อให้สามารถควบคุมความผันแปรของตัวแปรตาม Y ให้ได้มากที่สุดเท่าที่จะมากได้ กล่าวโดยสรุปแล้วตัวแปรดัมมี่ถูกนำมาใช้ประโยชน์ในกรณีต่าง ๆ ดังนี้

1. ตัวแปรดัมมี่ใช้เป็นตัวแปรทดแทนของตัวแปรเชิงคุณภาพ (Qualitative Variable หรือ Categorical Variable) เช่น อาชีพ ศาสนา ภาษา เพศ ถิ่นที่อยู่ ฯลฯ

2. ตัวแปรดัมมี่ใช้เป็นตัวแปรทดแทนของตัวแปรเชิงปริมาณที่เราไม่อาจวัดค่าได้ หรือไม่สะดวกที่จะใช้ค่าของตัวแปรดังกล่าวได้โดยตรง วิธีปฏิบัติทั่วไปสำหรับกรณีนี้คือให้จำแนกค่าของตัวแปรเชิงปริมาณเป็นกลุ่มการกระทำเช่นนี้มีผลให้ Contineous Variable เปลี่ยนสภาพเป็น Categorical Variable ซึ่งนักวิจัยสามารถจัดตัวแปรดัมมี่เข้ารับค่าของตัวแปรดังกล่าวได้ไม่เกินจำนวนกลุ่ม เช่น จำแนกอายุประชากรซึ่งเป็นตัวแปรต่อเนื่อง (Contineous Variable) เป็นกลุ่มอายุ กลุ่มละ 5 - 10 ปี จำแนกความพอใจซึ่งเป็นตัวแปรต่อเนื่องออกเป็น 5 กลุ่ม เป็นต้น

3. ตัวแปรดัมมี่ใช้ช่วยแสดงการเลื่อนที่ตั้ง (Shift) ของสมการถดถอยให้สอดคล้องกับสถานการณ์ขณะหนึ่ง ๆ

4. ตัวแปรดัมมี่ใช้ช่วยแสดงการเปลี่ยนความชันของ Hyperplane ที่วางขนานกันใน Space เป็นจำนวนทั้งสิ้น 2^m ชุด โดยมีจุดตัดบนแกนตั้ง (แกน Y) แตกต่างกัน แต่ในบางสถานการณ์เราไม่อาจเชื่อถือได้ว่าค่าความชันจะคงที่ตลอดไป เพราะอาจมีตัวแปรอิสระที่มีลักษณะของปฏิสัมพันธ์ (Interaction ระหว่างตัวแปร) แทรกปนอยู่ ด้วยเหตุนี้ ในทางปฏิบัติหลายสถานการณ์

---

[1] **ตัวแปรทดแทน** (Proxy Variable) คือตัวแปรที่เราสามารถนำค่าไปใช้ทดแทนตัวแปรจริง (true Variable) ได้ เช่น การศึกษาเป็นพฤติกรรมของมนุษย์ที่มีลักษณะของความต่อเนื่อง หากจะวัดกันจริง ๆ ต้องใช้เครื่องมือที่มีความไวและถี่ถ้วนมาก ทางออกสำหรับปัญหานี้ก็คือการนำ "จำนวนปีที่เคยได้รับการศึกษา ในสถาบันการศึกษา" เป็นตัวแปรทดแทน ดังนี้เป็นต้น สิ่งที่พึงระลึกไว้สำหรับการใช้ Proxy Variable ก็คือ Proxy Variable สามารถใช้ทดแทนตัวแปรจริงได้เท่านั้น มิใช่แทนกันได้อย่างแท้จริง ด้วยเหตุนี้ Proxy Variable จึงให้ Measurement Error ที่เรียกว่า Error in Variable เสมอ

เราจึงจำเป็นต้องเพิ่มตัวแปรลักษณะดังกล่าวเข้าไว้ในสมการ ตัวแปรนี้อาจเป็นปฏิสัมพันธ์ระหว่างตัวแปรดัมมี่ด้วยกันหรือระหว่างตัวแปรดัมมี่กับตัวแปรอิสระ หรือแม้แต่ในระหว่างตัวแปรอิสระด้วยกัน ทั้งนี้ขึ้นอยู่กับความรู้และวิจารณญาณของนักวิจัยเป็นสำคัญ[1]

5. ตัวแปรดัมมี่ใช้เป็นตัวแปรทดแทนของตัวแปรตาม

6. ตัวแปรดัมมี่ใช้สำหรับปรับสมการให้สอดคล้องกับจังหวะเวลา (Seasonal Adjustment) ของตัวแปรอนุกรมเวลา

8.5 การใช้ตัวแปรเวลา (t) เป็นตัวแปรอิสระ

ในสถานการณ์ทางปฏิบัติหลายสถานการณ์เรานิยมนำตัวแปรเวลา (t) ร่วมเป็นตัวแปรอิสระเพื่อให้สามารถมองเห็นภาพของค่า Y ว่ามีความเคลื่อนไหวไปตามเวลาที่เปลี่ยนแปลงไปหรือไม่มากน้อยเพียงใด การนำตัวแปรเวลาเข้าเป็นตัวแปรอิสระดังกล่าวสามารถกระทำได้หลายวิธีดังนี้

1. นำตัวแปร t เป็นตัวแปรอิสระโดยตรง

การนำตัวแปร t เป็นตัวแปรอิสระโดยตรงนั้นนักวิจัยสามารถกำหนดค่าของ t เป็น 1,2,...,n โดยที่ $t = 1$ ถ้าค่าสังเกต Z เป็นค่าสังเกตสำหรับวาระที่ 1 $t = 2$ ถ้าค่าสังเกต Z เป็นค่าสังเกตในวาระที่ 2,..., $t = n$ ถ้าค่าสังเกต Z เป็นค่าสังเกตสำหรับวาระที่ n สัมประสิทธิ์

---

[1]ในการพัฒนา Best Fitted Model เรามีวิธีนำตัวแปรอิสระลักษณะต่าง ๆ ซึ่งแสดงปฏิสัมพันธ์ระหว่างตัวแปร เช่น $Z = X_iX_j$, $W = \frac{X_i}{X_j}$ ซึ่งสร้างขึ้นใหม่เข้าสู่สมการ โดยการทดสอบ Lack of Fit ในแต่ละขั้นตอน อ่าน Drapper, N. and Smith, H, Op. cit. P. 26 - 29, 217 - 227 (CH. 7) หลักการโดยสรุปของวิธีดังกล่าวปรากฏดังนี้

(1) นำตัวแปรอิสระที่มีสหสัมพันธ์เชิงเส้นกับตัวแปรตาม Y สูงที่สุดเข้าสู่สมการเป็นตัวแรก

(2) ทดสอบ Lack of Fit ถ้ามี Lack of Fit (l. o. f) แสดงว่าเราจำเป็นต้องปรับปรุงสมการต่อไป SS(l.o.f) = SS(Residual) - SS(Pure Error)

มี df = df (Residual) - df (Pure error) โดยที่

SS(Pure Error) = $\Sigma\Sigma(Y_{iu} - \bar{Y}_i)$, df = $\Sigma n_i - c$ โดยที่ Y_{iu} คือค่าสังเกตของ Y หน่วยที่ u ที่สอดคล้องกับค่าของ X ที่วัดค่าซ้ำ ๆ กัน n_i ครั้ง

(3) นำ Residual ของสมการมา Plot ร่วมกับตัวแปรอิสระอื่น ๆ หรือการผสมผสานกันของตัวแปรอิสระแล้วตรวจสอบดูจากกราฟว่า Residual มีลักษณะที่สัมพันธ์กับตัวแปรดังกล่าวหรือไม่ ถ้าสัมพันธ์กันให้นำตัวแปรนั้นเข้าสู่สมการ

(4) ดำเนินการข้อ (1) - (3) ซ้ำ ๆ จนพบว่าไม่มีตัวแปรอิสระใดพึงเข้าสู่สมการอีก

ของ t ในสมการ $Y = f(X's, t)$ จะแสดงอัตราความเติบโตของ Y(Autonomeous Growth) ในขณะเดียวกันถ้าผู้วิจัยต้องการแสดงให้เห็นถึงอัตราการเติบโตของ X ผู้วิจัยสามารถกำหนดตัวแปรอิสระเป็น tX

2. นำตัวแปรเวลา t เข้าสู่ฟังค์ชัน $Y = f(X's)$ ในลักษณะของ First Difference ระหว่างตัวแปร

จากสมการ $$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \gamma t + u_t \qquad \ldots\ldots\ldots\ldots(1)$$

เมื่อ lag ไป 1 ช่วงเวลาสมการ (1) จะเปลี่ยนรูปเป็น

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \ldots + \beta_k X_{k,t-1} + \gamma(t-1) + u_{t-1} \qquad \ldots\ldots\ldots(2)$$

(1) - (2) ได้

$$(Y_t - Y_{t-1}) = \beta_2(X_{2t} - X_{2,t-1}) + \ldots + \beta_k(X_{kt} - X_{k,t-1}) + \gamma(t-t+1) + (u_t - u_{t-1})$$

$$= \gamma + \beta_2(X_{2t} - X_{2,t-1}) + \ldots + \beta_k(X_{kt} - X_{k,t-1}) + (u_t - u_{t-1}) \qquad \ldots\ldots\ldots(3)$$

สมการที่ (3) คือสมการอีกรูปหนึ่งซึ่งแสดงถึงการนำตัวแปรเวลาร่วมเป็นตัวแปร แม้มิใช่การร่วมเป็นตัวแปรอิสระโดยตรงแต่ก็ยังคงแสดงให้เห็นถึงอัตราการเติบโตของ Y (Autonomeous Growth) ได้เช่นเดียวกัน[1]

3. นำตัวแปรเวลาเข้าร่วมเป็นตัวแปรในลักษณะแฝงเร้น (Implicit Variable) ในรูปของ Lagged Variable

Lagged Variable คือตัวแปรอิสระหรือตัวแปรตามที่ผูกพันตัวเองกับเงื่อนไขเวลาในอดีต หรือกล่าวอีกนัยหนึ่ง Lagged Variable คือค่าของตัวแปรอิสระและตัวแปรตามในอดีต เรานำ Lagged Variable มาใช้เพื่อให้แบบจำลองมีลักษณะยืดหยุ่นหรือเคลื่อนไหวไปได้ตามกาลเวลา นักศึกษาจะได้ศึกษาเรื่องราวของ Distributed Lag โดยละเอียดในบทต่อไป

4. นำตัวแปรเวลาเข้าร่วมเป็นตัวแปรในลักษณะของตัวแปรดัมมี่

---

[1]ความจริงแล้วสัมประสิทธิ์ของตัวแปรเวลาไม่อาจตีความว่าเป็น Growth Factor ของ Y ได้เสมอไป ความหมายตามปกติของสัมประสิทธิ์ดังกล่าวก็คืออิทธิพลร่วมของปัจจัยต่าง ๆ ที่มิได้นำเข้าสู่สมการ

3.6 ค่าสังเกตที่เป็นชุดหรือจัดเป็นหมวดหมู่ (Grouping Observations)

ในบางสถานการณ์นักวิจัยอาจจำเป็นต้องนำค่าเฉลี่ยของกลุ่มค่าสังเกตของตัวแปรมาใช้ในการวิเคราะห์สมการถดถอยแทนค่าเดี่ยว ๆ ของตัวแปรทั้งนี้อาจด้วยเหตุผลใดเหตุผลหนึ่งต่อไปนี้คือ

1. แหล่งข้อมูลเดิมบันทึกข้อมูลเก็บไว้ในรูปค่าเฉลี่ย[1] เช่นปริมาณน้ำฝนโดยถัวเฉลี่ย ปริมาณผลผลิตถัวเฉลี่ยต่อไร่ รายได้ถัวเฉลี่ย ฯลฯ

2. ผู้วิจัยเองต้องการจัดหมวดหมู่ข้อมูลออกเป็นกลุ่ม ๆ ตามดัชนีจำแนกต่าง ๆ เช่น จำแนกรายได้-รายจ่ายผู้บริโภคตามระดับการศึกษา จำแนกตามความต้องการซื้อสินค้าที่สนใจตามระดับรายได้ของผู้บริโภค ฯลฯ การกระทำดังกล่าวมีผลให้ผู้วิจัยต้องใช้ข้อมูลที่เป็นตัวแทนกลุ่ม (โดยปกติเราใช้ค่าเฉลี่ยเป็นตัวแทนกลุ่ม) เป็นข้อมูลในการวิเคราะห์สมการถดถอย

เทคนิคการวิเคราะห์สมการถดถอยโดยอาศัยข้อมูลทั้งสองลักษณะนี้คือ GLS ที่กล่าวถึงแล้วในตอน 3.3 สำหรับการพัฒนาวิธีการวิเคราะห์สมการถดถอยในลำดับต่อไปนี้ผู้เขียนจะกล่าวถึงเฉพาะข้อมูลในลักษณะที่ 2 เท่านั้น สำหรับข้อมูลลักษณะที่ 1 ได้กล่าวถึงแล้วในตัวอย่าง 3.4

ให้ $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \, ; \, i = 1,2,\ldots n$

จำแนกค่าสังเกต $Z_i = (Y_i, X_{2i}, X_{3i}, \ldots, X_{ki}) \; ; \; i = 1,2\ldots,n$ ออกเป็น m กลุ่ม ๆ ละ n_ℓ ค่า ; $\ell = 1,2,\ldots m$ โดยที่ $k < m < n$

ดังนั้นค่าเฉลี่ยของค่าสังเกต Z ของกลุ่มที่ ℓ จึงมีลักษณะดังนี้

$$\bar{Z}_\ell = (\bar{Y}_\ell, \bar{X}_{2\ell}, \bar{X}_{3\ell}, \ldots \bar{X}_{k\ell}) \, ; \, \ell = 1,2,\ldots m$$

โดยที่

$$\bar{Y}_\ell = \frac{1}{n_\ell} \Sigma Y_{\ell i} = \left(\frac{1}{n_\ell}, \frac{1}{n_\ell}, \ldots \frac{1}{n_\ell}\right) \begin{bmatrix} Y_{\ell 1} \\ Y_{\ell 2} \\ \vdots \\ Y_{\ell_n} \end{bmatrix} \; ; \; \ell = 1,2,\ldots,m$$

---

[1] ข้อมูลลักษณะนี้จะนำมาใช้ในการวิเคราะห์ได้อย่างมีประสิทธิภาพก็ต่อเมื่อผู้วิจัยทราบขนาดตัวอย่าง n ที่ผู้บันทึกข้อมูลใช้เป็นฐานในการคำนวณค่าเฉลี่ย เพราะขนาดตัวอย่างดังกล่าวมีความจำเป็นต้องนำมาใช้เป็นตัวถ่วงน้ำหนักในการวิเคราะห์แบบ GLS หรือ WLS การวิเคราะห์สมการถดถอยสำหรับข้อมูลลักษณะนี้ได้กล่าวถึงแล้วในตัวอย่าง 3.3

$= G_\ell Y$ เมื่อ G_ℓ คือแถวที่ ℓ ของเมตริกซ์ G

$$\bar{X}_{j\ell} = \frac{1}{n_\ell}\Sigma X_{j\ell i} = \left(\frac{1}{n\ell}, \frac{1}{n\ell}, \ldots, \frac{1}{n\ell}\right) \begin{bmatrix} X_{j\ell 1} \\ X_{j\ell 2} \\ \vdots \\ X_{j\ell n\ell} \end{bmatrix} \quad ; l = 1,2,\ldots,m : j = 2,3,\ldots,k$$

$$= G_\ell X_j$$

$$\bar{U}_\ell = \frac{1}{n\ell}\Sigma u_{\ell i} = \left(\frac{1}{n\ell}, \frac{1}{n\ell}, \ldots, \frac{1}{n\ell}\right) \begin{bmatrix} u\ell_1 \\ u\ell_2 \\ \vdots \\ u_{\ell n\ell} \end{bmatrix} \quad ; \ell = 1,2,\ldots m$$

$$= G_\ell U$$

ดังนั้น

$$\bar{Y} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_m \end{bmatrix} = \begin{bmatrix} 0 & 0 & \ldots & 0 & \frac{1}{n_1} & \frac{1}{n_1} & \ldots & \frac{1}{n_1} & \ldots & 0 \\ 0 & 0 & \ldots & 0 & \frac{1}{n_2} & \frac{1}{n_2} & \ldots & \frac{1}{n_2} & \ldots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 & \frac{1}{n_m} & \frac{1}{n_m} & \ldots & \frac{1}{n_m} & \ldots & 0 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = GY$$

โดยที่ G เป็นเมตริกซ์ขนาด $m \times n$

ในทำนองเดียวกันเราจะพบว่าเมตริกซ์ $\bar{X} = GX$ และเวคเตอร์ $\bar{U} = GU$

ดังนั้นสมการถดถอยที่ต้องการคือ

$$\bar{Y} = \bar{X}\beta + \bar{U} \text{ โดยที่ } \bar{Y} = GY, \bar{X} = GX \text{ และ } \bar{U} = GU$$

หมายเหตุ เมตริกซ์ G ในที่นี้เป็นรูปทั่วไป ในทางปฏิบัติตำแหน่งที่ตั้งของสมาชิกของ G อาจอยู่ ณ ที่ใด (ในแต่ละแถว) ก็ได้โดยสลับอยู่กับ 0 ทั้งนี้ขึ้นอยู่กับข้อกำหนดของผู้วิจัยเองว่าจะให้สมาชิกของกลุ่มค่าสังเกตประกอบไปด้วยสมาชิกในลำดับที่เท่าไรของเวคเตอร์ Z หรือขึ้นอยู่กับข้อมูลว่าสมาชิกตัวที่ (เลขที่) เท่าไรปรากฏอยู่ในกลุ่มค่าสังเกตที่จำแนกเอาไว้ ตัวอย่างเช่น จากการสำรวจผลสัมฤทธิ์ทางการศึกษาของนักศึกษา (Y) โดยกำหนดให้ X_2 = จำนวนหน่วยกิตที่ลงทะเบียนในภาคการศึกษาที่ 2 X_3 = ระยะเวลาที่นักศึกษาอุทิศตนเพื่อการศึกษา (วัดเป็นชั่วโมง) ทั้งนี้กำหนดเอา IQ เป็นตัวแปรสำหรับจำแนกกลุ่มตัวอย่าง สมมุติว่าผลการบันทึกข้อมูลปรากฏดังนี้

ลำดับที่ของ Z	Y ผลสัมฤทธิ์ทางการศึกษา GPA	X_2 จำนวนหน่วยกิจที่ลงทะเบียนเรียน	X_3 เวลาที่ใช้ศึกษาค้นคว้า	IQ
1	2.20	18	3	90
2	3.10	15	5	100
3	2.05	20	3	99
4	2.80	22	4	120
5	3.00	24	2	110
6	3.15	22	4	100
7	1.90	28	1	100
8	2.78	30	3	130
9	2.20	20	8	100
10	2.30	15	7	99
11	3.50	24	10	130
12	3.00	27	10	120
13	2.78	21	2	110
14	2.88	15	5	120
15	3.90	22	4	145
16	3.30	25	7	135

สมมุติว่านักวิจัยจัดจำแนกกลุ่มตัวอย่างนี้ออกเป็น 3 กลุ่มตามระดับ IQ มี $IQ \leq 100$, $100 < IQ \leq 120$, $120 < IQ \leq 140$, $IQ > 140$ ดังนี้แสดงว่าค่าสังเกต $Z_i = (Y_i, X_{2i}, X_{3i})$ ชุดที่ 1, 2, 3, 6, 7, 9, 10 อยู่ในกลุ่มที่ 1 ($IQ \leq 100$) ค่าสังเกตชุดที่ 4, 5, 12, 13, 14 อยู่ในกลุ่มที่ 2 ($100 < IQ \leq 120$) ค่าสังเกตชุดที่ 8,11, 16 อยู่ในกลุ่มที่ 3 ($120 < IQ \leq 140$) และค่าสังเกตชุดที่ 15 อยู่ในกลุ่มที่ 4 ($IQ > 140$) ดังนั้น

$$\bar{Z}_l = (\bar{Y}_l, \bar{X}_{2l}, \bar{X}_{3l}) \text{ โดยที่}$$

$$\bar{Y}_1 = \frac{Y_{11}+Y_{12}+Y_{13}+Y_{16}+Y_{17}+Y_{19}+Y_{1,10}}{7} = \frac{16.90}{7} = 2.414$$

$$\text{หรือ } \bar{Y}_1 = \left(\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, 0, 0, \frac{1}{7}, \frac{1}{7}, 0, \frac{1}{7}, \frac{1}{7}, 0, 0, 0, 0, 0\right) \begin{bmatrix} 2.20 \\ 3.10 \\ 2.05 \\ \vdots \\ 2.90 \\ 3.30 \end{bmatrix} = G_1Y$$

$$\bar{Y}_2 = \frac{Y_{24} + Y_{25} + Y_{2,12} + Y_{2,13} + Y_{2,14}}{5} = \frac{14.46}{5} = 2.892$$

$$\text{หรือ } \bar{Y}_2 = \left(0, 0, 0, \frac{1}{5}, \frac{1}{5}, 0, 0, 0, 0, 0, 0, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, 0, 0\right) \begin{bmatrix} 2.20 \\ 3.10 \\ 2.05 \\ \vdots \\ 2.90 \\ 3.30 \end{bmatrix} = G_2Y$$

$$\bar{Y}_3 = \frac{Y_{38} + Y_{3,11} + Y_{3,16}}{3} = \frac{9.58}{3} = 3.193$$

$$\text{หรือ } \bar{Y}_3 = \left(0, 0, 0, 0, 0, 0, 0, \frac{1}{4}, 0, 0, \frac{1}{4}, 0, 0, 0, 0, \frac{1}{4}\right) \begin{bmatrix} 2.20 \\ 3.10 \\ 2.05 \\ \vdots \\ 2.90 \\ 3.30 \end{bmatrix} = G_3Y$$

$$\bar{Y}_4 = Y_{15} = 3.9 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,)\ Y = G_4Y$$

เราสามารถหา $\bar{X}_{21}$, $\bar{X}_{22}$, $\bar{X}_{23}$, $\bar{X}_{31}$, $\bar{X}_{32}$ และ $\bar{X}_{33}$ ได้ในทำนองเดียวกัน

ดังนั้น

$$G = \begin{bmatrix} G_1 \\ G_2 \\ G_3 \\ G_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & 0 & \frac{1}{7} & \frac{1}{7} & 0 & \frac{1}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

พิจารณา $\bar{Y} = \bar{X}\beta + \bar{U}$ โดยที่

$$\bar{Y} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_m \end{bmatrix}, X = \begin{bmatrix} 1 & \bar{X}_{11} & \bar{X}_{21} & \dots & \bar{X}_{k1} \\ 1 & \bar{X}_{12} & \bar{X}_{22} & \dots & \bar{X}_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \bar{X}_{1m} & \bar{X}_{2m} & \dots & \bar{X}_{km} \end{bmatrix}_{m \times k}, \bar{U} = \begin{bmatrix} \bar{U}_1 \\ \bar{U}_2 \\ \vdots \\ \bar{U}_m \end{bmatrix}$$ จะพบว่า

$$\begin{aligned} V(\bar{U}) &= E(\bar{U}\bar{U}') = E\{(GU)(GU)'\} = E(GUU'G') \\ &= GE(UU')G' = G\sigma_u^2 I_n G' = \sigma_u^2(GG') \neq \sigma_u^2 I_n \end{aligned}$$

แสดงว่าสมการถดถอยที่ใช้ค่าเฉลี่ยของกลุ่มเป็นค่าสังเกตในการวิเคราะห์ข้อมูลมีผลให้ข้อตกลงว่า $V(U) = \sigma_u^2 I_n$ ไม่เป็นจริง (Heteroscedasticity) ซึ่ง OLS ไม่อาจให้คำตอบที่ถูกต้องได้ วิธีที่สามารถใช้ได้ก็คือ GLS

ดังนั้น เมื่อประมาณค่าโดยวิธี GLS จะพบว่า (ให้ $GG' = \Omega$)

$$\hat{\beta} = [\bar{X}'(GG')^{-1}\bar{X}]^{-1}\bar{X}'(GG')^{-1}\bar{Y}$$

ตาราง ANOVA สำหรับกรณีนี้ปรากฏดังนี้

SOV	df	SS		MS	F
Mean	1	$m\bar{\bar{Y}}^2$			
Regression	k - 1	$\hat{\beta}'\bar{X}'(GG')^{-1}\bar{Y} - m\bar{\bar{Y}}^2$	$= Q_1$	$Q_1/k-1$	F_d
Residual	m - k	$\bar{Y}'(GG')^{-1}\bar{Y} - \hat{\beta}'\bar{X}'(GG')^{-1}\bar{Y}$	$= Q_2$	$Q_2/m-k$	R^2
Total (adj)	m - 1	$\bar{Y}'(GG')^{-1}\bar{Y} - m\bar{\bar{Y}}^2$	$= Q$	$Q/m-1$	
Total (unadj)	m	$\bar{Y}'(GG')^{-1}\bar{Y}$			

โดยที่ $\bar{\bar{Y}} = \frac{1}{m}(\sqrt{n_1}\,\bar{Y}_1 + \sqrt{n_2}\,\bar{Y}_2 + \ldots + \sqrt{n_m}\,\bar{Y}_m)$

ตามตัวอย่างที่ผ่านมานักศึกษาจะพบว่า

$$GG' = \begin{bmatrix} \frac{1}{7} & 0 & 0 & 0 \\ 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{และ } (GG')^{-1} = \begin{bmatrix} 7 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} & & & \\ & & & \\ & & & \\ n_1 & n_2 & n_3 & 0\ \ _4 \end{bmatrix}$$

$$\bar{Y} = GY = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \\ \bar{Y}_4 \end{bmatrix} = \begin{bmatrix} 2.414 \\ 2.892 \\ 3.193 \\ 3.900 \end{bmatrix}$$

$$\bar{X} = GX = \begin{bmatrix} 1 & \bar{X}_{21} & \bar{X}_{31} \\ 1 & \bar{X}_{22} & \bar{X}_{32} \\ 1 & \bar{X}_{23} & \bar{X}_{33} \\ 1 & \bar{X}_{24} & \bar{X}_{34} \end{bmatrix} = \begin{bmatrix} 1 & 19.714 & 4.428 \\ 1 & 21.800 & 4.600 \\ 1 & 26.333 & 6.666 \\ 1 & 22.000 & 4.000 \end{bmatrix}$$

$$\text{ดังนั้น } \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = [\bar{X}'(GG')^{-1}\bar{X}]^{-1}\bar{X}'(GG')^{-1}\bar{Y}$$

$$\hat{\sigma}^2 = \frac{1}{m-k}\{\bar{Y}'(GG')^{-1}\bar{Y} - \hat{\beta}'\bar{X}'(GG')^{-1}\bar{Y}\} \text{ โดยที่ } m = 4,\ k = 3$$

SOV แหล่งต่าง ๆ รวมทั้ง F - Ratio และ R^2 ย่อมเป็นสิ่งที่เราคำนวณหาได้โดยไม่ยากนัก เรื่องนี้ผู้เขียนขอเว้นการคำนวณไว้เป็นแบบฝึกหัด

ข้อสังเกตที่สำคัญ 2 ประการที่พึงคำนึงถึงเสมอในการใช้ตัวแทนกลุ่มค่าสังเกตเป็นข้อมูลในการวิเคราะห์สมการถดถอยก็คือ

1. เราจะต้องแบ่งกลุ่มค่าสังเกต (m) ให้มีจำนวนกลุ่มสูงกว่าจำนวนพารามิเตอร์ (k) เสมอ เพราะถ้า $m \leqslant k$ จะทำให้เราไม่อาจคำนวณหา σ^2 ได้ทั้งนั้น เพราะ

$$\hat{\sigma}^2 = \frac{\bar{Y}'(GG')^{-1}\bar{Y} - \hat{\beta}'\bar{X}'(GG')^{-1}\bar{Y}}{m - k}$$

ถ้า $m = k$ ตัวหารจะเป็น 0 มีผลให้ $\sigma^2 \rightarrow \infty$ และเมื่อใช้คอมพิวเตอร์ช่วยในการวิเคราะห์ เครื่องจะแสดง Error ให้ปรากฏ แสดงให้เห็นว่าได้เกิดความขัดแย้งกับนิยามของการหารเกิดขึ้น ขณะเดียวกันถ้า $m < k$ ตัวหารจะติดลบซึ่งเป็นสิ่งที่เป็นไปไม่ได้ที่ $\hat{\sigma}^2$ จะมีค่าติดลบ

2. ถ้าขนาดของกลุ่มเท่ากันตลอดคือ $n_1 = n_2 = \ldots = n_m$ จะไม่เกิดปัญหา Heteroscedasticity แต่ประการใด กรณีที่ขนาดตัวอย่างย่อยไม่เท่ากัน เราจะพบว่า $V(\bar{u}) = \frac{\sigma_u^2}{n_i}$ แสดงว่า $V(\bar{u})$ ไม่คงที่ แต่จะผันแปรค่าไปตามขนาดตัวอย่าง ซึ่งขัดแย้งกับข้อตกลงเดิมที่กำหนดไว้ว่า $V(u)$ ต้องคงที่เสมอ แต่ในกรณีขนาดตัวอย่างย่อยเท่ากันตลอดคือ $n_1 = n_2 = \ldots n_m = c$ จะพบว่า $V(\bar{u})$ คงที่เท่ากันเสมอคือเท่ากับ $\frac{\sigma_u^2}{c}$ ซึ่งไม่ขัดแย้งกับข้อตกลงเดิม ด้วยเหตุนี้ ถ้า $n_1 = n_2 = \ldots = n_m = c$ เราสามารถใช้ WLS ตามธรรมดาหรือปรารถนาจะใช้ GLS ตามวิธีที่ผ่านมาก็สามารถกระทำได้ (อ่าน Johnston, J., Op. Cit., p.228-238)

แบบฝึกหัดบทที่ 3

1. การอยู่ตัว (firmness) ของเนยแข็งขึ้นอยู่กับเวลาที่ใช้ในกระบวนการผลิตบางขั้นตอน เพื่อศึกษาเรื่องนี้เราได้ทำการศึกษาเกี่ยวกับเนยแข็ง 18 ชนิด โดยทำการบันทึกเวลาที่ใช้ในขั้นการผลิต พร้อมทั้งบันทึกค่าการอยู่ตัวเอาไว้ดังนี้

เวลา (au.)	½	1	1½	2	3	4
การอยู่ตัว	102	110	126	132	160	164
	105	120	128	143	149	166
	115	115	119	139	147	172

จงวิเคราะห์สมการถดถอย แล้วทดสอบว่าการกำหนดให้เป็น linear regression นั้นเหมาะสมเพียงใด (ทดสอบ linearity) (ให้นำ $\hat{y}$ มาพล็อตกับ e ถ้ามี pattern แสดงว่าไม่เป็น linear)

2. ระดับอัตราภาษีกับร้อยละของการไปเสียภาษีปรากฏดังนี้

ค่ากลางอัตราภาษี (X) (ปอนด์/ปี)	3	7	12	17	25	35	45	55	70	120
ร้อยละของการไปเสียภาษี	86	79	76	69	65	62	52	51	51	48

จงประมาณค่าสมการถดถอย

$$\frac{100}{100-Y} = a + \frac{b}{X}$$

3. สมการ $Y = a + bX + ct$ คือสมการแสดงความสัมพันธ์ระหว่าง Y กับ X และเวลา จงแสดงให้เห็นว่าค่าประมาณของ b ในสมการนี้ จะมีค่าเดียวกันกับ b ที่ประมาณได้จากสมการถดถอยอย่างง่าย (Simple regression) ภายหลังจากที่ได้จำกัด time trend ออกจากตัวแปร X แล้ว

4. x,, X_2 และ X_3 เป็นตัวแปรที่มีความเกี่ยวข้องกัน มีข้อมูลดังนี้ $s_1 = 1$, $s_2 = 1.3$, $s_3 = 1.9$, $r_{12} = 0.37$, $r_{13} = -0.641$,$r_{23} = -0.736$ จงคำนวณหา partial correlation $r_{13\cdot2}$

ถ้า $X_4 = X_1 + X_2$ จงคำนวณหา r_{42}, r_{43} และ $r_{43.2}$ อยากทราบว่าเพราะเหตุใด $r_{13.2}$ กับ $r_{13.2}$ จึงเท่ากัน

5. ในการศึกษาความแข็งของแท่งเหล็ก เราวัดความแข็งของเหล็กขนาดความยาว 18 นิ้ว โดยทดสอบความแข็งทุก ๆ ระยะ 1.8 นิ้ว ปรากฏข้อมูลดังนี้

ระยะทางนับจากปลาย (นิ้ว),d	0	1.8	3.6	5.4	7.2	9.0	10.8	12.6	14.4	16.2	18
ตัวเลขแสดงความแข็ง,h	250	276	298	335	374	414	454	503	558	604	671

สมการแสดงความสัมพันธ์ที่แสดงถึงความแข็งที่ค่อย ๆ เพิ่มขึ้นตามความยาว (ระยะจากปลาย) คือ

1) $h = A+Bd+Cd^2$

2) $h = \alpha e^{\beta d}$

อยากทราบว่าสมการใดเหมาะสมกว่า

6. ในสมการ $y_t = \beta x_{1t} + \gamma x_{2t} + u_t$ โดยที่ตัวแปรทุกตัวอยู่ในรูปเบี่ยงเบนจากค่าเฉลี่ย (คือ $y_t = Y_t - \bar{Y}$, $x_{1t} = X_{1t} - \bar{X}_1$, $x_{2t} = X_{2t} - \bar{X}_2$) จงประมาณค่า β โดยวิธีต่อไปนี้

1) คำนวณหา $\hat{\beta}$ และ $\hat{\gamma}$ จากสมการถดถอย y on x_1, x_2

2) วิเคราะห์สมการถดถอย y on x_2 แล้วหาส่วนเหลือ y^*_t

วิเคราะห์สมการถดถอย x_1 on x_2 แล้วคำนวณหาส่วนเหลือ x^*_{1t}

วิเคราะห์สมการถดถอย y^* on x^*_1 แล้วหาค่า $\hat{\beta}$

จงแสดงให้เห็นว่าค่า $\hat{\beta}$ ตามข้อ 1) และ 2) เท่ากัน และแสดงให้เห็นว่าส่วนเหลือ $y_t-\hat{\beta}x_{1t}-\hat{\gamma}x_{2t}$ กับ $y^*_t-\hat{\beta}x^*_{1t}$ มีค่าเท่ากัน

7. $r_{12.3}$ คือค่าสหสัมพันธ์บางส่วน (partial correlation) ระหว่าง Y กับ X_2 (ภายหลังจากที่ได้กำจัดอิทธิพลของ X_3 ออกจากทั้ง Y และ X_2) การหาค่า $r_{12.3}$ กระทำดังนี้

1) วิเคราะห์สมการถดถอย Y on X_3 คือ $y_i = b_{13}x_{3i}+u_i$ แล้วหาส่วนเหลือ $u_i = y_i-b_{13}x_{3i}$

2) วิเคราะห์สมการถดถอย X2 on X3 คือ $x_{2i} = b_{23}x_{3i} + v_i$ แล้วหาส่วนเหลือ

$v = x_{2i} - b_{23}x_{3i}$

3) หาค่าสหสัมพันธ์ระหว่าง u กับ v ได้

$$r_{12.3} = \frac{\Sigma u_i v_i}{\sqrt{\Sigma u_i^2 \Sigma v_i^2}}$$

จงพิสูจน์ว่า $r_{12.3}$ ในขั้นที่ 3 คือ

$$r_{12.3} = \frac{\Sigma u_i v_i}{\sqrt{\Sigma u_i^2 \Sigma v_i^2}} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

ทำนองเดียวกันจงแสดงให้เห็นว่า

$$r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}}$$

$$r_{23.1} = \frac{r_{23} - r_{12} r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}}$$

8. จากการติดตามศึกษาความต้องการซื้อสินค้าชนิดหนึ่ง ปรากฏข้อมูลดังนี้ (n = 21 ปี)

ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สปส. สหสัมพันธ์
$\bar{X}$ = 51.843	s_x = 9.205	r_{xy} = -0.9158
$\bar{Y}$ = 8.313	s_Y = 1.780	r_{yt} = -0.8696
$\bar{T}$ = 0	s = 6.057	r_{xt} = 0.9304

โดยที่ X = ปริมาณการบริโภคต่อบุคคล (เปานด์)
Y = ราคา (เซนต์ต่อเปานด์)
T = เวลา (ปี)

ก. ประมาณค่าสมการถดถอย Y = f (X,T)
ข. ตรวจสอบนัยสำคัญของ สปส. ความถดถอย
ค. อธิบายให้เข้าใจว่าการผนวกตัวแปร T เข้าไว้ในสมการแปลว่าอะไร

9. ข้อมูลปริมาณน้ำตาลที่ผลิตได้ อุณหภูมิเฉลี่ยของเดือนกลางปี (มิถุนายน) และปริมาณน้ำฝนเฉลี่ย ปรากฏดังนี้ จงคำนวณหาค่า สปส. สหสัมพันธ์บางส่วนคือ $r_{yx_1x_2}$ และ $r_{yx_2x_1}$ พร้อมทั้งแปลผล

ปีที่	ปริมาณผลผลิต น้ำตาล (1000 ตัน)	อุณหภูมิถั่วเฉลี่ย ในเดือน มิถุนายน (°F)	ปริมาณน้ำฝนถั่วเฉลี่ย (นิ้ว)
1	470	62	33
2	520	62	42
3	560	63	32
4	510	61	38
5	500	64	31
6	550	61	40
7	630	62	44
8	640	63	36
9	650	61	30
10	620	58	43

9. ถ้าตัวแปรทุกตัวเป็น standardized form และเราทราบค่า corretation matrix ได้ จงหาสูตรประมาณค่า สปส.สมการถดถอยและ สปส.สหสัมพันธ์บางส่วน

10. ในการศึกษาว่ากาลอากาศมีผลต่อการเป็นโรคของผึ้งหรือไม่ เรากำหนดให้

X_1 = อุณหภูมิเฉลี่ยในเดือนมกราคม (°F)
X_2 = อุณหภูมิเฉลี่ยในเดือนมิถุนายน (°F)
X_3 = วันที่ดอกไม้เริ่มบาน (นับจากวันที่ 1 มกราคม)
Y = ส่วนร้อยของรวงผึ้งที่ติดเชื้อ

จงวิเคราะห์สมการถดถอย $Y = f(X_1, X_2, X_3)$ จากข้อมูลต่อไปนี้

X_1	X_2	X_3	Y
35	53	200	49
35	53	212	40
38	50	211	41
40	64	212	46
40	70	203	52
42	68	194	59
44	59	194	53
46	73	188	61
50	59	196	55
50	71	190	64

11. ตารางต่อไปนี้คือปริมาณแร่เหล็กที่ขุดได้ในแต่ละปี (Y_t) และค่าพยากรณ์ของปริมาณดังกล่าวที่ใช้วิธีพยากรณ์ต่างกัน (อนุกรม F_{1t} และ F_{2t})

	ปีที่								
	1	2	3	4	5	6	7	8	9
Y	6.5	6.8	8.7	8.7	8.4	8.8	9.7	11.0	11.6
F_{1t}	5.0	6.3	8.0	8.6	8.7	8.9	9.4	9.9	11.0
F_{2t}	6.2	6.7	8.5	9.0	9.0	9.2	9.9	11.1	12.0

อยากทราบว่าวิธีพยากรณ์ใดดีกว่า? หากท่านประสงค์จะใช้ weighted average เพื่อสร้างค่าพยากรณ์ ท่านจะใช้สูตร weighted average ใด? มีผู้กล่าวว่าเราอยากทราบว่าวิธีพยากรณ์ใดดีกว่าให้เปรียบเทียบผลพยากรณ์แต่ละวิธีกับวิธีมาตรฐานใดหนึ่ง ในที่นี้แนะนำให้ใช้ naive method คือ $F_t = Y_{t-1}$ ท่านมีความเห็นเรื่องนี้อย่างไร?

12. ข้อมูลน้ำหนักสุกรที่วัดทุกสัปดาห์ปรากฏดังนี้

สัปดาห์ที่	น.น.(เปานด์)	สัปดาห์ที่	น.น.	สัปดาห์ที่	น.น.
1	48	7	94	13	158
2	54	8	104	14	170
3	60	9	112	15	181
4	67	10	124	16	192
5	76	11	134	17	204
6	86	12	144		

จงวิเคราะห์สมการถดถอยโดยอาศัยโพลิโนเมียล โดยให้ค่อยเพิ่มดีกรีไปเรื่อยจนถึงดีกรีที่ 3 ทั้งนี้ให้ตรวจสอบนัยสำคัญเป็นระยะๆ

13. ในกระบวนการทางเคมี พบว่าสารสกัด Y ผันแปรไปตามความเข้มข้นของสาร X_1 จากข้อมูล 67 ชุด ปรากฏค่าสถิติดังนี้

$$\Sigma y^2 = 1.16, \quad \Sigma x_1 y = -8, \quad \Sigma x^2 = 100$$

ก. จงวิเคราะห์สมการถดถอย $Y = f(X_1)$

นอกนั้นยังพบว่าเวลา (X_2) ที่ใช้ในการเพิ่มความร้อนให้แก่ส่วนผสมจนกระทั่งเกิดปฏิกิริยาที่สมบูรณ์ก็ผันแปรไป พบข้อมูลดังนี้

$$\Sigma x_2^2 = 100, \quad \Sigma x_1 x_2 = -80, \quad \Sigma x_2 y = 10$$

ข. จงวิเคราะห์สมการถดถอย $Y = f(X_1, X_2)$

ค. ทำไม สปส. ความถดถอยในข้อ ก. และ ข. ของตัวแปร X_1 จึงไม่เท่ากัน?

จากผลในข้อ ข. ท่านสรุปผลกระบวนการเคมีนี้อย่างไร?