บทที่ ๒

การวิเคราะห์และผลการสังเคราะห์ข้อมูล

๑. ประวัติและพัฒนาการของวิธีกำลังสองน้อยที่สุด

วิธีกำลังสองน้อยที่สุดได้คิดขึ้นเป็นครั้งแรกในปี ค.ศ. ๑๘๐๖ โดยเลอชองดร์ (Adrian Legendre), (๑๗๕๒ – ๑๘๓๓)[๑] โดยนำมาใช้กับข้อมูลที่ได้จากการสังเกตทางดาราศาสตร์ ต่อมาในปี ค.ศ. ๑๘๐๙ เกาส์ (Gauss) ได้ตั้งทฤษฎีเกี่ยวกับวิธีกำลังสองน้อยที่สุดที่ใช้กับสมการกำลังหนึ่ง และในปี ค.ศ. ๑๙๐๐ มากอฟฟ์ (Mar koff) ก็ได้ตั้งทฤษฎีนี้ขึ้นเช่นกัน จึงรวมเรียกว่า " Gauss-Markoff Theorem "[๒] เป็นทฤษฎีที่แสดงคุณสมบัติของตัวประมาณค่ากำลังสองน้อยที่สุดในสมการ $y = x\beta + e$ เราจะได้ว่าเวคเตอร์สุ่ม e มีคุณสมบัติดังนี้[๓]

ก. $E(e) = 0$

ข. $E(ee^1)\ \sigma^2 I$

ค่าประมาณที่เที่ยงตรงและมีความแปรปวนน้อยที่สุดของ β ของฟังค์ชั่นกำลังหนึ่ง y_i หาได้โดยวิธีกำลังสองน้อยที่สุด นั่นคือ $\hat{\beta} = s^{-1}x'y$ เป็นค่าประมาณที่เที่ยงตรงที่สุดของ β ในเมื่อ $S = x'x$

---

[1] Joln I. Griffin, Statistics, Methods and Applications (New York: Holt Rinehartand Winston Inc., c 1962), p. 230

[2] Samuel S. Wilks, Mathematical Statistics (New York: John Wiley & Sons, Inc., c 1962), p. 285.

[3] Franklin A. Graybill, An Introduction to linear Statistics Models, Vol. I. (New York : Mc Graw-Hill Book Company, Inc., 1961), pp. 114-16.

พิสูจน์

ให้ A เป็นตัวคงที่ p x n เมตริกและให้ $\beta^* = AY$

β^* เป็นฟังก์ชั่นของเส้นตรงทั่วไปของ y ซึ่งเราจะใช้ประมาณ

ค่า β เราจะต้องจำกัด A จนกระทั่ง β^* เป็นตัวประมาณค่าที่เที่ยงตรงที่สุดของ β

ให้ $A = S^{-1}X' + B$

$S^{-1}X'$ ทราบค่า เราจะหา B เมื่อจำกัด A สำหรับความเที่ยงตรง

(Unbiasedness) จะได้ว่า

$$E(\beta^*) = E(AY) = E\left[(S^{-1}X' + B)Y\right]$$

$$= (S^{-1}X' + B)\, X\beta$$

$$= \beta S^{-1}X'X + BX\beta$$

$$E(\beta^*) = \beta + BX\beta$$

แต่ถ้าเป็นตัวประมาณค่าที่เที่ยงตรง $E(\beta^*)$ ต้องเท่ากับ β และได้ว่า

$BX\beta = 0$ ทุกค่าของ β และทำให้ $BX = 0$

เมื่อหาคุณสมบัติที่ดีที่สุด เราจะต้องหาเมตริก B เพื่อที่จะทำให้ความ

แปรปวนของ β_i^* มีค่าน้อยที่สุดเมื่อ i = 1, 2, ...p ในการตรวจสอบ

คุณสมบัตินี้จะต้องพิจารณา Covariance

$$\mathrm{Cov}(\beta^*) = E\left[(\beta^* - \beta)(\beta^* - \beta)'\right]$$

$$= E\left[\{(S^{-1}X' + B)Y - \beta\}\{(S^{-1}X' + B)Y - \beta\}'\right]$$

แทนค่า $y = X\beta + e$ และ $BX = 0$

$$\mathrm{Cov}(\beta^*) = E\left[\{(S^{-1}X' + B)(X\beta + e) - \beta\}\{(S^{-1}X' + B)(X\beta + e) - \beta\}'\right]$$

$$= E\,(S^{-1}X'ee'XS^{-1} + Bee'B' + S^{-1}X'ee'B' + Bee'XS^{-1})$$

$$= \sigma^2 (S^{-1} + BB')$$

ให้ $BB' = G(g)$ จะได้ว่า

$$Cov(\beta^*) = \sigma^2 (S^{-1} + G)$$

ไดอะโกนัลของ $Cov(\beta^*)$ จะเป็นความแปรปวนของ β_i^* ตามลำดับในการทำให้ความแปรปวนของ β_i^* มีค่าน้อยที่สุด เราจะต้องทำให้แต่ละค่าของไดอะโกนัลมีค่าน้อยที่สุด ดังนั้น σ^2 และ S^{-1} เป็นตัวคงที่ เราจะต้องหาเมตริก G ที่ทำให้แต่ละค่าของไดอะโกนัลมีค่าน้อยที่สุด

แต่ $G = BB'$ เป็นบวกดังนั้น $g_{ii} \geqslant 0$ ทำให้ไดอะโกนัลของ $Cov(\beta^*)$ มีค่าน้อยที่สุดเมื่อ $g_{ii} = 0$ สำหรับ $i = 1, 2, \ldots, p$ แต่ถ้า $B = (b_{ij})$, ดังนั้น

$g_{ii} = \sum_{j=1}^{n} b_{ij}^2$ ฉะนั้นถ้า g_{ii} เท่ากับ 0 ทุกค่าของ i จะต้องได้ว่า $b_{ij} = 0$ สำหรับทุกค่าของ i และ j ด้วย หมายความว่า $B = 0$ ซึ่งสอดคล้องกับเกณฑ์ตัวประมาณค่าที่เที่ยงตรงคือ $BX = 0$ ฉะนั้น

$$A = S^{-1}X' \quad \text{และ} \quad \beta^* = \hat{\beta}$$

ในปี ค.ศ. ๑๙๓๑ เออร์วิน (Irwin)[๔] ได้พิสูจน์คุณสมบัติที่สำคัญของวิธีกำลังสองน้อยที่สุดว่า ถ้ามีกลุ่มของตัวแปรอิสสระ N ค่ากระจายเป็นโค้งปกติ u เป็นตัวคงที่ของกลุ่มสมการประมาณค่าอย่างอิสสระโดยใช้ผลบวกกำลังสองของส่วนที่เบี่ยงเบนหารด้วย N - u เป็นตัวประมาณค่าที่เที่ยงตรงของความแปรปวนของประชากร เช่นกระจายค่าที่วัดได้ N ค่าอย่างสุ่ม เป็น p แถวและ q สดมภ์ของสองแฟกเตอร์ ฉะนั้น $pq = N$ จะมี p ค่าของ $\tilde{X}_s'$ ที่ให้ตัวคงที่อิสสระ p-1 ตัวในแถว (ผลบวกของ $\tilde{X}_s'$ เป็น ๐) และมีตัวคงที่อิสสระ q - 1 ตัวในสดมภ์ แต่ละค่าของ $\tilde{X}$ จะให้ตัวคงที่ p + q -1 ตัว และผลบวกกำลังสองของส่วนที่เบี่ยงเบนหารด้วย N - (p + q - 1) จะเป็นค่าความแปรปวนที่เที่ยงตรงซึ่งอาจหาได้โดยใช้ $S(X - \tilde{X})^2$ หารด้วย N - 1 ค่าความแปรปวนที่หาได้ทั้งสองวิธีนี้จะแตกต่างกันโดยมีนัยสำคัญหรือไม่ ขึ้นกับปฏิกิริยาร่วมระหว่างแถวและสดมภ์

---

[4] L.H.C. Tippett, The Methods of Statistics (4th ed.; New York : Dover Publications, Inc., 1952), p. 214.

ต่อมาในปี ค.ศ. ๑๙๓๗ ไอเคน (Aiken) ได้คิดวิธีหาค่าน้ำหนักในสมการถดถอย (regression weight) เมื่อมีตัวแปรมากกว่าสาม[5] สมการสำหรับตัวแปร k ตัวสมการในแบบคะแนนมาตรฐานคือ $z_1' = \beta_2 z_2 + \beta_3 z_3 + \dots\dots\dots\dots \beta_k z_k$ สมการที่ใช้คะแนนดิบจะเปลี่ยนเป็นคะแนนมาตรฐานได้โดยให้ $z_i = (x_i - \bar{x}_i)/S_i$ จัดรูปใหม่ $x_1' = \beta_2 \frac{S_1}{S_2} x_2 + \beta_3 \frac{S_1}{S_3} + \dots\dots + \beta_k \frac{S_1}{S_k} x_k + A \dots\dots (1)$
เมื่อ A จะหาได้จาก

$$A = \bar{x}_1 - \beta_2 \frac{S_1}{S_2} \bar{x}_2 - \beta_3 \frac{S_1}{S_3} \bar{x}_3 - \dots\dots\dots - \beta_k \frac{S_1}{S_k} \bar{x}_k \dots\dots (2)$$

สัมประสิทธิสหสัมพันธ์หาได้จาก

$$R = \sqrt{\beta_2 r_{12} + \beta_3 r_{13} + \dots\dots\dots + \beta_k r_{1k}} \qquad \dots\dots (3)$$

วิธีคำนวณหาค่าสัมประสิทธิ, A และสัมประสิทธิสหสัมพันธ์เพื่อให้ได้ค่าตามสมการ (๑), (๒), และ (๓) ดังแสดงไว้ในภาคผนวก[๖]

ในปีเดียวกันนั้นเองวัตสัน (Watson) ได้นำวิธีกำลังสองน้อยที่สุดมาใช้ประมาณค่าพื้นที่เฉลี่ยของใบไม้จากน้ำหนักโดยใช้กลุ่มตัวอย่างเล็ก ๆ เป็นการประมาณค่าโดยใช้สมการถดถอย ซึ่งจะช่วยให้ประหยัดเวลาได้มาก

ในปี ค.ศ. ๑๙๔๒ ไอเคน (Aitken) และซิลเวอร์สโตน (Silverstone)[๗] พบว่าถ้าความแปรปวนของตัวประมาณค่าที่มีความแปรปวนต่ำสุดมีค่าเท่ากับ λ แสดงว่าการกระจายทั่วไปส่วนมากซึ่งมัชฌิม (Mean) ของกลุ่มตัวอย่างเป็นสัมประสิทธิ์ตัวประมาณค่าที่เที่ยงตรง จะมีลักษณะดังนี้

$$f(x,\theta) = c(x, \tau) \exp\left\{ -\frac{1}{2\tau^2}(x-\theta)^2 \right\}$$

---

[5] George A. Furgurson, Statistical Analysis in Psychology and Education (New York : Mc Graw-Hill Book Company, Inc., 1959), pp. 297 - 98.

[๖] ภาคผนวก ๑

[๗] Maurice G. Kendall, The Advanced Theory of Statistics, Vol. 2, (3 d ed., London: Charles Griffin & Company Limited, 1955), p. 61.

เมื่อ c เป็นฟังก์ชั่นใด ๆ และ σ^2 เป็นค่าความแปรปวนของ f และถ้าฟังค์ชั่น λ มีจริงและ

$$\alpha(\theta) = \int_0^\theta \frac{d\theta}{\lambda(\theta)}$$

ความแปรปวนของตัวประมาณค่าจะเท่ากับ $-\frac{1}{n} \frac{\partial^2 q}{\partial x^2}$ ฟังก์ชั่น q ได้จากการทำให้สมการข้างล่างมีค่าต่ำสุด

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left\{ x_i - \sum_{j=1}^{s} (a_{ij}\, q_j) \right\}^2$$

ต่อมาในปี ค.ศ. ๑๙๔๓ แมน (Mann) และ วาลด์ (Wald)[๘] ได้คิดทฤษฎีพื้นฐานในการประมาณค่า antoregressive series ขึ้น โดยพิสูจน์ให้เห็นชัดว่า คุณสมบัติของกลุ่มตัวอย่างของตัวประมาณค่ากำลังสองน้อยที่สุด $\hat{\alpha}$ ของ α จะเหมือนกับตัวประมาณค่ากำลังสองน้อยที่สุดในสมการถดถอยหลายตัวแปรที่กระจายเป็นโค้งปกติ ผลลัพธ์นี้ใช้ได้ดีในทางปฏิบัติ

สำหรับ Linear antoregressive series ทั่ว ๆ ไป

$$\sum_{j=0}^{k} \alpha_j\, u_{t-j} = \varepsilon_t$$

ในทางการทดลองศึกษาอนุกรม (series) ที่เกิดจากการกระจายเป็นสี่เหลี่ยมของ ε's สำหรับ n ที่มี ๖๐ เทอมพบว่าสมการของยูลวอคเคอร์ (Yule Walker) ใช้ได้ดี สมการนี้ได้จากการทำให้ฟังก์ชั่นต่อไปนี้มีค่าน้อยที่สุด

$$\sum_{t=k+1}^{n} \left\{ \sum_{j=0}^{k} \alpha_j\, u_{t-j} \right\}^2$$

---

[8]Manrice G. Kendall & Alan, Staurt, The Advanced Theory of Statistics, Vol.3.(2 ed.; London: Charles Griffin & Company Limited, 1968), p. 476.

ในปี ค.ศ. ๑๙๔๗ เจสเสน (Grecce Jessen)[๙] พบว่าความแปรปวนของ y_i เพิ่มตามค่าของ x_i^2 ซึ่งทำให้เกิด Weighted regression $w_i \propto \frac{1}{x_i^2}$ สำหรับค่าประมาณกำลังสองน้อยที่สุดของ b ได้จาก

$$b = \frac{\sum wi \;\; yi \; xi}{\sum wi \;\; xi^2} = \frac{1}{n} \sum \left(\frac{yi}{xi} \right)$$

ในกรณีนี้ตัวประมาณค่าที่ดีที่สุดของ $\bar{y}$ คือ $b\bar{x}$ เมื่อ b เป็นมัธฌิมของอัตราส่วน $\frac{y_i}{x_i}$ ของแต่ละหน่วยในกลุ่มตัวอย่างสำหรับการสุ่มตัวอย่างจากประชากร อัตราส่วนนี้จะเป็นตัวประมาณค่าที่เที่ยงตรงถ้า

๑. ความสัมพันธ์ระหว่าง y_i และ x_i เป็นเส้นตรงผ่านจุดกำเนิด

๒. ความแปรปวนของ y_i เป็นสัดส่วนโดยตรงกับ x_i

ปีต่อมา ค.ศ. ๑๙๔๘ ฮันส์เนอร์ (G.W. Honsner) และเบรนเนน (J.F.Brennan)[๑๐] ได้ศึกษาการใช้วิธีกำลังสองน้อยที่สุด ในสมการถดถอยสองตัวแปร ซึ่งตามวิธีของไอเคน, ววล์ด และคนอื่น ๆ จะต้องทราบค่าความแปรปวนของตัวแปรสุ่ม นอกจากนี้จะทำให้ประสิทธิภาพลดลงในทางปฏิบัติ วิธีของฮันส์เนอร์จะมีความคงที่ (Consistency) และมีประสิทธิภาพ (Efficiency) สูงกว่าวิธีอื่น ๆ พิจารณาตัวแปร ๒ ตัว

$$\mathcal{O}_\eta(x_{ik}, y_{ik}) , \quad i = 1, 2, \ldots\ldots\ldots\ldots, n$$
$$k = 1, 2, \ldots\ldots\ldots\ldots, N_i$$

---

[9] William G. Cochran, Sampling Techniques (New York : John Wiley & Sons, Inc., c 1953), p. 124.

[10] G.W. Honsner and J.F. Brennan, "The Estimation of Linear Trends", The Annals of Mathematics Statistics, Vol.19 (1948), pp. 380 - 87.

เมื่อ N_i เป็นตัวอย่างที่มีค่า x_i , y_i กระจายรอบมัชฌิม X_i, Y_i ให้มัชฌิมมีความสัมพันธ์กันดังนี้

$$Y_i = a + bX_i$$

และให้ตัวแปรสุ่ม x_i เป็นตัวแปรอิสสระ การกระจายความถี่โดยมีความแปรปวน $= \epsilon_y^2$ สำหรับทุกค่าของ i และตัวแปรสุ่ม y_i มีการกระจายความถี่โดยมีความแปรปวน $= \epsilon_y^2$ สำหรับทุกค่าของ i สถิติที่ใช้ประมาณค่า b ได้จากการสังเกตตัวอย่างคู่หนึ่ง

(x_{ik}, y_{ik}), (x_{jl}, y_{jl})

(1) .... $b_{ik,jl} = \dfrac{y_{ik} - y_{jl}}{x_{ik} - x_{jl}}$

และได้ว่า

(2) .... $y_{ik} = a + bx_{ik} + \eta_{ik} - b\,\epsilon_{ik}$

สมการ (๑) จะเขียนได้ดังนี้

$$(x_{ik} - x_{jl})\,\hat{b}_{ik,jl} = (x_{ik} - x_{jl})b + (\eta_{ik} - \eta_{jl}) - b(\epsilon_{ik} - \epsilon_{jl})$$

บวกสมการนี้ทุกจุดจะได้

(3)...... $\ddot{b} = \dfrac{\sum_i \sum_j \sum_k \sum_l (y_{ik} - y_{jl})}{\sum_i \sum_j \sum_k \sum_l (x_{ik} - x_{jl})} - \dfrac{\sum_i \sum_j \sum_k \sum_l \left((\eta_{ik} - \eta_{jl}) - b(\epsilon_{ik} - \epsilon_{jl})\right)}{\sum_i \sum_j \sum_k \sum_l (x_{ik} - x_{jl})}$

ผลบวกข้างบนจะได้จาก

l = 1, 2, 3.....Nj; k = 1, 2,....., Ni ; j = 1,2....(i -1); i = 1,2,..n

เทอมแรกทางด้านขวาของสมการ (๓) คือค่าประมาณของ b และเทอมที่สองแทนส่วนที่เบี่ยงเบนของค่าประมาณจากค่าจริง จากนี้เราจะได้ค่าประมาณ b

(4)....... $\hat{b} = \dfrac{\sum_i \sum_j \sum_k \sum_l (y_{ik} - y_{jl})}{\sum_i \sum_j \sum_k \sum_l (x_{ik} - x_{jl})}$

บวกตาม k และ l จะได้

$$\hat{b} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{i-1} N_i N_j (\bar{y}_i - \bar{y}_j)}{\sum_{i=1}^{n} \sum_{j=1}^{i-1} N_i N_j (\bar{x}_i - \bar{x}_j)}$$

เมื่อ $\bar{y}_i$ เป็นมัชฌิมของ y_i บวกตาม j ให้

$$(5) \ldots\ldots\ \hat{b} = \frac{\sum_i \left(N_i \bar{y}_i \sum_{i=1}^{k-1} N_j - N_i \sum^{i-1} N_j \bar{y}_i \right)}{\sum_i \left(N_i \bar{x}_i \sum_i N_j - N_i \sum N_j \bar{x}_j \right)}$$

สามารถเขียนให้สะดวกดังนี้

$$\sum_{i=1}^{n} \left(N_i \sum_{i=1}^{i-1} N_j \bar{N}_j \right) = \sum_{i=1}^{n} \left(N_i \bar{y}_i \sum_{i+1}^{n} N_j \right)$$

005860

$$= \sum_{i=1}^{N} \left(N_i \bar{y}_i \left(\sum^{n} N_j - \sum^{i} N_j \right) \right)$$

แทนค่าใน (๑) จะได้

$$(6) \ldots\ldots\ b = \frac{\sum_{i=1}^{n} \left[N_i \bar{y}_i \left(\sum^{n} N_j - 2 \sum^{i} N_j + N_i \right) \right]}{\sum_{i=1}^{n} \left[N_i \bar{x}_i \left(\sum^{n} N_j - 2 \sum^{i} N_j + N_i \right) \right]}$$

เป็นการประมาณค่าของแนวโน้มของสมการกำลังหนึ่งเมื่อมีสองตัวแปร ข้อสังเกตเส้น

$y = \hat{a} + \hat{b}x$ จะพอเหมาะกับจุดในกลุ่มตัวอย่างโดยให้ผ่านมัชฌิมของกลุ่มตัวอย่างนั้นคือ

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x}$$

เมื่อ $\bar{y}$ และ $\bar{x}$ เป็นมัชฌิมของทุก y_{ik} และ x_{ik} ตามลำดับ

นอกจากนี้เขายังได้แสดงให้เห็นว่าตัวประมาณค่านี้มีความคงที่ (Consistent) และหาประสิทธิภาพเปรียบเทียบกับวิธีอื่นโดยมีตัวอย่างเป็นตัวเลขประกอบสูตรที่ใช้เปรียบเทียบกันคือ

๑. สถิติจากวิธีนี้ $\hat{b}_1 = \dfrac{\sum^{n} y_i (i - \bar{i})}{\sum^{n} x_i (i - \bar{i})}$

๒. สถิติที่ได้จากการทำให้ผลบวกกำลังสองของส่วนที่เบี่ยงเบนจาก y มีค่าน้อยที่สุด $\hat{b}_2 = \dfrac{\sum^{n} y_i (x_i - \bar{x})}{\sum x_i (x_i - \bar{x})}$

๓. สถิติที่ได้จากการทำให้ผลบวกกำลังสองของส่วนที่เบี่ยงเบนจากออทอโกนัลมีค่าน้อยที่สุด

$$\hat{b}_3 = \frac{\pm \left[n \sum^{n} (y_i - \bar{y})^2 - n \sum^{n} (x_i - \bar{x})^2 + 4 (\sum^{n} (y_i - \bar{y})(x_i - \bar{x}) \right]^{\frac{1}{2}}}{\sum^{n} (y - \bar{y})(x - \bar{x})}$$

๔. สถิติที่ได้จากวิธีของวอลด์ (Wald)

$$\hat{b}_4 = \frac{\sum^{n/2} y_i - \sum_{n/2}^{n} y_i}{\sum^{n/2} x_i - \sum_{n/2}^{n} x_i}$$

พบว่า $\hat{b}_1$ ได้ค่าละเอียดที่สุด และได้ค่าใกล้เคียงกว่า $\hat{b}_3$ ความแปรปวนของกลุ่มตัวอย่างเท่ากับ ๐.๖๕ นอกจากนี้ $\hat{b}_1$ ยังคำนวณง่ายกว่า $\hat{b}_3$

ในปี ค.ศ. ๑๙๔๙ เยทส์ (Yates)[๑๑] ก็ได้ใช้สมการถดถอยในการประมาณค่าบริมาตรของไม้สักอย่างคร่าว ๆ ในบริเวณ ๑/๑๐ เอเคอร์ แล้วหาบริมาตรที่แท้จริงจากกลุ่มตัวอย่างสร้างสมการถดถอยสำหรับประมาณค่าคือ $\bar{Y} + b(\bar{X} - \bar{x}_i)$ การประมาณค่าอย่างคร่าว ๆ มีค่าผิดพลาด $D = x_i - y_i$ ถ้า $b = 1$

$$\bar{Y} + (\bar{X} - \bar{x}_i) = \bar{X} + (\bar{y} - \bar{x}_i)$$

= มัชฌิมของการประมาณคร่าว ๆ + ค่าแก้

ต่อมาในปี ค.ศ. ๑๙๖๐ ดูบิน (Dubin)[๑๒] ได้ขยายทฤษฎีแมนวอลด์ (Mann-Wald theorem) โดยศึกษาคุณสมบัติของตัวประมาณค่ากำลังสองน้อยที่สุดในระบบที่ผสม

---

11 Cocharn, Opcit., p. 141.

12 Kendall, Opcit., p. 498.

ระหว่าง antoregressive และ regressive เมื่อมี antoregressive set ของ y เปลี่ยนค่าตามค่า x ที่คงที่

$$\sum_{j=0}^{k} \alpha_j y_{t-j} = \sum_{i=1}^{q} \beta_i x_{it} + \epsilon_t$$

จะได้ $$y_t = -\sum_{j=1}^{k} \alpha_j y_{t-j} + \sum_{i=1}^{q} \beta_i x_{it} + \epsilon_t$$

พบว่าตัวประมาณกำลังสองน้อยที่สุดในระบบเช่นนี้มีคุณสมบัติเช่นเดียวกับเมื่อไม่มีตัวแปร y มาเกี่ยวข้อง โดยไม่คำนึงว่า residuals จะกระจายเป็นโค้งปกติหรือไม่

ในปีเดียวกันนั้น เวบสเตอร์[๑๓] (John Thomas Webster) ก็ได้ศึกษาตัวแปรอิสสระในสมการกำลังหนึ่ง โดยพิจารณาสมการถดถอยหลายตัวแปร

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

ปัญหาที่เขาศึกษาคือเปรียบเทียบตัวประมาณค่า ๒ ตัวว่าตัวประมาณค่าใดจะเป็นตัวประมาณค่าของ E (y) ที่ดีกว่า

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots\cdots\cdots\cdots \; b_p x_p$$

หรือ $$\hat{\hat{y}} = c_0 + c_1 x_1 + c_2 x_2 + \cdots\cdots\cdots\cdots \; c_{p-1} x_{p-1}$$

ซึ่งในแต่ละกรณีสัมประสิทธิ์เป็นตัวประมาณค่ากำลังสองน้อยที่สุด เขาได้ชี้ให้เห็นว่าตัวประมาณค่าทั้งสองถ้ามี mean square error น้อยกว่าจะใช้ได้ดีกว่าและควรจะนำมาใช้มากกว่าตัวประมาณค่าที่เที่ยงตรงและมีความแปรปวนน้อยที่สุด

พิจารณา $E\left[\hat{y} - E(y)\right]^2$ เทียบกับ $E\left[\hat{\hat{y}} - E(y)\right]^2$

พบว่า mean square error ของ $\hat{\hat{y}}$ มีค่าน้อยกว่าเมื่อ

$$\delta = \frac{\beta_p}{\sqrt{Var\,(b_p)}} < 1$$

---

13 John Thomas Webster, "A Decision Procedure for The inclusion of An Independent Variable in Linear Estimator", Dissertation Abstracts, 22(January, 1961), p. 275.

จากนี้ก็ได้มีการนำวิธีกำลังสองน้อยที่สุดไปใช้โดยวิธีการต่าง ๆ กันเช่นเมื่อปีที่แล้ว (ค.ศ. ๑๙๗๒) มิตรา (S.K. Mitra)[๑๔] ได้ศึกษาการประมาณค่าในสมการกำลังหนึ่งโดยใช้วิธีกำลังสองน้อยที่สุดตามสมการของเกาส์มากอฟฟ์ดังนี้ สำหรับสมการเส้นตรง $(y, x\beta, \sigma^2 v)$ V เป็นค่าที่ทราบ (อาจเป็นค่าเดียว) ซายคินส์ (Zyskind), มาร์ติน (Martin) และราว (Rao) ได้แสดงว่าการเลือกค่าประมาณของ M ซึ่งเป็นตัวประมาณค่าที่เที่ยงตรงของฟังค์ชั่น $p'\beta$ จะได้จาก $p'\hat{\beta}$ ถ้า $\hat{\beta}$ เป็นผลลัพธ์ของสมการปกติ $X' MX \beta = X' MY$ และได้เลือกค่า M เมื่อแสดงว่า

$$\frac{R_o^2}{\sigma^2} = (Y - X\hat{\beta})' M (Y - X\hat{\beta}) \sim \chi^2$$

แต่มิตรา (Mitra) ได้แสดงว่าไม่ว่าจะเลือกค่า M อย่างไร

$$V(p'\hat{\beta}) = p'(X'MX)^{-}p\,\sigma^2$$

จะไม่เป็นจริงทุกค่าของ $p'\beta$ นอกจาก $m(x) \subset m(v)$ ตามทฤษฎีกำลังสองน้อยที่สุดเดิมและสำหรับสมมุติฐานเส้นตรงซึ่งคงที่และตรวจสอบได้ H_o ถ้า $R^2_{H_o}$ เป็นค่าน้อยที่สุดของ $(Y - X\beta)' M(Y - X\beta)$ขึ้นกับ H_o เมื่อ $(R^2_{H_o} - R^2_o)/\sigma^2 \sim \chi^2$ โดยไม่ขึ้นกับ R_o^2 ถ้า M ได้จากการเลือก V^- ตามคำแนะนำของซายคินส์ (Zykind) และมาร์ติน (Martin) ผลลัพธ์นี้จะไม่เป็นจริงสำหรับผลลัพธ์ที่ได้จากการคำนวณที่ง่ายกว่าของราว (Rao) คือ $M = (V + c\,XX')^-$ และได้แสดงว่า aV^- สามารถจะคำนวณได้ใกล้เคียงกับ $M = (v + c\,XX')^-$ มีขั้นต่าง ๆ ของการคำนวณแสดงด้วยตัวอย่างที่เป็นตัวเลข (V^- หมายถึงเมตริกที่มีคุณสมบัติว่า $VV^- \neq I$)

๒. หลักการประมาณค่าโดยวิธีกำลังสองน้อยที่สุด

๒.๑ การประมาณค่าทางสถิติ (Statistical estimation) เป็นวิธีการใช้กลุ่มตัวอย่างประมาณค่าพารามิเตอร์ของประชากรทั้งหมด สถิติที่ใช้ประมาณค่าพารามิเตอร์

---

14 S.K. Mitra, "Unified Least Squares Approach to Linear Estimation in a General Gauss Markov Model," The Institute of Mathematical Statistics Bulletin, Vol. 1 No. 4 (1972) p. 169.

เรียก "ตัวประมาณค่า (Estimator)" ค่าที่ได้จากตัวประมาณค่า เรียก "ค่าประมาณ (Estimate)" ค่าประมาณของพารามิเตอร์สามารถจะหาได้ ๒ วิธี คือการประมาณค่าเป็นจุด (Point Estimation) และการประมาณค่าเป็นช่วงตัวเลขหรืออันตรภาพ (Interval Estimation).

การประมาณค่าเป็นจุด (Point Estimation) จะได้ค่าตัวเลขตัวเดียวที่ใช้แทนพารามิเตอร์ซึ่งจะไม่เท่ากับตัวพารามิเตอร์ ดังนั้นแทนที่จะใช้ตัวเลขใด ๆ แสดงคุณสมบัติของประชากรที่เราไม่ทราบจึงควรใช้เป็นช่วงตัวเลข

การประมาณค่าเป็นช่วงตัวเลขหรืออันตรภาค (Interval Estimation) เป็นการตั้งขอบเขตระหว่างค่าที่ประมาณได้เป็นจุด ซึ่งหาได้โดยใช้ขอบเขตของความเชื่อมั่นที่จะมีพารามิเตอร์ของประชากรทั้งหมดอยู่ภายในขอบเขตนี้

แม้ว่าจะแสดงให้เห็นว่าตัวกลางเลขคณิตของกลุ่มตัวอย่าง $\bar{x}$ เป็นค่าประมาณที่ดีของ μ แต่ก็เป็นการยากที่ $\bar{x}$ จะเท่ากับ μ ด้วยเหตุนี้การประมาณค่าเป็นช่วงตัวเลขจึงดีกว่า การประมาณค่าเป็นช่วงเริ่มต้นจากการประมาณค่าเป็นจุด และขยายออกเป็นช่วงโดยใช้ค่าที่ยอมให้ผิดพลาด ถ้า s =ค่าที่ยอมให้ผิดพลาด ช่วงตัวเลขที่ประมาณได้จะ= $\bar{x} \pm s$

๒.๑.๑ คุณสมบัติของตัวประมาณค่า (Estimator) ที่ดี

การเลือกตัวประมาณค่าต้องพิจารณาคุณสมบัติต่อไปนี้

ก. ความไม่เอนเอียง (Unbiasedness) ตัวประมาณค่าจะเที่ยงตรงถ้าค่าที่คาดหวัง (Expected value) ของสถิติที่ใช้มีค่าเท่ากับพารามิเตอร์ของประชากรที่จะประมาณค่า $E(\hat{\theta}) = \theta$

θ = แทนค่าพารามิเตอร์ซึ่งจะประมาณค่า

$\hat{\theta}$ = แทนตัวประมาณค่า (Estimator)

เช่น ค่าที่คาดหวังของค่าสถิติ $\bar{x}$ ได้จากมัชฌิมเลขคณิตของการกระจายของกลุ่มตัวอย่างและมีค่าเท่ากับมัชฌิมเลขคณิตของประชากร (μ) เสมอ

$$E(\bar{x}) = E\left(\frac{x_1 + x_2 + \dots\dots\dots x_n}{n}\right) = \mu$$

มัชฌิมเลขคณิตของกลุ่มตัวอย่างจึงเป็นตัวประมาณค่าที่เที่ยงตรงของมัชฌิมเลขคณิตของประชากร

แต่ค่าคาดหวังของความแปรปวนจากกลุ่มตัวอย่างจะไม่เท่ากับความแปรปวน

ของประชากรนั่นคือ

$E(s^2) \neq \sigma^2$

s^2 = ความแปรปวนของกลุ่มตัวอย่าง (Sample Variance)

σ^2 = ความแปรปวนของประชากร (Population Variance)

ความแปรปวนของกลุ่มตัวอย่างจึงไม่เป็นตัวประมาณค่าที่เที่ยงตรงของความแปรปวนของประชากร

ถ้า $E(s^2) = E\left(\frac{\Sigma(x - \bar{x})}{n}\right)^2 = \frac{(n-1)\ \sigma^2}{n}$

s^2 และ σ^2 คำนวณได้จาก

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n}$$

$$= \frac{\Sigma(x - \mu)^2}{N}$$

แต่ถ้าใช้สูตร $\hat{s}^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$

และ $\hat{\sigma}^2 = \frac{\Sigma(x - \mu)^2}{N - 1}$

จะได้ว่า $E(\hat{s}^2) = \hat{\sigma}^2$

ดังนั้น $\hat{s}^2$ เป็น ตัวประมาณค่าที่ไม่เอนเอียงของ σ^2

$\hat{s}^2$ =ความแปรปวนของกลุ่มตัวอย่างที่เรียกว่า " Modified sample variance"

และ $\hat{\sigma}^2$ = ความแปรปวนของประชากรที่เรียกว่า " Modified population variance"

<u>ตัวอย่าง</u>

สมมุติว่ากลุ่มประชากรมี ๕ หน่วย A,B,C,D,E แต่ละหน่วยมีค่าต่าง ๆ กันเป็น ๒, ๔, ๖, ๘, ๑๐

ตารางที่ ๑

	ค่าของแต่ละหน่วย	$x - \mu$	$(x - \mu)^2$
A	๒	−๔	๑๖
B	๔	−๒	๔
C	๖	๐	๐
D	๘	๒	๔
E	๑๐	๔	๑๖
	๓๐		๔๐

$$\mu = \frac{\sum x}{N} = \frac{๓๐}{๕} = ๖$$

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} = \frac{๔๐}{๕} = ๘$$

$$\hat{\sigma}^2 = \frac{\sum (x - \mu)^2}{N - 1} = \frac{๔๐}{๔} = ๑๐$$

ถ้ามีตัวอย่างสุ่ม (Random sample) ขนาด n = ๒ หามัชฌิมเลขคณิตและความแปรปวนของกลุ่มตัวอย่างทั้งหมดที่เป็นไปได้ จะได้ค่าต่าง ๆ ตามตาราง.

ตารางที่ ๒

กลุ่มตัวอย่าง	มัชฌิมเลขคณิต $\bar{x}$	ความแปรปวน s^2	ความแปรปวนแบบ Modified Variance ($\hat{s}^2$)
A,B	(2 + 4)/2 = 3	1	2
A,C	(2 + 6)/2 = 4	4	8
A,D	(2 + 8)/2 = 5	9	18
A,B	(2 + 10)/2 = 6	16	32
B,C	(4 + 6)/ 2 = 5	1	2
B,D	(4 + 8)/2 = 6	4	8
B,E	(4, + 10)/2 = 7	9	18
C,D	(6 + 8)/2 = 7	1	2
C. E	(6 + 10)/2 = 8	4	8
D,E	(8 + 10)/2 = 9	1	2
รวม	60	50	100
ค่าที่คาดหวัง	$E(\bar{x}) = \frac{60}{10} = 6$	$E(s^2) = \frac{50}{10} = 5$	$E(\hat{s}^2) = \frac{100}{10} = 10$

จากค่าตัวเลขที่คำนวณได้จะเห็นว่า $E(\bar{x}) = \mu$

$$E(s^2) \neq \sigma^2$$

$$E(\hat{s}^2) = \hat{\sigma}^2$$

$\hat{s}^2$ และ $\hat{\sigma}^2$ จะหาได้จาก s และ σ ดังนี้

$$\hat{s}^2 = s^2\left(\frac{n}{n-1}\right) \quad \text{และ} \quad \hat{\sigma}^2 = \sigma\left(\frac{N}{N-1}\right)$$

ถ้าขนาดของประชากร (N) เพิ่มขึ้น $\frac{N}{N-1} \longrightarrow 1$

ดังนั้น $\hat{\sigma}^2 \longrightarrow \sigma^2$

ดังนั้น $\hat{s}^2$ จะเป็นตัวประมาณค่าที่เที่ยงตรงของ σ^2 ถ้าทั้ง n และ N มากขึ้น $\frac{n}{n-1}$ และ $\frac{N}{N-1} \longrightarrow 1$ และ s^2 จะเป็น ตัวประมาณค่าที่ไม่เอนเอียง ของ σ^2

บางครั้งอาจจะหาตัวประมาณค่าไม่เอนเอียงของพารามิเตอร์ได้มากกว่า ๑ ตัว เช่น ถ้าประชากรมีการกระจายสม่ำเสมอทั้งมัชฌิมเลขคณิตของกลุ่มตัวอย่าง (Sample mean) และมัธยฐานของกลุ่มตัวอย่าง (Sample median) จะเป็นตัวประมาณค่าที่เที่ยงตรงของ μ ต้องพิจารณาคุณสมบัติข้ออื่นของตัวประมาณค่า (Estimator) ต่อไป

ข. ความคงเส้นคงวา (Consistency) คือตัวประมาณค่าที่ค่าจากกลุ่มตัวอย่างเข้าใกล้พารามิเตอร์เมื่อขนาดของกลุ่มตัวอย่างเพิ่มขึ้น ๆ หรือจะกล่าวว่าตัวประมาณค่า $\hat{\theta}_n$ จะเป็นตัวประมาณค่าที่คงเส้นคงวา ของ θ ถ้าขอบเขตของความน่าจะเป็นที่ $|\hat{\theta}_n - \theta| < \epsilon$ เป็น ๑ เมื่อ $n \longrightarrow \infty$ เขียนเป็นสัญญลักษณ์ได้ดังนี้

$$\lim_{n \to \infty} \text{Prob} \left\{ \left| \hat{\theta}_n - \theta \right| < \epsilon \right\} = 1$$

$\hat{\theta}_n$ เป็นตัวประมาณค่าที่ได้จากข้อมูล n หน่วย

θ เป็นพารามิเตอร์ที่ต้องการจะประมาณค่า

ϵ เป็นค่าบวกใด ๆ

ตามตัวอย่างที่แล้ว $\bar{x}$ และ s^2 เป็นตัวประมาณค่าที่คงเส้นคงวาของ μ และ σ^2 แต่มัธยฐานของกลุ่มตัวอย่าง (Sample median) จะเป็น ตัวประมาณค่าที่คงเส้นคงวาของ μ เมื่อการกระจายของประชากรสมมาตรกัน (Symmetry) เท่านั้นและมัธยฐานของกลุ่มตัวอย่าง (Sample median) จะเข้าใกล้มัธยฐานของประชากร (Population median) เมื่อขนาดของกลุ่มตัวอย่างเพิ่มขึ้น ฉะนั้นมัธยฐานของกลุ่มตัวอย่างจะเป็นตัวประมาณค่าที่คงเส้นคงวา ของมัธยฐานของประชากร

ถ้าใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ เพื่อประมาณค่า θ **Consistency** ของตัวประมาณค่าเป็นคุณสมบัติที่มีความสำคัญมาก แต่ถ้าใช้กลุ่มตัวอย่างที่มีขนาดเล็กมาก consistency ของตัวประมาณค่าจะไม่มีความสำคัญมากนัก นอกจากขอบเขตของความน่าจะเป็นข้างบน จะมีค่าเกือบเท่า ๑ เมื่อ n เล็กมาก

ค. ประสิทธิภาพ (Efficiency) ตัวประมาณค่า $\hat{\theta}_n$ จะมีประสิทธิภาพเมื่อเป็นไปตามข้อแม้ ๒ ประการคือ

๑. การกระจายของ $\sqrt{n}\ (\hat{\theta}_n - \theta)$ เป็นโค้งปกติ โดยมีมัชฌิมเลขคณิตเป็น ๐ และความแปรปวน = σ^2 เมื่อ n เป็นขนาดของกลุ่มตัวอย่าง

๒. ถ้าความแปรปวนของตัวประมาณค่าน้อย การกระจายของตัวประมาณค่า (Estimator) จะรวมกลุ่มอยู่ใกล้พารามิเตอร์ตัวประมาณค่านี้จะมีค่าใกล้เคียงกับพารามิเตอร์มากกว่า

ถ้ามีตัวประมาณค่าที่เที่ยงตรง ๒ ตัว เราจะเลือกตัวประมาณค่าได้โดยดูประสิทธิภาพเปรียบเทียบกัน (Relative Efficiency)

ให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณค่าที่เที่ยงตรงของพารามิเตอร์ตัวเดียวกัน θ ถ้าความแปรปวนของ $\hat{\theta}_1$ น้อยกว่าความแปรปวนของ $\hat{\theta}_2$ เรากล่าวว่า $\hat{\theta}_1$ จะมีประสิทธิภาพสูงกว่า $\hat{\theta}_2$

สำหรับการกระจายของประชากรที่สมมาตรกัน (Symmetry) ทั้งมัชฌิมเลขคณิตและมัธยฐานของกลุ่มตัวอย่างเป็น Unbiased และ Consistent estimator ของ μ เราจะเลือกระหว่างตัวประมาณค่าทั้งสองโดยดูประสิทธิภาพเปรียบเทียบ แล้วเลือกตัวที่มีความแปรปวนน้อยกว่า

ความแปรปวนของ $\bar{x}$ = $\frac{\sigma^2}{n}$

ความแปรปวนของมัธยฐาน= $\frac{\pi}{2}\frac{\sigma^2}{n}$ สำหรับ n ที่ใหญ่มาก

$$\therefore \quad \frac{\text{Var (Med.)}}{\text{Var }(\bar{x})} = \frac{(\frac{\pi}{2})(\frac{\sigma^2}{n})}{(\frac{\sigma^2}{n})} \approx 1.57$$

$\bar{x}$ จึงเป็นตัวประมาณค่าของ μ ที่ดีกว่ามัธยฐาน เพราะเป็นตัวประมาณค่าที่มีประสิทธิภาพมากกว่า

ฉะนั้นในจำนวนตัวประมาณค่าที่เที่ยงตรง จะมีตัวประมาณค่าที่มีความแปรปรวนน้อยที่สุด (Minimum variance estimator) ซึ่งจะเป็นตัวประมาณค่าที่มีประสิทธิภาพสูงที่สุดด้วย ในตัวอย่างที่กล่าวแล้ว $\bar{x}$ จะเป็น Minimum variance estimator ของ μ

ง. Sufficiency ตัวประมาณค่าที่ใช้คุณสมบัติทุกอย่างของพารามิเตอร์ที่มีอยู่ในกลุ่มตัวอย่างเรียก " Sufficient Estimator " เช่นถ้าใช้ค่าแรกที่ได้จากการทดลองหรือค่าสุดท้ายหรือค่าเฉลี่ยระหว่าง ๒ ค่านี้ เป็นตัวประมาณค่าพารามิเตอร์ ตัวประมาณค่านี้จะไม่เป็น sufficient estimator เพราะตัวประมาณค่าเหล่านี้ไม่ได้พิจารณาทุกหน่วยในกลุ่มตัวอย่าง $\bar{x}$ คำนวณจากทุกค่าในกลุ่มตัวอย่างนั้นจึงเป็น sufficient estimator ของ μ

จะเห็นว่า $\bar{x}$ มีคุณสมบัติของตัวประมาณค่าที่ดีครบทั้ง ๔ อย่าง

๒.๑.๒ วิธีหาตัวประมาณค่าเป็นจุดที่ดี (Methods of Finding A Good Point Estimator)

วิธีหาตัวประมาณค่าที่ดีของพารามิเตอร์โดยทั่วไปหาได้ ๔ วิธี

ก. วิธีโมเมนต์ (Method of Moment) วิธีโมเมนต์เป็นวิธีประมาณค่าพารามิเตอร์จากกลุ่มตัวอย่างโดยทำให้โมเมนต์ของกลุ่มตัวอย่างเป็นไปตามโมเมนต์ของประชากร โมเมนต์ของประชากรเป็นฟังก์ชั่นของตัวพารามิเตอร์ที่ไม่ทราบค่า จำนวนโมเมนต์ที่ใช้ขึ้นกับจำนวนพารามิเตอร์ที่ต้องการจะประมาณค่า เช่นถ้าจะประมาณค่ามัชฌิมเลขคณิตซึ่งเรียกว่า "โมเมนต์ที่ ๑ (First moment)" เราใช้มัชฌิมเลขคณิตของกลุ่มตัวอย่าง ถ้าจะประมาณค่าความแปรปวนซึ่งเรียกว่า "โมเมนต์ที่ ๒ (Second moment) รอบมัชฌิมเลขคณิตของตัวแปรสุ่ม จะประมาณค่าได้โดยใช้ความแปรปรวนของกลุ่มตัวอย่าง

โดยทั่วไปค่าที่ได้จากวิธีโมเมนต์จะกระจายเป็นโค้งปกติ เมื่อขนาดของกลุ่มตัวอย่าง (n) เพิ่มขึ้น แม้ว่าวิธีโมเมนต์จะเป็นวิธีที่ไม่ค่อยมีประสิทธิภาพมากนัก และไม่ให้ตัวประมาณค่าที่เที่ยงตรง แต่สามารถจะทำให้ส่วนที่คลาดเคลื่อน (Bias)นี้ มีค่าน้อยมากได้เมื่อเพิ่มขนาดของกลุ่มตัวอย่าง

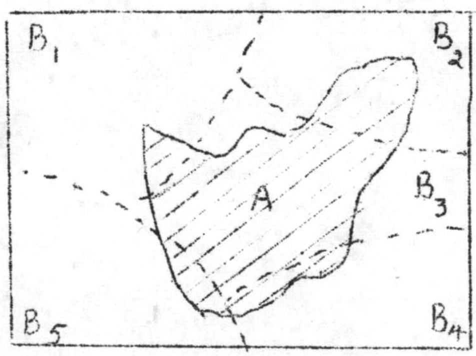
ข. วิธีของเบย์ (Bayesian method) ในปี ค.ศ. ๑๗๖๓ โทมัสเบย์ (Thomas Bayes) นักคณิตศาสตร์ชาวอังกฤษใช้วิธีนี้เป็นครั้งแรก วิธีนี้เรียก "Conditional probability "

สมมุติว่าเหตุการณ์ A จะเกิดขึ้นเมื่อมีเหตุการณ์ที่ไม่ขึ้นแก่กันและได้เกิดขึ้นแล้ว B_1 , B_2 ....B_n เกิดขึ้น โอกาสที่ A จะเกิดขึ้นหาได้จาก

$P(A) = P(B_1)\,P(A/B_1) + P(B_2)\,P(A/B_2) + \ldots + P(B_n)\,P(A/B_n)$

เพราะไม่ทราบว่าเหตุการณ์ใดใน B_1 , B_2 ,............B_n จะเกิดขึ้นจึงเป็นสมมุติฐานเป็นความน่าจะเป็นที่เกิดขึ้นก่อนการทดลอง ได้จากวัตถุประสงค์ที่กำหนดขึ้นก่อนหรือประสพการณ์สะท้อนให้เป็นสมมุติฐานที่เชื่อถือได้

รูปที่ ๑



$A = (B_1 \text{ และ } A) + (B_2 \text{ และ } A) + (B_3 \text{ และ } A) + (B_4 \text{ และ } A) + (B_5 \text{ และ } A)$

เมื่อทำการทดลองครั้งหนึ่งเกิดเหตุการณ์ A ขึ้นเราจะหาว่าความน่าจะเป็นที่จะเกิดตามสมมุติฐานจะเปลี่ยนแปลงเนื่องจากเหตุการณ์ A อย่างไร เราจะปรับความน่าจะเป็นเดิมคือ $P(B_1)$, $P(B_2)$ ....$P(B_n)$ เมื่อมีเหตุการณ์ใหม่เกิดขึ้น ผลการทดลองจะเปลี่ยนเป็น $P(B_1/A)$, $P(B_2/A)$......$P(B_n/A)$

Joint Probability ของ A และ B_1 เป็น

$P(AB_1) = P(A)\ P(B_1/A) = P(B_1)\ P(A/B_1)$, เอา P(A)

หารตลอด

$P(B_1/A) = P(B_1)\ P(A/B_1)/\ P(A)$ แทนค่า P(A)

ซึ่งเป็นความน่าจะเป็นทั้งหมดที่จะเกิดเหตุการณ์ A เราจะได้

$$P(B_1/A) = \frac{P(B_1)\ P(A/B_1)}{P(B_1)\ P(A/B_1) + P(B_2)\ P(A/B_2) + \ldots\ldots P(B_n)\ P(A/B_n)}$$

∴ เมื่อ $\{B_1, B_2 \ldots\ldots\ldots\ldots B_n\}$ เป็นกลุ่มของเหตุการณ์ โดยมี $P(B_i) \neq 0$

$i = 1, 2, \ldots\ldots\ldots\ldots n$ และ A เป็นเหตุการณ์ซึ่งมี $P(A) \neq 0$ สำหรับ

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

$$P(B_k/A) = \frac{P(B_k)\ P(A/B_k)}{\sum_{i=1}^{n} P(B_i)\ P(A/B_i)}$$

สูตรนี้เป็นทฤษฎีของเบย์ (Bayes's Theorem)

<u>ตัวอย่าง</u> เช่นโยนเหรียญ ๑๐๐ ครั้งได้หัว ๕๕ ครั้ง โอกาสที่จะได้หัวหรือก้อยเท่ากัน หรือโอกาสที่จะได้หัวในการโยนแต่ละครั้ง = $\frac{๑}{๒}$

สมมุติว่าโยน N ครั้ง ได้ความน่าจะเป็น = $p_1, p_2 \ldots\ldots\ldots p_n$ ความน่าจะเป็นแต่ละครั้งจะอยู่ระหว่าง ๐ และ ๑

ให้ B_i = เหตุการณ์ที่จะเลือกเหรียญที่มีความน่าจะเป็นที่จะได้หัว = p_i

และ A = เหตุการณ์ที่ได้หัว ๕๕ ครั้งจากการโยน ๑๐๐ ครั้ง

ให้ k = $1, 2, \ldots\ldots\ldots N$

$p_k = \frac{1}{2}$

จะหา $P(B_k / A)$ ซึ่งเป็น Conditional probability ซึ่งเลือกเหรียญที่มีความน่าจะเป็นเท่ากัน และได้หัว ๕๕ ครั้งในการโยน ๑๐๐ ครั้ง จะต้องทราบ $P(B_i)$ และ $P(A/B_i)$ เมื่อ $i = 1, 2, \ldots\ldots N$ จากกฎของไบโนเมียล (Binomail law)

$$P(A/B_i) = \binom{100}{55} (p_i)^{55} (1 - p_i)^{45}$$

$P(B_i)$ ไม่สามารถจะคำนวณได้ แต่ประมาณได้จะเห็นได้ว่า $P(A/B_k)$ หาได้จาก $P(B_1)$ ....$P(B_n)$ ถ้าไม่ทราบ $P(B_1)$ ......$P(B_n)$ จะหา $P(A/B_k)$ ไม่ได้ จึงต้องหา $P(A/B_k)$ โดยใช้สมมุติฐานว่า $P(B_1) = P(B_2) = \ldots = P(B_n) = \frac{1}{N}$ ดังนั้นความน่าจะเป็นทุกครั้ง $p_1, p_2 \ldots\ldots p_n$ เกือบเท่ากัน

$$\text{จาก } P(B_k/A) = \frac{P(B_k)\, P(A/B_k)}{\sum_{i=1}^{n} P(B_i)\, P(A/B_i)}$$

$$= \frac{\frac{1}{N}\binom{100}{55}(P_k)^{55}(1 - P_k)^{45}}{\frac{1}{N}\sum_{i=1}^{n}\binom{100}{55}(p_i)^{55}(1 - p_i)^{45}}$$

ถ้าให้ $N = 9$, $p_i = \frac{i}{10}$ เมื่อ $i = 1, 2, \ldots\ldots 9$ และ $K = 5$

$$\therefore P(B_5/A) = \frac{\binom{100}{55}(\frac{1}{2})^{55}(\frac{1}{2})^{45}}{\sum_{i=1}^{9}\binom{100}{55}\left(\frac{i}{10}\right)^{55}\left(\frac{10 - i}{10}\right)^{45}}$$

$$= \frac{0.048475}{0.097664} = 0.496$$

ความน่าจะเป็น $P(B_5) = \frac{1}{9}$ เรียก " Prior probability " ของเหตุการณ์ B_5 Conditional probability $P(B_5/A) = 0.496$ เรียก " Posterior probability " ของเหตุการณ์ B_5

ค. วิธีทำให้ความน่าจะเป็นมีค่าสูงสุด (Method of Maximum Likeli hood) วิธีการหาค่าพารามิเตอร์ที่จะทำให้ความน่าจะเป็นที่ได้ของผลการสังเกตจากกลุ่มตัวอย่างมีค่าสูงที่สุดเท่าที่จะเป็นไปได้เรียก "Maximum likelihood estimate"

ถ้า x เป็นตัวแปรสุ่มที่ไม่ต่อเนื่อง (Discrete random variable) ฟังก์ชั่นความน่าจะเป็นคือ $f(x : \theta)$ มีพารามิเตอร์ตัวเดียวคือ θ ความน่าจะเป็นของกลุ่มตัวอย่างสุ่ม $x_1, x_2, \ldots x_n$ คือ

$$L(x_1, x_2, \ldots x_n \mid \theta) = f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta)$$

และแทนความน่าจะเป็นที่ได้จากการสุ่มตัวอย่าง $x_1, x_2, \ldots x_n$ ดังนั้นค่าของพารามิเตอร์ θ ซึ่งจะทำให้ฟังก์ชั่นความน่าจะเป็นมีค่าสูงสุดคือ Maximum likelihood estimate

<u>ตัวอย่าง</u> เช่นการกระจายของไบโนเมียล สมมุติว่ามีคนออกเสียง ๑๐ คน พบว่า ๖ คนเสนอให้แก้ไขเกี่ยวกับภาษี ให้สัดส่วนจากกลุ่มตัวอย่าง $p = \frac{x}{n} = \frac{6}{10} = 0.6$ สัดส่วนตัวประมาณค่าของประชากร π จะมีค่าเท่าไร เมื่อให้ $p = 0.6$ โอกาสที่ $\pi = 0.3, 0.5$ และ 0.9 จะเป็นเท่าไร

ให้สูตรหา Binomial probability ของ x เมื่อกำหนด π

$$P(x ; \pi) = \binom{n}{x} \pi^{x} (1 - \pi)^{n - x}$$

ถ้า $\pi = 0.3$ ความน่าจะเป็น ที่จะได้คะแนนเสียง ๖ จาก ๑๐ คือ

$$P(6; \pi) = \binom{10}{6} (0.3)^6 (0.7)^4 = 0.0368$$

กล่าวได้ว่า ถ้า $\pi = 0.3$ โอกาสที่จะได้คะแนนเสียง ๖ ใน ๑๐ จะน้อยกว่า ๔ ใน ๑๐๐

ถ้ากลุ่มตัวอย่างได้จากประชากรที่มีพารามิเตอร์ $\pi = 0.5$ เราจะได้

$$P(6 ; 0.5) = \binom{10}{6} (0.5)^6 (0.5)^4 = 0.2051$$

ในทำนองเดียวกัน ถ้าเราพิจารณาทุกค่าของพารามิเตอร์ที่เป็นไปได้ ดังนั้นถ้า n = ขนาดของกลุ่มตัวอย่าง และ x = จำนวนครั้งที่ทำสำเร็จเป็นค่าคงที่สำหรับแต่ละกลุ่มตัวอย่าง ตัวแปรที่จะคำนวณคือพารามิเตอร์ตามทฤษฎี π ฟังก์ชั่นความน่าจะเป็นจะเขียนได้ดังนี้

$$L(\pi) = \binom{10}{6} \pi^6 (1 - \pi)^4$$

จะประเมินค่าฟังก์ชั่นนี้สำหรับค่า π อื่นที่เป็นไปได้ เราจะได้ความน่าจะเป็นดังนี้

π	$L(\pi)$	
0.0	0.0000	
0.1	0.0001	
0.2	0.0055	
0.3	0.0368	
0.4	0.1115	
0.5	0.2051	
0.6	0.2508	← ความน่าจะเป็นสูงสุด
0.7	0.2001	
0.8	0.0881	
0.9	0.0112	
1.0	0.0000	

ในระหว่างค่าต่าง ๆ ของ π ค่า $\pi = 0.6$ ทำให้ผลของกลุ่มตัวอย่างใกล้เคียงที่สุด จึงเป็น Maximum likelihood estimate

โดยทั่วไปการหาค่าซึ่งทำให้ฟังก์ชั่นความน่าจะเป็นมีค่าสูงสุดทำได้โดยดิฟเฟอเรนทิเอทเทียบกับพารามิเตอร์ θ แล้วให้เท่ากับ ๐ **และถ้าหาดิริเวทีฟที่สองของฟังก์ชั่นเทียบกับ** θ จะเป็น $-$ จึงจะเป็นจุดสูงสุดที่แท้จริง

$$L(\pi) = \binom{n}{x} \pi^{x} (1 - \pi)^{n-x}$$

$$\frac{\partial L(\pi)}{\partial \pi} = \binom{n}{x} \left[\pi^{x} (n - x)(1 - \pi)^{n-x-1} (-1) + x \pi^{x-1} (1 - \pi)^{n-x} \right] = 0$$

หารด้วย $\binom{n}{x} \pi^{x-1} (1 - \pi)^{n-x-1}$ เราจะได้

$$-\pi (n - x) + x (1 - \pi) = 0$$

$$\frac{x}{n} = p = \pi$$

ถ้าสุ่มตัวอย่างจากประชากรที่มีพารามิเตอร์ ๒ ตัวจะมีวิธีหาพารามิเตอร์ทั้งสองตัวพร้อมกันได้โดยดิฟเฟอเรนทิเอทเทียบกับพารามิเตอร์ทีละตัวแล้วให้เท่ากับศูนย์

วิธีทำให้ความน่าจะเป็นมีค่าสูงสุด นี้จะให้ตัวประมาณค่าที่มีคุณสมบัติดังนี้ Consistent, Efficient และ Sufficient แต่ทำให้เกิดความไม่เที่ยงตรงได้ นอกจากนี้ถ้าขนาดของกลุ่มตัวอย่างใหญ่มาก จะให้ค่าประมาณที่กระจายเป็นโค้งปกติ และมีค่าความแปรปวนต่ำที่สุดด้วย การประมาณค่าวิธีนี้จะใช้ได้เมื่อทราบลักษณะของฟังก์ชั่นความถี่แล้วเท่านั้น

ง. วิธีกำลังสองน้อยที่สุด (Method of Least Squares) เป็นวิธีประมาณค่าโดยใช้หลักว่า จะหาเส้นกราฟจากจุดต่าง ๆ ที่ได้จากการทดลองโดยให้ส่วนที่เบี่ยงเบนไปจากเส้นกราฟนี้มีค่าน้อยที่สุด เกณฑ์ที่จะทำให้ค่าผิดพลาดนี้น้อยที่สุดคือ ทำให้ผลบวกกำลังสองของค่าเบี่ยงเบนนี้มีค่าน้อยที่สุดเรียก "วิธีกำลังสองน้อยที่สุด (Method of Least Squares)"๑๕

ถ้าจากการทดลองได้ข้อมูลชุดหนึ่งมีค่าเป็น (x_1, y_1), (x_2, y_2)............., (x_n, y_n)

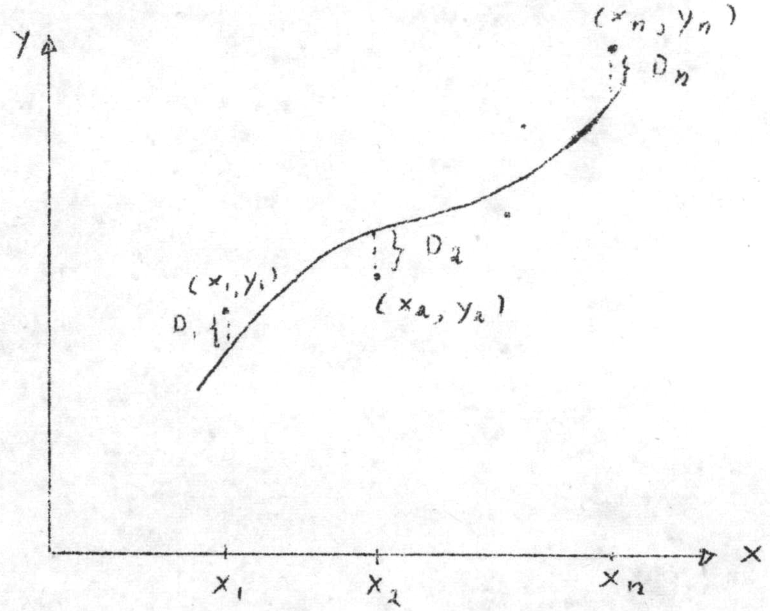
ให้ D_1 , D_2 , ...D_N เป็นค่าที่จุดต่าง ๆ เบี่ยงเบนไปจากเส้นกราฟตามรูป๑๖

---

15 Curtis f. Gerald, Applied Numcrical Analysis (Massachusetts: Addison-wesley publishing company, c 1970), p. 284.

16 Murray R. Spiegel, Theory and Problems of Statistics (New York : Schaum Publishing Company, c 1961), p. 219.

รูปที่ ๒



ถ้าเส้นกราฟนั้นมีคุณสมบัติว่า $D_1^2 + D_2^2 + \ldots\ldots\ldots + D_N^2$ มีค่าน้อยที่สุด เส้นกราฟนั้นจะเป็นเส้นที่เหมาะสมที่สุด D อาจมีค่าเป็นบวก, ศูนย์ หรือ ลบก็ได้ เส้นกราฟที่มีคุณสมบัตินี้จะเรียก "เส้นโค้งกำลังสองน้อยที่สุด (Least squares Curve)" เส้นตรงที่มีคุณสมบัตินี้เรียก "เส้นตรงกำลังสองน้อยที่สุด (Least Squares Line)" พาราโบล่าที่มีคุณสมบัตินี้เรียก "พาราโบล่ากำลังสองน้อยที่สุด (Least Squares Parabola)" เป็นต้น

การแสดงหลักการของวิธีกำลังสองน้อยที่สุดโดยใช้สมการเส้นตรงเป็นตัวอย่าง

ให้ Y_i แทนค่าที่ได้จากการทดลองและ y_i เป็นค่าที่คำนวณได้จากสมการ $y_i = ax_i + b + e_i$

$\therefore\ e_i = Y_i - y_i$ และ ให้ $SS = \sum_{i=1}^{N} e_i^2$

เราจะหาค่า SS ที่น้อยที่สุดได้โดยเลือกพารามิเตอร์ a และ b ที่เหมาะสม

สมมุติว่ามีข้อมูลชุดหนึ่ง[๑๗] ได้จากการตรวจสอบการจราจรผ่านอุโมงค์ลินคอน ซึ่งเชื่อมต่อระหว่างนิวยอร์คและนิวเจอร์ซี่ลอดใต้แม่น้ำฮัดสันดังตาราง

ตารางที่ ๓

ค่าเฉลี่ยความหนาแน่นของยานพาหนะต่อไมล์ x_i	ความเร็วเฉลี่ยเป็นไมล์ต่อชั่วโมง Y_i
๖๙	๑๖.๘
๘๘	๑๔.๘
๖๕	๑๙.๑
๘๔	๑๔.๙
๖๓	๒๑.๖
๕๓	๒๕.๗
๘๒	๑๖.๘
๕๓	๒๕.๑
๗๐	๑๙.๗
๓๔	๓๑.๘

ถ้าเลือกเส้นที่แสดงความสัมพันธ์ระหว่าง x_i และ y_i เป็นเส้นขนานกับแกน x โดยมีความชันเป็น ๐ สมการจะเป็น

$$y_i = b$$

17 Philip J. Mc Carthy, <u>Introduction to Statistical Reasoning</u> (New York : Mc Graw-Hill Book Company, Inc., c 1957), p. 343.

$$\therefore SS = \Sigma (Y_i - b)^2$$
$$= (Y_1 - b)^2 + (Y_2 - b)^2 + \dots\dots\dots\dots (Y_{10} - b)^2$$
$$= (16.8 - b)^2 + (14.8 - b)^2 + \dots\dots (31.8 - b)^2$$

ถ้าจะกำหนดค่า b เพียง ๕ ค่าคือ ๑๗, ๑๙, ๒๐.๖๓, ๒๒, ๒๕ เส้น y = ๑๗

$$SS = (๑๖.๘ - ๑๗)^{๒} + (๑๔.๘ - ๑๗)^{๒} + \dots\dots\dots (๓๑.๘ - ๑๗)^{๒}$$
$$= ๔๐๒.๕๓$$

เส้น y = ๑๙

$$SS = (๑๖.๘ - ๑๙)^{๒} + (๑๔.๘ - ๑๙)^{๒} + \dots\dots\dots\dots + (๓๑.๘ - ๑๙)^{๒}$$
$$= ๒๙๗.๓๓$$

เส้น y = ๒๐.๖๓

$$SS = (๑๖.๘ - ๒๐.๖๐)^{๒} + (๑๔.๘ - ๒๐.๖๓)^{๒} + \dots\dots\dots + (๓๑.๘ - ๒๐.๖๓)^{๒}$$
$$= ๒๗๐.๗๖$$

เส้น y = ๒๒

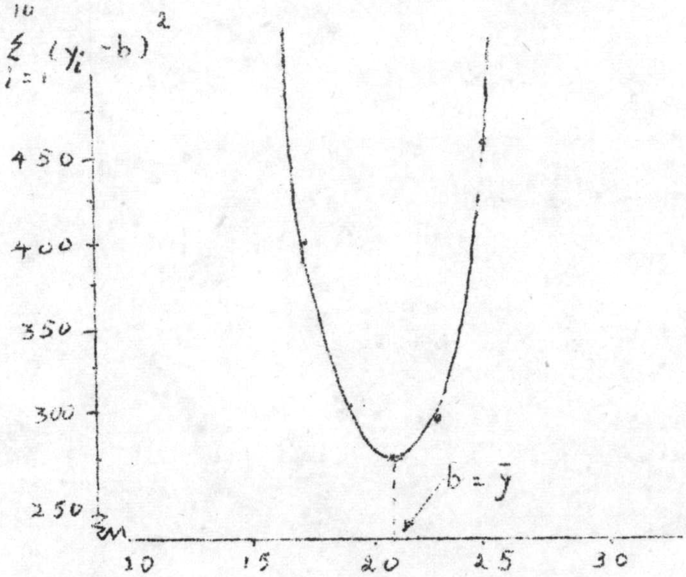
$$SS = (๑๖.๘ - ๒๒)^{๒} + (๑๔.๘ - ๒๒)^{๒} + \dots\dots\dots\dots + (๓๑.๘ - ๒๒)^{๒}$$
$$= ๒๘๙.๕๓$$

เส้น y = ๒๕

$$SS = (๑๖.๘ - ๒๕)^{๒} + (๑๔.๘ - ๒๕)^{๒} + \dots\dots\dots + (๓๑.๘ - ๒๕)^{๒}$$
$$= ๔๖๑.๗๓$$

นำค่า $\Sigma (Y_i - b)^2$ ทั้ง ๕ ค่ามาเขียนกราฟจะได้เส้นโค้งตามรูป ค่า b ที่ทำให้ SS มีค่าน้อยที่สุดคือ b = ๒๑ หลักการนี้สามารถจะใช้การคำนวณโดยคิฟเฟอเรนทิเอทเทียบกับพารามิเตอร์ ดังนี้

รูปที่ ๓



ค่าของ b

$$\frac{\partial SS}{\partial b} = \frac{\partial \left[\sum_{i=1}^{N} (y_i - b)^2 \right]}{\partial b} = 0$$

$$-2 \sum_{i=1}^{N} Y_i + 2Nb = 0$$

$$b = \frac{\sum_{i=1}^{N} Y_i}{N}$$

$$= \frac{206.3}{10}$$

$$= 20.63$$

คุณสมบัติของวิธีกำลังสองน้อยที่สุด

วิธีกำลังสองน้อยที่สุดจะใช้ได้ดีเมื่อทราบก่าฟังก์ชั่นความถี่ หรือฟังก์ชั่นความหนาแน่น (Density function) ของตัวแปรตาม แม้ว่าไม่ทราบลักษณะของฟังก์ชั่นวิธีกำลังสองน้อยที่สุดก็ยังสามารถให้ตัวประมาณค่าที่ดีได้ โดยมีคุณสมบัติที่เป็นไปตามเกณฑ์ของตัวประมาณค่าที่ดีดังนี้

๑. เป็นตัวประมาณค่าไม่เอนเอียง (Unbiased estimator)

๒. เป็นตัวประมาณค่าที่คงเส้นคงวา (Consistency)

๓. เป็นตัวประมาณค่าที่มีประสิทธิภาพโดยมีความแปรปวนน้อยที่สุด (Minimum Variance) ในหลายกรณี

วิธีกำลังสองน้อยที่สุดจะใช้ได้ดีโดยมีคุณสมบัติครบถ้วนตามนี้ในกรณีสำคัญ ๒ กรณี[๑๘]

ก. หาตัวประมาณค่าที่เที่ยงตรงและมีความแปรปวนน้อยที่สุดในสมการเส้นตรง

ข. เมื่อสมการมีค่าผิดพลาดซึ่งกระจายเป็นโค้งปกติและการประมาณค่าโดยวิธีกำลังสองน้อยที่สุดให้ผลเหมือนกับการประมาณค่าโดยทำให้ฟังก์ชั่นความน่าจะเป็นมีค่าสูงสุด

ในกรณีดังกล่าวตามทฤษฎีของเกาส์มาคอฟฟ์ (Gauss-Markoff Theorem)[๑๙] ถ้าสมการทั่วไปเป็น

$$Y = X\beta + e$$

e = เวคเตอร์ของตัวแปรสุ่มที่ไม่ขึ้นแก่กันหมายถึงส่วนที่เบี่ยงเบนไปจากค่าที่คาดหวัง (Expected value) e จะต้องมีคุณสมบัติดังนี้

๑. $E(e) = 0$

๒. $E(ee^1) = \sigma^2 I$

ข้อตกลงเบื้องต้นที่ไม่จำเป็นในการประมาณค่า b แต่ใช้ในการทดสอบสมมุติฐานในกรณีที่ใช้ t - test หรือ F - test เพื่อหาขอบเขตของความเชื่อมั่นคือ

๑. y ซึ่งแปรตาม x จะต้องกระจายเป็นโค้งปกติและไม่ขึ้นแก่กันโดยมีมัชฌิมเลขคณิต= μ ความแปรปวน = σ^2

---

18 Manrice G. Kendall and William R. Bnekland, <u>A Dictionary of Statistical Terms</u> (2 d ed.; New York : Hafner Publishing company, 1966), p. 160.

19 Alexander M. Mood and Franklin A. Graybill, <u>Introduction to the Theory of Statistics</u> (New York : Mc Graw Hill Book Company, Inc., c 1963), p. 349.

๒. e_i กระจายเป็นโค้งปกติและไม่ขึ้นแก่กันโดยมีมัชฌิมเลขคณิตเป็น ๐ ความแปรปวน = σ_ϵ^2 อาจเขียนสั้น ๆ ดังนี้ $E_i \sim N(0, \sigma_\epsilon^2)$

เส้นตรงกำลังสองน้อยที่สุดจะมีคุณสมบัติดังนี้[๒๐]

๑. ผลบวกของส่วนเบี่ยงเบนระหว่างค่าที่สังเกตได้กับค่าที่ประมาณเป็น ๐ ซึ่งเหมือนกับคุณสมบัติของมัชฌิมเลขคณิต

$$\Sigma \ (y - y_c) = 0$$

๒. ผลบวกกำลังสองของส่วนเบี่ยงเบนของค่าที่วัดได้จากค่าที่ประมาณจะน้อยที่สุด

$$\Sigma \quad (y - y_c)^2 \quad = \text{น้อยที่สุด}$$

๓. เส้นตรงนี้จะต้องผ่านตัวกลางของข้อมูลทั้งหมด $(\bar{x}, \bar{y})$

๔. ค่าที่ได้จากวิธีกำลังสองน้อยที่สุดจะเป็นค่าประมาณที่เที่ยงตรง (Unbiased estimates) ของพารามิเตอร์และจะมีความแปรปวนน้อยที่สุด

---

20 Ann Hughes, & Dennis Grawoig, Statistics : A Foundation for Analysis (Massachusetts: Addison-Wesley Publishing Company, c 1971), p. 322.

๒.๒ การหาสมการปกติ (Normal equation)

๒.๒.๑ สมการกำลังหนึ่ง (Lin ear Model)

ก. เมื่อมีตัวแปรเพียง ๒ ตัว

ให้ $(y_1, x_1), (y_2, x_2), \ldots\ldots\ldots (y_n, x_n)$ เป็นค่าที่ได้จากการสังเกตของตัวแปร y ที่สัมพันธ์กับค่า x ซึ่ง

$$E\left[\, y \mid x \,\right] = a + bx$$

ให้ Q = ผลบวกกำลังสองของผลต่างระหว่างค่าที่วัดได้และค่าที่คาดหวัง

$$Q = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

ในการหาค่า a และ b ให้พาเชียลดิริเวทีพเทียบกับ a และ b เท่ากับศูนย์จะได้ Normal equation สำหรับหาค่า a และ b และถ้า

$$\frac{\partial^2 Q}{\partial a^2} > 0$$

และถ้าดีเทอมิแนนท์ของ

$$\begin{vmatrix} \dfrac{\partial^2 Q}{\partial a^2} & \dfrac{\partial^2 Q}{\partial a \partial b} \\[2ex] \dfrac{\partial^2 Q}{\partial a \partial b} & \dfrac{\partial^2 Q}{\partial b^2} \end{vmatrix}$$

เป็นบวกที่ a, b แสดงว่า Q เป็นค่าต่ำสุด ดังนั้น

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\frac{\partial Q}{\partial b} = -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) = 0$$

จะได้สมการทั่วไปที่ใช้หา a และ b

$$n_a + b\sum x_i = \sum y_i \quad \ldots\ldots\ldots\ldots\ldots (1)$$

$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i \quad \ldots\ldots\ldots\ldots\ldots (2)$$

จาก (๑) $n a = \sum y_i - b\sum x_i$

$$a = \bar{y} - b\bar{x} \quad \ldots\ldots\ldots\ldots\ldots (3)$$

แทนค่า a ใน (๒)

$$(\bar{y} - b\bar{x})\sum x_i + b\sum x_i^2 = \sum x_i y_i$$

$$b(\sum x_i^2 - \bar{x}\sum x_i) = \sum x_i y_i - \bar{y}\sum x_i$$

$$b = \frac{\sum x_i y_i - \bar{y}\sum x_i}{\sum x_i^2 - \bar{x}\sum x_i}$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad \ldots\ldots\ldots\ldots\ldots (4)$$

$$\frac{\partial^2 Q}{\partial a^2} = 2n, \quad \frac{\partial^2 Q}{\partial a \partial b} = 2\sum x_i, \quad \frac{\partial^2 Q}{\partial b^2} = 2\sum x_i^2$$

แสดงว่า $\frac{\partial^2 Q}{\partial a^2} > 2$ และ

$$\begin{vmatrix} 2n & 2\sum x_i \\ 2\sum x_i & 2\sum x_i^2 \end{vmatrix} = 4n\sum x_i^2 - 4(\sum x_i)^2$$

$$= 4n\sum (x_i - \bar{x})^2$$

ฉะนั้นเส้นตรง $y = a + bx$ จะให้ค่า Q ที่น้อยที่สุด

ข. เมื่อมีตัวแปรมากกว่าสองตัวแปร

ถ้ามีตัวแปรจำนวนมากเกี่ยวข้องกัน เราประสงค์จะทำนายตัวแปรหนึ่งจากตัวแปรอิสสระอื่น ๆ โดยมีสมการทั่วไปเป็น

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots\cdots + \beta_k x_{ki} + e_i \quad \ldots (1)$$

ให้ขนาดของกลุ่มตัวอย่างเป็นN นั่นคือกลุ่มตัวอย่างหนึ่งมีค่า $(y_i, x_{1i}, x_{2i} \ldots\ldots x_{ki})$ N ชุด เมื่อ $i = 1, 2, \ldots\ldots N$ เราจะได้ N สมการ

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots\cdots + \beta_k x_{k1} + e_1 \\
y_2 &= \beta_0 + \beta x_{12} + \beta_2 x_{22} + \cdots\cdots + \beta_k x_{k2} + e_2 \quad \ldots\ldots (2) \\
y_3 &= \beta_0 + \beta_1 x_{13} + \beta_2 x_{23} + \cdots\cdots + \beta_k x_{k3} + e_3 \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 x_{1N} + \beta_3 x_{2N} + \cdots\cdots + \beta_k x_{kN} + e_N
\end{aligned}$$

โดยใช้เวกเตอร์และเมตริก เราสามารถจะเขียนสมการ N สมการนี้ง่าย ๆ ดังนี้ ใช้อักษรเส้นหนักแทนเวคเตอร์และเมตริก

ให้ $\mathbf{y}$ เป็นแถวของเวคเตอร์ที่มี N ค่า $y_1, y_2, y_3 \ldots\ldots y_N$

$\mathbf{e}$ " " N " $e_1, e_2, e_3 \ldots\ldots e_N$

β " " k+1 " $\beta_0, \beta_1, \beta_2 \ldots\ldots \beta_k$

$\mathbf{x}$ เป็นเมตริก $N(k + 1)$ ซึ่งมีค่าในแถวที่ i และสดมภ์ที่ j เป็น x_{ij} ดังนั้น

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & x_{32} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & x_{33} & \cdots & x_{k3} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1N} & x_{2N} & x_{3N} & \cdots & x_{kN} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \mathbf{e} = n \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_N \end{pmatrix}$$

โดยใช้สัญญลักษณ์ใหม่ชุดของสมการ N สมการจะเขียนได้ดังนี้

$$\mathbf{y} = \mathbf{x}\,\beta + \mathbf{e} \quad \ldots\ldots\ldots\ldots\ldots (3)$$

สมการ (๒) และ (๓) เหมือนกันสามารถคูณตามกฎการคูณของเมตริก ข้อตกลงเบื้องต้นที่ยังใช้อยู่คือ ค่าผิดพลาด e เป็นเทอมที่สุ่มและไม่ขึ้นแก่กัน มีการกระจายเป็นโค้งปกติมีมัชฌิม

เลขคณิตเป็น ๐ และความแปรปวนเป็น δ^2_e

ในการหาค่า $\beta_0, \beta_1, \beta_2 \ldots\ldots\ldots \beta_k$ ทำได้โดยประมาณค่าของเวคเตอร์ β สมมุติให้ตัวประมาณค่าของ β เป็น b ซึ่งเป็นเวกเตอร์ที่มี k + 1 ค่า $b_0, b_1, b_2, \ldots\ldots\ldots b_k$ สมการที่ใช้ประมาณค่าคือ

$$y = xb + e \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

ตามเกณฑ์ของวิธีกำลังสองน้อยที่สุดจะต้องทำให้ผลบวกกำลังสองของค่าผิดพลาด ($\sum_{i=1}^{N} e_i^2$) มีค่าน้อยที่สุด ผลบวกกำลังสองของค่าผิดพลาดสามารถแสดงได้โดยเมตริก $e'e$

$$ee' = \begin{bmatrix} e_1, e_2 \ldots\ldots\ldots e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix} = \sum_{i=1}^{N} e_i^2$$

ตามเกณฑ์เราจะต้องเลือกค่า b ที่จะทำให้ $e'e$ มีค่าน้อยที่สุด

จากสมการ (๔) $e = y - xb$

นั่นคือ $\sum_{i=1}^{n} e_i^2 = (y - xb)'(y - xb)$, $S = x'x$

$$\frac{\partial (e'e)}{\partial b_t} = 0 \text{ , } t = 0, 1, 2, \ldots\ldots\ldots k$$

$$\frac{\partial}{\partial b_t}(y - xb)'(y - xb) = 2at + \sum_{\substack{i=1 \\ i \neq t}}^{k} b_i S_{it} + \sum_{\substack{j=1 \\ j \neq t}}^{k} b_j S_{jt} + 2\beta_t S_{tt}$$

$$= 2a_t + 2\sum_{i=1}^{k} b_i S_{it}$$

ฉะนั้น $\frac{\partial (y - xb)' (y - bx)}{\partial b_1} = 2a_1 + 2 \sum_i b_i S_{i1} = 0$

$\frac{\partial (y - xb)' (y - bx)}{\partial b_2} = 2a_2 + 2 \sum_i b_i S_{i2} = 0$

$\vdots$

$\frac{\partial (y - xb)' (y - bx)}{\partial b_k} = 2a_k + 2 \sum_i b_i S_{ik} = 0$

เราจะได้ค่า $2A' + 2Sb = 0$

หรือ $-2x'y + 2x'xb = 0$

$2x'xb - 2x'y = 0$

ฉะนั้นตัวประมาณค่ากำลังสองน้อยที่สุดของ β คือ

$$b = S^{-1} x'y$$

๒.๒.๒ สมการกำลังมากกว่าหนึ่ง (Nonlinear Model)

สมการที่ไม่ใช่เส้นตรงมีหลายแบบ

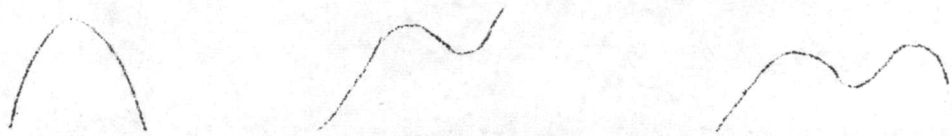
ก. โพลีโนเมียล (Polynomial) ซึ่งมีสมการทั่วไปดังนี้

$$y(x) = b_0 + b_1 x_1 + b_2 x_2^2 + b_3 x_3^3 + \ldots\ldots\ldots + b_k x^k \ \ldots. (1)$$

ถ้าสัมประสิทธิ์ทั้งหมดตั้งแต่ $b_2, b_3, \ldots\ldots, b_k$ เป็น ๐ สมการจะเป็นสมการเส้นตรงทำนองเดียวกันถ้าสัมประสิทธิ์ทั้งหมดเป็น ๐ ยกเว้น b_0, b_1, b_2 จะได้ โพลีโนเมียลกำลังสองลักษณะของเส้นกราฟจะโค้งครั้งหนึ่งและมีชื่อเรียกเฉพาะว่า " Parabola " ถ้าเป็น โพลีโนเมียล กำลัง ๓ จะโค้ง ๒ ครั้ง และกำลัง ๔ จะโค้ง ๓ ครั้ง ตามรูปทิศทางของเส้นโค้งขึ้นกับเครื่องหมายของค่าสัมประสิทธิ์ และจำนวนโค้งจะน้อยกว่ากำลังของอยู่ ๑ เสมอ[๒๑]

---

[21] Hubert M. Blalack, Jr., Social Statistics (London: Mc Graw-Hill Book Company, c 1960), p. 352.



$y = a+b_1x+b_2x^2$ $\quad$ $y = a+b_1x+b_2x^2+b_3x^3$ $\quad$ $y=a+b_1x+b_2x^2+b_3x^3+b_4x^4$

สมการที่ใช้ทั่วไปซึ่งสามารถแทนได้ถึง n สมการคือ

$$y(x_i) = b_o + b_1x_i + b_2x_i^2 + \ldots\ldots\ldots + b_kx_i^k \quad \ldots\ldots\ldots (2)$$

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

Q = ผลบวกกำลังสองของส่วนที่เบี่ยงเบนจากค่าที่คาดหวังตามแนวดิ่ง

$$Q = \sum_{i=1}^{n} \left[y_i - y(x_i) \right]^2$$

Q จะมีค่าน้อยที่สุดเมื่อสัมประสิทธิ์ $b_o, b_1, b_2, \ldots\ldots\ldots b_k$ เป็นไปตามข้อแม้ต่อไปนี้

$$\frac{\partial Q}{\partial b_0} = 0, \quad \frac{\partial Q}{\partial b_1} = 0, \quad \frac{\partial Q}{\partial b_2} = 0, \quad \ldots \quad \frac{\partial Q}{\partial b_k} = 0$$

จะได้สมการ $k + 1$ สมการมีตัวที่ไม่ทราบค่าเป็น $b_o, b_1 \ldots\ldots\ldots b_k$ สมการเหล่านี้เรียก " Normal equation " จะใช้สมการกำลังสามเป็นตัวอย่างในการหาสมการทั่วไป

$$y(x) = b_o \; b_1x + b_2x^2 + b_3x^3 \quad \ldots\ldots\ldots\ldots (3)$$

เราได้

$$y =(x_i) = b_o + b_1x_i + b_2x_i^2 + b_3x_i^3$$

$$\therefore \quad Q = \sum_{i=1}^{n} (y_i - b_o - b_1 x_i - b_2x_i^2 - b_3x_i^3)$$

ดีพเฟอเรนทิเอทเทียบกับ b_o, b_1 และ b_2 ให้เท่ากับ o

$$\frac{\partial Q}{\partial b_0} = -2 \sum (y_i - b_o - b_1x_i - b_2x_i^2 - b_3x_i^3) = 0$$

$$\frac{\partial Q}{\partial b_1} = -2 \sum x_i (y_i - b_o - b_1x_i - b_2x_i^2 - b_3x_i^3) = 0$$

$$\frac{\partial Q}{\partial b_2} = -2 \sum x_i^2 (y_i - b_o - b_1x_i - b_2x_i^2 - b_3x_i{}^3) = 0$$

$$\frac{\partial Q}{\partial b_3} = -2 \sum x_i^3 (y_i - b_o - b_1 x_i - b_2 x_i^2 - b_3 x_i^3) = 0$$

จะได้สมการเส้นตรงโดยมีตัวไม่ทราบค่า ๓ ตัว คือ b_o, b_1, b_2 เราจะได้ Normal equation ดังนี้

$$\begin{aligned}
b_o n + b_1 \sum x_i + b_2 \sum x_i^2 + b_3 \sum x_i^3 &= \sum y_i \\
b_o \sum x_i + b_1 \sum x_i^2 + b_2 \sum x_i^3 + b_3 \sum x_i^4 &= \sum x_i y_i \quad \ldots(5) \\
b_o \sum x_i^2 + b_1 \sum x_i^3 + b_2 \sum x_i^4 + b_3 \sum x_i^5 &= \sum x_i^2 y_i \\
b_o \sum x_i^3 + b_1 \sum x_i^4 + b_2 \sum x_i^5 + b_3 \sum x_i^6 &= \sum x_i^3 y_i
\end{aligned}$$

ข. ถ้าลักษณะของข้อมูลอยู่ในรูปที่แปลงเป็นสมการเส้นตรงได้ เช่นอยู่ในของกำลัง, เศษส่วนหรือรูต สามารถจะเปลี่ยนเป็นสมการเส้นตรงโดยแปลงให้อยู่ในรูปของ log เช่น

$$Y_c = AB^x$$

ผลบวกกำลังสองที่ต้องทำให้น้อยที่สุดคือ $\sum (y - A B^x)^2$ จะได้ Normal equation ที่คำนวณได้ยาก อาจทำให้ง่ายโดยใส่ log ทั้งสองข้าง

$$\log Y_c = \log A + x \log B$$

จะได้สมการซึ่งอยู่ในรูปสมการเส้นตรงซึ่งคำนวณง่ายกว่ามาก การแปลงเช่นนี้ มิได้ใช้ค่าจริง ทำให้เส้นตรงกำลังสองน้อยที่สุดมีคุณสมบัติเปลี่ยนไปดังนี้

๑. เส้นตรงกำลังสองน้อยที่สุดจะเป็นการทำให้กำลังของของ log y ต่างจาก log y_c ที่คำนวณได้มีค่าน้อยที่สุดไม่ใช่ทำให้ $(y - y_c)^2$ มีค่าน้อยที่สุด

๒. ในการกำหนดค่าที่กระจายออกไปจะต้องบวกหรือลบจากค่าเบี่ยงเบนมาตรฐานของสมการถดถอยจากค่า y เมื่อทั้งสองอยู่ในค่าของ log ผลลัพธ์ที่ได้จึงจะนำมาเปลี่ยนเป็นค่า y ตามเดิม

ถ้าพลอตค่า x และ y ในกระดาษที่เป็นเสกล log ทั้งแกนนอนและแกนตั้ง จะได้สมการเส้นตรง สมการที่แปลงแล้วควรจะเป็นดังนี้

$$(\log y)_c = \log a + b \log x$$

ตัวคงที่ log a และ b จะหาได้จาก Normal equation ของสมการเส้นตรงคือ $\Sigma \log y = n \log a + b \log x$

และ $\Sigma (\log x \cdot \log y) = \log a \, \Sigma \log x + b \, \Sigma (\log x)^2$

ส่วนเบี่ยงเบนทั้งหมดได้จาก

$$\Sigma(\log y)^2 = \Sigma (\log y)^2 - (\overline{\log y}) \, \Sigma \log y$$

เมื่อ $(\overline{\log y}) = (\Sigma \log y)/n$

ผลบวกกำลังสองของส่วนที่เกิดจากการเปลี่ยนแปลงที่ทราบหาได้จาก

$$\Sigma (\log y)_c^2 = \log a \, \Sigma \log y + b \, \Sigma (\log x . \log y) - (\overline{\log y}) \, \Sigma \log y$$

ผลบวกกำลังสองของส่วนที่ผิดพลาดได้จากผลต่างระหว่างผลบวกกำลังสองของทั้งหมดกับผลบวกกำลังสองของส่วนที่เกิดจากการเปลี่ยนแปลง

๓. <u>การแปลความหมายทางเรขาคณิต</u>

๓.๑ สมการกำลังหนึ่ง (Linear Model) ถ้าสมการเส้นตรงมีสองตัวแปร โดยมีสมการทั่วไปทางคณิตศาสตร์ดังนี้

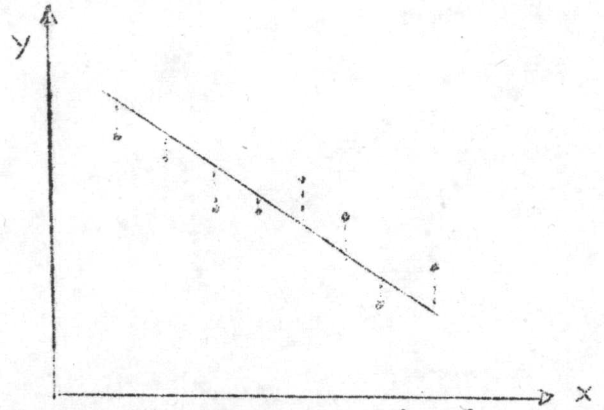
$$y = \theta_1 + \theta_2 x + \epsilon$$

เราสามารถจะถ่ายทอดเป็นกราฟให้เกิดความเข้าใจหลักการของวิธีกำลังสองน้อยที่สุดได้ดังนี้[๒๒]

---

22 Bernard. Ostle, Statistics in Research (Iowa: Oxford & IBH Publishing Co., 1966), p. 162.

รูปที่ ๔



เส้นตรงตามรูปได้จากการให้ผลบวกกำลังสองของส่วนเบี่ยงเบนตามแนวดิ่งมีค่าน้อยที่สุด สมการถดถอยของเส้นตรงนี้จะเขียนได้ดังนี้

$$\hat{y} = \hat{\theta}_1 + \hat{\theta}_2 x$$

เมื่อ $\hat{\theta}_j$ เป็นค่าประมาณของ θ_j $(j = 1, 2)$ และ $\hat{y}$ เป็นค่าประมาณของ y

สมการทั่วไปของสมการเส้นตรงหลายตัวแปรจะเขียนได้ดังนี้ [๒๓]

$$y = f\,(\xi\,,\,\theta\,) + \epsilon$$

$$= \theta_1 x_1 + \theta_2 x_2 + \ldots\ldots\ldots\ldots " \theta_p x_p + \epsilon$$

เมื่อ x_i เป็นฟังค์ชั่นของ ξ ถ้าค่าที่วัดได้ y_u มีค่าผิดพลาด ϵ_u เมื่อให้ x_i เป็น x_{1u}, x_{2u}, x_{3u} ....x_{pu}, $u = 1, 2, \ldots\ldots n$ เรา สามารถจะเขียนสมการเป็นแบบเวคเตอร์คือ $y = x\theta + \epsilon$

เมื่อ

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad x = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ x_{12} & x_{22} & \cdots & x_{p2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

---

23 Draper N.R. & Smith H., Applied Regression Analysis (New York : John Wiley & Sons, Inc., 1966), p. 285.

เราสามารถจะหาค่าของ " β_0 " ตามสมการนี้โดยใช้ $x_{1j} = 1$, $j = 1, 2, \ldots\ldots n$
ผลบวกกำลังสองของส่วนที่เบี่ยงเบนสามารถเขียนได้ดังนี้

$$S(\theta) = \sum_{j=1}^{n} \left[y_j - f(\xi_j, \theta) \right]^2 \qquad \ldots\ldots\ldots (1)$$

$$= \sum_{j=1}^{n} \left[y_j - \sum_{i=1}^{p} \theta_i \; x_{ij} \right]^2$$

$$= (y - x\theta)' (y - x\theta)$$

$$= y'y - 2\theta'x'y + \theta'x'x\theta \qquad \ldots\ldots\ldots (2)$$

ดีพเฟอเรนทิเอทสมการนี้เทียบกับ θ แล้วให้เท่ากับ ๐ และใช้ $\hat{\theta}$ แทน θ เราจะได้สมการปกติดังนี้

$$x' x\hat{\theta} = x'y \qquad \ldots\ldots\ldots (3)$$

ถ้า $x'x$ เป็น Nonsigular คือเมื่อ $|x'x| \neq 0$ จะได้ว่า

$$\hat{\theta} = (x'x)^{-1} \; x'y$$

จาก (๒)

$$S(\hat{\theta}) = y'y - 2\hat{\theta}'x'y + \hat{\theta}'x'x\hat{\theta}$$

$$= y'y - \hat{\theta}'x'y + \hat{\theta}'(x'y - x'x\hat{\theta})$$

จาก (๓)

$$x'y - x'x\hat{\theta} = 0$$

$$S(\hat{\theta}) = y'y - \hat{\theta}'x'y$$

ดังนั้น $\hat{\theta}$ เป็นไปตาม Normal equation $S(\hat{\theta})$ จะเป็นค่าน้อยที่สุดของ $S(\theta)$

ถ้าค่าผิดพลาด ϵ_j ไม่ขึ้นแก่กันและมีการกระจายดังนี้ $N(0, \sigma^2)$ นั่นคือถ้า $\epsilon \sim N(0, I\sigma^2)$ ดังนั้นจะสามารถแสดงว่า ถ้าสมการถูกต้องจะได้ว่า

1. $\hat{\theta} \sim N\left[\theta, (x'x)^{-1}\sigma^2\right]$
2. $S(\hat{\theta}) \sim \sigma^2 \chi^2_{n-p}$
3. $S(\hat{\theta}) - S(\hat{\theta}) \sim \sigma^2 \chi^2_p$

4. S (θ) - S ($\hat{\theta}$) และ S ($\hat{\theta}$) จะกระจายอย่างอิสสระ

ดังนั้น

$$\frac{[S(\theta) - S(\hat{\theta})] \;/p}{S(\hat{\theta}) \;/n - p} \sim F\ (p, n-p)$$

สามารถจะตรวจสอบ S(θ)ซึ่งเป็นค่าคงที่ได้ ๒ วิธี คือพิจารณาในทางกลุ่มตัวอย่าง (Sample space) และพิจารณาในแง่ของพารามิเตอร์ (Parameter Space) แต่การพิจารณาในด้านกลุ่มตัวอย่างจะทำให้เข้าใจดีกว่า จึงใช้ Sample space

สำหรับ n - มิติ เวคเตอร์ที่ได้จากการสังเกต $y = (y_1, y_2 \ldots\ldots y_n)$ เขียนในรูปเวคเตอร์ $\overrightarrow{oy}$ จากจุดกำเนิด o ถึงจุด y โดยมีระยะตามแกนเป็น ($y_1, y_2 \ldots\ldots y_n$) เมตริก x มี p สดมภ์แต่ละสดมภ์มี n ค่า ในสดมภ์ที่ j มีระยะตามแกนเป็น ($x_{j1}, x_{j2}, \ldots\ldots\ldots x_{jn}$) ของจุด x_j x เขียนเป็นเวคเตอร์ $\overrightarrow{ox}$, p เวคเตอร์คือ $\overrightarrow{ox_1}$, $\overrightarrow{ox_2} \ldots\ldots \overrightarrow{ox_p}$ เป็นส่วนหนึ่งของ p มิติ เรียก " Estimation space " ซึ่งอยู่ใน sample space จุดใด ๆ ใน subspace นี้ สามารถแทนได้โดยจุดปลายของเวคเตอร์ที่เป็นเส้นตรงของสดมภ์ x เช่น $x\theta$ เมื่อ $\theta = (\theta_1, \theta_2, \ldots\ldots\ldots \theta p)$ เป็น (px 1) เวคเตอร์

สมมุติว่าเวคเตอร์ $x\theta$ แทนจุด T ดังนั้น $y\,T^2$ เขียนได้ดังนี้

$$(y - x\theta)'\ (y - x\theta) = S\ (\theta)$$

S(θ) แทน กำลังสองของระยะทางจาก y ถึงจุดที่จะประมาณค่า ทำให้ S (θ) มีค่าน้อยที่สุดโดยเทียบกับ θ หาค่า θ ให้ = $\hat{\theta}$ ซึ่งทำให้จุด p (เขียนเวคเตอร์ $\hat{y} = x\hat{\theta}$) ในการประมาณค่าใกล้จุด y ที่สุด ตามเรขาคณิต p จะเป็นโคนของเส้นตั้งฉากจาก y ไปยัง estimation space นั่นคือโคนเส้นที่ผ่าน y และตั้งฉากกับทุกสดมภ์ของเมตริก x ในเทอมของเวคเตอร์จากจุดกำเนิดจะเขียนได้ดังนี้

$$y = \hat{y} + (y - \hat{y})$$
$$= \hat{y} + e$$

e เป็นเวคเตอร์ของ residuals y สามารถแบ่งเป็นเวคเตอร์ประกอบสองเวคเตอร์ซึ่งตั้งฉากกัน คือ $\hat{y}$ ซึ่งอยู่ใน estimation space และ $y - \hat{y} = e$ เป็นเวคเตอร์ของ residuals ซึ่งอยู่ใน Error space จะเขียนได้เป็น $(n - p)$ - มิติเป็น subspace ใน n - มิติ estimation space และ error space เป็นออร์ทอโกนัล[๒๔] จะแสดงให้เห็นได้ว่า $\hat{y}$ และ e เป็นออร์ทอโกนัล ดังนั้น

$$\hat{y}' e = (x\hat{\theta})' (y - x\hat{\theta})$$

$$= \hat{\theta}' x' (y - x\hat{\theta})$$

$$\hat{y}' e = \hat{\theta}'(x'y - x' x \hat{\theta})$$

ตาม (๓) $\quad x'y - x'x\hat{\theta} = 0$

$$\hat{y}'e = 0$$

$\hat{\theta}$ เป็นไปตาม Normal equation เวคเตอร์ e เป็นเวคเตอร์ $\overrightarrow{oR}$ จากจุดกำเนิด o โดยมีความยาว $oR = yp$ และ cR ขนานกับ py ถ้า T เป็นจุดทั่วไปของ estimation space และ yp เป็นออร์ทอโกนัล ดังนั้น

$$yT^2 = yp^2 + pT^2$$

หรือ

$$S(\theta) = S(\hat{\theta}) + pT^2$$

ดังนั้นโครงร่างเมื่อ $S(\theta)$ เป็นตัวคงที่จะต้องอยู่ในลักษณะที่

$$pT^2 = S(\theta) - S(\hat{\theta}) = \text{ตัวคงที่}$$

ใน sample Space โครงร่างที่อธิบาย $S(\theta)$ คงที่ประกอบด้วยทุกจุดของ T ซึ่ง pT^2 มีค่าเท่ากับตัวคงที่ นั่นคือจุดที่อยู่ใน estimation space ในรูป $x\theta$ ซึ่งอยู่ใน p - มิติ ทรงกลมที่จุดศูนย์กลางอยู่ที่ p อธิบายโดย $x\hat{\theta}$ รัศมีของทรงกลมเป็น $\left[S(\theta) - S(\hat{\theta})\right]^{\frac{1}{2}}$ เราหาได้แล้วว่า

---

[๒๔] เวคเตอร์ A และ B จะเป็นออทอโกนัล ถ้าผลคูณสเกลล่าเท่ากับศูนย์นั่นคือ $A'B = 0$

$$\frac{[\, S(\theta) - S(\hat{\theta}) \,]/p}{S(\hat{\theta})/(n-p)} \sim F(p, n-p)$$

เราสามารถจะหาขอบเขตของ $(1 - \alpha)$ % ของขอบเขตความเชื่อมั่นสำหรับจุด $x\theta$ ซึ่งได้จากค่าจริงของ θ (ซึ่งไม่ทราบค่า) โดย

$$\frac{[\, S(\theta) - S(\hat{\theta}) \,]/p}{S(\hat{\theta})/(n-p)} = F(p, n-p, 1 - \alpha)$$

$$S(\theta) = S(\hat{\theta}) \left[1 + \frac{p}{n-p} F(p, n-p, 1-\alpha) \right]$$

$S(\hat{\theta})(1 + q^2)$ เป็นค่าของ $S(\theta)$ ซึ่งมีค่ามากกว่าค่าต่ำสุด $S(\hat{\theta})$ ขอบเขตของความเชื่อมั่นประกอบด้วยส่วนในทรงกลมใน estimation space มีจุดศูนย์กลางอยู่ที่ P และมีรัศมี

$$[\, S(\theta) - S(\hat{\theta}) \,]^{\frac{1}{2}} = \left[S(\hat{\theta}) \frac{n}{n-p} F(p, n-p, 1 - \alpha) \right]$$

<u>ตัวอย่างเมื่อ $n = 3,\ p = 2$</u>[๒๕]

การแสดงหลักการนี้โดยใช้แผนภาพ เราจะสมมุติให้ $n = 3$ เมื่อ $n > 3$ ไม่สามารถจะเขียนเป็นภาพได้ แต่สามารถเข้าใจได้

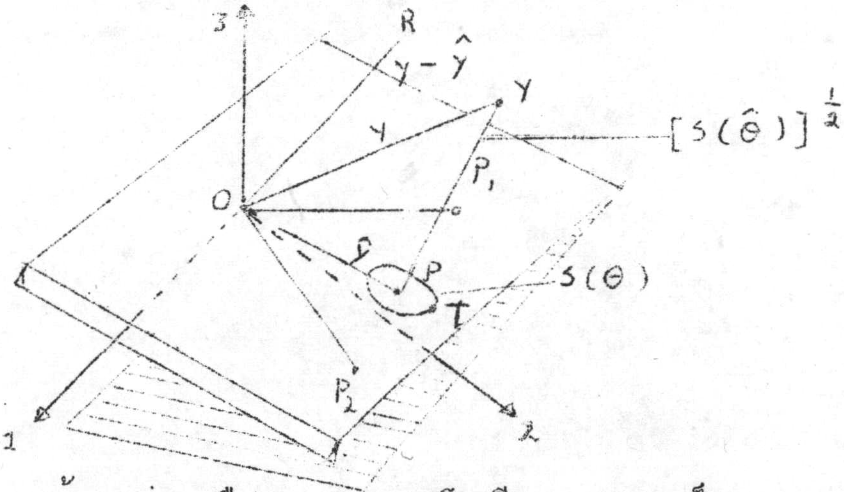
ตามรูปที่ ๕ ให้แกน ๑, ๒, ๓ เป็นส่วนประกอบของ $(y_1\ y_2,\ y_3)$ ของเวคเตอร์ y' สมมุติให้ $p = 2$ พารามิเตอร์ θ_1 และ θ_2 ดังนั้น x จะเป็นเมตริก 3×2

---

[25] I bid., p. 289.

$$X = \begin{vmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{vmatrix}$$

รูปที่ ๕



สดมภ์ของ x แทนด้วยจุด ๒ จุดคือ P_1 และ P_2 โดยมีระยะตามแกนเป็น (x_{11}, x_{12}, x_{13}) และ (x_{21}, x_{22}, x_{23}) และเวคเตอร์ $\overrightarrow{OP_1}$ และ $\overrightarrow{OP_2}$ แทนระนาบซึ่งแทนด้วย estimation space ๒ มิติ เวคเตอร์ $\hat{y} = x\,\hat{\theta}$ อยู่บนระนาบนี้จุด y อยู่เหนือระนาบนี้ และตั้งฉากกับ yP จาก y ไปยัง O P_1P_2 ที่จุด P ดังนั้น **yp** เป็นระยะที่สั้นที่สุดจาก y ถึงจุดใด ๆ ใน estimation space

P แทนโดย $\hat{y} = x\,\hat{\theta}$

$$S(\hat{\theta}) = yP^2$$

$$O\,y^2 = y'\,y\ ,$$

$$y'\,y = \hat{\theta}\,x'\,y + (y'\,y - \hat{\theta}'x'y)$$

$$= \hat{\theta}'\,x'\,y + S(\hat{\theta})$$

ซึ่งตรงกับ $Oy^2 = OP^2 + y\,P^2$

ถ้าเราเขียน OR ผ่าน O ซึ่งความยาว $OR^2 = S(\hat{\theta})$ และขนานกับ P y ดังนั้น $\overrightarrow{OR}$ แทนเวคเตอร์ของ residuals $e = y - \hat{y}$ เวคเตอร์ $\overrightarrow{OP}$ เป็น $\hat{y}$

เราได้สมการเวคเตอร์ดังนี้

$$\overrightarrow{Oy} = \overrightarrow{OP} + \overrightarrow{OR}$$

หรือ $$y = \hat{y} + (y - \hat{y})$$

ตัวคงที่ S (θ) แทนโดยทรงกลม p - มิติใน estimation space ดังนั้นจะมีวงกลมบนระนาบ O P_1 P_2 , T เป็นจุด X θ บนระนาบ S (θ) เป็นตัวคงที่หมายถึง YT^2 ดังนั้น

$$PT^2 = yT^2 - yP^2 = \text{ค่าคงที่}$$

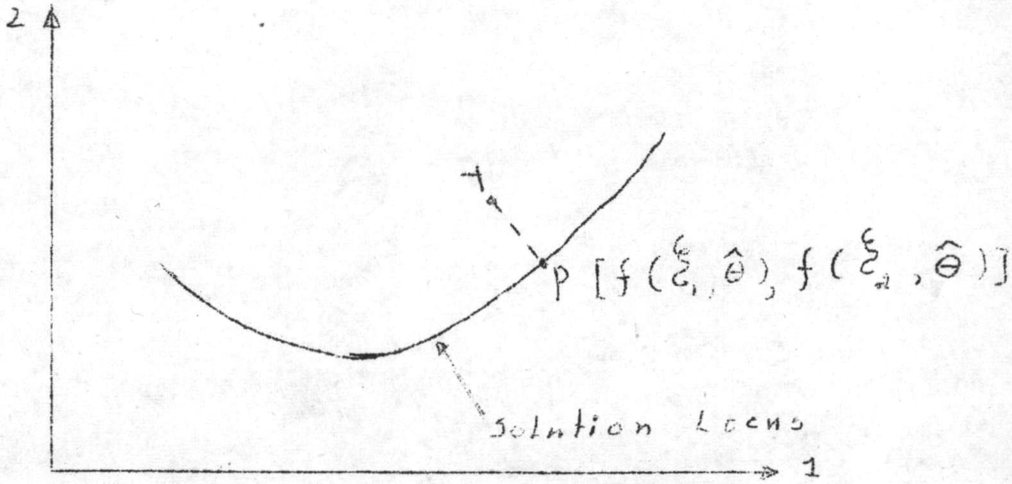
เราจะได้วงกลมรอบ P ซึ่งแสดงในรูปเป็นขอบเขตของความเชื่อมั่น ๑๐๐ (๑ − α)% สำหรับจุดจริง X θ มีรัศมีซึ่งหาได้จาก $[2S(\hat{\theta}) F(2, 1, 1-\alpha)]^{\frac{1}{2}}$ ได้จากการแทนค่า n = 2, p = 3 ในสูตรทั่วไป

๓.๒ สมการกำลังมากกว่าหนึ่ง (Nonlinear Model)

เมื่อรูปของสมการไม่ใช่สมการเส้นตรง ไม่สามารถจะเขียนเป็น x - เมตริกได้ แต่ยังมี**estimation space**ซึ่งเรียกว่า "ทางเดินของผลลัพธ์ (Solution locus)" ซึ่งประกอบด้วยจุดทั้งหมดโดยมีระยะตามแกนเป็น $[f(\xi_1, \theta), f(\xi_2, \theta) \ldots\ldots f(\xi_n, \theta)]$ ดังนั้นผลบวกกำลังสองของฟังก์ชั่น S (θ) ยังคงแทนกำลังสองของระยะทางจากจุด (y_1, y_2.........y_n) ไปยังจุดของestimation space การทำให้ S (θ) มีค่าน้อยที่สุดยังคงใช้หลักทางเรขาคณิตในการหาจุด P ใน Sample space ซึ่งใกล้จุด y ที่สุด

ตัวอย่างของสมการที่ไม่ใช่สมการเส้นตรงซึ่งง่ายที่สุดคือ n = 2 การวัดy_1 และy_2 ที่ $\xi = \xi_1$ และ $\xi = \xi_2$ ตามลำดับ พารามิเตอร์ θ ดังแสดงในรูปที่ ๖ estimation space ประกอบด้วยเส้นโค้งซึ่งประกอบด้วยจุด $[f(\xi_1, \theta), f(\xi_2, \theta)]$ เมื่อ θ เปลี่ยน ξ_1, ξ_2 คงที่ y มีระยะตามแกน x เป็น (y_1,y_2) และเป็นจุดของ estimation space ที่ใกล้ค่า y ที่สุด

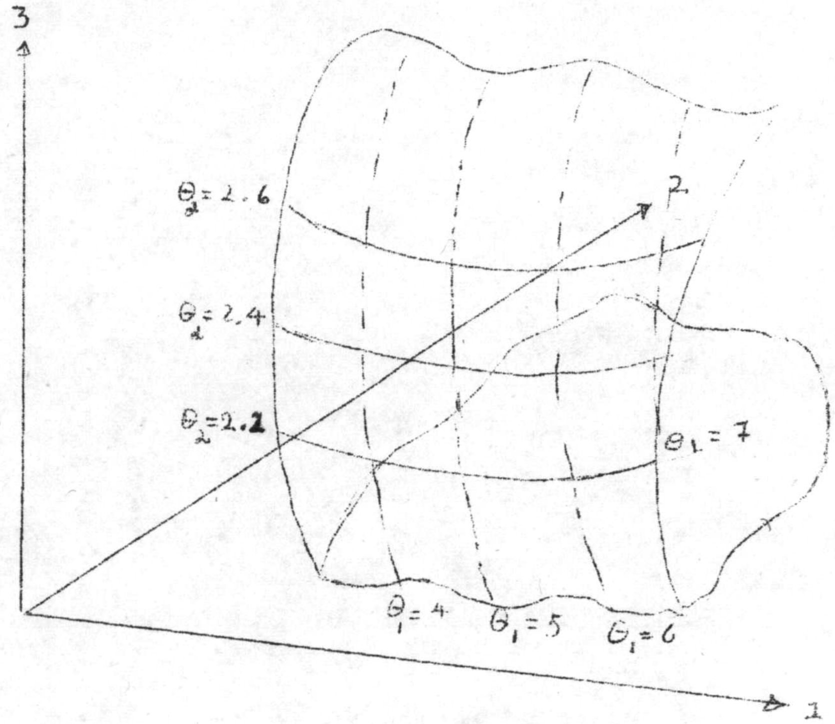
รูปที่ ๖



ตามรูปที่ ๗ แสดง sample space สำหรับตัวอย่างซึ่งมี $n = 3$ ค่าที่วัดได้ y_1, y_2, y_3 ที่ $\xi = \xi_1, \xi_2$ และ ξ_3 ตามลำดับ พารามิเตอร์ ๒ ตัว θ_1 และ θ_2 เส้นโค้งแสดง coordinate systems ของพารามิเตอร์ใน estimation space หรือทางเดินของผลลัพธ์ซึ่งประกอบด้วยทุกจุดที่อยู่ในรูป $\{f(\xi_1, \theta_1, \theta_2), f(\xi_2, \theta_1, \theta_2), f(\xi_3, \theta_1, \theta_2)\}$ ขณะที่ θ_1 และ θ_2 เปลี่ยนค่า ξ_1, ξ_2, ξ_3 จะคงที่ y มีระยะตามแกนเป็น (y_1, y_2, y_3) และ P เป็นจุดของ estimation space ที่ใกล้ y ที่สุด

เมื่อใช้เทคนิคของสมการเส้นตรงกับปัญหาที่ไม่ใช่เส้นตรงเราเลือกจุด θ_o ใน sample space เป็นจุดกำเนิดใหม่ กำหนด estimation spaceแบบเส้นตรงในรูปของเส้นสัมผัสกับ space ที่ θ_o และหากำลังสองน้อยที่สุดแบบเส้นตรง จะได้ผลลัพธ์ซึ่งให้อยู่ในหน่วยของอัตราการเปลี่ยนของ θ ซึ่งเริ่มจาก θ_o เท่านั้น ผลลัพธ์นี้ใช้กับปัญหาที่ไม่ใช่เส้นตรงจะไม่ถูกต้องนัก

รูปที่ ๗ sample space เมื่อ $n = 2$, $f(\xi, \theta)$ ไม่ใช่เส้นตรง

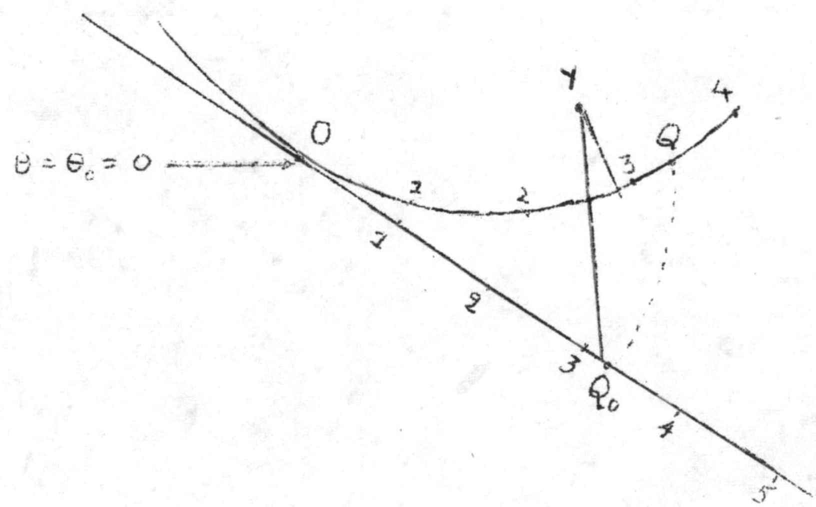


สำหรับปัญหาที่ไม่ใช่เส้นตรงมีค่าที่ได้จากการสังเกตเพียง ๒ ค่า และมีพารามิเตอร์ ๑ ตัว จะให้ผลตามที่แสดงในรูปที่ ๘ ตามรูปแสดง estimation space หรือทางเดินของผลลัพธ์ในหน่วยของ θ ข้างบน เราสมมุติว่า $\theta_o = 0$ และจุดซึ่ง $\theta = 1$ เป็นจุดของ estimation space ได้จาก $\theta = 1$ และค่าอื่น ๆ ให้สังเกตว่า θ ไม่เท่ากันเนื่องจากระบบของแกนไม่เป็นเส้นตรง และไม่สม่ำเสมอ เส้นที่สัมผัสกับเส้นโค้ง estimation space ที่ $\theta = \theta_o = 0$ ดังแสดงจะค่อย ๆ เพิ่มทีละหน่วย $\theta = 0, 1, 2, \ldots\ldots\ldots$ ซึ่งได้จากอัตราการเปลี่ยนแปลงที่ θ_o หน่วยเหล่านี้จะกระจายเท่ากัน เราจะหาตัวประมาณกำลังสองน้อยที่สุดของ θ โดยใช้ข้อตกลงเบื้องต้นของเส้นตรง

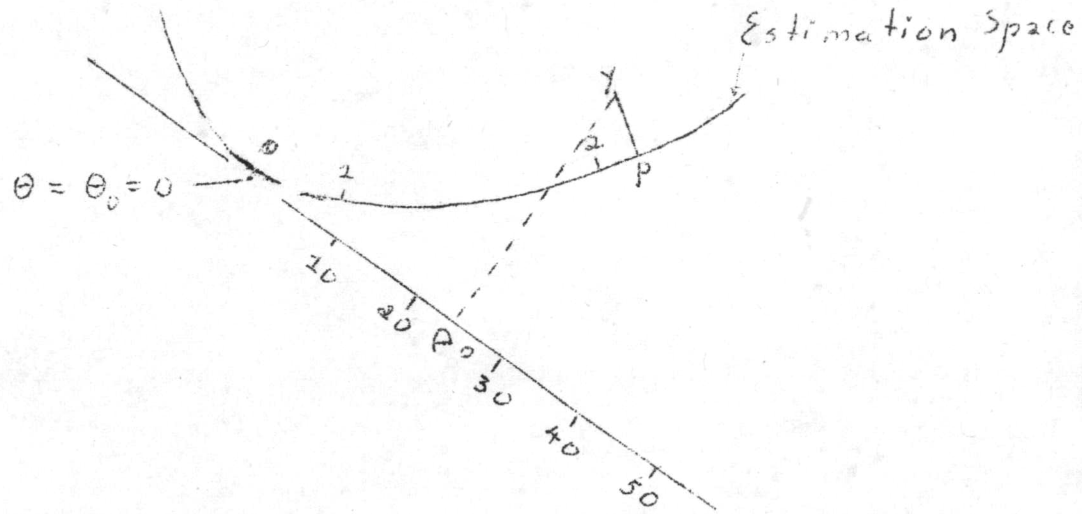
ตามเรขาคณิตหมายความว่าจะหาจุด Q_o เพื่อว่า yQ_o ตั้งฉากกับเส้นสัมผัส เราจะเห็นว่าในหน่วยเส้นตรงค่าของ θ ประมาณ ๓.๒ (ที่ Q_o) ดังแสดง ต่อไปจะใช้แบบเส้นตรงโดยใช้เส้นสัมผัสที่จุดซึ่ง θ = ๓.๒ บน estimation space curve นั่นคือที่จุด Q

จากเหตุผลนี้แสดงว่าการใช้วิธีการของเส้นตรงบางครั้งไม่ได้ผล ถ้าอัตราการเปลี่ยนของ $f(\xi, \theta)$ น้อยมากที่ θ_o แต่เพิ่มอย่างรวดเร็ว หน่วยบนเส้นสัมผัสจะไม่เป็นจริงดังตัวอย่างในรูปที่ ๘ อัตราการเปลี่ยนที่ $\theta_o = 0$ น้อยมากและหน่วยของ θ บนเส้นตรงน้อยมาก

รูปที่ ๘ การแปลความหมายทางเรขาคณิตโดยวิธีเส้นตรง (n = 2, p = 1)



รูปที่ ๔ ผลของวิธีเส้นตรงของส่วนที่เท่ากัน ($n = 2$, $P = 1$)



หน่วยที่แท้จริงจะเพิ่มขึ้นอย่างเที่ยงตรง ดังนั้นถ้าเราเริ่มใช้ค่าของ θ ประมาณ ๒๖ ที่จุด Q_0 จุดเริ่มต้นของเราใน estimation space จะห่างจากจุด P มากกว่าตอนแรกที่ใช้ $\theta = \theta_0 = 0$ สภาพที่เกิดขึ้นอาจจะถูกหรือผิดก็ได้ แม้ว่าจะใช้ $\theta_0 = 0$ และหน่วยเป็น ๑, ๒, ...เมื่อให้ใช้ทั่วไปได้ง่าย ไม่ว่าค่า θ_0 ตอนแรกจะเป็นเท่าไร และไม่ว่าระบบของหน่วยที่ใช้จะใกล้ θ_0 เท่าไร

เมื่อมีค่าที่วัดได้มากกว่า ๒ และพารามิเตอร์มากกว่า ๑ แนวความคิดคงเดิมแต่สภาพการณ์ยุ่งยากกว่าเขียนยากหรือเขียนไม่ได้เลย ถ้าสมการเป็นเส้นตรง $S(\theta)$ ใน sample space ประกอบด้วยทรงกลม ในกรณีที่ไม่ใช่เส้นตรงภายในจะต้องประกอบด้วยทุกจุดของ estimation space ระยะเท่ากัน ค่าที่เลือกจากจุด y $(y_1, y_2, \ldots y_n)$

การหาขอบเขตของความเชื่อมั่นเมื่อสมการไม่เป็นเส้นตรงผลลัพธ์ที่ได้จะแตกต่างกันออกไปเช่น เมื่อค่าผิดพลาด ϵ กระจายเป็นโค้งปกติ $\hat{\theta}$ จะไม่กระจายเป็นโค้งปกติและ $s^2 = S(\hat{\theta})/(n-p)$ ก็ไม่ใช่ค่าประมาณที่เที่ยงตรงของ σ^2 แม้ว่าจะเขียนขอบเขตของความเชื่อมั่นในรูป

$$S(\theta) = S(\hat{\theta}) \left[1 + \frac{p}{n-p} F(p, n-p, 1-\alpha) \right]$$

ซึ่งจะให้ขอบเขตของความเชื่อมั่น ๑๐๐ (๑ − α)% สำหรับสมการเส้นตรง สำหรับค่าผิดพลาดปกติในกรณีที่ไม่ใช่สมการเส้นตรงสัมประสิทธิความเชื่อมั่นจะไม่เท่ากับ ๑ − α โดยทั่วไปเราไม่ทราบขอบเขตของความเชื่อมั่นแน่นอนแต่ถือว่าเป็นขอบเขตของความเชื่อมั่นโดยประมาณของ θ เป็น ๑๐๐ (๑ − α)%

๔. <u>การนำวิธีกำลังสองน้อยที่สุดไปประยุกต์</u>

๔.๑ ปัญหาการประมาณค่า (Estimation Problem)

๔.๑.๑ การประมาณค่าตัวพารามิเตอร์ในสมการถดถอย (Regression Analysis)

เมื่อมีข้อมูลสองชุดที่มีความสัมพันธ์กันเป็นเส้นตรง เราจะหาเส้นตรงซึ่งเหมาะสมกับข้อมูลมากที่สุด ให้สมการเป็น

$$y = \alpha + \beta x + e$$

เมื่อ y เป็นตัวแปรสุ่มใด ๆ ที่วัดได้โดย $E(y) = \alpha + \beta x$

x = เป็นตัวแปรที่ทราบค่า (ไม่ใช่ตัวแปรสุ่ม)

α และ β เป็นพารามิเตอร์ที่ต้องการหาค่า

เลือก x, n ค่าและสำหรับแต่ละค่าของ x_i, y_i จะถูกเลือกจากdensity ซึ่งมีมัชฌิม (Mean) เป็น $\alpha + \beta x_i$ และความแปรปวนเป็น σ^2 จะต้องมีข้อตกลงเบื้องต้นต่อไปนี้

ก. ตัวแปรสุ่ม $y_1, y_2, \ldots\ldots y_n$ เป็นตัวแปรที่ไม่ขึ้นแก่กันโดยมีมัชฌิม $E(y_i) = \alpha + \beta x_i$ และความแปรปวน = σ^2 ซึ่งไม่ขึ้นกับ x_i

ข. α และ β จะทำให้ e_i เป็นตัวแปรสุ่มที่ไม่ขึ้นแก่กัน $E(e_i) = 0$ และความแปรปวนของ $e_i = \sigma^2$

ข้อตกลงเบื้องต้นทั้งสองข้ออาจเขียนสรุปได้ดังนี้

$$E(y_i) = \alpha + \beta x_i$$

$$\text{Cov}(y_i\, y_j) = 0 \qquad i \neq j \quad \text{..........(1)}$$

$$\text{Var}(y_i) = \sigma^2$$

จะเห็นว่าไม่ได้ระบุความหนาแน่นของตัวแปรสุ่ม y_i และ e_i ด้วย ในกรณีนี้ความน่าจะเป็นสูงสุดของ α และ β จะหาไม่ได้ แต่สามารถใช้วิธีกำลังสองน้อยที่สุดได้[๒๖]

การหาตัวพารามิเตอร์โดยให้มีคุณสมบัติตามเกณฑ์ของตัวประมาณค่าที่ดี ให้ (y_i, x_i), $i = 1, 2, \ldots\ldots n$ เป็นค่า n จุดซึ่งสอดคล้องกับสมการเส้นตรง $y_i = \alpha + \beta x_i + e_i$ ตัวประมาณค่ากำลังสองน้อยที่สุดเป็นค่า α และ β ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อน $\sum_{i=1}^{n} e_i^2$ มีค่าน้อยที่สุด แทนค่า e_i จะได้ว่า

$$L(\alpha, \beta) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

ตัวประมาณค่ากำลังสองน้อยที่สุดจะเป็น

$$\hat{\alpha} = \bar{y} + \hat{\beta}\,\bar{x} \qquad \text{...........(2)}$$

$$\hat{\beta} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

วิธีกำลังสองน้อยที่สุดไม่ให้ตัวประมาณค่า σ^2 แต่ตัวประมาณค่าของ σ^2 ขึ้นกับตัวประมาณค่าน้อยที่สุดของ α และ β คือ

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \text{.........(3)}$$

เกณฑ์ที่แสดงคุณสมบัติที่ดีของตัวประมาณค่าคือ จะต้องมีความเที่ยงตรง (Unbiasedness) และมีประสิทธิภาพหรือมีความแปรปวนต่ำสุด (Minimum Variance) เราจะพิสูจน์คุณสมบัตินี้สำหรับสมการเส้นตรง ทฤษฎีนี้เป็นที่รู้จักกันทั่วไปเรียก "ทฤษฎีของเกาส์มาคอฟฟ์"

---

26 Alexander M. Mood & Franklin A. Gray-bill, Introducation to the Theory of statistics (New York: Mc Graw-Hill Book Company, In., c 1963), p. 340.

<u>ทฤษฎี</u>

ให้สมการ $y_i = \alpha + \beta x_i + e_i$ เป็นสมการที่มี x_i เป็นตัวแปรที่ทราบค่า และ y_i เป็นตัวแปรสุ่มที่วัดได้ สมมุติให้ตัวแปรสุ่ม e_i วัดไม่ได้และไม่ขึ้นแก่กันมีมัชฌิมเป็นศูนย์และมีความแปรปวน $= \sigma^2$ ตัวประมาณค่ากำลังสองน้อยที่สุดตามสมการ (๒) สำหรับ α และ β เป็นตัวประมาณค่าที่เที่ยงตรงที่สุด

<u>พิสูจน์</u>

จะแสดงการพิสูจน์ α สำหรับการพิสูจน์ β จะเหมือนกัน เราจะจำกัดตัวประมาณค่าเป็นเส้นตรงและได้ว่า

$$\hat{\alpha} = \Sigma \text{ ap yp}$$

จะต้องหา ap เพื่อให้

ก. $E(\hat{a}) = \alpha$ ซึ่งเที่ยงตรง

ข. ความแปรปวนของ ($\hat{\alpha}$) เป็นค่าต่ำสุดของตัวประมาณค่าทุกตัวที่เป็นไปตาม (a)

สำหรับ (a) จะต้องมี

$$\alpha = E(\hat{\alpha}) = \Sigma \text{ ap } (E \text{ yp}) = \Sigma \text{ ap } (\alpha + \beta x_p)$$

ซึ่งจะให้สมการ ๒ สมการสอดคล้องกัน

$$\Sigma \text{ ap} = 1 \qquad \ldots\ldots\ldots\ldots (4)$$

$$\Sigma \text{ ap xp} = 0$$

$$\text{Var}(\hat{\alpha}) = E(\hat{\alpha} - \alpha)^2 = E(\Sigma \text{ p } y_p - \alpha)^2 = E\left[\Sigma \text{ ap}(\alpha + \beta x_p + e_p) - \alpha\right]^2$$

$$= E(\alpha \Sigma \text{ap} + \beta \Sigma \text{ ap } x_p + \Sigma \text{ ap } e_p - \alpha)^2$$

ใช้ข้อจำกัดตามสมการ (๔) จะได้ว่า

$$\text{Var}(\hat{\alpha}) = E(\Sigma \text{ ap ep})^2 = E(\Sigma \text{ ap}^2 \text{ ep}^2 + \sum_p \sum_q a_p a_q e_p e_q),\ p \neq q$$

$E(e_p\, e_q)$ เป็น 0 ถ้า $p \neq q$ ดังนั้นจากข้อตกลงเบื้องต้นว่า e_i ไม่มีความสัมพันธ์กันและมีมัชฌิมเป็น 0 ดังนั้นจะได้ว่า

$$\text{Var}(\hat{\alpha}) = \sigma^2 \sum a_p^2$$

σ^2 เป็นตัวคงที่ เมื่อจะทำให้ var ($\hat{\alpha}$) มีค่าน้อยที่สุด $\sum ap^2$ ต้องมีค่าน้อยที่สุด ดังนั้นจะต้องหาตัวคงที่ a_p ซึ่งจะทำให้ $\sum a_p^2$ มีค่าน้อยที่สุดตามสมการ (๔) ใช้ทฤษฎี La. grange multiplier เราจะต้องทำให้ L มีค่าน้อยที่สุด

$$L = \sum ap^2 - \lambda_1(\sum ap - 1) - \lambda_2(\sum ap\, x_p)$$

$$\frac{\partial L}{\partial a_t} = 2\, a_t - \lambda_1 - \lambda_2 x_t \qquad t_1 = 1, 2, \ldots\ldots, n$$

$$\frac{\partial L}{\partial \lambda_1} = -\sum ap + 1 = 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots(5)$$

$$\frac{\partial L}{\partial \lambda_2} = -\sum ap\, x_p = D$$

ถ้าบวก n สมการจะได้ $\sum at = 1$

$$2 = n\lambda_1 + \lambda_2 \sum x_i \qquad \ldots\ldots\ldots\ldots\ldots\ldots(6)$$

คูณสมการที่ p ใน (๕) ด้วย xp จะได้

$$\sum a_p x_p = x_p \qquad \ldots\ldots\ldots\ldots\ldots\ldots(7)$$

(6) x (7)

$$2 \sum a_p x_p = \lambda_1 \sum x_p + \lambda_2 \sum x_p^2$$

แต่ $\sum a_p x_p = 0$

$$\lambda_1 = -\lambda_2 \frac{\sum x_i^2}{\sum x_i} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(8)$$

แทนค่า λ_1 ใน (๖) จะได้ว่า

$$\lambda_2 = \frac{-2 \sum x_i/n}{\sum x_i^2 - n\bar{x}^2} = \frac{-2\bar{x}}{\sum (x_i - \bar{x})^2}$$

และ

$$\lambda_1 = \frac{2 \sum x_i^2 / n}{\sum (x_i - \bar{x})^2}$$

แทนค่า λ_1 และ λ_2 ในสมการที่ t ของ (๕) เมื่อหาค่า a_t จะได้

$$a_t = \frac{(\sum x_i^2/n) - \bar{x}\, x_t}{\sum (x_i - \bar{x})^2}$$

∴ ตัวประมาณค่าในสมการเส้นตรงที่เที่ยงตรงที่สุดของ α คือ

$$\hat{\alpha} = \sum a_t\, y_t = \frac{\bar{y} \sum x_i^2 - \bar{x} \sum y_t\, x_t}{\sum (x_i - \bar{x})^2}$$

$$= \bar{y} - \hat{\beta}\, \bar{x}$$

ซึ่งเหมือนกับที่ได้จากวิธีกำลังสองน้อยที่สุด ฉะนั้นจึงพิสูจน์ได้ว่าค่าประมาณที่ได้จากวิธีกำลังสองน้อยที่สุดเป็นค่าประมาณที่เที่ยงตรงและมีความแปรปวนต่ำที่สุด

ในการคำนวณเราจะได้สมการถดถอยที่ใช้ทำนายค่า y สำหรับค่า x ใด ๆ ดังนี้[๒๗]

$$\bar{y}_x = a + bx$$

เมื่อ

$$a = \hat{\alpha} = \bar{y} + b\bar{x}$$

$$b = \hat{\beta} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$$= \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

และตัวประมาณค่าที่เที่ยงตรงของ σ^2 โดยมีชั้นแห่งความอิสสระ (degree of freedom) เป็น n - 2 และหาได้จากสูตร

$$S^2 = \frac{n - 1}{n - 2} (S_y^2 - b^2 S_x^2)$$

---

27 Ronald E. Walpole, Introduction to Statistics (New York : The Macmillian Company, c 1968), p. 270.

เมื่อ $S_x^2 = \dfrac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}$

$S_y^2 = \dfrac{n \sum y_i^2 - (\sum y_i)^2}{n(n-1)}$

<u>ตัวอย่าง ๑</u> ข้อมูลได้จากคะแนนทดสอบเชาว์ปัญญาและคะแนนวิชาเคมีปีที่ ๑ ของมหาวิทยาลัย

นักเรียนคนที่	คะแนนทดสอบ x	คะแนนเคมี y
๑	๖๕	๘๕
๒	๕๐	๗๔
๓	๕๕	๗๖
๔	๖๕	๙๐
๕	๕๕	๘๕
๖	๗๐	๘๗
๗	๖๕	๙๔
๘	๗๐	๙๘
๙	๕๕	๘๑
๑๐	๗๐	๙๑
๑๑	๕๐	๗๖
๑๒	๕๕	๗๔
	๗๒๕	๑๐๑๑

จากข้อมูลนี้จะคำนวณได้ดังนี้

$\sum_{i=1}^{12} x_i = 725$, $\sum_{i=1}^{12} y_i = 1011$, $\sum_{i=1}^{12} x_i y_i = 61685$

$\sum_{i=1}^{12} x_i^2 = 44,475$, $\bar{x} = 60.417$, $\bar{y} = 84.250$

ดังนั้น $b = \dfrac{(12)(61685) - (725)(1011)}{(12)(44,475) - (725)^2} = 0.897$

$$a = 84.250 - (0.897)(60.417) = 30.056$$

∴ สมการถดถอยคือ

$$\bar{y}_x = 30.056 + 0.897x$$

เช่นถ้านักศึกษาคนหนึ่งได้คะแนนจากแบบทดสอบเชาว์ปัญญาเท่ากับ ๘๐ จะทำนายได้ว่าคะแนนวิชาเคมีปีที่ ๑ จะได้ $= \bar{y}_x$

$$\bar{y}_x = 30.056 + 0.897(80)$$

$$= 101.816$$

และจะหาความแปรปวนได้จาก

$$S = \frac{n - 1}{n - 2} \left(S_y^2 - b^2 S_x^2\right)$$

$$S_x^2 = \frac{(12)(44,475) - (725)^2}{(12)(11)} = 61.174$$

$$S_y^2 = \frac{(12)(85,905) - (1011)^2}{(12)(11)} = 66.205$$

แทนค่า S_y^2 , S_x^2 และ b^2 จะได้

$$S^2 = \frac{11}{10} \left[66.205 - (0.805)(61.174) \right] = 18.656$$

ซึ่งสามารถจะนำไปทดสอบสมมุติฐานหรือหาขอบเขตของความเชื่อมั่นต่อไปได้

<u>ตัวอย่าง ๒</u> แสดงการคำนวณโดยใช้คะแนนที่เบี่ยงเบน (Deviation score) เพื่อให้การคำนวณง่ายขึ้น[๒๘]

ข้อมูลข้างล่างเป็นคะแนนที่ได้จากการทดสอบนักศึกษา ๑๐ คน โดยผู้ทดสอบสองคน

ตารางที่ ๕

นักเรียนคนที่	คะแนนดิบ		คะแนนที่เบี่ยงเบน		คะแนนมาตรฐาน	
	X_1	X_2	x_1	x_2	z_1	z_2
๑	๗๐	๖๙	+๒.๖	+๓.๐	+๐.๕๙	+๐.๘๖
๒	๗๒	๖๕	+๔.๖	–๑.๐	+๑.๐๔	–๐.๒๙
๓	๖๖	๖๗	–๑.๔	+๑.๐	– ๐.๗๒	+๐.๒๙
๔	๖๗	๖๙	–๐.๔	+๓.๐	– ๐.๐๙	+๐.๘๖
๕	๖๗	๖๓	–๐.๔	– ๓.๐	– ๐.๐๙	–๐.๘๖
๖	๗๔	๖๘	+๖.๖	+ ๒.๐	+ ๑.๔๙	+๐.๕๗
๗	๗๑	๗๑	+๓.๖	+ ๕.๐	+ ๐.๘๑	+๑.๔๓
๘	๖๒	๖๖	–๕.๔	๐.๐	– ๑.๒๒	๐.๐๐
๙	๖๐	๖๒	–๗.๔	– ๔.๐	– ๑.๖๗	–๑.๑๔
๑๐	๖๕	๖๐	–๒.๔	– ๖.๐	– ๐.๕๔	–๑.๗๑
ผลบวก	๖๗๔	๖๖๐	๐.๐	๐.๐	๐.๐๐	๐.๐๑
ผลบวกของกำลังสอง	๔๕,๖๐๔	๔๓,๖๗๐	๑๗๖.๔	๑๑๐.๐	๘.๙๙	๘.๙๘

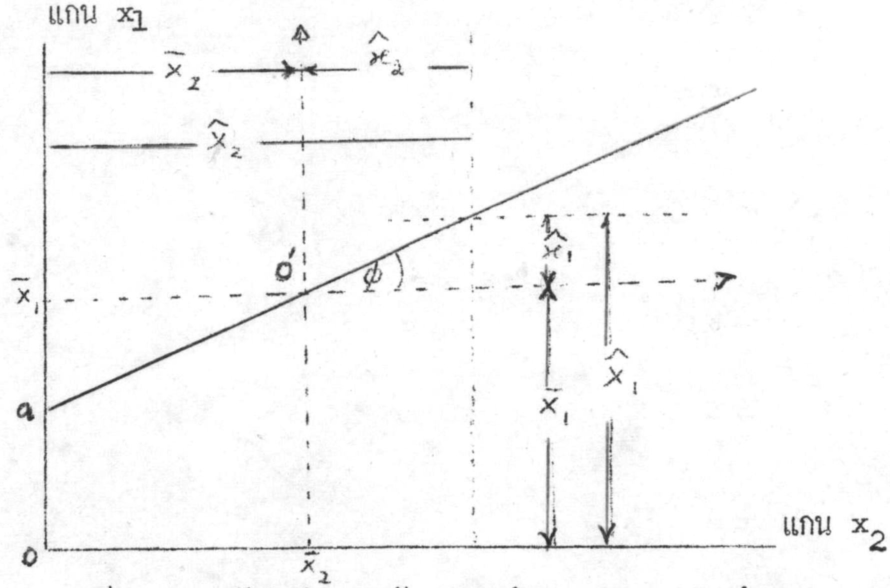
$\bar{X}_1$ = เป็นมัชฌิมของคนที่ได้จากผู้ทดสอบคนที่ ๑

$\bar{X}_2$ " " " " ๒

---

[28] Francis G. Cornell, <u>The Essentials of Educational Statistics</u> (New York : John Wiley & Sons, Inc., c 1956), p. 149.

รูปที่ ๑๐



จากกราฟแสดงให้เห็นว่าการใช้คะแนนเบี่ยงเบนเป็นการเปลี่ยนจุดกำเนิดจาก o เป็น o' ตัวแปรอิสสระ X_1 มี ๓ ค่า

ก. คะแนนของตัวแปรอิสสระ X_1

ข. มัชฌิมของ X_1 ทั้งหมดคือ $\bar{X}_1$

ค. ระยะจากจุดบนเส้นตรงไปยังแกน X_2 หรือค่าที่ทำนาย$\hat{X}_1$ ส่วนที่เบี่ยงเบนมี ๓ ชนิดด้วย

ก. ส่วนเบี่ยงเบนของ X_1 จากมัชฌิม $\bar{X}_1$

ข. ส่วนเบี่ยงเบนของค่าบนเส้นตรงกับมัชฌิม ($\hat{x}_1$)

ค. ส่วนเบี่ยงเบนของ X_1 จากเส้นตรง

ส่วนเบี่ยงเบนชนิด ค. อาจเป็นบวกถ้าอยู่เหนือเส้นตรงและเป็นลบเมื่ออยู่ใต้เส้นตรงเป็นค่าที่จะต้องทำให้กำลังสองของค่านี้มีค่าน้อยที่สุด

a = จุดตัดของเส้นตรงบนแกน X_1

b = ความชัน (Slope) ของเส้นตรงกำลังสองน้อยที่สุด

ในกรณีที่ใช้คะแนนเบี่ยงเบนจุดกำเนิดเป็น O' จะทำให้ a = 0 และสมการจะเหลือเพียง

$$\hat{x}_1 = b_1 x_2$$

$$b_1 = \frac{\Sigma x_1 x_2}{\Sigma x_2^2}$$

$$= \frac{77.0}{110} = .70$$

$$\hat{x}_1 = .70\, x_2 \qquad \ldots\ldots\ldots\ldots\ldots (1)$$

ถ้าจะทำนาย x_2 จาก x_1

$$b_2 = \frac{\Sigma x_1 x_2}{\Sigma x_1^2}$$

$$= \frac{77.0}{176.4} = .437$$

$$\hat{x}_2 = b_2 x_1$$

$$\hat{x}_2 = .437 x_1 \qquad \ldots\ldots\ldots\ldots\ldots (2)$$

สมการ (๑) และ (๒) เป็น Deviation scores form ถ้าจะทำเป็นสูตรที่ใช้กับคะแนนดิบได้เลย (Raw score form) แทนค่า

$$\hat{x}_1 = b_1 x_2$$

$$(\hat{X}_1 - \bar{X}_1) = b_1 (X_2 - \bar{X}_2)$$

$$\hat{X}_1 = b_1 (X_2 - \bar{X}_2) + \bar{X}_1$$

$$= b_1 X_2 + \bar{X}_1 - b_1 \bar{X}_2$$

$$\therefore \quad a_1 = \bar{X}_1 - b_1 \bar{X}_2$$

สมการสำหรับ Raw score form คือ

$$\hat{X}_1 = b_1 X_2 + a_1$$

$$\hat{X}_2 = b_2 X_1 + a_2$$

$$b_1 = \frac{\Sigma x_1 x_2}{\Sigma x_1^2}$$

$$= r_{12} \frac{S_1}{S_2}$$

และ $$b_2 = \frac{\Sigma x_1 x_2}{\Sigma x_2^2}$$

$$= r_{21} \frac{S_2}{S_1}$$

$$b_1 = .55 \left(\frac{4.43}{3.50}\right) = .700$$

และ $$b_2 = .55\left(\frac{3.5}{(4.43)}\right) = .435$$

$$a_1 = \bar{X}_1 - b_1 \bar{X}_2$$

$$a_2 = \bar{X}_2 - b_2 \bar{X}_1$$

แทนค่า

$$a_1 = 67.4 - (.70)(66.0) = 21.2$$

และ $$a_2 = 66.0 - (.44)(67.4) = 36.3$$

จะได้สมการถดถอยดังนี้

$$\hat{X}_1 = .70\, X_2 + 21.2$$

และ $$\hat{X}_2 = .44\, X_1 + 36.3$$

$$r_{12} = \sqrt{b_1 b_2}$$

สมการถดถอยในรูปของคะแนนมาตรฐานจะเป็น

$$\hat{z}_1 = r_{12} z_2$$

และ $$\hat{z}_2 = r_{12} z_1$$

ในการนี้สัมประสิทธิ์สหสัมพันธ์จะเป็นความชันของเส้นตรงด้วย

<u>ตัวอย่าง ๓</u>

การนำวิธีกำลังสองน้อยที่สุดไปใช้ในการรวบรวมข้อมูล เช่นมีกลุ่มตัวอย่างที่ได้จากการสำรวจ ๒ แบบ คือการสัมภาษณ์และการส่งแบบสอบถามทางไปรษณีย์ เราเชื่อว่าการสัมภาษณ์ให้ผลดีกว่า เราจะสุ่มตัวอย่างมาใช้ทั้งสองวิธี แล้วสร้างสมการแสดงความสัมพันธ์ระหว่าง ๒ วิธี ต่อไปจะใช้แบบสอบถามส่งทางไปรษณีย์ แล้วมาหาค่าที่ถูกต้องจากสมการ ตัวอย่างข้างล่างเป็นการเก็บข้อมูลเกี่ยวกับการฟังรายการวิทยุ[๒๙]

---

[29] Morris H. Hansen, William H. Hurwitz, William G. Madow, <u>Sample Survey Methods and Theory</u> (New York : John Wiley & Sons, Inc., c 1953), p. 484.

ตารางเปรียบเทียบผลที่ได้จากการสัมภาษณ์กับ
การส่งแบบสอบถามทางไปรษณีย์

ตารางที่ ๖

จำนวนสถานี ที่รับได้	เปอร์เซ็นต์ของแม่บ้านที่รายงานจำนวนสถานีที่รับได้			
	กลางวัน		กลางคืน	
	ส่งไปรษณีย์(%)	สัมภาษณ์ (%)	ส่งไปรษณีย์ (%)	สัมภาษณ์ (%)
๐	๐	๐	๐	๐
๑	๔	๖	๖	๖
๒	๑๑	๑๕	๑๕	๑๖
๓	๑๒	๒๒	๑๖	๒๒
๔	๑๗	๒๔	๑๖	๒๓
๕	๑๖	๑๕	๑๕	๑๖
๖	๑๒	๘	๑๑	๙
๗	๑๐	๕	๘	๕
๘	๗	๒	๕	๒
๙	๕	๒	๓	๑
๑๐หรือมากกว่า	๖	๑	๕	๑
รวม	๑๐๐	๑๐๐	๑๐๐	๑๐๐
มัชฌิมจำนวนสถานีที่ได้ยิน	๔.๙	๓.๘	๔.๓	๓.๘

ตามตารางเปรียบเทียบข้อมูล ๒ ชุดนี้เมื่อจะหาแนวโน้มที่จะได้จากการสัมภาษณ์ มีวิธีที่จะปรับข้อมูลที่ได้จากการส่งทางไปรษณีย์หลายวิธี แต่วิธีกำลังน้อยที่สุดทำได้โดยหาเส้นตรง $\hat{y} = a + bx$ เมื่อหาค่า a และ b ได้แล้วจะนำข้อมูลที่ได้จากการส่งทางไปรษณีย์ x มาปรับหาค่าให้ใกล้เคียงกับค่าที่ได้จากการสัมภาษณ์ y ได้โดยแทนค่า x ในสมการข้างบน การหาค่า a และ b ใช้สูตรเช่นเดียวกับตัวอย่าง ๑ และ ๒ คือ

$$a = \bar{y} - b\bar{x}$$

$$b = \Sigma (x_i - \bar{x}) (y_i - \bar{y}) / \Sigma (x_i - \bar{x})^2$$

การหาระดับของความเชื่อมั่นของค่าที่ประมาณได้จากการส่งทางไปรษณีย์ ถ้ามีการสำรวจข้อมูลทั้งหมดโดยการสัมภาษณ์แบบสุ่มตัวอย่าง ค่าจริงของแต่ละเมืองจะไม่เบี่ยงเบนจากเส้นตรงนี้มากนัก ขอบเขตของความเชื่อมั่นคือ $\hat{y} \pm k \sqrt{MSE}$

k = โอกาสที่ต้องการให้ได้ค่าที่ถูกต้อง

$\sqrt{MSE}$ = root mean square error

$\hat{y}$ = เป็นค่าที่ได้จากสมการถดถอย

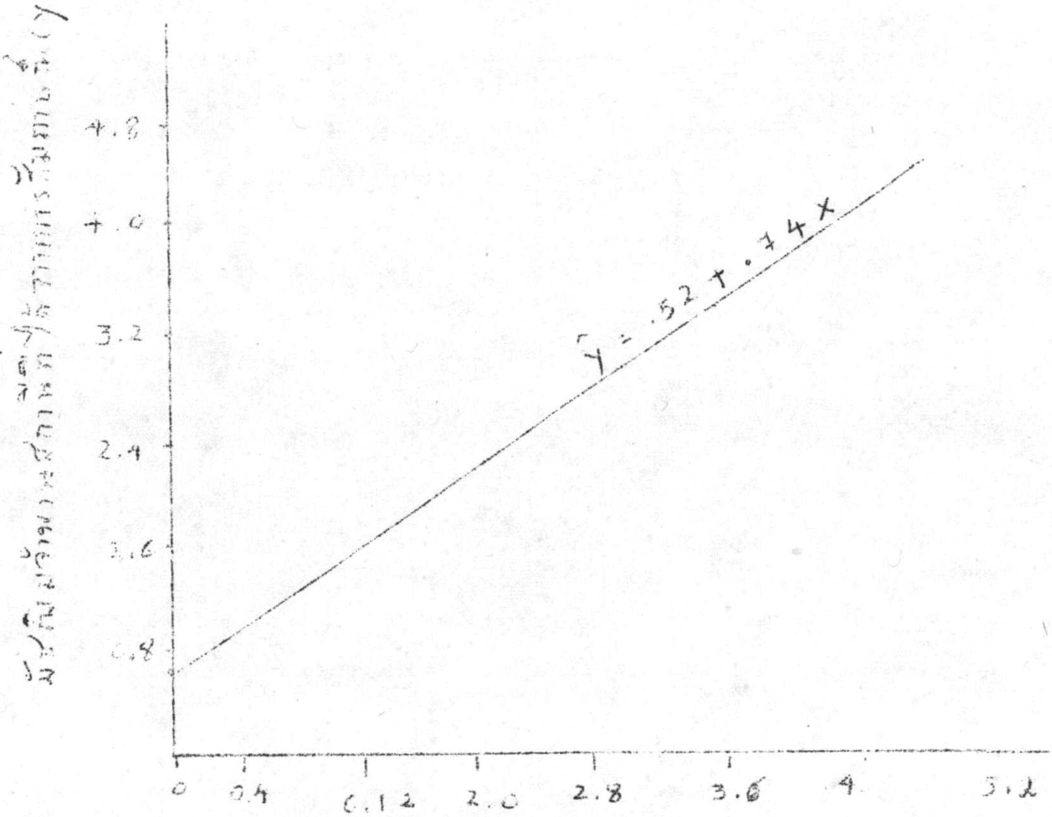
ถ้าการสุ่มนี้ทำในแต่ละกลุ่มของประชากร k = 1

จากข้อมูลตามตารางจะได้ a = .52, b = .74

∴ สมการถดถอยที่ใช้ทำนายคือ

$$\hat{y} = .52 + .74\,x$$

รูปที่ ๑๑



มัชฌิมจำนวนสถานีจากการส่งทางไปรษณีย์ (x)

ส่วนเบี่ยงเบนมาตรฐาน $\sigma_{\hat{y} - y} = 0.50$

สัมประสิทธิสหสัมพันธ์ $r_{xy} = 0.70$

ถ้าสัมประสิทธิสหสัมพันธ์ค่อนข้างสูงค่าที่วัดได้จากข้อมูลที่สุ่มอย่างหยาบ ๆ (ตามตัวอย่างใช้ส่งทางไปรษณีย์) จะให้ค่าใกล้เคียงกับมัชฌิมของแต่ละบริเวณที่สำรวจ เมื่อสัมประสิทธิสหสัมพันธ์ค่อนข้างต่ำ ข้อมูลจากการวัดอย่างหยาบ ๆ จะผิดจากค่าจริง อย่างไรก็ตาม ค่าประมาณที่ปรับแล้วสำหรับค่าเฉพาะใด ๆ ยังใกล้เคียงกับค่าเฉลี่ยที่ได้จากการสัมภาษณ์ทั้งหมด

การคำนวณหาสมการถดถอยกำลังหนึ่งอาจใช้เรื่องเมตริก[๓๐] ช่วยในการคำนวณ โดยเฉพาะเมื่อมีตัวแปรมากกว่า ๒ ตัว ซึ่งจุดต่าง ๆ กระจายอยู่ในที่ว่าง (space)

<u>ตัวอย่าง ๔</u> การใช้เมตริกกับข้อมูลที่มีตัวแปรเพียง ๒ ตัว[๓๑]

y	0	0	1	1	3
x	-2	-1	0	1	2

สมการที่แสดงความสัมพันธ์ระหว่างข้อมูล $y = \beta_0 + \beta_1 x + \epsilon$

$$y = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix}, \quad x = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$x'x = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}$$

$$x'y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \end{pmatrix}$$

$$(x'x)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{10} \end{pmatrix}$$

จาก Normal equation เราได้

$$\hat{\beta} = (x'x)^{-1} x'y$$

---

[๓๐] ภาคผนวก ก.

[๓๑] William, Mendenhall, <u>Introduction to Linear Models and The Design and Analysis of Experiments</u> (California : Wadsworth Publishing Company, Inc., c 1968), p.126 .

แทนค่า $\hat{\beta} = \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{10} \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 \\ .7 \end{bmatrix}$

จะได้สมการที่ใช้ทำนายคือ $\hat{y} = 1 + 0.7\,x$

วิธีกำลังสองน้อยที่สุดอาจใช้หาค่าตัวคงที่ทางวิทยาศาสตร์ จากการทดลองหลาย ๆ ครั้งได้

ถ้า $y = \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots\ldots \beta_m x_m$

ถ้าทำการทดลอง n ครั้ง เมื่อ $n > m$ และจากการทดลองได้ค่า $x_{i1}, x_{i2}, \ldots\ldots x_{im}$ สำหรับค่า y_i เราจะได้ n สมการและได้ normal equation ตามที่กล่าวแล้ว

<u>ตัวอย่าง ๕</u> ให้ $y = \beta_1 x_1 + \beta_2 x_2$ และผลจากการทดลอง ๔ ครั้งได้ค่า $y = (15, 12, 10, 0)$, $x_1 = (1, 2, 1, 1)$, $x_2 = (2,1,1,-1)$ และได้สมการ ๔ สมการดังนี้ [๓๒]

$$15 = \beta_1 + 2\beta_2$$
$$12 = 2\beta_1 + \beta_2$$
$$10 = \beta_1 + \beta_2$$
$$0 = \beta_1 - \beta_2$$

ถ้า x_1 และ x_2 ไม่ขึ้นแก่กันจะได้ Normal equation ดังนี้

$$(x_1 \cdot x_1)\,\hat{\beta}_1 + (x_1 \cdot x_2)\,\hat{\beta}_2 = x_1 y$$

$$7\,\hat{\beta}_1 + 4\,\hat{\beta}_2 = 49 \quad \ldots\ldots\ldots\ldots\ldots (1)$$

และจาก $(x_2 . x_1)\,\hat{\beta}_1 + (x_2 . x_2)\,\hat{\beta}_2 = x_2 y$

$$4\,\hat{\beta}_1 + 7\,\hat{\beta}_2 = 52 \quad \ldots\ldots\ldots\ldots\ldots (2)$$

จาก (๑) และ (๒) จะได้ $\hat{\beta}_1$ และ $\hat{\beta}_2$ ซึ่งเป็นตัวคงที่ทางวิทยาศาสตร์ดังนี้

$$\hat{\beta}_1 = \frac{45}{11} \qquad \text{และ} \qquad \hat{\beta}_2 = \frac{56}{11}$$

จะได้ว่า $11y = 45\,x_1 + 56x_2$

---

[32] Donald R. Ostberg & Fred W. Perkins, <u>An Introduction to Linear Analysis</u> (Massachusetts : Addison Wesley Publishing Company, Inc., 1966), pp. 294 - 95.

<u>ตัวอย่าง ๖</u> เมื่อมีตัวแปร ๓ ตัวดังตารางข้างล่าง[๓๓]

y	2	5	7	8	5
x_1	8	8	6	5	3
x_2	0	1	1	3	4

สมการถดถอยจะอยู่ในรูป $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$$\sum_{i=1}^{5} x_{1i} = 30,\ \sum_{i=1}^{5} x_{2i} = 9,\ \sum_{i=1}^{5} x_{1i} x_{2i} = 41,\ \sum_{i=1}^{5} x_{1i}^2 = 198$$

$$\sum_{i=1}^{5} x_{2i}^2 = 27,\ \sum_{i=1}^{5} y_i = 27,\ \sum_{i=1}^{5} x_{1i} y_i = 153,\ \sum_{i=1}^{5} x_{2i} y_i = 56$$

แทนค่าจะได้ Normal equation ดังนี้

$5\,b_0 + 30\,b_1 + 9\,b_2 = 27$ ....(1)

$30b_0 + 198b_1 + 41b_2 = 153$ ....(2)

$9\,b_0 + 41b_1 + 27\,b_2 = 56$ ....(3)

จากสมการทั้ง ๓ จะได้ $b_0 = 4.486$, $b_1 = -0.039$ และ $b_2 = 0.638$

ฉะนั้นจะได้สมการถดถอยดังนี้

$$\hat{y} = 4.486 - 0.039x_1 + 0.638x_2$$

หรือจะคำนวณโดยใช้เมตริกก็จะได้ผลอย่างเดียวกัน

ในกรณีที่สมการถดถอยไม่ใช่สมการกำลังหนึ่งแต่สามารถแปลงให้เป็นสมการเส้นตรงได้เช่นอยู่ในรูปรูต, เศษส่วน หรือค่า log ตัวอย่างเช่นสมการอยู่ในรูปกำลัง[๓๔]

$$y_c = AB^x$$

---

33 Walpole, <u>Opcit.</u>, pp. 283 - 84.

34 Hughes, Ann & Grawoig, Dennis, Statistics. <u>A Foundation for Analysis</u> (Massachusetts : Addison-Wesly Publishing Company, c 1971), p. 390.

การแปลงเป็นรูปสมการเส้นตรงมีประโยชน์มากเพราะการทำให้ $\Sigma(y - AB^x)^2$ มีค่าน้อยที่สุดจะได้ Normal Equation ที่ไม่สามารถจะคำนวณได้ แก้ไขได้โดยใช้ log ดังนี้

$$\log y_c = \log A + x \log B$$

ข้อควรระวังในการใช้เส้นตรงกำลังสองน้อยที่สุดนี้

๑. เส้นตรงกำลังสองน้อยที่สุดได้จากการทำให้ผลบวกกำลังสองของ log ของค่าที่ได้จากการสังเกต y เบี่ยงเบนจาก log ของค่าที่คำนวณได้น้อยที่สุด

๒. การคำนวณหาค่า y ต้องใช้การบวกลบ log ของค่าต่าง ๆ โดยตลอด ผลลัพธ์ที่ได้ต้องเปลี่ยนกลับเป็นค่า y ที่แท้จริงอีกครั้งหนึ่ง

$$\log y_i = \log A + x \log B + e_i$$

เทียบกับสมการ

$$y_i = a + bx_i + e_i$$

$a = \log A$, $b = \log B$ สูตรที่ใช้คำนวณจะเป็น

$$b = \frac{n \Sigma x_i \log y_i - \Sigma x_i \Sigma \log y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2}$$

$$a = \bar{y} - b\bar{x}$$

ตัวอย่าง ๗ ข้อมูลต่อไปนี้เป็นจำนวนผู้ที่สมัครเข้าเรียนในคณะศิลปศาสตร์ในช่วงเวลา ๗ ปี[๓๕]

x ปี	๑	๒	๓	๔	๕	๖	๗
y จำนวนสมัครเข้าเรียน	๓๐๔	๓๔๑	๓๙๓	๔๕๗	๕๔๘	๖๗๐	๘๘๒

ใช้วิธีกำลังสองน้อยที่สุดประมาณเส้นโค้ง $y_c = AB^x$ และสามารถจะประมาณค่าจำนวนนักศึกษาที่จะเข้าเรียนใน ๕ ปีข้างหน้าได้

---

35 Walpole, Op. cit., pp. 281 - 82.

ค่าของ log y = 2.483, 2.533, 2.594, 2.660, 2.739, 2.826

และ 2.945 ตามลำดับและจะได้

$$\sum_{i=1}^{7} x_i = 28, \quad \sum_{i=1}^{7} \log y_i = 18.780, \quad \sum_{i=1}^{7} x_i^2 = 140$$

$$\sum_{i=1}^{7} x_i \log y_i = 77.237, \quad \bar{x} = 4 \quad \overline{\log y} = 2.683$$

ใช้สูตรดังกล่าวแล้วแทนค่า

$$b = \frac{7\,(77.237) - (28)\,(18.780)}{7(140) - (28)^2} = 0.076$$

และ $a = 2.683 - (0.076)(4) = 2.379$

$$A = 10^{2.379} = 239$$

$$B = 10^{0.076} = 1.19$$

สมการถดถอยกำลังสองน้อยที่สุดคือ

$$y_c = (239)\,(1.19)^x$$

จากอัตราการเพิ่มตามสมการข้างบน คาดว่าจำนวนผู้ที่จะสมัครเข้าเรียนในอีก ๕ ปีข้างหน้าจากปัจจุบัน (x = 12) จะเป็น

$$y_c = (239)\,(1.19)^{12} = 1954$$

นอกจากนี้ยังมีสมการเส้นตรงชนิดที่เป็นความสัมพันธ์ระหว่างค่าของ log ทั้งตัวแปรอิสสระและตัวแปรตามคือ $(\log y)_c = \log a + b \log x$

ตัวคงที่ log a และ b หาได้จาก Normal equation ต่อไปนี้

$$\sum \log y = n \log a + b \sum \log x \quad \ldots\ldots\ldots\ldots (1)$$

$$\sum (\log x . \log y) = \log a \sum \log x + b \sum (\log x)^2 \quad \ldots\ldots (2)$$ [36]

---

36 Hughes, Op.cit., pp. 290 - 92.

ความแปรปวนทั้งหมดคำนวณได้จาก

$$\sum (\log y_i)^2 = \sum (\log y)^2 - (\overline{\log y}) \sum \log y$$

เมื่อ $(\overline{\log y}) = \left[(\sum \log y)/n \right]$

Explained sum of squares หาได้จาก

$$\sum(\log y)_c^2 = \log a \sum \log y + b \sum (\log x \cdot \log y) - (\overline{\log y}) \sum \log y$$

กำลังสองของค่าผิดพลาดที่ไม่ทราบ หาได้จากผลต่างระหว่างผลบวกกำลังสองทั้งหมดกับผลบวกกำลังสองของส่วนที่ทราบหรืออธิบายได้

<u>ตัวอย่าง ๘</u> คำนวณหาสมการ $(\log y)_c = \log a + b \log x$ ซึ่งเป็นความสัมพันธ์ระหว่างค่ารักษารถเป็นเซ็นต่อไมล์กับจำนวนเดือนหลังจากตรวจสภาพครั้งที่แล้ว

ตารางที่ ๗

x	y	log x	log y	log x.log y	$(\log x)^2$	$(\log y)^2$
4	5	0.6021	0.6990	0.4209	0.3625	0.4886
6	9	0.7782	0.9542	0.7426	0.6056	0.9105
12	13	1.0792	1.1139	1.2021	1.1647	1.2408
7	11	0.8451	1.0404	0.8801	0.7142	1.0845
9	13	0.9542	1.1139	1.0629	0.9105	1.2408
4	6	0.6021	0.7782	0.4686	0.3625	0.6056
3	5	0.4771	0.6990	0.3335	0.2276	0.4886
7	12	0.8451	1.0792	0.9120	0.7142	1.1647
11	12	1.0414	1.0792	1.1239	1.0845	1.1647
4	4	0.6021	0.6021	0.3625	0.3625	0.3625
4	8	0.6021	0.9031	0.5438	0.3625	0.8156
9	12	0.9542	1.0792	1.0298	0.9105	1.1647
10	14	1.0000	1.1461	1.1461	1.0000	1.3135
7	10	0.8451	1.0000	0.8451	0.7142	1.0000
11	15	1.0414	1.1761	1.2248	1.0845	1.3832
8	11	0.9031	1.0414	0.9405	0.8156	1.0845
5	9	0.6990	0.9542	0.6670	0.4886	0.9105
6	11	0.7782	1.0414	0.8104	0.6056	1.0845
8	13	0.9031	1.1139	1.0060	0.8156	1.2408
5	7	0.6990	0.8451	0.5907	0.4882	0.7142
140	200	16.2518	19.4606	16.3133	13.7941	19.4628

แทนค่าใน Normal equation ได้

$$19.4606 = 20 \log a + 16.2518\, b \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$16.3133 = 16.2518 \log a + 13.7941\, b \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

จากสมการทั้งสองจะได้

$$\log a = 0.2824$$

$$b = 0.8499$$

สมการที่ใช้ประมาณค่าคือ

$$(\log y)_c = 0.2824 + 0.8499 \log x$$

ผลบวกกำลังสองของทั้งหมดคำนวณได้จาก

(log y)² log y/n (log y)

$$\Sigma(\log y_i)^2 = 19.4628 - (19.4606/20)(19.4606) = 0.5271$$

Explained sum of squares คือ

log a log y b log x log y

$$\Sigma(\log y)_c^2 = (0.2824)(19.4606) + (0.8499)(16.3133) - \frac{19.4604}{20}(19.4604)$$

$$= 0.4246$$

สัมประสิทธิ์คือ

$$r_{\log y . \log x} = \frac{0.4246}{0.5271} = 0.806$$

ในการประมาณค่า y จะได้ค่า log y เมื่อ x = 5, log x = 0.69897 ต้องแทนค่าดังนี้

$$(\log y)_c = 0.2824 + (0.8499)(0.69897) = 0.87645$$, anti log ของ 0.87645 คือ 7.525 ดังนั้นเมื่อ $x = 5$, $y_c = 7.525$

ถ้าข้อมูลนั้นมีความสัมพันธ์กันเป็นเส้นโค้งและไม่สามารถจะเปลี่ยนให้เป็นสมการเส้นตรงได้โดยมีสมการทั่วไปดังนี้[๓๗]

---

[37] Erwin, Kreyszig, Introductory Mathematical Statistics (New York : John Wiley & Sons, Inc., c 1970), p. 319-21.

$$y(x) = b_o + b_1x + b_2x^2 + \ldots\ldots\ldots\ldots\ldots b_mx^m$$

สมการนี้เรียกสมการโพลี่โนเมียล ถ้าสมการมีเพียง

$$y\ (x) = b_o + b_1\ x$$

เราเรียกว่า "สมการโพลีโนเมียลกำลังหนึ่ง" และถ้าสมการเป็น

$$y(x) = b_o + b_1x + b_2x^2$$

เราเรียกว่า "สมการโพลี่โนเมียลกำลังสอง" ถ้าสมการมีกำลังสูงขึ้นจะได้เส้นโค้งลักษณะต่าง ๆ กัน[๓๘] สำหรับข้อมูลที่ได้จากการทดลองหรือการสังเกต

$$y\ (x_i) = b_o + b_1x_i + b_2x_i^2 + \ldots\ldots\ldots\ldots\ldots b_mx_i^m$$

กำลังสองของส่วนที่เบี่ยงเบนตามแนวดิ่ง

$$q = \sum_{i=1}^{n} \left[(y_i - y(x_i) \right]^2$$

q จะมีค่าน้อยที่สุดเมื่อ $\frac{\partial q}{\partial b_o} = 0, \quad \frac{\partial q}{\partial b_1} = 0, \ldots\ldots\ldots \quad \frac{\partial q}{\partial b_m} = 0$

จะได้ Normal equation กำลังหนึ่ง m + 1 สมการและมีตัวพารามิเตอร์ m + 1 ตัวคือ b_o , b_1, b_2, ...b_m ตามทฤษฎีการคำนวณหาตัวพารามิเตอร์เหล่านี้ไม่มีความยุ่งยาก ในทางปฏิบัติ การคำนวณเสียเวลามาก วิธีของเกาส์ (Gauss's elimination method) และวิธีอื่น ๆ จะแสดงการหาพารามิเตอร์ในสมการโพลีโนเมียลกำลังสอง $y(x) = b_o + b_1\ x + b_2x^2$ ซึ่งจะได้ Normal equation ดังนี้

$$\begin{aligned}
\sum y_i &= b_on + b_1 \sum x_i + b_2 \sum x_i^2 \\
\sum x_iy_i &= b_o \sum x_i + b_1 \sum x_i^2 + b_2 \sum x_i^3 \\
\sum x_i^2y_i &= b_o \sum x_i^2 + b_1 \sum x_i^3 + b_2 \sum x_i^4
\end{aligned}$$

---

[๓๘] ภาคผนวก ข.

ตัวอย่าง ๙ การตายของทารกสามารถแสดงได้โดยสมการพาราโบล่าซึ่งเป็นสมการโพลี่โนเมียลกำลังสอง

การตายของทารกเป็นเปอร์เซ็นต์กับความสูงเมื่อแรกเกิด

ตารางที่ ๘

ความสูง		จำนวนทารกในชั้นนี้	จำนวนทารกที่ตายในชั้นนี้	การตายเป็นเปอร์เซ็นต์ (y)
อันตรภาคชั้น	จุดกึ่งกลางชั้น (x_i)			
๔๖ – ๔๙	๔๗.๕	๑๒๔	๑๔	๑๑.๒๙
๔๙ – ๕๒	๕๐.๕	๑๒๕๕	๑๓	๑.๐๔
๕๒ – ๕๕	๕๓.๕	๓๑๔๙	๒๘	๐.๘๙
๕๕ – ๕๘	๕๖.๕	๑๔๔๑	๓๑	๒.๑๕
๕๘ – ๖๑	๕๙.๕	๑๗๕	๑๖	๙.๑๔

เพื่อให้สดวกในการคำนวณให้ $x_i = x_i^* + 53.5$

$$x_i^* = x_i - 53.5$$

$$\Sigma x_i^* = 124(-6) + 1255(-3) + 1441(3) + 175(6) = 864$$

$$\Sigma x_i^{*2} = 124(36) + 1255(9) + 1441(9) + 175(36) = 35,028$$

$$\Sigma x_i^{*3} = 124(-216) + 1255(-27) + 1441(27) + 175(216) = 16,038$$

$$\Sigma x_i^{*4} = 124(1296) + 1255(81) + 1441(81) + 175(1296) = 605,880$$

กลุ่มตัวอย่างนี้ประกอบด้วย $n = 6144$ ใช้การคำนวณตามข้างบนจะได้

$$\Sigma y_i = 10,205.42$$

$$\Sigma x_i^* y_i = 6,576.09$$

$$\Sigma x_i^{*2} y_i = 147,610.71$$

แทนค่าในสมการปกติจะได้

(1) .... $10,205.42 = 6144\ b_o + 864\ b_1 + 35,028\ b_2$

(2) .... $6,576.09 = 864\ b_o + 35,028\ b_1 + 16,038\ b_2$

(3) .... $147,610.71 = 35,028\ b_o + 16,038\ b_1 + 605,880\ b_2$

ขั้นแรกทำให้ b_o หมดไปจาก (๒) โดยคูณ (๑) ด้วย ๘๖๔/๖๑๔๔ จะได้

$$864b_o + 121.500b_1 + 4925.813b_2 = 1435.137$$

ลบจากสมการ (๒) จะได้

(4)......... $34,906.500b_1 + 11,112.187b_2 = 5140.953$

กำจัด b_o จากสมการ (๓) โดยคูณ (๑) ด้วย ๓๕,๐๒๘/๖๑๔๔ จะได้

$$35,028b_o + 4925.813b_1 + 199,700.64\ b_2 = 58,182.853$$

ลบจาก (๓) จะได้

(5)....... $11,112.187\ b_1 + 406,179.352b_2 = 89,427.857$

ขั้นที่ ๒ กำจัด b_1 จาก (๕)

คูณ (๔) ด้วย ๑๑, ๑๑๒,๑๘๗/๓๔,๙๐๖.๕๐๐ จะได้

(6) .... $11, 112, 187b_1 + 3537.470b_2 = 1636.579$

ลบจาก (๕) จะได้

(7).......$402,641.882b_2 = 87,791.278$

$$b_2 = \frac{87,791.278}{402,641.882} = 0.218038$$

แทนค่า b_2 ใน (๔)

$$b_1 = \frac{1}{34,906.5}(5140.953 - 11,112.187b_2)$$

$$= 0.077867$$

แทนค่า b_2 ใน (๑)

$$b_o = \frac{1}{6144}(10,205.42 - 864b_1 - 35,028b_2)$$

$$= 0.407016$$

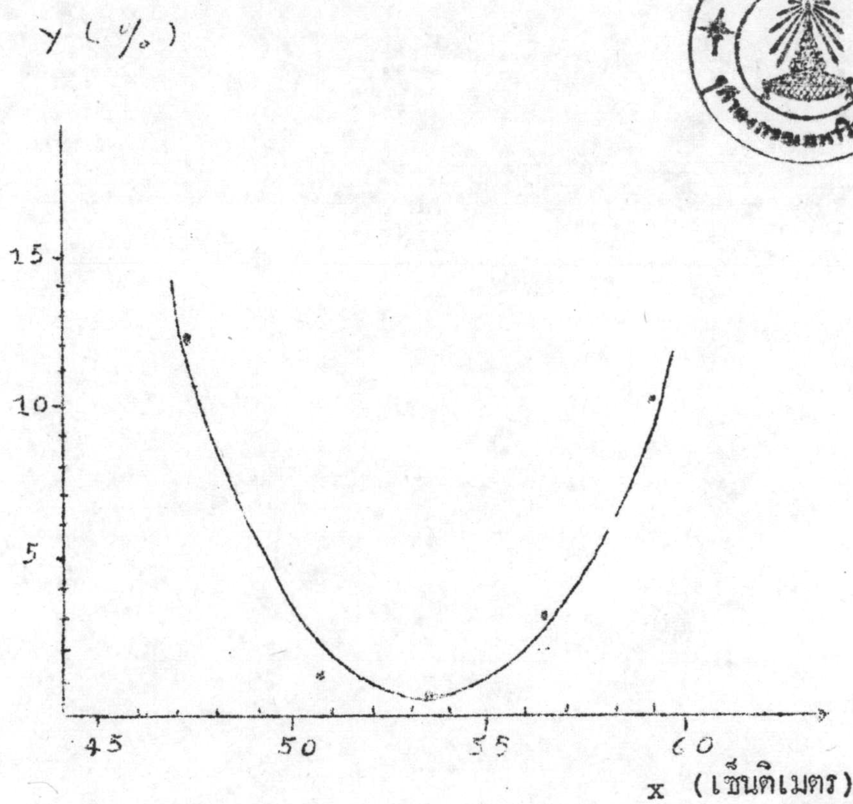
ดังนั้นสมการพาราโบล่าจะเป็น

$$y = 0.407 + 0.078x^* + 0.218x^{*2}$$

แทนค่า $x^* = x - 53.5$ จะได้

$$y = 620.205 - 23.248x + 0.218x^2$$

รูปที่ ๑๒



การประมาณค่าโดยวิธีกำลังสองน้อยที่สุดสำหรับโพลี่โนเมียลที่กำลังมากกว่า ๒ การหาสมการปกติจะมีเทคนิคแตกต่างกันออกไป[๓๙]

ให้ $\phi_j(x)$ เป็นโพลี่โนเมียลกำลัง j

$$\hat{f}_i \approx \sum_{j=0}^{m} a_j^{(m)} \phi_j(x_i) \qquad i = 1, \ldots\ldots\ldots\ldots n$$

$a_j^{(m)}$ หาได้จาก

$$H(a_o^{(m)}, \ldots\ldots\ldots a_m^{(m)}) = \sum_{i=1}^{n} W(x_i) \left[\hat{f}_i - \sum_{j=0}^{m} a_j^{(m)} \phi_j(x_i) \right]^2$$

$W(x)$ เรียก "Weight function" $w(x_i) \geq 0$

เราจะประมาณค่า

$$y_m(x) = \sum_{j=0}^{m} a_j^{(m)} \phi_j(x)$$

เมื่อ $w(x) = 1$

$$\frac{\partial H}{\partial a^{(m)}} = -2 \sum_{i=1}^{n} w_i \left[\hat{f} - \sum_{j=0}^{m} a_j^{(m)} \phi_j(x_i) \right] \phi_k(x_i) = 0$$

$$k = 0, \ldots\ldots\ldots m, \qquad w_i = w(x_i)$$

$$(1) \ldots\ldots\ldots \sum_{i=1}^{n} \left[\hat{f}_i - \sum_{j=0}^{m} a_j^{(m)} x_i^j \right] x_i^k = 0, \qquad k = 0 \ldots\ldots m$$

$$(2) \ldots\ldots\ldots \sum_{j=0}^{m} a_j^{(m)} \left(\sum_{i=1}^{n} x_i^{j+k} \right) = \sum_{i=1}^{n} \hat{f}_i x_i^k$$

---

39 Anthony, Ralston, A First Course in Numerical Analysis (New York : Mc Graw-Hill Book Company, c 1965), p. 232.

ใช้สัญญลักษณ์ $g_{ik} = \sum_{i=1}^{n} x^{j+k}$ และ $\rho_k = \sum_{i} \hat{f}_i \, x_i^{k}$

สมการปกติจะเขียนได้ดังนี้

$$(3) \,\ldots\ldots\ldots \sum_{j=0}^{m} g_{ik} \, a_j^{(m)} = \rho_k \qquad k = 0 \,\ldots\ldots\ldots\ldots m$$

การแก้สมการกำลังสองน้อยที่สุดกรณี $\phi_j(x) = x^j$ $w(x) = 1$ สำหรับค่า m น้อย ๆ ถึง ๕ หรือ ๖ ผลลัพธ์ของสมการ (๓) จะให้ตัวประมาณค่าที่ดี ถ้า m มาก ๆ ผลลัพธ์จากสมการ (๓) จะไม่ดี

การเลือกกำลังของโพลี่โนเมียล

กำหนดค่า n ให้จะเลือก m กำลังของการประมาณค่าโพลี่โนเมียล สมมุติฐานเบื้องต้นของเราคือฟังค์ชั่นจริง f(x) เป็นโพลี่โนเมียลที่มีกำลัง M น้อยกว่า n ถ้าเราเลือก m น้อยกว่า M จะไม่สามารถแทนฟังค์ชั่นจริงได้ จะต้องเลือก m = n - 1 เราได้

$$(4) \ldots\ldots \delta_m^2 = \sum_{i=1}^{n} w_i R_i^2 = \sum_{i=1}^{n} w_i \left[f_i^2 - \sum_{j=0}^{m} a_j^{(m)} x_i^j \right]^2$$

มีค่าเท่ากับศูนย์ ค่าใด ๆ ของ m มากกว่า M จะทำให้กำลังของ x ตาม (๔) ได้

$$(5) \ldots\ldots\ldots f(x) = \sum_{j=0}^{m} a_j^{(m)} x^j$$

เมื่อเรารู้ค่า M จะประมาณค่ากำลังสองน้อยที่สุดได้ดังนี้

$$(6) \ldots\ldots\ldots\ldots y_{M+1}(x) = \sum_{j=0}^{M+1} a_j^{(m+1)} x^j$$

ข้อมูลที่ได้จากการสังเกต $\{ \bar{f}_i \}$ ค่าสถิติ $a_{(M+1)}^{(M+1)}$ จะเป็นศูนย์ นั่นคือข้อมูลไม่ผิดพลาด ถ้ามีสมมุติฐานว่า f(x) มีรูปตาม (๕) เราจะทดสอบว่า $a_{M+1}^{(M+1)} = 0$ และมีข้อตกลงเบื้องต้นต่อไปว่า ค่าผิดพลาด E_i กระจายเป็นโค้งปกติโดยมีมัชฌิม = ๐ และความแปรปวนเท่ากับ $\frac{\sigma^2}{w_i}$ ถ้าสมมุติฐานถูกต้องเราจะได้ว่า

(7)......... $\sigma_m^2 = \delta_m^2/(n-m-1)$

จะไม่ขึ้นกับ m เมื่อ m = M, M + 1, ...n - 1 ดังนั้นในทางปฏิบัติถ้าเราไม่รู้ค่า M จะหาสมการทั่วไป (๓) สำหรับ m = 1, 2, ...

ถ้า $p_j(x)$ เป็นโพลี่โนเมียลกำลัง j การประมาณค่ากำลังสองน้อยที่สุดกำลัง m จะเขียนได้ดังนี้

(8) ....$y_{(m)}(x) = \sum_{j=0}^{m} b_j^{(m)} \, p_j(x)$

ในทางปฏิบัติการแก้ปัญหาเรามักจะเริ่มจากค่า m > 1 เพื่อที่จะทำให้ H มีค่าน้อยที่สุดเราได้

(9)......... $H = \left[b_0^{(m)}, \ldots\ldots\ldots, b_m^{(m)} \right] = \sum_{i=1}^{n} w_i \left[\hat{f} - y_m (x_i) \right]^2$

และได้ว่า

(10)......... $\sum_{j=0}^{m} d_{jk} \, b_j^{(m)} = w k$, $k = 0, \ldots\ldots\ldots m$

เมื่อ $d_{jk} = \sum_{i=1}^{n} w_i p_j(x_i) \, p_k(x_i)$

(11)........ $w_k = \sum_{i=1}^{n} w_i \bar{f_i} \, p_k (x_i)$

ถ้า $\{ p_j(x) \}$ เป็นออร์ทอโกนัลตลอดกลุ่มของจุด $\{ x_i \}$

(12) .... $\sum_{i=1}^{n} w_i \, p_j^{(m)} (x_i) p_k^{(m)} (x_i) = 0$ **ถ้า** $j \neq k$

สมมุติว่า $w (x_i) > 0$ ทุกค่าของ i ถ้า $\{ p_j^{(n)} (x) \}$ เป็นออร์ทอโกนัล $d_{jk} = 0, j \neq k$ จะได้

(13) ....$d_{kk} b_k^{(m)} = w_k$, $k = 0, \ldots\ldots\ldots, m$

จะให้ผลดังนี้

(14) .... $b_k^{(m)} = w_k / d_{kk}, \quad k = 0, \ldots\ldots, m$

เป็นการแก้ปัญหาข้อบกพร่องของ Normal equation และผลลัพธ์เมื่อแทนด้วย m + 1 จะได้ดังนี้

(15) .... $b_k^{(m+1)} = w_k / d_{kk}, \quad k = 0, \ldots\ldots, m$

แสดงว่าการใช้ออทอโกนัลโพลี่โนเมียลสามารถแก้ปัญหาความยุ่งยากในการหาสมการปกติสำหรับโพลี่โนเมียลกำลังสูง ๆ ได้ จากสัญญลักษณ์ที่กำหนดไว้โดยให้มี N + 1 จุดเท่ากัน $x_i = x_o + ih$, $i = 0, \ldots\ldots, N$ สมการ (๘) เขียนใหม่ได้ดังนี้

(16)........ $\sum_{i=0}^{N} w_i \, P_j^{(N+1)} (x_o + ih) p_k^{N+1} (x_o + ih) = 0, \quad j \neq k$

เพื่อความสะดวกเราแทน $p_j^{(N+1)} (x_o + ih)$ โดย $p_j (i, N)$ เราให้

(17)........ $w_i \, p_i (i, N) = \Delta^j U_j (i,N)$

$U_j (i, N)$ เป็นฟังค์ชั่นที่ต้องการหา ถ้า $q_{j-1} (x)$ เป็นโพลี่โนเมียลกำลัง $j - 1$ ใด ๆ สมการ (๑๖) จะเท่ากับ

(18) .... $\sum_{i=1}^{N} q_{j-1} (i) \; \Delta^j \; U_j (i, N) = 0$

ในการประมาณค่ากำลังสองน้อยที่สุดโดยข้อมูลมีการกระจายเท่ากัน เป็นการสะดวกที่จะใช้จำนวนจุดเป็นเลขคี่ และให้จุดกึ่งกลางเป็นศูนย์ ดังนั้นถ้าเราให้ N = 2L และกำหนดตัวแปร S ใหม่ ซึ่ง i = L + S สมการ

(19)......... $p_j (i, N) = B_{jN} \sum_{k=0}^{j} (-1)^k \dfrac{(j + k)^{(2k)} \; i^{(k)}}{(k\,!)^2 \; N^{(k)}}$

$$B_{jN} = (-1)^{i} \; j! \; N^{(j)} \; A_{jN}$$

สมการ (๑๙) เมื่อกำหนดช่วงตัวแปรนี้จะได้

$$(20) \ldots\ldots\ldots p_i(S, 2L) = \sum_{k=0}^{j} (-1)^{k+j} \frac{(j+k)^{(2k)} \; (L+S)^{(k)}}{(k!)^{2} \; (2L)^{(k)}}$$

โพลีโนเมียล p_i (S, 2L) เป็นที่รู้จักกันว่าเป็น Gram polynomials เพื่อความสดวกในการหา p_i (S, 2L) จะเขียนเป็นตารางจาก

$$(21) \ldots\ldots\ldots p_j(S, 2L) = \epsilon_j \sum_{k=0}^{j} \delta_{jk} \, S^{k}, \quad \delta_{ij} = 1$$

สัมประสิทธิ์ใน (๒๑) สำหรับ j = 1, 2, 3, 4, 5 แสดงในตารางข้างล่างทุกค่าของ δ_{jk} ที่ไม่เท่ากับศูนย์

ตารางที่ ๔

ตารางสัมประสิทธิ์ของ p_j (S, 2L) ตามที่หาได้จากสูตร (๒๑)	
$j = 1, \; \epsilon_1 = 1/L$	
$j = 2, \; \epsilon_2 = 3/L\,(2L-1),$	$\delta_{20} = -\dfrac{L(L+1)}{3}$
$j = 3, \; \epsilon_3 = 5/L\,(L-1)(2L-1),$	$\delta_{31} = \dfrac{-3L^2 + 3L - 1}{5}$
$j = 4, \; \epsilon_4 = \dfrac{35}{2L(L-1)(2L-1)(2L-3)},$	$\delta_{40} = \dfrac{3L(L^2-1)(L+2)}{35}$
	$\delta_{42} = \dfrac{-6L^2 + 6L - 5}{7}$
$j = 5, \; \epsilon_5 = \dfrac{63}{2L(L-1)(L-2)(2L-1)(2L-3)},$	$\delta_{51} = \dfrac{15L^4 + 30L^3 - 35L^2 - 50L + 12}{63}$
	$\delta_{52} = \dfrac{-5(2L^2 + 2L - 3)}{9}$

จาก Gram polynomial จะได้ว่า

$$(22) \ldots\ldots \sum_{S=-L}^{L} p_i (S, L)\, p_j (S, 2L) = 0 \qquad i \neq j$$

ดังนั้นถ้าเราให้ค่าของฟังค์ชั่น f (S) ที่ S = -L, -L + 1, ... 0, 1, 2, ...L การประมาณค่ากำลังสองน้อยที่สุดของ f (S) ใช้สมการ (๘)

$$(23) \ldots\ldots y_m(S) = \sum_{j=0}^{m} b_j\, p_i (S, 2L)$$

$$(24) \ldots\ldots b_j = w_j / \gamma_j$$

$$(25) \ldots\ldots w_j = \sum_{S=-L}^{L} \left[\bar{f}S, \quad p_j (S, 2L) \right]$$

และ $\gamma_j = \delta_{jj} = \sum_{S=-L}^{L} p_j^2 (S, 2L) = \dfrac{(2L + J + 1)!\ (2L - j)!}{(2j + 1) \left[(2L)! \right]^2}$

การเปลี่ยนตัวแปรให้อยู่ในเทอม x ใช้ $S = \left(\dfrac{x - x_o}{h} \right) - L$

<u>ตัวอย่าง ๑๐</u> สมมุตว่าข้อมูลที่ได้เป็นดังนี้ [๔๐]

x_i	.1	.2	.3	.4	.5	.6	.7	.8	.9
f_i	5.1234	5.0357	5.5687	5.9378	6.4370	7.0978	7.9493	9.0253	10.3627

เราต้องการจะประมาณค่าโพลิโนเมียลกำลังสองน้อยที่สุดสำหรับ f(x) เมื่อจุดต่าง ๆ กระจายเท่ากัน ให้จุดกำเนิดใน S โคออนิเนทอยู่ที่ x = 0.5 เราได้

x = .5 + .1S ....(1)

---

40 <u>I did.</u>, p. 243.

ใช้ตารางที่ได้จากสูตร (๒๑) โดยไม่ใช้ 2L จะได้ว่า

$p_o(S) = 1$ $\qquad p_1(S) = \frac{S}{4}$

$p_2(S) = (3S^2 - 20)/28$ $\qquad p_3(S) = (5S^3 - 59S)/84$ ....(2)

$p_4(S) = (7S^4 - 115S^2 + 216)/168$ $\qquad p_5(S) = (9S^5 - 185S^3 + 716S)/240$

ตารางข้างล่างจะใช้ค่า $p_j(S)$, j = 0, 1, 2..........5 หาค่าของ S ที่ตรงกับค่าของ γ_j

ตารางค่าของ $p_j(S)$

ตารางที่ ๑๐

S	p_o	p_1	p_2	p_3	p_4	p_5
-4	1	-1	1	-1	1	-1
-3	1	$-\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	-3/2	11/4
-2	1	$-\frac{1}{2}$	-2/7	13/14	-11/14	-1
-1	1	$-\frac{1}{4}$	-17/28	9/14	9/14	-9/4
0	1	0	-5/7	0	9/7	0
1	1	$\frac{1}{4}$	-17/28	-9/14	9/14	9/4
2	1	$\frac{1}{2}$	-2/7	-13/14	-11/14	1
3	1	$\frac{3}{4}$	$\frac{1}{4}$	$-\frac{1}{2}$	-3/2	-11/4
4	1	1	1	1	1	1
γ_j	9	15/4	99/28	495/98	143/14	117/4

จากสูตร (๒๔) และ (๒๕) จะคำนวณได้ว่า

$$
\begin{aligned}
w_o &= 62.80770 & \qquad b_o &= 6.97863 \\
w_1 &= 9.50930 & b_1 &= 2.53581 \\
w_2 &= 2.69424 & b_2 &= .76201 \quad \ldots\ldots\ldots\ldots\ldots(3) \\
w_3 &= .42323 & b_3 &= .08379 \\
w_4 &= .02449 & b_4 &= .00240 \\
w_5 &= .00100 & b_5 &= .00003
\end{aligned}
$$

ใช้ผลจากสมการ (๒) + (๓) เราจะแทนค่าในสูตร (๔) คือ

$$\delta_m^2 = \sum_{S=-4}^{4} \left(\hat{f}_s - \sum_{j=0}^{m} b_j p_j(S) \right)^2 \quad \ldots\ldots\ldots(4)$$

จะหา δ_m^2 ได้ดังนี้

$$\delta_0^2 = 26.202 \qquad \delta_1^2 = 2.089 \qquad \delta_2^2 = .0355 \quad \ldots\ldots\ldots\ldots(5)$$

$$\delta_3^2 = .000059 \qquad \delta_4^2 = .0000049$$

$$\delta_5^2 = .00000045$$

จากสูตร (๗) เราจะคำนวณได้ดังนี้

$$\zeta_0^2 = 3.275 \qquad \zeta_1^2 = .289 \qquad \zeta_2^2 = .0059 \quad \ldots\ldots\ldots\ldots(6)$$

$$\zeta_3^2 = .000012 \qquad \zeta_4^2 = .00000012 \qquad \zeta_5^2 = .00000015$$

จะสรุปได้ว่า m = 4 จะให้ค่าประมาณกำลังสองน้อยที่สุดที่ดีที่สุดคือ

$$y_4(S) = \sum_{j=0}^{4} b_j p_j(S) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

จากค่า b_j ที่ได้ตามสมการ (๓) และ $p_j(S)$ ตามสมการ (๒) ใช้สมการ (๑) และ (๒) เปลี่ยนกลับเป็นค่า x และได้

$$y_4(x) = 1.0000x^4 + 2.9875x^3 + 2.0188x^2 + .9915x + 5.0010 \ldots\ldots. (8)$$

๔.๑๒ การใช้วิธีกำลังน้อยที่สุดวิเคราะห์อนุกรมเวลา (Time Series Analysis)

ถ้าตัวแปรอิสสระ x เป็นเวลา ข้อมูลที่แทนด้วยค่า y จะเปลี่ยนแปลงตามเวลา ลักษณะของข้อมูลเช่นนี้เรียก "อนุกรมเวลา (Time Series)" เส้นตรงหรือเส้นโค้งที่แสดงความสัมพันธ์ระหว่าง y และ x เรียก "แนวโน้ม (Trend line or Trend curve)" ซึ่งจะใช้ในการประมาณค่าหรือทำนายหรือคาดการณ์ล่วงหน้า เพื่อให้ทราบปัญหาการใช้วิธีกำลังสองน้อยที่สุดวิเคราะห์อนุกรมเวลา ควรทราบคุณสมบัติ ลักษณะ ช่วงเวลาที่มีการเปลี่ยนแปลงครบรอบและความสม่ำเสมอในการเปลี่ยนแปลงของอนุกรมเวลาแบบต่าง ๆ ดังนี้[๔๑]

๑. แนวโน้มระยะยาว (Secular or Long-time trends) ปรากฏการณ์ทางสังคมศาสตร์มักมีแนวโน้มที่จะเพิ่มหรือลดไปทางเดียวกันในช่วงเวลาค่อนข้างนาน

๒. การเปลี่ยนแปลงเป็นคาบ (Periodic fluctuations) แบบที่พบกันทั่วไปคือปรากฏการณ์ทางสังคมศาสตร์ที่มีการเปลี่ยนแปลงตามฤดูกาล มีค่าสูงสุดและต่ำสุดสม่ำเสมออาจเป็นช่วงวันหรือรอบสัปดาห์

๓. การเปลี่ยนแปลงแบบวัฏจักร (Undulatory or Cyclical movement) การเปลี่ยนแปลงลักษณะนี้เหมือนคลื่นแต่ไม่มีระยะเวลาแน่นอนเช่นวงจรธุรกิจมีการเปลี่ยนแปลงระหว่างเศรษฐกิจตกต่ำและเศรษฐกิจเฟื่องฟู

๔. การเปลี่ยนแปลงที่ไม่แน่นอน (Irregular Variation) การเปลี่ยนแปลงแบบนี้อาจเกิดเป็นครั้งคราวหรือเกิดโดยบังเอิญ เนื่องจากแฟคเตอร์ต่าง ๆ เช่น การนัดหยุดงาน, การปิดโรงงาน การเกิดอัคคีภัย, แผ่นดินไหว หรือการจลาจล ทั้งที่เกิดจากบุคคลและ

---

[41] Pau-Line V. Young, Scientific Social Survey and Research (4th ed; New Jersey : Prentice-Hall, Inc., c 1939), pp. 342-43.

ธรรมชาติ การเปลี่ยนแปลงแบบนี้จะเป็นผลให้หยุดชะงักการเปลี่ยนแปลงตามปกติ และทำให้ไม่สามารถจะเห็นแนวโน้มในช่วงต่อไป การเปลี่ยนแปลงโดยบังเอิญนี้จะพบน้อยกว่าการเปลี่ยนแปลงแบบอื่น และวิเคราะห์หาสาเหตุได้ยาก

การประมาณค่าแนวโน้มทำได้หลายวิธี[๔๒]

๑. วิธีกำลังสองน้อยที่สุด ใช้หาสมการเส้นตรงหรือเส้นโค้งเพื่อใช้ประมาณค่าแนวโน้ม

๒. เขียนเส้นโดยประมาณด้วยสายตา ซึ่งจะไม่แน่นอนขึ้นกับการตัดสินใจของแต่ละบุคคล

๓. วิธีเฉลี่ยในช่วงเวลาต่าง ๆ กัน (The Moving Average Method) วิธีนี้ทำให้การเปลี่ยนแปลงแบบวัฏจักร และการเปลี่ยนแปลงที่ไม่แน่นอนหมดไป ข้อเสียคือขาดข้อมูลจุดเริ่มต้นและจุดสุดท้ายของอนุกรม

๔. วิธีเฉลี่ยแบ่งครึ่ง (The Method of Semi-averages) แบ่งข้อมูลเป็นสองส่วน แล้วลากเส้นผ่านจุดทั้งสอง ซึ่งให้ผลไม่ค่อยดีนักและใช้ได้กับเส้นตรงเท่านั้น

วิธีกำลังสองน้อยที่สุดเป็นวิธีที่ใช้ประมาณค่าแนวโน้มได้ดีที่สุดและใช้กับสมการได้หลายแบบเช่น

---

[42] Spiegel, Opcit., p. 286

<u>ตัวอย่าง ๑๑</u> แสดงแนวโน้มระยะยาวของอัตราการขาย เของบริษัทและห้างร้านต่าง ๆ ใน สหรัฐอเมริการะหว่างปี ค.ศ. ๑๙๔๑ ถึง ๑๙๕๙ ค่า y ใช้หน่วยเป็นล้านดอลลาร์[๔๓]

<u>ตารางที่ ๑๑</u>

ปี ค.ศ.	y	x	x^2	xy	y_c	อัตราการลด y/y_c
1941	110	-9	81	-990	118.1	93.1
1942	128	-8	64	-1,024	129.3	99.0
1943	137	-7	49	-959	140.5	97.5
1944	148	-6	36	-888	151.7	97.6
1945	154	-5	25	-770	162.9	94.5
1946	180	-4	16	-720	174.1	103.4
1947	194	-3	9	- 582	185.3	104.7
1948	209	-2	4	-418	196.5	106.4
1949	209	-1	1	-209	207.7	100.6
1950	216	0	0	0	218.9	98.7
1951	238	+1	1	+238	230.1	103.3
1952	250	+2	4	+500	241.3	103.6
1953	258	+3	9	+774	252.5	102.2
1954	252	+4	16	+1,008	263.7	95.6
1955	275	+5	25	+1,375	274.9	100.0
1956	285	+6	36	+1,710	286.1	99.6
1957	294	+7	49	+2,058	297.3	98.9
1958	301	+8	64	+2.408	308.5	97.6
1959	321	+9	81	+2,889	319.7	100.4
	4,159		570	+12,960	4,159.1	
				- 6,560		
				+ 6,400		

43 John I., Griffin, <u>Statistics method and Applications</u> (New York : Holt, Richart and Winston, c 1962), pp. 327-29.

แทนค่าใน Normal equation ของสมการสองตัวแปรจะได้

$$a = 4,159/19 = 218.9$$

$$b = + 6,400/570 = + 11.2$$

สมการแสดงแนวโน้มจะเขียนได้ดังนี้

$$y = 218.9 + 11.2\,x$$ เมื่อจุดกำเนิดอยู่ที่ปี ๑๙๕๐

สมการนี้จะหาแนวโน้มในปีต่าง ๆ ได้เช่น ปี ค.ศ. ๑๙๔๑ จะได้

$$y_c = 218.9 + 11.2(-9) = 118.1$$

จากสมการจะเห็นว่าอัตราการเพิ่มต่อปี = ๑๑.๒ ถ้าต้องการจะลดแนวโน้มลงจะหาอัตราการลดได้จาก y/y_c ดังแสดงในตาราง ในทางปฏิบัติมักใช้จำนวนปีเป็นเลขคี่ถ้าต้องการหาแนวโน้มระยะยาวอนุกรมสั้น ๆ จำเป็นต้องใช้จำนวนปีเป็นเลขคู่และค่าของ x ใช้เป็นหน่วยครึ่งปี เปลี่ยนจุดกำเนิดใหม่เป็น –๑, –๓, –๕ และ +๑, +๓, +๕, เมื่อแสดงหลักการนี้ได้นำข้อมูล ๖ ปี ท้ายมาหาแนวโน้มดังตาราง

ตารางที่ ๑๒

ปี ค.ศ.	y	x	x^2	xy	y_c	y/y_c
1954	252	-5	25	-1,260	257.0	98.0
1955	275	-3	9	- 825	269.4	102.0
1956	285	-1	1	- 285	281.9	101.1
1957	294	+1	1	+ 294	294.2	99.9
1958	301	+3	9	+ 903	306.6	98.2
1959	321	+5	25	+1,605	319.0	100.6
	1,728		70	+2,802	1,728.1	
				-2,370		
				+ 432		

แทนค่าใน Normal equations จะได้

$$a = 1,728/6 = 288$$

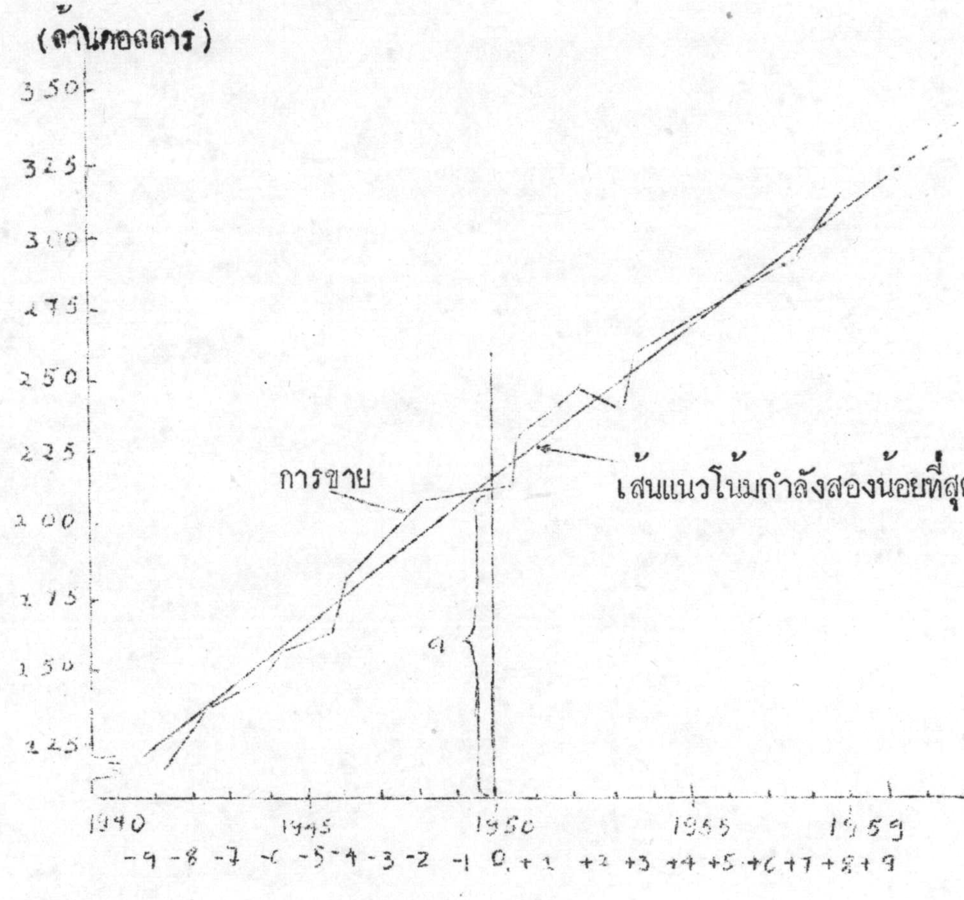
$$b = +432/70 = + 6.2$$

สมการแสดงแนวโน้มกำลังน้อยที่สุดคือ $y_c = 288 + 6.2x$ จุดกำเนิดอยู่ที่ปี ค.ศ. ๑๙๕๗

แนวโน้มที่เพิ่มขึ้นต่อปี = 2b หรือ ๑๒.๔

กราฟแสดงแนวโน้มระยะยาว

รูปที่ ๑๓



ในการวิเคราะห์อนุกรมเวลามักจะพบแนวโน้มที่เป็นเส้นตรงหรือโพลี่โนเมียลกำลังไม่เกิน ๓ ซึ่งสามารถจะใช้วิธีกำลังสองน้อยที่สุดได้ทั้งสิ้น ถ้าโพลี่โนเมียลกำลังมากกว่า ๒ ใช้ออทอโกนัลโพลี่โนเมียลสดวกกว่าและยังลดความยุ่งยากในการคำนวณเมื่อจะเพิ่มหรือจะลดกำลังของโพลี่โนเมียลจากตอนเริ่มต้น นอกจากนี้แนวโน้มที่อยู่ในรูปของกำลัง (Exponential equation)[๔๔] จะเกิดขึ้นในอนุกรมเวลาทางธุรกิจ, สังคม, เศรษฐศาสตร์และชีววิทยา อนุกรมเหล่านี้จะเพิ่มอัตรากำลังมากกว่าเป็นจำนวนเท่า เช่นการลงทุนจะมีการเพิ่มตามจำนวนปี x อัตราการเพิ่ม ๔% จะได้ว่า

$$\text{ต้นทุนปัจจุบัน} = (\text{ต้นทุนเดิม})(๑.๐๔)^x$$

สมการทั่วไปอยู่ในรูป

$$y = ab^x$$

$$\log y = \log a + x \log b$$

ตัวอย่าง ๑๒ ข้อมูลข้างล่างอยู่ในรูปของกำลัง

ตารางที่ ๑๓

x	y_c	$\log y_c$
0	10.0	1.000
1	15.0	1.176
2	22.5	1.352
3	33.75	1.528
4	50.625	1.704
5	75.9375	1.880

---

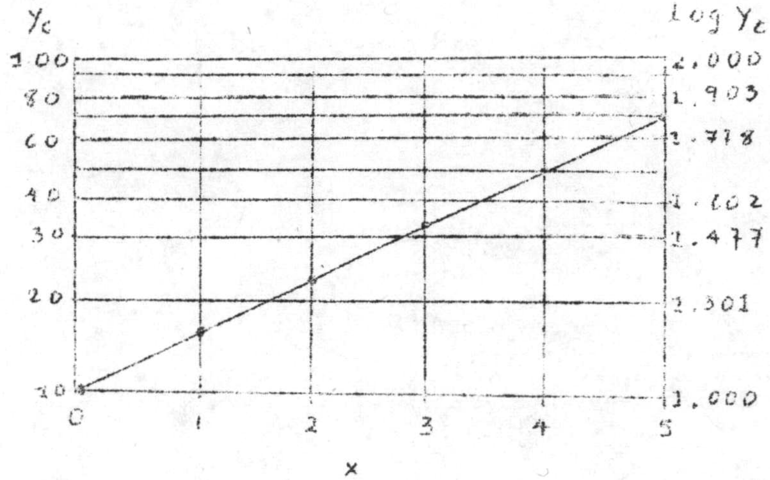
44 Samuel B., Richmond, Statistical Analysis (2 d. ed., New York : The Ronald Press Company, c 1964), pp. 368-70.

ตารางข้างบนได้จากสมการ $y_c = 10\ (1.5)^x$ เมื่อเปลี่ยนค่า y_c เป็นค่า log แล้วจะได้สมการเส้นตรงซึ่ง สมการที่ได้จะเป็น

$$\log y_c = 1.000 + 0.176x$$

นำค่า x และ y_c มาพลอตต์ในกระดาษ Semi-logarithm จะได้ดังนี้

รูปที่ ๑๔



สูตรและการคำนวณเหมือนกับในหัวข้อ ๘.๑๑ และหลักในการเปลี่ยนจุดกำเนิดเช่นเดียวกับอนุกรมเวลาแบบเส้นตรงเช่นถ้าเดิมจุดกำเนิดเป็นปี ค.ศ. ๑๙๕๘ และหน่วยของ x = ๑ ปี ต้องการจะเลื่อนจุดกำเนิดไป ๒ ปีเป็นปี ค.ศ. ๑๙๖๐ จากสมการเดิม

$$y_c = (10)(1.5)^x \qquad \text{จะเปลี่ยนเป็น}$$

$$y_c = (10)(1.5)^{x+2} = (10)(1.5)^x\ (1.5)^2$$

$$y_c = (22.5)(1.5)^x$$

<u>ตัวอย่าง ๑๓</u> แสดงการหาแนวโน้มของอนุกรมเวลาซึ่งเหมาะกับสมการกำลังสองมากที่สุด ข้อมูลที่ใช้เป็นตัวอย่างคือรายได้จากผู้โดยสารขององค์การรถไฟนอกเมืองและชานเมืองอเมริการะหว่างปี ค.ศ. ๑๙๐๕ ถึง ๑๙๔๕ ซึ่งเป็นระยะเวลาที่มีทางรถไฟเพิ่มขึ้นอย่างรวดเร็วและลดลงในสงครามโลกครั้งที่สอง แนวโน้มลักษณะนี้ใช้พาราโบล่าจึงจะสมเหตุผล ได้แสดงการคำนวณ

เปรียบเทียบกับแนวโน้มแบบเส้นตรงด้วย[๔๕]

ตารางที่ ๑๔

ปี ค.ศ.	y (ล้าน)	x	x^2	x^4	xy	x^2y	แนวโน้มเส้นตรง	a+bx	cx^2	แนวโน้มกำลังสอง
1905	6.0	-4	16	256	-24.0	96.0	9.12	10.7	-3.8	6.9
1910	8.6	-3	9	81	-25.8	77.4	9.14	10.7	-2.2	8.5
1915	9.9	-2	4	16	-19.8	39.6	9.16	10.8	-1.0	9.8
1920	12.3	-1	1	1	12.3	12.3	9.18	10.8	-0.2	10.6
1925	21.1	0	0	0	0	0	9.20	10.8	-	10.8
1930	10.4	+1	1	1	+10.4	10.4	9.22	10.8	-0.2	10.6
1935	7.4	+2	4	16	+14.8	29.6	9.24	10.8	-1.0	9.8
1940	6.5	+3	9	81	+19.5	58.5	9.26	10.9	-2.2	8.7
1945	9.6	+4	16	256	+38.4	153.6	9.28	10.9	-3.8	7.1
	82.8		60	708	+83.1	477.4	82.80			82.8
					-81.9					
					+ 1.2					

45 Richmond , Op.cit., pp. 330 - 32.

จะได้ Normal equation ดังนี้

$$\Sigma y = na + c \Sigma x^2$$

$$b = \frac{\Sigma xy}{\Sigma x^2}$$

$$\Sigma x^2 y = a \Sigma x^2 + c \Sigma x^4$$

จากการคำนวณจะหา a, b และ c แทนค่าในสมการ $y_c = a + bx + cx^2$

$$a = 10.8, \quad b = 0.02, \quad c = -0.24$$

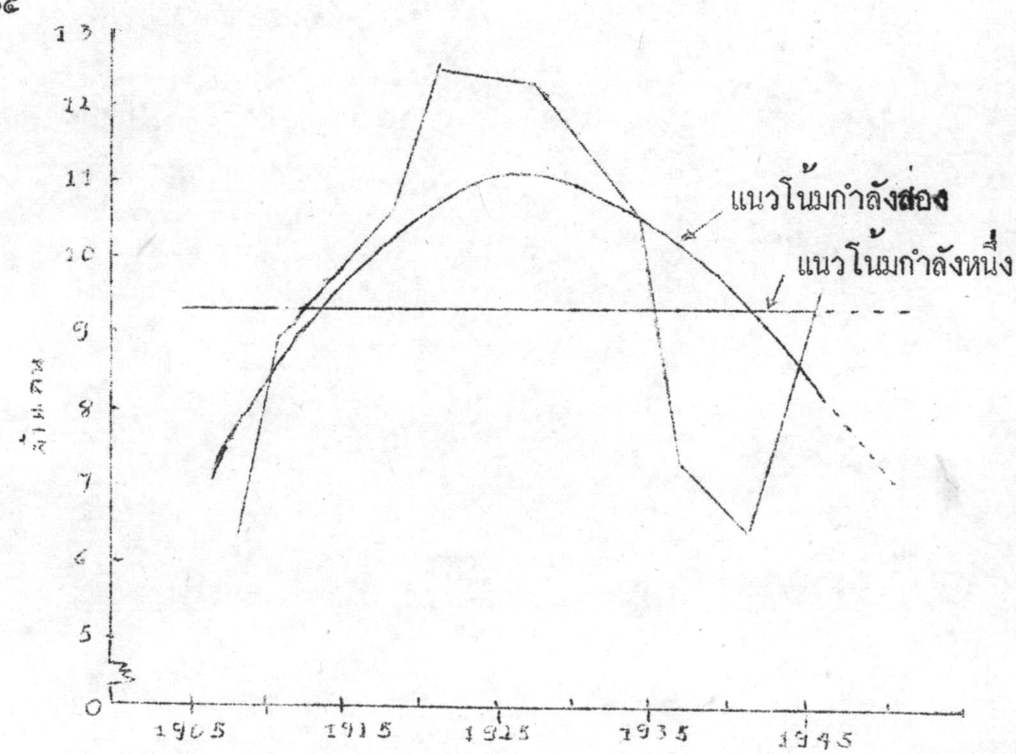
จะได้ $y_c = 10.8 + 0.02x - 0.24x^2$ ....(1)

แนวโน้มเส้นตรงได้ $a = \frac{82.8}{9} = 9.2$ และ $b = +\frac{1.2}{60} = +0.2$

$y_c = 9.2 + 0.02x$ ....(2)

สมการ (๑) เป็นแนวโน้มกำลังสองหรือพาราโบล่า สมการ (๒) เป็นแนวโน้มกำลังหนึ่งหรือเส้นตรงดังแสดงในรูปที่ ๑๕

รูปที่ ๑๕



การวิเคราะห์อนุกรมเวลาแม้จะใช้วิธีกำลังสองน้อยที่สุด แต่พิจารณาในช่วงเวลาต่างกัน อาจจะได้แนวโน้มต่าง ๆ กันจึงเกิดปัญหาในการเลือกแนวโน้มที่ถูกต้องเพื่อใช้ทำนายในระยะยาวได้ ทั้งนี้เพราะมีแนวโน้มแบบวัฏจักรและแบบที่เกิดโดยบังเอิญเข้ามามีอิทธิพล หลักทั่วไปที่ช่วยในการวิเคราะห์หาแนวโน้ม[๔๖]

๑. ระยะเวลาที่ใช้หาแนวโน้มต้องนานเพราะแนวโน้มระยะยาว (Secular Trend) จะไม่เกิดในเวลาสั้น ๆ และป้องกันการเปลี่ยนแปลงแบบวัฏจักรเข้ามามีส่วนร่วม ปกติอย่างน้อยต้องใช้ ๒๐ ปี หรือมากกว่านั้นยิ่งดี

๒. จะต้องพิจารณาว่าการเปลี่ยนแปลงแบบวัฏจักรที่แทรกเข้ามาในช่วงที่นำมาพิจารณาหาแนวโน้มนั้นครบรอบพอดีหรือไม่เช่นถ้าเริ่มจากยอดของวัฏจักรต้องจบลงที่ยอดด้วย

๓. การเลือกฟังค์ชั่นทางคณิตศาสตร์ที่จะใช้แสดงแนวโน้มต้องระมัดระวังให้ตรงกับการเปลี่ยนแปลงของข้อมูล นอกจากใช้หลักการทางคณิตศาสตร์แล้วยังต้องใช้ประสพการณ์ประกอบการพิจารณาเช่น แนวโน้มการเพิ่มจำนวนทารกที่ตายต่อทารกที่เกิดใหม่ ๑๐๐๐ คนและแนวโน้มแสดงการเพิ่มจำนวนการขายมักจะเป็นเส้นตรง จำนวนนักศึกษาที่สมัครเข้าในปีต่าง ๆ, การเพิ่มจำนวนของแบ็คทีเรีย, การสลายตัวของสารกัมมันตรังสี โดยทั่วไปอยู่ในรูปกำลัง (Exponential curve) คือ $y = AB^{x}$ เป็นต้น

การใช้กำลังสองน้อยที่สุดจะให้ผลดีเมื่อข้อมูลเป็นไปตามข้อตกลงเบื้องต้นดังนี้[๔๗]

๑. ตัวแปรตามของประชากรกระจายเป็นโค้งปกติ

๒. ความแปรปวนของการกระจายของประชากรสม่ำเสมอ

๓. ค่าที่เบี่ยงเบนจากเส้นกำลังสองน้อยที่สุดไม่ขึ้นแก่กัน สำหรับอนุกรมเวลาข้อตกลงเบื้องต้นข้อ ๓ นี้มักจะเป็นไปไม่ได้เช่น จำนวนนักเรียนในปีหน้าจะขึ้นกับจำนวนนักเรียน

---

46 William Addison, Neiswanger, Elementary Statistical Method (New York : The Macmillan Company, c 1956), pp. 518 - 19.

47 Werner Z. Hirsch, Introduction to Modern Statistics (New York : John Wiley & Sons, Inc., c 1957), p. 301.

ในปีนี้อย่างแน่นอน การวิเคราะห์อนุกรมเวลาจึงต้องพิจารณาอย่างรอบคอบ

๔.๑๓ การใช้วิธีกำลังสองน้อยที่สุดในการวางแผนการวิจัย (Research Design) การหาสูตรสำหรับการวางแผนวิจัยแบบต่าง ๆ [๔๘]

๑. สำหรับ Randomized Group Design ซึ่งมีตัวแปรสุ่มที่เป็นไปตาม

สมการ $x_{ij} = \mu + t_i + e_{ij}$, $i = 1, \ldots\ldots\ldots k$, $j = 1, \ldots\ldots\ldots n$ ....(1)

$\mu =$ ตัวคงที่สำหรับทุกค่าของ t_i เป็นตัวคงที่สำหรับการวัด n ครั้ง k treatment

$e_{ij} =$ ค่าผิดพลาดสุ่มของการวัดแต่ละครั้ง

เราจะใช้วิธีกำลังสองน้อยที่สุดหา t_i และ μ สำหรับกลุ่มตัวอย่างที่ได้จากการสุ่ม

$$e_{ij} = x_{ij} - t_i - \mu \qquad \ldots\ldots\ldots\ldots (2)$$

ให้ $Q = \sum e_{ij}^2 = \sum (x_{ij} - t_i - \mu)^2$ ....(3)

เราให้ $\bar{x}$ เป็นค่าประมาณของ μ และ $\bar{x}_i - \bar{x}$ เป็นค่าประมาณของ t_i ตัวประมาณค่าเหล่านี้จะเป็นตัวประมาณค่ากำลังสองน้อยที่สุด ดีฟเฟอเรนทิเอท (๒) เทียบกับ μ และ (๓) เทียบกับ e จะได้ -1 และ $2\sum e$ เมื่อคูณดิริเวทีฟทั้งสองจะได้

$$\frac{dq}{d\mu} = (-1)\ 2 \sum e$$

$$= -2(\sum x_{..} - n \sum t_i - k_n \mu) \qquad \ldots\ldots\ldots\ldots (4)$$

เราให้ $\sum t_i = 0$ ถ้าดิริเวทีฟ $= 0$ จะหา μ ได้ดังนี้

$$2 k_n \mu = 2 \sum x_{..}$$

หรือ $\mu = \bar{x}_{..}$

---

[48] Allenl. Edwards, Experimental Design in Psychological Research (3 ded. ; New York : Holt, Rinechart and Winston, Inc., c 1968), pp. 353 - 58.

ฉะนั้น $\bar{x}$ เป็นค่าประมาณกำลังสองน้อยที่สุดของ μ

ถ้าดีฟเฟอเรนทิเอท (๒) เทียบกับ t_i และ (๓) เทียบกับ e คูณกัน

จะได้ $\frac{dQ}{dt_i} = -2 (\sum x_{i.} - nt_i - n\mu) \quad \ldots\ldots\ldots\ldots\ldots (5)$

ให้ดิริเวทีพนี้เท่ากับศูนย์และหาค่า t_i เราได้

$$2\,nt_i = 2 \sum x_i - 2 \sum n\mu$$

$$t_i = \bar{x}_{k.} - \mu$$

แทนค่าตัวประมาณค่ากำลังสองน้อยที่สุดของ μ จะได้

$$t_i = \bar{x}_{k.} - \bar{x}_{..}$$

ฉะนั้น $\bar{x}_k - \bar{x}$ เป็นตัวประมาณค่ากำลังสองน้อยที่สุดของ t_i ถ้าแทนตัวประมาณค่ากำลังสองน้อยที่สุดของ μ และ t_i ใน (๓) เราได้

$$\sum_{i}^{ij} e_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} \left[x_{ij} - (\bar{x}_{i.} - \bar{x}_{..}) - \bar{x}_{..} \right]^2$$

$$= \sum_{i}^{k} \sum_{i}^{n} \left[x_{ij} - \bar{x}_{i.} \right]^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

สมการ (๖) เป็นผลบวกกำลังสองของในระหว่าง treatments

๒. การประมาณค่าสำหรับ two-factor experiment โดยใช้ randomized group design กำหนดให้

$$x_{abn} = \mu + t_a + t_b + t_{ab} + e_{abn} \quad \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

และ $e_{abn} = x_{abn} - t_a - t_b - t_{ab} - \mu \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$

$$Q = \sum e_{abn}^2 = \sum (x_{abn} - ta - t_b - t_{ab} - \mu)^2 \quad \ldots\ldots\ldots\ldots (9)$$

ดีฟเฟอเรนทิเอ (๘) เทียบกับค่าที่ไม่ทราบ t_a, t_b, t_{ab} หรือ μ เราได้ -๑ และดีฟเฟอเรนทิเอท (๙) เทียบกับ e เป็น $2 \sum e$ คูณกันจะได้

$$-2 \sum e = -2 \sum (x_{abn} - t_a - t_b - t_{ab} - \mu) \text{..........(10)}$$

บวกสมการที่ได้จากการสังเกตทั้งหมด nab เราได้

$$\frac{dQ}{d\mu} = -2 \left(\sum x_{\ldots} - nb \sum_{1}^{a} t_a - n_a \sum_{1}^{b} t_b - n \sum_{1}^{ab} t_{ab} - nab\mu \right) \text{..(11)}$$

เรากำหนดไว้ล่วงหน้าและโดยใช้ side condition จะได้

$$\sum_{1}^{a} t_a = \sum_{1}^{b} t_b = \sum_{1}^{a} t_{ab} = \sum_{1}^{b} t_{ab} \simeq \sum_{1}^{ab} t_{ab} = 0 \text{.........(12)}$$

สมการ (๑๑) จะเป็น

$$\frac{dQ}{d\mu} = -2 \sum X_{\ldots} + a\, nab\mu \qquad \text{.............(13)}$$

ให้ดิริเวทีพเท่ากับ ๐ จะได้ $\mu = \bar{x}$ และ $\bar{x}_{\ldots}$ เป็นตัวประมาณค่ากำลังสองน้อยที่สุดของ μ

ในการหาตัวประมาณค่ากำลังสองน้อยที่สุดของ ta กำหนดให้ค่าที่สังเกตได้มี nb สมการซึ่งประกอบด้วย t_a และสมการ (๑๐) จะเป็น

$$\frac{dQ}{dt_a} = -2\left(\sum x_{a..} - nbt_a - n \sum_{1}^{b} t_b - n \sum_{1}^{b} t_{ab} - nb\mu \right) \text{..(14)}$$

เราจะได้ $$\frac{dQ}{dt_a} = -2 \sum x_{a..} + 2\, nbt_a + 2nb\mu \qquad \text{..........(15)}$$

ให้ดิริเวทีพเท่ากับศูนย์และแทนค่าตัวประมาณค่ากำลังสองน้อยที่สุดของ μ ด้วย $\bar{x}_{\ldots}$ เราจะได้

$$t_a = \bar{x}_{a..} - \bar{x}_{\ldots}$$

และ $\bar{x}_{a..} - \bar{x}_{\ldots}$ เป็นตัวประมาณค่ากำลังสองน้อยที่สุดของ ta และเราจะหาตัวประมาณค่ากำลังสองน้อยที่สุดของ t_b ได้เป็น

$$t_b = \bar{x}_{.b.} - \bar{x}_{\ldots}$$

ในการหาตัวประมาณค่ากำลังสองน้อยที่สุดของ t_{ab} เราได้

$$\frac{dQ}{dt_{ab}} = -2 \left(\sum x_{ab.} - nt_a - nt_b - nt_{ab} - n\mu \right) \quad \ldots\ldots(16)$$

แทนค่าตัวประมาณค่ากำลังสองน้อยที่สุดของ ta, t_b และ μ แล้วให้ดิริเวทีพเท่ากับศูนย์เราจะได้

$$\sum e^2_{abn} = \sum \Big[x_{abn} - (\bar{x}_{a..} - \bar{x}_{...}) - (\bar{x}_{.b.} - \bar{x}_{...}) - (\bar{x}_{ab.} - \bar{x}_{a..} - \bar{x}_{.b.} + \bar{x}_{...}) - \bar{x}_{...} \Big]^2$$

$$\sum e^2_{abn} = \sum_{1}^{abn} (x_{abn} - \bar{x}_{ab.})^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(17)$$

สมการ (๑๗) จะเป็นผลบวกกำลังสองในระหว่าง treatments

๓. การประมาณค่าสำหรับ Randomized Block Design

ถ้ามี k treaments และ n blocks สมการคณิตศาสตร์จะเป็น

$$x_{kn} = \mu + t_n + t_k + e_{kn} \quad \ldots\ldots\ldots\ldots\ldots(18)$$

เมื่อ t_n เป็น block effect และ t_k เป็น treatment effect และ e_{kn} เป็น random error ดังนั้น

$$e_{kn} = x_{kn} - t_k - t_n - \mu \quad \ldots\ldots\ldots\ldots\ldots(19)$$

$$Q = \sum e^2_{kn} = \sum (x_{kn} - t_k - t_n - \mu)^2 \quad \ldots\ldots\ldots\ldots\ldots(20)$$

ใช้วิธีเดียวกันหาว่าถ้า Q มีค่าน้อยที่สุดแล้วตัวประมาณค่ากำลังสองน้อยที่สุดของ t_k, t_n และ μ จะหาได้จาก side condition ว่า $\sum t_k = \sum t_n = 0$ ดังนั้น

$$t_k = \bar{x}_{k.} - \bar{x}_{..}$$

$$t_n = \bar{x}_{.n} - \bar{x}_{..}$$

$$\mu = \bar{x}_{..}$$

แทนค่า (๒๐) ด้วยตัวประมาณค่ากำลังสองน้อยที่สุดของ t_k, t_n และ μ จะได้ว่า

$$\sum e^2_{kn} = \sum (x_{kn} - \bar{x}_{k.} - \bar{x}_{.n} + \bar{x}_{..})^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (21)$$

สมการ (๒๑) เป็น error sum square สำหรับ randomized block design.

๔. การประมาณค่าสำหรับ Latin Square Design

สำหรับ Latin square design ซึ่งมี r แถวและ c สดมภ์ และ t treatments เมื่อ r = c = t เราจะได้สมการทางคณิตศาสตร์ดังนี้

$$x_{rct} = \mu + t_r + t_c + t_t + e_{rct} \quad \ldots\ldots\ldots\ldots\ldots\ldots (22)$$

เมื่อ t_r เป็นผลของแถว, t_c เป็นผลของสดมภ์, t_t เป็นผลของ treatment และ e_{rct} เป็นค่าผิดพลาดสุ่มดังนั้น

$$e_{rct} = x_{rct} - t_r - t_c - t_t - \mu \quad \ldots\ldots\ldots\ldots\ldots (23)$$

และ $Q = \sum e^2_{rct} = \sum (x_{rct} - t_r - t_c - t_t - \mu)^2$ ....(24)

side conditions คือ

$$\sum_{1}^{r} t_r = \sum_{1}^{t} t_t = \sum_{1}^{c} t_c = 0$$

เราจะได้ตัวประมาณค่ากำลังสองน้อยที่สุด ดังนี้

$$t_r = \bar{x}_{r..} - \bar{x}_{...}$$

$$t_c = \bar{x}_{.c.} - \bar{x}_{...}$$

$$t_t = \bar{x}_{..t} - \bar{x}_{...}$$

$$\mu = \bar{x}_{...}$$

แทนค่าตัวประมาณค่าเหล่านี้ในสมการ (๒๔) เราจะได้

$$\sum e^2_{rct} = \sum (x_{rct} - \bar{x}_{r..} - \bar{x}_{.c.} - \bar{x}_{..t} + 2\bar{x}_{...})^2 \quad \ldots\ldots\ldots (25)$$

<u>ตัวอย่าง ๑๔</u> แสดงการคำนวณใน ๒^{๓} factorial design เพื่อให้ได้สมการกำลังหนึ่ง[๔๙]
จากการทดลองสมมุติว่ามีแฟคเตอร์ ๓ อย่างที่มีผลต่อการทดลองทางเคมี คือ

A = x_1 = เวลาที่ใช้ในปฏิกริยา (๕.๕ และ ๖.๕ ชั่วโมง)

B = x_2 = อุณหภูมิ (๒๖๐° และ ๓๔๐° c)

C = x_3 = จำนวนสารที่ใช้ (๑.๔ และ ๑.๘ กรัม)

สมการทางคณิตศาสตร์ที่ใช้ คือ

$$\hat{y}_i = b_0 + b_1x_{1i} + b_2x_{2i} + b_3x_{3i} \qquad (i = 1,2......8)$$

$\hat{y}_i$ = เป็นค่าที่ทำนายค่า y

เพราะว่าค่า x ทุกตัวมี ๒ ค่า เพื่อสะดวกจะได้สเกลใหม่ให้ค่า x เป็น +๑ และ −๑ ทุกตัวดังนี้

$$x_1 = \text{๒ (เวลา −๖)}, \quad x_2 = \frac{(\text{อุณหภูมิ} - \text{๓๐๐})}{\text{๔๐}}, \quad x_3 = \frac{(\text{จำนวนสาร} - \text{๑.๖})}{\text{๐.๒}}$$

การคำนวณใช้เมตริกข้อมูลดังแสดงในตาราง

---

49 William G. Cochran & Gertrude M. Cox, <u>Experimental Designs</u> (2d ed.; New York : John Wiley & Sons Inc., c 1957), pp. 338 - 40.

ตารางที่ ๑๕

	x (โคดสเกล)				Y
Treatment	x_0	x_1	x_2	x_3	y
(1)	1	-1	-1	-1	161
a	1	+1	-1	-1	183
b	1	-1	+1	-1	151
ab	1	+1	+1	-1	170
c	1	-1	-1	+1	166
ac	1	+1	-1	+1	192
bc	1	-1	+1	+1	156
abc	1	+1	+1	+1	183

Normal Equation $\hat{b} = [x'x]^{-1} x'y$

$$x'x = \begin{bmatrix} 8 & & & \\ & 8 & & \\ & & 8 & \\ & & & 8 \end{bmatrix}, \qquad x'y = \begin{bmatrix} 1362 \\ 94 \\ -42 \\ 32 \end{bmatrix}$$

$$[x'x]^{-1} = \begin{bmatrix} \frac{1}{8} & & & \\ & \frac{1}{8} & & \\ & & \frac{1}{8} & \\ & & & \frac{1}{8} \end{bmatrix}$$

$$\therefore \hat{b} = \begin{bmatrix} \frac{1}{8} & & & \\ & \frac{1}{8} & & \\ & & \frac{1}{8} & \\ & & & \frac{1}{8} \end{bmatrix} \begin{bmatrix} 1362 \\ 94 \\ -42 \\ 32 \end{bmatrix} = \begin{bmatrix} 170.2 \\ 11.8 \\ -5.2 \\ 4.0 \end{bmatrix}$$

สมการที่ได้จากการทดลองนี้คือ

$$\hat{y} = 170.2 + 11.8x_1 - 5.2x_2 + 4.0x_3$$

การใช้วิธีกำลังสองน้อยที่สุดในการวิเคราะห์ข้อมูลที่ได้จากการทดลองใช้หลักการอย่างเดียวกับการคำนวณหาสมการถดถอย คือจะได้ Normal equation $\hat{b} = [X'X]^{-1}X'Y$ เมื่อได้สมการที่ใช้แทนผลการทดลองแล้ว ต้องตรวจสอบดูว่าข้อมูลที่ได้จากการทดลองใช้สมการกำลัง หนึ่งแทนได้จริงหรือไม่ โดยหาค่าที่ได้จากสมการที่เบี่ยงเบนไปจากค่าจริง ค่าที่เบี่ยงเบนนี้เกิดจากสาเหตุ ๒ อย่างคือ ความผิดพลาดเนื่องจากการทดลอง และความผิดพลาดที่เกิดขึ้นเนื่องจากเลือกสมการทางคณิตศาสตร์ไม่เหมาะสมกับข้อมูล ถ้าค่าผิดพลาดชนิดหลังมากกว่า แสดงว่าสมการกำลังหนึ่งใช้ไม่ได้ ค่าผิดพลาดที่เนื่องจากเลือกสมการผิดนี้บางทีเรียก " lack of fit "

<u>ตัวอย่างที่ ๑๕</u> แสดงการวิเคราะห์ 2 x 3 factorial experiment ข้อมูลเป็นผลของความดันและอุณหภูมิต่อผลลัพธ์ที่ได้จากปฏิกิริยาทางเคมีซึ่งความดันและอุณหภูมิที่ใช้มีดังนี้[๕๐]

ความดัน (ปอนด์ต่อตารางนิ้ว)		อุณหภูมิ (°F)	
๕๐	−๑	๑๐๐	−๑
๘๐	๑	๒๐๐	๐
		๓๐๐	๑

---

[50] Mendenhall, <u>Op. cit</u>., pp. 128-31.

ถ้าอุณหภูมิและความดันไม่มีปฏิกิริยาร่วมจะได้ข้อมูลดังตารางข้างล่าง

ตารางที่ ๑๖

Y	ความดัน (P)		อุณหภูมิ (T)	
	จริง	โค้ด	จริง	โค้ด
21	50	-1	100	-1
23	50	-1	200	0
26	50	-1	300	1
22	80	1	100	- 1
23	80	1	200	0
28	80	1	300	1

$$x_1 = \frac{P - 65}{15} \quad , \qquad x_2 = \frac{T - 200}{100}$$

สมการทางคณิตศาสตร์เมื่อไม่มีปฏิกิริยาร่วม

$$y = b_o + b_1x_1 + b_2x_2 + b_3x_2^2 + e$$

$$y = \begin{bmatrix} 21 \\ 23 \\ 26 \\ 22 \\ 23 \\ 28 \end{bmatrix} \quad , \qquad x = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & +1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & +1 & 1 \end{bmatrix}$$

$$x'x = \begin{bmatrix} 6 & 0 & 0 & 4 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 4 & 0 & 0 & 4 \end{bmatrix} \quad , \qquad x'y = \begin{bmatrix} 143 \\ 3 \\ 11 \\ 97 \end{bmatrix}$$

$$[X'X]^{-1} \begin{bmatrix} \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{3}{4} \end{bmatrix}$$

และ

$$\hat{b} = (X'X)^{-1} X'y = \begin{bmatrix} \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{3}{4} \end{bmatrix} \begin{bmatrix} 143 \\ 3 \\ 11 \\ 97 \end{bmatrix} = \begin{bmatrix} 23 \\ .5 \\ 2.75 \\ 1.25 \end{bmatrix}$$

สมการที่ใช้ทำนายคือ $\hat{y} = 23 + 0.5x_1 + 2.75x_2 + 1.25x_2^2$

เราอาจจะตรวจสอบค่าจากการทำนายกับค่าจริงที่ได้จากการทดลองได้

x_1	x_2	y	$\hat{y}$	$y - \hat{y}$
-1	-1	21	21.0	0
-1	0	23	22.5	0.5
-1	1	26	26.5	-0.5
1	-1	22	22.0	0
1	0	23	23.5	-0.5
1	1	28	27.5	0.5

การใช้โค้ดจะช่วยให้การคำนวณง่ายขึ้นโดยไม่ทำให้ค่าที่ทำนายผิดไป

นอกจากนี้ในการวิเคราะห์ข้อมูลมักจะพบปัญหาเกี่ยวกับข้อมูลบางค่าขาดหายไป เนื่องจากความบกพร่องในการรวบรวมข้อมูลหรือมีเหตุการณ์ที่ไม่คาดหวังเกิดขึ้น ซึ่งจะมีผลต่อการวิเคราะห์ข้อมูล การแก้ไขคือพยายามประมาณค่าข้อมูลที่ขาดหายไปโดยเขียนสมการ

ทางคณิตศาสตร์ที่สามารถจะแทนข้อมูลที่ได้จากการสำรวจ แล้วใช้วิธีกำลังสองน้อยที่สุดหา Normal equation จะได้สมการเช่นเดียวกับที่มีครบทุกค่าดังนี้

$$\Sigma(y_{ijk} - m - t_i - b_j) = 0$$

จะมีสมการที่ขาดข้อมูลบางตัว ทำให้การคำนวณยุ่งยาก ในการวิเคราะห์ความแปรปวน (Analysis of Variance) ทำให้เกิดการเปลี่ยนแปลง ๒ อย่าง

๑. treatment และ block SS สับสนจะต้องคำนวณ treatment SS. หลัง block SS.

๒. จำนวนขั้นแห่งความอิสสระทั้งหมดจะลดไป นอกจากว่าจะหายไปทั้ง treatment หรือทั้ง block จำนวนพารามิเตอร์ที่ต้องการจึงจะเท่าเดิม จำนวนชั้นแห่งความอิสสระที่หายไป มีผลจากผลบวกของค่าผิดพลาดกำลังสอง อาจกล่าวได้ว่าข้อมูลที่หายไปสามารถจะใช้วิธีกำลังสองน้อยที่สุดหาได้จากข้อมูลที่มีอยู่ทั้งหมด วิธีนี้เรียก "ขบวนการแก้ไขกำลังสองน้อยที่สุด (Correct least squares procedure)"[๕๑]

อย่างไรก็ตามการคำนวณตามวิธีดังกล่าวยุ่งยาก เยทส์ (yates) และฟิสเชอร์ (Fisher) ได้คิดหาวิธีหาข้อมูลที่ขาดหายไปมาเพิ่มให้ครบเช่นข้อมูลหายไป ๑ ค่าแทนด้วย x ในการวิเคราะห์ความแปรปวน ผลบวกกำลังสองของค่าผิดพลาดของข้อมูลทั้งหมดจะเขียนในรูป $AX^2 - 2BX + C$, A,B,C, เป็นไปตามแบบของการวิจัยและข้อมูลที่มีอยู่เมื่อ A เป็นค่าบวกเสมอ

การหาค่า x เยทส์แนะให้ใช้วิธีทำให้ผลบวกกำลังสองของค่าผิดพลาดน้อยที่สุด นั่นคือ $x = \frac{B}{A}$ ถ้าใช้ค่านี้แทนค่าที่ขาดหายไป เมื่อนำข้อมูลที่เพิ่มค่าที่ขาดหายไปแล้วมาวิเคราะห์จะมีคุณสมบัติต่าง ๆ เปรียบเทียบกับวิธีกำลังสองน้อยที่สุดดังนี้

๑. การประมาณค่า treatment และ block effects จะเหมือนที่

---

51 Cochran, Op.cit., pp. 80 - 82.

คำนวณได้จากขบวนการแก้ไขกำลังสองน้อยที่สุด

๒. ผลบวกของค่าผิดพลาดกำลังสองเท่ากัน

๓. จำนวนขั้นแห่งความผิสสระถูกต้องเนื่องจากลบจากทั้งหมด ๑ และลบจากผลบวกกำลังสองของค่าผิดพลาดอีก ๑

การแก้ไขดังกล่าวทำให้จำนวนขั้นแห่งความอิสสระเปลี่ยนไปเท่านั้น การคำนวณเหมือนเดิม ถ้ามีข้อมูลหลายค่าหายไป ก็ใช้วิธีดังกล่าวซ้ำเพื่อให้ได้ข้อมูลครบทุกค่า อย่างไรก็ตามวิธีแก้ไขนี้เพื่อให้การวิเคราะห์ข้อมูลทั้งหมดสะดวกเท่านั้น แต่ไม่สามารถจะทำให้ผลการวิเคราะห์เหมือนกับที่มีข้อมูลครบ เพียงแต่ลดข้อผิดพลาดลงเท่านั้น

๔.๒ การทดสอบสมมุติฐาน (Testing Hypothesis)

การประมาณค่าเป็นจุดมักจะได้ค่าไม่เท่ากับตัวพารามิเตอร์ จึงควรจะประมาณค่าเป็นอันตรภาคดังได้กล่าวไว้แล้ว การประมาณค่าเป็นอันตรภาคมีข้อตกลงเบื้องต้นว่า การกระจายของค่า y สำหรับค่า x ใด ๆ ต้องเป็นโค้งปกติ กลุ่มตัวอย่างจึงจะกระจายแบบ t [๕๒]

ถ้าสมการเป็น $y = a + bx + \epsilon$

$\hat{y}$ เป็นค่าประมาณของ y ความแปรปวนของตัวประมาณค่าจะช่วยให้หาขอบเขตของความเชื่อมั่นได้ ความแปรปวนของ y ได้จากผลบวกกำลังสองของส่วนเบี่ยงเบนหารด้วยจำนวนขั้นของความอิสสระคือ $Sy^2 = \frac{\Sigma \epsilon_i^2}{\nu}$ ในการหาสมการถดถอยจะต้องหาจุดกึ่งกลาง$(\bar{x}, \bar{y})$, ความชัน b และจุดตัด a ถ้า n เป็นจำนวนจุดที่ได้จากการสังเกต (x_i, y_i) จำนวนขั้นของความอิสสระจะเป็น [๕๓]

$$\nu = n - 2$$

ดังนั้น $$S^2y = \frac{\Sigma \epsilon_i^2}{n - 2} \qquad \text{....................(1)}$$

ความแปรปวนของ $\bar{y}$ ได้จาก

$$S^2_{\bar{y}} = S^2_y / n \qquad \text{....................(2)}$$

จะเขียนขอบเขตของความเชื่อมั่นของ $\bar{y}$ ได้ดังนี้ $\bar{y} \pm t\, S_{\bar{y}}$ โอกาสที่จะผิดเท่ากับระดับความเชื่อมั่นของ t ความแปรปวนของความชัน b ได้จาก

---

52 Wilfrid J. Dixon & Frank J. Massey Jr., Introduction to Statistical Analysis (2 d ed.; New York : Mc Graw-Hill Book Company, Inc., c 1957), p. 193.

53 Adam M. Neville, John B. Kennedy, Basics Statistical Methods for Engineers and Scientists (Pennsylvania : International Text Book Company, c 1964), pp. 178 - 83.

$$S_b^2 = \frac{S_y^2}{\Sigma (x - \bar{x})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

ขอบเขตของความชันจะเป็น $b \pm t\, S_b$

ถ้าจะหาขอบเขตของความเชื่อมั่นที่คำนวณได้จากสมการถดถอย $\hat{y}_i$ สำหรับค่า x_i ต้องหาความแปรปวนของ $\hat{y}_i$ ถ้าต้องการหาขอบเขตของมัชฌิมของ y_i คือ $\bar{y}_i$ จะต้องหาความแปรปวนดังนี้

$$S_{\bar{y}i}^2 = S_y^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{(x - \bar{x})^2} \right] \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

ขอบเขตของมัชฌิมที่ประมาณได้คือ $\bar{y}_i \pm tS_{\bar{y}i}$

เมื่อจะประมาณขอบเขตของค่าที่ทำนาย $\hat{y}_i$ ค่าเดียวจะหาความแปรปวนได้

จาก $$Sy_i^2 = Sy^2 \left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{(x - \bar{x})^2} \right) \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

และ $Sy^2{}_i = S_{\bar{y}i}^2 + S_y^2$ ……………… (6)

ขอบเขตความเชื่อมั่นของค่า y ใด ๆ $= y \pm ts_{yi}$ ……………… (7)

ถ้าต้องการจะหาจุดตัด a แทนค่า $x_i = 0$ ใน (๔) จะได้ค่าความแปรปวนของ a

$$S_a^2 = S_y^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\Sigma (x - \bar{x})^2} \right) \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

การทดสอบความมีนัยสำคัญของความชัน ในบางกรณีเรามีค่าความชันตามทฤษฎี b_o เมื่อความชันในสมการถดถอย b ต้องการจะทดสอบว่า b_o และ b ต่างกันมีนัยสำคัญหรือไม่ ส่วนเบี่ยงเบนมาตรฐานของ $b - b_o$ คือ S_b เราต้องหาค่า t จาก

$$t = \frac{|b - b_o|}{S_b} \quad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

ถ้า t ที่คำนวณได้มากกว่าค่า t ในตาราง ความแตกต่างนั้นจะมีนัยสำคัญและเราจะสรุปว่า b และ b_o ต่างกัน ใช้จำนวนชั้นแห่งความอิสสระเป็น n - 2 การทดสอบ a ก็เช่นเดียวกัน

$$t = \frac{|a - a_o|}{S_a} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

ถ้าจะทดสอบว่าความชันต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ใช้

$$t = \frac{b}{S_b} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

ถ้า b=0 y จะไม่ขึ้นกับ x

<u>ตัวอย่าง ๑๖</u> สมการถดถอยระหว่าง x และ y คือ $y = 4.089 + 1.026x$

ก. จงตรวจสอบว่าจุดตัด a และความชัน b ต่างจากศูนย์มีนัยสำคัญหรือไม่

ข. จงหาว่าความชันจะต่างจากความชันตามทฤษฎี b=1 อย่างมีนัยสำคัญหรือไม่

ค. จงหาขอบเขตของความเชื่อมั่น ๙๕ เปอร์เซ็นต์ของ a และ b

ง. สมมุติว่าสมการถดถอยนี้มีค่า x=25 จงประมาณค่ามัชฌิมของ $\hat{y}_i$ เมื่อ x=25 และหาขอบเขตของความเชื่อมั่น ๙๕ เปอร์เซ็นต์สำหรับค่าที่คำนวณได้

จ. หาขอบเขตของความเชื่อมั่น ๙๕ เปอร์เซ็นต์สำหรับค่า y เมื่อ x = 25

ตารางที่ ๑๗

x	$\hat{y}$	$\epsilon_i = y_i - \hat{y}_i$	ϵ_i^2
11	15.375	-0.175	0.0306
13	17.427	+0.273	0.0745
15	19.479	-0.179	0.0320
17	21.531	-0.031	0.0010
19	23.583	+0.317	0.1005
21	25.635	-0.235	0.0552
$n = 6$			$\sum \epsilon_i^2 = 0.2938$

ก. จากสมการ (๑) $S_y = \sqrt{\frac{\sum \epsilon_i^2}{n - 2}} = \sqrt{\frac{0.2938}{6 - 2}} = 0.271$

จากสมการ (๓) $S_b = \frac{0.271}{\sqrt{70}} = 0.0324$

จากสมการ (๔) $S_a = S_y \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}} = 0.271 \sqrt{\frac{1}{6} + \frac{16^2}{70}} = 0.5298$

ทดสอบ b $t = \frac{b}{S_b} = \frac{1.026}{0.0324} = 31.667$

จากตารางที่ d.f. = 4 ค่า t จะมีนัยสำคัญที่ระดับ ๐.๑ เปอร์เซ็นต์อย่างต่ำ

ทดสอบ a $t = \frac{a}{S_a} = \frac{4.089}{0.5298} = 7.718$

a = 4.089 ที่ระดับความมีนัยสำคัญที่ ๐.๑ เปอร์เซ็นต์เช่นเดียวกัน

ข. จากสมการ (๕)

$$t = \frac{|b - b_o|}{S_b} = \frac{1.026 - 1}{0.0324} = 0.8025$$

ที่ d.f. = 4, t ไม่มีนัยสำคัญค่า b และค่าที่ทำนาย $b_o = 1$ ไม่ต่างกัน

ค. ที่ d.f. = 4 และระดับความมีนัยสำคัญ ๕ เปอร์เซ็นต์, t = 2.776

ขอบเขตของ a จะเป็น $4.089 \pm 2.776 \times 0.5298 = (2.618, 5.560)$

" b " $1.026 \pm 2.776 \times 0.0324 = (0.936, 1.116)$

ง. จากสมการ $\bar{y}_i = 4.089 + 1.026 \times 25 = 29.739$

$$S_{\bar{y}i} = S_y \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{(x - \bar{x})^2}} = 0.271 \sqrt{\frac{1}{6} + 1} = 0.2927$$

ขอบเขตของ $\bar{y}_i = 29.739 \pm 2.776 \times 0.2927 = (30.651, 28.927)$

จ. จากสมการ (๖) $S_{yi} = \sqrt{S_{\bar{y}i}^2 + S_y^2} = \sqrt{.07345 + .08568} = .399$

จากสมการถดถอยได้ $y_i = 29.739$ เมื่อ $x = 25$

ขอบเขตของ $y_i = 29.739 \pm 2.776 \times .399 = (30.847, 28.631)$

การทดสอบว่าตัวแปร y ไม่ขึ้นกับตัวแปร x คือมัชฌิมของ y จะเท่ากันทุกค่าของ x ในสมการถดถอยคือ b=0 มีขั้นของการทดสอบดังนี้

๑. ตั้งสมมุติฐาน $H_o : b = 0$

๒. เลือกระดับความมีนัยสำคัญ (α)

๓. คำนวณหาค่า t จาก $t = \frac{|b - 0|}{S_b}$

๔. ถ้าการกระจายของ y สำหรับแต่ละค่าของ x เป็นโค้งปกติมีความแปรปวนเท่ากัน โดยมีมัชฌิมเลขคณิต $\mu = A$ กลุ่มตัวอย่างจะกระจายแบบ t โดยมีขั้นแห่งความอิสสระ=n - 2

๕. จะไม่ยอมรับสมมุติฐานเมื่อ t ที่คำนวณได้มากกว่าค่าจากตาราง

ตัวอย่าง ๑๗ แสดงลำดับขั้นการทดสอบด้วยตัวเลข

$$H_o : b = 0$$

$$\alpha = .05$$

$$t = \frac{(5.029 - 0)}{\frac{9.61}{13.1}} = 6.9$$

ถ้าการกระจายของกลุ่มตัวอย่างเป็นแบบ t ตามข้อตกลงเบื้องต้นโดยใช้ d.f. = 10 จากตารางจะได้ $-2.3 < t < 2.3$ ค่าที่คำนวณได้อยู่นอกขอบเขตของ t สรุปได้ว่า $b \neq 0$ ที่ระดับความเชื่อมั่น ๕ เปอร์เซ็นต์ ∴ y ขึ้นกับ x

การทดสอบว่าสมการถดถอยเป็นสมการกำลังหนึ่งหรือมัชฌิมของแต่ละกลุ่มอยู่บนเส้นตรงนี้หรือไม่ เราใช้เทคนิคการวิเคราะห์ความแปรปวนคือเปรียบเทียบความแปรปวนในกลุ่มกับส่วนเบี่ยงเบนของกลุ่มจากเส้นตรงถดถอยที่ประมาณค่า แล้วใช้ F-test ทดสอบดังแสดงในตารางข้างล่าง ๕๔

ผลบวกกำลังสองของมัชฌิมแต่ละกลุ่มของ y เบี่ยงเบนจากเส้นตรงถดถอยได้จากผลต่างระหว่างผลบวกกำลังสองทั้งหมดกับผลบวกกำลังสองถดถอย ค่าที่ใช้ทดสอบคืออัตราส่วนระหว่างความแปรปวนของสมการถดถอยและความแปรปวนในกลุ่มโดยมีข้อตกลงเบื้องต้นเกี่ยวกับการกระจายเป็นโค้งปกติ ถ้าสมมุติฐานเป็นจริงการกระจายของกลุ่มตัวอย่างจะเป็น F (k - 2, N - k)

---

[54] Dixon, Op.cit., pp. 197 - 98.

ตารางที่ ๑๘

	ผลบวกของกำลังสอง	d.f.	Mean Square
ระหว่างกลุ่ม	$S_1 = \sum\sum y_{ij}^2 - \sum \frac{T_{yi+}^2}{n_i}$	N - k	$S_1/N - k$
สมการถดถอย	$S_2 = b^2\left[\sum\sum x_{ij}^2 - \frac{(\sum\sum x_{ij})^2}{N}\right]$	1	
ระหว่างสมการถดถอย	$S_3 =$ ผลต่าง	k - 2	$S_3/k - 2$
ผลรวม	$S_4 = \sum \frac{T_{yi+}^2}{n_i} - \frac{T_{y++}^2}{N}$	k - 1	

ตัวอย่าง ๑๘ แสดงการทดสอบสมมุติฐานว่าสมการถดถอยนั้นเป็นสมการเส้นตรงโดยกำหนดผลบวกต่าง ๆ ดังนี้

$$\frac{T_{y++}^2}{N} = 240{,}833.3 \quad , \quad \sum\sum y_{ij}^2 = 246{,}100$$

$$\sum \frac{T_{yi+}^2}{n_i} = \frac{(365)^2}{3} + \frac{(525)^2}{4} + \frac{(295)^2}{2} + \frac{(515)^2}{3} = 245{,}235.41$$

$$S_1 = 246{,}\ 100 - 245{,}235.41 = 864.59$$

$$S_2 = (5.028)^2(171.67) = 4340.00$$

$$S_4 = 245{,}235.41 - 240{,}833.3 = 4402.1$$

$$S_3 = 4402.1 - 4340.0 = 62.1$$

ตารางที่ ๑๙

	ผลบวกกำลังสอง	d.f	Mean Square
ระหว่างกลุ่ม	864.6	8	108.1
สมการถดถอย	4340.0	1	
ระหว่างสมการถดถอย	62.1	2	31.0
ผลรวม	4402.1	3	

จากตาราง $F = 31.0 / 108.1 = 0.29$ เทียบกับ F จากตาราง $F_{.95}(2, 8) = 4.46$ ไม่มีนัยสำคัญทางสถิติแสดงว่ายอมรับสมมุติฐานว่าสมการถดถอยนี้เป็นสมการเส้นตรง

การทดสอบสมการถดถอยที่กำลังมากกว่าหนึ่งขึ้นไป หลักการจะแตกต่างไปโดยเพิ่มกำลังของโล่ในเมียลคือ m จะใช้การกระจายของ F โดยจำนวนชั้นแห่งความอิสสระเปลี่ยนเป็น $(k - m - 1, N - k)$ วิธีการยังคงใช้การวิเคราะห์ความแปรปวนดังตัวอย่างข้างล่าง[๕๕]

ตัวอย่าง ๑๙ ข้อมูลได้จากการทดลองการขยายของกล้องโทรทัศน์ x_i มีหน่วยเป็นเมตรแทนระยะวัตถุ (สเกล) มองผ่านกล้องโทรทัศน์โดยใช้ตาเดียว ตาอีกข้างมองสเกลนั้นโดยตรงระยะที่เห็นเท่ากับมองผ่านกล้องเป็น y_{ij} เปลี่ยนค่า x_i ๕ ค่าวัด y_{ij} ครั้งละ ๒ ค่า ขนาดของกลุ่มตัวอย่างจึงเป็น ๑๐ การวัด ๒ ครั้งไม่เท่ากันเนื่องจากความผิดพลาดและมีสิ่งรบกวนในการวัด

---

[55] Kreyszig, Op.cit., pp. 325 - 27.

ในการหาสมการแทนกลุ่มตัวอย่าง ถ้าเราแน่ใจว่าความสัมพันธ์นี้ไม่เป็นเส้นตรง และอาจจะเป็นพาราโบล่า จึงต้องการทดสอบ สมการถดถอยที่ได้จากสมมุติฐานว่าข้อมูลสัมพันธ์กันเป็นพาราโบล่าคือ $y = 32.450 - 1.021x + 0.037x^2$ และมีข้อตกลงเบื้องต้นว่าสำหรับค่า x ใด ๆ y จะกระจายเป็นโค้งปกติและมีความแปรปวนไม่ขึ้นกับ x

การคำนวณขั้นแรกเปลี่ยน y ให้คำนวณง่าย $y^*_{ij} = y_{ij} - 25$ สมการถดถอยจะเปลี่ยนเป็น $y^* = 7.450 - 1.021x + 0.037x^2$

ตารางที่ ๒๐

x_i	y_{ij}		y^*_{ij}		$\bar{y}^*_i$	$0.037x_i^2$	$1.021x_i$	$y^*(x_i)$	$(\bar{y}^*_i - y^*(x_i))^2$
3	29.7	30.1	4.7	5.1	4.90	0.333	3.063	4.72	0.032
5	28.1	27.8	3.1	2.8	2.95	0.925	5.105	3.27	0.102
7	27.4	26.9	2.4	1.9	2.15	1.813	7.147	2.12	0.001
10	26.0	26.2	1.0	1.2	1.10	3.700	10.210	0.94	0.026
14	25.5	25.2	0.5	0.2	0.35	7.252	14.294	0.41	0.004
									0.165

ตารางวิเคราะห์ความแปรปวน

ตารางที่ ๒๑

	d.f	ผลบวกกำลังสอง	Mean Square
มัชฌิมระหว่างสมการถดถอย	2	0.330	0.165
ในระหว่างกลุ่ม	5	0.315	0.063
ผลรวม	7	0.645	

$$n_i = 2,\ S_1 = 2 \times \sum \left[(\bar{y}_i^* - y^*(x_i) \right]^2 = 2 \times 0.165 = 0.330$$

$$S_2 = \sum_{i=1}^{5} \sum_{j=1}^{2} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{5} \sum_{j=1}^{2} (y_{ij}^* - \bar{y}_{ij}^*)^2 = 0.315$$

$$F = \frac{0.165}{0.063} = 2.62, \quad k - m - 1 = 5 - 2 - 1 = 2$$

$$N - k = 10 - 5 = 5$$

ที่ระดับความเชื่อมั่น ๙๕ เปอร์เซ็นต์ $F(2, 5) = 5.79$ F ที่คำนวณได้น้อยกว่า ๕.๗๙ ยอมรับสมมุติฐานหมายความว่า สามารถใช้สมการพาราโบลาแทนข้อมูลชุดนี้ซึ่งมีค่า $x = 3$ ถึง 14 แต่มิได้หมายความว่า ข้อมูลชุดนี้จะแทนได้ด้วยสมการพาราโบลาเท่านั้น และมิได้หมายความว่าถ้าขยายช่วงของ x ให้กว้างออกไปข้อมูลเหล่านั้นจะแทนได้ด้วยสมการพาราโบลาอีกสมการ การแปลความหมายต้องระมัดระวังในเรื่องนี้