บทที่ ๓

เครื่องมือที่ใช้ในการวิจัย

ในบทที่ ๓ นี้เราจะกล่าวถึงเครื่องมือที่ใช้ช่วยในการวิจัย ในการวิจัยนี้ส่วนใหญ่เป็นเรื่องเกี่ยวกับตัวเลข ซึ่งมีความจำเป็นอย่างมากที่จะต้องอาศัยเครื่องมือในการคำนวณที่มีประสิทธิภาพ

ในปัจจุบันเครื่องมือที่นิยมใช้กันมากและมีประโยชน์ในเรื่องเกี่ยวกับตัวเลขก็คือ เครื่องคอมพิวเตอร์ ในบางครั้งเราไม่สามารถแก้ปัญหาทางคณิตศาสตร์โดยตรงได้ (Analytical Analysis) จึงมีความจำเป็นต้องอาศัยเครื่องคอมพิวเตอร์ประกอบกันกับวิธีการที่เรียกว่า "การวิเคราะห์ตัวเลข" (Numerical Analysis) เข้าช่วยในการแก้ปัญหา

บางคนอาจถามว่าการวิเคราะห์ตัวเลข (Numerical Analysis) คืออะไร? ผู้ชำนาญในสาขาวิชานี้มักจะถือว่าการวิเคราะห์ตัวเลขเป็นทั้งวิทยาศาสตร์และศิลปศาสตร์ ทั้งนี้เนื่องจากว่าการแก้ปัญหาทางการวิเคราะห์ตัวเลขนั้นบ่อยครั้งต้องประสบกับปัญหาการวิเคราะห์ฟังก์ชั่นซึ่งเราไม่อาจแสดงให้เห็นจริงได้ทั้งในทางทฤษฎีและในทางปฏิบัติ

๓.๑ <u>การหาค่าคาดหมายของจำนวนข้อมูล</u> (Expected Frequency)

การหาค่าคาดหมายของจำนวนข้อมูลในฟังก์ชั่นการแจกแจงใด ๆ ถ้าลักษณะของกราฟของฟังก์ชั่นเป็นแบบไม่ต่อเนื่อง การหาค่าคาดหมายของจำนวนข้อมูลในแต่ละช่วงย่อยไม่ยุ่งยากอะไร แต่สำหรับฟังก์ชั่นการแจกแจงที่มีลักษณะกราฟแบบต่อเนื่อง การหาค่าคาดหมายของจำนวนข้อมูลจะเริ่มยุ่งยากเพราะในช่วงของค่าของข้อมูลหนึ่ง ๆ การหาแต่ละช่วงต้องใช้การอินทิเกรต (integrate) สำหรับฟังก์ชั่นการแจกแจงบางอย่าง การอินทิเกรตในช่วงหนึ่ง ๆ จะทำได้ยาก

หรือบางครั้งก็ทำไม่ได้ ในกรณีเช่นนี้การวิเคราะห์ตัวเลขจะช่วยประมาณค่าอินทิเกรตที่ต้อง-การได้

สำหรับในการวิจัยนี้ใช้ตัวสถิติทดสอบไคสแควร์ (χ^2) ในการทดสอบลักษณะการแจกแจงของข้อมูลตัวอย่างโดยใช้สูตร

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad \text{..........(3.1)}$$

ตามที่กล่าวเอาไว้ในบทที่ ๒

สำหรับค่า E_i คำนวณจาก $E_i = N \times \int_{t_i}^{t_{i+1}} f(x)\,dx$

$f(x)$ คือฟังก์ชั่นการแจกแจงที่คาดว่าจะเข้ากับลักษณะการแจกแจงของข้อมูลในการวิจัยนี้ $f(x)$ คือฟังก์ชั่นการแจกแจงแบบปกติดังนั้น

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ และ σ^2 คือ ค่าเฉลี่ยและค่าความแปรปรวนซึ่งเป็นพารามิเตอร์ของประชากรที่มีการแจกแจงแบบปกติ

ซึ่งในการปฏิบัติแล้วเราไม่สามารถศึกษาทั้งประชากรได้ เพียงแต่ศึกษาบางส่วนของประชากรเท่านั้น โดยการสุ่มตัวอย่างมาศึกษา ดังนั้นค่าเฉลี่ยและค่าความแปรปรวนจึงเป็นเพียงตัวประมาณค่าของพารามิเตอร์เท่านั้น เพราะหามาจากตัวอย่างโดยใช้สูตร

ค่าเฉลี่ยจากตัวอย่างใช้สัญญลักษณ์ $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$

ค่าความแปรปรวนจากตัวอย่างใช้สัญลักษณ์

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$$

นั่นคือ

$$E_i = N \times \int_{t_i}^{t_{i+1}} \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2S^2}(x - \bar{X})^2} dx$$

จากสูตรการหา E_i จะพบว่าเราจะหาค่า E_i ได้ถ้าเราคำนวณ

$$\int_{t_i}^{t_{i+1}} \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2S^2}(x - \bar{X})^2} dx \quad \text{ได้}$$

แต่การที่เราจะอินทิเกรตฟังก์ชั่นเอ็กโพเน้นเชียลที่มีกำลังเป็นตัวแปรยกกำลังสอง โดยใช้สูตรการอินทิเกรตโดยตรงไม่ได้ จุดนี้เองที่เราจำเป็นต้องใช้การวิเคราะห์ทางตัวเลขช่วยในการคำนวณ วิธีการวิเคราะห์ทางตัวเลขในการแก้ปัญหาอันนี้มีหลายวิธี คือ

ก. กฎเกณฑ์การหาพื้นที่สี่เหลี่ยมคางหมู[๑]

ข. กฎเกณฑ์การหาพื้นที่โดยใช้กราฟเส้นโค้ง[๒]

ค. การประมาณค่าจากการแจกแจงปกติที่มีค่าเฉลี่ยเป็นศูนย์ และค่าความ[๓] เบี่ยงเบนมาตรฐานเป็นหนึ่ง

สำหรับในการวิจัยนี้จะใช้วิธีในข้อ ค. แต่ในวิทยานิพนธ์นี้ได้กล่าวเอาไว้ทั้ง ๓ วิธี เพื่อเป็นแนวทางในการวิจัยสำหรับฟังก์ชั่นการแจกแจงในรูปแบบอื่น ๆ

---

[๑]George B. Thomas, JR. "calculus and Analytic Geometry", Addison - Wesley Publishing Company, Inc. London, p. 207-209, 1967.

[๒]Ibid., p. 386-387.

[๓]IBM. Application program System/360 Scientific Subroutine Package (360 A-CX-03X) Version III, 1968.

๓.๑.๑ กฎเกณฑ์การหาพื้นที่สี่เหลี่ยมคางหมู (Trapezoidal Rule) [๑]

ถ้า $Y = f(x)$

$$\int_{X_1}^{X_2} f(x)\,dx = (X_2 - X_1)\frac{(Y_1 + Y_2)}{2}$$

เมื่อ $Y_1 = f(X_1)$

$Y_2 = f(X_2)$

ค่าคาดหมายของจำนวนข้อมูลในช่วงที่ i (E_i) จะได้

$$E_i = N \times \int_{X_i}^{X_{i+1}} f(x)\,dx$$

N = จำนวนข้อมูลทั้งหมด (Total Observation)

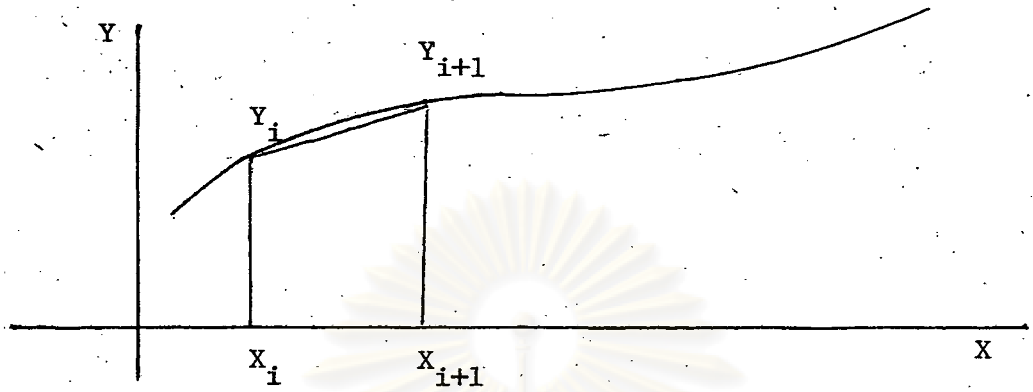
$\int_{X_i}^{X_{i+1}} f(x)\,dx$ = ความน่าจะเป็น (Probability) ที่อยู่ในช่วง (X_i , X_{i+1}) ซึ่งความหมาย $\int_{X_i}^{X_{i+1}} f(x)\,dx$ ก็คือ พื้นที่ภายใต้กราฟ f(x) เมื่อ x อยู่ในช่วง (X_i ,X_{i+1}) ดังรูป ๓.๑ ถ้า (X_i, X_{i+1}) เป็นช่วงเล็ก ๆ ของค่าของข้อมูล พื้นที่ภายใต้กราฟจะประมาณ $\frac{1}{2}(Y_{i+1} + Y_i)(X_{i+1} - X_i)$ ซึ่งเป็นสูตรที่ใช้หาพื้นที่ของรูปสี่เหลี่ยมคางหมู

ถ้า (X_i , X_{i+1}) เป็นช่องกว้างมาก ๆ การแทนพื้นที่ภายใต้กราฟด้วยสูตรดังกล่าวจะเกิดความผิดพลาดมาก ดังนั้น ถ้าช่องของค่าของข้อมูลกว้าง การหาพื้นที่ภายใต้กราฟในช่วงดังกล่าว เอาผลที่ได้ในแต่ละช่วงมารวมกัน

---

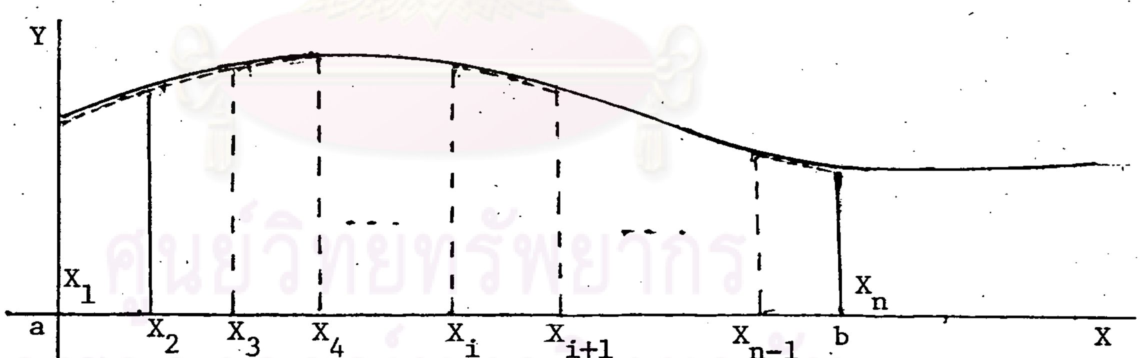
[๑] George B. Thomas, JR. "Calculus and Analytic Geometry", Addison-Wesley Publishing Company, Inc. London, p. 207-209, 1967.

รูปที่ ๓.๑



สำหรับความกว้างของช่องเล็ก ๆ จะมากน้อยแค่ไหนขึ้นอยู่กับลักษณะของกราฟของฟังก์ชั่น ถ้าแบ่งการหาพื้นที่เป็นช่องกว้าง ๆ จะได้ค่าประมาณพื้นที่ภายใต้กราฟดังกล่าวมีความคลาดเคลื่อนจากค่าที่ควรจะเป็นมาก ดังนั้นถ้าจะอินทิเกรตฟังก์ชั่นในช่วง $[a,b]$ ซึ่งห่างกันตามรูปที่ ๓.๒

รูปที่ ๓.๒



จะแบ่งช่วงย่อย ๆ โดยสมมติให้ในช่วง $[a,b]$ มี n-1 ช่องหรือมี n จุด โดยที่แต่ละช่วงเท่ากันดังนี้ $X_2 - X_1 = X_3 - X_2 = \ldots = X_n - X_{n-1} = h$ เมื่อ $h = \frac{b-a}{n-1}$

$$Y_1 = f(X_1)\ ,\ Y_2 = f(X_2)\ \ldots\ldots,\ Y_i = f(X_i)$$

$$\int_a^b f(x)\,dx = \int_{X_1}^{X_2} f(x)\,dx + \int_{X_2}^{X_3} f(x)\,dx + \ldots + \int_{X_i}^{X_{i+1}} f(x)\,dx + \ldots + \int_{X_{n-1}}^{X_n} f(x)\,dx$$

$$\approx \frac{1}{2}\left[(Y_1+Y_2)(X_2-X_1)+(Y_2+Y_3)(X_3-X_2)+\ldots+(Y_{n-1}+Y_n)(X_n-X_{n-1})\right]$$

$$\approx \frac{h}{2}\left[Y_1+2Y_2+\ldots+2Y_{n-1}+Y_n\right]$$

ดังนั้น

$$\int_a^b f(x)\,dx \approx \frac{b-a}{2(n-1)}(Y_1+2Y_2+2Y_3+\ldots+2Y_{n-1}+Y_n)$$

(สำหรับช่วงย่อยของค่าของข้อมูลที่กว้างเท่า ๆ กัน)

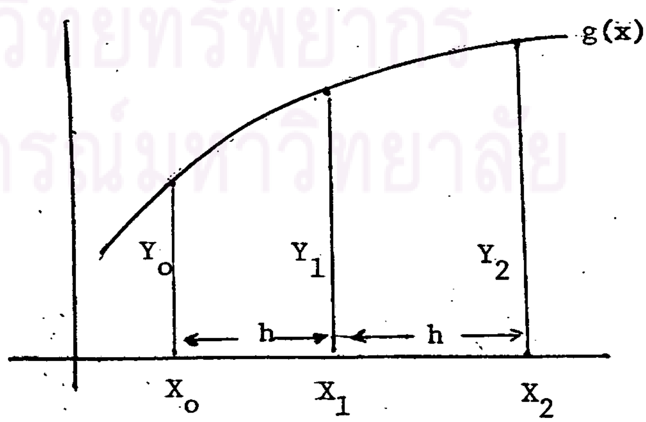
ถ้า $X_i - X_{i-1} \neq X_j - X_{j-1}$; $i \neq j$

$$\int_a^b f(x)dx \approx \frac{1}{2}\left[(Y_1+Y_2)(X_2-X_1)+(Y_2+Y_3)(X_3-X_2)+\ldots+(Y_{n-1}+Y_n)(X_n-X_{n-1})\right]$$

๓.๑.๒ กฎเกณฑ์การหาพื้นที่โดยใช้กราฟเส้นโค้ง (Simpson's Rule)[๒]

รูปที่ ๓.๓

$$E_i = N \times \int_{X_i}^{X_{i+1}} g(x)\,dx$$



---

[๒] George B. Thomas, JR. "Calculus and Analytic Geometry", Addison-Wesley Publishing Company, Inc. London, p.386-387, 1967.

ถ้า

$$g(x) = p + q(x) + rx^2$$

$$x_1 = x_o + h$$

$$x_2 = x_1 + h$$

$$y_o = g(x_o) = p + q(x_o) + rx_o^2$$

$$y_1 = g(x_1) = p + q(x_1) + r(x_1)^2$$

$$y_2 = g(x_2) = p + q(x_2) + r(x_2)^2$$

$$\int_{x_o}^{x_2} g(x)\,dx = \int_{x_o}^{x_2} (p + qx + rx^2)\,dx$$

$$= p\int_{x_o}^{x_2} dx + q\int_{x_o}^{x_2} x\,dx + r\int_{x_o}^{x_2} x^2\,dx$$

$$= p(x_2 - x_o) + q\left(\frac{x_2^2}{2} - \frac{x_o^2}{2}\right) + r\left(\frac{x_2^3}{3} - \frac{x_o^3}{3}\right)$$

$$= p.2\,h + \frac{q}{2}(x_2 - x_o)(x_2 + x_o) + \frac{r}{3}(x_2 - x_o)(x_2^2 + x_2\,x_o + x_o^2)$$

$$= 2ph + \frac{2qh}{2}(x_2 + x_o) + \frac{2rh}{3}(x_2^2 + x_2\,x_o + x_o^2)$$

$$= 2h\left[p + \frac{q}{2}(x_2 + x_o) + \frac{r}{3}(x_2^2 + x_2\,x_o + x_o^2)\right]$$

$$= \frac{2h}{6}\left[6p + 3q(x_2 + x_o) + 2r(x_2^2 + x_2\,x_o + x_o^2)\right]$$

$$= \frac{h}{3}\left[6p + 3qx_2 + 3qx_o + 2rx_2^2 + 2r\,x_2\,X_o + 2r\,X_o^2\right]$$

$$= \frac{h}{3}\left[(p + qx_o + rx_o^2) + (p + qx_2 + rx_2^2) + 4p + 2qx_o + 2qx_2 + rx_o^2 + 2rx_o\,x_2 + rx_2^2\right]$$

$$= \frac{h}{3}\left[y_o + y_2 + 4p + 2q(x_1 - h) + 2q(x_1 + h) + r(x_1 - h)^2\right.$$

$$\left. + r(x_1 + h)^2 + 2r(x_1 + h)(x_1 - h)\right]$$

$$= \frac{h}{3}\left[y_o + y_2 + 4p + 4qx_1 + 4rx_1^2\right]$$

$$= \frac{h}{3}\left[Y_o + y_2 + 4(p + qx_1 + rx_1^2)\right]$$

$$= \frac{h}{3}\left[y_o + y_2 + 4y_1\right] = \frac{h}{3}\left[y_o + 4y_1 + y_2\right]$$

ถ้าช่องการหาพื้นที่กว้างมาก ทำได้โดยการแบ่งเป็นช่องย่อย ๆ แล้วแต่ละช่วงย่อยก็หาพื้นที่ตามที่กล่าวมาแล้วข้างต้น แล้วนำพื้นที่ทั้งหมดมารวมกัน เช่น ต้องการหา

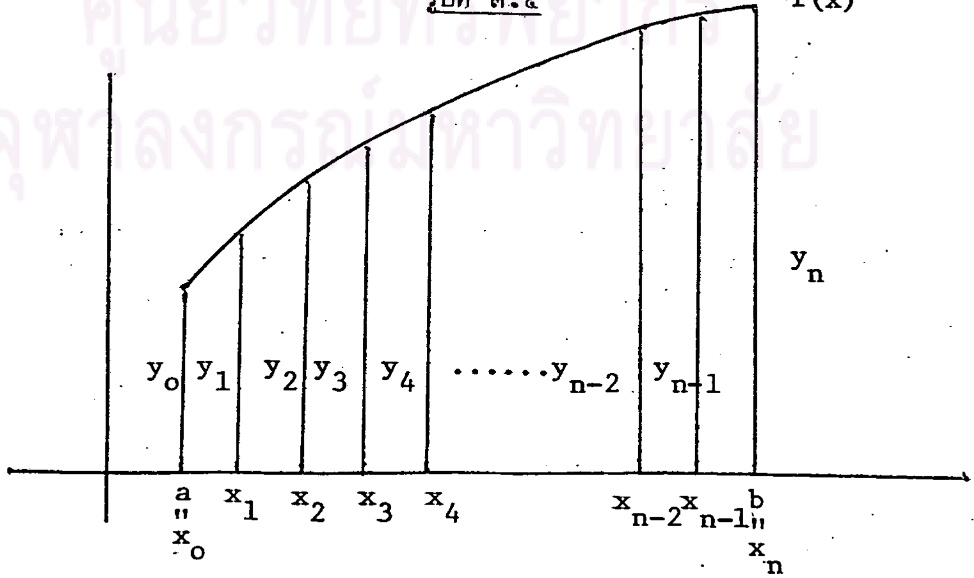
$$\int_a^b f(x)\,dx$$

แบ่งช่อง [a,b] ออกเป็น n ช่องเท่า ๆ กันด้วยจุด

$$a = x_o < x_1 < x_2 < \ldots < x_n = b$$

$$\therefore\ h = \frac{b-a}{n} \quad \text{และให้} \quad \int_a^b f(x)\,dx = \int_a^b f$$

รูปที่ ๓.๔



จากรูปที่ ๓.๔ $\int_a^b f = \int_{x_o}^{x_2} f + \int_{x_2}^{x_4} f + \ldots + \int_{x_{n-2}}^{x_n} f$

เราจะสร้างกราฟของฟังก์ชั่นบนแต่ละช่วงย่อย $|x_o, x_2|$, $|x_2, x_4|, \ldots, |x_{n-2}, x_n|$ โดยให้เป็นกราฟของฟังก์ชั่นกำลังสอง $g(x) = p + q(x) + rx^2$ แล้วหาอินทิกรัลของ $\int_a^b g$

$$\therefore \int_a^b g = \int_{x_o}^{x_2} g + \int_{x_2}^{x_4} g + \ldots + \int_{x_{n-2}}^{x_n} g$$

$$= \frac{h}{3}\left[y_o + 4y_1 + y_2\right] + \frac{h}{3}\left[y_2 + 4y_3 + y_4\right] + \ldots$$

$$\ldots + \frac{h}{3}\left[y_{n-2} + 4y_{n-1} + y_n\right]$$

$$= \frac{h}{3}\left[y_o + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \ldots + y_{n-2} + 4y_{n-1} + y_n\right]$$

$$= \frac{h}{3}\left[y_o + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \ldots + 2y_{n-2} + 4y_{n-1} + y_n\right]$$

จะเห็นได้ว่า $\int_a^b f$ $\int_a^b g$ และบรรดา $y_o, y_1, \ldots, y_n$ ต่างก็เป็นค่าของฟังก์ชั่น ซึ่งอยู่บนกราฟของ $f(x)$ ดังนั้นเราสามารถหาค่า $y_o, y_1, \ldots, y_n$ เหล่านี้ได้จาก $f(x)$ โดยตรง โดยที่ $y_o = f(x_o)$, $y_1 = f(x_1), \ldots, y_n = f(x_n)$

ดังนั้นเราสรุปสูตรได้ดังนี้

$$\int_a^b f(x)\, dx \simeq \frac{h}{3} \left|y_o + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \ldots + y_n\right|, n \text{ เลขคู่}$$

๓.๑.๓ การหาพื้นที่โดยการประมาณค่าจากการแจกแจงปกติที่มีค่าเฉลี่ยเป็นศูนย์และค่าความเบี่ยงเบนมาตรฐานเป็นหนึ่ง ๑

ถ้าเราต้องการคำนวณค่า $Y = p(x) = p(X < x)$

$$p(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2)\, du$$

[๑] IBM Application program System/360 Scientific Subroutine Package (360A-CX-03X) Version III, p. 78, 1968.

ฟังก์ชั่นที่เราจะหาค่านี้ อินทิเกรตโดยตรงไม่ได้ เราอาจใช้วิธีที่ในหัวข้อ ๓.๑.๑ หรือ ๓.๑.๒ ประมาณค่าได้เช่นกัน แต่สำหรับในที่นี้เราใช้การประมาณค่าด้วยสูตรดังนี้

$$p(x) = 1 - f(x) \sum_{i=1}^{5} a_i w^i ; \quad x > 0$$

โดยที่ :

$$w = 1/(1 + px)$$

$$f(x) = \exp(-x^2/2)/\sqrt{2\pi}$$

$$p = 0.2316419$$

$$\left.\begin{aligned} a_1 &= 0.3193815 \\ a_2 &= -0.3565638 \\ a_3 &= 1.781478 \\ a_4 &= -1.821256 \\ a_5 &= 1.330274 \end{aligned}\right\} \text{คือค่าคงที่ซึ่งนำไปใช้งานได้เลย}$$

การประมาณค่าโดยวิธีนี้จะมีความผิดพลาดประมาณ $7(10^{-7})$[๑]

$$\therefore \text{ ถ้าจะหาค่า } \int_a^b f(x)\,dx \simeq P(a < x < b)$$

$$\simeq P(\hat{b}) - P(\hat{a})$$

โดยที่ $P(\hat{b})$ และ $P(\hat{a})$ คำนวณค่ามาจากสูตร $P(x)$ ดังกล่าวข้างต้น โดยถือว่าข้อมูลของช่อง $|a,b|$ ต้องทำเป็นการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็นศูนย์และค่าความเบี่ยงเบนเป็นหนึ่ง นั่นคือ $\hat{a} = \dfrac{a - \bar{x}}{S_x}$, $\hat{b} = \dfrac{b - \bar{x}}{S_x}$ โดยที่ $\bar{x}$, S_x เป็นค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของข้อมูลตัวอย่างทั้งหมด ซึ่งการประมาณค่าโดยวิธีนี้ได้เขียนเป็นโปรแกรมไว้แล้ว (ดูในภาคผนวก ก) โดยใช้ชื่อว่า NDTR

---

[๑] เรื่องเดียวกัน, หน้า ๗๘.

ในการวิจัยนี้ผู้วิจัยมีความประสงค์ที่จะศึกษาถึงวิธีการ optimization technique ที่จะทำให้ได้ตัวประมาณค่าของพารามิเตอร์ที่ดีที่สุดหรือเหมาะสมที่สุดในฟังก์ชั่นการแจกแจงของข้อมูลที่ต้องการจะศึกษา ว่ามีประสิทธิภาพมากน้อยเพียงใด จึงมีความจำเป็นที่จะต้องสร้างตัวเลขสุ่ม (Random number) ที่เข้ากับการแจกแจงมาตรฐานชนิดใดชนิดหนึ่งขึ้นมาเสียก่อน เพื่อใช้เป็นตัวแบบในการจัดความถูกต้องของค่าประมาณของพารามิเตอร์ที่ได้ ว่ามีความถูกต้องมากน้อยเพียงใด การแจกแจงมาตรฐานที่สร้างขึ้นในการวิจัยนี้คือการแจกแจงแบบปกติ (Normal Distribution) ตัวเลขสุ่มแบบปกตินี้จะใช้เป็นประชากรที่เราสนใจจะศึกษา ซึ่งจะกล่าวถึงวิธีการสร้างเลขสุ่มที่มีการแจกแจงแบบปกติในหัวข้อ ๓.๒

ในทางปฏิบัติเราไม่สามารถศึกษาประชากรทั้งหมดได้ แต่ศึกษาตัวอย่างที่เลือกมาจากประชากรเพียงบางส่วนเท่านั้น ซึ่งวิธีการเลือกตัวอย่างในการวิจัยนี้จะได้กล่าวถึงวิธีการเลือกตัวอย่างในหัวข้อ ๓.๓

๓.๒ การสร้างเลขสุ่ม

ในวิทยานิพนธ์ฉบับนี้เรามีความประสงค์จะนำเลขสุ่มที่มีการแจกแจงแบบปกติไปใช้ในการวิจัย อย่างไรก็ดีเนื่องจากตัวเลขสุ่มที่จะมีการแจกแจงแบบปกติได้นั้น จำเป็นต้องสร้างตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ (Uniform Distribution) ขึ้นมาก่อน[๑] แล้วจึงนำเลขสุ่มนี้มาสร้างเป็นเลขสุ่มที่มีการแจกแจงแบบปกติอีกครั้งหนึ่ง เพราะเหตุว่าในการสร้างเลขสุ่มหรือการจำลองจะประสบผลสำเร็จเพียงใดนั้น ขึ้นอยู่กับคุณภาพของตัวเลขสุ่มเป็นสำคัญในการสร้างเลขสุ่มใด ๆ จะต้องสามารถสร้างอนุบรรพสุ่ม (Random Sequence) ตามการกระจายความน่าจะเป็นของขบวนการนั้น โดยปกติจะใช้เลขสุ่มที่มีการกระจายสม่ำเสมอเพื่อกำหนดค่าของตัวแปรเชิงสุ่มตามการกระจายที่ต้องการ ดังนั้นวิธีกำเนิดเลขสุ่มที่มีการกระจายสม่ำเสมอจึงเป็นรากฐานที่มีความสำคัญต่อการนำไปใช้ในการวิจัยในวิทยานิพนธ์ฉบับนี้ต่อไป

---

[๑]จักร ตามประทีป, "นโยบายที่ดีที่สุดสำหรับโรงไฟฟ้าพลังกังหันลม" วิทยานิพนธ์คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์, หน้า ๗๖, พ.ศ. ๒๕๒๐.

๓.๒.๑ หลักการสร้างเลขสุ่มแบบสม่ำเสมอ (Uniform Random Number)

เทคนิคการสร้างและสุ่มแบ่งกว้าง ๆ ได้เป็น ๒ ประเภทดังนี้

ประเภทที่ ๑ เทคนิคโปรแกรม[๑] วิธีนี้เลขสุ่มจะกำเนิดขึ้นโดย Recursive Relation ซึ่งหมายความว่าตัวเลขสุ่มตัวถัดไปจะต้องใช้ตัวเลขสุ่มตัวปัจจุบันหรือกลุ่มตัวเลขสุ่มในอดีต เทคนิคโปรแกรมนี้จะให้อนุบรรพอันตะ (Finite Sequence) ของตัวเลขเท่านั้น ความจริงอนุบรรพอันตะจะเป็นอนุบรรพสุ่มตามความหมายที่แท้จริงของการสุ่มไม่ได้ แต่ถ้าอนุบรรพอันตะใดที่มีคุณสมบัติบางประการของอนุบรรพสุ่ม จะถูกยอมรับว่าเป็นอนุบรรพคล้ายสุ่ม (Pseudo-Random Sequence)

ประเภทที่ ๒ เครื่องมือเลขสุ่ม[๒] (Random Number Device) วิธีนี้จะแปลงกระบวนการสุ่มกายภาพให้เป็นอนุบรรพของเลขฐานสอง ซึ่งเป็นเลขสุ่ม

$$IX_{n+1} = C \times IX_n \pmod{m}$$

โดยที่ IX_o , C , m เป็นเลขจำนวนเต็มบวก เมื่อนำความสัมพันธ์นี้มาใช้กับเครื่องคอมพิวเตอร์ที่ทำงานด้วยเลขฐาน ๒ แล้ว จะกำหนดให้ $m = 2^k$ ถ้า IX_o เป็นเลขคู่ เลขในหลักท้าย ๆ ของ IX_o ในฐาน ๒ จะเป็น ๐ จึงเป็นการสิ้นเปลืองหลักในเลขฐาน ๒ โดยมิได้ประโยชน์แต่อย่างใด และจำนวนเลข ๐ ในหลักท้าย ๆ ของเลข IX มีโอกาสสะสมมากขึ้น จึงควรเลือก IX_o เป็นเลขคี่ ในทำนองเดียวกัน C ก็ควรเป็นเลขคี่ หาก C เป็นเลขคู่ การสะสมเลข ๐ ในหลักท้าย ๆ ของ IX จะเป็นไปอย่างรวดเร็ว จนทำให้อนุบรรพกลายเป็นอนุบรรพเลข ๐ หมด ในการวิจัยนี้ใช้ $m = 2^{31} = 2,147,483,648$

---

[๑] เรื่องเดียวกัน, หน้า ๗๑

[๒] เรื่องเดียวกัน, หน้า ๗๒

๓.๒.๒ หลักการสร้างเลขสุ่มแบบปกติ (Normal Random Number)

ในการสร้างเลขสุ่มแบบปกติ จำเป็นจะต้องกำหนดพารามิเตอร์ค่าเฉลี่ยและค่าความแปรปรวนของตัวเลขสุ่มตามที่ต้องการ และตัวเลขสุ่มจะมีลักษณะการแจกแจงเข้าใกล้การแจกแจงแบบปกติได้นั้น จะพบว่าตัวเลขที่ได้จะได้มาจากตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ ซึ่งเป็นตัวเลขสุ่มที่ได้จากหัวข้อ ๓.๒.๑ โดยใช้สูตร

$$Y = \frac{\sum_{i=1}^{k} X_i - \frac{k}{2}}{\sqrt{k/12}}$$

โดยที่ Y = เลขสุ่มที่มีการแจกแจงแบบปกติ

X_1 = เลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ

k = จำนวนค่าของ X_i ที่จะถูกนำมาใช้

โดยปกติแล้ว ตัวเลขสุ่ม Y จะมีค่าเข้าใกล้เลขสุ่มที่มีการแจกแจงแบบปกติที่แท้จริงนั้น เมื่อค่าของ k เข้าใกล้ค่าอนันต์ (Infinity)

สำหรับโปรแกรมที่ใช้สร้างเลขสุ่มนี้จะเลือก k เป็น ๑๒ เพื่อลดเวลาในการคำนวณ (Execution) ในเครื่องคอมพิวเตอร์ ดังนั้น จากสูตรข้างบนจะได้สูตรใหม่ดังนี้

$$Y = \sum_{i=1}^{12} X_i - 6.0$$

และเพื่อให้ตัวเลขสุ่มที่สร้างขึ้นมาแจกแจงเข้าใกล้การแจกแจงแบบปกติโดยที่มีค่าเฉลี่ยและค่าความแปรปรวนตามที่ต้องการดังนั้น $Y' = Y \times S + AM$

โดยที่ Y' = คือตัวเลขสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยและค่าความแปรปรวนที่ต้องการ

S = ค่าเบี่ยงเบนมาตรฐานที่กำหนด

AM = ค่าเฉลี่ยที่กำหนด

ในวิทยานิพนธ์นี้การสร้างเลขสุ่มแบบสม่ำเสมอ (Uniform Random Number) และการสร้างเลขสุ่มแบบปกติ (Normal Random Number) ผู้วิจัยไม่ได้เขียนโปรแกรมขึ้นเอง แต่ใช้โปรแกรมสำเร็จรูปชื่อ GAUSS[๑] สำหรับสร้างเลขสุ่มแบบปกติ ซึ่งจะต้องเรียกโปรแกรมการสร้างเลขสุ่มแบบสม่ำเสมอโดยใช้โปรแกรมชื่อ RANDU[๒] รายละเอียดของโปรแกรมอยู่ในภาคผนวก ก

๓.๓ <u>การสุ่มตัวอย่างแบบมีระบบ</u> (Systematic Sampling)[๓]

ถ้าประชากรประกอบด้วยสมาชิก N หน่วยเรียงกันจาก ๑ ถึง N แบบสุ่ม การสุ่มตัวอย่างอาจจะใช้วิธีการสุ่มแบบมีระบบ ซึ่งเป็นวิธีการสุ่มที่ง่ายและสะดวกกับการใช้เครื่องคอมพิวเตอร์

สมมติว่าต้องการสุ่มตัวอย่างมาเพียง ๕ เปอร์เซนต์ของจำนวนประชากรทั้งหมด ก็เลือกตัวอย่างจากสมาชิกของประชากรที่เรียงกันอยู่นั้น จำนวน ๑ คน ต่อประชากรจำนวนทุก ๆ ๒๐ คน โดยเริ่มต้นด้วยการสุ่มหาตัวเลขตั้งต้น (random start) เสียก่อน ซึ่งทำได้ง่าย ๆ โดยการสุ่มเลือกเลข ๑ ตัวมาจากเลข ๑ - ๒๐ สมมติว่าได้ ๓ ตัวอย่างที่จะเลือกมาทั้งหมดก็คือประชากรลำดับที่ ๓, ๒๓, ๔๓, ๖๓, ๘๓, ... นั่นคือตัวอย่างที่ต้องการมีขนาด n จะต้องหาตัวเลข sampling interval I ให้มีค่าใกล้เคียง $\frac{N}{n}$ แล้วสุ่มเลือก random start มาจากเลขระหว่าง ๑ - I สมมติได้ ๑๐ ประชากรลำดับที่ ๑๐ + I, ๑๐ + ๒I, ๑๐ + ๓I, ... ก็จะถูกบังคับให้เลือกออกมาโดยอัตโนมัติ

ผู้วิจัยได้อาศัยหลักการข้างบนนี้ในการสุ่มตัวอย่างจากประชากรที่ได้จากการใช้วิธีที่ได้อธิบายไว้ในหัวข้อ ๓.๒ และได้เขียนโปรแกรมการสุ่มตัวอย่างด้วยวิธีนี้ขึ้นเองโดยใช้ชื่อโปรแกรมว่า SMPLNG สำหรับรายละเอียดของโปรแกรมอยู่ในภาคผนวก ก

---

[๑]System/360 Scientific Subroutine Package (360A-CM-03X) Version III Programmer's Manual page 77-78.

[๒]เรื่องเดียวกัน.

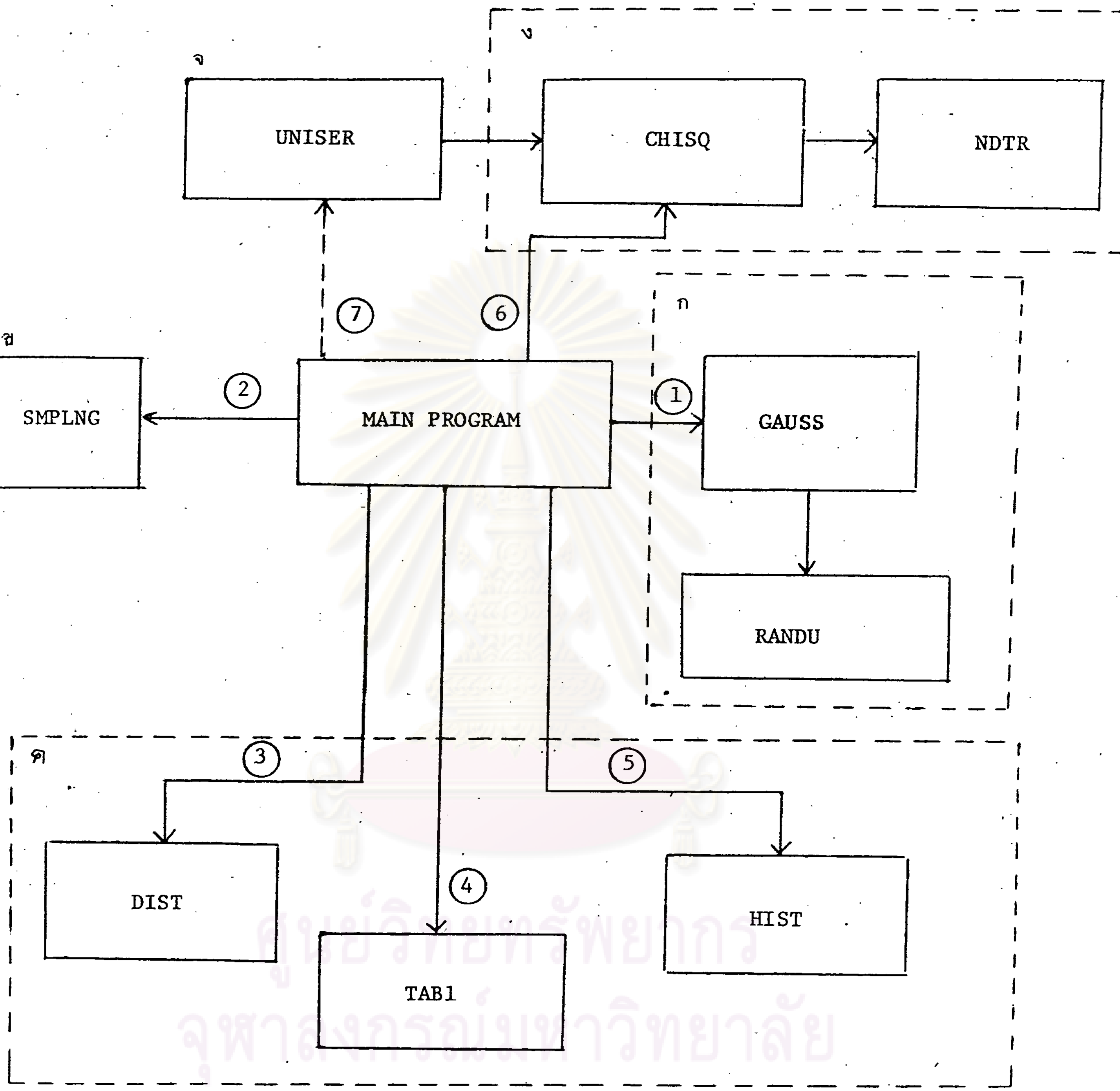
[๓]วีนัส มีชวณิชย์ และ สมจิต วัฒนาชยกุล, "สถิติสำหรับนักสังคมศาสตร์", (มหาวิทยาลัยธรรมศาสตร์ พ.ศ. ๒๕๑๙) หน้า ๑๗๒ - ๑๗๓.

๓.๔ โปรแกรมที่ใช้งานทั้งหมด

ลำดับที่	ชื่อโปรแกรม	คุณสมบัติของโปรแกรม	จำนวนคำสั่ง	แหล่งที่มา	ชื่อโปรแกรมย่อยที่เรียกใช้
๑	GAUSS	สร้างเลขสุ่มที่มีการแจกแจงแบบปกติ	๑๑	*	RANDU
๒	RANDU	สร้างเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ		*	-
๓	SMPLNG	วิธีการสุ่มตัวอย่างจากเลขสุ่มโดยใช้วิธีการสุ่มตัวอย่างแบบมีระบบ	๘	เขียนเอง	-
๔	DIST	สร้างตารางแจกแจงความถี่	๕๕	เขียนเอง	-
๕	TAB1	หาค่าเฉลี่ยค่าความเบี่ยงเบนมาตรฐาน ค่าสูงสุดและค่าต่ำสุดของข้อมูล	๕๖	*	-
๖	HIST	จากกราฟหรือสร้างฮิสโตแกรมให้กับข้อมูลที่แจกแจงความถี่แล้ว	๔๒	*	-
๗	CHISQ	ใช้ทดสอบภาวะสารูปสนิทดี (goodness of fit test)	๔๖	เขียนเอง	NDTR
๘	NDTR	ใช้ประมาณพื้นที่ใต้กราฟของฟังก์ชั่น การแจกแจงแบบปกติเพื่อใช้หาค่าคาดหมายของความถี่		*	-
๙	UNISER	ใช้ในการหาค่าเฉลี่ยและค่าความเบี่ยงเบนที่ทำทำให้ไคสแควร์ต่ำสุด	๘๑	เขียนเอง	CHISQ

สำหรับวิธีการใช้ตลอดจนรายละเอียดของแต่ละโปรแกรมมีอยู่ในภาคผนวก ก

*IBM Application program System/360 Scientifies Subroutine Package (360A-CM-03X) Version III



รูปที่ ๓.๕

ผังภาพแสดงขั้นตอนต่าง ๆ ในการทำงานของโปรแกรมในการวิจัยทั้งหมด

๓.๕ โปรแกรมที่ใช้งานทั้งหมด

รูปที่ ๓.๕ นี้ เป็นผังภาพแสดงขั้นตอนต่าง ๆ ที่จะต้องนำมาใช้ในวิทยานิพนธ์ฉบับนี้ทั้งหมด ได้แก่

(ก) การจัดเตรียมข้อมูลทั้งหมด ๒๐๐๐ ตัว เพื่อใช้เป็นแบบจำลองของประชากรที่ต้องการจะศึกษา

(ข) การสุ่มตัวอย่างจากประชากรที่จัดเตรียมได้ในข้อ (ก) เพื่อใช้เป็นตัวอย่างสำหรับทำการศึกษาจริง ๆ แทนประชากรที่เรากำลังสนใจ

(ค) โปรแกรมช่วยงาน ช่วยคำนวณค่าเฉลี่ยค่าความเบี่ยงเบนมาตรฐาน (μ , σ) สร้างฮิตโตแกรมให้กับข้อมูลในข้อ (ก) และคำนวณค่าเฉลี่ยค่าความเบี่ยงเบนมาตรฐาน ($\bar{X}$, S) สร้างฮิตโตแกรมให้กับข้อมูลที่ได้ในข้อ (ข) ด้วย

(ง) โปรแกรมคำนวณค่าไคสแควร์เพื่อทดสอบลักษณะการแจกแจงของข้อมูลว่าเข้ากับลักษณะการแจกแจงของฟังก์ชั่นการแจกแจงชนิดใด (Goodness of fit test)

(จ) โปรแกรมคำนวณค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานตัวใหม่ ($\bar{X}^*$, S^*) ของข้อมูลที่ได้ในข้อ (ข) โดยการ Optimize ค่าไคสแควร์ให้มีค่าต่ำสุด สำหรับขั้นตอนนี้ UNISER ไม่ใช่โปรแกรมย่อยของ MAINPROGRAM โดยตรง เพียงแต่นำเอาค่า AM, S, O, W, PF จาก MAINPROGRAM มาใช้เท่านั้น

(ฉ) เพื่อความมั่นใจว่าวิธีการ Optimization Method ใช้คำนวณค่าประมาณของพารามิเตอร์ คือค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานที่ทำให้ค่าสถิติทดสอบไคสแควร์ต่ำที่สุดนั้น ยังเป็นตัวประมาณค่าพารามิเตอร์ที่ดีพอหรือไม่นั้น การที่จะทราบว่าตัวประมาณค่าพารามิเตอร์ตัวใหม่ที่ได้เป็นตัวประมาณค่าพารามิเตอร์ที่ดีพอนั้น โดยปกติถ้าเราสามารถหาสูตรในทางคณิต-ศาสตร์ของค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานตัวใหม่ได้ เราก็จะทำการพิสูจน์ได้ตามหลักทางคณิตศาสตร์และสถิติ แต่เนื่องจากว่าค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานค่าใหม่ที่ได้จากการใช้วิธี Optimization Method นี้ไม่สามารถเขียนออกมาเป็นสูตรได้ จึงมีความจำเป็นต้องใช้ตัวอย่างในการศึกษาหลาย ๆ ตัวอย่าง เพื่อนำผลทั้งหมดมาทดสอบทางสถิติโดยใช้การทดสอบแบบ t-test ซึ่งในการวิจัยนี้การทดสอบแบบ t-test ไม่ได้ใช้คอมพิวเตอร์คำนวณ แต่ใช้เครื่องคิดเลขช่วยในการคำนวณ