บทที่ 2

ทฤษฎีที่เกี่ยวข้องในการวิจัย

การวิจัยเรื่องนี้จำเป็นต้องใช้ทฤษฎีทางสถิติที่เกี่ยวข้องดังต่อไปนี้

2.1 การแจกแจงแบบปกติสองตัวแปร (bivariate normal distribution)

ถ้า (X, Y) เป็นตัวแปรที่มีการแจกแจงแบบปกติฟังก์ชันการแจกแจงความน่าจะเป็นของตัวแปรทั้งสองคือ

$$f(X,Y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[\frac{1}{-2(1-\rho^2)}\left\{\left(\frac{X-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{X-\mu_x}{\sigma_x}\right)\cdot\left(\frac{Y-\mu_y}{\sigma_y}\right) + \left(\frac{Y-\mu_y}{\sigma_y}\right)^2\right\}\right]$$

เมื่อ $-\infty < x < \infty$; $-\infty < y < \infty$ และ $-1 \leqslant \rho \leqslant 1$; $\sigma_x^2 \geqslant 0$; $\sigma_y^2 \geqslant 0$

$-\infty < \mu_x < \infty$; $-\infty < \mu_y < \infty$

โดยมีค่าเฉลี่ยและความแปรปรวนดังนี้

$$\underset{\sim}{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\,\sigma_x\sigma_y \\ \rho\,\sigma_y\,\sigma_x & \sigma_y^2 \end{bmatrix}$$

เมื่อ ρ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปร X และ Y

μ_x คือ ค่าเฉลี่ยของตัวแปร X

μ_y คือ ค่าเฉลี่ยของตัวแปร Y

σ_x^2 คือ ความแปรปรวนของตัวแปร X

σ_y^2 คือ ความแปรปรวนของตัวแปร Y

สำหรับการวิจัยนี้ ศึกษาเฉพาะข้อมูลที่มีการแจกแจงแบบปกติสองตัวแปร (X_i, Y_i) เป็น independent identically distributed

โดย $i = 1,2,\ldots, N$

$$(X_i , Y_i) \sim N(\mu_1 , \mu_2 , \sigma_1^2 , \sigma_2^2 , \rho)$$

ทำการจัดเรียงลำดับข้อมูล (Ordered data) ของตัวแปร X จากน้อยไปมาก (X_{Nk}) เป็น X order statistic ที่ k

$$X_{N1} < X_{N2} < \ldots < X_{Nk} < \ldots < X_{NN}$$

และค่าของตัวแปร Y จะสัมพันธ์กับค่าของตัวแปร X ที่เรียงลำดับแล้วแบบ Concomitant of order statistic ที่ k เป็นค่า Y_{Nk} $(Y_{N1}, Y_{N2}, \ldots, Y_{Nk}, \ldots, Y_{NN})$ โดยให้

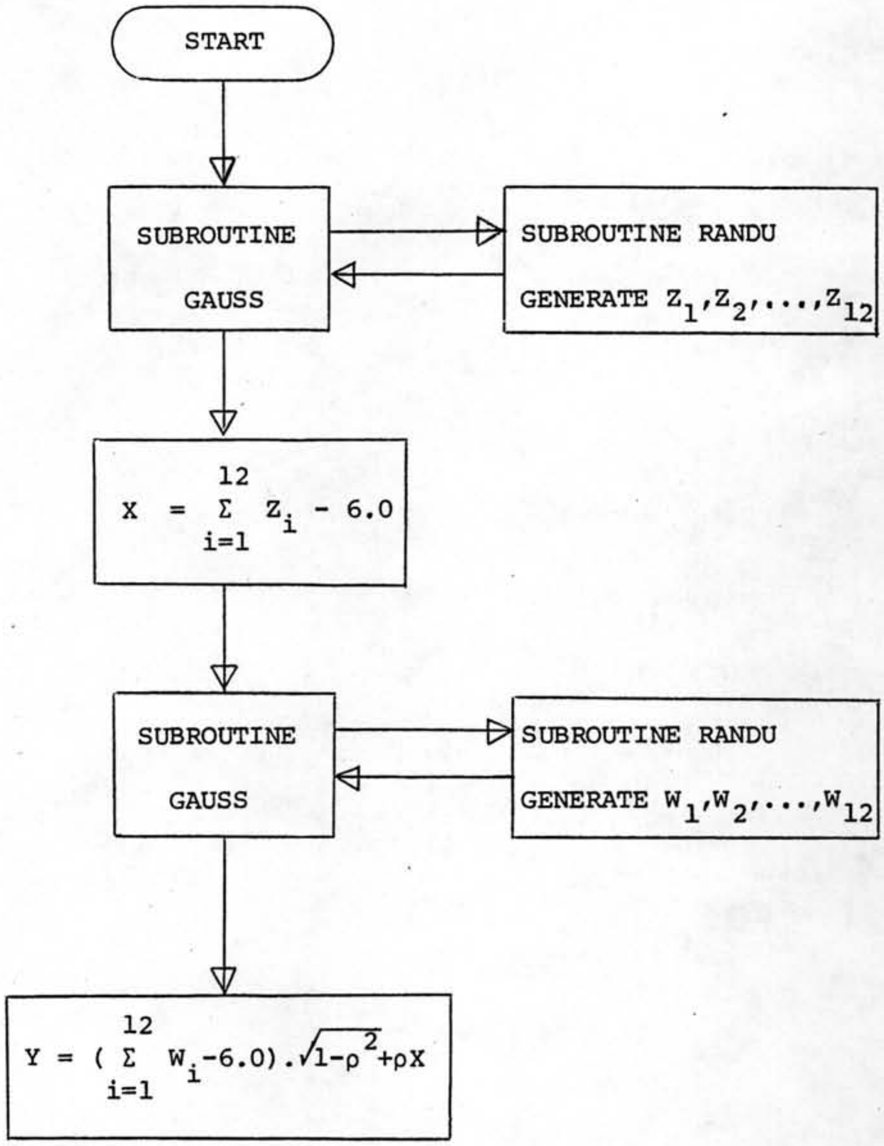
$$Y_{Nk} = Y_j \quad \text{เมื่อ}$$

$$X_{Nk} = X_j$$

2.2 การสร้างข้อมูลที่มีการแจกแจงแบบปกติสองตัวแปรโดยวิธีซิมูเลชั่น (Simulation)

ในการวิจัยครั้งนี้ ผู้วิจัยสร้างตัวแปรที่มีการแจกแจงปกติสองตัวแปร โดยการแทนค่า ρ ระดับต่าง ๆ แล้วหาตัวแปร X และ Y ที่มีความสัมพันธ์กันในระดับของ ρ ที่กำหนดขึ้น โดยเริ่มจากการสร้างตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ (Uniform distribution) ขึ้นมาก่อน โดยใช้โปรแกรม SUBROUTINE RANDU จากนั้นนำผลที่ได้จากการสร้างตัวเลขสุ่มแบบสม่ำเสมอ ไปสร้างตัวแปรแบบปกติมาตรฐาน โดยใช้โปรแกรม SUBROUTINE GAUSS โดยค่าที่ได้จากโปรแกรมนี้แทนด้วยตัวแปร X จากนั้นสร้างตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอขึ้นมาอีกชุดหนึ่ง ซึ่งไม่ขึ้นอยู่กับตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอชุดก่อน โดยใช้โปรแกรม SUBROUTINE RANDU จากนั้นนำค่าตัวเลขสุ่มมาสร้างตัวแปรที่มีการแจกแจงแบบปกติ โดยเรียกค่าที่ได้ว่า Y ค่าของ Y นี้จะมีลักษณะการแจกแจงที่ขึ้นอยู่กับค่าของ X ที่ได้จากครั้งแรก

การสร้างตัวแปรปกติมาตรฐาน 1 ตัว โดยใช้โปรแกรม SUBROUTINE GAUSS นี้ โปรแกรม SUBROUTINE GAUSS จะเรียกตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ 12 ตัว มาใช้ (ดูโปรแกรมในภาคผนวก) ผลลัพธ์ที่ได้จะเป็นค่า (X ,Y) ที่เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติสองตัวแปร และมีความสัมพันธ์ในระดับของ ρ ตามที่กำหนดดังแผนผังต่อไปนี้



<u>รูปที่ 1</u> ผังงานการสร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติสองตัวแปร

ผู้วิจัยได้อาศัยการสร้างเลขสุ่มจากเครื่องคอมพิวเตอร์ IBM/3031 มาใช้ในการสร้างตัวแปรปกติชนิดสองตัวแปร โดยใช้ SUBPROGRAM GAUSS และ RANDU

ทฤษฎีและข้อสมมติที่ใช้ประกอบเพื่อสรุปความถูกต้องของตัวแปรสุ่มแบบปกติสองตัวแปร คือ

1. ทฤษฎีลิมิตส่วนกลาง (Central limit theorem)
2. ตัวแปรมีลักษณะสุ่ม และมีความน่าจะเป็นเท่ากัน

มีสาระสำคัญดังนี้

ถ้า $Y_1, Y_2, Y_3, \ldots, Y_n$ แทนตัวอย่างสุ่มขนาด n ที่เป็นอิสระต่อกัน ซึ่งเลือกมาจากประชากรที่มีการแจกแจงแบบปกติ โดยมีค่าเฉลี่ย μ และค่าความแปรปรวน σ^2 แล้ว การแจกแจงของตัวสถิติ

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

ประมาณได้ด้วยการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2/n

เนื่องจากโปรแกรม GAUSS ใช้ตัวแปรที่มีการแจกแจงแบบสม่ำเสมอในช่วง $[0,1]$ 12 ตัว โดยมีค่าเฉลี่ยเท่ากับ 1/2 และค่าความแปรปรวนเท่ากับ 1/144

จะได้ $$\bar{Y} = \frac{1}{12} \sum_{i=1}^{12} Y_i \sim N(1/2 , 1/144)$$

ดังนั้น

$$X = \frac{\bar{Y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}}$$

$$= \frac{\bar{Y} - 1/2}{\sqrt{1/144}}$$

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

จะมีการแจกแจงโดยประมาณเป็นแบบปกติมาตรฐาน

2.3 ตัวสถิติที่ใช้วัดความสัมพันธ์

การศึกษาเรื่อง ความสัมพันธ์ระหว่างตัวแปร สำหรับการวิจัยนี้ ใช้ตัวสถิติที่ประมาณสัมประสิทธิ์สหสัมพันธ์ 4 วิธี คือ

2.3.1 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จากวิธี Grouped data ($\hat{\rho}_{grouped}$)

2.3.2 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จากวิธี Least Square ($\hat{\rho}_{LS}$)

2.3.3 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จากวิธีของ Bartlett and Wald ($\hat{\rho}_{BW}$)

2.3.4 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จากวิธีของ Cramer ($\hat{\rho}_{Cramer}$)

สูตรและที่มาของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์มีดังนี้

2.3.1 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จากวิธี Grouped data

$$\hat{\rho}_{grouped} = \frac{\sum_{i=1}^{n} (\bar{X}_i - \bar{X}^*)(\bar{Y}_i - \bar{Y}^*)}{\left[\sum_{i=1}^{n} (\bar{X}_i - \bar{X}^*)^2 \sum_{i=1}^{n} (\bar{Y}_i - \bar{Y}^*)^2\right]^{\frac{1}{2}}}$$

เมื่อ

$$\bar{X}^* = \frac{\sum_{i=1}^{N} X_i}{N}$$

$$\bar{Y}^* = \frac{\sum_{i=1}^{N} Y_i}{N}$$

$$\bar{X}_i = \frac{\sum_{k=1}^{m} X_k}{m}$$

$$\bar{Y}_i = \frac{\sum_{k=1}^{m} Y_k}{m}$$

จะเห็นได้ว่า ตัวประมาณค่าโดยวิธีนี้ มีสูตรในทำนองเดียวกับตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ แบบข้อมูลที่ยังไม่ได้จัดกลุ่ม (ตัวประมาณค่า ρ โดยวิธีของ Pearson) เพียงแต่ว่า ค่าสังเกตแต่ละค่าจะถูกแทนที่โดยค่าเฉลี่ยของข้อมูลในแต่ละกลุ่ม

2.3.2 <u>ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จากวิธี Least Square</u>

ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์จากวิธีของ Least Square นี้จะเกี่ยวข้องกับสัมประสิทธิ์ของสมการถดถอย (β) สำหรับข้อมูล X และ Y ที่มีการแจกแจงแบบปกติสองตัวแปร (Bivariate Normal distribution)

จาก

$$\rho = \beta \frac{\sigma_x}{\sigma_y}$$

โดยที่ค่า β นี้ได้จากข้อมูลที่ถูกจัดกลุ่มแล้ว โดยเรียงลำดับค่า ตัวแปร X ก่อน จากข้อมูล X และ Y ค่าคำนวณที่ได้ของ β มาจากการใช้วิธีของ least square หรือ modified version of estimators ซึ่งเสนอโดย Bartlett (1949) และ Wald (1940)

เมื่อใช้วิธี Least Square เพื่อประมาณค่าของ ρ ตัวประมาณค่า ρ คือ

$$\hat{\rho}_{LS} = \hat{\beta}_{LS} \frac{\hat{\sigma}_x}{\hat{\sigma}_y}$$

$$= \frac{\sum_{i=1}^{n} (\bar{X}_i-\bar{X}^*)(\bar{Y}_i-\bar{Y}^*) \left[\sum_{i=1}^{N} (X_i-\bar{X}^*)^2 /m\right]^{\frac{1}{2}}}{\sum_{i=1}^{n} (\bar{X}_i-\bar{X}^*)^2 \left[\sum_{i=1}^{N} (Y_i-\bar{Y}^*)^2/m\right]^{\frac{1}{2}}}$$

เนื่องจากค่าของตัวแปร X ถูกจัดลำดับและถูกจัดกลุ่ม ทำให้ค่า

$$\frac{\sum_{i=1}^{N} (X_i-\bar{X}^*)^2/m}{\sum_{i=1}^{n} (\bar{X}_i-\bar{X}^*)^2} \quad \text{มีค่า} \geq 1$$

ดังนั้น

$$\hat{\rho}_{LS} = \frac{\sum_{i=1}^{n} (\bar{X}_i-\bar{X}^*)(\bar{Y}_i-\bar{Y}^*)}{\left[\sum_{i=1}^{N} (X_i-\bar{X}^*)^2/m \quad \sum_{i=1}^{N} (Y_i-\bar{Y}^*)^2/m\right]^{\frac{1}{2}}}$$

โดย $\bar{X}^*$ และ $\bar{Y}^*$ คือ ค่าเฉลี่ยของ X และ Y ทุก ๆ ค่าตามลำดับ

$\bar{X}_i$ และ $\bar{Y}_i$ คือ ค่าเฉลี่ยของข้อมูลในแต่ละกลุ่มของตัวแปร X และ Y ตามลำดับ

2.3.3 <u>ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จากวิธีของ Bartlett and Wald</u>

จาก $\rho = \beta \dfrac{\sigma_X}{\sigma_Y}$

สำหรับกรณีที่ข้อมูลถูกจัดเป็นกลุ่ม Bartlett and Wald ได้ประยุกต์ตัวประมาณ β เป็น

$$\hat{\beta}_{BW} = \frac{\bar{Y}_{U3} - \bar{Y}_{L3}}{\bar{X}_{U3} - \bar{X}_{L3}}$$

เมื่อ $\bar{Y}_{U3}$ = ค่าเฉลี่ยของ $\bar{Y}_i$ สำหรับ $i \geq 2n/3$

$\bar{Y}_{L3}$ = ค่าเฉลี่ยของ $\bar{Y}_i$ สำหรับ $i \leq n/3$

$\bar{X}_{U3}$ = ค่าเฉลี่ยของ $\bar{X}_i$ สำหรับ $i \geq 2n/3$

$\bar{X}_{L3}$ = ค่าเฉลี่ยของ $\bar{X}_i$ สำหรับ $i \leq n/3$

ดังนั้น สำหรับกรณีนี้ ตัวประมาณของ ρ คือ

$$\hat{\rho}_{BW} = \hat{\beta} \frac{\hat{\sigma}_X}{\hat{\sigma}_Y}$$

$$= \left[\frac{\bar{Y}_{U3} - \bar{Y}_{L3}}{\bar{X}_{U3} - \bar{X}_{L3}}\right] \frac{\left[\sum_{i=1}^{N} (X_i - \bar{X}^*)^2/m\right]^{\frac{1}{2}}}{\left[\sum_{i=1}^{N} (Y_i - \bar{Y}^*)^2/m\right]^{\frac{1}{2}}}$$

2.3.4 <u>ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จากวิธีของ Cramer</u>

ตัวประมาณโดยวิธีนี้ Cramer ได้ประยุกต์ใช้ในการหาความสัมพันธ์ของข้อมูล เมื่อปี 1964 ซึ่งสูตรของตัวประมาณ จะใช้ค่าเฉลี่ยของข้อมูลในแต่ละกลุ่ม (grouped mean) แต่ไม่ใช้ความแปรปรวนของข้อมูลในแต่ละกลุ่ม (grouped Variance) กล่าวคือ

$$\hat{\rho}_{Cramer} = \left[\frac{1}{\frac{C^*}{\hat{\rho}_{grouped}} - C^* + 1}\right]^{\frac{1}{2}}$$

เมื่อ $\hat{\rho}_{grouped}$ คือ ตัวประมาณสัมประสิทธิ์สหสัมพันธ์จาก grouped data

$$C^* = \frac{\sum_{i=1}^{n} (\bar{X}_i - \bar{X}^*)^2}{\sum_{i=1}^{N} (X_i - \bar{X}^*)^2} \qquad \frac{nm - 2}{n - 2}$$

โดย N คือ จำนวนข้อมูลทั้งหมด

n คือ จำนวนกลุ่ม

m คือ จำนวนข้อมูลในแต่ละกลุ่ม

$\bar{X}_i$ คือ ค่าเฉลี่ยของข้อมูลในแต่ละกลุ่มของตัวแปร X

$\bar{X}^*$ คือ ค่าเฉลี่ยของ X ทุก ๆ ค่า

2.4 ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (Minimum mean-squared-error estimator)

ในการพิจารณาคุณสมบัติของตัวประมาณ มักจะพบว่าคุณสมบัติที่ดีบางประการมักจะไม่ไปทางเดียวกันเช่น ความไม่เอนเอียงของตัวประมาณ กับความคลาดเคลื่อนต่ำสุดของตัวประมาณ

ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ ความคลาดเคลื่อนกำลังสองเฉลี่ย หรือ MSE ของ $\hat{\theta}$ คือ

$$\begin{aligned} \text{MSE} &= E(\hat{\theta} - \theta)^2 \\ &= V(\hat{\theta}) + (\text{Bias})^2 \\ &= V(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 \end{aligned}$$

ตัวประมาณ $\hat{\theta}$ เป็นตัวประมาณที่มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด ถ้า $\hat{\theta}$ มีค่า $E(\hat{\theta} - \theta)^2$ ต่ำที่สุด หรือ

$$\text{MSE} \simeq \frac{\Sigma\ (\hat{\theta} - \theta)^2}{n}$$

เรียกตัวประมาณที่มีค่า MSE ต่ำสุด ในบรรดาตัวประมาณทั้งหมดสำหรับทุกค่า ของ θ ว่าเป็น ตัวประมาณที่ดีที่สุด (best estimator)

2.5 ตัวประมาณที่มีประสิทธิภาพ (Efficient estimators)

โดยทั่วไปเราถือว่าตัวประมาณเป็นตัวประมาณที่ดีหากมีความคลาดเคลื่อนต่ำ ความคลาดเคลื่อนของตัวประมาณนี้ดูได้จากความแปรปรวนหรือส่วนเบี่ยงเบนมาตรฐาน จึงอาจใช้ความแปรปรวนของตัวประมาณเป็นเครื่องมือในการพิจารณาประสิทธิภาพของตัวประมาณได้ เช่น ให้

$\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็น ตัวประมาณที่ไม่เอนเอียง (Unbiased estimator) ของ θ โดยที่ $\hat{\theta}_2$ เป็นตัวประมาณ ที่มีความแปรปรวนต่ำสุด ของ θ การเปรียบเทียบประสิทธิภาพของตัวประมาณ $\hat{\theta}_1$ เทียบกับ $\hat{\theta}_2$ คือ

$$\frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = RE(\hat{\theta}_1)$$

$$= re(\hat{\theta}_1|\hat{\theta}_2)$$

เรียกสัดส่วนของการเปรียบเทียบนี้ว่า ประสิทธิภาพสัมพันธ์ (Relative efficiency) ของตัวประมาณ $\hat{\theta}_1$ เทียบกับตัวประมาณ $\hat{\theta}_2$ ของพารามิเตอร์ θ เดียวกัน

ถ้าสัดส่วนนี้มีค่าน้อยกว่า 1 แสดงว่า $\hat{\theta}_2$ ดีกว่า $\hat{\theta}_1$ และถ้าสัดส่วนมีค่าเท่ากับ 1 แสดงว่า $\hat{\theta}_1$ เป็นตัวประมาณที่มีประสิทธิภาพของ θ

2.6 การแจกแจงแบบปกติเมื่อใกล้อนันต์ (Asymptotic Normality)

การแจกแจงแบบปกติเมื่อใกล้อนันต์ เป็นคุณสมบัติหนึ่งที่ได้จากทฤษฎีการใช้ตัวอย่างขนาดใหญ่ (large Sample theory) โดยที่ต้องทราบรูปแบบของการแจกแจงของตัวสถิติที่ต้องการทดสอบเสียก่อน และเมื่อใช้ตัวอย่างขนาดใหญ่ขึ้นแล้ว การแจกแจงของตัวสถิตินั้นจะเข้าสู่ลักษณะของการแจกแจงแบบปกติ

สมมติว่า T_n เป็น Asymptotic Normal ดังนั้นการแจกแจงของ

$$T_n \xrightarrow{n \text{ Large}} N\left(\mu_n(\theta)\ ,\ \sigma_n^2(\theta)\right)$$

T_n จะมีลักษณะการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเป็น $\mu_n(\theta)$ และความแปรปรวนเป็น $\sigma_n^2(\theta)$ เมื่อตัวอย่าง (n) มีขนาดใหญ่

ปรับเป็นการแจกแจงแบบปกติมาตรฐาน (Standard Normal) ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน เท่ากับ 1

มีลำดับ (sequence) $\mu_n(\theta)$, $\sigma_n^2(\theta)$ ที่ทำให้

$$\Phi\left[\frac{t-\mu_n(\theta)}{\sigma_n(\theta)}\right] \text{ เป็นค่าประมาณที่ใกล้เคียงของ } P(T_n \leqslant t)$$

$$z \sim N(0,1)$$

$$\mu_n(\theta) = E\left[T_n(\underset{\sim}{X})\right]$$

$$\sigma_n^2(\theta) = V\left[T_n(\underset{\sim}{X})\right]$$

$$\lim_{n \to \infty} P\left[\frac{T_n(\underset{\sim}{X}) - \mu_n(\theta)}{\sigma_n(\theta)} \leq z\right] = \Phi(z)$$

เมื่อ $\mu_n(\theta)$ เป็น Asymptotic mean ของ T_n

$\sigma_n^2(\theta)$ เป็น Asymptotic Variance ของ T_n

พิจารณา T_n ซึ่งเป็นตัวประมาณของ $q(\theta)$ จะได้ว่า

1) asymptotic Variance ของ T_n ที่มี order $\frac{1}{n}$ เมื่อ $n \to \infty$

$$n\,\sigma_n^2(\theta) \to \sigma^2(\theta) > 0$$

2) T_n เป็น asymptotic Unbiased ของ $q(\theta)$

เมื่อ $n \to \infty$, $\sqrt{n}\left[\mu_n(\theta) - q(\theta)\right] \to 0$

ถ้าคุณสมบัติ 2 ข้อนี้เป็นจริงอาจแทน

$\mu_n(\theta)$ ด้วย $q(\theta)$ และแทน

$\sigma_n^2(\theta)$ ด้วย $\frac{\sigma^2(\theta)}{n}$

บางครั้งตัวประมาณเป็นตัวประมาณที่เอนเอียง (biased estimator) แต่จะเป็นตัวประมาณที่ไม่เอนเอียงเมื่อใกล้อนันต์ (Asymptotic Unbiased estimator) ในกรณีที่ n มีขนาดใหญ่ขึ้นเข้าสู่อนันต์ ($n \to \infty$) จะได้ว่า

$$\frac{R(\theta, T_n)}{V(T_n)} = \frac{MSE(T_n)}{V(T_n)}$$

$$= \left[1 + \frac{(\mu_n(\theta) - q(\theta))^2}{\sigma_n^2(\theta)}\right] \rightarrow 1$$

โดย $V(T_n) = \sigma_n^2(\theta)$ คือ ความแปรปรวนของ T_n

$R(\theta, T_n) = MSE(T_n)$ คือ ค่าคลาดเคลื่อนกำลังสองของ T_n

$MSE(T_n) = V(T_n) + (\text{bias})^2$

$\text{bias} = \mu_n(\theta) - q(\theta)$

$\mu_n(\theta)$ คือ ค่าเฉลี่ยของ T_n

$q(\theta)$ คือ ค่าพารามิเตอร์

2.7 ตัวประมาณที่มีประสิทธิภาพเมื่อใกล้อนันต์ (Asymptotic efficient estimator)

บางครั้งตัวประมาณไม่เป็นตัวประมาณไม่เอนเอียง (Unbiased estimator) แต่จะเป็น Asymptotic Unbiased estimator เมื่อ n เข้าสู่อนันต์

$$T^{(1)} = \{T_n^{(1)}\}$$

$$T^{(2)} = \{T_n^{(2)}\}$$

โดย $T^{(1)}$, $T^{(2)}$ เป็นลำดับของตัวประมาณที่มี asymptotic Variance คือ

$\sigma^2_{n_1}(\theta)$ และ $\sigma^2_{n_2}(\theta)$ ตามลำดับ

และทำให้ $n\,\sigma^2_{n_i}(\theta) \rightarrow \sigma^2_i(\theta)$ เมื่อ $n \rightarrow \infty$ for $i = 1,2$

และเป็น Asymptotic Unbiased estimator ด้วย แล้วจะนิยามประสิทธิภาพของ $T^{(1)}$ เทียบกับ $T^{(2)}$ ได้ดังนี้

$$e\,(\theta,\ T^{(1)}\ ,\ T^{(2)}) \quad = \quad \frac{\sigma^2_2\ (\theta)}{\sigma^2_1\ (\theta)} \quad = \quad \frac{\mathrm{Var}(T_2)}{\mathrm{Var}(T_1)}$$

การเปรียบเทียบประสิทธิภาพของตัวประมาณ โดยวิธีดังกล่าวมานี้ใช้สำหรับ กรณีที่ใช้ขนาดตัวอย่างเท่า ๆ กัน ของตัวประมาณทั้ง 2 ตัว แต่ถ้าใช้ขนาดตัวอย่างไม่เท่ากัน เป็น $T_n^{(1)}$, $T_m^{(2)}$ ประสิทธิภาพของ $T_n^{(1)}$ เทียบกับ $T_m^{(2)}$ จะเป็นลำดับที่ขึ้นอยู่กับ n และ m เนื่องจากความแปรปรวนของตัวประมาณขึ้นกับขนาดตัวอย่าง ซึ่งเป็นสัดส่วนผกผันกัน ดังนั้น เมื่อพิจารณาสัดส่วนของขนาดตัวอย่าง m/n จึงทำให้ Asymptotic Variance ทั้ง 2 ตัว เท่ากันโดยประมาณ และสามารถหาค่าประสิทธิภาพของ $T_n^{(1)}$ เทียบกับ $T_m^{(2)}$ ได้

2.8 อัตราส่วนของความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean squared error ratio)

การเปรียบเทียบประสิทธิภาพของตัวประมาณในท่ามกลางตัวประมาณที่ไม่เอนเอียงนั้น จะใช้ค่าความแปรปรวนมาพิจารณา แต่ถ้าเป็นตัวประมาณที่เอนเอียง และจะเป็นตัวประมาณที่ไม่เอนเอียงได้ ก็ต่อเมื่อใช้ขนาดตัวอย่างใหญ่พอ (n → เข้าสู่อนันต์) ค่าของความคลาดเคลื่อนกำลังสองจะมีค่าเท่ากับค่าความแปรปรวน โดยค่าความเอนเอียงเท่ากับศูนย์

จาก $\quad \mathrm{MSE} \quad = \quad \mathrm{Var} + (\mathrm{bias})^2$

$$\text{bias} = E(\hat{\theta}) - \theta$$

แต่การที่จะตอบว่า ควรใช้ n ขนาดใหญ่แค่ไหนนั้น จึงล่มควรประมาณด้วยการแจกแจงแบบปกติได้ ไม่มีหลักเกณฑ์ที่แน่นอน ดังนั้น การเปรียบเทียบประสิทธิภาพของตัวประมาณ จึงพิจารณาในรูปของอัตราส่วนของความคลาดเคลื่อนกำลังสอง เฉลี่ย

ให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ เป็นตัวประมาณของ θ

ต้องการหาประสิทธิภาพของ $\hat{\theta}_1$ เปรียบเทียบกับ $\hat{\theta}_2$ จะแสดงเป็นอัตราส่วนของ

$$\frac{MSE_{\hat{\theta}_2}}{MSE_{\hat{\theta}_1}}$$

ถ้า $MSE_{\hat{\theta}_2} < MSE_{\hat{\theta}_1}$ จะได้ค่าอัตราส่วน < 1 แสดงว่า $\hat{\theta}_2$ มีประสิทธิภาพดีกว่า $\hat{\theta}_1$

ถ้า $MSE_{\hat{\theta}_2} > MSE_{\hat{\theta}_1}$ จะได้ค่าอัตราส่วน > 1 แสดงว่า $\hat{\theta}_1$ ดีกว่า $\hat{\theta}_2$

ถ้า $MSE_{\hat{\theta}_2} = MSE_{\hat{\theta}_1}$ ได้ค่าอัตราส่วน $= 1$ แสดงว่า $\hat{\theta}_1$ เป็นตัวประมาณที่มีประสิทธิภาพเทียบเท่ากับ $\hat{\theta}_2$