บทที่ 2

ระเบียบวิธีที่ใช้ในการวิจัย

2.1 ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีที่ใช้ในการวิจัยครั้งนี้คือ ทฤษฎีการสุ่มตัวอย่าง (Sampling Theory) ซึ่งจะเกี่ยวข้องกับการกำหนดวิธีการสุ่มเลือกหน่วยตัวอย่าง (Sampling Procedure) การกำหนดขนาดของตัวอย่าง (Sample Size Determination) และกำหนดวิธีการประมาณค่า (Estimation Method) ให้ได้ค่าประมาณที่มีความแม่นยำมากที่สุด สำหรับการศึกษาครั้งนี้จะเน้นเฉพาะการสุ่มแบบให้มีค่าน่าจะเป็นไม่เท่ากัน (Unequal Probability Sampling, UPS) และเป็นการสุ่มแบบใส่คืน (With Replacement) เท่านั้น เพราะสูตรการคำนวณค่าประมาณและค่าความแปรปรวนสำหรับการสุ่มแบบ UPS และไม่ใส่คืนนั้นค่อนข้างจะยุ่งยากมากไม่เหมาะกับภาคปฏิบัติ

กำหนดให้

N = จำนวนหน่วยสุ่ม (Sampling Units) ทั้งหมดของประชากร หรือขนาดของประชากร (Population Size)

n = จำนวนหน่วยสุ่มที่เป็นตัวอย่างหรือขนาดของตัวอย่าง

y_i = ค่าของลักษณะที่ศึกษาของหน่วยที่ "i"

Y = $\sum_{i=1}^{N} y_i$ คือ ค่ายอดรวมทั้งหมดของประชากร

$\hat{Y}$ = ตัวประมาณค่า (Estimator) ของ Y

P_i = ค่าความน่าจะเป็น (Probability) ในการเลือกหน่วยที่ i ซึ่ง P_i มีคุณสมบัติ ดังนี้

$$0 < P_i < 1$$

$$\sum_{i=1}^{N} P_i = 1$$

ในการสุ่มโดยมีค่าความน่าจะเป็นไม่เท่ากัน และสุ่มแบบใส่คืน ตัวประมาณค่าของ Y แยกเป็นกรณีได้ ดังนี้

สำหรับ $n = 1$ จะได้

$$\hat{Y}_i = \frac{y_i}{P_i} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2.1)$$

โดยที่

$\hat{Y}_i$ = ตัวประมาณค่าที่ได้จากหน่วยที่ "i"

ค่าคาดหวัง (Expected Value) ของ $\hat{Y}_i$ จะเท่ากับ

$$E(\hat{Y}_i) = E\left(\frac{y_i}{P_i}\right)$$

$$= \sum_{i=1}^{N} P_i\left(\frac{y_i}{P_i}\right)$$

$$= \sum_{i=1}^{N} y_i = Y$$

แสดงว่า $\hat{Y}_i$ เป็นตัวประมาณค่าที่ไม่เอนเอียง (Unbiased Estimator) ของ Y

ค่าความแปรปรวนของ $\hat{Y}_i$ คำนวณโดยใช้สูตร

$$V(\hat{Y}_i) = E(\hat{Y}_i - Y)^2$$

$$= \sum_{i=1}^{N} P_i(\hat{Y}_i - Y)^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2.2)$$

สำหรับ $n > 1$ จะได้ตัวประมาณค่า คือ

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2.3)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{P_i} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2.4)$$

ค่าคาดหวังของ $\hat{Y}$ จะเท่ากับ

$$E(\hat{Y}) = E(\frac{1}{n} \sum_{i=1}^{n} \hat{Y}_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(\hat{Y}_i)$$

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

$$= \frac{n}{n} Y = Y$$

แสดงว่า $\hat{Y}$ ซึ่งเป็นค่าเฉลี่ยของ $\hat{Y}_i$ เป็นตัวประมาณค่าที่ไม่เอนเอียงของ Y

ค่าความแปรปรวนของ $\hat{Y}$ เท่ากับ

$$V(\hat{Y}) = V(\frac{1}{n} \sum_{i=1}^{n} \hat{Y}_i)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} V(\hat{Y}_i)$$

$$= \frac{n}{n^2} \sum_{i=1}^{N} P_i(\hat{Y}_i - Y)^2$$

$$= \frac{1}{n} \sum_{i=1}^{N} P_i(\hat{Y}_i - Y)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.5)$$

$$= \frac{1}{n} V(\hat{Y}_i)$$

ตัวประมาณค่าความแปรปรวนของ $V(\hat{Y})$ คือ

$$\hat{V}(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{Y}_i - \hat{Y})^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.6)$$

$$= \frac{1}{n(n-1)} (\sum_{i=1}^{n} \hat{Y}_i^2 - n\hat{Y}^2) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.7)$$

และ $E(\hat{V}(\hat{Y})) = \frac{1}{n(n-1)} E(\sum_{i=1}^{n} \hat{Y}_i^2 - n\hat{Y}^2)$

$$= \frac{1}{n(n-1)} (\sum_{i=1}^{n} E(\hat{Y}_i^2) - nE(\hat{Y}^2))$$

$$= \frac{1}{n(n-1)} (nV(\hat{Y}_i) + nY^2 - nV(\hat{Y}) - nY^2)$$

$$= \frac{1}{n(n-1)} (n^2V(\hat{Y}) - nV(\hat{Y}))$$

$$= \frac{n(n-1)}{n(n-1)} V(\hat{Y}) = V(\hat{Y})$$

ซึ่งแสดงว่า $\hat{V}(\hat{Y})$ ในสมการที่ (2.6) เป็นตัวประมาณค่าที่ไม่เอนเอียงของ $V(Y)$

จาก (2.3) ถ้าหากเป็นการสุ่มโดยให้มีความน่าจะเป็นเท่ากันทุกหน่วยหรือให้

$P_i = \frac{1}{N}$ สำหรับทุกหน่วยของประชากร จะได้

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} y_i \qquad \text{.......................(2.8)}$$

เมื่อแทนค่า $P_i = \frac{1}{N}$ ในสมการที่ (2.5) จะได้

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^{N} \frac{1}{N}(Ny_i - Y)^2$$

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

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

$$= N^2 \frac{\sigma^2}{n} \qquad \text{.....................(2.9)}$$

โดยที่

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

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y})^2$$

ซึ่งจะเท่ากับการสุ่มอย่างง่ายและใส่คืน (Simple Random Sampling With Replacement) นั่นเอง

ข้อได้เปรียบเสียเปรียบของการสุ่มแบบ UPS นั้น ในทางทฤษฎีจะขึ้นกับค่า "P_i" ที่ใช้ในการเลือกตัวอย่าง กล่าวคือ ค่าความแปรปรวนของตัวประมาณค่าจากการสุ่มแบบ UPS จะมีค่าน้อยหรือมีความแม่นยำสูง ถ้าหากว่าค่าของ P_i เป็นปฏิภาคกับค่า y_i โดยเฉพาะอย่างยิ่งถ้า $P_i = \frac{y_i}{Y}$ แล้ว $V(\hat{Y})$ จะมีค่าเท่ากับ "o" ทั้งนี้จะเห็นได้โดยการแทนค่า $P_i = \frac{y_i}{Y}$ ใน (2.5) ซึ่งจะได้

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^{N} \frac{y_i}{Y} (Y - Y)^2 = 0$$

อย่างไรก็ตาม ปัญหาในทางปฏิบัติก็คือ เรามักจะไม่ทราบค่าของ y_i หรือ Y ก่อนการสำรวจเพราะเป็นสิ่งที่เราต้องการประมาณค่า ดังนั้น จำเป็นจะต้องเลือกตัวแปร (Variable) อื่นๆ คือ "x_i" ที่คาดว่าจะมีความสัมพันธ์ในทางบวกกับ "y_i" มากที่สุด (Highly Positive Correlation) ซึ่งจะได้

$$P_i = \frac{x_i}{X}$$

โดยที่ x_i = ตัวแปรที่ใช้วัดขนาด (Measure of Size) ของหน่วยที่ i

$$X = \sum_{i=1}^{N} x_i$$

และการสุ่มแบบนี้เรียกว่าการสุ่มแบบให้มีความน่าจะเป็นในการเลือกเป็นสัดส่วนกับขนาดของหน่วยสุ่ม (Probability Proportional to Size Sampling, PPS)

2.2 <u>การประยุกต์</u>

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบวิธีการสุ่มเลือกกิ่งตัวอย่างโดยวิธีการต่างๆ กัน เพื่อหาวิธีที่เหมาะสมในการประมาณผลผลิต (จำนวนผล) ต่อต้นของส้มเขียวหวาน

กำหนดให้

N = จำนวนกิ่งปลาย (Terminal Branch) ทั้งหมดของต้นส้มเขียวหวาน

n = จำนวนกิ่งปลายที่ใช้เป็นกิ่งตัวอย่าง

y_i = จำนวนผลส้มของกิ่งปลาย

Y = จำนวนผลส้มทั้งหมดของต้น

$$= \sum_{i=1}^{N} y_i$$

P_i = ค่าความน่าจะเป็นในการเลือกกิ่งปลายที่ "i"

$\hat{Y}_i = \dfrac{y_i}{P_i}$ คือ ค่าประมาณจำนวนผลส้มทั้งต้นที่ได้จากกิ่งที่ "i"

$\hat{Y} = \dfrac{1}{n}\sum_{i=1}^{n}\hat{Y}_i$ คือ ค่าประมาณผลส้มทั้งหมดที่ประมาณได้เฉลี่ยจากกิ่งตัวอย่าง "n" กิ่ง

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{P_i} \quad \text{..................(2.10)}$$

สำหรับวิธีการสุ่มเลือกที่จะศึกษาในครั้งนี้มีอยู่ 4 วิธีด้วยกัน ซึ่งจะทำให้ได้ค่า P_i แตกต่างกันไปในแต่ละวิธี ดังนี้

<u>วิธีที่</u> 1 สุ่มกิ่งปลายโดยตรงและมีความน่าจะเป็นในการเลือกกิ่งเท่ากันทุกกิ่ง (Direct Selection - Equal Probability With Replacement, DS-EP) ซึ่งจะได้

$$P_i = \frac{1}{N} \qquad \text{..................(2.11)}$$

วิธีที่ 2 สุ่มกิ่งปลายโดยตรงและมีค่าความน่าจะเป็นในการเลือกกิ่งเป็นสัดส่วนกับขนาด (พื้นที่หน้าตัด) ของกิ่ง (Direct Selection - Probability Proportional to Cross Section Area With Replacement, DS-PPS) ซึ่งจะได้

$$P_i = \frac{x_i}{X} \qquad \text{..................(2.12)}$$

โดยที่ x_i = พื้นที่หน้าตัดของกิ่งปลายที่ "i"

$$X = \sum_{i=1}^{N} x_i$$

วิธีที่ 3 สุ่มหลายขั้นตอนจากจุดที่แตกแขนงตั้งแต่โคนต้นไปจนถึงกิ่งปลายและมีค่าความน่าจะเป็นในการเลือกกิ่งแต่ละขั้นตอนเท่ากันทุกกิ่ง (Random Path - Equal Probability With Replacement, RP - EP) ซึ่งจะได้

$$P_i = P_i^1\, P_i^2\, P_i^3 \text{ ....} P_i^k \qquad \text{..............(2.13)}$$

โดยที่

P_i^j = ค่าความน่าจะเป็นในการเลือกกิ่งปลายที่ i ในขั้นตอนการแตกกิ่งที่ "j"

$= \frac{1}{N^j}$

N^j = จำนวนกิ่งทั้งหมดที่แตกแขนง ในขั้นที่ "j"

k = จำนวนขั้นตอนที่เลือก (ที่แตกแขนง) จนกระทั่งถึงกิ่งปลายที่ "i" (k อาจไม่เท่ากันทุกค่าของ i)

วิธีที่ 4 สุ่มหลายขั้นตอนจากจุดที่แตกแขนงตั้งแต่โคนต้นไปจนถึงกิ่งปลายและมีค่าความน่าจะเป็นในการเลือกกิ่งแต่ละขั้นตอนเป็นสัดส่วนกับขนาดของกิ่ง (Random Path - Probability Proportional to Cross Section Area With Replacement, RP - PPS) ซึ่งจะได้

$$P_i = P_i^1 \; P_i^2 \; P_i^3 \; \ldots\ldots. P_i^k \qquad \ldots\ldots\ldots\ldots(2.14)$$

โดยที่

$$P_i^j = \frac{x_h^j}{X^j}$$

x_h^j = พื้นที่หน้าตัดของกิ่งที่ "h" ในขั้นตอนการแตกกิ่ง "j"

$$X^j = \sum_{h=1}^{N_j} x_h^j$$

ในการเปรียบเทียบความแม่นยำของตัวประมาณค่าที่ได้จากการสุ่มตัวอย่าง 2 วิธี ซึ่งมีขนาดของตัวอย่างเท่ากันจะใช้ค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency) เป็นตัววัด ซึ่งจะเท่ากับ

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

โดยที่

$V(\hat{Y}_1)$ = ค่าความแปรปรวนจากการสุ่มโดยวิธีที่ 1

$V(\hat{Y}_2)$ = ค่าความแปรปรวนจากการสุ่มโดยวิธีที่ 2

ส่วนการศึกษาหาขนาดของตัวอย่าง (จำนวนกิ่งตัวอย่าง) ที่เหมาะสมจะพิจารณาจากขนาดของตัวอย่างที่ให้ค่าสัมประสิทธิ์ของความแปรปรวน (Coefficient of Variation, C.V.) มีค่าไม่เกินที่กำหนด ซึ่งมีสูตรคำนวณ ดังนี้

$$\%\ CV(\hat{Y}) = \frac{SE(\hat{Y})}{Y} \times 100$$

โดยที่

$$SE(\hat{Y}) = \sqrt{V(\hat{Y})}$$

= ค่าความคลาดเคลื่อนมาตรฐาน (Standard Error) ของ $\hat{Y}$

ในการศึกษาครั้งนี้ แม้ว่าจะมีวัตถุประสงค์หลักเพื่อเปรียบเทียบเทคนิคการสุ่มตัวอย่างทั้ง 4 วิธีก็ตาม แต่เพื่อที่จะให้ทราบค่าต่างๆ ที่แท้จริงของประชากร (Population Values) เช่น จำนวนผลทั้งหมดในแต่ละต้น ขนาดของกิ่งปลายแต่ละกิ่ง รวมทั้งค่าความแปรปรวนของจำนวนผลบนกิ่งปลายแต่ละกิ่ง จึงจำเป็นจะต้องทำการวัดขนาดของทุกกิ่งทุกแขนงตั้งแต่โคนต้นจนถึงกิ่งปลาย และนับผลบนกิ่งปลายทุกกิ่งด้วย