# บทที่ 3

# ทฤษฎีเกี่ยวกับการประมาณ
# (Estimation Theory)

## 3.1 คำจำกัดความของการประมาณค่า (Definition of Estimation)

องค์ประกอบที่สำคัญยิ่งของทฤษฎีการสุ่มตัวอย่างนั้น ประกอบด้วย 2 ส่วนใหญ่ ๆ คือ

ก) ระเบียบแบบแผนการสุ่มหน่วยตัวอย่าง ( Sample Design)

ข) วิธีการประมาณผล ( Estimation Procedure )

จากองค์ประกอบทั้ง 2 ส่วนดังกล่าวหากขาดตกบกพร่องส่วนหนึ่งส่วนใดแล้ว จะทำให้การนำเอาวิธีการสำรวจโดยการสุ่มหน่วยตัวอย่างมาใช้ไม่ได้ผลตามจุดมุ่งหมายที่ต้องการ กล่าวคือ ผลที่ประมวลได้จะคลาดเคลื่อน หรือ อาจใช้การไม่ได้เลยก็เป็นได้ เช่น เราจัดวางระเบียบแบบแผนการสุ่มหน่วยตัวอย่างไว้วิธีการหนึ่งอย่างดี และถูกต้องตามหลักวิชาการทุกประการ คือ สมมุติใช้วิธีเลือกหน่วยตัวอย่างใหม่บางส่วนและใช้ซ้ำบางส่วนในการสำรวจที่ทำมากกว่า 1 ครั้งขึ้นไป สำหรับข้อมูลเกี่ยวกับค่าใช้จ่ายหรือสินค้าประเภทเครื่องนุ่งห่มโดยเฉพาะของสตรี ซึ่งเป็นข้อมูลที่มีลักษณะจะเปลี่ยนแปลงเมื่อถูกกระทบกระเทือนโดยเวลา **เนื่องจากเครื่องนุ่งห่มของสตรีจะเปลี่ยนไปตามสมัยนิยม** ต่อมาเมื่อถึงขั้นการใช้วิธีการประมาณผล แทนที่เราจะใช้วิธีการประมาณคำนวนให้สอดคล้องตามแบบ P.R. แต่กลับคิดคำนวนตามแบบเสมือนการใช้หน่วยตัวอย่างชุดเดียวกันทั้ง 2 รอบ ซึ่งผลการประมาณที่ได้ก็จะไม่ได้ประโยชน์ดีขึ้นเลยจากการที่นำเอาแบบแผนดังกล่าวมาใช้ คือ ผลประมาณที่ได้ก็จะเท่ากับวิธีการใช้หน่วยตัวอย่างชุดเดียวกันโดยไม่เกิดผลแตกต่างอะไรขึ้น หากแต่จะได้ประโยชน์เพียงทางด้านปฏิบัติเท่านั้น โดยการไปลดความไม่ให้ความร่วมมือในการตอบข้อสอบถาม ในทางตรงข้าม/มีหากใช้วิธีการประมาณผลถูกต้องแต่แบบแผนที่วางไว้อีกอย่างหนึ่งก็ไม่สามารถจะกล่าวได้ว่าให้ผลอย่างมีประสิทธิภาพเช่นกัน ดังนั้นจึงกล่าวได้ว่าทั้ง2 ส่วนมีความสัมพันธ์อย่างใกล้ชิด และจะขาดเสียอย่างหนึ่งอย่างใดไม่ได้

<u>การประมาณค่า</u> (Estimation) คือ ขบวนการคำนวนตามสูตรของแต่ละทฤษฎี เพื่อคาดคะเนผลโดยอาศัยจากข้อมูลที่เก็บรวบรวมได้จากหน่วยตัวอย่างที่สุ่มขึ้นมา เป็นตัวแทนของประชากรทั้งหมด มาเป็นมาตรฐานในการคำนวน มิใช่เป็นการคาดคะเนจากประสพการณ์ หรือเดาเอา ตัวอย่างเช่น การประมาณค่าของ Sample Mean และค่าที่ประมาณได้สามารถจะให้ค่าเป็น Unbiased Estimate[1] และจะให้ผลประสิทธิภาพเพียงไรนั้น จะขึ้นอยู่กับแต่ละแบบแผนวิธีการที่ใช้ ซึ่งการวัดความคลาดเคลื่อนเป็นเครื่องที่จะเปรียบเทียบประสิทธิภาพกันได้

3.2 <u>ประเภทของการประมาณ</u> (Method of Estimation)

วิธีการประมาณในการสำรวจจากหน่วยตัวอย่างที่มีหน่วยตัวอย่างซ้ำ เพียงบางส่วนนั้นสามารถนำมาใช้ในการประมาณเกี่ยวกับ Population Characteristic ในรูปต่าง ๆ ได้ดังนี้ :-

ก) <u>Mean</u> ซึ่งอาจจะหา Mean ของแต่ละรอบของการสำรวจ หรือ ผลเฉลี่ยของทุกรอบการสำรวจ (Mean of Over all Occasions) โดยทั่วไปจะอยู่ในรูปของ

$$\bar{X} = \sum_{i=1}^{n} X_i / P_i nN$$ ในกรณีที่เป็น Unequal Prob. Sampling

ซึ่ง $P_i$ = Arbitrary Probability ที่จะกำหนด $0 < P_i < 1$

ถ้าหาก $P_i = \frac{1}{N}$ นั่นคือการเลือกตัวอย่างที่ทุกหน่วยตัวอย่างมีโอกาศถูกเลือกขึ้นมาเท่า ๆ กัน และจะอยู่ในรูปของ

$$\bar{X} = \sum_{i=1}^{n} X_i / n$$ ในกรณีที่เป็น Simple Random Sampling

ข) <u>Total</u> อาจจะหาได้โดยอาศัยจาก Mean ของแต่ละรอบ หรือ Mean of Over all Occ. แล้วแต่สิ่งที่ต้องการจะหา ซึ่งเขียนได้ในรูปของ

$$\hat{T} = N\bar{X}$$

ในเมื่อ $$\bar{X} = \sum_{i=1}^{n} X_i / n$$

นอกจากนั้นเรายังอาจคำนวนหาค่าประมาณในรูปของ Ratio,Propertion,percentile.

[1]Ibit P.2

ในที่นี้สิ่งที่เราจะกล่าวต่อไปก็เฉพาะแต่ Mean เพียงอย่างเดียว โดยจะเปรียบเทียบถึงประสิทธิภาพของแต่ละแบบแผน ส่วนค่าประมาณในด้าน Total นั้นก็คงมีวิธีการคำนวนแบบเดียวกับ Mean ดังที่ได้แสดงในข้อ (ข) คือ $\hat{T} = N\bar{X}$ และเพื่อความสะดวกในการเปรียบเทียบค่าMean ที่เป็น Unbiasedness และมี Variance ที่มีประสิทธิภาพกว่าแบบแผน No.R.U. หรือ C.R.

## 3.3 <u>พื้นความรู้ทั่วไปเกี่ยวกับการใช้จำนวนหน่วยตัวอย่างที่เลือกขึ้นใหม่</u>

เกี่ยวกับการสำรวจที่ทำมากกว่า 1 ครั้งขึ้นไป โดยใช้หน่วยตัวอย่างซ้ำกันบางส่วน และเลือกหน่วยตัวอย่างขึ้นใหม่บางส่วนนั้น อัตราส่วนของหน่วยตัวอย่างที่ซ้ำนั้นจะถูกกำหนดโดย $(1-\mu)$ คือ ค่าของ $\mu$ จะเป็น :

$0 < \mu < 1$ หมายความเป็นแบบแผนที่ใช้หน่วยตัวอย่างซ้ำบางส่วนและเลือกใหม่บางส่วน ซึ่งค่า $\mu$ จะเป็นเศษส่วน โดยทางปฏิบัติมักใช้เศษส่วนง่าย ๆ เช่น $\frac{1}{3}, \frac{1}{2}, \frac{2}{3}$ เป็นต้น จะเป็นอัตราส่วนเท่าใดนั้น ย่อมแล้วแต่ความเหมาะสม

ถ้า $\mu = 0$ หมายความเป็นแบบแผนที่ใช้หน่วยตัวอย่างชุดเดียวกันสำหรับทุกรอบ

ถ้า $\mu = 1$ หมายความเป็นแบบแผนที่ใช้หน่วยตัวอย่างใหม่ทุกครั้งสำหรับทุกรอบ

ดังนั้นอัตราส่วนของหน่วยตัวอย่างที่จะเลือกขึ้นมาใหม่ $= \mu$

<u>ตัวอย่างที่ 1</u> กรณีการสำรวจ 2 รอบติดต่อกัน



ให้ Universe : $(X_1, X_2, X_3, \ldots\ldots\ldots X_N)$

ในรอบที่ 1 หน่วยตัวอย่างที่ถูกเลือกขึ้นมาทั้งหมด : ( n = 8 units )

ในรอบที่ 2 หน่วยตัวอย่างที่ถูกเลือกขึ้นมาทั้งหมด : ( n = 8 units )

$X_i$ = ลักษณะของข้อมูลที่สนใจศึกษาที่ลำดับ i

$\mu$ = อัตราการเลือกตัวอย่างใหม่ $= \frac{1}{2}$

| | 4 หน่วย | 4 หน่วย | | |
|---|---|---|---|---|
| | $X_1\ X_2\ X_3\ X_4$ | $X_5 X_6\ X_7\ X_8$ | | รอบที่ 1 |
| | | $X_5\ X_6\ X_7\ X_8$ | $X_9\ X_{10} X_{11} X_{12}$ | รอบที่ 2 |
| | | 4 หน่วย | 4 หน่วย | |

จำนวนหน่วยตัวอย่างที่ใช้ทั้งหมดในที่นี้ $= n + (p-1)\mu n$

$n$ = จำนวนหน่วยตัวอย่างที่ใช้ในรอบหนึ่ง ๆ = 8 หน่วย

$p$ = จำนวนรอบของการสำรวจ = 2 รอบ

$\mu$ = **อัตราการเลือกหน่วยตัวอย่างใหม่** = $\frac{1}{2}$

ดังนั้น**หน่วยตัวอย่างทั้งหมด**ที่ใช้ $= 8+(2-1)\,\frac{1}{2}\cdot 8 =$ 12 หน่วย

ซึ่งมิใช่จำนวน $= 8 \cdot 2 =$ 16 "

และจำนวนหน่วยตัวอย่างที่เลือกขึ้นใหม่ $= \mu n = \frac{1}{2}\cdot 8 =$ 4 "

" ที่ซ้ำกัน $= n(1-\mu) = (1-\frac{1}{2})\cdot 8 = 4$ หน่วย

<u>ตัวอย่างที่ 2</u> กรณีการทำสำรวจ 4 รอบติดต่อกัน และมีแบบแผน ดังในรูป

| | | | | | | |
|---|---|---|---|---|---|---|
| 30 | 30 | 30 | 30 | 30 | 30 | |
| $X_1$ | $X_2$ | $X_3$ | | | | รอบที่ 1 |
| | $X_2$ | $X_3$ | $X_4$ | | | รอบที่ 2 |
| | | $X_3$ | $X_4$ | $X_5$ | | รอบที่ 3 |
| | | | $X_4$ | $X_5$ | $X_6$ | รอบที่ 4 |

$X_i$ = กลุ่มของหน่วยตัวอย่างกลุ่มละ 30 หน่วย ซึ่ง $i = 1,2,3,4,5,6$

$n_t$ = จำนวนหน่วยตัวอย่างที่ใช้ในแต่ละรอบการสำรวจ = 90 หน่วย

$t$ = 1,2,3,4

$\mu$ = **อัตราการเลือกหน่วยตัวอย่างใหม่** = $\frac{1}{3}$

$p$ = จำนวนรอบของการสำรวจ = 4

ดังนั้นจำนวนหน่วยตัวอย่างที่ใช้ทั้งหมด $= n+(p-1)\mu n = 90+3\cdot\frac{1}{3}\cdot 90 = 180$ หน่วย

ซึ่งมิใช่ $= 90 \cdot 4 = 360$ "

จำนวนหน่วยตัวอย่างที่เลือกขึ้นใหม่ในแต่ละรอบ $= \mu n = \frac{1}{3}\cdot 90 = 30$ "

จำนวนหน่วยตัวอย่างที่ซ้ำกันในแต่ละรอบ $= n(1-\mu) = (1-\frac{1}{3})90 = 60$ "

<u>ตัวอย่างที่3</u> กรณีเหมือน ตัวอย่าง 5 แต่มีแบบแผนที่ต่างกันออกไป คือในรอบสุดท้ายของการสำรวจแทนที่จะเลือกหน่วยตัวอย่างขึ้นมาใหม่ กลับใช้หน่วยตัวอย่างซ้ำกับรอบที่1 ในกลุ่มที่ $X_1$

| 30 | 30 | 30 | | | | |
|---|---|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_3$ | 30 | | | รอบที่ 1 |
| | $X_2$ | $X_3$ | $X_4$ | 30 | | " 2 |
| | | $X_3$ | $X_4$ | $X_5$ | | " 3 |
| | | | $X_4$ | $X_5$ | $X_1$ | " 4 |

$X_i$ = กลุ่มของหน่วยตัวอย่างกลุ่มละ 30 หน่วย
ซึ่ง $i = 1,2,3,4,5$

$n_t$ = จำนวน **ตัวอย่าง** ที่ใช้ในแต่ละรอบการสำรวจ = 90 หน่วย
ซึ่ง $t = 1,2,3,4$

$\mathcal{M}$ = **อัตราการเลือกตัวอย่างใหม่** = $\frac{1}{3}$

p = จำนวนรอบของการสำรวจ

ดังนั้น จำนวนหน่วยตัวอย่างที่ใช้ทั้งหมด $= \left\{ n + (p-1)\mathcal{M}n \right\} - 30$

$= \left\{ 90+(4-1)\frac{90}{3} \right\} - 30$

$= 150$ หน่วย

จะเห็นว่าน้อยกว่าแบบแผน ที่ใช้ในตัวอย่าง 2 อยู่ 30 หน่วย นั่นก็คือหนึ่งกลุ่มซึ่งไปใช้ซ้ำในรอบแรก

จากตัวอย่างที่ 1 - 3 สามารถนำมาเขียนเป็นรูปของ <u>Generalized Formula</u> ได้ดังนี้ .

| | | | |
|---|---|---|---|
| $\mathcal{M}n$ | $(1-\mathcal{M})n$ | | รอบที่ 1 |
| | $(1-\mathcal{M})n$ | $\mathcal{M}n$ | " 2 |

$\mathcal{M}n$ = จำนวน **ตัวอย่าง ใหม่**ที่เลือกขึ้นมาทดแทน

$(1-\mathcal{M})n$ = " " ซ้ำกันในการสำรวจทั้ง 2 ติดต่อกัน

และ**ตัวอย่างทั้งหมด** ที่ใช้จะ $= n + (p-1)\mathcal{M}n$ - - - -(3.3.1)

## 3.4 การประมาณค่ามัชฌิมจากรอบสุดท้ายกรณีการสำรวจ 2 รอบติดต่อกัน
(Estimated Mean for Last Occasion)

| $\bar{X}_{1,1}$ | $\bar{X}_{1,2}$ | | |
|---|---|---|---|
| $\mu n$ units | $(1-\mu)n$ units | | รอบที่ 1 |
| | $(1-\mu)n$ units | $\mu n$ units | " 2 |
| | $\bar{X}_{2,1}$ | $\bar{X}_{2,2}$ | |

ให้ $\bar{X}_{1,1}$ = Sample mean ของการสำรวจในรอบที่ 1 สำหรับหน่วยตัวอย่าง $\mu n$ หน่วย
$\bar{X}_{1,2}$ = " " ซ้ำ $(1-\mu)n$ "
$\bar{X}_{2,1}$ = " " 2 " $(1-\mu)n$ "
$\bar{X}_{2,2}$ = " " เลือกใหม่ $\mu n$ "

ในแต่ละรอบของการสำรวจนั้นเราสามารถที่จะประมาณค่าของ Mean ซึ่งเป็น Unbiased Estimate ของ Population Mean ได้ตามทฤษฎีของ Sampling แต่ในที่นี้เราสนใจที่จะได้รับประโยชน์จากการใช้หน่วยตัวอย่างที่ซ้ำกันบางส่วนในรอบที่ 2 มาประมาณค่าของ Population Mean ของ การสำรวจรอบที่ 2 โดยวิธีดังต่อไปนี้

<u>General Linear Combination Method</u>

$$\bar{X}_2 = a\bar{x}_{1,1}+b\bar{x}_{1,2}+c\bar{x}_{2,1}+d\bar{x}_{2,2} \quad \text{------} \quad (1)$$

โดยที่ $\bar{X}_2$ = Estimated Mean for second Occasion

a,b,c และ d = Arbitrary Constants

จาก (1) เราอาจเขียนได้ว่า

$$E(\bar{X}_2) = E\left[a\bar{x}_{1,1}+b\bar{x}_{1,2}+c\bar{x}_{2,1}+d\bar{x}_{2,2}\right]$$
$$= aE(\bar{x}_{1,1})+bE(\bar{x}_{1,2})+cE(\bar{x}_{2,1})+dE(\bar{x}_{2,2}) \quad \text{-----} \quad (2)$$

เนื่องจากค่า

$$\left.\begin{array}{l} E(\bar{x}_{1,1}) = E(\bar{x}_{1,2}) = \theta_1 \\ E(\bar{x}_{2,1}) = E(\bar{x}_{2,2}) = \theta_2 \end{array}\right\}^{2}$$

ค่าของ Population Mean $\theta_1$ และ $\theta_2$ จะไม่เท่ากันเนื่องเปลี่ยนแปลงไปตามการเปลี่ยนแปลงของเวลา

จาก (2)

ดังนั้น $E(\bar{X}_2) = (a+b)\,\theta_1 + (c+d)\,\theta_2 \quad \text{-----} \quad (3)$

และ $E(\bar{X}_2)$ จะเป็น Unbiased Estimate ของ $\theta_2$ ก็ต่อเมื่อ

[2] ดูการพิสูจน์ในภาคผนวกเชิงคณิตศาสตร์หัวข้อที่ 1 หน้า 52

$$a + b = 0 \quad \text{หรือ} \quad b = -a$$

$$c + d = 1 \quad \text{"} \quad d = 1 - c$$



จาก (2) เขียนใหม่จะได้

$$E(\bar{x}_2) = a(\bar{x}_{1,1} - \bar{x}_{1,2}) + c\,\bar{x}_{2,1} + (1-c)\,\bar{x}_{2,2} \qquad \text{-------}(4)$$

ในที่นี้ (4) เป็น Linear function ซึ่งเป็น Unbiased Estimate ของ $\theta_2$ เท่านั้น แต่จะเป็น Best Linear Unbiased Estimator (BLUE) หรือไม่นั้น จะต้องอาศัยการพิสูจน์ต่อไปนี้เป็นส่วนประกอบ

3.5 <u>การประมาณค่า Variance จากรอบสุดท้าย</u> (Estimated Variance for the Last Occasion)

$$V(\bar{x}_2) = V\left[a(\bar{x}_{1,1} - \bar{x}_{1,2}) + c\,\bar{x}_{2,1} + (1-c)\,\bar{x}_{2,2}\right] \qquad \text{-------} (1)$$

การคำนวนหา $V(\bar{x}_2)$ จะต้องอาศัยทฤษฎีเกี่ยวกับการกระจายเทอมของ Variance และ Covariance[3] บางประการ ซึ่งจะได้ความสัมพันธ์หลังจากที่ได้กระจายแล้วดังนี้

$$V(\bar{x}_2) = a^2 V\underbrace{(\bar{x}_{1,1} - \bar{x}_{1,2})}_{A} + c^2 V(\bar{x}_{2,1}) + (1-c)^2 V(\bar{x}_{2,2}) + 2ac\,Cov\{\underbrace{(\bar{x}_{1,1} - \bar{x}_{1,2}), \bar{x}_{2,1}}_{B}\}$$
$$+ 2a(1-c)\,Cov\{\underbrace{(\bar{x}_{1,1} - \bar{x}_{1,2}), \bar{x}_{2,2}}_{C}\} + 2c(1-c)\,Cov\underbrace{(\bar{x}_{2,1}, \bar{x}_{2,2})}_{D} \qquad \text{----}(2)$$

---

3 <u>ทฤษฎีบางประการเกี่ยวกับการกระจายเทอมของ Variance และ Covariance</u>

$$V\left[\sum W_h Y_h\right] = \sum W_h^2 V(Y_h) + 2 \sum_{h<g} W_h W_g \, Cov(Y_h, Y_g)$$
$$V(Y_1 \pm Y_2) = V(Y_1) + V(Y_2) \pm 2\, Cov(Y_1, Y_2)$$

ในกรณีที่ $Y_1$ และ $Y_2$ ไม่มีความสัมพันธ์กันเราจะได้

$$V\left(\sum W_h Y_h\right) = \sum W_h^2 V(Y_h)$$
$$Cov\left[\sum W_h Y_h, \sum V_h X_h\right] = \sum W_h V_h \, Cov(Y_h, X_h)$$

เทอมของ Cov. จะเป็นรูปของ Var. เมื่อ :

$$V(Y_h) = Cov(Y_h, Y_h)$$

Leslie Kish, <u>Survey Sampling</u> (New York : John Wiley & Sons Inc., 1965) P.69

จาก (2) พิจารณาเฉพาะเทอม A[4]

$$V(\bar{x}_{1,1}-\bar{x}_{1,2}) = V(\bar{x}_{1,1})+V(\bar{x}_{1,2})-2\,Cov(\bar{x}_{1,1},\bar{x}_{1,2})$$
$$= V(\bar{x}_{1,1})+V(\bar{x}_{1,2}) \quad \text{------------- (3)}$$

พิจารณาเฉพาะเทอม B[4]

$$Cov[(\bar{x}_{1,1}-\bar{x}_{1,2}),\bar{x}_{2,1}] = Cov(\bar{x}_{1,1},\bar{x}_{2,1}) - Cov(\bar{x}_{1,2},\bar{x}_{2,1})$$
$$= -\frac{\sigma_{1.2}}{(1-\mu)n} \quad \text{------------- (4)}$$

พิจารณาเฉพาะเทอม C

$$Cov[(\bar{x}_{1,1}-\bar{x}_{1,2}),\bar{x}_{2,2}] = Cov(\bar{x}_{1,1},\bar{x}_{2,2}) - Cov(\bar{x}_{1,2},\bar{x}_{2,2})$$
$$= 0 \quad \text{------------- (5)}$$

พิจารณาเฉพาะเทอม D

$$Cov(\bar{x}_{2,1},\bar{x}_{2,2}) = 0 \quad \text{------------- (6)}$$

แทนค่า (3),(4),(5)และ(6) ใน (2) จะได้

$$V(\bar{x}_2) = a^2[V(\bar{x}_{1,1})+V(\bar{x}_{1,2})] + e^2V(\bar{x}_{2,1})+(1-e)^2V(\bar{x}_{2,2}) - \frac{2ae\sigma_{1.2}}{(1-\mu)n} \quad \text{----- (7)}$$

เนื่องจากว่า [4.1]

$$V(\bar{x}_1) = \frac{S_1^2}{n} \quad ; \quad V(\bar{x}_2) = \frac{S_2^2}{n}$$
$$V(\bar{x}_{1,1}) = \frac{S_1^2}{\mu n} \quad ; \quad V(\bar{x}_{1,2}) = \frac{S_1^2}{(1-\mu)n}$$
$$V(\bar{x}_{2,1}) = \frac{S_2^2}{(1-\mu)n} \quad ; \quad V(\bar{x}_{2,2}) = \frac{S_2^2}{\mu n}$$

---

[4] ดูการพิสูจน์ในภาคผนวกเชิงคณิตศาสตร์หน้า 53

[4.1] ในที่นี้เราจะพิจารณาค่าของเทอม Variance โดยไม่คำนึงถึง $\frac{N-n}{N-1}$ ซึ่งเป็นที่รู้จักกันดีในเทอมของ Finite Population Correction ( f.p.c.)เหตุผลที่เราไม่คำนึงถึง f.p.c.ในการคำนวน คือ :

1. เมื่อ N เป็นจำนวนใหญ่มาก ๆ โดยเทียบกับ n เราก็จะไม่คำนึงถึง f.p.c.เพื่อความสะดวกแก่การคำนวน

2. เมื่อ $N \to \infty, \lim_{N\to\infty}(\frac{N-n}{N-1}) = 1$ ดังนั้นการทำ Sampling จาก Infinite Population แล้ว f.p.c. จะหายไป

จาก (7) เมื่อเขียนเสียใหม่จะได้

$$V(\bar{x}_2) = a^2\left\{\frac{S_1^2}{\mu n} + \frac{S_1^2}{(1-\mu)n}\right\} + c^2\frac{S_2^2}{(1-\mu)n} + (1-c)^2\frac{S_2^2}{\mu n} - \frac{2ac\sigma_{1,2}}{(1-\mu)n} \qquad (3.5.1)$$

เพื่อความสะดวกในการคำนวณขั้นต่อไปและเป็นประโยชน์ต่อการเปรียบเทียบประสิทธิภาพ ค่าของ Variance ระหว่างแบบแผน P.R. ต่อ แบบแผน NO,R.และ C.R. ได้ ถ้าเราสมมุติให้

(1) $S_1^2 = S_2^2 = S^2$ [5]

และ(2)ภายใต้ Stationary Assumption $\sigma_{1,2} = \rho S^2$ [6]

$$V(\bar{x}_2) = \frac{a^2 S^2}{\mu n(1-\mu)} + \frac{c^2 S^2}{(1-\mu)n} + (1-c)^2\frac{S^2}{\mu n} - \frac{2ac\rho S^2}{(1-\mu)n} \qquad (3.5.2)$$

การพิจารณาเลือกค่า Combination ของ a และ c ที่จะ Minimized $V(\bar{x}_2)$ นั้น จะมีค่าอยู่คู่หนึ่งเท่านั้นที่จะให้ค่า Variance เล็กสุด วิธีทำก็โดยอาศัยหลักการ Partial Differentiation ดังนี้

$$\frac{\partial V(\bar{x}_2)}{\partial a} = \frac{2aS^2}{\mu n(1-\mu)} - \frac{2c\rho S^2}{(1-\mu)n} = 0 \qquad \text{-------------}(8)$$

$$\frac{\partial V(\bar{x}_2)}{\partial c} = \frac{2cS^2}{(1-\mu)n} - 2(1-c)\frac{S^2}{\mu n} - \frac{2a\rho S^2}{(1-\mu)n} = 0 \qquad \text{-------------}(9)$$

จาก(8) และ (9) เมื่อ Simplify ตามหลักคณิตจะได้

$$a = \frac{(1-\mu)\mu\rho}{1-\mu^2\rho^2}$$

$$c = \frac{1-\mu}{1-\mu^2\rho^2}$$

ซึ่ง $\rho$ คือค่าของ Correlation Coefficient ระหว่างลักษณะที่จะได้ จากข้อมูลทั้ง 2 รอบ ดังนั้นเมื่อแทนค่า a และ c ใน (3.5.2) จะได้

---

[5] เนื่องจากในการทำ Sampling จาก Population เดียวกัน ส่วนที่จะแตกต่างกันนั้นก็คือส่วนที่เป็น Sample mean เท่านั้น ส่วน Sample Variance อาจจะไม่แตกต่างกัน และมักจะเป็นจริงเสมอ

[6] โดยทั่ว ๆ ไปจากสูตร $\frac{Cov}{S_1 S_2}(\bar{x}_1,\bar{x}_2) = \rho$

จากข้อสมมุติที่ว่า $S_1 = S_2 = S$

$Cov(\bar{x}_1,\bar{x}_2) = S_1 S_2 \rho$

$Cov(\bar{x}_1,\bar{x}_2) = \rho S^2 = \sigma_{1,2}$

$$V(\bar{x}_2) = \frac{(1-\mu)^2\mu^2\rho^2}{(1-\mu^2\rho^2)^2}\frac{S^2}{\mu n(1-\mu)} + \frac{(1-\mu)^2}{(1-\mu^2\rho^2)^2}\frac{S^2}{(1-\mu)n} + \left[1-\frac{(1-\mu)}{(1-\mu^2\rho^2)}\right]^2\frac{S^2}{\mu n} - 2\left[\frac{(1-\mu)\mu\rho}{1-\mu^2\rho^2}\cdot\frac{(1-\mu)}{1-\mu^2\rho^2}\right]\frac{\rho S^2}{(1-\mu)n}$$

หลังจาก Simplifly แล้วเราจะได้

$$V(\bar{x}_2) = \frac{s^2}{n}\left\{\frac{1-\mu\rho^2}{1-\mu^2\rho^2}\right\} \qquad (3.5.3)$$

และ Unbiased Estimated Mean หลังจากแทนค่า a และ c แล้วจะได้

$$\bar{x}_2 = \frac{(1-\mu)\mu\rho}{1-\mu^2\rho^2}(\bar{x}_{1,1} - \bar{x}_{1,2}) + \frac{(1-\mu)}{1-\mu^2\rho^2}\bar{x}_{2,1} + \left[1-\frac{(1-\mu)}{1-\mu^2\rho^2}\right]\bar{x}_{2,2} \qquad (3.5.4)$$

3.6 <u>การเปรียบเทียบค่า Variance กรณีสำรวจ 2 รอบติดต่อกันระหว่างแบบแผน P.R. กับแบบแผน NO. R.U หรือ C.R.</u>

จาก (3.5.3) $V(\bar{x}_2) = \frac{S^2}{n}\left\{\frac{1-\mu\rho^2}{1-\mu^2\rho^2}\right\}$

ซึ่งเป็น Variance ของแบบแผน P.R. นั้น จะสังเกตเห็นได้ว่า $\mu$ ที่เป็นค่าของอัตราส่วนในการกำหนดจำนวนหน่วยตัวอย่างที่จะเลือกขึ้นใหม่ ในรอบที่ 2 โดยที่ $0 < \mu < 1$ และสามารถให้ประโยชน์ คือได้ค่า Variance ที่เล็กกว่าทั้งแบบแผน NO. R.U และ C.R. ดังที่จะพิสูจน์ให้เห็นดังนี้ :

เมื่อ $\mu = 0$ หมายถึงการใช้แบบแผน No. R.U.

แทนค่า $\mu = 0$ ใน (3.5.3) $V(\bar{x}_2) = \frac{S^2}{n}\frac{(1-0)}{(1-0)} = \frac{S^2}{n}$

หรือเมื่อ $\mu = 1$ หมายถึงการใช้แบบแผน C.R.

แทนค่า $\mu = 1$ ใน (3.5.3)

$$V(\bar{x}_2) = \frac{s^2}{n}\frac{(1-\rho^2)}{(1-\rho^2)} = \frac{s^2}{n}$$

จะเห็นได้ว่าทั้งแบบแผน NO. R.U. และ C.R. ให้ค่า Variance ที่เท่ากันคือ $\frac{S^2}{n}$ ซึ่งจะต่างกับแบบแผน P.R. อยู่ที่เทอม $\frac{1-\mu\rho^2}{1-\mu^2\rho^2}$ และจะเป็นเทอมที่ทำให้ค่าของ Variance เล็กลงเพราะเหตุว่า

$$0 < \frac{1-\mu\rho^2}{1-\mu^2\rho^2} < 1 \qquad \text{เสมอ}$$

เมื่อ $\rho \neq 0$

และ $\rho^2$ นี้จะเป็นค่าบวกเสมอเมื่อยกกำลัง 2

ดังนี้ $V(\bar{x}_2) < \frac{S^2}{n}$

<u>ตัวอย่างที่ 1</u> สมมุติให้ $\mu = \frac{1}{2}$ ; $\rho = 0.9$

$$V(\bar{x}_2) = \frac{S^2}{n}\left[\frac{1-\frac{1}{2}(0.81)}{1-\frac{1}{4}(0.81)}\right] = \frac{(0.38)}{3.19}\frac{S^2}{n}$$

$$= 0.74\frac{S^2}{n} < \frac{S^2}{n}$$

### 3.7 วิธีการหาค่า Optimum $\mu$ กรณีการสำรวจ 2 รอบติดต่อกัน

เราสามารถจะกำหนดค่าของ $\mu$ เพื่อที่จะให้ได้ค่าการประมาณที่มีประสิทธิภาพสูงสุดได้ โดยอาศัยหลักการของ Partial Differentiation ดังนี้ :-

จาก (3.5.3)
$$\frac{\partial V(\bar{x}_2)}{\partial \mu} = \frac{S^2}{n}\left[\frac{(1-\mu^2\rho^2)(-\rho^2)-(1-\mu\rho^2)(-2\mu\rho^2)}{(1-\mu^2\rho^2)^2}\right] = 0$$
$$-\rho^2+\mu^2\rho^4+2\mu\rho^2-2\mu^2\rho^4 = 0$$
$$-\mu^2\rho^2+2\mu-1 = 0$$
$$\mu_{opt.} = -2\pm\sqrt{4-4(-\rho^2)(-1)}\,/\,2(-\rho^2)$$
$$= -1\pm\sqrt{1-\rho^2}\,/\,-\rho^2$$

ในที่นี้ Optimum $\mu$ มีราก 2 ค่า รากอันหนึ่ง คือ $\frac{1+\sqrt{1-\rho^2}}{\rho^2} > 1$ ซึ่งอยู่นอกเหนือขอบเขตของ P.R. ที่กำหนดไว้ว่า $0 < \mu < 1$ ดังนี้รากค่านี้จึงไม่พิจารณา

$$\mu_{opt.} = \frac{1-\sqrt{1-\rho^2}}{\rho^2} \qquad (3.7.1)$$

ดังนั้น จาก (3.7.1) ถ้าเรากำหนดจำนวนหน่วยตัวอย่างขึ้นจำนวนหนึ่ง คือ ให้เท่ากับ n แล้ว $V(\bar{X}_2)$ จะให้ค่า Variance น้อยที่สุดเท่าใดนั้นขึ้นอยู่กับค่าของ $\rho$ (Correlation Coefficient) ซึ่งจะกล่าวในหัวข้อต่อไป

### 3.8 การประมาณค่าที่เกี่ยวกับค่า $\rho$ (Estimation of Correlation Coefficient)

ในทางปฏิบัติก่อนที่จะกำหนดค่าของ $\mu$ เพื่อให้ผลการประมาณมีประสิทธิภาพสูงสุดนั้นจำเป็นจะต้องคำนวนหาค่าของ $\rho$ ซึ่งโดยคำจำกัดความโดยทั่วไปหมายความถึงการวัดค่าสหสัมพันธ์ (Correlation Coefficient) ระหว่างข้อมูล 2 ชุด แต่ค่าของ $\rho$ ที่ใช้ในวิทยานิพนธ์นี้จะเป็นการวัด (Correlation Coefficient) ระหว่างข้อมูลชุดเดียวกันแต่ต่างกันในระยะเวลา เช่น การวัดค่า $\rho$ ระหว่างข้อมูล ${}_t x_i$ กับ ${}_{t+1}x_i$ โดยที่

${}_t x_i$ = ข้อมูลที่เก็บรวบรวมจากการสำรวจในรอบที่ t อันดับที่ i จากตัวอย่างชุดหนึ่ง ; i = 1,2, ...., n'

${}_{t+1}x_i$ = ข้อมูลที่เก็บรวบรวมจากการสำรวจในรอบที่ t+1 อันดับที่ i จากตัวอย่างชุดเดียวกัน ; i = 1,2, ...., n'

n' = จำนวนตัวอย่างซ้ำและชุดเดียวกันที่ใช้ในการหา Correlation Coefficient

ซึ่ง Correlation Coeficient จะมีค่า $-1 \leq \rho \leq 1$ แต่ในที่นี้เราสนใจแต่เฉพาะค่าบวกเท่านั้น คือ $0 \leq \rho \leq 1$ ยิ่งข้อมูลมีความสัมพันธ์กันมาก ค่า $\rho$ จะยิ่งใกล้หนึ่ง ในทางตรงข้ามค่า $\rho$ จะใกล้ศูนย์ เมื่อข้อมูลทั้งสองมีความสัมพันธ์กันน้อย

ค่าของ $\rho$ บางครั้งเรียกว่า Theoretical or Population Correlation Coefficient ซึ่งในการปฏิบัติยากที่จะคำนวนได้ แต่เราสามารถประมาณค่าของ $\rho$ ได้ด้วย r (Sample Correlation Coefficient) โดยคำจำกัดความ

$$r = \sum_{i=1}^{n'} ({}_t x_i - {}_t\bar{x})({}_{t+1}x_i - {}_{t+1}\bar{x}) \Big/ (n'-1)\, {}_tS_x \, {}_{t+1}S_x$$

$${}_tS_x = \sqrt{\sum_{i=1}^{n'} ({}_t x_i - {}_t\bar{x})^2 / n'-1}$$

$${}_{t+1}S_x = \sqrt{\sum_{i=1}^{n'} ({}_{t+1}x_i - {}_{t+1}\bar{x})^2 / n'-1}$$

และ r มีคุณสมบัติเป็น Unbiased Estimate ของ $\rho$[7]

3.9 <u>วิธีการหา Optimum Variance และ Efficiency Gain กรณีการสำรวจ 2 รอบติดต่อกัน</u>

โดยการแทนค่า $\mu_{opt.}$ ที่ได้ใน (3.7.1.) ใน (3.5.3.) ก็จะได้ Optimum Variance ที่ต้องการสำหรับการสำรวจ 2 รอบติดต่อกัน

$$V_{opt.}(\bar{x}_2) = \frac{S^2}{n}\left[1-\left(\frac{1-\sqrt{1-\rho^2}}{\rho^2}\right)\rho^2 \Big/ 1-\left(\frac{1-\sqrt{1-\rho^2}}{\rho^2}\right)^2\rho^2\right]$$

$$= \frac{S^2}{n}\left[1-1+\sqrt{1-\rho^2} \Big/ (\rho^2-1-2\sqrt{1-\rho^2}-1+\rho^2)/\rho^2\right]$$

$$= \frac{S^2}{2n}\left[\rho^2\sqrt{1-\rho^2} \Big/ \rho^2+\sqrt{1-\rho^2} -1\right] \quad (3.9.1.)$$

การหาผลประสิทธิภาพที่เพิ่มขึ้นเป็นอัตราร้อยละ (Percent Gain in Efficiency) โดยการเปรียบเทียบระหว่าง Variance ของแบบแผน P.R. กับแบบแผน No.R.U. หรือ C.R. สามารถแบ่งการพิจารณาออกเป็น 2 กรณี คือ :-

ก) <u>Percent Gain in Efficiency For Optimum case</u>

---

[7]Paul G.Hoel, <u>Introduction to Mathematical Statistics</u> (New York : John Wiley & Sons Inc., <u>3rd</u> Edition 1962) P.203-205

สูตรการหา Percent Gain in Efficiency For Optimum Case คือ

$$PGE_{opt.} = \frac{V(\bar{x}_0) - V_{opt.}(\bar{x}_2)}{V_{opt.}(\bar{x}_2)} \times 100 \qquad (3.9.2.)$$

ในที่นี้ $V(\bar{x}_0) = \frac{S^2}{n}$ = Variance ของแบบแผน No.R.U. หรือ C.R.

และ $V_{opt.}(\bar{x}_2) = \frac{S^2}{2n}\left[\frac{\rho^2\sqrt{1-\rho^2}}{\rho^2+\sqrt{1-\rho^2}-1}\right]$

แทนค่าดังกล่าวใน (3.9.2)

$$PGE_{opt.} = \left[\frac{S^2}{n} - \left\{\frac{S^2}{2n}\,\frac{\rho^2\sqrt{1-\rho^2}}{\rho^2+\sqrt{1-\rho^2}-1}\right\}\right]\cdot 100 \Big/ \frac{S^2}{2n}\,\frac{\rho^2\sqrt{1-\rho^2}}{\rho^2+\sqrt{1-\rho^2}-1}$$

$$= \left[1 - \frac{\rho^2\sqrt{1-\rho^2}}{2(\rho^2+\sqrt{1-\rho^2}-1)}\right]\cdot 100 \Big/ \frac{\rho^2\sqrt{1-\rho^2}}{2(\rho^2+\sqrt{1-\rho^2}-1)}$$

$$= \frac{2(\rho^2+\sqrt{1-\rho^2}-1) - \rho^2\sqrt{1-\rho^2}}{\rho^2\sqrt{1-\rho^2}} \times 100 \qquad (3.9.3.)$$

จากสูตรที่คำนวนได้ใน (3.7.1) เกี่ยวกับ $\mu_{opt.}$ และ $PGE_{opt.}$ ใน (3.9.3) สามารถนำมาแสดงเปรียบเทียบหาแนวโน้มของ $\mu_{opt.}$ และ $PGE_{opt.}$ มีความสัมพันธ์กันอย่างไร เมื่อเรากำหนดค่า $\rho$ โดยให้อยู่ในอันตรภาค ( Interval ) $0 < \rho \leq 1$ ดังที่แสดงในตาราง ที่ 1 ข้างล่างนี้ : โดยคำนวนจากสูตร (3.9.3)

<u>ตารางที่1</u> การเปรียบเทียบระหว่าง $\mu_{opt.}$ , $(1-\mu)_{opt.}$ และ $PGE_{opt.}$

กรณีสำรวจ 2 รอบติดต่อกัน

| $\rho$ | Optimum % of $\mu$ | Optimum % of $(1-\mu)$ | % Gain in Efficiency |
|---|---|---|---|
| 0.1 | 100.00 | 0 | 0 |
| 0.2 | 52.50 | 47.50 | 1.02 |
| 0.3 | 51.11 | 48.89 | 2.33 |
| 0.4 | 52.50 | 47.50 | 3.40 |
| 0.5 | 53.60 | 46.40 | 6.91 |
| 0.6 | 55.56 | 44.44 | 11.11 |
| 0.7 | 58.36 | 41.64 | 16.57 |
| 0.8 | 62.50 | 37.50 | 25.00 |
| 0.9 | 69.62 | 30.38 | 39.37 |
| 0.95 | 76.23 | 23.77 | 52.48 |
| 1.00 | 100.00 | 0 | 100.00 |

จากตารางที่ 1 แสดงให้เห็นว่าค่าของ Correlation Coefficient ตั้งแต่ 0.1 ถึง 1.0 นั้น ซึ่งค่าของ $\rho$ สูงขึ้น ร้อยละของอัตราการใช้หน่วยตัวอย่างใหม่สูงสุด ( Optimum Rate ) จะสูงตาม และการกำหนดร้อยละของอัตราการใช้หน่วยตัวอย่างสูงขึ้นจะทำให้ประสิทธิภาพความเชื่อมั่น------------------------------สูงขึ้นตาม ในทางตรงกันข้ามกับกรณีการกำหนดอัตราร้อยละ การใช้หน่วยตัวอย่างเก่ายิ่งมากขึ้น ค่าความเชื่อมั่นกลับจะลดลงเมื่อค่าของ $\rho$ สูงขึ้น ดังนั้นเพื่อให้ได้มาซึ่งประสิทธิภาพสูงสุด จึงควรจะกำหนด $\mu_{opt}$ มากกว่า 50 % ขึ้นไปเสมอ

<u>ข้อสังเกต</u>

1) ในกรณีที่เมื่อ $\rho$ =1 ตามสูตร (3.9.3) จะได้ $\mu_{opt.} = 100$ หรืออัตราการเลือกตัวอย่างใหม่ =1 นั่นคือ $(1-\mu)=0$ เช่นนี้หากตัวอย่างที่เลือกขึ้นมาใหม่มีค่าสหสัมพันธ์กับตัวอย่างชุดก่อนมาก ๆ คือ $\rho=1$ แล้ว $PGE_{opt.}$ =100% แต่ภายใต้แบบแผน P.R. มีข้อจำกัดว่า $0<\mu<1$ เท่านั้น ดังนั้นกรณีนี้จึงไม่นำมาพิจารณา

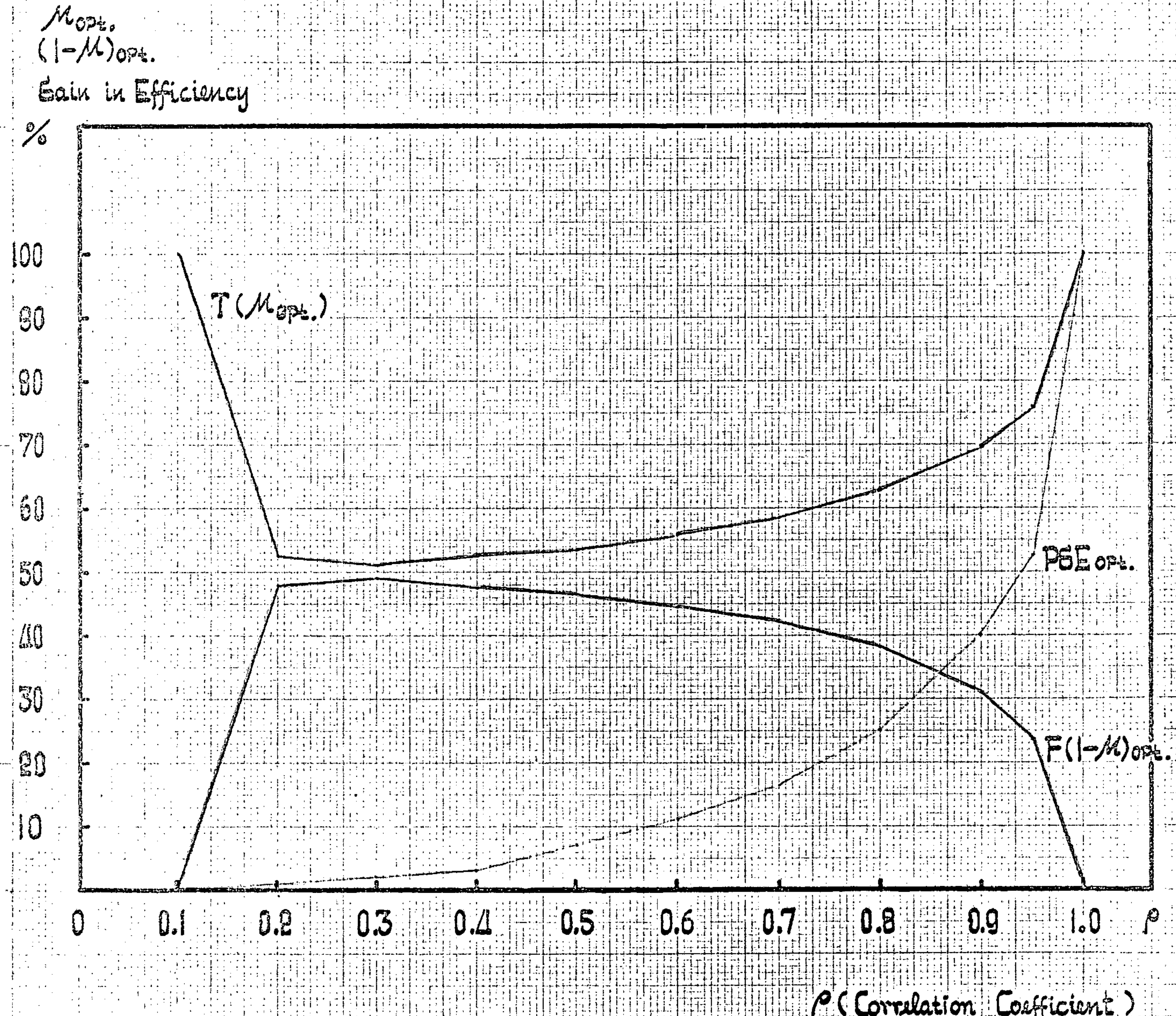
2) แม้ว่าเราสามารถจะคำนวนหา $PGE_{opt}$ ได้โดยง่ายจากสูตร (3.9.3) แต่เมื่อมาคำนึงถึงค่า $\mu$ ที่จะใช้และค่า $PGE_{opt}$ นั้นเป็นค่าไม่ลงตัวซึ่งไม่สะดวกต่อทางปฏิบัติ เช่น $\mu = 0.63$ ที่ $\rho=0.8$ ให้ $PGE_{opt.}=25$ เราอาจใช้เป็นค่า $\mu=\frac{3}{5}$ เป็นค่าใกล้เคียงกับ 0.63 ก็ได้ $PGE_{opt}$ ใกล้เคียงกันมากและสะดวกในทางปฏิบัติมากกว่าเนื่องจากการเลือกหน่วยตัวอย่างใหม่ ไม่สามารถจะกำหนดได้เป็นจุดทศนิยม ในทางปฏิบัติ ดังเช่นทางทฤษฎี

<u>การอธิบายแผนภูมิ 1</u>

การเปรียบเทียบกราฟในแผนภูมิที่ 1 หน้า 27 จะต้องพิจารณาควบคู่กันไปทีละ 2 curves คือ curve $T(\mu)_{opt.}$ กับ curve $PGE_{opt.}$ หรือ curve $F(1-u)_{opt.}$ กับ $PGE_{opt.}$ ซึ่งจะเห็นภาพได้ชัดเจนยิ่งขึ้นเมื่อประกอบกับคำอธิบายของตารางที่ 1 ข้างต้น

แผนภูมิที่ 1 เปรียบเทียบกราฟระหว่าง $\mu_{opt.}$, $(1-\mu)_{opt.}$ และ PGE opt.

กรณีการสำรวจ 2 รอบติดต่อกัน



ข) <u>Percent Gain in Efficiency For Normal Case</u>

สูตรการหา Percent Gain in Efficiency For Normal case

คือ

$$PGE = \frac{V(\bar{x}_0) - V(\bar{x}_2)}{V(\bar{x}_2)} \times 100 \qquad (3.9.4)$$

ในที่นี้ $V(\bar{x}_0) = \frac{S^2}{n}$ Variance ของแบบแผน No.R.U. หรือ C.R.

$$V(\bar{x}_2) = \frac{S^2}{n}\left\{\frac{1-\mu\rho^2}{1-\mu^2\rho^2}\right\}$$

แทนค่าดังกล่าวใน (3.9.4)

$$PGE = \left[\frac{S^2}{n} - \frac{S^2}{n}\left\{\frac{1-\mu\rho^2}{1-\mu^2\rho^2}\right\}\right] \times 100 \Big/ \frac{S^2}{n}\left\{\frac{1-\mu\rho^2}{1-\mu^2\rho^2}\right\}$$

$$= \left[1 - \frac{1-\mu\rho^2}{1-\mu^2\rho^2}\right] \times 100 \Big/ \left\{\frac{1-\mu\rho^2}{1-\mu^2\rho^2}\right\}$$

$$= \frac{\mu\rho^2(1-\mu)}{1-\mu\rho^2} \times 100 \qquad (3.9.5)$$

จากสูตร (3.9.5) ที่คำนวณได้เราสามารถนำมาสร้างตารางหรือเปรียบเทียบที่ได้จากแบบแผน P.R. เมื่อเทียบกับแบบแผน NO.R.U. หรือ C.R. โดยกำหนดค่าของ $\rho$ ให้อยู่ในอันตรภาค $0<\rho\leq 1$ และอัตรา ( $1-\mu$ ) ที่ต่าง ๆ กันเพื่อการพิจารณาเลือกเอาแบบแผนที่จะให้ประสิทธิภาพมากที่สุด จากการางต่อไปนี้โดยคำนวณจากสูตร (3.9.5)

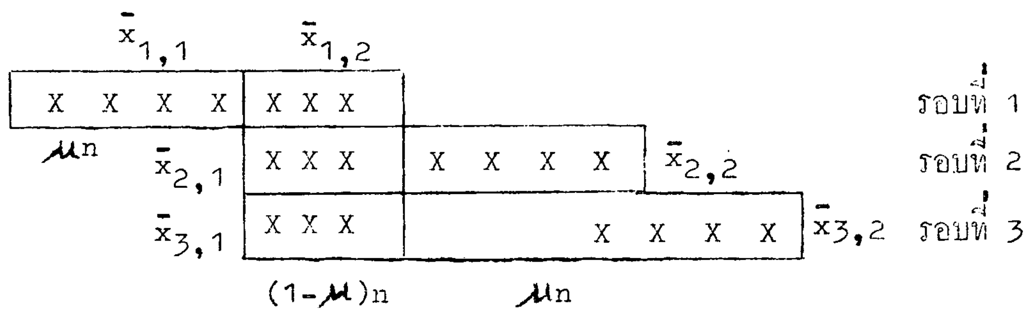
<u>ตารางที่ 2</u> การเปรียบเทียบ PGE กรณีการสำรวจ 2 รอบติดต่อกัน

| $\rho$ | % Gain in Efficiency with $(1-\mu)$ Matched Portion | | | | |
|---|---|---|---|---|---|
| | $(1-\mu)=3/4$ | $(1-\mu)=1/2$ | $(1-\mu)=1/3$ | $(1-\mu)=1/4$ | $(1-\mu)=1/5$ |
| 0.1 | 0.19 | 0.25 | 0.22 | 0.19 | 0.16 |
| 0.2 | 0.76 | 1.02 | 0.91 | 0.77 | 0.66 |
| 0.3 | 1.73 | 2.36 | 2.13 | 1.81 | 1.55 |
| 0.4 | 3.13 | 4.35 | 3.98 | 3.41 | 2.94 |
| 0.5 | 5.00 | 7.14 | 6.66 | 5.77 | 5.00 |
| 0.6 | 7.42 | 10.98 | 10.53 | 9.25 | 8.09 |
| 0.7 | 10.47 | 16.22 | 16.17 | 14.53 | 12.89 |
| 0.8 | 14.29 | 23.53 | 24.81 | 23.08 | 20.98 |
| 0.9 | 19.04 | 34.03 | 39.13 | 38.69 | 36.82 |
| 0.95 | 21.85 | 41.12 | 50.35 | 52.37 | 51.94 |
| 1.0 | 25.00 | 50.00 | 66.67 | 75.00 | 80.00 |

จากตารางที่ 2 จะสังเกตเห็นได้ระหว่างค่าอัตรา ( $1-\mu$ ) ที่ต่าง ๆ กันนั้น จะให้ค่า PGE ในระดับต่าง ๆ กัน ซึ่งพอจะสรุปนำเอาส่วนที่จะเป็นประโยชน์มาใช้เป็นแนวทางได้ คือ

ถ้าหากเรามีข้อมูล 2 ชุด ที่มีความสัมพันธ์ $0.1 \leq \rho \leq 0.7$ เราก็ควรที่จะเลือกใช้แบบแผนที่กำหนดอัตรา $(1-\mu) = \frac{1}{2}$ ซึ่งจะให้ค่า PGE สูงกว่าการกำหนด $(1-\mu)$ อัตราอื่น ถ้าหาก $0.8 \leq \rho \leq 1.0$ ก็ควรหันกลับมาใช้แบบแผนที่กำหนดอัตรา $(1-\mu) = \frac{1}{3}$ เพราะกลับให้ค่า PGE สูงกว่าเมื่อกำหนด $(1-\mu) = \frac{1}{2}$

### 3.10 ทฤษฎีการประมาณค่ามัชฌิมจากรอบสุดท้ายในกรณีการสำรวจ 3 รอบติดต่อกัน




ในการสุ่มตัวอย่างเพื่อการประมาณค่ามัชฌิมจากรอบสุดท้าย โดยให้เป็นไปตาม**แบบแผนในรูป-----** ดังกล่าวข้างต้น คือ จะใช้หน่วยตัวอย่างส่วนหนึ่งจากรอบที่หนึ่งเป็นตัวอย่างซึ่งมีจำนวนเท่ากับ $(1-\mu)n$ ไปใช้ในรอบที่ 2 และ 3 **ตามลำดับเหมือนกันทั้ง** 3 รอบ ส่วนหน่วยตัวอย่างที่เหลือจะเลือกขึ้นมาใหม่ทั้งในรอบที่ 2 และ 3 เท่ากับ $\mu n$ หน่วย ซึ่งวิธีการประมาณค่า Mean จากรอบที่ 3 นั้น สามารถคำนวนได้ในรูปของ Linear Combination ดังนี้ :

$$\bar{X}_3 = a\bar{x}_{1,1} + b\bar{x}_{1,2} + c\bar{x}_{2,1} + d\bar{x}_{2,2} + e\bar{x}_{3,1} + f\bar{x}_{3,2}$$

$\bar{X}_3$ = Estimated mean for third occasion

a, b, c, d, e, f = Arbitrary Constants

$$E(\bar{X}_3) = E(a\bar{x}_{1,1}+b\bar{x}_{1,2}+c\bar{x}_{2,1}+d\bar{x}_{2,2}+e\bar{x}_{3,1}+f\bar{x}_{3,2})$$

$$E(\bar{x}_{1,1}) = E(\bar{x}_{1,2}) = \theta_1 \; ; \; E(\bar{x}_{2,1}) = E(\bar{x}_{2,2}) = \theta_2 \; ; \; E(\bar{x}_{3,1}) = E(\bar{x}_{3,2}) = \theta_3$$

ดังนั้น $E(\bar{X}_3) = (a+b)\,\theta_1 + (c+d)\,\theta_2 + (e+f)\,\theta_3$

$E(\bar{X}_3)$ จะเป็น Unbiased Estimate ก็ต่อเมื่อ :

$$a+b = 0 \; ; \quad c+d = 0 \; ; \quad e+f = 1$$

นั่นคือ $b = -a \; ; \quad d = -c \; ; \quad f = 1-e$

$$\therefore \bar{X}_3 = a(\bar{x}_{1,1}-\bar{x}_{1,2}) + c(\bar{x}_{2,1}-\bar{x}_{2,2}) + e\bar{x}_{3,1} + (1-e)\bar{x}_{3,2}$$

$$= a(\bar{x}_{1,1}-\bar{x}_{1,2}) + c(\bar{x}_{2,1}-\bar{x}_{2,2}) + e(\bar{x}_{3,1}-\bar{x}_{3,2}) + \bar{x}_{3,2} \quad (3.10.1)$$

3.11 การประมาณค่า Variance จากรอบสุดท้ายกรณีการสำรวจ 3 รอบติดต่อกัน

$$V(\bar{x}_3) = V\left[a(\bar{x}_{1,1}-\bar{x}_{1,2})+c(\bar{x}_{2,1}-\bar{x}_{2,2})+e(\bar{x}_{3,1}-\bar{x}_{3,2})+\bar{x}_{3,2}\right]$$
$$= a^2V(\bar{x}_{1,1}-\bar{x}_{1,2})+c^2V(\bar{x}_{2,1}-\bar{x}_{2,2})+e^2V(\bar{x}_{3,1}-\bar{x}_{3,2})+V(\bar{x}_{3,2})$$
$$+ 2ac\ Cov\left\{(\bar{x}_{1,1}-\bar{x}_{1,2}),(\bar{x}_{2,1}-\bar{x}_{2,2})\right\}+2ae\ Cov\left\{(\bar{x}_{1,1}-\bar{x}_{1,2}),(\bar{x}_{3,1}-\bar{x}_{3,2})\right\}$$
$$+ 2a\ Cov\left\{(\bar{x}_{1,1}-\bar{x}_{1,2}),\bar{x}_{3,2}\right\}+2ce\ Cov\left\{(\bar{x}_{2,1}-\bar{x}_{2,2}),(\bar{x}_{3,1}-\bar{x}_{3,2})\right\}$$
$$+ 2c\ Cov\left\{(\bar{x}_{2,1}-\bar{x}_{2,2}),\bar{x}_{3,2}\right\}+2e\ Cov\left\{(\bar{x}_{3,1}-\bar{x}_{3,2}),\bar{x}_{3,2}\right\}$$

โดยอาศัยทฤษฎีการกระจายเทอม Variance และ Covariance [8] และ Simplify แล้วจะได้

$$V(\bar{x}_3) = a^2\left\{\frac{S_1^2}{\mu n}+\frac{S_1^2}{(1-\mu)n}\right\}+c^2\left\{\frac{S_2^2}{(1-\mu)n}+\frac{S_2^2}{\mu n}\right\}+e^2\left\{\frac{S_3^2}{(1-\mu)n}+\frac{S_3^2}{\mu n}\right\}+\frac{S_3^2}{\mu n}$$
$$-2ac\frac{\sigma_{1.2}}{(1-\mu)n}-2ae\frac{\sigma_{1.3}}{(1-\mu)n}+2ce\frac{\sigma_{2.3}}{(1-\mu)n}-2e\frac{S_3^2}{\mu n}$$

หลังจาก Simplify แล้วจะได้

$$V(\bar{x}_3) = \frac{a^2S_1^2}{\mu n(1-\mu)}+\frac{c^2S_2^2}{\mu n(1-\mu)}+\frac{e^2S_3^2}{(1-\mu)n}+(1-e)^2\frac{S_3^2}{\mu n}-2ae\frac{\sigma_{1.2}}{(1-\mu)n}-2ae\frac{\sigma_{1.3}}{(1-\mu)n}+2ce\frac{\sigma_{2.3}}{(1-\mu)n}$$

ภายใต้ข้อสมมุติฐานที่ว่า [8.1]

$$\left.\begin{array}{c} S_1^2 = S_2^2 = S_3^2 \\ \sigma_{tt'} = \rho S^2 \end{array}\right\} \quad (t \neq t' = 1,2,3)$$

$$V(\bar{x}_3) = \frac{a^2S^2}{\mu n(1-\mu)}+\frac{c^2S^2}{\mu n(1-\mu)}+\frac{e^2S^2}{(1-\mu)n}+(1-e)^2\frac{S^2}{\mu n}-2ac\frac{\rho S^2}{(1-\mu)n}-2ae\frac{\rho S^2}{(1-\mu)n}+2ce\frac{\rho S^2}{(1-\mu)n} \quad *** (3.11.1)$$

จะพิจารณาเลือกค่าของ a,c,e ที่จะ Minimized $V(\bar{x}_3)$ โดยอาศัยวิธีการ Partial Differentiation ดังนี้ :

$$\frac{\partial V(\bar{x}_3)}{\partial a} = \frac{2aS^2}{\mu n(1-\mu)}-\frac{2c\rho S^2}{(1-\mu)n}-\frac{2e\rho S^2}{(1-\mu)n} = 0 \quad \text{-----------} (1)$$

$$\frac{\partial V(\bar{x}_3)}{\partial c} = \frac{2cS^2}{\mu n(1-\mu)}-\frac{2a\rho S^2}{(1-\mu)n}+\frac{2e\rho S^2}{(1-\mu)n} = 0 \quad \text{-----------} (2)$$

$$\frac{\partial V(\bar{x}_3)}{\partial e} = \frac{2eS^2}{(1-\mu)n}-2(1-e)\frac{S^2}{\mu n}-2a\frac{\rho S^2}{(1-\mu)n}+\frac{2c\rho S^2}{(1-\mu)n} \quad \text{---------} (3)$$

---

8 Ibid P.19

8.1 Ibid P.21

จาก (1),(2) และ (3) เมื่อ Simplify แล้ว จะได้

$$e = (1-\mu)(1-\mu^2\rho^2) / 1+\mu^2\rho^2(2\mu\rho-3) \quad \text{---- (4)}$$

$$a = (1-\mu)\mu\rho(1-\mu\rho) / 1+\mu^2\rho^2(2\mu\rho-3) \quad \text{---- (5)}$$

$$e = -(1-\mu)\mu\rho(1-\mu\rho) / 1+\mu^2\rho^2(2\mu\rho-3) \quad \text{---- (6)}$$

นั่นคือ $a = -e$

จาก (3 .11.1) สามารถเขียนสมการใหม่เป็น

$$V(\bar{x}_3) = \frac{S^2}{\mu n(1-\mu)}\left[a^2+e^2+\mu e^2+(1-\mu)(1-e)^2-2ae\mu\rho-2ae\mu\rho+2ee\mu\rho\right]$$

แทนค่า $e = -a$ จะได้

$$V(\bar{x}_3) = \frac{S^2}{\mu n(1-\mu)}\left[1+2a^2(1+\mu\rho)-4ae\mu\rho-2e(1-\mu)+e^2-\mu\right] \text{-- (7)}$$

แทนค่า $a$, $e$ และ $e$ ใน (7) หลังจาก Simplify แล้วจะได้

$$V(\bar{x}_3) = \frac{S^2}{n}\frac{1}{\mu(1-\mu)}\left[(1-\mu) - \frac{(1-\mu)^2(1-\mu^2\rho^2)}{1+\mu^2\rho^2(2\mu\rho-3)}\right]$$

$$= \frac{S^2}{\mu n}\left[1-\left\{(1-\mu)(1-\mu^2\rho^2) / 1+\mu^2\rho^2(2\mu\rho-3)\right\}\right]$$

$$= \frac{S^2}{\mu n}\left[\mu(2\mu^2\rho^3-2\mu\rho^2-\mu^2\rho^2+1) / 1+\mu^2\rho^2(2\mu\rho-3)\right]$$

$$= \frac{S^2}{n}\left[\frac{1+\mu\rho^2(2\mu\rho-\mu-2)}{1+\mu^2\rho^2(2\mu\rho-3)}\right] \quad *** (3.11.2)$$

และเมื่อแทนค่า $a$, $e$ และ $e$ ใน (3.10.1) จะได้ Unbiased Estimated Mean ดังนี้ :

$$\bar{x}_3 = \frac{(1-\mu)\mu\rho(1-\mu\rho)}{1+\mu^2\rho^2(2\mu\rho-3)}(\bar{x}_{1,1}-\bar{x}_{2,1}) - \frac{(1-\mu)\mu\rho(1-\mu\rho)}{1+\mu^2\rho^2(2\mu\rho-3)}(\bar{x}_{2,1}-\bar{x}_{2,2})$$
$$+ \frac{(1-\mu)(1-\mu^2\rho^2)}{1+\mu^2\rho^2(2\mu\rho-3)}(\bar{x}_{3,1}-\bar{x}_{3,2}) + \bar{x}_{3,2} \quad *** (3.11.3)$$

## 3.12 การเปรียบเทียบค่า Variance ระหว่างแบบแผน P.R. กับ NO.R.U. หรือ C.R. กรณีการสำรวจ 3 รอบติดต่อกัน

จาก (3.11.2) $\quad V(\bar{x}_3) = \frac{S^2}{n}\left[\frac{1+\mu\rho^2(2\mu\rho-\mu-2)}{1+\mu^2\rho^2(2\mu\rho-3)}\right]$

ค่าของ Variance ใน (3.11.2) สามารถจะให้ค่าซึ่งเล็กกว่าทั้งแบบ NO.R.U. และ CR. ดังจะพิสูจน์ให้เห็นได้ คือ :

เมื่อ $\mu = 0$ หมายถึงการใช้แบบแผน NO.R.U.

แทนค่า $\mu = 0$ ใน (3.11.2) $V(\bar{x}_3) = \frac{s^2}{n}\left[\frac{1+0}{1+0}\right] = \frac{s^2}{n}$

หรือเมื่อ $\mu = 1$ หมายถึงการใช้แบบแผน C.R.

แทนค่า $\mu = 1$ ใน (3.11.2) $V(\bar{x}_3) = \frac{s^2}{n}\left[\frac{1+\rho^2(2\rho-3)}{1+\rho^2(2\rho-3)}\right] = \frac{s^2}{n}$

จะเห็นได้ว่าทั้งแบบแผน No.R.U. และ C.R. ให้ค่า Variance ที่เท่ากันคือ $\frac{s^2}{n}$ ซึ่งจะต่างกับแบบแผน P.R. อยู่ที่เทอม $\frac{1+\mu\rho^2(2\mu\rho-\mu-2)}{1+\mu^2\rho^2(2\mu\rho-3)}$ และจะเป็นเทอมที่ทำให้ Variance เล็กลง เพราะเหตุว่า $0 < \frac{1+\mu\rho^2(2\mu\rho-\mu-2)}{1+\mu^2\rho^2(2\mu\rho-3)} < 1$ เสมอ

เมื่อ $\rho \neq 0$

ดังนั้นจึงกล่าวได้ว่าแบบแผนP.R. สามารถกำหนดค่าการประมาณที่ได้มีประสิทธิภาพดีขึ้น และ Estimated Mean ใน (3.11.3) นอกจากจะเป็น Unbiased Estimate แล้ว ยังมี Estimated Variance ที่ค่าเล็กด้วย คือ :

$$V(\bar{x}_3) < \frac{s^2}{n} \qquad \text{เมื่อ} \quad \rho \neq 0$$

เพื่อให้เห็นได้ชัดเจนจึงได้นำเอาตัวอย่างโดยสมมุติตัวเลขซึ่งทางปฏิบัติมักนิยมใช้มาแสดงให้เห็น ดังนี้ :

<u>ตัวอย่างที่</u> 2 สมมุติ $\mu = \frac{1}{2}$ ; $\rho = 0.9$

$$V(\bar{x}_3) = \frac{s^2}{n}\cdot\left[\frac{1+\frac{1}{2}(0.81)(2.\frac{1}{2}.(0.9)-\frac{1}{2}-2)}{1+\frac{1}{4}(0.81)(2.\frac{1}{2}.9-3)}\right]$$

$$\approx \frac{0.36}{0.58}\ \frac{s^2}{n} \approx 0.63\ \frac{s^2}{n}$$

<u>ตัวอย่างที่ 3</u> สมมุติ $\mu = \frac{3}{4}$ ; $\rho = .9$

$$V(\bar{x}_3) = \frac{S^2}{n}\left[\frac{1+\frac{3}{4}(0.81)\ (2.\frac{3}{4}\ (0.9)-\frac{3}{4}-2)}{1+\frac{9}{16}(0.81)\ (2.\frac{3}{4}\ (.9)\ -3)}\right]$$

$$\approx \frac{0.16}{0.26}\cdot\frac{S^2}{n}$$

$$\approx 0.61\ \frac{S^2}{n}$$

## 3.13 <u>วิธีการหาค่า Optimum $\mu$ กรณีการสำรวจ 3 รอบติดต่อกัน</u>

โดยอาศัยทฤษฎีเกี่ยวกับ Partial Differentiation ในการหา Optimum $\mu$ ได้ดังนี้ : จาก (3.11.2) $\frac{\partial V(\bar{X}_3)}{\partial \mu} = 0$

$$\frac{\partial V(\bar{X}_3)}{\partial \mu} = \frac{S^2}{n}\left[\frac{(1+2\mu^3\rho^3-3\mu^2\rho^2)(4\mu\rho^3-2\mu\rho^2-2\rho^2)-(1+2\mu^2\rho^3-\mu^2\rho^2-2\mu\rho^2)(6\mu^2\rho^3-6\mu\rho^2)}{(1+2\mu^3\rho^3-3\mu^2\rho^2)^2}\right] = 0$$

$$4\mu\rho^3+4\mu\rho^2-2\rho^2-4\mu^4\rho^6+2\mu^4\rho^5+8\mu^3\rho^5-6\mu^2\rho^4-6\mu^2\rho^3 = 0 \quad \text{------}(1)$$

$2\rho^2$ หาร (1) ตลอด $2\mu\rho+2\mu-1-2\mu^4\rho^4+\mu^4\rho^3+4\mu^3\rho^3-3\mu^2\rho^2-3\mu^2\rho = 0$

$$(\rho^3-2\rho^4)\mu^4+4\rho^3\mu^3-3(\rho^2+\rho)\mu^2+2(\rho+1)\mu-1 = 0 \quad \text{----}(3.13.1)$$

จาก (3.13.1) เป็นสมการที่จะคำนวนหาค่า Optimum $\mu$ ได้ด้วยการหารากที่ 4 ของ $\mu$ ซึ่งในที่นี้เราไม่สามารถจะคำนวนหารากที่ 4 ได้โดยวิธีแยกแฟคเตอร์หรือวิธีอื่นใด นอกจากวิธีการหาค่ารากโดยประมาณในแบบ Iteration วิธีการดังกล่าวมีอยู่หลายวิธี แต่มีวิธีหนึ่งซึ่งสามารถจะให้ค่าได้ใกล้เคียงมากและเป็นที่รู้จักในนามของ " <u>Newton - Raphson Method</u>"[9] โดยมีวิธีการคำนวนดังต่อไปนี้ จาก (3.13.1) สามารถเขียนใหม่เป็น

$$A\mu^4 + B\mu^3 - C\mu^2 + D\mu - E = 0 \qquad (3.13.2)$$

where A,B,C,D และ E = Arbitrary Constants

เนื่องจากเทอมของ $\rho$ ซึ่งเป็นค่า Constant คือ $0 < \rho \leq 1$

---

[9] Churchman, Ackoff, Arnoff, <u>Introduction to Operations Research</u> (New york : John Wiley & Sons, Inc., 1963) P. 191-192

<u>วิธีการหา</u> ให้ $f(Y)$ เป็นฟังชั่นที่ต้องการจะหาราก

สูตร $$Y_{n+1} = Y_n - \frac{f(Y_n)}{f'(Y_n)} \qquad \text{***}$$

ในที่นี้ $Y_n$ = ค่าใด ๆ ที่สมมุติขึ้นที่คิดว่าจะเป็นราก

$f(Y_n)$ = $f(Y)$ ที่แทนค่า Y ด้วย $Y_n$

$f'(Y_n)$ = First derivative $f(Y)$ ของ Y ที่แทนค่าด้วย $Y_n$

$Y_{n+1}$ = รากประมาณที่ต้องการจะหา เมื่อ $\frac{f(Y_n)}{f'(Y_n)}$ Converge เข้าสู่ 0 แล้ว

กรณีที่ยังไม่ Converge เข้าสู่ 0 แล้ว ก็จะต้องทำไปเรื่อย ๆ ตามวิธีเดิมจนกว่าจะได้ค่ารากประมาณที่ต้องการ จากวิธีข้างต้นสามารถนำมาคำนวนหา $\mu_{opt.}$ ซึ่งโดยปกติแล้วจะได้ค่าราก 4 ราก แต่จะมีรากเพียงรากเดียวที่จะให้ค่า PGEopt. ได้ นอกจากนี้ $0 < \mu < 1$ ซึ่งเป็นข้อจำกัดอีกข้อหนึ่งด้วย การคำนวนหา $\mu_{opt.}$ และ PGEopt. ได้แสดงให้เห็นตามตารางที่ 3 นี้

<u>ตารางที่ 3</u> การเปรียบเทียบระหว่าง $\mu_{opt.}$, $(1-\mu)_{opt.}$ และ $PGE_{opt.}$ กรณีการสำรวจ 3 รอบติดต่อกัน

| $\rho$ | Optimum % of $\mu$ | Optimum % of $(1-\mu)$ | % Gain in Efficiency |
|---|---|---|---|
| 0.5 | 50 | 50 | 12.5 |
| 0.6 | 52 | 48 | 19.2 |
| 0.7 | 55 | 45 | 28.7 |
| 0.8 | 58 | 42 | 43.1 |
| 0.9 | 65 | 35 | 69.3 |
| 1.0 | 100 | 0 | 100.0 |

จากตารางที่ 3 ค่า $0 < \rho \leq 1$ นั้น/เนื่องจากในทางปฏิบัติค่า $\rho$ ของข้อมูลชุดเดียวกัน แต่ในระยะเวลาที่ต่างกัน ส่วนใหญ่มักจะมีค่ามากกว่า 0.5 ขึ้นไป ดังที่ได้แสดงตัวอย่างให้เห็นในตัวอย่างที่ 4 หน้า 39

เมื่อค่า $\rho$ สูงขึ้น $\mu_{opt.}$ จะมากขึ้นและ$PGE_{opt.}$ ก็จะเพิ่มขึ้นตามด้วย ในทางตรงข้าม " $(1-\mu)_{opt.}$ น้อยลงแต่ " กลับเพิ่มขึ้น .

หน้า 36

พิจารณาดูกราฟในแผนที่ 2 และ 3/ประกอบ จากแผนภูมิที่ 3 จะเห็นว่า Curve ของ $PGE_{opt.}$ ของการสำรวจ 3 รอบติดต่อกันอยู่ในระดับที่สูงกว่า $PGE_{opt.}$ ของการสำรวจ 2 รอบติดต่อกัน เนื่องจาก $PGE_{opt.}$ ของ 3 รอบ ให้ประสิทธิภาพที่ดีกว่า คือ Curve จะค่อย ๆ กางออกเมื่อค่า $\rho$ มากขึ้น และจะมาบรรจบกันเมื่อ $\rho = 1$

## 3.14 การหา Percent Gain in Efficiency ของการสำรวจ 3 รอบติดต่อกัน

สูตรการคำนวนหา Percent Gain in Efficiency คือ

$$PGE = \frac{V(\bar{x}_o) - V(\bar{x}_3)}{V(\bar{X}_3)} \times 100$$

ในที่นี้ $V(\bar{X}_o) = \frac{S^2}{n}$ = Variance ของแบบแผน No. R.หรือ CR.

และ $V(\bar{X}_3) = \frac{S^2}{n}\left[\frac{1+\mu\rho^2(2\mu\rho-\mu-2)}{1+\mu^2\rho^2(2\mu\rho-3)}\right]$

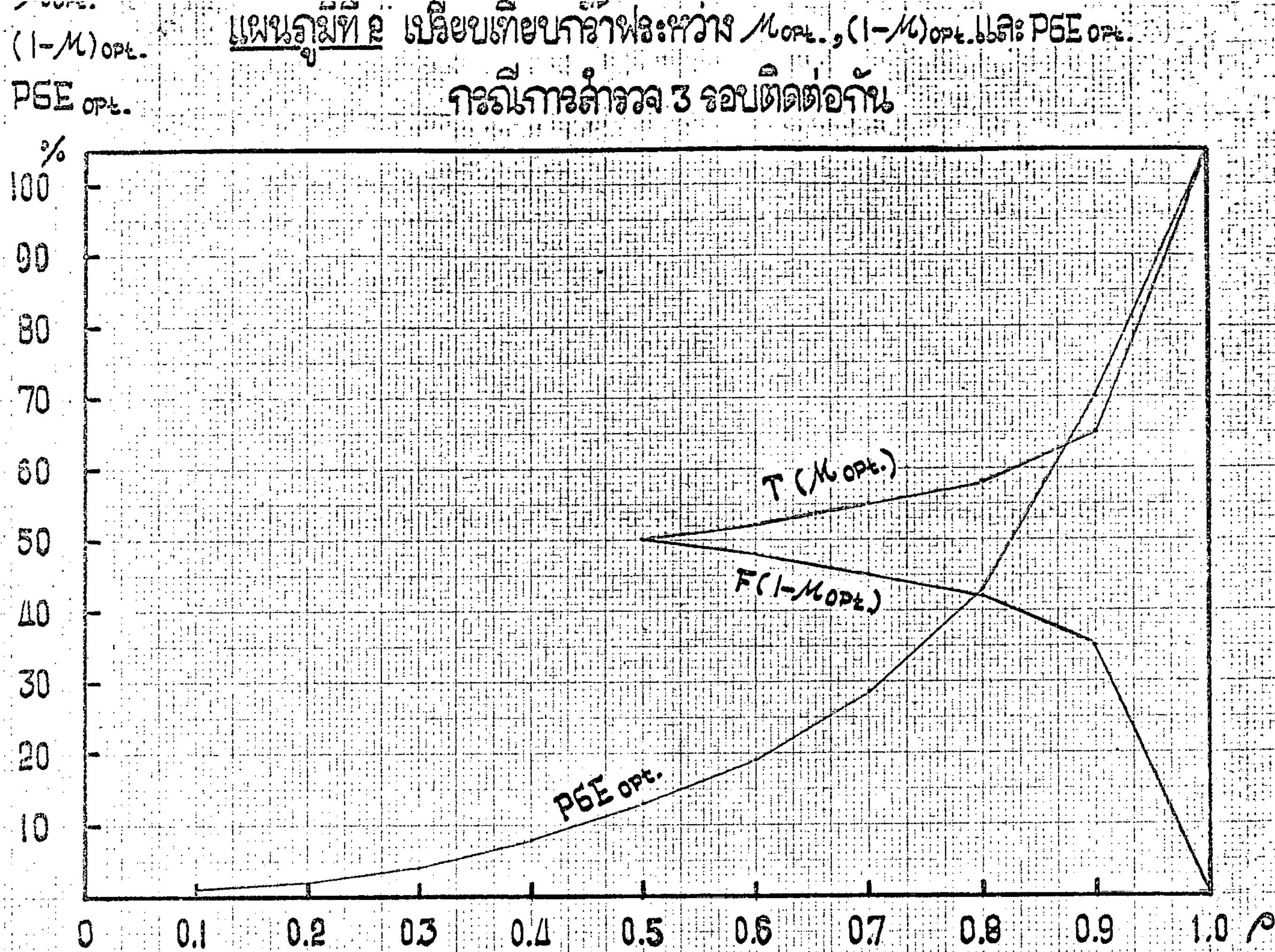
$$PGE = \left[\frac{S^2}{n} - \frac{S^2}{n}\left\{\frac{1+\mu\rho^2(2\mu\rho-\mu-2)}{1+\mu^2\rho^2(2\mu\rho-3)}\right\}\right]100 \Big/ \frac{S^2}{n}\left\{\frac{1+\mu\rho^2(2\mu\rho-\mu-2)}{1+\mu^2\rho^2(2\mu\rho-3)}\right\}$$

หลังจาก Simplify แล้วได้เป็น

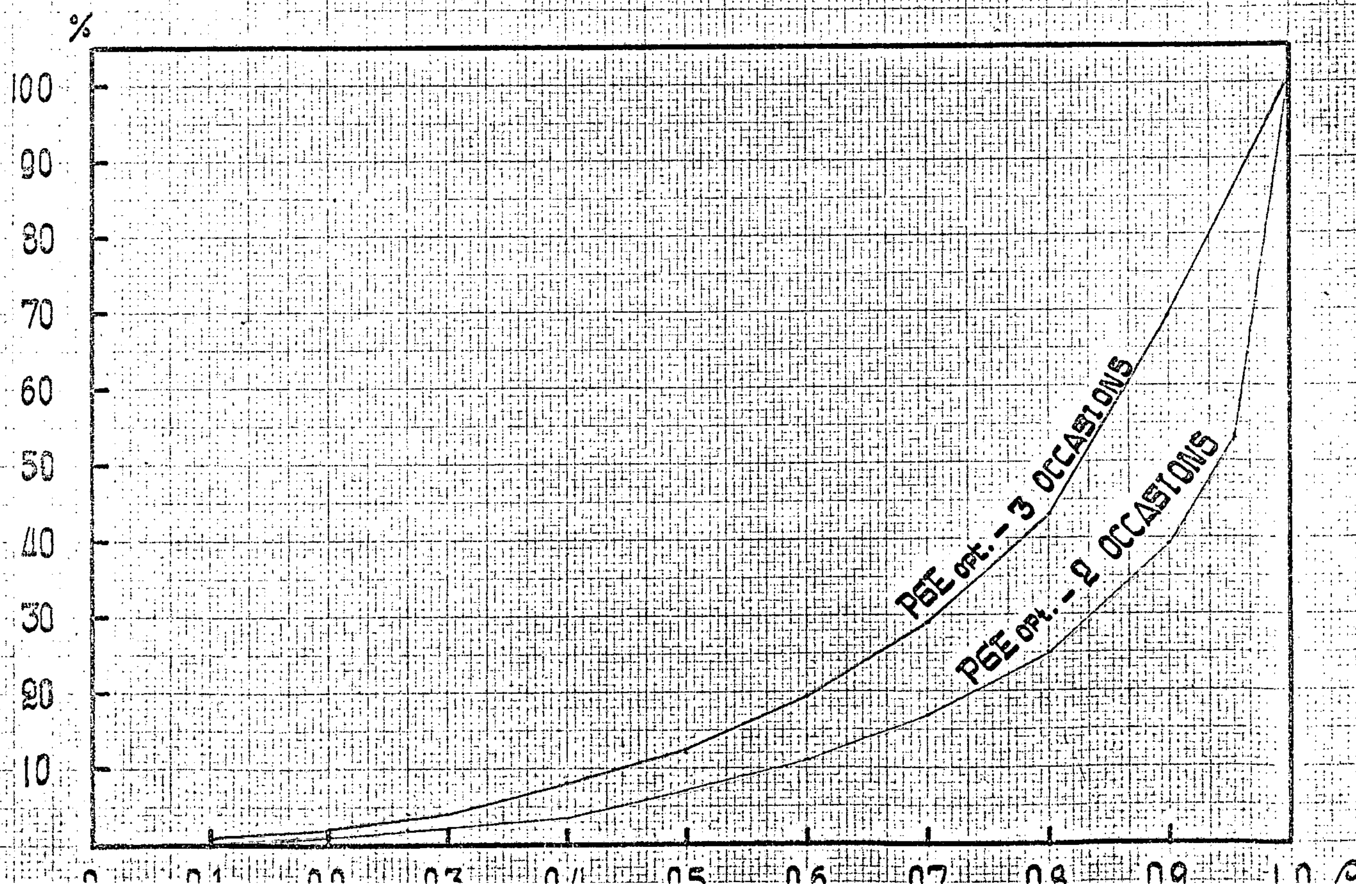
$$PGE = 2\mu\rho^2(\mu^2\rho-\mu-\mu\rho+1)\ 100 \Big/ 1+\mu\rho^2(2\mu\rho-\mu-2) \quad *** \quad (3.14.1)$$

แผนภูมิที่ 2 เปรียบเทียบกราฟระหว่าง $\mu_{opt.}$, $(1-\mu)_{opt.}$ และ PSE opt.

กรณีการสำรวจ 3 รอบติดต่อกัน




แผนภูมิที่ 3 เปรียบเทียบกราฟ PSE opt. ระหว่างการสำรวจ 2 และ 3 รอบติดต่อกัน



จากสูตร (3.14.1) ที่คำนวนได้ เราสามารถนำมาสร้างตารางเพื่อเปรียบเทียบ PGE ที่ได้จากแบบแผน P.R. เมื่อเทียบกับแบบแผน No. R.U. หรือ CR. โดยกำหนดค่า $\rho$ ให้อยู่ในอันตรภาค $0<\rho \leq 1$ และอัตรา $(1-\mu)$ ที่ต่าง ๆ กัน เพื่อการพิจารณาเลือกเอาแบบแผนที่จะให้ประสิทธิภาพสูงสุด จากตารางที่ 4 ดังต่อไปนี้โดยอาศัยสูตร (3.14.1)

<u>ตารางที่ 4</u> การเปรียบเทียบ PGE ในกรณีการสำรวจ 3 ครั้งติดต่อกัน

| $\rho$ | % Gain in Efficiency with $(1-\mu)$ Matched Portion | | | | |
|---|---|---|---|---|---|
| | $(1-\mu)=3/4$ | $(1-\mu)=1/2$ | $(1-\mu)=1/3$ | $(1-\mu)=1/4$ | $(1-\mu)=1/5$ |
| 0.1 | 0.37 | 0.48 | 0.42 | 0.35 | 0.30 |
| 0.2 | 1.46 | 1.89 | 1.64 | 1.38 | 1.17 |
| 0.3 | 3.28 | 4.25 | 3.70 | 3.10 | 2.63 |
| 0.4 | 5.88 | 7.69 | 6.75 | 5.66 | 4.81 |
| 0.5 | 9.38 | 12.50 | 11.11 | 9.37 | 8.00 |
| 0.6 | 13.92 | 19.15 | 17.39 | 14.84 | 12.74 |
| 0.7 | 19.76 | 28.49 | 26.78 | 23.26 | 20.21 |
| 0.8 | 27.27 | 42.10 | 41.83 | 37.50 | 33.25 |
| 0.9 | 37.04 | 63.28 | 69.23 | 66.03 | 54.39 |
| 0.95 | 43.04 | 78.82 | 93.28 | 94.34 | 91.39 |
| 1.00 | 50.00 | 100.00 | 133.33 | 150.00 | 200.00 |

ตัวอย่างที่ 4 จากตัวเลขจริงในทางปฏิบัติ[10]

จากตารางที่ 5 เป็นข้อมูลเกี่ยวกับการสำรวจแรงงานในเขตเทศบาลพระนคร-ธนบุรี พ.ศ.2510 ซึ่งได้นำเข้ามาเป็นตัวอย่างประกอบเพื่อให้เห็นถึงประโยชน์จากสูตรที่ได้สร้างขึ้นในตอนต้นบทที่ 3 เกี่ยวกับการประมาณค่า Correlation Coefficient และ Estimated Variance

ตารางที่ 5 ข้อมูลการสำรวจแรงงานในเขตเทศบาลพระนคร-ธนบุรี พ.ศ.2510

| เขต | จำนวนตัวอย่าง Segments ที่แจงนับซ้ำกัน | จำนวนครัวเรือน | | จำนวนประชากร | |
|---|---|---|---|---|---|
| | | รอบที่ 1 | รอบที่ 2 | รอบที่ 1 | รอบที่ 2 |
| พระนคร | 90 | 1,952 | 1,910 | 11,757 | 11,525 |
| ธนบุรี | 27 | 616 | 628 | 3,673 | 3,803 |
| รวม | 117 | 2,568 | 2,538 | 15,430 | 15,328 |

ที่มาแห่งข้อมูล การสำรวจแรงงาน พ.ศ.2510 สำนักงานสถิติแห่งชาติ

หมายเหตุ รายละเอียดของข้อมูลที่แต่ละ Segment ทำการแจงนับจำนวนครัวเรือนและจำนวนประชากรของทั้ง 2 รอบ ไม่ขอนำมาแสดงไว้เนื่องจากจะกินเนื้อที่กระดาษมาก

วิธีการคำนวน

ก) การประมาณเกี่ยวกับ Correlation Coefficient และ Estimated Variance จากการสำรวจรอบที่ 2 ของครัวเรือน

---

[10] ข้อมูลของการสำรวจดังกล่าวสำนักงานสถิติแห่งชาติได้จัดทำไว้ 2 ครั้ง ๆ ละ 3 เดือน คือระหว่างเดือน ก.พ.-เม.ย.และ ส.ค.-ต.ค.2510 อีกครั้งหนึ่ง และให้ตัวอย่างที่แจงนับนั้นซ้ำกันครั้งหนึ่งในระหว่างรอบ ซึ่งมีหน่วยตัวอย่าง Segments ในรอบหนึ่ง ๆ เป็นจำนวน 234 หน่วย เพื่อใช้ในการศึกษาเปรียบเทียบลักษณะต่าง ๆ

$_1\bar{X}_h$ = Mean ของการสำรวจจำนวน HH's ในรอบที่ 1 = 22 ครัวเรือนต่อเขต

$_2\bar{X}_h$ = Mean ของการสำรวจจำนวน HH's ในรอบที่ 2 = 22 ครัวเรือนต่อเขต

$$S_{_1x_h} = \sqrt{\sum_{i=1}^{n} ({}_1X_{ih} - {}_1\bar{X}_h)^2 / n'-1} = \sqrt{\frac{5,187}{116}} = 6.69$$

$$S_{_2x_h} = \sqrt{\sum_{i=1}^{n} ({}_2X_{ih} - {}_2\bar{X}_h)^2 / n'-1} = \sqrt{\frac{4,384}{116}} = 6.15$$

Sample Correlation Coefficient $r_h = \dfrac{\sum_{i=1}^{n} ({}_1X_{ih} - {}_1\bar{X}_h)({}_2X_{ih} - {}_2\bar{X}_h)}{(n'-1)\, {}_1S_{x_h}\, {}_2S_{x_{h_2}}} = \dfrac{4,551}{4,772} = .95$

$$V_h(\bar{X}_2) = \frac{S^2}{n}\left[\frac{1-\mu^2}{1-\mu^2\rho^2}\right] = \frac{37.79}{234}\left[\frac{1-\frac{1}{2}(.95)^2}{1-\frac{1}{4}(.95)^2}\right] = 0.16\left(\frac{2.18}{3.09}\right)$$

$$= 0.11 < 0.16$$

ข) <u>การประมาณเกี่ยวกับ Correlation Coefficientและ Estimated Variance</u>

<u>จากรอบที่ 2 ของประชากร</u>

$_1\bar{X}_p$ = Mean ของการสำรวจจำนวน Population ในรอบที่ 1 = 132 คนต่อเขต

$_2\bar{X}_p$ = Mean ของการสำรวจจำนวน Population ในรอบที่ 2 = 131 คนต่อเขต

$$S_{_1x_p} = \sqrt{\sum_{i=1}^{n'} ({}_1x_{ip} - {}_1\bar{x}_p)^2 / n'-1} = \sqrt{\frac{170,915}{116}} = 38.38$$

$$S_{_2x_p} = \sqrt{\sum_{i=1}^{n'} ({}_2x_{ip} - {}_2\bar{x}_p)^2 / n'-1} = \sqrt{\frac{161,110}{116}} = 37.27$$

Sample Correlation Coefficient $r_p = \sum_{i=1}^{n'} \dfrac{({}_1x_{ip} - {}_1\bar{x}_p)({}_2x_{ip} - {}_2\bar{x}_p)}{(n'-1)\, S_{_1x_p}\, S_{_2x_p}} = \dfrac{149,950}{165,929} = 0.90$

$$V_p(\bar{X}_2) = \frac{1,388.88}{234} \times \left[\frac{1-\frac{1}{2}(0.9)^2}{1-\frac{1}{4}(0.9)^2}\right]$$

$$= 5.94\left(\frac{2.366}{3.183}\right)$$

$$= 4.42 < 5.94$$

ข้อสังเกต ค่า Estimated Correlation Coefficient ที่คำนวนได้ในทั้ง 2 กรณีจะเห็นว่ามีค่ามากกว่า 0.9 ขึ้นไป ซึ่งในทางปฏิบัติมักจะเป็นจริงเสมอ ดังนั้นในการใช้แบบแผน Partial Replacement จึงมีประสิทธิภาพเสมอเช่นเดียวกัน