บทที่ ๓

การประมาณค่าความแปรปรวนของตัวประมาณค่าที่สำคัญ

ในบทนี้จะกล่าวถึงการประมาณค่าความแปรปรวนของตัวประมาณค่ายอดรวม โดยวิธีตรง และวิธีลัด

๓.๑ การหาสูตรประมาณ-ความแปรปรวนของตัวประมาณค่ายอดรวมโดยวิธีตรง

๓.๑.๑ สูตรการประมาณ-ความแปรปรวนของค่าประมาณของผลรวมในการเลือกตัวอย่างแบบขั้นเดียวโดยการสุ่มอย่างง่าย

สมมุติว่าประชากรทั้งหมดคือ U ประกอบด้วยสมาชิกทั้งหมด N หน่วย X เป็นลักษณะที่ต้องการศึกษาซึ่งประกอบด้วย $x_1, x_2, \ldots x_N$ ดังนั้นสูตรของผลรวม

$$T(x) = \sum_{i=1}^{N} x_i$$

ถ้าเลือกตัวอย่างจากประชากรทั้งหมดมา n หน่วย โดยให้โอกาสที่แต่ละหน่วยจะถูกเลือกเท่า ๆ กันจะได้ค่าประมาณของผลรวมเป็น

$$\hat{T}(x) = \frac{N}{n} \sum_{i=1}^{n} x_i$$

สูตรการประมาณความแปรปรวนของ $\hat{T}(x)$ จะหาได้โดย

$$V(\hat{T}(x)) = V\left[\frac{N}{n} \sum_{i=1}^{n} x_i\right]$$

$$= \frac{N^2}{n^2} V\left(\sum_{i=1}^{n} x_i\right)$$

$$= \frac{N^2}{n^2}\left[\sum_{i=1}^{n} V(x_i) + \sum_{i \neq j}^{n} cov.(x_i, x_j)\right]^{1/}$$

1/ $cov(x_i, x_j) = -\frac{\sigma^2}{N-1}$ เป็นการเลือกตัวอย่างโดยการไม่แทนที่

$= 0$ เป็นการเลือกตัวอย่างโดยการแทนที่

$$V(\hat{T}(x)) = \frac{N^2}{n^2} n\sigma^2 = \frac{N^2}{n}\sigma^2$$ (เลือกโดยการแทนที่)

หรือ $$V(\hat{T}(x)) = \left(\frac{N^2}{n}\right)\left[n\sigma^2 + n(n-1)\left(\frac{-\sigma^2}{N-1}\right)\right]$$ (เลือกโดยการไม่แทนที่)

$$= \frac{N^2}{n}\sigma^2 - \frac{N^2(n-1)}{n(N-1)}\sigma^2$$

$$= \frac{N^2\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

$$= \frac{N^2\sigma^2}{n}\left(\frac{N-1-n+1}{N-1}\right)$$

$$= \frac{N^2\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

ค่า $\frac{N-n}{N-1} < ๑$ เสมอเมื่อ $n > ๑$ ดังนั้นจึงทำให้ค่า $V(\hat{T}(x))$ ในกรณีที่ใช้แผนแบบการเลือกตัวอย่างเป็นแบบเลือกโดยการไม่แทนที่ มีค่าน้อยกว่าค่า $V(\hat{T}(x))$ ในกรณีที่เป็นแบบเลือกโดยการแทนที่ซึ่งถือว่าเป็นแผนแบบที่มีประสิทธิภาพกว่า และในทางปฏิบัติการเลือกตัวอย่างแบบเลือกโดยการแทนที่ไม่เป็นที่นิยมกันเพราะโอกาสที่แต่ละหน่วยจะถูกเลือกเป็นตัวอย่างมีได้มากกว่า ๑ ครั้ง ซึ่งอาจทำให้ข้อมูลที่ได้มาไม่มีประสิทธิภาพเท่าที่ควร เนื่องจากเกิดความคลาดเคลื่อนชนิดที่ควบคุมไม่ได้ ตัวอย่างเช่น การสำรวจรายได้รายจ่าย ต้องการทราบว่ารายได้รายจ่ายของแต่ละครัวเรือนในกรุงเทพมหานคร เผอิญตัวอย่างที่เลือกมาเกิดมีการเลือกซ้ำ ๒๐ % เมื่อพนักงานไปสัมภาษณ์ผู้ที่ถูกสัมภาษณ์อาจให้ข้อมูลที่ไม่ตรงกับความจริง ทั้งนี้เนื่องจากกลัวว่าจะถูกรัฐบาลเก็บภาษีรายได้หรือบุคคลอื่นจะรู้ข้อเท็จจริง การให้ข้อมูลซึ่งไม่ตรงกับข้อเท็จจริงซ้ำ ๆ กัน ทำให้ค่าประมาณที่ได้มีความคลาดเคลื่อนมาก

จากเหตุผลดังกล่าว สูตรที่ใช้ประมาณค่าผลรวมในการสำรวจขนาดใหญ่ จะหาเฉพาะพวกเลือกตัวอย่างโดยการแทนที่ ในการเลือกตัวอย่างแบบขั้นเดียวโดยวิธีการสุ่มอย่างง่าย การประมาณค่าความแปรปรวนทำได้ง่ายเพราะ

$$\hat{V}(\hat{T}(x)) = \frac{N^2}{n}\hat{\sigma}^2\left(\frac{N-n}{N-1}\right)$$

$$\hat{\sigma}^2 = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}$$

เมื่อพิจารณาจากสมการจะเห็นว่าการคำนวณไม่มีความยุ่งยากซับซ้อนเลย แต่ในทางปฏิบัติจริง ๆ แล้ว แผนแบบนี้ไม่นิยมใช้ในการสำรวจ เนื่องจากการสำรวจต่าง ๆ เป็นงานที่ทำในการสำรวจขนาดใหญ่ ซึ่งมีประชากรใหญ่มาก จึงนิยมใช้แผนแบบที่มีความยุ่งยากสลับซับซ้อน เช่น การสุ่มแบบ ๓ ขั้น เป็นแผนแบบซึ่งมีการเลือกตัวอย่างมากกว่า ๒ ขั้น ก่อให้เกิดความยุ่งยากสลับซับซ้อนในการประมาณผล ซึ่งจะต้องทำการประมาณหลายขั้นเท่ากับขั้นของการเลือกตัวอย่าง และความแปรปรวนของค่าประมาณก็จะมีหลายส่วนเท่ากับจำนวนขั้นของการเลือกตัวอย่างและประมาณผลด้วย เพื่อช่วยลดความยุ่งยากในการประมาณผล ควรใช้วิธีการประมาณผลที่มีขั้นตอนน้อย แผนแบบในการเลือกตัวอย่างควรเป็นแบบข้อมูลทุก ๆ ตัวจะถูกเลือกเป็นตัวอย่างเท่า ๆ กัน หรือโอกาสที่จะถูกเลือกเป็นตัวอย่างต่างกันน้อยสุด จะเป็นการสะดวก รวดเร็ว มีความยุ่งยากน้อยลงในการขั้นประมาณผล การประมาณความแปรปรวน จะทำให้ได้ผลรวดเร็ว เพื่อนำไปใช้ประโยชน์ในด้านอื่น ๆ ดังนั้นจึงต้องการศึกษาเปรียบเทียบการประมาณค่าความแปรปรวนของพารามิเตอร์ ผลรวม ($T(x)$) ด้วยวิธีทั่ว ๆ ไปกับวิธีลัด ว่าวิธีใดจะมีประโยชน์ต่อการปฏิบัติงานจริง ๆ อย่างไรบ้าง

ก่อนที่จะหาสูตรในการประมาณค่าความแปรปรวนของผลรวมในการเลือกตัวอย่างแบบหลายขั้น ซึ่งแผนแบบมีความยุ่งยากในการคำนวณ จะหาสูตรของความแปรปรวนในการเลือกตัวอย่างแบบสองขั้นซึ่งเป็นแผนแบบในการเลือกตัวอย่างที่ยุ่งยากกว่าการเลือกตัวอย่างแบบขั้นเดียว

๓.๑.๒ สูตรการประมาณความแปรปรวนของค่าประมาณของผลรวมในการเลือกตัวอย่างแบบสองขั้นโดยวิธีการสุ่มอย่างง่าย

ให้ x_{ij} เป็นลักษณะที่สนใจจะศึกษาของการเลือกตัวอย่างขั้นที่สอง หน่วยที่ j ($j = 1, 2, \ldots M_i$) ในขั้นที่หนึ่งหน่วยที่ i ($i = 1, 2, \ldots N$) ถ้าเลือกตัวอย่างแบบการสุ่มอย่างง่ายมา n หน่วยจากขั้นแรก และภายในแต่ละหน่วยตัวอย่างขั้นแรกเลือก m_i หน่วยจาก M_i หน่วยในขั้นที่สอง

$$T(x) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} x_{ij}$$

$$\hat{T}(x) = \frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij}$$

. $\hat{T}(x)$ เป็นตัวประมาณค่าที่ดีของ $T(x)$ ซึ่งสามารถจะแสดงให้เห็นได้ว่า $\hat{T}(x)$ เป็น unbiased estimator คือ $E(\hat{T}(x)) = T(x)$

ในการแสดงว่า $\hat{T}(x)$ เป็น unbiased estimator ของ $T(x)$ อาจพิสูจน์โดยอาศัยทฤษฎีเกี่ยวกับ Conditional expected value

จาก $E(X) = E(E(X/Y))$

ถ้าประยุกต์ ทฤษฎี $E(x) = E[E(x/y)]$ โดยให้ T แทน x และให้ y แทนด้วย s ซึ่งเป็นตัวอย่างขั้นแรก จะพิสูจน์ได้ว่า $E(\hat{T}(x)) = T(x)$

<u>พิสูจน์</u> จากทฤษฎี Conditional expected value จะได้

$$E(\hat{T}) = E[E(\hat{T}/s)]$$

$$E(\hat{T}/s) = E\left[\left(\frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij}\right)/s\right]$$

$$= \frac{N}{n} \sum_{i=1}^{n} E\left(\frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij}\right)$$

$\because$ สำหรับตัวอย่างขั้นแรกสิ่งที่เป็นตัวแปร คือ $\frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij}$

$$= \frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} E(x_{ij})$$

$$= \frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i}\, m_i \sum_{j=1}^{M_i} x_{ij} \cdot \frac{1}{M_i}$$

$$= \frac{N}{n} \sum_{i=1}^{n} \sum_{j=1}^{M_i} x_{ij}$$

$$E[E(\hat{T}/s)] = E\left[\frac{N}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{M_i} x_{ij}\right)\right]$$

$$= \frac{N}{n}\sum_{i=1}^{n} E\left(\sum_{j=1}^{M_i} x_{ij}\right)$$

$$= \frac{N}{n}\cdot n \sum_{i=1}^{N} \cdot \frac{1}{N}\sum_{j=1}^{M_i} x_{ij}$$

$$= \sum_{i}^{N}\sum_{j}^{M_i} x_{ij}$$

$$= T(x)$$

$\therefore$ จะได้ $E(\hat{T}(x)) = T(x)$

$\Rightarrow$ $\hat{T}(x)$ เป็น unbiased estimator ของ $T(x)$

ในการหาสูตรความแปรปรวนในการเลือกตัวอย่างแบบหลายขั้นต้องอาศัยทฤษฎี Conditional Variance คือ $V(X) = V[E(X/Y)] + E[V(X/Y)]$ ①

เมื่อประยุกต์ ① กับการหา $V(\hat{T}(x))$ อาจให้ $\hat{T}$ แทนค่า x

และให้ s = ตัวอย่างขั้นแรก แทนค่า y

$\therefore$ จะได้ $V(\hat{T}) = V[E(\hat{T}/s)] + E[V(\hat{T}/s)]$

จากที่พิสูจน์มาแล้วได้ $E(\hat{T}/s) = \frac{N}{n}\sum_{i=1}^{n}\sum_{j=1}^{M_i} x_{ij}$

$$= \frac{N}{n}\sum_{i=1}^{n} x_{i.} \quad ; \quad x_{i.} = \sum_{j=1}^{M_i} x_{ij}$$

$$V(E(\hat{T}/s)) = V\left(\frac{N}{n}\sum_{i=1}^{n} x_{i.}\right)$$

$$= \left(\frac{N}{n}\right)^2 V\left(\sum_{i=1}^{n} x_{i.}\right)$$

$$V(E(\hat{T}/s)) = \frac{N^2}{n^2}\left[\sum_{i=1}^{n} V(X_i) + \sum_{i\neq j=1}^{n} cov(X_i, X_j)\right]$$

$$= \frac{N^2}{n}\sigma_b^2\left(\frac{N-n}{N-1}\right) + 0 \quad ; \quad \sigma_b^2 = \sum_{i=1}^{N}\frac{(X_i - \bar{\bar{X}})^2}{N} \quad ①$$

$$V(\hat{T}/s) = V\left[\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i} x_{ij}/s\right]$$

$$= V\left[\frac{N}{n}\sum_{i=1}^{n} M_i\bar{x}_i\right]$$

$$= \frac{N^2}{n^2}\sum_{i=1}^{n} V(M_i\bar{x}_i)$$

$$= \frac{N^2}{n^2}\sum_{i=1}^{n} M_i^2\, V(\bar{x}_i)$$

$$= \frac{N^2}{n^2}\sum_{i=1}^{n} M_i^2\frac{\sigma_i^2}{m_i}\left(\frac{M_i - m_i}{M_i - 1}\right) \quad ; \quad \sigma_i^2 = \sum_{j=1}^{M_i}\frac{(X_{ij} - \bar{\bar{X}}_i)^2}{M_i}$$

$$E\left[V(\hat{T}/s)\right] = E\left[\frac{N^2}{n^2}\sum_{i=1}^{n}\frac{M_i^2}{m_i}\sigma_i^2\left(\frac{M_i - m_i}{M_i - 1}\right)\right]$$

$$= \frac{N^2}{n^2}\cdot n\, E\left[\frac{M_i^2}{m_i}\sigma_i^2\left(\frac{M_i - m_i}{M_i - 1}\right)\right]$$

$$= \frac{N^2}{n}\sum_{i=1}^{N}\frac{M_i^2}{m_i}\sigma_i^2\,\frac{1}{N}\left(\frac{M_i - m_i}{M_i - 1}\right)$$

$$= \frac{N}{n}\sum_{i=1}^{N}\frac{M_i^2}{m_i}\sigma_i^2\left(\frac{M_i - m_i}{M_i - 1}\right) \quad ②$$

$$\therefore V(\hat{T}(x)) = ① + ②$$

$$= N^2\frac{\sigma_b^2}{n}\left(\frac{N-n}{N-1}\right)^{๒/} + \frac{N}{n}\sum_{i=1}^{N}\frac{M_i^2}{m_i}\sigma_i^2\left(\frac{M_i - m_i}{M_i - 1}\right)^{๓/} \quad ③$$

๒/ σ_b^2 = ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นแรก

๓/ σ_i^2 = ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่สองในหน่วยตัวอย่างขั้นที่ i

จากสมการที่ ๓ นี้จะเห็นว่า $V(\hat{T}(X))$ ประกอบขึ้นด้วย ๒ ส่วน คือ σ_b^2 เป็นการหาการกระจายระหว่าง $x_{i.}$ กับ $\bar{\bar{x}}$ ส่วน σ_i^2 เป็นการหาความแตกต่างของ x_{ij} กับ $\bar{x}_i$ หรือ $V(\hat{T}(X))$ = ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่ ๑
+ ความแปรปรวนภายในหน่วยตัวอย่างขั้นที่ ๒

ในเทอมต่าง ๆ ของ ⑤ เมื่อหน่วยตัวอย่างในขั้นที่ ๒ (ระบบ) ถูกเลือกมาทั้งหมดคือ $M_i = m_i$ ค่า $V(\hat{T}(X))$ จะเป็น

$$V(\hat{T}(X)) = \frac{N^2\sigma_b^2}{n}\left(\frac{N-n}{N-1}\right) + 0$$

ในทางกลับกันถ้าหน่วยตัวอย่างในขั้นที่ ๒ (ระบบ) ถูกเลือกจาก N หน่วย ทั้งหมด คือ $N = n$ จะได้ $V(\hat{T}(X))$ คือ

$$V(\hat{T}(X)) = 0 + \frac{N}{n}\sum_{i=1}^{N}\frac{M_i^2}{m_i}\sigma_i^2\left(\frac{M_i - m_i}{M_i - 1}\right)$$

$$= \sum_{i=1}^{N}\frac{M_i^2}{m_i}\sigma_i^2\left(\frac{M_i - m_i}{M_i - 1}\right)$$

เมื่อความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่ ๑ มีค่า = 0 $V(\hat{T}(X))$ ที่เหลือจะมีค่าเท่ากับ $V(\hat{T}(X))$ ในแผนแบบซึ่งมีการแบ่งเป็นสตราตาและเลือกตัวอย่างโดยการสุ่มอย่างง่าย

จะเห็นได้ว่าความคลาดเคลื่อนที่เกิดขึ้น ๒ ค่านี้ไม่ขึ้นแก่กัน ดังนั้นความแปรปรวนของผลรวมจึงเท่ากับผลรวมของความแปรปรวนในขั้นที่ย่อยลงไป

ในทางปฏิบัติจะไม่สามารถทราบค่าต่าง ๆ ของประชากรทั้งหมดซึ่งใหญ่มากได้ ดังนั้นจึงไม่สามารถคำนวณหาค่า $V(\hat{T}(X))$ ได้ต้องใช้ค่าประมาณของ $V(\hat{T}(X))$ $(= \hat{V}(\hat{T}(X)))$ โดยอาศัยข้อมูลจากตัวอย่าง เราจะได้ $\hat{V}(\hat{T}(X))$ โดยการใช้ค่าประมาณของ σ_b^2 และ σ_i^2 จะได้

$$\hat{V}(\hat{T}(X)) = \frac{N^2\hat{\sigma}_b^2}{n}\left(\frac{N-n}{N-1}\right) + \frac{N}{n}\sum_{i=1}^{n}\frac{M_i^2}{m_i}\hat{\sigma}_i^2\left(\frac{M_i - m_i}{M_i - 1}\right)$$

$$\hat{\sigma}_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{X}_{i.} - \hat{\bar{\bar{X}}})^2$$

$$\hat{\sigma}_i^2 = \frac{1}{m_i - 1}\sum_{j=1}^{m_i}(x_{ij} - \bar{x}_i)^2$$

๓.๑.๓ สูตรการประมาณความแปรปรวนของค่าประมาณของผลรวมในการเลือกตัวอย่างแบบสามขั้นโดยการสุ่มอย่างง่าย

เป็นการเลือกตัวอย่างแบบ ๓ ขั้น คือ ขั้นที่หนึ่งจะเลือกหน่วยใหญ่ที่สุด แล้วเลือกหน่วยที่รองลงไปเป็นขั้นที่สอง ในขั้นที่สามเลือกหน่วยเล็กสุด ในการเลือกตัวอย่างโดยใช้วิธีการสุ่มอย่างง่ายแบบ ๓ ขั้นย่อมมีความยุ่งยากสลับซับซ้อนมากขึ้นในการประมาณผล และสูตรที่ใช้ประมาณความแปรปรวนก็มีลักษณะคล้ายสูตรหาความแปรปรวนโดยการสุ่มอย่างง่ายแบบ ๒ ขั้น

การหาสูตรความแปรปรวนโดยวิธีการสุ่มอย่างง่ายแบบ ๓ ขั้น ก็หาในลักษณะเดียวกับสูตรความแปรปรวนโดยวิธีการสุ่มอย่างง่ายแบบ ๒ ขั้น คือ

ระหว่างหน่วยตัวอย่างขั้นที่ ๑ กับหน่วยตัวอย่างขั้นที่ ๒ ของการประมาณความแปรปรวนของผลรวมโดยวิธีการสุ่มอย่างง่ายแบบ ๒ ขั้น

จะได้ $V(\hat{T}(x))$ = (ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่ ๑) + (ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่ ๒)

สำหรับสูตรหาความแปรปรวนโดยวิธีการสุ่มอย่างง่ายแบบ ๓ ขั้น อาจพิจารณาคล้ายในลักษณะการหาความแปรปรวนโดยวิธีการสุ่มอย่างง่ายแบบ ๒ ขั้นจะได้

$V(\hat{T}(x))$ = ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่ ๑ + ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่ ๒ ภายในหน่วยตัวอย่างขั้นที่ ๑ + ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่ ๓ ภายในหน่วยตัวอย่างขั้นที่ ๒ ภายในหน่วยตัวอย่างขั้นที่ ๑

สมมุติว่าในหน่วยตัวอย่างขั้นที่ ๑ ประกอบด้วยสมาชิกทั้งหมด N หน่วย เลือกตัวอย่างมา n หน่วยจากแต่ละ n หน่วยตัวอย่างซึ่งประกอบด้วย M_i หน่วย เลือกตัวอย่างมา m_i หน่วยเป็นหน่วยตัวอย่างขั้นที่ ๒ ในแต่ละ m_i หน่วยตัวอย่างเลือกตัวอย่างมา n_{ij} หน่วยจากทั้งหมด N_{ij} หน่วยเป็นหน่วยตัวอย่างขั้นที่สาม

x_{ijk} เป็นลักษณะที่สนใจจะศึกษาจากหน่วยตัวอย่างชั้นที่สามที่ k ในหน่วยตัวอย่างที่สองที่ j ในหน่วยตัวอย่างขั้นที่ ๑ ที่ i

$$\therefore \quad T(x) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} x_{ijk}$$

$$\hat{T}(x) = \frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}$$

จาก ⑤ $V(\hat{T}(x))$ โดยวิธีการสุ่มอย่างง่ายแบบ ๒ ขั้นระหว่างหน่วยตัวอย่างขั้นที่ ๑ และหน่วยตัวอย่างขั้นที่ ๒ ได้

$$\frac{N^2}{n} \sigma_b^2 \left(\frac{N-n}{N-1}\right) + \frac{N}{n} \sum_{i=1}^{N} M_i^2 \left(\frac{M_i - m_i}{M_i - 1}\right) \frac{\sigma_i^2}{m_i} \qquad ①$$

$$\sigma_b^2 = \frac{1}{N} \sum_{i=1}^{N} (X_{i.} - \bar{\bar{X}})^2$$

$$\sigma_i^2 = \frac{1}{M_i} \sum_{j=1}^{M_i} (X_{ij} - \bar{\bar{X}}_i)^2$$

ถ้าประยุกต์ค่า $V(\hat{T}(x))$ โดยวิธีการสุ่มอย่างง่ายแบบ ๒ ขั้นกับวิธีการสุ่มอย่างง่ายแบบ ๓ ขั้นระหว่างหน่วยตัวอย่างขั้นที่ ๒ และหน่วยตัวอย่างขั้นที่ ๓ ได้

$$\frac{M_i^2}{m_i} \left(\frac{M_i - m_i}{M_i - 1}\right) \sigma_i^2 + \frac{M_i}{m_i} \sum_{j=1}^{M_i} N_{ij}^2 \left(\frac{N_{ij} - n_{ij}}{N_{ij} - 1}\right) \frac{\sigma_{ij}^2}{n_{ij}} \qquad ②$$

$$\sigma_{ij}^2 = \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} (x_{ijk} - \bar{\bar{x}}_{ij})^2$$

จากการรวมสมการ ① และ ② $V(\hat{T}(x))$ โดยวิธีการสุ่มอย่างง่ายแบบ ๓ ขั้นจะได้

$$V(\hat{T}(x)) = N^2 \left(\frac{N-n}{N-1}\right) \frac{\sigma_b^2}{n} + \frac{N}{n} \sum_{i=1}^{N} \left[M_i^2 \left(\frac{M_i - m_i}{M_i - 1}\right) \frac{\sigma_i^2}{m_i} + \frac{M_i}{m_i} \sum_{j=1}^{M_i} N_{ij}^2 \left(\frac{N_{ij} - n_{ij}}{N_{ij} - 1}\right) \frac{\sigma_{ij}^2}{n_{ij}} \right]$$

$$V(\hat{T}(x)) = N^2 \left(\frac{N-n}{N-1}\right) \frac{\sigma_b^2}{n} + \frac{N}{n} \sum_{i=1}^{N} M_i^2 \left(\frac{M_i - m_i}{M_i - 1}\right) \frac{\sigma_i^2}{m_i} + \frac{N}{n} \sum_{i=1}^{N} \frac{M_i}{m_i} \sum_{j=1}^{M_i} N_{ij}^2 \left(\frac{N_{ij} - n_{ij}}{N_{ij} - 1}\right) \frac{\sigma_{ij}^2}{n_{ij}}$$

ในทางปฏิบัติจะไม่สามารถหาค่า $V(\hat{T}(x))$ ดังนั้นจึงต้องประมาณค่าความแปรปรวนของค่าประมาณผลรวมด้วย $\hat{V}(\hat{T}(x))$

$$\hat{V}(\hat{T}(x)) = N^2\left(\frac{N-n}{N-1}\right)\frac{\hat{\sigma}_b^2}{n} + \frac{N}{n}\sum_{i=1}^{n} M_i^2\left(\frac{M_i-m_i}{M_i-1}\right)\frac{\hat{\sigma}_i^2}{m_i} + \frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i} N_{ij}^2\left(\frac{N_{ij}-n_{ij}}{N_{ij}-1}\right)$$

$$\hat{\sigma}_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{X}_{i\cdot}-\hat{\hat{\bar{X}}})^2$$

$$\hat{\sigma}_i^2 = \frac{1}{m_i-1}\sum_{j=1}^{m_i}(\hat{X}_{ij}-\hat{\bar{X}}_i)^2$$

$$\hat{\sigma}_{ij}^2 = \frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}}(X_{ijk}-\hat{\bar{X}}_{ij})^2$$

ซึ่งค่าความแปรปรวนโดยการสุ่มอย่างง่ายแบบ ๓ ขั้น คือ

$V(\hat{T}(x))$ = ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นแรก + ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่ ๒ ภายในหน่วยตัวอย่างขั้นแรก + ความแปรปรวนระหว่างหน่วยตัวอย่างขั้นที่สามภายในหน่วยตัวอย่างขั้นที่สอง

จะเห็นได้ว่าแผนแบบในการสุ่มแบบหลายขั้นนั้นไม่จำเป็นจะต้องมี สามขั้นอาจจะเป็นสี่หรือห้าขั้นก็ได้ แต่ตามหลักแล้วไม่ควรจะใช้แผนแบบที่มีหลายขั้นเกินไปเพราะจะให้ค่าความคลาดเคลื่อนในการประมาณสูง และก่อให้เกิดความยุ่งยากในการประมาณผลและการคำนวณค่าความแปรปรวนก็มีความยุ่งยากสลับซับซ้อนมาก

โดยทั่วไปแผนแบบการเลือกตัวอย่างแบบการจัดสตราตาก่อน เป็นแผนแบบชนิดหนึ่งที่ช่วยลดความคลาดเคลื่อนในการเลือกตัวอย่างได้ ดังนั้นในการสำรวจสถิติระดับใหญ่ผู้วางแผนจึงนิยมใช้ร่วมกับแผนแบบอื่น ๆ เช่น อาจจะเป็นแบบการจัดสตราตาของการเลือกตัวอย่างแบบสองขั้นโดยวิธีการสุ่มอย่างง่าย เช่นในงานสำรวจแรงงาน แผนแบบที่ใช้เป็นแบบการจัดสตราตาของการ

เลือกตัวอย่างแบบหลายขั้นและโอกาสที่แต่ละหน่วยถูกเลือกไม่เท่ากัน กล่าวคือ เราสามารถจะประมาณผลในระดับสตราตาก่อนแล้วหาผลรวมของทุก ๆ สตราตาซึ่งก็เป็นค่าทั้งหมดที่ต้องการประมาณ เช่น ถ้าแบบแผนการเลือกตัวอย่างเป็นแบบการจัดสตราตาของการเลือกตัวอย่างแบบสองขั้นโดยวิธีการสุ่มอย่างง่ายซึ่งสูตรที่ใช้ประมาณผลรวมคือ

$$\hat{T}_h(x) = \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} x_{hij}$$

$$v(\hat{T}_h(x)) = \sum_{h=1}^{L} \left[\frac{N_h^2}{n_h} \sigma_{hb}^2 \left(\frac{N_h - n_h}{N_h - 1} \right) + \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}^2}{m_{hi}} \sigma_{hi}^2 \left(\frac{M_{hi} - m_{hi}}{M_{hi} - 1} \right) \right]$$

ในเมื่อ $\hat{T}_h(x)$ = ค่าประมาณผลรวมของ x ในสตราตาที่ h, h=1,2,........L

<u>แผนแบบการเลือกตัวอย่างโดยให้โอกาสที่แต่ละหน่วยจะถูกเลือกไม่เท่ากัน</u> (Unequal Probability Sampling)

เป็นวิธีการเลือกตัวอย่างชนิดหนึ่งโดยหน่วยต่าง ๆ มีโอกาสที่จะถูกเลือกไม่เท่ากันซึ่งขึ้นกับขนาดของหน่วยนั้น ๆ เช่น ในการสำรวจผลิตผลของเกษตรในหมู่บ้านซึ่งมีขนาดใหญ่มากและมีประชากรอาศัยอยู่มากในแต่ละหมู่บ้าน ดังนั้นในการที่จะประมาณผลรวมของผลิตผลที่ผลิตได้ การเลือกตัวอย่างจะเลือกหมู่บ้านโดยให้โอกาสที่แต่ละหมู่บ้านจะถูกเลือกขึ้นกับขนาดของประชากรในแต่ละหมู่บ้านนั้นหรือขึ้นกับพื้นที่ก็ได้ วิธีการเลือกตัวอย่างแบบนี้เรา เรียกว่าวิธีการเลือกตัวอย่างซึ่งขึ้นกับขนาดของประชากร (Sampling With Probability Propertional to Size หรือเรียกย่อ ๆ ว่า PPS. Sampling)

จากหลักทฤษฎีของสถิติวิธีการเลือกตัวอย่างแบบนี้จะช่วยทำให้การประมาณผลลดค่าความแปรปรวนหรือความคลาดเคลื่อนของค่าประมาณได้โดยไม่ต้องเพิ่มขนาดตัวอย่างเป็นการยุ่งยากและสิ้นเปลืองค่าใช้จ่ายด้วย เป็นวิธีการที่มีประสิทธิภาพสูงกว่าแบบที่ทุกหน่วยมีโอกาสจะถูกเลือกเท่า ๆ กันแต่เป็นสิ่งที่จำเป็นที่สุดที่ผู้วางแผนการสำรวจและผู้ออกแบบวิธีการเลือกตัวอย่างต้องกำหนด selection probability ให้เหมาะสม วิธีที่ดีคือใช้แบบ PPS. เพราะถ้าสามารถกำหนดให้ P_i เป็นปฏิภาค X_i ได้มากเท่าไรความแปรปรวนของค่าประมาณก็ยิ่งมีขนาดเล็กลงได้มากเท่านั้น ทำให้ค่าประมาณมีความเชื่อถือได้มากขึ้น

ได้ทราบแล้วว่าแผนแบบที่มีการจัดสตราตาก่อนเป็นแผนแบบการเลือกตัวอย่างที่ดีอันหนึ่ง ซึ่งสามารถช่วยลดค่าแปรปรวนหรือความคลาดเคลื่อนของค่าประมาณเมื่อเรานำเอาแผนแบบที่มีการจัดสตราตาก่อนกับแผนแบบที่ให้โอกาสของแต่ละหน่วยที่จะถูกเลือกไม่เท่ากันมาใช้ด้วยกัน จะเป็นวิธีการที่ดีที่จะทำให้ค่าประมาณของพารามิเตอร์ที่น่าสนใจ มีความเชื่อถือได้มากขึ้น

๓.๑.๔ <u>สูตรการประมาณความแปรปรวนของค่าประมาณของผลรวมในการเลือกตัวอย่างแบบสองขั้นโดยความน่าจะเป็นที่แต่ละหน่วยจะถูกเลือกไม่เท่ากัน</u>

ให้ x_{ij} เป็นลักษณะที่สนใจจะศึกษาของขั้นที่สองหน่วยที่ j, $j=1,2,\ldots,M_i$ ในขั้นแรกหน่วยที่ i; $i=1,2,\ldots,N$ ถ้าเลือกตัวอย่างมา n หน่วยในขั้นแรกโดยให้ความน่าจะเป็นที่จะถูกเลือก P_i และในขั้นที่สองเลือกหน่วยตัวอย่างมา m_i หน่วยจากในแต่ละหน่วยตัวอย่างของขั้นแรกจะได้

$$T(X) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} X_{ij}$$

$$\hat{T}(X) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{P_i} \left(\frac{M_i}{m_i} \sum_{j=1}^{m_i} X_{ij} \right)$$

ซึ่ง $\hat{T}(x)$ เป็นตัวประมาณค่าที่ดีของ $T(x)$ และสามารถพิสูจน์ได้ว่าเป็น unbiased estimater ของ $T(x)$ การพิสูจน์เช่นเดียวกันกับแผนแบบของการเลือกตัวอย่างที่โอกาสทุกๆ หน่วยจะถูกเลือกเท่ากันโดยอาศัย conditional expected value $E(\hat{T})=E\left[E(\hat{T}/S)\right]$

<u>พิสูจน์</u> $$E(\hat{T}/s) = E\left[\frac{1}{n} \sum_{i=1}^{n} \frac{1}{P_i} \frac{M_i}{m_i} \sum_{j=1}^{m_i} X_{ij} \,\middle|\, s \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{P_i} \frac{M_i}{m_i} E\left(\sum_{j=1}^{m_i} X_{ij} \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{P_i} \frac{M_i}{m_i} \cdot m_i \sum_{j=1}^{M_i} X_{ij} \cdot \frac{1}{M_i}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{P_i} \sum_{j=1}^{M_i} X_{ij}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{P_i} X_{i.}$$

$$E\left[E(\hat{T}/s)\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\sum_{j=1}^{M_i} X_{ij}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{p_i}\sum_{j=1}^{M_i} X_{ij}\right)$$

$$= \frac{1}{n}\cdot n\sum_{i=1}^{N} p_i\cdot\frac{1}{p_i}\sum_{j=1}^{M_i} X_{ij}$$

$$= \sum_{i}^{N}\sum_{j}^{M_i} X_{ij} \qquad = T(X)$$

แต่ $E(\hat{T}) = E\left[E(\hat{T}/s)\right]$

$$= T(X)$$

$\therefore E(\hat{T}(X)) = T(X)$

$\Rightarrow$ แสดงว่า $\hat{T}(X)$ เป็น unbiased estimator ของ $T(X)$ โดยอาศัยทฤษฎี

Conditional Variance ได้

$$V[\hat{T}(X)] = V\left[E(\hat{T}/s)\right] + E\left[V(\hat{T}/s)\right]$$

$$V\left[E(\hat{T}/s)\right] = V\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}X_{i.}\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V\left(\frac{X_{i.}}{p_i}\right)$$

$$= \frac{1}{n^2}\cdot n\left[E\left(\frac{X_{i.}}{p_i} - E\left(\frac{X_{i.}}{p_i}\right)\right)^2\right]$$

แต่ $E\left(\frac{X_{i.}}{p_i}\right) = \sum_{i=1}^{N}\frac{p_i X_{i.}}{p_{i.}} = \sum_{i=1}^{N} X_{i.} = T$

$$V\left[E(\hat{T}/s)\right] = \frac{1}{n}\left[E\left(\frac{X_{i.}}{p_i} - T\right)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} p_i\left(\frac{X_{i.}}{p_i} - T\right)^2 \qquad 1'$$

$$E\left[V(\hat{T}/s)\right] = E\left[V\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\frac{M_i}{m_i}\sum_{j=1}^{m_i}x_{ij}\,\Big|\,s\right)\right]$$

$$= E\left[\frac{1}{n^2}\sum_{i=1}^{n}V\left(\frac{1}{p_i}\frac{M_i}{m_i}\sum_{j=1}^{m_i}x_{ij}\right)\right]$$

$$= E\left[\frac{1}{n^2}\cdot n\left(\frac{M_i}{p_i}\right)^2 V(\bar{x}_i)\right]$$

$$= E\left[\frac{1}{n}\left(\frac{M_i}{p_i}\right)^2\frac{\sigma_i^2}{m_i}\left(\frac{M_i-m_i}{M_i-1}\right)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{M_i}{p_i}\right)^2\frac{\sigma_i^2}{m_i}\left(\frac{M_i-m_i}{M_i-1}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\frac{\sigma_i^2}{m_i}\left(\frac{M_i-m_i}{M_i-1}\right) \qquad (2)$$

$$V(\hat{T}(x)) = (1) + (2)$$

$$V(\hat{T}(x)) = \frac{1}{n}\sum_{i=1}^{N}p_i\left(\frac{X_{i.}}{p_i}-T\right)^2 +$$

$$\frac{1}{n}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\frac{\sigma_i^2}{m_i}\left(\frac{M_i-m_i}{M_i-1}\right)$$

ซึ่ง $\sigma_i^2 = \sum_{j=1}^{M_i}\frac{(X_{ij}-\bar{\bar{X}}_i)^2}{M_i}$

แต่ในทางปฏิบัติความแปรปรวนตัวนี้ไม่สามารถทราบค่าได้ เนื่องจากเราไม่สามารถทราบค่าต่าง ๆ ของประชากรทั้งหมดได้ ดังนั้นค่า σ_i^2 จะไม่สามารถทราบค่าได้จึงต้องใช้สูตรการประมาณค่าความแปรปรวนของ $\hat{T}(x)$

๓.๑.๕ สูตรการประมาณความแปรปรวนของค่าประมาณของผลรวมในการเลือกตัวอย่างแบบสามขั้นโดยความน่าจะเป็นที่แต่ละหน่วยจะถูกเลือกไม่เท่ากัน

การหาสูตรก็มีลักษณะคล้ายกับสูตรการประมาณความแปรปรวนของการประมาณค่าผลรวมโดยโอกาสที่จะถูกเลือกเท่า ๆ กัน

สมมุติว่าในขั้นที่หนึ่งมี N หน่วยเลือกตัวอย่างมา n หน่วย จากแต่ละ n หน่วยตัวอย่างซึ่งประกอบด้วยสมาชิก M_i หน่วยเลือกตัวอย่างมา m_i หน่วยเป็นขั้นที่สอง จากแต่ละ m_i หน่วยตัวอย่างซึ่งประกอบด้วยสมาชิก N_{ij} เลือกตัวอย่างมา n_{ij} เป็นขั้นที่สามโดยในขั้นแรกและขั้นสองโอกาสที่แต่ละหน่วยจะถูกเลือกไม่เท่ากัน ส่วนขั้นที่สามเลือกโดยวิธีการสุ่มอย่างง่าย

ให้ X_{ijk} = ลักษณะที่สนใจจะศึกษาของขั้นที่สามหน่วยที่ k ในขั้นที่สองของหน่วยที่ j ในขั้นแรกของหน่วยที่ i

จะได้

$$T(X) = \sum_{i=1}^{N}\sum_{j=1}^{M_i}\sum_{k=1}^{N_{ij}} X_{ijk}$$

$$\hat{T}(X) = \frac{N}{n}\sum_{i=1}^{n}\frac{1}{P_i}\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{1}{P_{ij}}\frac{N_{ij}}{n_{ij}}\sum_{k=1}^{n_{ij}} X_{ijk}$$

จากค่าความแปรปรวนของค่าประมาณของผลรวมในการเลือกตัวอย่างแบบสองขั้นโดยที่แต่ละหน่วยจะถูกเลือกไม่เท่ากันเมื่อนำมาประยุกต์กับการเลือกตัวอย่างแบบสามขั้นระหว่างหน่วยตัวอย่างขั้นที่ ๑ และที่ ๒ จะได้

$$\frac{1}{n}\sum_{i=1}^{N} P_i\left(\frac{X_{i.}'}{P_i} - T\right)^2 + \frac{1}{n}\sum_{i=1}^{N}\frac{M_i^2}{P_i}\frac{\sigma_i^2}{m_i}\left(\frac{M_i - m_i}{M_i - 1}\right) \qquad (1)$$

ประยุกต์สูตรความแปรปรวนของการเลือกตัวอย่างแบบสองขั้นโดยโอกาสที่แต่ละหน่วยจะถูกเลือกไม่เท่ากันกับการเลือกตัวอย่างแบบสามขั้นและโอกาสของแต่ละหน่วยจะถูกเลือกไม่เท่ากันระหว่างหน่วยตัวอย่างขั้นที่ ๒ กับขั้นที่ ๓ จะได้

$$\frac{M_i^2}{P_i}\frac{\sigma_i^2}{m_i}\left(\frac{M_i - m_i}{M_i - 1}\right) + \frac{1}{m_i}\sum_{j=1}^{M_i}\frac{N_{ij}^2}{P_{ij}}\frac{\sigma_{ij}^2}{n_{ij}}\left(\frac{N_{ij} - n_{ij}}{N_{ij} - 1}\right) \qquad (2)$$

ดังนั้นสูตรหาความแปรปรวนของการเลือกตัวอย่างแบบสามขั้นโดยโอกาสที่แต่ละหน่วยจะถูกเลือกไม่เท่ากันเช่นเดียวกับในสูตรของการหาความแปรปรวนโดยการสุ่มอย่างง่ายแบบสามขั้น

ดังนั้น $$V(\hat{T}(x)) = \frac{1}{n}\sum_{i=1}^{N} P_i\left(\frac{X_{i.}}{P_i} - T\right)^2 + \frac{1}{n}\sum_{i=1}^{N}\left[\frac{M_i^2}{P_i}\frac{\sigma_i^2}{m_i}\left(\frac{M_i - m_i}{M_i - 1}\right) + \frac{1}{m_i}\sum_{j=1}^{M_i}\frac{N_{ij}^2}{P_{ij}}\frac{\sigma_{ij}^2}{n_{ij}}\left(\frac{N_{ij} - n_{ij}}{N_{ij} - 1}\right)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} P_i\left(\frac{X_{i.}}{P_i} - T\right)^2 + \frac{1}{n}\sum_{i=1}^{N}\frac{M_i^2}{P_i}\frac{\sigma_i^2}{m_i}\left(\frac{M_i - m_i}{M_i - 1}\right) + \frac{1}{n}\sum_{i=1}^{N}\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{N_{ij}^2}{P_{ij}}\frac{\sigma_{ij}^2}{n_{ij}}\left(\frac{N_{ij} - n_{ij}}{N_{ij} - 1}\right)$$

การประมาณค่าความแปรปรวนของยอดรวมนี้เป็นการประมาณค่าแบ่งเป็นส่วน ๆ (Compenent) ถ้าไม่ต้องการที่จะประมาณค่าความแปรปรวนตาม compenent ก็ได้โดยหาในลักษณะรวม การใช้ก็แล้วแต่ความเหมาะสมหรือสะดวก ตัวอย่างเช่น แผนแบบแบบสองขั้นค่าประมาณของผลรวมเป็น

$$\hat{T}(X) = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{P_i m_i}\sum_{j=1}^{m_i} X_{ij}$$

$$V(\hat{T}(X)) = \frac{1}{n}\left[\sum_{i=1}^{N} P_i\left(\frac{X_{i.}}{P_i} - T\right)^2 + \sum_{i=1}\frac{M_i^2}{P_i m_i}\sigma_i^2\left(\frac{M_i - m_i}{M_i - 1}\right)\right]$$

ค่าประมาณ $V(\hat{T}(x))$ ซึ่งถ้าไม่ประมาณค่าตาม compenent จะได้

$$\hat{V}(\hat{T}(X)) = \frac{1}{n(n-1)}\left[\sum_{i=1}^{n}\frac{M_i}{P_i m_i}\sum_{j=1}^{m_i} X_{ij} - \frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{P_i m_i}\sum_{j=1}^{m_i} X_{ij}\right]^2$$

$$= \frac{1}{n(n-1)}\left[z_i - \bar{z}\right]^2$$

ในเมื่อ $$z_i = \sum_{i=1}^{n}\frac{M_i}{P_i m_i}\sum_{j=1}^{m_i} X_{ij}$$

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{P_i m_i}\sum_{j=1}^{m_i} X_{ij}$$

ส่วนสูตรที่คำนวณหาความแปรปรวนของตัวพารามิเตอร์ที่มีรูปเป็น non - linear function เช่น อัตราส่วนซึ่งรูปของตัวพารามิเตอร์นี้ก็มีลักษณะยุ่งยากแม้แต่ในแผนแบบซึ่งเป็นแบบง่ายๆ การที่จะดูประสิทธิภาพสัมพัทธ์ของค่าที่จะประมาณก็มีความยุ่งยากมากขึ้นด้วย โดยเฉพาะในการสำรวจสถิติระดับใหญ่ การประมาณค่าความแปรปรวนจะมีความสลับซับซ้อนมากขึ้น ตัวอย่างการหาความแปรปรวนของอัตราส่วนซึ่งแผนแบบเป็นแบบการสุ่มอย่างง่ายในการเลือกตัวอย่างขั้นเดียว

ให้ $R = \dfrac{\bar{\bar{y}}}{\bar{\bar{x}}}$

$$\hat{R} = \frac{\bar{y}}{\bar{x}}$$

$$= \frac{\bar{y} - \bar{\bar{y}} + \bar{\bar{y}}}{\bar{x} - \bar{\bar{x}} + \bar{\bar{x}}}$$

$$= \frac{\bar{\bar{y}}}{\bar{\bar{x}}}\left[\frac{1 + \frac{\bar{y} - \bar{\bar{y}}}{\bar{\bar{y}}}}{1 + \frac{\bar{x} - \bar{\bar{x}}}{\bar{\bar{x}}}}\right]$$

$$= \frac{\bar{\bar{y}}}{\bar{\bar{x}}}\left[\frac{1 + \Delta_y}{1 + \Delta_x}\right]$$

$$= \frac{\bar{\bar{y}}}{\bar{\bar{x}}}(1 + \Delta_y)(1 - \Delta_x + (\Delta_x)^2 - (\Delta_x)^3 + \cdots\cdots)$$

$$= R(1 - \Delta_x + (\Delta_x)^2 - (\Delta_x)^3 + \Delta_y - \Delta_x \Delta_y + \cdots\cdots)$$

$$\Rightarrow \quad \hat{R} \doteq R(1 - \Delta_x + \Delta_y)$$

$$V(\hat{R}) \doteq V[R(1 - \Delta_x + \Delta_y)]$$

$$\doteq V\left[R\left(1 - \frac{\bar{x} - \bar{\bar{x}}}{\bar{\bar{x}}} + \frac{\bar{y} - \bar{\bar{y}}}{\bar{\bar{y}}}\right)\right]$$

$$\doteq R^2\left[0 + V\left(\frac{\bar{y} - \bar{\bar{y}}}{\bar{\bar{y}}} - \frac{\bar{x} - \bar{\bar{x}}}{\bar{\bar{x}}}\right)\right]$$

$$\doteq R^2\left[\frac{V(\bar{x})}{\bar{\bar{x}}^2} + \frac{V(\bar{y})}{\bar{\bar{y}}^2} - \frac{2\,COV(\bar{x}, \bar{y})}{\bar{\bar{x}}\,\bar{\bar{y}}}\right]$$

$$\doteq R^2\left[\frac{\sigma^2_{\bar{x}}}{\bar{\bar{x}}^2} + \frac{\sigma^2_{\bar{y}}}{\bar{\bar{y}}^2} - 2\frac{\sigma_{\bar{x}\bar{y}}}{\bar{\bar{x}}\,\bar{\bar{y}}}\right]$$

$$\hat{V}(\hat{R}) \doteq \hat{R}^2\left[\frac{\hat{\sigma}^2_{\bar{x}}}{\bar{\bar{x}}^2} + \frac{\hat{\sigma}^2_{\bar{y}}}{\bar{\bar{y}}^2} - 2\frac{\hat{\sigma}_{\bar{x}\bar{y}}}{\bar{\bar{x}}\,\bar{\bar{y}}}\right]$$

จะเห็นได้ว่า $\hat{V}(\hat{R})$ ในการสำรวจด้วยตัวอย่างขั้นเดียว ก็มีความยุ่งยากเนื่องจากตัว $\hat{R}$ มีลักษณะเป็นตัวประมาณที่มีความยุ่งยาก ถ้าแผนแบบการเลือกตัวอย่างมีมากกว่าสองขั้นปัญหาของการประมาณค่าความแปรปรวนของค่าพารามิเตอร์ต่าง ๆ โดยวิธีตรงก็จะมีมากขึ้น เพื่อเป็นการประหยัดเวลา ค่าใช้จ่าย ลดความยุ่งยาก และสะดวกต่อการใช้เครื่องจักรกลคำนวณ จะใช้วิธีซึ่งเรียกว่า วิธีลัด ถึงแม้การประมาณจะให้ค่าเพียงคร่าว ๆ แต่ก็เหมาะสมกับงานที่เร่งด่วนหรือมีงบประมาณน้อย และสามารถหาพิสัยของความแปรปรวนของตัวประมาณค่าได้ซึ่งจะเป็นเครื่องตัดสินได้ว่าแผนแบบการเลือกตัวอย่างหรือสูตร ที่ใช้ควรมีการปรับปรุงแก้ไขหรือไม่

๓.๒ วิธีการประมาณค่าความแปรปรวนโดยวิธีลัด

๓.๒.๑ วิธีการจัดหมู่แบบสุ่ม[4] (Random Group)

จากประชากรทั้งหมด N หน่วย เลือกตัวอย่างมาชุดหนึ่ง n หน่วยด้วยวิธีการสุ่มอย่างง่าย สมมุติว่าเป็น $x_1, x_2 \ldots\ldots x_n$ ดังนั้นจะได้ค่าความแปรปรวนจากตัวอย่าง

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

ถ้านำตัวอย่าง n หน่วยมาแบ่งออกเป็น t หมู่โดยแต่ละหมู่ประกอบด้วยสมาชิก k ตัว ซึ่ง $k = \frac{n}{t}$ ดังตาราง

หมู่ที่	1	2		g		t
	x_{11}	x_{21}		x_{g1}		x_{t1}
	x_{12}	x_{22}		x_{g2}		x_{t2}
ผลรวม	x_1	x_2		x_g		x_t

4/ รวบรวมจากหนังสือ Hansen, Hurvitz & Madow, Sample Survey Methods and Theory (Vol. I, II) (New York: John Wiley & Sons Inc., 1953)
Des Raj, Sampling Theory (New York: McGraw-Hill Book Company, 1968)
Leslie Kish, Survey Sampling (New York: John Wiley & Sons, Inc., 1965)
Atlantida Series ISPO 1, No. I.E (V.S. Bureau of the Census, 1966)

ดังนั้นความแปรปรวนของผลรวม (s_k^2) ของ t หมู่จะเป็นค่าประมาณที่เป็น unbiased estimator ของ σ^2 นั่นคือ $E(s_k^2) = \sigma^2$ โดยที่

$$s_k^2 = \sum_{g=1}^{t} \frac{(x_g - \bar{x}')^2}{k(t-1)}$$

x_{gi} = ค่าที่สนใจจะศึกษาจากตัวอย่างที่ i ในหมู่ที่ g

x_g = ค่าของผลรวมของ k หน่วยในหมู่ที่ g

$\bar{x}'$ = ค่าเฉลี่ยของ t หมู่ ; $\bar{x}' = \sum_{g=1}^{t} x_g / t$

t = จำนวนหมู่ทั้งหมด

k = จำนวนหน่วยในแต่ละหมู่

วิธีการคือใช้แบ่งตัวอย่างที่เลือกมาเป็นหมู่แบบสุ่ม การแบ่งตัวอย่างออกเป็นหมู่ต่างๆ ทำได้หลายวิธี คือ

๑. ทำโดยวิธีจับสลาก เป็นวิธีที่ง่าย สมมุติว่ามีครัวเรือนทั้งหมด ๕๐๐ ครัวเรือน ต้องการจัดเป็น ๑๐ หมู่ แต่ละหมู่จะประกอบด้วยครัวเรือนจำนวน ๕๐ ครัวเรือน ดังนั้นจึงให้เลขที่จำนวนครัวเรือนตั้งแต่ ๑ ถึง ๕๐๐ แล้วเขียนสลากขึ้นชุดหนึ่งมีเลขที่ตั้งแต่ ๑ ถึง ๕๐๐ เช่นเดียวกันแล้วจึงจับสลากเหล่านั้นมาครั้งละ ๕๐ ชิ้น จนกว่าจะหมด ๕๐๐ ชิ้น สลาก ๕๐ ชิ้นแรกจะเป็นเลขที่ครัวเรือนของหมู่ที่หนึ่ง สลาก ๕๐ ชิ้นที่สองก็จะเป็นเลขที่ของครัวเรือนของหมู่ที่สอง เช่นนี้เรื่อย ๆ ไปจนครบ ๑๐ หมู่

๒. ทำโดยอาศัยตารางเลขสุ่มตัวอย่าง ตารางเลขสุ่มประกอบขึ้นด้วยตารางหลักเดียว สองหลัก ฯลฯ การใช้ก็ขึ้นกับตัวเลขที่เราต้องการถ้ามีสองหลักก็ใช้ตารางสองหลัก ตัวอย่างเช่นเดียวกับข้อ ๑ ดังนั้นตารางที่ใช้ควรจะใช้ตาราง ๓ หลัก ซึ่งการใช้เราจะเริ่มที่ใดของตารางก็ได้ การอ่านจะอ่านไปทางขวาหรืออ่านขึ้น หรืออ่านลงก็ได้ โดยดูทีละ ๓ หลัก ค่าที่ใช้ได้คือ ๑ ถึง ๕๐๐ ถ้าเกิน ๕๐๐ ก็ไม่ใช้ หมู่แรกมี ๕๐ ครัวเรือนจากตารางเลขสุ่มก็เลือกมา ๕๐ จำนวน เช่น ๑๒ ก็ถือว่าตรงกับครัวเรือนที่ ๑๒ หรือ ๑๘๕ ก็คือครัวเรือนที่ ๘๕ ทำเรื่อย ๆ ไปจนได้ ๕๐ จำนวนก็ถือว่าอยู่ในหมู่แรก ส่วนหมู่อื่น ๆ ก็ทำเช่นเดียวกัน

๓. ทำโดยกำหนด Running Number ของจำนวนตัวอย่างทั้งหมด เช่น มีจำนวนครัวเรือนทั้งหมด ๕๐๐๐ ครัวเรือนก็เขียนเลขกำกับตั้งแต่ ๑ ถึง ๕๐๐๐ ถ้าต้องการแบ่งเป็น ๑๐ หมู่ก็กำหนดว่าครัวเรือนที่ลงท้ายด้วยเลข ๑ ก็อยู่หมู่ที่ ๑ ลงท้ายเลข ๒ ก็อยู่หมู่ที่ ๒ เรื่อย ๆ ไป เป็นการสะดวกในการที่จะใช้เครื่องคอมพิวเตอร์ช่วยคำนวณ

โดยทั่ว ๆ ไปแต่ละหมู่ควรมีขนาดเท่า ๆ กัน การที่จะแบ่งจำนวนหมู่เป็นเท่าไรก็ไม่จำกัด ทั้งนี้ขึ้นกับขนาดของจำนวนตัวอย่างทั้งหมดที่ใช้ แต่ที่นิยมใช้กันมักแบ่งตั้งแต่ ๑๐ – ๔๐ หมู่

ในกรณีที่แผนแบบในการเลือกตัวอย่างเป็นแบบใช้วิธีการสุ่มตัวอย่างหลายขั้น การคำนวณหาความแปรปรวนของตัวประมาณค่าต่าง ๆ โดยวิธีการจัดหมู่แบบสุ่มนั้นทำได้โดยนำหน่วยตัวอย่างขั้นสุดท้ายมาจัดเป็นหมู่ตามแบบที่กล่าวมาแล้วข้างต้นแบบใดแบบหนึ่ง

ในกรณีที่มีสตราตา การจัดหมู่แบบสุ่มให้ทำในแต่ละสตราตาก่อนโดยวิธีเดียวกัน การประมาณค่าความแปรปรวนของตัวประมาณค่าอาจหาในแต่ละสตราตาได้ หรือในกรณีที่มีสตราตาเป็นจำนวนมาก การประมาณค่าความแปรปรวนอาจจะไม่ทำในแต่ละสตราตาดังกล่าวข้างต้นก็ได้ โดยเอาตัวอย่างทั้งหมดมาแบ่งเป็นหมู่ ตัวอย่างเช่น ในการสำรวจโครงการหนึ่ง แผนแบบที่ใช้แบ่งเป็น ๒๐ สตราตา แต่ละสตราตาประกอบด้วยครัวเรือนตัวอย่าง ๑๐๐ ครัวเรือน ประสงค์จะคำนวณหาค่าความแปรปรวนโดยวิธีการจัดหมู่แบบสุ่ม จึงแบ่งข้อมูลแต่ละสตราตาออกเป็น ๕๐ หมู่ แต่ละหมู่จึงประกอบด้วย ๒ ครัวเรือนจากแต่ละสตราตา ดังนั้นจะได้ครัวเรือนตัวอย่างในแต่ละหมู่โดยไม่คำนึงถึงสตราตาเป็น ๔๐ (= ๒๐ × ๒) ครัวเรือน

วิธีการจัดหมู่แบบสุ่มนี้มีข้อดีคือ สะดวก ง่าย รวดเร็ว เสียค่าใช้จ่ายน้อยในการคำนวณ แต่ค่าที่ได้เป็นเพียงค่าคร่าว ๆ ซึ่งจะให้ค่าประมาณความแปรปรวนสูงกว่าวิธีตรง และเป็นวิธีหนึ่งซึ่งเป็นที่นิยมกัน

๓.๒.๒ วิธีการสุ่มอย่างง่าย (Simple Random)[๖]

แม้แผนแบบที่ใช้จะมีความซับซ้อน เพื่อความสะดวกในการประมาณค่าความแปรปรวน อาจใช้วิธีลัดซึ่งเรียกว่าวิธีการสุ่มอย่างง่าย คือ อาจข้ามขั้นตอนต่าง ๆ ในแผนแบบการเลือกตัวอย่าง โดยใช้หน่วยตัวอย่างขั้นสุดท้ายมาใช้ประมาณ

$$\hat{T}(X) = \frac{N}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

เมื่อ N = จำนวนหน่วยตัวอย่างทั้งหมดของประชากร

n = ขนาดตัวอย่างทั้งหมด

x_i = ลักษณะที่สนใจจะศึกษาที่ i

ค่าประมาณความแปรปรวนของค่าประมาณยอดรวมเป็น

$$\hat{V}(\hat{T}(X)) = \frac{N^2}{n}\hat{\sigma}_x^2 \qquad (2)$$

$$\hat{\sigma}_x^2 = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}$$

ในกรณีที่ขนาดตัวอย่างที่ใช้ใหญ่มากเพื่อความสะดวกในการคำนวณหาค่าประมาณ อาจใช้วิธีลดขนาดตัวอย่างโดยใช้วิธีเลือกตัวอย่างมาส่วนหนึ่ง (Sub - Sample) เช่นตัวอย่างที่เลือกมาจากประชากรทั้งหมดมี n หน่วย ต้องการลดขนาดตัวอย่างเหลือเพียง m หน่วย โดยให้แต่ละหน่วยมีความน่าจะเป็นในการถูกสุ่มเท่า ๆ กัน ดังนั้นค่าประมาณที่ได้ตาม (1) จึงเปลี่ยนเป็น

$$\hat{V}(\hat{T}(X)) = \frac{N^2}{n}\sum_{i=1}^{m}\frac{(x_i-\bar{x})^2}{m-1}$$

แม้ว่าวิธีข้างต้นนี้จะให้ความสะดวกและรวดเร็วในการคำนวณ แต่ค่าที่ประมาณได้ก็เป็นไปอย่างคร่าว ๆ และได้ค่าความแปรปรวนสูงกว่าวิธีตรง

---

๖ รวบรวมมาจาก Cochran, W.G., Sampling Techniques (New York; John Wiley & Sons Inc., 1963

Leslie Kish, Survey Sampling (New York : John Wiley & Sons, Inc., 1965

วิธีลัดทั้งสองวิธีนี้จะนำไปทำการประยุกต์กับงานสำรวจแรงงานซึ่งสำนักงานสถิติแห่งชาติเป็นผู้ดำเนินการสำรวจ เนื่องจากแผนแบบของงานสำรวจแรงงานเหมาะกับวิธีการทั้งสองนี้ และต้องการเปรียบเทียบประสิทธิภาพกับวิธีตรงเพื่อจะหาวิธีลัดที่เหมาะกับงานสำรวจแรงงาน นอกจากวิธีลัดทั้งสองวิธีนี้แล้ว ยังมีวิธีลัดอื่น ๆ อีกหลายวิธีที่น่าสนใจ

๓.๒.๓ <u>วิธีรวมสตราตา</u>[๖] (Grouped Strata)

เป็นวิธีการที่คำนวณหาความแปรปรวนของตัวประมาณค่าที่ช่วยให้การคำนวณสะดวกและรวดเร็วขึ้นวิธีหนึ่งเหมาะสำหรับในการสำรวจซึ่งแผนแบบที่ใช้ประกอบด้วยสตราตาเป็นจำนวนมาก เป็นต้นว่า ๕๐ สตราตา การประมาณค่าความแปรปรวนของตัวประมาณค่าต่าง ๆ ย่อมยุ่งยาก เพราะค่าความแปรปรวนย่อมประกอบขึ้นด้วยค่าความแปรปรวนภายในหน่วยตัวอย่างขั้นแรกและค่าความแปรปรวนระหว่างหน่วยตัวอย่างขั้นแรก ดังนั้นการที่เราจะประมาณผลรวมค่าความแปรปรวนทั้งหมด โดยไม่จำเป็นต้องคำนวณแต่ละส่วน ก็โดยจัดรวมสตราตาเป็นหมู่ ๆ หรือในกรณีที่แต่ละสตราตาประกอบด้วย ๑ หน่วย การประมาณค่าความแปรปรวนย่อมทำไม่ได้ ดังนั้นจึงใช้วิธีรวมสตราตาเช่นเดียวกับการรวมสตราตาเข้าด้วยกันนั้น จะรวมเป็นสตราใหม่เท่าใดนั้นแล้วแต่ความเหมาะสม ทั้งนี้ควรให้สตราตาที่ได้ใหม่นี้มีความแตกต่างกันภายในน้อยที่สุด และถ้าเป็นไปได้ขนาดตัวอย่างในแต่ละสตราตาใหม่ก็ไม่ควรจะแตกต่างกันมากนัก เพื่อ BIAS ที่เกิดขึ้นมีน้อยที่สุด

ในกรณีที่แต่ละสตราตาประกอบด้วยตัวอย่าง ๑ หน่วย วิธีการรวมก็เช่นเดียวกันแต่การจัดทำเป็นสตราตาใหม่นั้น ในแต่ละสตราตาใหม่ควรประกอบขึ้นด้วยสตราตาเดิมสองสตราตาทั้งนี้ผู้เชี่ยวชาญสถิติแนะนำว่าแผนแบบที่ใช้นั้นควรมีสตราตาเดิมไม่น้อยกว่า ๒๐ สตราตา เมื่อนำ

---

๖ รวบรวมมาจาก Des Raj, <u>Sampling Theory</u> (New York:McGraw - Hill Book Company, 1968.

Cochran,W.G., <u>Sampling Techniques</u> (New York; John Wiley & Sons Inc.,1963

มารวมเป็นสตราใหม่แล้วจะมีประมาณ ๑๐ สตราตา เพื่อให้ค่าอัตรา...ที่มีขึ้นในทางบวกภาย หลังการจัดรวมสตราตาใหม่นี้มีค่าน้อยลงถ้าหากค่าเฉลี่ยในแต่ละสตราตาไม่ต่างกันมาก

๓.๒.๔ วิธีการใช้โค้งแห่งความถดถอย[ป] (Regression Curve Method)

การสำรวจด้วยวิธีตัวอย่างขนาดใหญ่นั้น จำเป็นต้องคำนวณหาตัวประมาณค่าเป็นจำนวนมาก และในบางครั้งตัวรวมเสนอผลการสำรวจก็มีเป็นจำนวนมากด้วย ดังนั้นการจะประมาณค่าความแปรปรวนของตัวประมาณค่าทั้งหมดที่สนใจนั้นย่อมเป็นไปไม่ได้ ในทางปฏิบัติแล้วอาจทำได้แต่เพียงลักษณะหลักเท่านั้น แม้กระนั้นก็ตามเรายังสนใจที่จะทราบว่าค่าความแปรปรวนของตัวประมาณค่า ลักษณะย่อย ๆ นั้นเป็นอย่างไร ด้วยเหตุนี้การใช้โค้งแห่งความถดถอยจึงเป็นวิธีที่น่าสนใจในการประมาณค่าความแปรปรวนอีกวิธีหนึ่งที่สะดวกและประหยัดเวลาเหมาะสมกับลักษณะย่อย ๆ ของประชากร

นักสถิติได้ค้นพบว่า ในการสำรวจซึ่งประกอบด้วยลักษณะต่าง ๆ นั้น บางลักษณะมีการเปลี่ยนแปลงไปในทางเดียวกันหรือเปลี่ยนแปลงไปในทางคงที่ เช่นจำนวนคนเป็นต้น นอกจากนี้อาจกล่าวได้ว่าค่าประมาณความแปรปรวน จะมีการเปลี่ยนแปลงอย่างมีระบบคือค่าประมาณความแปรปรวนของลักษณะเดียวกันมักมีค่าคงที่หรือมีการเปลี่ยนแปลงน้อยมาก ด้วยเหตุผลดังกล่าวนี้การใช้โค้งแห่งความถดถอยจะช่วยให้ทราบค่าประมาณความแปรปรวนของลักษณะทั้งหมดโดยไม่ต้องทำการประมาณทุกลักษณะ

วิธีการทำก็คือประมาณค่าความแปรปรวนของตัวประมาณค่าในกลุ่มซึ่งเป็นเรื่องที่เกี่ยวข้องกัน โดยประมาณค่าความแปรปรวนเฉพาะลักษณะหลักสำคัญ ๆ ก่อน แล้วหาค่าความแปรปรวนสัมพัทธ์[๒๑] (Relative Variance) จากนั้นนำค่าของตัวประมาณค่าและค่าความแปรปรวน

ป รวบรวมมาจาก

Hansen,Hurwitze Madow; Sample Survey Methods and Theory (Vol.I,II) (New York:John Wiley & Sons Inc.1953,

Atlantida Series ISPO 1, No.I.E (V.S.Bureau of the Census, 1966,

๒๑ ค่าความแปรปรวนสัมพัทธ์คือค่าสัมประสิทธิ์ของความแปรปรวนยกกำลังสอง

สัมพัทธ์ไปเขียนกราฟโดยให้แกน x แทนตัวประมาณค่าแกน y แทนค่าความแปรปรวนสัมพัทธ์ จากกราฟนี้ก็จะทราบ——ความสัมพันธ์ของค่าทั้งสอง ดังนั้นจึงอ่านค่าความแปรปรวนสัมพัทธ์โดยประมาณของลักษณะย่อย ๆ จากกราฟ

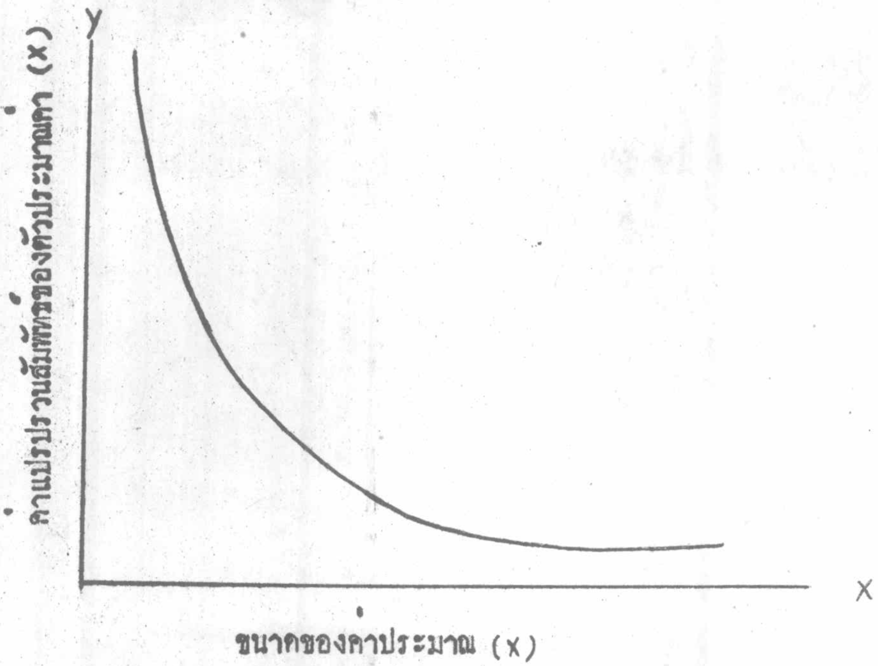
ค่าประมาณของตัวประมาณค่าและค่าประมาณของความแปรปรวนมีความสัมพันธ์กันดังนี้คือ ถ้าค่าของตัวประมาณค่าเพิ่มขึ้นค่าความแปรปรวนสัมพัทธ์จะลดลง ถ้าค่าความแปรปรวนสัมพัทธ์ลดลงค่าประมาณของตัวประมาณค่าจะเพิ่มขึ้น แต่ค่าที่เพิ่มขึ้นหรือลดลงจะไม่เปลี่ยนเร็วนัก ซึ่งความสัมพันธ์ของทั้งสองสิ่งนี้จะอยู่ในรูปฟังค์ชั่นของ $v_x^2 = a + \frac{b}{x}$

เมื่อ v_x^2 = ค่าความแปรปรวนสัมพัทธ์ของตัวประมาณค่า (x)

x = ค่าประมาณของลักษณะที่สนใจจะศึกษา เช่น ผลรวม ฯลฯ

a, b = ค่าคงที่ ซึ่งคำนวณได้โดยวิธีกำลังสองน้อยที่สุด (Least square method)

ซึ่งลักษณะของความสัมพันธ์จะอยู่ในรูปเส้นโค้งดังนี้



ค่าความแปรปรวนที่ได้จะให้ความเชื่อถือได้มากน้อยเพียงไรย่อมขึ้นกับการเลือกลักษณะของค่าประมาณของเรื่องนั้น ๆ ว่ามีความเกี่ยวข้องกันเพียงไร ถ้าลักษณะของกลุ่มที่ทำการศึกษานั้นมีความสัมพันธ์กันมากค่าความแปรปรวนที่ได้ก็จะให้ความเชื่อถือได้มาก นอกจากนี้การประมาณค่าความแปรปรวนโดยวิธีใช้เส้นโค้งแห่งความถดถอยควรจะได้รับตรวจสอบบ้างโดยคำนวณหาค่าความแปรปรวนสัมพัทธ์จากสูตรประมาณจริง ๆ แล้วนำค่านั้นมาเปรียบเทียบกับค่าที่อ่านได้จากเส้นโค้ง

วิธีการประมาณค่าความแปรปรวนของตัวประมาณค่าต่อไปนี้ได้รวบรวมมาจากหนังสือ American Statistical Association, Procudings of the Social Statistics Section, 1971 (P 20-27) โดยมากใช้ในการประมาณหาค่าความแปรปรวนของค่าเฉลี่ยอัตราส่วน ความแตกต่างของค่าเฉลี่ยอัตราส่วน ยอดรวม Partial Correlations, สหสัมพันธ์แบบธรรมดา สหสัมพันธ์เชิงซ้อน สัมประสิทธิ์ของความถดถอย

๓.๒.๕ <u>วิธีการกระจายแบบเทเลอร์</u> (Taylor Expansion Method)

วิธีนี้อาจเรียกอีกอย่างหนึ่งได้ว่าวิธีเดลตา (Delta-Method) เหมาะสมที่จะใช้กับแผนแบบการเลือกตัวอย่างประเภทที่มีการแบ่งสตราตา สมมุติว่ามีจำนวนสตราตาทั้งหมด H สตราตา ตัวอย่างในแต่ละสตราตาจะถูกเลือกมาด้วยวิธีการสุ่มอย่างง่ายจำนวน ๒ หน่วยตัวอย่างขั้นต้น (PSU) จากจำนวนทั้งหมด A หน่วย

ให้ $y = (y_1, y_2, \ldots\ldots, y_k)$ เป็นเวคเตอร์ของค่าจากตัวอย่างซึ่งค่าเหล่านี้ประกอบด้วยค่าของหน่วยตัวอย่างขั้นต้น y_{iha} ในเชิงเส้นตรง

ซึ่ง $$y_i = \sum_{h=1}^{H}\sum_{a=1}^{2} y_{iha} = \sum_{h=1}^{H} y_{ih.}$$

โดยที่ $E(y) = Y = (Y_1, Y_2, \ldots\ldots Y_i, \ldots. Y_k)$

ให้ $g(y)$ เป็นตัวประมาณค่าของพารามิเตอร์ $g(Y)$ โดยอาศัยวิธีการกระจายแบบเทเลอร์จะได้

$$V[g(y)] \doteq V\left[g(Y) + \sum_{i=1}^{k} (y_i - Y_i) \frac{\partial g(Y)}{\partial Y_i}\right]\Bigg|_{y=Y}$$

เมื่อ $g(Y)$ และ $Y_i \frac{\partial g(Y)}{\partial Y_i}$ คงที่

$$\therefore V(g(y)) \doteq V\left(\sum_{i=1}^{k}\left(\frac{\partial g(Y)}{\partial Y_i}\right) y_i\right)$$

แต่ $y_i = y_{ih.}$, $Y_i = Y_{ih.}$

เนื่องจากการเลือกตัวอย่างในแต่ละสตราตาเป็นอิสระต่อกัน

$$V(g(y)) \doteq \sum_{h=1}^{H} W_h^2 \, V\left(\sum_{i=1}^{k} \frac{\partial g(Y)}{\partial Y_i} y_{ih.}\right)$$

W_h = ตัวถ่วงน้ำหนักในสตราตาที่ h

เนื่องจากในแต่ละสตราตาเลือกมา ๒ หน่วยตัวอย่างข้างบนโดยการไม่ใส่แทนที่และให้ความน่าจะเป็นในการถูกเลือกของแต่ละหน่วยเท่ากันและเท่ากับ f_h ดังนั้นเราจะได้ค่าประมาณของค่าความแปรปรวนของ Y_{ih} เป็น

$$\hat{V}\left(\sum_{i=1}^{k} \frac{\partial g(Y)}{\partial Y_i} y_{ih}\right) = (1-f_h)\left(\sum_{i=1}^{k} \frac{\partial g(Y)}{\partial Y_i} y_{ih1} - \sum_{i=1}^{k} \frac{\partial g(Y)}{\partial Y_i} y_{ih2}\right)^2$$

ซึ่ง y_{ih1} และ y_{ih2} เป็นค่าของผลรวมจากตัวอย่างที่ได้มาจากหน่วยตัวอย่างข้างต้นในสตราตาที่ h

ถ้า $W_h = 1$ และ f_h เท่ากันในทุกสตราตา $= f$ แล้วค่าประมาณของ $V[g(y)]$ เป็น

$$\hat{V}(g(y)) = (1-f)\sum_{h=1}^{H}\left(\sum_{i=1}^{k} \frac{\partial g(Y)}{\partial Y_i} y_{ih1} - \sum_{i=1}^{k} \frac{\partial g(Y)}{\partial Y_i} y_{ih2}\right)^2$$

แต่ในทางปฏิบัติจริง ๆ แล้วค่าคงที่ $\frac{\partial g(Y)}{\partial Y_i}$ ไม่สามารถทราบค่าได้ ซึ่งต้องทำการประมาณจาก $g(y)$

ดังนั้น $$\hat{V}(g(y)) = (1-f)\sum_{h=1}^{H}\left(\sum_{i=1}^{k}\frac{\partial g(y)}{\partial y_i}\, y_{ih_1} - \sum_{i=1}^{k}\frac{\partial g(y)}{\partial y_i}\, y_{ih_2}\right)^2$$

๓.๒.๖ วิธีบาลานซ์รีพีทเตดเรพพลิเคชัน (Balanced Repeated Replication (BRR) Method)

สมมุติว่าแผนแบบในการเลือกตัวอย่างเป็นแบบการจัดสตราตาซึ่งมีทั้งหมด H สตราตา ในแต่ละสตราตาเลือกมา ๒ หน่วยตัวอย่างข้างต้น โดยให้โอกาสที่จะถูกเลือกเท่า ๆ กันเป็น f

s = ตัวอย่างทั้งหมด

H_i = half - sample ที่ i ที่เกิดจากการเลือกหน่วยตัวอย่างข้างต้นมา ๑ ใน ๒ ของแต่ละสตราตา

C_i = ส่วนที่ไม่ใช่ half - sample ที่ i (the ith complement half - sample) ที่เกิดจากการรวมหน่วยตัวอย่างข้างต้นที่เหลือซึ่งอยู่ใน s เมื่อจัด half - sampleแล้ว

ถ้าต้องการจัด k half - sample คือ $H_1, H_2, \ldots\ldots\ldots H_k$ ซึ่งจะ correspond กับส่วนที่ไม่ใช่ half - sample $C_1, C_2 \ldots\ldots C_k$ ดังนั้นจะได้วิธีประมาณความแปรปรวนของ g(s) มี ๔ ชนิดคือ

ชนิดที่ ๑ Half - total : $$V_{BRR-H}(g(s)) = \frac{(1-f)}{k}\sum_{i=1}^{k}(g(H_i) - g(s))^2$$

ชนิดที่ ๒ Complement - total : $$V_{BRR-C}(g(s)) = \frac{(1-f)}{k}\sum_{i=1}^{k}(g(C_i) - g(s))^2$$

ชนิดที่ ๓ ผลบวกของ BRR-H และ BRR-C : $$V_{BRR-s}(g(s)) = \frac{1}{2}\left\{V_{BRR-H}(g(s)) + V_{BRR-C}(g(s))\right\}$$

ชนิดที่ ๔ Half-complement: $$V_{BRR-D}(g(s)) = \frac{(1-f)}{4k}\sum_{i=1}^{k}(g(H_i) - g(C_i))^2$$

วิธีลัดต่าง ๆ เหล่านี้มีประโยชน์มากเพราะสามารถลดขั้นการทำงานต่าง ๆ ลงไป เพื่อความสะดวก รวดเร็ว ในกรณีที่เร่งด่วน หรืองบประมาณน้อย ถึงแม้ค่าที่ได้ลดความเชื่อถือลงไปบ้าง