แทน (2) และ (3) ใน (1)

ดังนั้น $V(\bar{x}) = \frac{1}{n}\frac{N-1}{N}\cdot S^2 - \frac{n-1}{n}\cdot\frac{1}{N}\cdot S^2$

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

$$= \frac{N-1-n+1}{nN}\cdot S^2$$

$$= \frac{N-n}{N}\frac{S^2}{n}$$

คำอธิบายเพิ่มเติม 2

1. $(\sum_{i}^{n} x_i)^2$ สามารถเสนอได้เป็น 3 ลักษณะซึ่งเราสามารถเลือกใช้ลักษณะใด-ลักษณะหนึ่งได้ คือ

ก. $(\sum_{i}^{n} x_i)^2 = \sum_{i}^{n} x_i^2 + \sum_{i\neq j}^{n}\sum^{n} x_i x_j$

ข. $(\sum_{i}^{n} x_i)^2 = \sum_{i}^{n} x_i^2 + 2\sum_{\substack{i=1\\i>j}}^{n-1}\sum_{j=2}^{n} x_i x_j$

ค. $(\sum_{i}^{n} x_i)^2 = \sum_{i}^{n} x_i^2 + 2\sum_{\substack{i=2\\i>j}}^{n}\sum_{j=1}^{n-1} x_i x_j$

ในที่นี้จะแสดงให้เห็นโดยจะเริ่มจากกรณีที่ $n = 3$ ก่อนแล้วค่อยขยายความสู่กรณีทั่วไปในภายหลัง จะพบว่า

$$(\sum_{i}^{3} x_i)^2 = (x_1 + x_2 + x_3)^2$$

$$= x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_2x_1 + x_1x_3 + x_3x_1 + x_2x_3 + x_3x_2$$

$$= \sum_{i}^{3} x_i^2 + \sum_{i=1}^{3} \sum_{\substack{j=1 \\ i \neq j}}^{3} x_i x_j$$

หรือ $= x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3$

$$= \sum_{i}^{3} x_i^2 + 2 \sum_{i=1}^{2} \sum_{\substack{j=2 \\ i<j}}^{3} x_i x_j$$

หรือ $= x_1^2 + x_2^2 + x_3^2 + 2x_2x_1 + 2x_3x_1 + 2x_3x_2$

$$= \sum_{i}^{3} x_i^2 + 2 \sum_{i=2}^{3} \sum_{\substack{j=1 \\ i>j}}^{2} x_i x_j$$

ถ้าพิจารณา $\sum_{i=1}^{3} \sum_{\substack{j=1 \\ i \neq j}}^{3} x_i x_j$, $2 \sum_{i=1}^{2} \sum_{\substack{j=2 \\ i>j}}^{3} x_i x_j$, $2 \sum_{i=2}^{3} \sum_{\substack{j=1 \\ i>j}}^{2} x_i x_j$ จะพบว่า

$$\sum_{i=1}^{3} \sum_{\substack{j=1 \\ i \neq j}}^{3} x_i x_j = \sum_{\substack{i=1 \\ i \neq j}}^{3} (x_i x_1 + x_i x_2 + x_i x_3)$$

$$= (x_1x_2 + x_1x_3) + (x_2x_1 + x_2x_3) + (x_3x_1 + x_3x_2)$$

$$= (x_1x_2 + x_1x_3 + x_2x_1 + x_2x_3 + x_3x_1 + x_3x_2)$$

$$2 \sum_{\substack{i=1 \\ i>j}}^{2} \sum_{j=2}^{3} x_i x_j = 2 \sum_{\substack{i=1 \\ i<j}}^{2} (x_i x_2 + x_i x_3)$$

$$= 2\left\{(x_1x_2 + x_1x_3) + x_2x_3)\right\}$$

$$= 2x_1x_2 + 2x_1x_3 + 2x_2x_3$$

$$2 \sum_{\substack{i=2 \\ i>j}}^{3} \sum_{j=1}^{2} x_i x_j = 2 \sum_{\substack{i=2 \\ i>j}}^{3} (x_i x_1 + x_i x_2)$$

$$= 2\left\{(x_2x_1 + (x_3x_1 + x_3x_2)\right\}$$

$$= 2x_2x_1 + 2x_3x_1 + 2x_3x_2$$

ดังนั้นในกรณีทั่วไปจึงพบว่า

$$\left(\sum_{i}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^{n} \sum_{j=1}^{n} x_i x_j$$

$$= \sum_{i=1}^{n} x_i^2 + 2 \sum_{\substack{i=1 \\ i<j}}^{n-1} \sum_{j=2}^{n} x_i x_j .$$

หรือ

$$= \sum_{i=1}^{n} x_i^2 + 2 \sum_{\substack{i=2 \\ i>j}}^{n} \sum_{j=1}^{n-1} x_i x_j$$

2. $\sum_{\substack{i=1\\ i\neq j}}^{n} \sum_{j=1}^{n} S^2 = n(n-1)S^2$

ให้ $n = 3$ และ $S_{12}^2 = S_{21}^2 = S_{13}^2 = S_{31}^2 = S_{23}^2 = S_{32}^2 = S^2$
ดังนั้น

$$\sum_{\substack{i=1\\ i\neq j}}^{3} \sum_{j=1}^{3} S_{ij}^2 = \sum_{i=1}^{3} (S_{i1}^2 + S_{i2}^2 + S_{i3}^2)$$

$$= (S_{12}^2 + S_{13}^2) + (S_{21}^2 + S_{23}^2) + (S_{31}^2 + S_{32}^2)$$

$$= (S^2 + S^2) + (S^2 + S^2) + (S^2 + S^2)$$

$$= 3(2S^2)$$

$$= 3(3-1)S^2$$

ดังนั้นสำหรับกรณีทั่วไปคือ $\sum_{\substack{i=1\\ i\neq j}}^{n} \sum_{j=1}^{n} S^2$ จึงพบว่า

$$\sum_{\substack{i=1\\ i\neq j}}^{n} \sum_{j=1}^{n} S_{ij}^2 = (\quad + S_{12}^2 + \ldots + S_{1n}^2) + (S_{21}^2 + \quad + S_{23}^2 + S_{24}^2 + \ldots + S_{2n}^2)$$

$$+ (S_{31}^2 + S_{32}^2 + \quad + S_{34}^2 + S_{35}^2 + \ldots + S_{3n}^2) + (S_{41}^2 + S_{42}^2 + \ldots + S_{4n}^2)$$

$$+ \ldots + (S_{n1}^2 + S_{n2}^2 + \ldots + S_{n,(n-1)}^2)$$

$$= \underbrace{(n-1)S^2 + (n-1)S^2 + (n-1)S^2 + \ldots + (n-1)S^2}_{n \text{ เทอม}}$$

$$= n(n-1)S^2$$

3. $$2 \sum_{\substack{i=1 \\ i<j}}^{n-1} \sum_{j=2}^{n} S^2 = n(n-1)S^2$$

ให้ $n = 3$ และ $S_{12}^2 = S_{13}^2 = S_{23}^2 = S^2$

ดังนั้น $$2 \sum_{\substack{i=1 \\ i<j}}^{2} \sum_{j=2}^{3} S_{ij}^2 = 2 \sum_{\substack{i=1 \\ i<j}}^{2} (S_{i2}^2 + S_{i3}^2)$$

$$= 2\left\{(S_{12}^2 + S_{13}^2) + S_{23}^2\right\}$$

$$= 2(1+1+1)S^2$$ [1]

$$= 2 \cdot \frac{3}{2}(1+1)S^2$$

$$= 3.2S^2$$

$$= 3(3-1)S^2$$

---

[1] ผลรวมของอนุกรมเลขคณิตคือ $S_n = \frac{n}{2}(a+L)$ เมื่อ a คือเทอมแรก L คือเทอมสุดท้าย

ดังนั้นในกรณีทั่วไปจะพบว่า

$$2 \sum_{\substack{i=1 \\ i<j}}^{n-1} \sum_{j=2}^{n} S_{ij}^2 = 2 \sum_{\substack{i=1 \\ i<j}}^{n-1} (S_{i2}^2 + S_{i3}^2 + \ldots + S_{in}^2)$$

$$= 2\{\underbrace{S_{12}^2 + S_{13}^2 + \ldots + S_{1n}^2)}_{(n-1)\text{ เทอม}} + \underbrace{(S_{23}^2 + S_{24}^2 + \ldots + S_{2n}^2)}_{(n-2)\text{ เทอม}} +$$

$$\underbrace{(S_{34}^2 + S_{35}^2 + \ldots + S_{3n}^2)}_{(n-3)\text{ เทอม}} + \ldots + (S^2_{n-1,n})\}$$

$$= 2 \cdot \frac{n-1}{2} \cdot \{(n-1)+1\}S^2$$

$$= \frac{2(n-1)n}{2} \cdot S^2$$

$$= n(n-1)S^2$$

4. ในทำนองเดียวกัน เราสามารถแสดงให้เห็นได้ว่า $2 \sum_{\substack{i=2 \\ i>j}}^{n} \sum_{j=1}^{n-1} S^2 = n(n-1)S^2$

5. $\bar{x} = \frac{1}{n} \sum_{i}^{n} x_i$ เป็นตัวสถิติ (Statistics) ถือว่าเป็นตัวแปรสุ่มเพราะค่าของ $\bar{x}$ ย่อมเปลี่ยนไปได้มากมายตามกลุ่มตัวอย่าง (Possible Sample) และค่าของ $\bar{x}$ เกิดขึ้นได้เพราะมีความน่าจะเป็นกำกับอยู่ในกรณีนี้มีค่าเท่ากับ $\frac{1}{\binom{N}{n}}$ ขอให้สังเกตว่าความน่าจะเป็นที่จะได้ $\bar{x}$ ใด ๆ มีค่าเดียวกับความน่าจะเป็นที่จะได้กลุ่มตัวอย่างใด ๆ ทั้งนี้เพราะ $\bar{x}$ เป็นค่าที่ได้มาจากกลุ่มตัวอย่าง ความน่าจะเป็นที่จะได้ตัวสถิติอื่น ๆ ก็มีค่าเท่ากับ $\frac{1}{\binom{N}{n}}$ เช่นกัน

เช่นค่าสูงสุด (Y_n) ค่าต่ำสุด (Y_1) พิสัย ($Y_n - Y_1$) ความแปรปรวน (s^2) ฯลฯ เพราะค่าเหล่านี้เป็นค่าที่ได้มาจากกลุ่มตัวอย่างและตัวอย่างแต่ละกลุ่มมีความน่าจะเป็นที่จะได้รับเลือกเท่ากับ

$$\frac{1}{\binom{N}{n}}$$

เมื่อ $\bar{x}$ เป็นตัวแปรสุ่ม $\bar{x}$ ย่อมมี pdf ซึ่งเรียกว่า Sampling Distribution ทำให้เราทราบค่าเฉลี่ยและความแปรปรวนของ $\bar{x}$ ได้ ความแปรปรวนของ $\bar{x}$ คือ $V(\bar{x})$ ใช้เป็นเครื่องมือวัดการกระจายของ $\bar{x}$ ในระหว่างค่าที่เป็นไปได้ทั้งหมดซึ่งมีอยู่ $\binom{N}{n}$ ค่า รอบ ๆ ค่าเฉลี่ยคือ $E(\bar{x}) = \bar{X}$ ถ้า $V(\bar{x})$ มีค่าต่ำแสดงว่าโดยถัวเฉลี่ยแล้ว $\bar{x}$ เป็นค่าที่เบนห่างไปจาก $\bar{X}$ ไม่มากนัก ถ้า $V(\bar{x})$ มีค่าสูงแสดงว่าโดยถัวเฉลี่ยแล้ว $\bar{x}$ จะเบนห่างไปจาก $\bar{X}$ มาก ดังนั้น $V(\bar{x})$ จึงใช้วัดความแม่นยำของการประมาณค่า $\bar{X}$ ด้วย $\bar{x}$ ถ้า $V(\bar{x})$ มีค่าต่ำแสดงว่า $\bar{x}$ สามารถประมาณค่า $\bar{X}$ ได้แม่นยำหรือค่อนข้างแม่นยำถ้า $V(\bar{x})$ มีค่าสูงก็จะมีความหมายโดยนัยตรงข้ามกัน

เมื่อกล่าวถึงกรณีทั่วไป ถ้าเราใช้ $\hat{\theta}$ เป็นตัวประมาณค่าของพารามิเตอร์ θ แล้ว-$V(\hat{\theta})$ จะใช้เป็นเครื่องมือวัดความแม่นยำของ $\hat{\theta}$ ที่ใช้ประมาณค่าพารามิเตอร์ θ ดังกล่าวและในงานวิจัยชิ้นหนึ่ง ๆ พารามิเตอร์ที่นักวิจัยให้ความสนใจอาจมีมากมายหลายตัว ดังนั้น θ อาจหมายถึงสัดส่วน (P) อัตราส่วน (R) ค่าเฉลี่ย ($\bar{X}$) ร้อยละ (100P%) ค่าสูงสุด (Max.) ค่าต่ำสุด (Min.) พิสัย (Range) และอื่น ๆ พารามิเตอร์เหล่านี้ย่อมมีตัวประมาณค่าต่างกันไปเฉพาะเรื่องเฉพาะตัว สิ่งที่นักวิจัยพึงทราบก็คือตัวประมาณค่ามีลักษณะรูปร่างอย่างไรและมีความแม่นยำในการประมาณค่ามากน้อยเพียงใด เกี่ยวกับเรื่องราวที่น่าสนใจเหล่านี้นักศึกษาจะได้พบในลำดับต่อ ๆ ไป

6. $E(x_i - \bar{X})(x_j - \bar{X}) = \sum_{i \neq j}^{N}\sum^{N} (x_i - \bar{X})Pr(x_i - \bar{X})(x_j - \bar{X})Pr(x_j - \bar{X})$ ขอให้ทำความเข้าใจว่า $Pr(x_i - \bar{X}) = Pr(x_i)$ และ $Pr(x_j - \bar{X}) = Pr(x_j)$ ทั้งนี้เพราะ $\bar{X}$ เป็นค่าของพารามิเตอร์ซึ่งถือว่าเป็นค่าคงที่ การนำค่าคงที่ไปหักออกจากค่าของตัวแปรสุ่มไม่ทำให้ความสุ่ม (Randomness) เสียไป

พิจารณา $Pr(x_i)$ และ $Pr(x_j)$,x_i และ x_j เป็นตัวแปรสุ่มที่ x_i หรือ x_j อาจได้รับเลือกก่อนหลังกันได้ x_i อาจได้รับเลือกก่อน หรือ x_j อาจได้รับเลือกก่อนก็ได้ ถ้า x_i ได้รับเลือกก่อนแสดงว่า $Pr(x_i) = Pr(x_i - \overline{X}) = \frac{1}{N}$ เมื่อ x_i ได้รับเลือกไปแล้วขนาดประชากรจะลดลงเหลือเพียง $N-1$ ทำให้ $Pr(x_j) = Pr(x_j - \overline{X}) = \frac{1}{N-1}$ ถ้า x_j ได้รับเลือกก่อน $Pr(x_j) = Pr(x_j - \overline{X}) = \frac{1}{N}$ และ $Pr(x_i) = Pr(x_i - \overline{X}) = \frac{1}{N-1}$

อนึ่ง ในทางทฤษฎีเราใช้อักษรตัวเล็กในความหมายของค่าของตัวแปรสุ่มซึ่งโดยปกติจะใช้แทนด้วยอักษรตัวใหญ่ เช่น $Pr(X=x)$ อ่านว่า ความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าเท่ากับ x ตัวอย่างเช่น $Pr(X=2)$, $Pr(X=3)$ เป็นต้น ในที่นี้จึงขอให้นักศึกษาทำความเข้าใจร่วมกันไว้ในที่นี้ว่า เราใช้อักษรตัวเล็กในความหมายของทั้งตัวแปรสุ่มและค่าของตัวแปรสุ่ม ทั้งนี้เพราะไม่ปรารถนาให้การใช้อักษรไปปะปนกันกับพารามิเตอร์ซึ่งใช้อักษรตัวใหญ่ขอให้ทำความเข้าใจง่าย ๆ ดังนี้ คือ ขณะดำเนินการพิสูจน์ อักษรตัวเล็กทุกตัว (ยกเว้นอักษรใหญ่ซึ่งถือว่าเป็นพารามิเตอร์) จะเป็นตัวแปรสุ่ม เมื่อจะนำสูตรนั้นไปใช้วิเคราะห์หมายความว่าได้รับข้อมูลมาเรียบร้อยแล้ว อักษรตัวเล็กทุกตัวจะใช้ในความหมายของค่าตัวแปรสุ่ม ความจริงเรื่องนี้ ถ้านักศึกษามีพื้นความรู้ทางทฤษฎีสถิติและความน่าจะเป็นอย่างเพียงพอแล้วจะไม่รู้สึกสับสนประการใด[1] เพราะในทางทฤษฎีนี้เราพูดถึงเฉพาะกับตัวแปรสุ่มทั้งสิ้นไม่เคยพูดถึงค่าของตัวแปรสุ่มเลย จนกว่าจะถึงขั้นปฏิบัติ

2. พิสูจน์ $V(\hat{T}) = N^2 \frac{N-n}{N} \cdot \frac{S^2}{n}$

การพิสูจน์ตอนนี้เราสามารถอาศัยผลการพิสูจน์ของทฤษฎี 1.1 และ 1.2 รวมทั้งทฤษฎีสถิติบางเรื่องรวมกัน ดังนี้

$$
\begin{aligned}
V(\hat{T}) &= V(N\bar{x}) \\
&= N^2 V(\bar{x})
\end{aligned}
$$

[1] อ่าน "ทฤษฎีสถิติ 2 เล่มที่ 1"

$$= N^2 \frac{N-n}{N} \cdot \frac{S^2}{n}$$

บทแทรก 2.2 เมื่อดำเนินการสุ่มตัวอย่างโดยใช้แผนสำรวจ SRS ช่วงเชื่อมั่น $100(1-\alpha)\%$ ที่เชื่อว่าค่าจริงของ $\bar{X}$ และ T จะปรากฏอยู่คือ $(\bar{x} - Z_{1-\alpha/2}\sqrt{V(\bar{x})},\ \bar{x}+Z_{1-\alpha/2}\sqrt{V(\bar{x})})$ และ $(\hat{T} - Z_{1-\alpha/2}\sqrt{V(\hat{T})},\ \hat{T} + Z_{1-\alpha/2}\sqrt{V(\hat{T})})$ หรือ

$$\bar{x} - Z_{1-\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}} < \bar{X} < \bar{x} + Z_{1-\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

และ

$$\hat{T} - Z_{1-\alpha/2}\frac{NS}{\sqrt{n}}\sqrt{\frac{N-n}{N}} < T < \hat{T} + Z_{1-\alpha/2}\frac{NS}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

และถ้า $N >> n$ ช่วงเชื่อมั่น $100(1-\alpha)\%$ โดยประมาณที่คาดว่าค่าจริงของ $\bar{X}$ และ T จะปรากฏอยู่คือ

$$\bar{x} - Z_{1-\alpha/2}S/\sqrt{n} < \bar{X} < \bar{x} + Z_{1-\alpha/2}S/\sqrt{n}$$

และ

$$\hat{T} - Z_{1-\alpha/2}NS/\sqrt{n} < T < \hat{T} + Z_{1-\alpha/2}NS/\sqrt{n}$$

บทแทรก 2.2 แสดงให้เห็นถึงวิธีการประมาณค่าเฉลี่ยและยอดรวมของกลุ่มประชากรด้วยช่วงเชื่อมั่น (Interval Estimation) ในขณะที่ $\bar{x} = \frac{1}{n} \sum_{i}^{n} x_i$ และ $T = N\bar{x}$ คือค่าประมาณของพารามิเตอร์ $\bar{X}$ และ T ตามวิธีของการประมาณค่าพารามิเตอร์ด้วยค่าคงที่ การประมาณค่าพารามิเตอร์ด้วยค่าคงที่เป็นวิธีที่เสี่ยงต่อความผิดพลาดได้ง่ายเพราะถ้าค่าประมาณคลาดเคลื่อนกับค่าจริงเพียงเล็กน้อยก็ถือว่าประมาณผิดพลาดแล้ว ส่วนวิธีประมาณค่าด้วยช่วงเชื่อมั่นความเสี่ยงต่อความผิดพลาดย่อมมีได้น้อยกว่า ในทางปฏิบัติการใช้ทั้งสองวิธี

ช่วงเชื่อมั่นที่ดีคือช่วงเชื่อมั่นที่แคบกว่า ณ. ระดับความเชื่อมั่นเดียวกันเพราะคาดหมายไว้ได้กระชับรัดกุมแม่นยำกว่า แสดงให้เห็นถึงประสิทธิภาพของงานที่สูงกว่าและการรู้จักใช้ข้อมูลข้อสนเทศ (n) มากกว่า จำนวนข้อมูลข้อสนเทศ (ขนาดตัวอย่าง) มีส่วนอย่างสำคัญที่ทำให้ความผิดพลาดของการประมาณค่าลดลงซึ่งมีส่วนให้ช่วงเชื่อมั่นแคบลงด้วย ตัวอย่างเช่น เรื่องการกะประมาณอายุ ท่านเห็นชายผู้หนึ่งในระยะห่างโดยอาศัยข้อมูลในขณะนั้นคือท่าทางการเดิน ลักษณะการแต่งกาย และขนาดของร่างกาย ท่านอาจประมาณอายุของชายผู้นั้นได้ว่าคงจะอยู่ระหว่าง 25 ถึง 35 ปี แต่ถ้าท่านได้พบชายผู้นั้นในระยะใกล้กันได้มีโอกาศสนทนา ฟังเสียง เห็นสีผม ลักษณะของผิว สีหน้า หนวดเครา ฟัน การแสดงทัศนะ ฯลฯ ท่านอาจกะประมาณอายุของชายผู้นั้นได้รัดกุมกว่าเดิมเพราะได้รับข้อมูลข้อสนเทศมากกว่าเดิม เช่น อาจกะประมาณช่วงอายุได้แคบเข้าเป็น 29 - 31 ปี ดังนี้เป็นต้น จึงเห็นได้ว่าการได้รับข้อสนเทศมากขึ้นจะทำให้การประมาณค่าแม่นยำและรัดกุมขึ้น หรือนัยหนึ่งการทราบข้อมูลข้อสนเทศมากขึ้นจะทำให้ความแปรปรวนของตัวประมาณค่าต่ำลง ความแปรปรวนของตัวประมาณค่ามีส่วนสำคัญต่อการประมาณค่าพารามิเตอร์ด้วยช่วงเชื่อมั่นและความแปรปรวนดังกล่าวจะมีค่าลดลงได้นอกจากเพราะการเลือกตัวอย่างเป็นไปโดยสุ่มแล้ว อิทธิพลที่สำคัญที่ทำให้ความแปรปรวนของการประมาณค่าลดลงก็คือขนาดตัวอย่าง ขนาดของตัวอย่างโตขึ้นมากเพียงใดค่าความแปรปรวนจะลดลงมากเพียงนั้น ความจริงนี้เราสามารถพิสูจน์ให้เห็นด้วยกฎการใช้กลุ่มตัวอย่างขนาดใหญ่ (Law of Large Number) ซึ่งจะไม่ขอกล่าวถึงในที่นี้อีกเพราะได้อธิบายและพิสูจน์พร้อมคำอธิบายโดยละเอียดไว้แล้วในหนังสือ "ทฤษฎีสถิติ 2 " นักศึกษาที่สนใจสามารถศึกษาได้ด้วยตนเอง

ปัญหาปัจจุบันก็คือ ทั้ง $V(\bar{x})$ และ $V(\hat{T})$ เป็นฟังก์ชันของ S^2 ซึ่งเป็นพารามิเตอร์ที่ไม่อาจทราบค่าได้ คือ $V(\bar{x}) = \frac{N-n}{N}\frac{S^2}{n}$ และ $V(\hat{T}) = N^2\frac{N-n}{N}\cdot\frac{S^2}{n}$ โดยที่ $S^2 = \frac{1}{N-1}\sum_i^N (x_i-\bar{X})^2$ ทำให้ไม่อาจทราบช่วงเชื่อมั่นของ $\bar{X}$ และ T ได้ตามต้องการ สิ่งที่เราต้องการทำในชั้นนี้จึงเป็นเรื่องของการหาค่าประมาณของ $V(\bar{x})$ และ $V(\hat{T})$ แต่ขอให้สังเกตว่า $V(\bar{x})$ และ $V(\hat{T})$ เป็นฟังก์ชันของ S^2 ดังนั้น ถ้าเราสามารถประมาณค่าของ S^2 ได้การหาค่าประมาณของ $V(\bar{x})$ และ $V(\hat{T})$ ก็มิใช่ปัญหาอีกต่อไป เพราะเพียงแต่แทนที่ S^2 ด้วยตัวประมาณค่าก็จะได้ $\hat{V}(\bar{x})$ และ $\hat{V}(\hat{T})$ ตามต้องการ

ทฤษฎี 2.3 ถ้าดำเนินการสำรวจโดยใช้แผนสำรวจแบบ SRS แล้วค่าความแปรปรวนจากกลุ่มตัวอย่าง (Sample Variance) จะเป็นตัวประมาณค่าที่ปราศจากอคติของค่าความแปรปรวนของกลุ่มประชากร (Population Variance) กล่าวคือ

$$E(s^2) = S^2$$

โดยที่ $s^2 = \frac{1}{n-1}\sum_i^n (x_i-\bar{x})^2, \; S^2 = \frac{1}{N-1}\sum_i^N (x_i-\bar{X})^2$

พิสูจน์ $E(s^2) = E\left\{\frac{1}{n-1}\sum_i^n (x_i-\bar{x})^2\right\}$

$$= \frac{1}{n-1}E\left\{\sum_i^n (x_i-\bar{X}-\bar{x}+\bar{X})^2\right\}$$: บวกเข้าและลบออกด้วย $\bar{X}$

$$= \frac{1}{n-1}E\left\{\sum_i^n (x_i-\bar{X})^2 - 2(\bar{x}-\bar{X})\sum_i^n (x_i-\bar{X}) + \sum_i^n (\bar{x}-\bar{X})^2\right\}^1$$

---

[1] $2(\bar{x}-\bar{X})\sum_i^n (x_i-\bar{X}) = 2(\bar{x}-\bar{X})\left(\sum_i^n x_i - \sum_i^n \bar{X}\right) = 2n(\bar{x}-\bar{X})^2$

$$= \frac{1}{n-1} E\left\{\sum_{i}^{n} (x_i-\overline{X})^2 - 2(\bar{x}-\overline{X})(n\bar{x}-n\overline{X}) + n(\bar{x}-\overline{X})^2\right\}$$

$$= \frac{1}{n-1} E\left\{ \sum_{i}^{n} (x_i-\overline{X})^2 - n(\bar{x}-\overline{X})^2\right\}$$

$$= \frac{1}{n-1} \sum_{i}^{n} E(x_i-\overline{X})^2 - \frac{n}{n-1} E(\bar{x}-\overline{X})^2 \quad {}^{1}$$

$$= \frac{1}{n-1} \sum_{i}^{n} \left\{ \sum_{i}^{N} (x_i-\overline{X})^2 \cdot \frac{1}{N} \right\} - \frac{n}{n-1} V(\bar{x})$$

$$= \frac{1}{n-1} \sum_{i}^{n} (N-1)S^2 \cdot \frac{1}{N} - \frac{n}{n-1} V(\bar{x})$$

$$= \frac{1}{n-1} \cdot \frac{N-1}{N} \sum_{i}^{n} S^2 - \frac{n}{n-1} \left(\frac{N-n}{N} \frac{S^2}{n} \right) \quad {}^{2}$$

$$= \frac{n(N-1)}{(n-1)N} \cdot S^2 - \frac{n(N-n)}{(n-1)Nn} \cdot S^2$$

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

$$= \frac{n^2N - nN}{n^2N - nN} \cdot S^2 = S^2$$

แสดงว่า $E(s^2) = S^2$ หรือนัยหนึ่งเราสามารถใช้ s^2 แทน S^2 ได้

---

[1] นิยาม $V(\bar{x}) = E(\bar{x}-\overline{X})^2$

[2] ทฤษฎี 1.2

จากผลการพิสูจน์ทฤษฎี 2.3 ทำให้เราสามารถคำนวณหาค่าประมาณของ $V(\bar{x})$ และ $V(\hat{T})$ ได้ดังนี้

$$\hat{V}(\bar{x}) = \frac{N-n}{N}\frac{S^2}{n}$$

และ

$$\hat{V}(\hat{T}) = N^2 \cdot \frac{N-n}{N} \cdot \frac{s^2}{n}$$

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

ตัวอย่าง 2.1 สุ่มตัวอย่างครัวเรือนมา 30 ครัวเรือนจากในพื้นที่แห่งหนึ่งซึ่งประกอบไปด้วยครัวเรือน 14,848 ครัวเรือน ปรากฏจำนวนบุคคลต่อครัวเรือนเป็นดังนี้

5,6,3,3,2,2,2,2,4,4,3,2,7,4,3,5,4,4,3,3,4,3,3,1,2,4,3,4,2,4 คน ตามลำดับ

จงกะประมาณจำนวนประชากรทั้งหมดในพื้นที่ดังกล่าว ทั้งนี้ค่าประมาณจะคลาดเคลื่อนไปจากค่าจริงไม่เกิน ± 10%

วิธีทำ ยอดรวม $T = N\bar{x}$ ในที่นี้ $N = 14,848,\ n = 30$

$$\bar{x} = \frac{1}{n}\sum_{i}^{n} x_i$$

$$= \frac{1}{30}(5+6+3+3+\ldots+4+2+4)$$

$$= \frac{104}{30} = 3.467$$

หรือโดยถัวเฉลี่ยแล้วครัวเรือนหนึ่ง ๆ จะประกอบด้วยสมาชิก 3.467 คน ดังนั้นประชากรทั้งหมดในพื้นที่ดังกล่าวจึงมีจำนวนทั้งสิ้นเท่ากับ $N\bar{x} = 14,848 \times 3.467 = 51,473.07$ หรือประมาณ 51,473 คน

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

$$= \frac{1}{30-1} \left(404 - \frac{10,816}{30}\right) = 1,304/29 = 44.965$$

$$s = 6.7056$$

ดังนั้น $\hat{V}(\bar{x}) = \frac{N-n}{N} \frac{s^2}{n}$

$$= \frac{14,848-30}{14848} \left(\frac{44.965}{30}\right)$$

$$= 1.4958$$

$$\hat{V}(\hat{T}) = N^2\hat{V}(\bar{x})$$

$$= (14.848)^2 \times 1.4985 = 329,773,602.1$$

$$\sqrt{\hat{V}(\hat{T})} = 18,159.6596$$

ดังนั้นช่วงเชื่อมั่น 90% ที่คาดว่าค่าจริงของยอดรวมจะปรากฏอยู่คือ

$$51{,}473 - Z_{.95}\sqrt{\hat{V}(\hat{T})} < T < 51{,}473 + Z_{.95}\sqrt{\hat{V}(\hat{T})}$$

$$51{,}473 - (1.645)(18{,}159.6696) < T < 51{,}473 + (1.645)(18{,}159.6696)$$

$$51{,}473 - 29{,}872.65 < T < 51{,}473 + 29{,}872.65$$

$$21{,}600.34 < T < 81{,}345.65$$

$$21{,}600 < T < 81{,}346$$

นั่นคือเราสามารถคาดหมายได้ว่าจำนวนประชากรทั้งหมดของพื้นที่ดังกล่าวจะมีอยู่ในระหว่าง 21,600 ถึง 81,346 คน ทั้งนี้การประมาณดังกล่าวนี้จะคลาดเคลื่อนไปได้ไม่เกิน 10% หรือเราสามารถกล่าวด้วยความเชื่อมั่นถึง 90% (หรือผลที่ยอมให้ผิดพลาดได้ 10%) ได้ว่า จำนวนประชากรทั้งหมดของพื้นที่ดังกล่าวจะมีอยู่ระหว่าง 21,600 คน ถึง 81,346 คน

ตัวอย่าง 2.2 คลังสินค้าแห่งหนึ่งจัดสินค้าสำรองไว้เป็น 36 หิ้ง และในการสำรวจมูลค่าสินค้าสำรองฝ่ายบริหารต้องการให้ใช้วิธีสำรวจด้วยกลุ่มตัวอย่าง ทั้งนี้เพื่อเป็นการประหยัดเวลาและกำลังแรงงาน

จากสถิติมูลค่าสินค้าสำรองครั้งล่าสุด ปรากฏว่าสินค้าแต่ละหิ้งมีมูลค่าดังนี้

29, 38, 42, 44, 45, 57, 51, 53, 53, 54, 56, 56, 56, 58, 59, 60, 60

60, 60, 61, 61, 61, 62, 64, 65, 65, 67, 68, 69, 71, 74, 77, 82, 85

$\Sigma x_i = 2{,}138 \qquad \Sigma x_i^2 = 131{,}682$

การประมาณครั้งนี้ผู้บริหารต้องการให้กะประมาณคลาดเคลื่อนไปไม่เกิน 200 บาท ด้วยโอกาสที่เป็นไปได้ 1 ใน 20 (หรือ 5%) การนี้ที่ปรึกษาแนะนำว่า ถ้าจะให้การประมาณค่ามูลค่าของสินค้าสำรองสอดคล้องกับความต้องการดังกล่าว เราสุ่มตัวอย่างสินค้ามาเพียง 12 หิ้ง ก็เป็นการเพียงพอ ท่านเห็นด้วยหรือไม่กับคำแนะนำเช่นนี้

วิธีทำ สิ่งที่ต้องการคือ จากสมการ

$$\Pr\{|\hat{T} - E(\hat{T})| \leq 200\} \geq 1 - .05 \text{ [1]} \qquad \text{..........(1)}$$

อยากทราบว่า $n = ?$

โดยอาศัยอสมภาพของเชบบิเชฟ

$$\Pr\{|\hat{T} - E(\hat{T})| \leq 200\} \geq 1 - \frac{V(\hat{T})}{(200)^2}$$

$$V(\hat{T}) = N^2V(\bar{x}) = N^2 \cdot \frac{N-n}{N}\,\frac{S^2}{n} \text{ [2]}$$

$$S^2 = \frac{1}{N-1}\left\{\sum_{i}^{N} x_i^2 - \frac{(\sum^{N} x_i)^2}{N}\right\}$$

$$= \frac{1}{35}\left(131{,}682 - \frac{4{,}571{,}044}{36}\right) = 134.53$$

ดังนั้น

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

$$= 36^2(134.53/n) - (36)(134.53)$$

$$V(\hat{T}) = \frac{174{,}351.086}{n} - 4{,}843.086$$

---

[1] อ่านว่า โอกาสที่ค่าประมาณยอดรวม (มูลค่ารวมของสินค้าสำรอง) จะคลาดเคลื่อนไปจากมูลค่าจริงในช่วง ± 200 บาท ควรสุ่มตัวอย่างมากี่ชุดจึงจะเชื่อถือได้ถึง 95% เป็นอย่างน้อย

[2] ในที่นี้ใช้ S^2 เพราะข้อมูลที่บันทึกไว้เป็นข้อมูลจากกลุ่มประชากรทั้งหมด (N = 36)

$$Pr\{|\hat{T} - E(\hat{T})| \leqslant 200\} \geqslant 1 - \frac{174,351.086}{40,000n} + \frac{4,843.086}{40,000}$$

$$\geqslant 1 - \frac{4.3587}{n} + .12108$$

$$\geqslant 1.12108 - \frac{4.3587}{n}$$

จาก (1) แสดงว่า

$$1 - V(\hat{T})/(200)^2 \geqslant .95$$

$$=> \frac{V(\hat{T})}{(200)^2} \leqslant .05$$

$$1.12108 - \frac{4.3587}{n} < .05$$

$$\frac{4.3587}{n} \geqslant 1.07107$$

$$=> n \leqslant 4.06$$ หรือ $n \leqslant 4$

2.4.2 การประมาณค่าอัตราส่วนของกลุ่มประชากร (Estimation of Population Ratio)

ดังที่ได้กล่าวมาแล้วในตอนต้น อัตราส่วนเป็นค่าที่แสดงอัตราเปรียบเทียบระหว่างตัวแปรที่แสดงคุณลักษณะทางประชากร 2 ประการ ข้อมูลที่นำมาเปรียบเทียบสามารถใช้ทั้งยอดรวมของตัวแปรทั้งสองและค่าเฉลี่ยของตัวแปรทั้งสองได้ กล่าวคือ $R = \sum_{i}^{N} x_i / \sum_{i}^{N} y_i$ หรือ $R = \frac{\overline{X}}{\overline{Y}}$ แต่ในทางทฤษฎีค่าของ R จะเป็นพารามิเตอร์ที่เราไม่ทราบค่าแต่สนใจ

ใคร่รู้ ทั้งนี้เพราะ R เป็นฟังก์ชันของ T(x) และ T(y) หรือ $\bar{X}$ และ $\bar{Y}$ ซึ่งเป็นค่าที่ต่างก็เป็นพารามิเตอร์ที่เราไม่ทราบค่า ดังนั้นในทางปฏิบัติเราสามารถประมาณค่าของ R ได้โดยอาศัยข้อมูลข้อสนเทศจากกลุ่มตัวอย่าง กล่าวคือ $\hat{R} = \sum_{i}^{n} x_i / \sum_{i}^{n} y_i = \frac{\bar{x}}{\bar{y}}$ และถ้าขนาดตัวอย่าง n มีขนาดใหญ่พอ $\hat{R}$ จะเป็นตัวประมาณค่าที่ปราศจากอคติ (Unbiased Estimator) ของ R แต่ถ้า n มีขนาดเล็ก $\hat{R}$ จะเป็นตัวประมาณค่าที่มีอคติเล็กน้อย (Slightly Bias)

ทฤษฎี 2.4 ถ้า x และ y เป็นตัวแปร 2 ตัวที่เราสนใจจากกลุ่มตัวอย่างขนาด n ชุดเดียวกัน และ n มีขนาดใหญ่ โดยที่กลุ่มตัวอย่างดังกล่าวได้รับเลือกมาโดยวิธี SRS แล้ว

1. $\hat{R} = \frac{\bar{x}}{\bar{y}}$ จะเป็นตัวประมาณค่าที่ปราศจากอคติ (Unbiased Estimator) ของ R กล่าวคือ $E(\hat{R}) = R$

2. $V(\hat{R}) = \frac{1}{\bar{Y}^2}\frac{N-n}{N}\frac{S_d^2}{n}$ โดยที่ $S_d^2 = \frac{1}{N-1}\sum_{i}^{N}(x_i - Ry_i)^2$

และ $\hat{V}(\hat{R}) = \frac{1}{\bar{y}^2}\frac{N-n}{N}\frac{s_d^2}{n}$ โดยที่ $s_d^2 = \frac{1}{n-1}\sum_{i}^{n}(x_i - \hat{R}y_i)^2$

$\hat{V}(\hat{R})$ เป็นตัวประมาณค่าของ $V(\hat{R})$

ข้อสังเกต ขอให้สังเกตว่าโครงสร้างของสูตร $V(\hat{R})$ จะต้องคล้ายคลึงกันกับโครงสร้างของสูตร $V(\bar{x})$ ในทฤษฎี 2.2 แตกต่างกันเพียงเล็กน้อยเฉพาะ $V(\hat{R})$ มี $\bar{Y}^2$ หารตลอด

พิสูจน์ 1. $E(\hat{R}) = R$

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

เมื่อ $n \to \infty$ ค่าของ $\bar{x}$ และ $\bar{y}$ จะไม่ต่างไปจากค่าจริง $\bar{X}$ และ $\bar{Y}$ มากนัก ดังนั้นเราจึงสามารถใช้ $\bar{Y}$ แทนที่ของ $\bar{y}$ หรือสับเปลี่ยนกันได้โดยประมาณ ดังนั้น

$$\hat{R} - R \simeq \frac{\bar{x} - R\bar{y}}{\bar{Y}}$$ [1]

$$E(\hat{R} - R) \simeq E\left(\frac{\bar{x} - R\bar{y})}{\bar{Y}}\right) \simeq \frac{1}{\bar{Y}} \{E(\bar{x}) - RE(\bar{y})\}$$

$$\simeq \frac{1}{\bar{Y}} (\bar{X} - R\bar{Y})$$

$$\simeq \frac{1}{\bar{Y}} (\bar{X} - \frac{\bar{X}}{\bar{Y}} \cdot \bar{Y}) \simeq 0$$

นั่นคือ $E(\hat{R} - R) \simeq 0$

แต่ $E(\hat{R} - R) = E(\hat{R}) - R$

ดังนั้น $E(\hat{R}) - R \simeq 0$ หรือ $E(\hat{R}) \simeq R$

นั่นคือ ถ้าใช้ขนาดตัวอย่าง n ให้ใหญ่พอแล้ว $\hat{R}$ จะเป็นตัวประมาณค่าที่ปราศจากอคติของ R หรือนัยหนึ่งเราสามารถใช้ $\hat{R} = \frac{\bar{x}}{\bar{y}}$ เป็นตัวประมาณค่าของ $R = \frac{\bar{X}}{\bar{Y}}$ ได้

2. $V(\hat{R}) = E(\hat{R} - R)^2 = E\left(\frac{\bar{x} - R\bar{y}}{\bar{Y}}\right)^2 \simeq \frac{1}{\bar{Y}^2} E(\bar{x} - R\bar{y})^2$

ให้ $d_i = x_i - Ry_i$; $i = 1,2,\ldots,n$ จะพบว่า $\bar{d} = \frac{1}{n} \sum_i^n d_i = \bar{x} - R\bar{y}$

โดยที่ $\bar{D} = \frac{1}{N} \sum_i^N d_i = \bar{X} - R\bar{Y} = \bar{X} - \frac{\bar{X}}{\bar{Y}} \cdot \bar{Y} = 0 = E(\bar{d})$

---

[1] อ่าน กฎการใช้กลุ่มตัวอย่างขนาดใหญ่ (Law of Large Number) จากทฤษฎีสถิติ 2 โดยผู้เขียนคนเดียวกัน

$$V(\hat{R}) \cong \frac{1}{\bar{Y}^2} E(\bar{x} - R\bar{y})^2 \simeq \frac{1}{\bar{Y}^2} E\{(\bar{x} - R\bar{y}) - 0\}^2$$

$$\simeq \frac{1}{\bar{Y}^2} E\{\bar{d} - E(\bar{d})\}^2$$

$$\simeq \frac{1}{\bar{Y}^2} V(\bar{d})$$

โดยอาศัยทฤษฎี 1.2 จะพบว่า

$$V(\bar{d}) = \frac{N-n}{N} \frac{S_d^2}{n} \text{ โดยที่ } S_d^2 = \frac{1}{N-1} \sum_i^N (d_i - \bar{D})^2$$

ดังนั้น $V(\hat{R}) = \frac{1}{\bar{Y}^2} V(\bar{d}) = \frac{1}{\bar{Y}^2} \frac{N-n}{N} \cdot \frac{S_d^2}{n}$

$$= \frac{1}{\bar{Y}^2} \frac{N-n}{N} \cdot \frac{1}{n} \cdot \frac{1}{N-1} \sum_i^N (x_i - Ry_i)^2$$

$$= \frac{1}{n\bar{Y}^2} \frac{N-n}{N} \cdot \frac{1}{N-1} \sum_i^N (x_i - Ry_i)^2$$

สำหรับค่าประมาณของ $V(\hat{R})$ คือ $\hat{V}(\hat{R})$ โดยที่

$$\hat{V}(\hat{R}) = \frac{1}{\bar{Y}^2} \hat{V}(\bar{d}) = \frac{1}{\bar{Y}^2} \cdot \frac{N-n}{N} \frac{s_d^2}{n}$$

แต่ $s_d^2 = \frac{1}{n-1} \sum_i^n (d_i - \bar{d})^2 \simeq \frac{1}{n-1} \sum_i^n \{(x_i - \hat{R}y_i) - (\bar{x} - \hat{R}\bar{y})\}^2$ ซึ่ง