psu ssu และ tsu มาเป็นจำนวนมาก หรือ ℓ, m_i และ n_{ij} มีค่าสูงแล้ว $\frac{S_b^2}{\ell}$ $\frac{S_i^2}{m_i}$ และ $\frac{S_{ij}^2}{n_{ij}}$ จะมีค่าต่ำลง ซึ่งชี้ให้เห็นว่าถ้าจะให้งานมีคุณภาพสูงเราควรสุ่ม psu, ssu และ tsu มาเป็นจำนวนมาก แต่สถานการณ์เช่นนี้มักเป็นไปได้ยากในทางปฏิบัติเพราะทำให้สิ้นเปลืองมากเกินไปและงบประมาณจำกัด หนทางที่เหมาะสมที่สุดก็คือควรเลือกดูเองว่าควรสุ่มใครในระหว่าง psu ssu และ tsu มามากกว่ากันจึงจะมีผลใน $V(\hat{T})$ ลดลงเท่านั้น เมื่อพิจารณา S_b^2, S_i^2 และ S_{ij}^2 จะพบว่าโดยธรรมชาติแล้ว S_b^2 จะมีค่าที่สูงที่สุด[1] ซึ่ง $V(\hat{T})$ จะมีค่าสูงต่ำเพียงใดย่อมขึ้นอยู่กับอิทธิพลของ S_b^2 เป็นสำคัญ ดังนั้นถ้าเราสามารถลดความผันแปรระหว่าง psu ลงได้ $V(\hat{T})$ ย่อมลดลงได้ซึ่งความผันแปรระหว่าง psu จะลดลงได้ก็ต่อเมื่อเราสุ่มตัวอย่าง psu มามาก หรือ ℓ มีค่าสูง เพราะเมื่อ ℓ มีค่าสูง $\frac{S_b^2}{\ell}$ จะมีค่าต่ำลง อันจะยังผลให้ $V(\hat{T})$ ลดลง ดังนั้นในทางปฏิบัติเราจึงนิยมสุ่มตัวอย่าง psu มาเป็นจำนวนมาก[2] ซึ่งเมื่อสุ่ม psu มาเป็นจำนวนมากเสียแล้ว จำนวน ssu และ tsu ก็ย่อมลดน้อยลงโดยอัตโนมัติทั้งนี้เพราะมีข้อจำกัดเรื่องงบประมาณเป็นตัวควบคุม

---

[1] จากบทที่ 2 และ 3 นักศึกษาจะพบว่า $V(\hat{T})$ จะมากกว่า $V(\bar{x})$ เสมอ ทั้งนี้เพราะ T_i มีค่าสูงกว่า x_i ความผันแปรระหว่าง T_i จึงสูงกว่าระหว่าง x_i ในทำนองเดียวกัน ในที่นี้ T_i ย่อมมีค่าสูงกว่า T_{ij} และ x_{ijk} ดังนั้นความผันแปรระหว่าง T_i จึงสูงที่สุด (ดูตัวอย่าง 5.8)

[2] สิ่งที่ควรคำนึงว่าจะสุ่ม psu มามากหรือน้อยเพียงใดอีกประการหนึ่งก็คือความคล้ายคลึงกันในระหว่างหน่วยภายใน psu ถ้าหน่วยภายใน psu คล้ายคลึงกันมากเราจะต้องสุ่มตัวอย่าง psu มามาก ถ้าหน่วยภายใน psu แตกต่างกันมากเราจะสุ่ม psu มาน้อย ทั้งนี้เพราะถ้าหน่วยภายใน psu คล้ายคลึงกันมากความแตกต่างระหว่าง psu จะมีมาก ถ้าหน่วยภายใน psu ต่างกันมากความแตกต่างระหว่าง psu จะมีน้อย ตัวอย่างเช่น นักเรียนซึ่งประกอบด้วยเด็กชายและเด็กหญิง ถ้าเราจัดเด็กชายไว้ห้องหนึ่งและเด็กหญิงไว้อีกห้องหนึ่ง ความแตกต่างภายในชั้นเรียนจะมีน้อยขณะที่ความแตกต่างระหว่างชั้นเรียนระหว่างห้องมีมาก ถ้าจัดชั้นเรียน 2 ชั้น แต่ละชั้นมีทั้งเด็กชายและเด็กหญิงความแตกต่างระหว่างชั้นจะมีน้อยเพราะความแตกต่างภายในชั้นมีมาก ซึ่งจะมีเห็นว่าในกรณีแรกเราจำเป็นต้องสุ่มตัวอย่างนักเรียนจากทั้งสองห้อง ขณะที่เราสุ่มตัวอย่างนักเรียนมาเพียงห้องเดียวสำหรับกรณีที่ 2

อนึ่ง สำหรับสูตร $V(\hat{T})$ หรือ $\hat{V}(\hat{T})$ นั้นเราสามารถลดรูปสู่รูปที่ง่าย ๆ ได้ ถ้าหากว่าจำนวน psu มีปริมาณสูงมากเมื่อเทียบต่อจำนวนตัวอย่าง psu หรือ $L >> \ell$

ดังนี้

เมื่อ $\ell << L$

$$\because \hat{V}(\hat{T}) = L^2 \frac{L-\ell}{L} \frac{S_b^2}{\ell} + \frac{L}{\ell} \sum_{i}^{L} M_i^2 \frac{M_i - m_i}{M_i} \frac{S_i^2}{m_i}$$

$$+ \frac{L}{\ell} \sum_{i}^{\ell} \frac{M_i}{m_i} \sum_{j}^{m_i} N_{ij}^2 \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{s_{ij}^2}{n_{ij}}$$

ดังนั้น

$$\hat{V}(\hat{T}) \simeq \left(\frac{L}{\ell}\right)^2 \left\{ \frac{L-\ell}{L} \ell s_b^2 + \frac{\ell}{L} \sum_{i}^{\ell} M_i^2 \frac{M_i - m_i}{M_i} \frac{s_i^2}{m_i}\right.$$

$$\left. + \frac{\ell}{L} \sum_{i}^{\ell} \frac{M_i}{m_i} \sum_{j}^{m_i} N_{ij}^2 \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{s_{ij}^2}{n_{ij}} \right\}$$

เมื่อ $L >> \ell$ ดังนั้น $\ell/L \cong 0$ และ $(L-\ell)/L \cong 1$

ดังนั้น

$$\hat{V}(\hat{T}) \simeq (L/\ell)^2 \ell s_b^2$$

$$= L^2 \frac{s_b^2}{\ell}$$

ตัวอย่าง 5.8 ในภาคกลางมีโรงเรียนมัธยมศึกษาทั้งสิ้น $L = 2500$ โรง สุ่มโรงเรียนมาเป็นตัวอย่าง $\ell = 5$ โรง จากแต่ละแห่งผู้วิจัยเตรียมกรอบตัวอย่างโดยจดรายชื่อของนักเรียนไว้ในสมุด โรงเรียนละ 1 เล่ม รวม 5 เล่ม ให้ M_i คือจำนวนหน้ากระดาษของสมุดรายชื่อ (กรอบตัวอย่าง) ของโรงเรียนตัวอย่างที่ i และ N_{ij} คือจำนวนรายชื่อนักเรียนในหน้ากระดาษที่ j ของบัญชีรายชื่อสำหรับโรงเรียนที่ i จากกรอบตัวอย่าง สุ่มหน้ากระดาษรายชื่อโรงเรียนมาโรงเรียนละ 2 หน้า แต่ละหน้ากระดาษสุ่มรายชื่อนักเรียนมาหน้าละ 10 ชื่อ จากนั้นจึงสัมภาษณ์นักเรียนที่ตกเป็นตัวอย่างจำนวนครั้งที่นักเรียนแต่ละคนขาดโรงเรียนใน 1 เดือน

ผลการบันทึกข้อมูลปรากฏดังนี้ ตารางจะแสดงถึงจำนวนหน้าและจำนวนรายชื่อในกรอบตัวอย่าง และผลการบันทึกข้อมูล

โรงเรียน (ℓ)	จำนวนหน้าของสมุดบัญชีรายชื่อนักเรียน (M_i)	จำนวนรายชื่อนักเรียนในหน้ากระดาษ (N_{ij})	$\sum_{k}^{n_{ij}} x_{ijk}$
1	40	30	20
		30	25
2	20	40	20
		40	30
3	120	40	25
		40	18
4	300	50	15
		50	25
5	220	40	20
		40	22

ก. จงประมาณยอดรวมจำนวนครั้งที่นักเรียนมัธยมศึกษาในจังหวัดภาคกลางขาดโรงเรียนใน 1 เดือน

ข. จงคำนวณหา $\hat{V}(\hat{T})$ และช่วงเชื่อมั่น 2σ ของยอดรวมประชากร T

วิธีทำ จากข้อมูลเราสามารถเตรียมตารางวิเคราะห์ได้ดังนี้ อนึ่งตามโจทย์ $m_i = 2, i=1,2,\ldots,5, n_{ij}=10; i=1,2,\ldots,5, j=1,2,$ หรือ $\overline{m} = 2, \overline{n} = 10$

โรงเรียน ℓ	หน้ากระดาษ M_i	รายชื่อ N_{ij}	$\sum_{k}^{n_{ij}} x_{ijk}$	$\bar{\bar{x}}_{ij}$	$\hat{T}_{ij} = N_{ij}\bar{\bar{x}}_{ij}$	$\hat{\bar{T}}_i = \frac{1}{m_i}\sum_{j}^{m_i}\hat{T}_{ij}$	$\hat{T}_i = M_i\hat{\bar{T}}_i$
1	40	30	20	2.0	60	67.5	2700
		30	25	2.5	70		
2	20	40	20	2.0	80	100	2000
		40	30	3.0	120		
3	120	40	25	2.5	100	86	10320
		40	18	1.8	72		
4	300	50	15	1.5	75	100	30000
		50	25	2.5	125		
5	220	40	20	2.0	80	84	18480
		40	22	2.2	88		
รวม							63500

ก. จากตาราง $\hat{\bar{T}} = \frac{1}{\ell}\sum_{i}^{\ell}\hat{T}_i = 63500/5 = 12700$

ดังนั้น $\hat{T} = L\hat{\bar{T}} = 2500 \times 12700 = 31750000$

นั่นคือ คาดว่าใน 1 เดือนจะมีนักเรียนในจังหวัดภาคกลางขาดโรงเรียนทั้งสิ้นประมาณ 31750000 ครั้ง

ข. เนื่องจาก $\ell = 5$ ขณะที่ $L = 2500$ ซึ่งจะเห็นได้ว่า $L >> \ell$ ดังนั้นสูตร $\hat{V}(\hat{T})$ จะลดรูปลงเป็น

$$\hat{V}(\hat{T}) \simeq L^2 \frac{s_b^2}{\ell}$$

โดยที่ $s_b^2 = \frac{1}{\ell - 1} \sum_{i}^{\ell} (T_i - \hat{\bar{T}})^2$

$$= \frac{1}{\ell-1} \left\{ \sum_{i}^{\ell} \hat{T}_i^2 - \frac{1}{\ell} \left(\sum_{i}^{\ell} \hat{T}_i \right)^2 \right\}$$

$$= \frac{1}{4} \left\{ (2700^2 + 2000^2 + \ldots + 18480^2) - \frac{63500^2}{5} \right\}$$

$$= \frac{1}{4} (1359302800 - 806450000)$$

$$= 138213200$$

ดังนั้น $\hat{V}(\hat{T}) = L^2 \frac{s_b^2}{\ell} = (2500)^2 \cdot \frac{138213200}{5} = 1.72766500 \times 10^{14}$

$$s_{\hat{T}} = 13144067.10269$$

ช่วงเชื่อมั่น 2σ ของ T คือ $\hat{T} - 1.96 s_{\hat{T}} < T < \hat{T} + 1.96 s_{\hat{T}}$

นั่นคือ

$$5987628.47873 < T < 57512371.52127$$

หรือสามารถกล่าวได้อีกนัยหนึ่งว่าด้วยความเชื่อมั่น 95% ยอดรวมจริงของจำนวนครั้งที่นักเรียนชั้นมัธยมศึกษาในจังหวัดภาคกลางจะขาดเรียนในเดือนหนึ่ง ๆ นั้นปรากฏอยู่ในช่วง 5987629 ถึง 57512312 ครั้ง

ค. ข้อนี้จะทำได้ภายหลังจากการศึกษาในตอน 5.3.3

5.3.3 การประมาณค่าเฉลี่ยประชากร $\bar{\bar{X}}$

ค่าเฉลี่ยประชากรคือ $\bar{\bar{X}} = \frac{1}{N} \sum_{i}^{L} \sum_{j}^{M_i} \sum_{k}^{N_{ij}} x_{ijk}$ เมื่อ $N = \sum_{i}^{L} \sum_{j}^{M_i} N_{ij}$

หรือ $\bar{\bar{X}} = T/N$ เมื่อ T คือยอดรวมประชากร $T = \sum_{i}^{L} \sum_{j}^{M_i} \sum_{k}^{N_{ij}} x_{ijk}$

การประมาณค่าเฉลี่ยสามารถกระทำได้โดยง่ายแต่นำขนาดประชากร N ไปหารค่าประมาณของยอดรวมคือ $\hat{T}$ ที่ได้จากตอน 5.3.2 ดังนั้น

$$\hat{\bar{\bar{X}}} = \frac{\hat{T}}{N} = \frac{1}{N}\frac{L}{\ell}\sum_{i}^{\ell}\frac{M_i}{m_i}\sum_{j}^{m_i}\frac{N_{ij}}{n_{ij}}\sum_{k}^{n_{ij}} x_{ijk}$$

และ
$$V(\hat{\bar{\bar{X}}}) = \frac{1}{N^2}V(\hat{T})$$

$$= \frac{1}{N^2}\{L^2\frac{L-\ell}{L}\frac{S_b^2}{\ell} + \frac{L}{\ell}\sum_{i}^{\ell}M_i^2\frac{M_i-m_i}{M_i}\frac{S_i^2}{m_i} +$$

$$\frac{L}{\ell}\sum_{i}^{\ell}\frac{M_i}{m_i}\sum_{j}^{M_i}N_{ij}^2\frac{N_{ij}-n_{ij}}{N_{ij}}\frac{S_{ij}^2}{n_{ij}}\}$$

$$\hat{V}(\hat{\bar{\bar{x}}}) = \frac{1}{N^2}\{L^2\frac{L-\ell}{L}\frac{s_b^2}{\ell} + \frac{L}{\ell}\sum_{i}^{\ell}M_i^2\frac{M_i-m_i}{M_i}\frac{s_i^2}{m_i} +$$

$$\frac{L}{\ell}\sum_{i}^{\ell}\frac{M_i}{m_i}\sum_{j}^{m_i}N_{ij}^2\frac{N_{ij}-n_{ij}}{N_{ij}}\frac{s_{ij}^2}{n_{ij}}\}$$

ปัญหาทางปฏิบัติที่พบอยู่เสมอในการประมาณค่ายอดรวมและค่าเฉลี่ยประชากรคือการไม่ทราบค่าที่แน่นอนของ N_{ij} และ M_i เพราะเมื่อใดที่ไม่ทราบค่า N_{ij}; $i = 1,2,\ldots,L$; $j = 1,2,\ldots,M_i$ แล้วย่อมหมายถึงการไม่ทราบค่าของ N ด้วยทั้งนี้เพราะ $N = \sum_{i}^{L}\sum_{j}^{M_i} N_{ij}$ วิธีหนึ่งที่สามารถแก้ปัญหานี้ได้ก็คือการวางแผนจัดให้ ssu มีขนาดเท่ากันหรือใกล้เคียงกัน คือจัดให้ $N_{ij} = \bar{N}$ เช่นแบ่งพื้นที่สำรวจออกเป็นหลาย ๆ psu แต่ละ psu ประกอบด้วย ssu ขนาดเดียวกัน ตัวอย่างเช่น แบ่งพื้นที่ในอำเภอออกเป็นส่วน ๆ แต่ละส่วนประกอบด้วยครัวเรือนส่วนละ 1000 ครัวเรือน เป็นต้น ซึ่งในกรณีเช่นนี้จะพบว่าเราไม่อาจใช้เขตการปกครอง (จังหวัด อำเภอ ตำบล หมู่บ้าน) เป็นเกณฑ์ในการจำแนกประชากรหากเป็นจำแนกประชากรตามความเหมาะสมที่เรากำหนดขึ้นเอง วิธีนี้อาจประสบปัญหาในทางปฏิบัติที่สำคัญคือการสร้างกรอบตัวอย่าง เพราะจะต้องสร้างกรอบตัวอย่างขึ้นมาใหม่ทั้งหมดซึ่งย่อมสิ้นเปลืองงบประมาณและเวลาไปมิใช่น้อยผิดกับวิธีที่ใช้เขตการปกครองเป็นตัวจำแนกประชากร ซึ่งกรอบตัวอย่างมีให้ไว้แล้ว เพียงปรับปรุงอีกเพียงเล็กน้อย วิธีหนึ่งที่จะช่วยลดภาระในส่วนนี้ลงได้คือการจำแนกประชากรตามเขตเลือกตั้ง เพราะเขตเลือกตั้งเป็นกรอบตัวอย่างที่สมบูรณ์และมักจะมีขนาดเท่ากันหรือเป็นพหุคูณของกันและกัน

ในกรณีที่ ssu มีขนาดเดียวกันคือประกอบด้วย tsu เท่ากันเท่ากับ $\overline{N}$ หน่วยจะพบว่า $N = \sum_{i}^{L} \sum_{j}^{M_i} \overline{N} = \overline{N} \sum_{i}^{\ell} M_i$ ค่า M_i มักไม่เป็นปัญหาเพราะ M_i มักจะมีค่าน้อย (M_i คือจำนวน ssu ใน psu ตัวอย่างที่ i เช่นจำนวนห้องเรียนชั้น ป.5 ในโรงเรียนตัวอย่างที่ 1 เท่ากับ 6 โรง)

อนึ่ง ในหลายสถานะการณ์เราสามารถสุ่ม ssu จากแต่ละ psu และ tsu จากแต่ละ ssu มาได้ในจำนวนเท่า ๆ กัน[1] ซึ่งในกรณีเช่นนี้การวิเคราะห์ย่อมเป็นไปได้ง่ายกว่ากรณีทั่วไป โดยเฉพาะอย่างยิ่งในเรื่องการกำหนดขนาดตัวอย่าง นั่นคือ เมื่อ $m_i = \overline{m}$ และ $n_{ij} = \overline{n}$

$$\hat{T} = L\hat{\overline{T}} = \frac{L}{\ell} \sum_{i}^{\ell} \frac{M_i}{\overline{m}} \sum_{j}^{\overline{m}} \frac{N_{ij}}{\overline{n}} \sum_{k}^{\overline{n}} x_{ijk}$$

$$\hat{\overline{\overline{X}}} = \hat{T}/N$$

$$\hat{V}(\hat{T}) = L^2 \frac{L-\ell}{L} \frac{s_b^2}{\ell} + \frac{L}{\ell} \sum_{i}^{\ell} M_i^2 \frac{M_i - \overline{m}}{M_i} \frac{s_i^2}{\overline{m}}$$

$$+ \frac{L}{\ell} \sum_{i}^{\ell} \frac{M_i}{\overline{m}} \sum_{j}^{\overline{m}} N_{ij}^2 \frac{N_{ij} - \overline{n}}{N_{ij}} \frac{s_{ij}^2}{\overline{n}}$$

5.3.4 การจัดสรรขนาดตัวอย่าง

ให้ C_1, C_2, C_3 เป็นค่าใช้จ่ายในการเลือก psu ssu และ tsu โดยมี Cost Function ดังนี้ $C = C_0 + C_1\ell + C_2\ell\overline{m} + C_3\ell\overline{m}\overline{n}$

หรือ $C' = C_1\ell + C_2\ell\overline{m} + C_3\ell\overline{m}\overline{n}$ = งบประมาณในการสำรวจ

จาก $\hat{V}(\hat{T}) = L^2 \frac{L-\ell}{L} \frac{s_b^2}{\ell} + \frac{L}{\ell} \sum_{i}^{\ell} M_i^2 \frac{M_i - \overline{m}}{M_i} \frac{S_i^2}{\overline{m}}$

$$+ \frac{L}{\ell} \sum_{i}^{\ell} \frac{M_i}{\overline{m}} \sum_{j}^{\overline{m}} N_{ij}^2 \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{s_{ij}^2}{\overline{n}}$$

ถ้าถือว่า $L >> \ell$, $M_i >> \overline{m}$ และ $N_{ij} >> \overline{n}$ ซึ่งเป็นจริงเสมอในทางปฏิบัติ

---

[1] อ่านตอน 5.2.6

$$=> \hat{V}(\hat{T}) = L^2 \frac{s_b^2}{\ell} + L \sum_i^{\ell} M_i^2 \frac{s_i^2}{\ell.\overline{m}} + L \sum_i^{\ell} M_i \sum_j^{\overline{m}} N_{ij}^2 \frac{s_{ij}^2}{\ell \overline{m}\overline{n}}$$

ให้ $\frac{1}{L} \sum_i^{\ell} M_i^2 s_i^2 = s_{2i}^2$ และ $\frac{1}{L} \sum_i^{\ell} M_i \sum_j^{\overline{m}} N_{ij}^2 s_{ij}^2 = s_{3j}^2$

$$=> \hat{V}(\hat{T}) = L^2 \left(\frac{s_b^2}{\ell} + \frac{s_{2i}^2}{\ell\overline{m}} + \frac{s_{3j}^2}{\ell\overline{m}\overline{n}} \right)$$

โดยอาศัย Lagrange Multiplier Method

$$=> F = \hat{V}(\hat{T}) + \lambda(C' - C_1\ell - C_2\ell\overline{m} - C_3\ell\overline{m}\overline{n})$$

$$= L^2 \left(\frac{s_b^2}{\ell} + \frac{s_{2i}^2}{\ell\overline{m}} + \frac{s_{3j}^2}{\ell\overline{m}\overline{n}} \right) + \lambda C' - \lambda C_1\ell - \lambda C_2\ell\overline{m} - \lambda C_3\ell\overline{m}\overline{n}$$

$$\frac{\partial F}{\partial \ell} = 0 => -L^2 \frac{s_b^2}{\ell^2} - L^2 \frac{s_{2i}^2}{\ell^2\overline{m}} - L^2 \frac{s_{3j}^2}{\ell^2\overline{m}\overline{n}} - \lambda C_1 - \lambda C_2\overline{m} - \lambda C_3\overline{m}\overline{n} = 0 \quad ...(1)$$

$$\frac{\partial F}{\partial \overline{m}} = 0 => -L^2 \frac{s_{2i}^2}{\ell\overline{m}^2} - L^2 \frac{s_{3j}^2}{\ell\overline{n}\overline{m}^2} - \lambda C_2\ell - \lambda C_3\ell\overline{n} = 0 \quad ...(2)$$

$$\frac{\partial F}{\partial \overline{n}} = 0 => -L^2 s_{3j}^2/\ell\overline{m}\overline{n}^2 - \lambda C_3\ell\overline{m} = 0 \quad ...(3)$$

$$\frac{\partial F}{\partial \lambda} = 0 => C' = C_1\ell + C_2\ell\overline{m} + C_3\ell\overline{m}\overline{n} = 0 \quad ...(4)$$ [1]

$$(3) \times \ell\overline{m}\overline{n}^2 => L^2 s_{3j}^2 + \lambda C_3(\ell\overline{m}\overline{n})^2 = 0 \quad ...(5)$$

$$(2) \times \ell\overline{n}\overline{m}^2 => L^2\overline{n}s_{2i}^2 + \{L^2 s_{3j}^2 + \lambda C_3(\ell\overline{m}\overline{n})^2\} + \lambda C_2\overline{n}(\ell\overline{m})^2 = 0 \quad ...(6)$$

แทน (5) ใน (6) $=> L^2\overline{n}s_{2i}^2 + \lambda C_2\overline{n}(\ell\overline{m}) = 0$

$$=> L^2 s_{2i}^2 + \lambda C_2(\ell\overline{m})^2 = 0 \quad ...(7)$$

---

1. ความจริงแล้วเราไม่จำเป็นต้อง differentiate F wrt.λ เพราะถึงแม้จะไม่ differentiate เราก็อาศัย Cost Function ได้ขอให้สังเกตว่า $\frac{\partial F}{\partial \lambda} = 0$ จะได้ Cost Function

$(1) \times \ell^2\overline{m}\overline{n}$

$$L^2\overline{m}\overline{n}s_b^2 + L^2\overline{n}s_{2i}^2 + L^2s_{3j}^2 + \lambda C_1\ell^2\overline{m}\overline{n} + \lambda C_2\overline{n}(\ell\overline{m})^2 + \lambda C_3(\ell\overline{m}\overline{n})^2 = 0$$

$$L^2\overline{m}\overline{n}s_b^2 + \{L^2\overline{n}s_{2i}^2 + \lambda C_2\overline{n}(\ell\overline{m})^2\} + \{L^2s_{3j}^2 + \lambda C_3(\ell\overline{m}\overline{n})^2\} +$$

$$\lambda C_1\ell^2\overline{m}\overline{n} = 0 \quad \ldots(8)$$

แทน (5) และ (7) ใน (8)

$$L^2\overline{m}\overline{n}s_b^2 + \lambda C_1\ell^2\overline{m}\overline{n} = 0$$

$$L^2s_b^2 + \lambda C_1\ell^2 = 0 \quad \ldots(9)$$

เพื่อให้ดูแล้วเข้าใจง่ายขอนำสมการ (5), (7) และ (9) มาเรียงลำดับกันใหม่ดังนี้

$$L^2s_{3j}^2 = -\lambda C_3(\ell\overline{m}\overline{n})^2 \quad \ldots(7)$$

$$L^2s_{2i}^2 = -\lambda C_2(\ell\overline{m})^2 \quad \ldots(8)$$

$$L^2s_b^2 = -\lambda C_1\ell^2 \quad \ldots(9)$$

$(7) \div (8)$ $\quad s_{3j}^2/s_{2i}^2 = C_3\overline{n}^2/C_2$

$$\overline{n}^2 = C_2s_{3j}^2/C_3s_{2i}^2$$

$$\overline{n} = \frac{s_{3j}}{s_{2i}}\sqrt{\frac{C_2}{C_3}} \; ***$$

$(8) \div (9)$ $\quad s_{2i}^2/s_b^2 = C_2\overline{m}^2/C_1$

$$\overline{m} = \frac{s_{2i}}{s_b}\sqrt{\frac{C_1}{C_2}} \; ***$$

แทนค่า $\overline{m}$ และ $\overline{n}$ ใน (4)

$$C' = \ell\{C_1 + C_2(\frac{s_{2i}}{s_b}\sqrt{\frac{C_1}{C_2}}) + C_3(\frac{s_{2i}}{s_b}\sqrt{\frac{C_1}{C_2}})(\frac{s_{3j}}{s_{2i}}\sqrt{\frac{C_2}{C_3}})\}$$

$$= \ell\{C_1 + \frac{s_{2i}}{s_b}\sqrt{C_1C_2} + \frac{s_{3j}}{s_b}\sqrt{C_1C_3}\}$$

ดังนั้น

$$\ell = C'/\{C_1 + \frac{s_{2i}}{s_b}\sqrt{C_1C_2} + \frac{s_{3j}}{s_b}\sqrt{C_1C_3}\}$$

ขณะเดียวกัน

$$\bar{m} = \frac{s_{2i}}{s_b}\sqrt{\frac{C_1}{C_2}}$$

และ $$\bar{n} = \frac{s_{3j}}{s_{2i}}\sqrt{\frac{C_2}{C_3}}$$

เมื่อ $$s_{2i}^2 = \frac{1}{L}\sum_i^{\ell} M_i^2 s_i^2;\ s_{3j}^2 = \frac{1}{L}\sum_i^{\ell} M_i \sum_j^{\bar{m}} N_{ij}^2 s_{ij}^2$$

$$s_b^2 = \frac{1}{\ell - 1}\sum_i^{\ell}(\hat{T}_i - \hat{T})^2,\ s_i^2 = \frac{1}{\bar{m}-1}\sum_j^{\bar{m}}(\hat{T}_{ij} - \hat{\bar{T}}_i)^2$$

และ $$s_{ij}^2 = \frac{1}{\bar{n}-1}\sum_k^{\bar{n}}(x_{ijk} - \bar{\bar{x}}_{ij})^2$$

$$\bar{\bar{x}}_{ij} = \frac{1}{n_{ij}}\sum_k^{n_{ij}} x_{ijk},\ \hat{T}_{ij} = N_{ij}\bar{\bar{x}}_{ij},\ \hat{\bar{T}}_i = \frac{1}{\bar{m}}\sum_j^{\bar{m}}\hat{T}_{ij};\ i = 1,2,\ldots,\ell$$

$$\hat{T}_i = M_i\hat{\bar{T}}_i,\ \hat{T} = \frac{1}{\ell}\sum_i^{\ell}\hat{T}_i$$

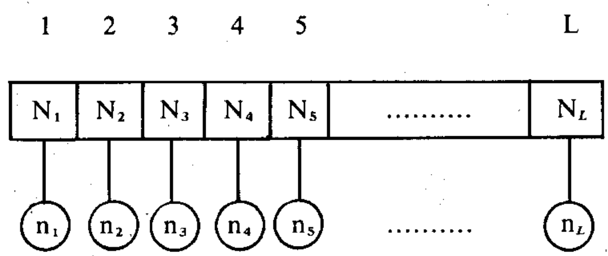
5.4 Stratified Cluster Sampling (SCS)

แผนสำรวจแบบ SCS คือ แผนสำรวจที่เกิดจากการผสมกันระหว่างแผนสำรวจแบบแบ่งชั้นภูมิกับแผนสำรวจแบบ CS และในที่นี้จะกล่าวเฉพาะในกรณีของการผสมระหว่างแผนสำรวจแบบแบ่งชั้นภูมิกับแผนสำรวจแบบ 2CS เท่านั้น กรณีอื่นที่นอกเหนือไปจากนี้จะขอเว้นไว้เป็นภาระของนักศึกษา อย่างใดก็ตามถ้านักศึกษามีความเข้าใจเทคนิคและวิธีการของ 2CS 3CS 4CS หรืออื่น ๆ ดี SCS ย่อมเป็นเรื่องง่ายเพราะ SCS ย่อมเป็นเพียงกรณีเฉพาะของเทคนิคดังกล่าว

ก่อนที่จะศึกษาถึง SCS ต่อไปขอสรุปลักษณะของแผนสำรวจที่แตกต่างกันระหว่างแผนสำรวจแบบแบ่งชั้นภูมิกับและ CS ดังนี้คือ

ก. ในแผนสำรวจแบบแบ่งชั้นภูมิเราจะจำแนกประชากรออกเป็นส่วน ๆ เรียกว่าชั้นภูมิ โดยที่หน่วยสำรวจที่มีลักษณะคล้ายคลึงกันจะถูกจัดในชั้นภูมิเดียวกัน การสุ่มตัวอย่าง

จะสุ่มหน่วยสำรวจมาจากแต่ละชั้นภูมิในปริมาณที่จัดสรรไว้ตามความเหมาะสม การประมวลผลให้นำผลลัพธ์จากค่าประมาณของแต่ละชั้นภูมิมารวมกัน ดังภาพและสูตรต่อไปนี้



$$\bar{x}_1 \quad \bar{x}_2 \quad \bar{x}_3 \quad \bar{x}_4 \quad \bar{x}_5 \qquad \bar{x}_L$$

$$\hat{T}_1 \quad \hat{T}_2 \quad \hat{T}_3 \quad \hat{T}_4 \quad \hat{T}_5 \quad \cdot \qquad \hat{T}_L \;;\; \hat{T}_h = N_h\bar{x}_h$$

ดังนั้น $$\hat{T} = \hat{T}_1 + \hat{T}_2 + \ldots + \hat{T}_L$$

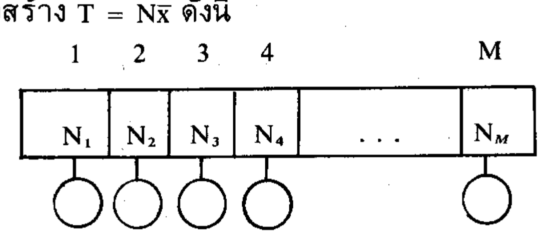
$$= N_1\bar{x}_1 + N_2\bar{x}_2 + \ldots + N_L\bar{x}_L$$

$$= \sum_h^L N_h\bar{x}_h \quad = \sum_h^L \frac{N_h}{n_h} \sum_i^{n_h} x_{hi}$$

$$\bar{x}_{st} = \hat{\bar{X}} = \hat{T}/N = \frac{1}{N}\sum_h^L N_h\bar{x}_h = \frac{1}{N}\sum_h^L \frac{N_h}{n_h}\sum_i^{n_h}$$

$$\hat{V}(\bar{x}_{st}) = \frac{1}{N^2}\sum_h^L \frac{N_h - n_h}{N_h}\,\frac{N_h^2 s_h^2}{n_h}$$

ข. ในแผนสำรวจแบบ 2CS เราจะเลือกชั้นภูมิ (psu) มาเป็นตัวอย่างบางชั้นภูมิแล้วสุ่มหน่วยสำรวจออกจากแต่ละชั้นภูมิ การประมวลผลให้หาค่าเฉลี่ยต่อชั้นภูมิแล้วจึงหายอดรวมโดยอาศัยโครงสร้าง $T = N\bar{x}$ ดังนี้



$$\bar{\bar{x}}_1 \quad \bar{\bar{x}}_2 \quad \bar{\bar{x}}_3 \quad \bar{\bar{x}}_4 \quad \cdot \qquad \bar{\bar{x}}_m$$

$$\hat{T}_1 \quad \hat{T}_2 \quad \hat{T}_3 \quad \hat{T}_4 \qquad \hat{T}_m \quad \hat{T}_i = N_i\bar{x}_i$$

ดังนั้น $$\hat{T} = M\hat{\bar{T}} = M\frac{1}{m}\sum_i^m \hat{T}_i = \frac{M}{m}\sum_i^m \frac{N_i}{n_i}\sum_j^{n_i} x_{ij}$$

$$\hat{\bar{\bar{X}}} = \hat{T}/N = \frac{1}{N}\frac{M}{m}\sum_{i}^{m}\frac{N_i}{n_i}\sum_{j}^{n_i} x_{ij}\ ;\ N = \sum_{i}^{M} N_i$$

$$\hat{V}(\hat{\bar{\bar{X}}}) = \frac{1}{N^2}\left(M^2\frac{M-m}{M}\frac{s_b^2}{m} + \frac{M}{m}\sum_{i}^{m}\frac{N_i-n_i}{N_i}N_i^2\frac{s_i^2}{n_i}\right)$$

ถ้าเปลี่ยน subscript i เป็น h และ j เป็น i และให้ $M = m$ หมายความว่าเลือกตัวอย่างจากทุกชั้นภูมิ (psu) แผนสำรวจ 2CS จะเป็นแผนเดียวกันกับแผนสำรวจแบบแบ่งชั้นภูมิ

จากแนวคิดในข้อ ก. และข้อ ข. ชี้ให้เห็นความเกี่ยวข้องกันในระหว่างแผนสำรวจแบบแบ่งชั้นภูมิและแบบ 2CS ได้เป็นอย่างดีและเห็นได้ว่าแผนสำรวจแบบแบ่งชั้นภูมิเป็นเพียงกรณีเฉพาะของแผนสำรวจแบบ 2CS กล่าวคือ ถ้าให้ $M = m$ แผนสำรวจแบบ 2CS จะลดรูปลงเป็นแผนสำรวจแบบแบ่งชั้นภูมิ สำหรับ subscript นั้น เราจะเปลี่ยน i ให้เป็น h และ j ให้เป็น i หรือไม่ก็ได้ เหตุที่เปลี่ยนมีจุดมุ่งหมายเพียงให้ subscript สอดคล้องกับที่เคยใช้มาแล้วในเรื่องของแผนสำรวจแบบแบ่งชั้นภูมิในบทที่ 3 เท่านั้น

ลองพิจารณาแผนสำรวจแบบ 3 CS จะพบว่า

$$\hat{T} = \frac{L}{\ell}\sum_{i}^{\ell}\frac{M_i}{m_i}\sum_{j}^{m_i}\frac{N_{ij}}{n_{ij}}\sum_{k}^{n_{ij}} x_{ijk}\ ,\ \hat{\bar{\bar{X}}} = \hat{T}/N \text{ เมื่อ } N = \sum_{i}^{L}\sum_{j}^{M_i} N_{ij}$$

$$\hat{V}(\hat{T}) = L^2\frac{L-\ell}{L}\frac{s_b^2}{\ell} + \frac{L}{\ell}\sum_{i}^{\ell}\frac{M_i-m_i}{M_i}\frac{M_i^2 s_i^2}{m_i} + \frac{L}{\ell}\sum_{i}^{\ell}\frac{M_i}{m_i}\sum_{j}^{m_i}\frac{N_{ij}-n_{ij}}{N_{ij}}\frac{N_{ij}^2 s_{ij}^2}{n_{ij}}$$

และ $\hat{V}(\bar{\bar{x}}) = \frac{1}{N^2}\hat{V}(\hat{T})$

ขอให้พิจารณาไดอะแกรมอีกครั้งหนึ่งจากตอน 5.3.1 จะพบว่า

1. เราใช้โรงเรียนเป็นชั้นภูมิ (psu) โดยสุ่มโรงเรียนมาบางส่วน จากนั้นจึงสุ่มชั้นเรียนและสุ่มนักเรียน ถ้าเราสุ่มโดยไม่สุ่มโรงเรียนแต่กลับใช้ทุกโรงเรียนคือให้ $L = \ell$ แล้วสุ่มชั้นเรียนจากทุกโรงเรียนทั้ง L โรง แต่ละโรงเรียนสุ่มนักเรียนมาเป็นตัวอย่าง

แผนสำรวจนี้เราเรียกว่า SCS โดยถือว่า psu เป็นชั้นภูมิ

ดังนั้น เมื่อ $L = \ell$

$$=> \quad \hat{T} = \sum_{i}^{\ell}\frac{M_i}{m_i}\sum_{j}^{m_i}\frac{N_{ij}}{n_{ij}}\sum_{k}^{n_{ij}} x_{ijk}\ ,\ \hat{\bar{\bar{X}}} = \hat{T}/N \text{ เมื่อ } N = \sum_{i}^{L}\sum_{j}^{M_i} N_{ij}$$

$$\hat{V}(\hat{T}) = \sum_{i}^{L} \frac{M_i - m_i}{M_i} \frac{M_i^2 s_i^2}{m_i} + \sum_{i}^{L} \frac{M_i}{m_i} \sum_{j}^{m_i} \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{N_{ij}^2 s_{ij}^2}{n_{ij}}$$ [2]

เมื่อเปลี่ยน subscript ให้สอดคล้องกับแผนสำรวจแบบแบ่งชั้นภูมิในบทที่ 3 คือ เปลี่ยน i เป็น h และเปลี่ยน subscript ให้สอดคล้องกับแผนสำรวจแบบ 2CS คือ เปลี่ยน j เป็น i และเปลี่ยน k เป็น j

จะพบว่า

$$\hat{T} = \sum_{h}^{L} \frac{M_h}{m_h} \sum_{i}^{m_h} \frac{N_{hi}}{n_{hi}} \sum_{j}^{n_{hi}} x_{hij} \ , \quad \hat{\bar{\bar{X}}} = \hat{T}/N$$

$$\hat{V}(\hat{T}) = \sum_{h}^{L} \frac{M_h - m_h}{M_h} \frac{M_h^2 s_h^2}{m_h} + \sum_{h}^{L} \frac{M_h}{m_h} \sum_{i}^{m_h} \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{N_{hi}^2 s_{hi}^2}{n_{hi}}$$

$$\hat{V}(\hat{\bar{\bar{X}}}) = \frac{1}{N^2} \hat{V}(\hat{T})$$

2. ถ้าสุ่มโรงเรียนมาบางส่วน แต่ละโรงเรียนประกอบด้วยชั้นเรียน (ssu) ถ้าเราไม่สุ่มชั้นเรียนมาเป็นบางส่วน แต่กลับใช้ทุกชั้นเรียน โดยถือว่าชั้นเรียนเป็นชั้นภูมิแล้วสุ่มตัวอย่างนักเรียน

แผนสำรวจนี้เรียกว่า SCS โดยที่ถือว่า ssu เป็นชั้นภูมิ ($M_i = m_i$)

ดังนั้น เมื่อ $M_i = m_i$

$$=> \quad \hat{T} = \frac{L}{\ell} \sum_{i}^{\ell} \sum_{j}^{M_i} \frac{N_{ij}}{n_{ij}} \sum_{k}^{n_{ij}} x_{ijk} \ ; \quad \hat{\bar{\bar{X}}} = \hat{T}/N$$

$$\hat{V}(\hat{T}) = L^2 \frac{L-\ell}{L} s_b^2 + \frac{L}{\ell} \sum_{i}^{\ell} \sum_{j}^{M_i} \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{N_{ij}^2 s_{ij}^2}{n_{ij}}$$

$$\hat{V}(\hat{\bar{\bar{X}}}) = \frac{1}{N^2} \cdot \hat{V}(\hat{T})$$

---

[1] เหลือเพียง " = " มิใช่ " ≡ " เพราะแผนนี้มิใช่ 3CS อีกต่อไปแต่เหลือเพียง 2CS ที่กระทำกับทุก psu โดยที่ psu ทำหน้าที่เป็นชั้นภูมิ

[2] $L = \ell$ ดังนั้น $L/\ell = 1$, $(L - \ell)/L = 0$