3.4.3 การประมาณค่าของอัตราส่วนโดยวิธี Separate Ratio Estimate

วิธีนี้ให้เริ่มด้วยการคำนวณหาอัตราส่วนจากแต่ละชั้นภูมิก่อน คือ $\hat{R}_h = \bar{x}_h/\bar{y}_h$ หรือ $\hat{R}_h = \sum_i^{n_h} x_{hi} / \sum_i^{n_h} y_{hi}$ แล้วจึงคำนวณหายอดรวม $\hat{T}_{xh}$ โดยถือว่าทราบค่าจริงของ T_{yh} และ T_y จากนั้นจึงคำนวณหา $\hat{R}_s$ ในภายหลัง นั่นคือ

$$\hat{T}_{xh} = \hat{R}_h T_{yh}$$

จากนั้นจึงรวมค่าของ $\hat{T}_{xh}$ เข้าด้วยกันในทุกชั้นภูมิ (ทุกค่าของ h) เพื่อประมาณค่า T_x

$$1.\ \hat{T}_x = \sum_h^L \hat{T}_{xh} = \sum_h^L \hat{R}_h T_{yh}$$

แล้วประมาณค่าอัตราส่วน R โดยการนำยอดรวมของ Y คือ T_y หารตลอด

นั่นคือ $\hat{R}_s = \hat{T}_x/T_y = \sum_h^L \hat{R}_h T_{yh}/T_y$

$$2.\ \hat{R}_s = \sum_h^L (T_{yh}/T_y)\hat{R}_h$$

และเราสามารถพิสูจน์ได้ว่า

$$3.\ V(\hat{R}_s) \simeq \frac{1}{N^2}\frac{1}{\bar{Y}^2}\sum_h^L \frac{N_h - n_h}{N_h}\frac{N_h^2}{n_h}\cdot(S_{xh}^2 + R_h^2 S_{yh}^2 - 2R_h \varrho_h S_{xh} S_{yh})$$

และ

$$4.\ \hat{V}(\hat{R}_s) \simeq \frac{1}{N^2}\frac{1}{\bar{y}_{st}^2}\sum_h^L \frac{N_h - n_h}{N_h}\frac{N_h^2}{n_h}(s_{xh}^2 + \hat{R}_h s_{yh}^2 - 2\hat{R}_h \varrho_h s_{xh} s_{yh})$$

ขอให้สังเกตว่าสูตร $V(\hat{R}_c)$ ต่างกับ $V(\hat{R}_s)$ เล็กน้อยตรงที่ค่า R ในสูตร $V(\hat{R}_s)$ ใช้ R_h ส่วนในสูตร $V(\hat{R}_c)$ ใช้ R เฉย ๆ ซึ่งเมื่ออาศัยข้อมูลจากกลุ่มตัวอย่างต้องประมาณค่า R นี้ด้วย $\hat{R}_c = \sum_h^L N_h\bar{x}_h / \sum_h^L N_h\bar{y}_h = \bar{x}_{st}/\bar{y}_{st}$ ส่วนใน $V(\hat{R}_s)$ เราประมาณค่า R_h ด้วย

$$\hat{R}_h = \bar{x}_h/\bar{y}_h \quad \text{หรือ} \quad \sum_{i}^{n_h} x_{hi} / \sum_{i}^{n_h} y_{hi}$$

การพิสูจน์สูตร $V(\hat{R}_s)$ จะขอเว้นไว้เป็นแบบฝึกหัด

สำหรับการประมาณค่าของยอดรวม $\hat{T}_x$ เราสามารถพิสูจน์ได้ว่า

5. $$V(\hat{T}_x) \simeq \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2}{n_h} (S_{xh}^2 + R_h^2 S_{yh}^2 - 2R_h \varrho_h S_{xh} S_{yh})$$

และ

6. $$\hat{V}(\hat{T}_x) \simeq \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2}{n_h} (s_{xh}^2 + \hat{R}_h^2 s_{yh}^2 - 2\hat{R}_h \hat{\varrho}_h s_{xh} s_{yh})$$

ตัวอย่าง 3.6 จากข้อมูลในตัวอย่างที่ 3.5 จงกะประมาณค่า R ด้วยช่วงเชื่อมั่น 99% ถ้าทราบว่ายอดขายรวมในเดือนเมษายนของแต่ละชั้นภูมิ (T_{yh}) เท่ากับ 13,000, 14,000 และ 11,000 หน่วยบาทตามลำดับ

วิธีทำ

ชั้นภูมิ	N_h	n_h	$\bar{x}_h$	$\bar{y}_h$	T_{yh}	s_{xh}^2	s_{yh}^2	$\hat{R}_h$	$\hat{\varrho}_h$	$\hat{R}_h^2$
1	1,000	50	10	10	13,000	1	1	10/10 = 1	1	1
2	500	50	33	30	14,000	2	2	33/30 = 1.1	1	1.21
3	200	50	60	50	11,000	3	3	60/50 = 1.2	1	1.44
รวม	1,700	150			38,000					

$$\text{ดังนั้น} \quad \hat{R}_s = \sum_{h}^{L} (\hat{R}_h T_{yh})/T_y$$

$$= (1 \times 13000 + 1.1 \times 14000 + 1.2 \times 11000)/38000$$

$$= 41600/38000 = 1.09$$

จาก $\hat{V}(\hat{R}_s) = \frac{1}{N^2}\frac{1}{\bar{y}_{st}^2}\sum_h^L \frac{N_h - n_h}{N_h}\frac{N_h^2}{n_h}(s_{xh}^2 + \hat{R}_h^2 s_{yh}^2 - 2\hat{\varrho}_h\hat{R}_h s_{yh}s_{xh})$

เราสามารถกะประมาณโดยการเตรียมตารางการวิเคราะห์ได้ดังนี้

ชั้นภูมิ	N_h	n_h	$\frac{(N_h - n_h)}{N_h}\frac{N_h^2}{n_h}$	s_{xh}^2	$R_h^2 s_{yh}^2$	$2\hat{R}_h\hat{\varrho}_h s_{xh}s_{yh}$	s_{zh}^2	$\frac{N_h - n_h}{N_h}\frac{N_h^2 s_{zh}^2}{n_h}$
1	1000	50	19000	1	1	2	0	0
2	500	50	4500	2	2.42	4.4	.02	90
3	200	50	600	3	4.32	7.2	.12	72
รวม	1700	150						162

และ $\because \bar{y}_{st} = \frac{1}{N}\sum_h^L N_h\bar{y}_h = (1000\times 10 + 500\times 30 + 200\times 50)/1700 = 20.588$

และ $N\bar{y}_{st} = \sum_h^L N_h\bar{y}_h \qquad N^2\bar{y}_{st}^2 = (35000)^2 = 1225000000$

$$\hat{V}(\hat{R}_s) = \frac{1}{(35000)^2}\times 162 = .0000001$$

$$s(\hat{R}_s) = .000316$$

ดังนั้น ช่วงเชื่อมั่น 99% ของ R คือ

$$\{1.09 - (2.58)(.000316),\ 1.09 + (25.8)(.000316)\} = (1.08918,\ 1.09082)$$

3.4.4 การเปรียบเทียบคุณภาพระหว่าง $\hat{R}_c$ และ $\hat{R}_s$

การเปรียบเทียบว่า $\hat{R}_c$ หรือ $\hat{R}_s$ จะดีกว่ากันนั้นเรานิยมพิจารณาจุดใหญ่ ๆ 3 จุด คือ ปริมาณความเอียงเฉ ปริมาณความแปรปรวนของการประมาณค่า และความเหมาะสมในทางปฏิบัติ โดยพิจารณาดูว่า $\text{Bias}(\hat{R}_c) - \text{Bias}(\hat{R}_s) = ?$ และ $V(\hat{R}_c) - V(\hat{R}_s) = ?$ ผลต่างของทั้ง Bias และ Variance มีอะไรเป็นที่น่าสังเกตได้บ้าง อย่างไรก็ตามเราสามารถสรุปผลจากผลต่างได้ดังนี้

1. ถ้า n_h มีค่าต่ำ $\text{Bias}(\hat{R}_s)$ จะมีค่าสูง ดังนั้น ถ้า n_h มีค่าต่ำเราควรใช้ Combined Ratio Estimate คือ $\hat{R}_c = \sum_h^L N_h\bar{x}_h / \sum_h^L N_h\bar{y}_h = \hat{T}_x/\hat{T}_y = \bar{x}_{st}/\bar{y}_{st}$

2. ถ้า n_h มีค่าสูงคือ $n_h \geqslant 50$; $h=1,2,\ldots,L$ เราจะถือว่าตัวอย่างมีขนาดใหญ่พอ ซึ่งในกรณีเช่นนี้ ไม่จำเป็นต้องกังวลถึงปริมาณความเอียงเฉต่อไป

3. ถ้าปรารถนาจะทราบอัตราส่วนของแต่ละชั้นภูมิคือ R_h ; $h=1,2,\ldots,L$ และ n_h มีค่าสูง ให้เลือกใช้ Separate Ratio Estimate

4. ถ้า $\hat{R}_h$ มีค่าผันแปรไปในระหว่างชั้นภูมิได้มาก จะมีผลให้ $V(\hat{R}_c) > V(\hat{R}_s)$ ในกรณีเช่นนี้ควรจะใช้ Separate Ratio Estimate ความจริงข้อนี้สามารถแสดงให้เห็นได้ดังนี้คือ

$$\begin{aligned}
V(\hat{R}_c) - V(\hat{R}_s) &= \frac{1}{N^2}\frac{1}{\bar{Y}^2}\sum_h^L N_h^2\,\frac{N_h-n_h}{N_h}\cdot\frac{1}{n_h}\{(R^2-R_h^2)S_{yh}^2 - 2(R-R_h) \\
&\quad \cdot\varrho_h S_{xh}S_{yh}\} \\
&= \frac{1}{N^2}\frac{1}{\bar{Y}^2}\sum_h^L N_h^2\,\frac{N_h-n_h}{N_h}\cdot\frac{1}{n_h}\{(R-R_h)^2S_{yh}^2 + 2RR_hS_{yh}^2 \\
&\quad - 2R_h^2\,S_{yh}^2 - 2(R-R_h)\varrho_h S_{xh}S_{yh}\} \\
&= \frac{1}{N^2}\frac{1}{\bar{Y}^2}\sum_h^L N_h^2\,\frac{N_h-n_h}{N_h}\,\frac{1}{n_h}\{(R-R_h)^2S_{yh}^2 + 2(R-R_h) \\
&\quad \cdot(2R_hS_{yh}^2 - 2\varrho_h S_{xh}S_{yh})\}
\end{aligned}$$

ถ้าสมการถดถอยผ่านจุดกำเนิดก็แสดงว่าในทุก ๆ ชั้นภูมินั้น $R_hS_{yh} = \varrho_h S_{xh}$ ซึ่งมีผลให้ $2R_hS_{yh}^2 - 2\varrho_h S_{xh}S_{yh} = 0$

$$=> V(\hat{R}_c) - V(\hat{R}_s) \simeq \frac{1}{N^2} \frac{1}{\bar{Y}^2} \sum_{h}^{L} N_h^2 \frac{N_h - n_h}{N_h} \frac{1}{n_h} (R - R_h)^2 S_{yh}^2$$

แสดงว่าผลต่างของความแปรปรวนจะต้องผันผวนไปตามค่าของ $(R - R_h)^2$

$$=> (V(\hat{R}_c) - V(\hat{R}_s)) \propto (R - R_h)^2$$

นั่นคือ ถ้า R_h มีค่าแตกต่างกันไปมาก $(R - R_h)^2$ จะมีค่ามาก ซึ่งมีผลให้ $V(\hat{R}_c) - V(\hat{R}_s)$ มีค่ามาก หรือ $V(\hat{R}_c) > V(\hat{R}_s)$

3.4.5 การจัดสรรขนาดตัวอย่าง

การจัดสรรขนาดตัวอย่างเพื่อกะประมาณสัดส่วนโดยใช้แผนการสำรวจแบบแบ่งชั้นภูมินั้น เราสามารถจัดสรรขนาดตัวอย่างได้ 4 วิธีคือ

1. จัดสรรอย่างเท่าเทียมกัน คือ $n_h = n/L \; ; h = 1,2,\ldots,L$
2. จัดสรรให้ได้สัดส่วนกับขนาดชั้นภูมิคือ $n_h = \frac{n}{N} \cdot N_h \; ; h = 1,2,\ldots,L$
3. จัดสรรแบบอุตมะ

จากกรณีทั่วไปคือ $V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{n}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 S_h^2}{n_h}$ ภายใต้ข้อกำหนดเกี่ยวกับงบประมาณคือ $C = c_0 + \sum_{h}^{L} n_h c_h$

เราสามารถพิสูจน์ได้ว่า $n_h = \dfrac{N_h S_h / \sqrt{c_h}}{\sum_{h}^{L} N_h S_h / \sqrt{c_h}} \cdot n \; ; h = 1,2,\ldots,L$

สำหรับกรณีของอัตราส่วน เราจะพบว่าโครงสร้างของสูตร $V(\hat{R}_c)$ หรือ $V(\hat{R}_s)$ คล้ายคลึงกันกับโครงสร้างของสูตร $V(\bar{x}_{st})$ ดังนั้น เราสามารถพิสูจน์ได้โดยง่ายว่า

ก. ในกรณีของ Combined Ratio Estimate

$$n_h = \frac{N_h (S_{ch}/\bar{Y}) / \sqrt{c_h}}{\sum_{h}^{L} N_h (S_{ch}/\bar{Y}) / \sqrt{c_h}} \cdot n \quad ; h = 1,2,\ldots,L$$

$$S_{ch}^2 = S_{xh}^2 + R^2S_{yh}^2 - 2R\rho_h S_{xh}S_{yh}$$

การพิสูจน์กระทำได้โดยง่ายเพียงแต่จัดรูป $V(\hat{R}_c)$ เสียใหม่ กล่าวคือ

จาก $$V(\hat{R}_c) = \frac{1}{N^2}\frac{1}{\overline{Y}^2}\sum_h^L \frac{N_h - n_h}{N_h}\frac{N_h^2}{n_h}(S_{xh}^2 + R^2S_{yh}^2 - 2R\rho_h S_{xh}S_{yh})$$

ให้ $$S_{ch}^2 = S_{xh}^2 + R^2S_{yh}^2 - 2R\rho_h S_{xh}S_{yh}$$

$$=> \quad V(\hat{R}_c) = \frac{1}{N^2}\sum_h^L \frac{N_h - n_h}{N_h}\frac{N_h^2(S_{ch}/\overline{Y})^2}{n_h}$$

จาก $n_h = \dfrac{N_hS_h/\sqrt{c_h}}{\Sigma N_hS_h/\sqrt{c_h}} \cdot n$ สำหรับการจัดสรรตัวอย่างเพื่อประมาณค่า $\overline{X}$

เราแทนที่ S_h ด้วย $S_{ch}/\overline{Y}$ ก็จะได้ n_h สำหรับกรณีจัดสรรตัวอย่างเพื่อประมาณค่า R

$$=> \quad n_h = \frac{N_h(S_{ch}/Y)/\sqrt{c_h}}{\Sigma N_h(S_{ch}/\overline{Y})/\sqrt{c_h}} \cdot n \quad ; \quad h = 1,2,\ldots,L$$

และถ้าไม่ทราบค่าจริงของ S_{ch} และ $\overline{Y}$ ให้ใช้

$$n_h = \frac{N_h(s_{ch}/\bar{y}_{st})/\sqrt{c_h}}{\Sigma N_h(s_{ch}/\bar{y}_{st})/\sqrt{c_h}} \cdot n \ ; h = 1,2,\ldots,L$$

เมื่อ $$s_{ch}^2 = s_{xh}^2 + \hat{R}_c^2 s_{yh}^2 - 2\hat{R}_c\hat{\rho}_h s_{xh}s_{yh} \ ; \ \hat{R}_c = \bar{x}_{st}/\bar{y}_{st} = \hat{T}_x/\hat{T}_y$$

$$\hat{\rho}_h = \frac{\sum_i^{n_h}(x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{(n_h - 1)s_{xh}s_{yh}}$$

และ $\bar{y}_{st} = \dfrac{1}{N}\sum_h^L N_h\bar{y}_h$ ซึ่งค่าเหล่านี้ทั้งหมดเป็นค่าที่ได้จากงานสำรวจเบื้องต้น

ข. ในกรณี Separate Ratio Estimate

$$n_h = \frac{N_h(S_{sh}/\overline{Y})/\sqrt{c_h}}{\sum_h^L N_h(S_{sh}/\overline{Y})/\sqrt{c_h}} \cdot n \ ; h = 1,2,\ldots,L$$

โดยที่ $S^2_{sh} = S^2_{xh} + R^2_h S^2_{yh} - 2R_h\varrho_h S_{xh}S_{yh}$ และ

$$V(\hat{R}_s) = \frac{1}{N^2} \cdot \frac{1}{\bar{Y}^2} \sum_h^L \frac{N_h - n_h}{N_h} \frac{N_h^2 S^2_{sh}}{n_h}$$

$$= \frac{1}{N^2} \sum_h^L \frac{N_h - n_h}{N_h} \frac{N_h^2 (S_{sh}/\bar{Y})^2}{n_h}$$

และในกรณีที่ไม่ทราบค่าจริงของ S^2_{sh} และ $\bar{Y}$ ให้ใช้

$$n_h = \frac{N_h(s_{sh}/\bar{y}_{st})/\sqrt{c_h}}{\sum_h^L N_h(s_{sh}/\bar{y}_{st})/\sqrt{c_h}} \cdot n \; ; \; h = 1,2,\ldots,L$$

โดยที่ $s^2_{sh} = (s^2_{xh} + \hat{R}_h s^2_{yh} - 2\hat{R}_h\hat{\varrho}_h s_{xh}s_{yh})$, $\hat{R}_h = \bar{x}_h/\bar{y}_h$

$$\bar{y}_{st} = \frac{1}{N} \sum_h^L N_h\bar{y}_h \, , \; \hat{\varrho}_h = \frac{\sum_i^{n_h} (x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{(n_h - 1)s_{xh}s_{yh}}$$

4. การจัดสรรแบบเนย์แมน

การจัดสรรแบบนี้อาจถือว่าเป็นกรณีเฉพาะของการจัดสรรแบบอุตมะ เมื่อ $c_h = c_f \; ; \; h = 1,2,\ldots,L$

ดังนั้น การจัดสรรขนาดตัวอย่างที่เหมาะสมสำหรับสุ่มตัวอย่างจากแต่ละชั้นภูมิเพื่อกะประมาณค่าของ R ก็คือ

ก. กรณี Combined Ratio Estimate

$$n_h = \frac{N_h(S_{ch}/\bar{Y})}{\sum_h^L N_h(S_{ch}/\bar{Y})} \cdot n \; ; \; h = 1,2,\ldots,L$$

และใช้ $n_h = \dfrac{N_h(s_{ch}/\bar{y}_{st})}{\sum_h^L N_h(s_{ch}/\bar{y}_{st})} \cdot n \; ; \; h = 1,2,\ldots,L$ เมื่อไม่ทราบค่า

S^2_{ch} และ $\bar{Y}$ โดยที่ s^2_{ch} และ $\bar{y}_{st}$ เป็นค่าที่ได้จากงานสำรวจเบื้องต้น

ข. กรณี Separate Ratio Estimate

$$n_h = \frac{N_h(S_{sh}/\bar{Y})}{\sum_h^L N_h(S_{sh}/\bar{Y})} \cdot n \,;\, h=1,2,\ldots,L$$

และใช้ $$n_h = \frac{N_h(s_{sh}/\bar{y}_{st})}{\sum_h^L N_h(s_{sh}/\bar{y}_{st})} \cdot n \,;\, h=1,2,\ldots,L$$ เมื่อไม่ทราบค่า

S^2_{sh} และ $\bar{Y}$ โดยที่ s^2_{sh} และ $\bar{y}_{st}$ เป็นค่าที่ได้จากงานสำรวจเบื้องต้น

3.4.6 การกำหนดขนาดตัวอย่าง

การกำหนดขนาดตัวอย่างเพื่อกะประมาณ R ให้ปฏิบัติเช่นเดียวกันกับการประมาณค่าเฉลี่ยและยอดรวมที่กล่าวถึงมาแล้วในตอน 3.3 ความสัมพันธ์ที่จะใช้คือ $D^2 = V(\hat{R}_c)$ หรือ $D^2 = V(\hat{R}_s)$ แล้วแต่กรณี โดยที่ $D^2 = d_0^2/Z^2_{1-\alpha_0/2}$.

ขนาดตัวอย่างจะต้องกำหนดให้สอดคล้องกับ D^2 และแผนการณ์จัดสรรตัวอย่าง ดังนั้นความสัมพันธ์ที่จะใช้จึงเปลี่ยนแปลงไปตามแผนการจัดสรรดังนี้

1. $D^2 = V(\hat{R}_{ceq})$ หรือ $D^2 = V(\hat{R}_{seq})$

โดยแทนที่ n_h ด้วย n/L แล้วแก้สมการหาค่า n

2. $D^2 = V(\hat{R}_{cprop})$ หรือ $D^2 = V(\hat{R}_{sprop})$

โดยแทนที่ n_h ด้วย $\frac{n}{N} \cdot N_h$ แล้วแก้สมการหาค่า n

3. $D^2 = V(\hat{R}_{copt})$ หรือ $D^2 = V(\hat{R}_{sopt})$

โดยแทนที่ n_h ด้วย $\frac{N_h(S_{ch}/\bar{Y})/\sqrt{c_h}}{\sum_h^L N_h(S_{ch}/\bar{Y})/\sqrt{c_h}} \cdot n$ หรือ $\frac{N_h(S_{sh}/\bar{Y})/\sqrt{c_h}}{\sum_h^L N_h(S_{sh}/\bar{Y})/\sqrt{c_h}} \cdot n$

แล้วแต่กรณี แล้วแก้สมการหาค่า n

4. $D^2 = V(\hat{R}_{cney})$ หรือ $D^2 = V(\hat{R}_{sney})$

โดยแทนที่ n_h ด้วย $\dfrac{N_h(S_{ch}/\overline{Y})}{\sum_h^L N_h(S_{ch}/\overline{Y})} \cdot n$ หรือ $\dfrac{N_h(S_{sh}/\overline{Y})}{\sum_h^L N_h(S_{sh}/\overline{Y})} \cdot n$

แล้วแก้สมการหาค่า n

ตัวอย่าง 3.7 ในท้องที่ (ตำบล) แห่งหนึ่งมีครัวเรือนทั้งสิ้น 460 ครัวเรือน จำแนกครัวเรือนออกเป็น 3 ชั้นภูมิตามเขตการปกครองท้องที่ (หมู่บ้าน) คือหมู่บ้านที่ 1 มี 200 ครัวเรือน หมู่บ้านที่ 2 100 ครัวเรือน หมู่บ้านที่ 3 100 ครัวเรือน

ต้องการกะประมาณอัตราส่วนที่เป็นภาระของตำบลนี้ โดยนิยามว่า บุคคลที่เป็นภาระคือบุคคลที่มีอายุต่ำกว่า 15 ปี และเกินกว่า 65 ปี โดยให้ x_i = จำนวนผู้ที่เป็นภาระในครัวเรือนที่ i y_i = จำนวนบุคคลที่เป็นสมาชิกของครัวเรือนในครัวเรือนที่ i

ผลการสำรวจเบื้องต้นปรากฏข้อมูลดังนี้

หมู่บ้านที่	n_h	x_{hi} (จำนวนผู้ที่เป็นภาระ)	y_{hi} (จำนวนสมาชิกครัวเรือน)
1	10	2,4,1,3,1,5,4,4,5,3	3,4,2,4,2,5,4,5,6,5
2	8	5,6,5,7,8,8,9,11	6,8,6,9,11,8,10,13
3	5	11,13,15,12,18	13,12,18,14,24

จงกำหนดขนาดตัวอย่างด้วยแผนสำรวจให้สอดคล้องกับแผนจัดสรรตัวอย่างทั้ง 4 แบบ พร้อมทั้งจัดสรรตัวอย่างให้แก่แต่ละชั้นภูมิ กำหนดให้ $D^2 = .01^2$ และให้ใช้ Combined Ratio Estimate

วิธีทำ จากโจทย์เราสามารถเตรียมตารางวิเคราะห์ได้ดังนี้

ชั้นภูมิ	N_h	n_h	x_{hi}	y_{hi}	$\bar{x}_h$	$\bar{y}_h$	$N_h\bar{x}_h$	$N_h\bar{y}_h$
1	200	10	2,4,1,3,1,5 4,4,5,3	3,4,2,4,2,5 4,5,6,5	3.2	4	640	800
2	160	8	5,6,5,7,8,8, 9,11	6,8,6,9,11, 8,10,13	7.38	8.875	1180	1420
3	100	5	11,13,15,12,18	13,12,18,14,24	13.8	16.2	1380	1620
รวม	460	23					3200	3840

s^2_{xh}	s^2_{yh}	s_{xyh}
$\frac{1}{9}(122-32^2/10) = 2.18$	$\frac{1}{9}(176-40^2/10) = 1.78$	$\frac{1}{9}(144-128) = 1.778$
$\frac{1}{7}(465-59^2/8) = 4.27$	$\frac{1}{7}(671-71^2/8) = 5.84$	$\frac{1}{7}(556-523.625) = 4.625$
$\frac{1}{4}(983-69^2/5) = 7.7$	$\frac{1}{4}(1409-81^2/5) = 24.2$	$\frac{1}{4}(1169-117.8) = 12.8$

หมายเหตุ

$$s_{xyh} = \frac{1}{n_h-1}\sum_{i}^{n_h}(x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h)$$

$$= \frac{1}{n_h-1}\sum_{i}^{n_h}(x_{hi}y_{hi}-n_h\bar{x}_h\bar{y}_h)$$

$$\hat{\varrho}_h = s_{xyh}/s_{xh}s_{yh}\ ,\ \hat{R}_c = N_h\bar{x}_h/N_h\bar{y}_h = 3200/3840 = .833$$

ดังนั้น

ชั้นภูมิ	s^2_{xh}	s^2_{yh}	$\hat{R}_c$	$\hat{R}^2_c s^2_{yh}$	s_{xh}	s_{yh}	$\hat{\varrho}_h$	$2\hat{\varrho}_h \hat{R}_c s_{xh} s_{yh}$	s^2_{ch}
1	2.18	1.78	.833	1.24	1.48	1.33	.903	2.961	.459
2	4.27	5.84	.833	4.05	2.07	2.42	.923	7.703	.617
3	7.7	24.2	.833	16.79	2.78	4.92	.936	21.329	3.161

ต่อไปนี้จะแสดงให้ดูเฉพาะกรณีการจัดสรรอย่างเท่าเทียมกัน (Equal Allocation) เท่านั้น กรณีอื่น ๆ ขอเว้นไว้เป็นแบบฝึกหัด

จาก
$$D^2 = V(\hat{R}_c) = \frac{1}{N^2\bar{y}^2_{st}} \sum_h^L \frac{N_h - n_h}{N_h} \frac{N_h^2 s^2_{ch}}{n_h}$$

$$=> \quad D^2 = \frac{1}{\hat{T}^2_y} \sum_h^L \frac{N_h^2 s^2_{ch}}{n/L} - \frac{1}{\hat{T}^2_y} \sum_h^L N_h s^2_{ch}$$

$$(.01)^2 = \frac{1}{(3840)^2} \left(\frac{3}{n}\right)(250^2 \times .459 + 160^2 \times .617 + 100^2 \times 3.161)$$

$$- \frac{1}{(3840)^2}(200 \times 0.459 + 160 \times .617 + 100 \times 3.161)$$

$$.0001 = \frac{197295.6}{(3840)^2 n} - \frac{506.62}{(3840)^2}$$

$$.0133799/n = .0001343$$

$$n = 99.6 \simeq 100$$

3.5 การประมาณค่าสัดส่วน, P

ดังที่ได้กล่าวมาแล้วในบทที่ 2 ว่าสัดส่วนก็คือกรณีเฉพาะของค่าเฉลี่ยเมื่อตัวแปร x มีค่าที่เป็นไปได้ (Possible Value) เพียง 2 ค่าคือ 0 หรือ 1 และได้ศึกษาถึงเรื่องสัดส่วนไว้

แล้วค่อนข้างจะละเอียดในบทดังกล่าว ดังนั้นในที่นี้จึงจะขอกล่าวถึงเรื่องของสัดส่วนโดยสรุปเท่านั้น การพิสูจน์และคำอธิบายเพิ่มเติมขอให้ยึดถือหลักเกณฑ์ที่ศึกษาแล้วในตอนที่ผ่านมา

3.5.1 การประมาณค่าสัดส่วน P

ให้ $x_{hi} = 0$ ถ้า $x_{hi} \notin c$ และ $x_{hi} = 1$ ถ้า $x_{hi} \in c$; $i = 1,2,\ldots,n_h$; $h = 1,2,\ldots,L$

ดังนั้น เมื่อสุ่มตัวอย่างโดยใช้แผนสำรวจแบบ SRS จากแต่ละชั้นภูมิ $\hat{P}_h$ จะเป็นตัวประมาณค่าที่ปราศจากอคติของ P_h เมื่อ $p_h = \hat{P}_h = \frac{1}{n_h} \sum_{i}^{n_h} x_{hi}$; $h = 1,2,\ldots,L$

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i}^{N_h} (x_{hi} - \bar{X}_h)^2 = \frac{1}{N_h - 1} \left\{ \sum_{i}^{N_h} x_{hi}^2 - \left(\sum_{i}^{N_h} x_{hi} \right)^2 / N_h \right\} = \frac{N_h P_h Q_h}{N_h - 1}$$

และ $$s_h^2 = \frac{1}{n_h - 1} \sum_{i}^{n_h} (x_{hi} - \bar{x}_h)^2 = \frac{1}{n_h - 1} \left\{ \sum_{i}^{n_h} x_{hi}^2 - \left(\sum_{i}^{n_h} x_{hi} \right)^2 / n_h \right\} = \frac{n_h p_h q_h}{n_h - 1}$$

ดังนั้นค่าประมาณของ P_{st} และ $V(\hat{P}_{st})$ ก็คือ

$$\hat{P}_{st} = \frac{1}{N} \sum_{h}^{L} N_h \hat{P}_h$$

$$V(\hat{P}_{st}) = \frac{1}{N^2} \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 S_h^2}{n_h}$$

$$= \frac{1}{N^2} \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 N_h P_h Q_h}{n_h (N_h - 1)}$$

$$= \frac{1}{N^2} \sum_{h}^{L} \frac{N_h - n_h}{N_h - 1} \frac{N_h^2 P_h Q_h}{n_h} \simeq \frac{1}{N^2} \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 P_h Q_h}{n_h}$$

และ $$\hat{V}(\hat{P}_{st}) = \frac{1}{N^2} \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 s_h^2}{n_h}$$

$$= \frac{1}{N^2} \sum_{h}^{L} \frac{N_h - n_h}{N_h} \cdot \frac{N_h^2 n_h p_h q_h}{n_h (n_h - 1)}$$

$$= \frac{1}{N^2} \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 p_h q_h}{n_h - 1}$$

$$\simeq \frac{1}{N^2}\sum_{h}^{L}\frac{N_h-n_h}{N_h}\frac{N_h^2p_hq_h}{n_h}$$

ดังนั้นช่วงชื่อมั่น $(1-\alpha)$ 100% ที่คาดว่าค่าจริงของ P จะปรากฏอยู่คือ

$(\hat{P}_{st}+Z_{\alpha/2}\sqrt{V(\hat{P}_{st})}\,,\ \hat{P}_{st}+Z_{1-\alpha/2}\sqrt{V(\hat{P}_{st})})$ ถ้าทราบค่า S_h^2 หรือทราบค่า P_h

หรือ

$(\hat{P}_{st}+Z_{\alpha/2}\sqrt{\hat{V}(\hat{P}_{st})}\,,\ \hat{P}_{st}+Z_{1-\alpha/2}\sqrt{\hat{V}(\hat{P}_{st})})$ เมื่อประมาณ S_h^2 ด้วย s_h^2 และ $n\to\infty$

หรือ

$(\hat{P}_{st}+t_{n-1,\alpha/2}\sqrt{\hat{V}(\hat{P}_{st})},\ \hat{P}_{st}+t_{n-1,1-\alpha/2}\sqrt{\hat{V}(\hat{P}_{st})})$ เมื่อประมาณค่า S_h^2 ด้วย s_h^2 และ $1 \leqslant n < \infty$

สูตรและโครงสร้างต่าง ๆ เหล่านี้ เป็นสิ่งที่เราสามารถพิสูตรได้โดยง่าย ในที่นี้จึงจะไม่พิสูจน์และจะขอเว้นไว้เป็นแบบฝึกหัด อนึ่ง ขอให้สังเกตไว้ประการหนึ่งก็คือการประมาณค่าพารามิเตอร์ด้วยช่วงเชื่อมั่นนั้น ในทุกตอนที่ผ่านมาจะพบว่าเรานิยมใช้การแจกแจงแบบปกติเข้าช่วยเหลือ ความจริงข้อนี้เราสามารถพิสูจน์ให้เห็นด้วยทฤษฎี CLT ซึ่งการจะอ้างทฤษฎีนี้ได้ต้องอาศัยกลุ่มตัวอย่างขนาดใหญ่คือ $n\to\infty$ ดังนั้น จึงขอให้นักศึกษาได้ทำความเข้าใจไว้ในที่นี้ด้วยว่า ในการสำรวจทุกครั้งควรใช้กลุ่มตัวอย่างขนาดใหญ่เสมอ เพราะจะได้อาศัยการแจกแจงแบบปกติเข้าช่วยเหลือได้ มิเช่นนั้นอาจจำเป็นต้องคำนวณหา Sampling Distribution ของตัวประมาณค่า ซึ่งโดยปกติก็ไม่ใช่เรื่องง่ายอยู่แล้ว ซ้ำยังจะต้องเสียเวลาไปโดยไม่จำเป็นอีกด้วย

3.5.2 การจัดสรรตัวอย่าง

การจัดสรรตัวอย่าง n ให้แก่แต่ละชั้นภูมิเพื่อการสำรวจและกะประมาณสัดส่วน P นั้น เราสามารถกระทำได้ 4 แบบเช่นเดียวกันกับที่เคยปฏิบัติมาแล้วคือ

1. $n_h = n/L\ ;\ h=1,2,\ldots,L$ (Equal Allocation)

2. $n_h = \frac{n}{N} \cdot N_h \; ; \; h = 1,2,\ldots,L$ (Proportional Allocation)

3. $$n_h = \frac{N_h(\sqrt{N_hP_hQ_h/N_h-1})/\sqrt{c_h})}{\sum_h^L N_h(\sqrt{(N_hP_hQ_h/N_h-1})/\sqrt{c_h})} \cdot n \text{ (Optimum Allocation)}$$

$$\simeq \frac{N_h\sqrt{P_hQ_h}/\sqrt{c_h}}{\sum_h^L N_h\sqrt{P_hQ_h}/\sqrt{c_h}} \cdot n$$

4. $$n_h = \frac{N_h\sqrt{(N_hP_hQ_h)/N_h-1}}{\sum_h^L N_h\sqrt{N_hP_hQ_h/N_h-1}} \cdot n \text{ (Neyman Allocation)}$$

$$\simeq \frac{N_h\sqrt{P_hQ_h}}{\sum_h^L N_h\sqrt{P_hQ_h}} \cdot n$$

สำหรับกรณีการจัดสรรแบบอุตมะและแบบเนย์แมน เราสามารถพิสูจน์ได้โดยอาศัย Lagrange Multiplier Method เช่นเดียวกันกับที่ได้แสดงการพิสูจน์แล้วในตอน 3.3 เพียงแต่เปลี่ยน $V(\bar{x}_{st})$ เป็น $V(\hat{P}_{st})$ เท่านั้น เช่น ในกรณีการจัดสรรแบบอุตมะจะพบว่า

$$F = V(\hat{P}_{st}) + \lambda(\sum_h^L n_hc_h - C)$$

$$= \frac{1}{N^2}\sum_h^L \frac{N_h-n_h}{N_h-1} N_h^2 \frac{P_hQ_h}{n_h} + \lambda(\sum_h^L N_hc_h - C)$$

แล้วดิฟเฟอเรนชิเอทเทียบต่อ n_h ก็จะได้สูตร n_h ตามต้องการ

หรืออาจใช้วิธีสรุปง่าย ๆ ว่า

เนื่องจาก $n_h = \frac{N_hS_h/\sqrt{c_h}}{\sum_h^L N_hS_h/\sqrt{c_h}} \cdot n \; ; \; h = 1,2,\ldots,L$

และ $\because S_h = \sqrt{N_hP_hQ_h/N_h-1}$ หรือ $S_h \simeq \sqrt{P_hQ_h}$

แทนค่า S_h ลงในสูตรข้างบนก็จะได้สูตร n_h สำหรับการประมาณสัดส่วนตามต้องการ

3.5.3 การกำหนดขนาดตัวอย่างเพื่อกะประมาณสัดส่วน P

ขนาดตัวอย่างที่จะกำหนดนี้เราต้องกำหนดให้สอดคล้องกับแผนการจัดสรรตัวอย่าง ซึ่งเราต้องกำหนดไว้ก่อนที่จะมีการกำหนดขนาดตัวอย่าง ตลอดจนต้องสอดคล้องกับระดับความแม่นยำและความน่าเชื่อถือ

และโดยอาศัยวิธีการเช่นเดียวกับที่ศึกษาแล้วในตอน 3.3 เราสามารถพิสูจน์ได้โดยง่ายว่า

1. $$n_{eq} = \frac{L\,\Sigma N_h^2(N_hP_hQ_h/N_h-1)}{N^2D^2+\sum_h^L N_h(N_hP_hQ_h/N_h-1)};\ D^2 = d_0^2/Z^2_{1-\alpha/2}$$

2. $$n_{prop} = \frac{N\,\Sigma N_h(N_hP_hQ_h/N_h-1)}{N^2D^2+\sum_h^L N_h(N_hP_hQ_h/N_h-1)}$$

$$\simeq \frac{N\,\Sigma N_hP_hQ_h}{N^2D^2+\sum_h^L N_hP_hQ_h}$$

3. $$n_{opt} = \frac{(\Sigma N_h(\sqrt{N_hP_hQ_h/N_h-1})\sqrt{c_h})(\Sigma N_h(\sqrt{N_hP_hQ_h/N_h-1})/\sqrt{c_h})}{N^2D^2+\sum_h^L N_h(N_hP_hQ_h/N_h-1}$$

$$\simeq \frac{(\Sigma N_h\sqrt{P_hQ_hc_h})(\Sigma N_h\sqrt{P_hQ_h/c_h})}{N^2D^2+\sum_h^L N_hP_hQ_h}$$

4. $n_{ney} = \dfrac{(\Sigma N_h\sqrt{N_hP_hQ_h/N_h-1})^2}{N^2D^2 + \sum\limits_h^L N_h(N_hP_hQ_h/N_h-1)}$

$$\simeq \frac{(\Sigma N_h\sqrt{P_hQ_h})^2}{N^2D^2 + \sum\limits_h^L N_hP_hQ_h}$$

ในกรณีที่อาศัยข้อมูลจากกลุ่มตัวอย่างเพื่อกะประมาณค่าของ S_h^2 ให้ใช้ s_h^2 แทนที่ S_h^2 หรือนัยหนึ่ง ใช้ $\dfrac{n_hp_hq_h}{n_h-1}$ แทน $\dfrac{N_hP_hQ_h}{N_h-1}$ หรือ p_hq_h แทน P_hQ_h สำหรับการพิสูจน์จะขอเว้นไว้เป็นแบบฝึกหัด

ตัวอย่าง 3.8 โรงงานแห่งหนึ่งต้องการกะประมาณสัดส่วนของพนักงานที่มีความรู้สึกขัดแย้งในวิธีบริหารงานของฝ่ายบริหาร และถ้าพบว่าสัดส่วนของพนักงานดังกล่าวมีค่าสูงกว่า .30 อย่างมีนัยสำคัญทางโรงงานจะปรับปรุงฝ่ายบริหารเสียใหม่

โดยที่โรงงานแห่งนี้สามารถจำแนกพนักงานออกได้เป็น 4 ฝ่ายตามประเภทของสินค้าที่ผลิต

และจากการทดลองสำรวจเบื้องต้น ปรากฏข้อมูลดังนี้

ฝ่าย	จำนวนพนักงาน (N_h)	ตัวอย่าง (n_h)	สัดส่วนของพนักงานที่รู้สึกขัดแย้ง (P_h)
1	2000	100	.20
2	1600	80	.30
3	1200	60	.40
4	1200	60	.30
รวม	6000	300	

ก. จงกะประมาณสัดส่วนของพนักงานที่มีความรู้สึกขัดแย้งกับฝ่ายบริหาร

ข. จงทดสอบสมมุติฐานว่าโรงงานแห่งนี้มีพนักงานเกินกว่า 30% ที่มีความรู้สึกขัดแย้งกับการบริหารงานของฝ่ายบริหาร (สมมุติว่าตัวแปรสุ่มมีการกระจายแบบปกติ σ^2 unknown)

ค. จงอาศัยข้อมูลการสำรวจเบื้องต้นกำหนดขนาดตัวอย่าง แล้วจัดสรรตัวอย่าง โดยขนาดตัวอย่างดังกล่าวต้องให้ผลด้านความแม่นยำสูงมาก คือผิดพลาดไม่เกิน ±4% และ มีความน่าเชื่อถือได้ถึง 99%

วิธีทำ ก. $\hat{P}_{st} = \frac{1}{N} \sum_{h}^{L} N_h p_h$

$$= \frac{1}{6000}(2000 \times .2 + 1600 \times .3 + 1200 \times .4 + 1200 \times .3)$$

$$= \frac{1720}{6000} = .2867$$

$$= 28.67\%$$

สำหรับ $\hat{V}(p_{st})$ นั้น เราสามารถเตรียมตารางวิเคราะห์ได้ดังนี้

ฝ่าย	N_h	n_h	$(N_h - n_h)/N_h$	N_h^2	p_h	q_h	$p_h q_h / n_h - 1$	$\frac{N_h - n_h}{N_h} N_h^2 \frac{p_h q_h}{n_h - 1}$
1	2000	100	.95	4000000	.2	.8	.00162	6141.41
2	1600	80	.95	2560000	.3	.7	.00266	6461.81
3	1200	60	.95	1440000	.4	.6	.00407	5564.75
4	1200	60	.95	1440000	.3	.7	.00356	4869.15
รวม	6000	300						23037.12

ดังนั้น $\hat{V}(\hat{P}_{st}) = \frac{1}{N^2}\sum_{n}^{L}\frac{N_h - n_h}{N_h} N_h^2 \frac{p_h q_h}{n_h - 1} = \frac{23037.12}{(6000)^2} = .00064$

$s(\hat{P}_{st}) = .0253$

ดังนั้นช่วงเชื่อมั่น 99% ที่คาดว่าค่าจริงของ P จะปรากฏอยู่หรือคาดว่าสัดส่วนจริงของพนักงานที่มีความรู้สึกขัดแย้งในการบริหารงานของฝ่ายบริหารจะปรากอยู่ในช่วง

$\{\hat{P}_{st} - (2.58)(s(\hat{P}_{st})),\ \hat{P}_{st} + (2.58)(s(\hat{P}_{st}))\}$

$= \{.2867 - (2.58)(.0253),\ .2867 + (2.58)(.0253)\}$

$= (.2214,\ .3520)$ หรือระหว่างประมาณ 22% - 35%

ข. $H_0 : P = .30$ vs $H_1 : P > .30$

กำหนดระดับนัยสำคัญ = 5%

test statistics $t_c = \frac{\hat{P}_{st} - P}{s(\hat{P}_{st})}$

ปฏิเสธสมมุติฐานหลักเมื่อ $t_c > t_{n-1,\alpha}$

จะพบว่า

$$t_c = \frac{.2867 - .30}{.0253} = -0.526$$

และจากตารางจะพบว่า $t_{299,.05} = 1.64$

จะเห็นว่า t_c ไม่ได้มีค่ามากกว่า 1.64 เราจึงไม่ปฏิเสธสมมุติฐานหลักได้ และเชื่อว่ามีพนักงานไม่เกิน 30% ที่สึกผิดหวังหรือรู้สึกขัดแย้งในวิธีบริหารงานของฝ่ายบริหาร

ค. การกำหนดขนาดตัวอย่าง

ในที่นี้จะแสดงเฉพาะกรณีของขนาดตัวอย่างที่สอดคล้องกับแผนการจัดสรรที่ได้สัดส่วนกับขนาดชั้นภูมิเท่านั้น อีก 3 วิธีที่เหลืออยู่จะขอเว้นไว้เป็นแบบฝึกหัด และในกรณี n_{opt} ให้ใช้ $c_1 = 10,\ c_2 = 20,\ c_3 = 30,\ c_4 = 40$

จาก $n_{prop} = \dfrac{N\Sigma N_h s_h^2}{N^2D^2 + \Sigma N_h s_h^2} = \dfrac{N\Sigma N_h(n_h p_h q_h/n_h - 1)}{N^2D^2 + \Sigma N_h(n_h p_h q_h/n_h - 1)}$

และ $D^2 = d_0^2/Z^2_{1-\alpha_0/2} = (.04)^2/(1.96)^2 = .0004164$, $N^2D^2 = 14990.4$

ดังนั้น $n_{prop} = \dfrac{6000(323.23 + 340.25 + 292.88 + 256.27)}{14990.4 + (323.23 + 340.25 + 292.88 + 256.27)}$

$$= \frac{7275780}{16203.03} = 450.04 \simeq 450$$

นั่นคือ $n_{prop} = 450$ และสามารถจัดสรรให้แก่แต่ละชั้นภูมิได้ดังนี้คือ

$n_1 = (450/6000) \times 2000 = (.075)(2000) = 150$

$n_2 = (450/6000) \times 1600 = (.075)(1600) = 120$

$n_3 = (450/6000) \times 1200 = (.075)(1200) = 90$

$n_4 = (450/6000) \times 1200 = (.075)(1200) = 90$