$$= \frac{1}{N^2}\sum_{h}^{L}\frac{N_h^2S_h^2}{n_h} - \frac{1}{N^2}\sum_{h}^{L}N_hS_h^2 \quad \text{เมื่อ } n_h = \frac{N_hS_h}{\Sigma N_hS_h}\cdot n$$

หรือแทนค่า $n_h = \frac{N_hS_h}{\Sigma N_hS_h}\cdot n$ ลงในสูตรจะพบว่า

$$V(\bar{x}_{ney}) = \frac{1}{N^2}\sum_{h}^{L} N_h^2S_h^2/(n.N_hS_h/\sum_{h}^{L}N_hS_h) - \frac{1}{N^2}\sum_{h}^{L}N_hS_h^2$$

$$= \frac{1}{N^2}\cdot\frac{1}{n}(\sum_{h}^{L}N_hS_h)^2 - \frac{1}{N^2}\sum_{h}^{L}N_hS_h^2$$

3. $$\hat{V}(\bar{x}_{ney}) = \frac{1}{N^2}\cdot\frac{1}{n}(\sum_{h}^{L}N_hs_h)^2 - \frac{1}{N^2}\sum_{h}^{L}N_hs_h^2$$

4. $$\hat{T}_{ney} = \sum_{h}^{L}N_h\bar{x}_h \;;\; \bar{x}_h = \frac{1}{n_h}\sum_{i}^{n_h}x_{hi} \quad \text{โดยที่ } n_h = \frac{N_hS_h}{\Sigma N_hS_h}\cdot n$$

5. $$V(\hat{T}_{ney}) = \frac{1}{n}(\sum_{h}^{L}N_hS_h)^2 - \sum_{h}^{L}N_hS_h^2$$

6. $$\hat{V}(\hat{T}_{ney}) = \frac{1}{n}(\sum_{h}^{L}N_hs_h)^2 - \sum_{h}^{L}N_hs_h^2$$

ตัวอย่าง ๕.๕ ต้องการกะประมาณยอดขายของร้านอาหารในเมืองแห่งหนึ่ง ซึ่งมีทั้งสิ้น 1,000 ร้าน ซึ่งสามารถจำแนกร้านค้าเหล่านี้เป็น 3 ชั้นภูมิตามจำนวนพนักงานบริการ คือ ชั้นภูมิที่ 1 ร้านขนาดเล็ก 600 ร้าน ชั้นภูมิที่ 2 ร้านขนาดกลาง 300 ร้าน และชั้นภูมิที่ 3 ร้านขนาดใหญ่ 100 ร้าน

จากการสำรวจเบื้องต้นพบว่า $s_1^2 = 100$ บาท $s_2^2 = 400$ บาท และ $s_3^2 = 900$ บาท และสมมุติว่า $\bar{x}_1 = 40$ หน่วยบาท/วัน $\bar{x}_2 = 60$ หน่วยบาท/วัน $\bar{x}_3 = 80$ หน่วยบาท/วัน และกำหนดให้ขนาดตัวอย่าง $n = 300$ ร้าน

ก. จงจัดสรรตัวอย่างให้แก่แต่ละชั้นภูมิโดยวิธีจัดสรรแบบอุตมะ พร้อมทั้งประมาณยอดขายเฉลี่ยของร้านอาหารหนึ่ง ๆ ในเมืองดังกล่าวในแต่ละวัน พร้อมทั้งช่วงเชื่อมั่น 95% กำหนดให้ค่าใช้จ่ายในการสำรวจร้านอาหารในแต่ละชั้นภูมิดังนี้ $c_1 = 9$ บาท $c_2 = 16$ บาท $c_3 = 25$ บาท

ข. จงจัดสรรตัวอย่างให้แก่แต่ละชั้นภูมิโดยวิธีจัดสรรแบบเนย์แมน พร้อมทั้งประมาณยอดขายเฉลี่ยและยอดขายรวมของร้านอาหารทุกร้านในแต่ละวัน พร้อมทั้งคำนวณหาช่วงเชื่อมั่น 99% ของการประมาณค่า กำหนดให้ $c_1 = 10$ บาท $c_2 = 10.25$ บาท และ $c_3 = 9.90$ บาท

วิธีทำ

ก. การจัดสรรแบบอุตมะ

จากโจทย์ เราสามารถเตรียมตารางข้อมูลสำหรับการจัดสรรตัวอย่างและประมาณค่าพารามิเตอร์ได้ดังนี้

ชั้นภูมิ (ร้านอาหาร)	N_h	s_h	$N_h s_h$	$\sqrt{c_h}$	$N_h s_h/\sqrt{c_h}$	$(N_h s_h/\sqrt{c_h})/(\Sigma N_h s_h/\sqrt{c_h})$
ขนาดเล็ก	600	10	6000	3	2000	2000/4100 = .483
ขนาดกลาง	300	20	6000	4	1500	1500/4100 = .366
ขนาดใหญ่	100	30	3000	5	600	600/4100 = .146
รวม	1000		15000		4100	

จาก $n_h = (N_h s_h/\sqrt{c_h}) / (\sum_{h}^{L} N_h s_h/\sqrt{c_h}).n \; ; h = 1,2,\ldots,L$

ดังนั้น $n_1 = (.488)(300) \simeq 146$ $n_2 = (.366)(300) \simeq 110$

$n_3 = (.146)(300) \simeq 44$

และสำหรับการประมาณค่าพารามิเตอร์เราสามารถเตรียมตารางวิเคราะห์ได้ดังนี้

ชั้นภูมิ (ร้านอาหาร)	N_h	n_h	$\bar{x}_h$	s_h^2	$N_h\bar{x}_h$	$\frac{N_h-n_h}{N_h}$	$\frac{N_h^2 s_h^2}{n_h}$	$\frac{N_h-n_h}{N_h}\frac{N_h^2 s_h^2}{n_h}$
ขนาดเล็ก	600	146	40	100	24000	.757	246575.34	186657.53
ขนาดกลาง	300	110	60	400	18000	.633	327272.72	207163.63
ขนาดใหญ่	100	44	80	900	8000	.560	204545.44	114545.44
รวม	1000	300			50000			508366.60

ดังนั้น

(1) $\bar{x}_{opt} = 50000/1000 = 50$

(2) $\hat{T}_{opt} = 50000$

(3) $\hat{V}(\bar{x}_{opt}) = \frac{1}{N^2} \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 s_h^2}{n_h} = \frac{508366.60}{(1000)^2} = .508$

(4) ช่วงเชื่อมั่น 95% ที่คาดว่าค่าจริงของ $\bar{X}$ จะปรากฏอยู่คือ

$$\{\bar{x}_{opt} - (1.96)(\sqrt{.508}), \bar{x}_{opt} + (1.96)(\sqrt{.508})\} = (48.60, 51.39)$$

(5) $\hat{V}(\hat{T}_{opt}) = \sum_{h}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 s_h^2}{n_h} = 508366.6$

(6) ช่วงเชื่อมั่น 95% ที่คาดว่าค่าจริงของ T จะปรากฏอยู่คือ

$$(\hat{T}_{opt} - 1.96\sqrt{508366.6}, \hat{T}_{opt} + 1.96\sqrt{508366.6})$$

$$= (50000 - 1397.48, 50000 + 1397.48)$$

$$= (48602.52, 51397.48)$$

ข. การจัดสรรแบบเนย์แมน[1]

อาศัยข้อมูล จากข้อ ก. (ตารางที่ 1) เราสามารถจัดสรรขนาดตัวอย่างตามวิธีของเนย์แมนได้ดังนี้

$$\because n_h = \frac{N_h s_h}{\Sigma N_h s_h} \cdot n \ ; h = 1,2,\ldots,L$$

ดังนั้น $n_1 = \frac{6000}{15000} \cdot 300 = 120$, $n_2 = \frac{6000}{15000} \cdot 300 = 120$, $n_3 = \frac{3000}{15000} \cdot 300 = 60$ และเราสามารถเตรียมตารางสำหรับประมาณค่าพารามิเตอร์ได้ดังนี้

ชั้นภูมิ (ร้านอาหาร)	N_h	n_h	$\bar{x}_h$	s_h^2	$N_h\bar{x}_h$	$\frac{N_h - n_h}{N_h}$	$\frac{N_h^2 s_h^2}{n_h}$	$\frac{N_h - n_h}{N_h} \frac{N_h^2 s_h^2}{n_h}$
ขนาดเล็ก	600	120	40	100	2400	0.8	300000	240000
ขนาดกลาง	300	120	60	400	18000	0.6	300000	180000
ขนาดใหญ่	100	60	80	900	8000	0.4	150000	60000
รวม	1000	300			50000			480000

[1] กรณีของข้อ ข. ควรใช้วิธีจัดสรรแบบเนย์แมนเพราะค่าใช้จ่ายในการสำรวจต่อหน่วยในชั้นภูมิทั้งสามมีค่าใกล้เคียงกัน

ดังนั้น

(1) $\bar{x}_{ney} = \dfrac{50000}{1000} = 50$

(2) $\hat{T}_{ney} = 50000$

(3) $\hat{V}(\bar{x}_{ney}) = \dfrac{480000}{(1000)^2} = .48$

(4) ช่วงเชื่อมั่น 99% ที่คาดว่าค่าจริงของ $\bar{X}$ จะปรากฏอยู่คือ

$$\{\bar{x}_{ney} - 2.58\sqrt{.48}, \bar{x}_{ney} + 2.58\sqrt{.48}\} = (48.21, 51.79)$$

(5) $\hat{V}(\hat{T}_{ney}) = 480000$

(6) ช่วงเชื่อมั่น 99% ที่คาดว่าค่าจริงของ T จะปรากฏอยู่คือ

$$\{\hat{T}_{ney} - 2.58\sqrt{480000}, \hat{T}_{ney} + 2.58\sqrt{480000}\} = (48212.52, 51787.48)$$

3.3.4 การกำหนดขนาดตัวอย่าง (Determination of Sample Size)

ในตอนที่ผ่านมาการศึกษาทั้งในเรื่องของคุณสมบัติของตัวประมาณค่าคือ $\bar{x}_{st}$ และ $\hat{T}_{st}$ และเรื่องการจัดสรรตัวอย่างให้แก่ชั้นภูมินั้นเราถือว่าทราบขนาดตัวอย่างเป็นเท่าไรแล้วโดยการสมมุติ ปัญหาก็คือขนาดตัวอย่างที่พึงใช้เพื่อการประมาณค่าเฉลี่ยและยอดรวมที่เหมาะสมควรมีค่าเท่าไร? ซึ่งจะเป็นเรื่องที่กล่าวถึงในลำดับต่อไป

ก่อนอื่นขอทำความเข้าใจไว้ในชั้นนี้เสียก่อนว่า ก่อนที่จะมีการกำหนดขนาดตัวอย่างได้นั้น นักวิจัยจะต้องทราบหรือต้องกำหนดให้ชัดเจนเสียก่อนว่าจะใช้แผนการจัดสรรตัวอย่างให้แก่ชั้นภูมิด้วยวิธีใด ซึ่งหลักเกณฑ์การจัดสรรได้กล่าวถึงมาแล้วโดยละเอียดในตอน 3.3.3 ทั้งนี้เพราะถ้าไม่กำหนดแผนการจัดสรรตัวอย่างไว้เสียแต่ต้นจะเกิดปัญหาในภายหลัง ทั้งนี้เพราะแผนการจัดสรรแต่ละแผนส่งผลสะท้อนต่อความแม่นยำของงานโดยตรง แต่ละวิธี

จะให้ค่า $V(\hat{\theta})$ มากน้อยแตกต่างกัน ซึ่งความเหลื่อมล้ำกันในค่าของ $V(\hat{\theta})$ ย่อมหมายถึงการใช้ขนาดตัวอย่างที่ต่างกัน[1] ดังนั้นถ้ากำหนดขนาดตัวอย่างไว้กลาง ๆ ไม่อ้างอิงแผนการจัดสรรแบบใดโดยมุ่งหมายว่าจะจัดสรรตัวอย่างเอาเองในภายหลังการกระทำดังนี้จะก่อให้เกิดความลำบากใจให้แก่นักวิจัยเพราะไม่อาจแน่ใจได้ว่างานหรือโครงการนั้นจะมีความแม่นยำและความน่าเชื่อถืออยู่ในระดับใด

ดังนั้น เราจึงจำเป็นต้องวางแผนไว้ก่อนว่าจะจัดสรรตัวอย่างด้วยวิธีใด จากนั้นจึงกำหนดขนาดตัวอย่างให้สอดคล้องกับแผนการนั้น เมื่อได้ขนาดตัวอย่างมาแล้วจึงค่อยจัดสรรให้แก่ชั้นภูมิตามแผนการณ์เดิม การกระทำเช่นนี้ย่อมมีผลให้นักวิจัยสามารถควบคุมระดับความแม่นยำและความน่าเชื่อถือของงานได้ตั้งแต่ต้นจนเสร็จสิ้นโครงการ

การกำหนดขนาดตัวอย่างให้สอดคล้องกับระดับความแม่นยำ ระดับความน่าเชื่อถือและแผนการการจัดสรรตัวอย่างให้แก่ชั้นภูมิจะอาศัยความสัมพันธ์เดิมคือ

$$d^2 = V(\bar{x})Z^2_{1-\alpha/2}$$

และเพื่อความสะดวกกำหนดให้ $D^2 = d^2/Z^2_{1-\alpha/2}$ ดังนั้น $D^2 = V(\bar{x})$ คือความสัมพันธ์ที่จะใช้เป็นหลักในการกำหนดขนาดตัวอย่างต่อไป การกำหนดขนาดตัวอย่างสามารถพัฒนาได้โดยง่ายทำนองเดียวกันกับที่แสดงให้ดูมาแล้วโดยละเอียดในบทที่ 2 ในที่นี้จะแสดงให้ดูเพียงบางตอนเท่านั้น สูตรที่เหลืออยู่จะเว้นไว้ให้นักศึกษาพิสูจน์เองเป็นแบบฝึกหัด

3.3.4.1 การกำหนดขนาดตัวอย่างเพื่อกะประมาณค่าเฉลี่ย $\bar{X}$

ขนาดตัวอย่างที่จะใช้ในการกะประมาณค่าเฉลี่ย $\bar{X}$ ให้สอดคล้องกับระดับความแม่นยำความน่าเชื่อถือและแผนการจัดสรรตัวอย่างปรากฏดังนี้

---

[1] เรื่องนี้เป็นการมองในมุมกลับ กล่าวคือ ถ้า $n \to \infty$ จะมีผลให้ $V(\bar{x}) \to 0$ เมื่อ $V(\bar{x}) = \sigma^2/n$ ขณะเดียวกันในมุมกลับ ถ้า $\sigma^2 \to \infty$ จะมีผลให้ $n \to \infty$ และถ้า $\sigma^2 \to 0$ จะมีผลให้ $n \to 1$

1. $n_{eq} \quad = \quad (L \sum_{h}^{L} N_h^2 S_h^2)/N^2D^2$ เมื่อ $D^2 = d_0^2/Z^2_{1-\alpha_0/2}$

2. $n_{prop} \quad = \quad (N \sum_{h}^{L} N_h S_h^2)/(N^2D^2 + \sum_{h}^{L} N_h S_h^2)$

3. $n_{opt} \quad = \quad (\sum_{h}^{L} N_h S_h \sqrt{c_h})(\sum_{h}^{L} N_h S_h/\sqrt{c_h})/(N^2D^2 + \sum_{h}^{L} N_h S_h^2)$

4. $n_{ney} \quad = \quad (\sum_{h}^{L} N_h S_h)^2/(N^2D^2 + \sum_{h}^{L} N_h S_h^2)$

หมายเหตุ

1. n_{eq} หมายถึงขนาดตัวอย่างที่กำหนดไว้เพื่อจัดสรรอย่างเท่าเทียมกัน
   n_{prop} หมายถึงขนาดตัวอย่างที่กำหนดไว้เพื่อจัดสรรอย่างได้สัดส่วนกับชั้นภูมิ
   n_{opt} หมายถึงขนาดตัวอย่างที่กำหนดไว้สำหรับจัดสรรแบบอุตมะ
   n_{ney} หมายถึงขนาดตัวอย่างที่กำหนดไว้สำหรับการจัดสรรแบบเนย์แมน

2. ถ้าไม่ทราบค่า S_h^2 ให้ใช้ s_h^2 ซึ่งได้จากงานสำรวจเบื้องต้นแทนต่อไปนี้ จะพิสูจน์ให้เห็นถึงความเป็นมาหรือการพัฒนาสูตร n_{opt} ส่วนการพิสูจน์สูตรสำหรับ n_{eq} และ n_{prop} และ n_{ney} จะขอเว้นไว้เป็นแบบฝึกหัด อนึ่ง ในทางปฏิบัติเราไม่มีความจำเป็นใด ๆ ที่จะต้องจดจำสูตรเหล่านี้ เพราะเราสามารถคำนวณหาค่า n ได้โดยตรงจากความสัมพันธ์ $d^2 = V(\bar{x})Z^2_{1-\alpha/2}$ หรือ $D^2 = V(\bar{x})$ แต่ ถ้าสามารถจำได้ก็เป็นสิ่งที่ดีเพราะช่วยประหยัดเวลาทำงานลดขั้นตอนการทำงานลงได้มาก

พิสูจน์ กำหนดให้ $d^2 = d_0^2$ และ $\alpha = \alpha_0$ ดังนั้น $D^2 = d_0^2/Z^2_{1-\alpha_0/2}$

$$\because \quad V(\bar{x}_{opt}) = \frac{1}{N^2}\cdot\frac{1}{n}(\sum_{h}^{L} N_h S_h \sqrt{c_h})(\sum_{h}^{L} N_h S_h/\sqrt{c_h}) - \frac{1}{N^2}\sum_{h}^{L} N_h S_h^2$$

ดังนั้น $D^2 = \frac{1}{N^2} \cdot \frac{1}{n} \left(\sum_h^L N_h S_h \sqrt{c_h} \right) \left(\sum_h^L N_h S_h / \sqrt{c_h} \right) - \frac{1}{N^2} \sum_h^L N_h S_h^2$

$$D^2 + \frac{1}{N^2} \sum_h^L N_h S_h^2 = \frac{1}{N^2} \cdot \frac{1}{n} \left(\sum_h^L N_h S_h / \sqrt{c_h} \right) \left(\sum_h^L N_h S_h \sqrt{c_h} \right)$$

ดังนั้น $N^2 D^2 + \sum_h^L N_h S_h^2 = \frac{1}{n} \left(\sum_h^L N_h S_h \sqrt{c_h} \right) \left(\sum_h^L N_h S_h / \sqrt{c_h} \right)$

$$n_{opt} = \left(\sum_h^L N_h S_h \sqrt{c_h} \right) \left(\sum_h^L N_h S_h / \sqrt{c_h} \right) / \left(N^2 D^2 + \sum_h^L N_h S_h^2 \right)$$

ในทางทฤษฎีนั้น การกำหนดขนาดตัวอย่าง (ของแผนสำรวจทุกแผน) สำหรับกรณีที่ไม่ทราบค่า S_h^2 ต้องเสียเวลาคำนวณหลายรอบหมายความว่าต้องสุ่มตัวอย่างเพื่อการคำนวณหา s_h^2 และ n สลับกันหลายครั้งหลายคราว และจะต้องกระทำติดต่อกันไปเรื่อย ๆ จนกว่า n ในรอบที่ติดต่อกันมีค่าเท่ากันจึงจะหยุด แล้วถือว่า n ค่านั้นคือค่าที่ถูกต้อง ตัวอย่างเช่นการคำนวณหา n_{opt} จะต้องดำเนินการดังนี้

รอบที่ 1 ให้เดาขนาดตัวอย่าง n ขึ้นมาซึ่งเป็นเท่าไรก็ได้ แล้วจัดสรรตัวอย่างให้แก่แต่ละชั้นภูมิตามที่คิดว่าเหมาะสม (เดาเช่นกัน) แต่ควรจะจัดสรรโดยวิธีที่เป็นสัดส่วนกับขนาดชั้นภูมิ ที่ต้องทำเช่นนี้เพราะเรายังไม่ทราบค่า S_h^2 หรือ s_h^2 ทำให้ไม่อาจจัดสรรตัวอย่างตามวิธีของการจัดสรรแบบอุตมะได้

หลังจากนั้นจึงดำเนินการสุ่มตัวอย่างโดยวิธี SRS จากแต่ละชั้นภูมิ นำข้อมูลมาวิเคราะห์หาค่า s_h^2; h = 1,2,...,L แล้วแทนค่า s_h^2 เหล่านี้ลงในสูตร

$$n_{opt} = \left(\sum_h^L N_h s_h \sqrt{c_h} \right) \left(\sum_h^L N_h s_h / \sqrt{c_h} \right) / \left(N^2 D^2 + \sum_h^L N_h s_h^2 \right)$$

ผลลัพธ์คือ n_{opt} รอบที่ 1

รอบที่ 2 นำ n_{opt} ในรอบที่ 1 มาจัดสรรแบบอุตมะดังนี้

$$n_h = n(N_h s_h/\sqrt{c_h})/(\sum_h^L N_h s_h/\sqrt{c_h})\ ;\ h = 1,2,\ldots,L$$

ทั้งนี้อาศัย s_h^2 จากรอบที่ 1

นำขนาดตัวอย่างที่จัดสรรแล้วไปสุ่มตัวอย่างโดยวิธี SRS จากแต่ละชั้นภูมิแล้วนำข้อมูลมาวิเคราะห์หาค่า s_h^2 ; $h = 1,2,\ldots,L$ แล้วแทนค่าลงในสูตรของ n_{opt} ผลลัพธ์คือ n_{opt} รอบที่ 2

ถ้า n_{opt} รอบที่ 1 $\neq n_{opt}$ รอบที่ 2 ให้ดำเนินการสำรวจรอบที่ 3

รอบที่ 3 นำ n_{opt} ในรอบที่ 2 มาจัดสรรแบบอุตมะโดยอาศัยค่า s_h^2 จากรอบที่ 2 แล้วนำขนาดตัวอย่างที่จัดสรรแล้วไปสุ่มตัวอย่างโดยวิธี SRS จากแต่ละชั้นภูมิ นำข้อมูลมาวิเคราะห์หา s_h^2 ; $h = 1,2,\ldots,L$ แล้วแทนค่าลงในสูตรของ n_{opt} ผลลัพธ์ก็คือ n_{opt} รอบที่ 3

ถ้า n_{opt} รอบที่ 3 $\neq n_{opt}$ รอบที่ 2 ให้ดำเนินการสำรวจรอบที่ 4

การดำเนินการให้วนเวียนเช่นนี้ไปจนกว่า n_{opt} ในรอบที่ j $\simeq n_{opt}$ ในรอบที่ j − 1 จึงหยุดดำเนินการและถือว่าค่า n_{opt} รอบที่ j คือขนาดตัวอย่างที่จะใช้สำรวจต่อไป

จะเห็นได้ว่าวิธีการเช่นนี้สิ้นเปลืองค่าใช้จ่ายและเวลามาก อย่าลืมว่าการสำรวจรอบหนึ่ง ๆ ต้องเสียค่าใช้จ่ายในการสำรวจและเวลาดำเนินการสำรวจและประมวลผลมิใช่น้อย ซึ่งถ้านำมาใช้ในทางปฏิบัติจะเกิดปัญหาหลายประการโดยเฉพาะเกี่ยวกับงบประมาณเพื่อการวิจัยและแผนการดำเนินงานซึ่งกำหนดเวลาหรือหมายกำหนดการไว้เรียบร้อยแล้ว ดังนั้นในทางปฏิบัติเราจึงไม่นิยม แต่จะเลือกปฏิบัติวิธีใดวิธีหนึ่งดังต่อไปนี้คือ

วิธีที่ 1 พยายามศึกษางานด้านเอกสารโดยพยายามค้นคว้าหรือเสาะหาข้อมูลในอดีตเพื่อเป็นแนวทางในการกำหนดขนาดตัวอย่าง ข้อมูลที่ปรารถนาก็คือ s^2 หรือ s_h^2 จากงานสำรวจ หรืองานวิจัยทำนองเดียวกันหรือโครงการที่คล้ายคลึงกัน และถ้าปรารถนาจะคำนวณหาขนาดตัวอย่างโดยอาศัย CV ให้พยายามเสาะหาค่า CV จากโครงการต่าง ๆ ที่มีลักษณะคล้ายคลึงกัน

วิธีที่ 2 ทำการทดลองสำรวจหรือการสำรวจเบื้องต้น (Preliminary Survey) การทดลองสำรวจหรือการสำรวจเบื้องต้นนั้นโดยปกติจะดำเนินการเพียงรอบเดียวแล้วนำค่า s^2 หรือ s_h^2 ; $h = 1,2,\ldots,L$ มาใช้ประโยชน์ทันทีไม่จำเป็นต้องดำเนินการรอบที่ 2,3,... ตามทฤษฎี และเพื่อให้การสำรวจเบื้องต้นได้ผลคุ้มค่านักวิจัยพึงเดาขนาดตัวอย่างให้ใหญ่พอสมควร จะให้ใหญ่เท่าไรก็ได้แต่ไม่ควรต่ำกว่า 30 หน่วย และควรดำเนินการสำรวจตามแผนการสำรวจที่วางไว้ทุกประการ การดำเนินการสำรวจตามแผนเดิมมีผลดีที่เด่นชัดคือ ถ้าขนาดตัวอย่างที่คำนวณได้ในภายหลังมีขนาดใหญ่กว่าขนาดตัวอย่างที่เดาไว้ในครั้งแรก เราจะถือว่าขนาดตัวอย่างที่ใช้ทดลองสำรวจเป็นส่วนหนึ่งของขนาดตัวอย่างที่จะใช้สำรวจจริงและถือว่าได้ดำเนินการสำรวจในส่วนนั้นไปแล้ว การสำรวจจริงจึงสำรวจเฉพาะหน่วยที่เหลือ แต่ถ้าขนาดตัวอย่างที่คำนวณได้มีค่าน้อยกว่าขนาดตัวอย่างทดลอง ให้ถือว่าขนาดตัวอย่างทดลองเป็นขนาดตัวอย่างจริงและให้ถือว่าข้อมูลจากการสำรวจเบื้องต้นเป็นข้อมูลของการสำรวจจริงไม่จำเป็นต้องทำการสำรวจอีกต่อไป แต่นักวิจัยจะต้องคำนวณหาค่า d^2 และ α ในภายหลังเพื่อสอบย้ำระดับความแม่นยำและความน่าเชื่อถือของงาน อนึ่งกรณีนี้เหมาะสำหรับงานสำรวจที่แบบสำรวจ (แบบสอบถาม) ผ่านการตรวจสอบมาดีแล้วเท่านั้น ถ้าแบบสอบถามที่ใช้เป็นเพียงแบบทดลองไม่ควรถือเอาผลการทดลองสำรวจเป็นผลของการสำรวจจริง

3.3.4.2 การกำหนดขนาดตัวอย่างเพื่อกะประมาณยอดรวม

การกำหนดขนาดตัวอย่างเพื่อกะประมาณยอดรวมให้สอดคล้องกับระดับความแม่นยำ ความน่าเชื่อถือและแผนการจัดสรรตัวอย่างนั้น ให้ดำเนินการในทำนองเดียวกันกับในตอน 3.3.4.1 คือ กำหนดขนาดตัวอย่างขึ้นโดยอาศัยความสัมพันธ์ $d^2 = V(\hat{T}_{st})Z^2_{1-\alpha/2}$ หรือ $D^2 = V(\hat{T}_{st})$ เมื่อ $D^2 = d_0^2/Z^2_{1-\alpha/2}$ กล่าวคือ

n_{eq} พัฒนาขึ้นมาจากความสัมพันธ์ $D^2 = V(\hat{T}_{eq})$

n_{prop} พัฒนาขึ้นมาจากความสัมพันธ์ $D^2 = V(\hat{T}_{prop})$

n_{opt} พัฒนาขึ้นมาจากความสัมพันธ์ $D^2 = V(\hat{T}_{opt})$

n_{ney} พัฒนาขึ้นมาจากความสัมพันธ์ $D^2 = V(\hat{T}_{ney})$

ในที่นี้จะไม่เสนอสูตรค่า n ต่าง ๆ เหล่านี้แต่จะขอเว้นไว้เป็นแบบฝึกหัดซึ่งนักศึกษาสามารถพิสูจน์หรือพัฒนาขึ้นมาได้เองโดยง่าย

ตัวอย่าง 3.4 ในการสำรวจภาวะเศรษฐกิจของครอบครัวในชุมชนแห่งหนึ่งซึ่งประกอบด้วยครัวเรือนทั้งสิ้น 1,000 ครัวเรือน จำแนกออกได้เป็น 3 ชั้นภูมิตามจำนวนสมาชิกของครัวเรือนคือ ครัวเรือนที่มีสมาชิกไม่เกิน 3 คนถือว่าเป็นครัวเรือนขนาดเล็กมีอยู่ทั้งสิ้น 600 ครัวเรือน ครัวเรือนที่มีสมาชิก 4-8 คนถือว่าเป็นครัวเรือนขนาดกลางมีอยู่ทั้งสิ้น 300 ครัวเรือน และครัวเรือนที่มีสมาชิกตั้งแต่ 9 คนขึ้นไปถือว่าเป็นครัวเรือนขนาดใหญ่มีอยู่ทั้งสิ้น 100 ครัวเรือน

จากการทดลองสำรวจเบื้องต้นเฉพาะรายการข้อมูล (Field) เกี่ยวกับค่าใช้จ่ายด้านอาหาร พบว่าแต่ละชั้นภูมิมีความแปรปรวนภายใน (Stratum Variance) เป็น 400 บาท 900 บาท และ 2,500 บาท ตามลำดับ

ก. จงกำหนดขนาดตัวอย่างเพื่อกะประมาณค่าใช้จ่ายเฉลี่ยด้านอาหารของแต่ละครัวเรือนให้สอดคล้องกับระดับความถูกต้องแม่นยำ $d = \pm 3$ บาท และระดับความน่าเชื่อถือ 99% ทั้งนี้ให้คำนวณขนาดตัวอย่างไว้สำหรับแผนจัดสรรตัวอย่างทุกแผนพร้อมทั้งจัดสรรตัวอย่างให้แก่แต่ละชั้นภูมิ สำหรับกรณีของแผนจัดสรรแบบอุตมะให้ถือว่า $c_1 = 25$ บาท- $c_2 = 30$ บาท $c_3 = 44$ บาท

ข. จงกำหนดขนาดตัวอย่างเพื่อประมาณยอดรวมค่าใช้จ่ายด้านอาหารของท้องที่ดังกล่าว ทั้งนี้ให้อาศัยเงื่อนไขเดียวกันกับข้อ ก.

วิธีทำ

ก. จากข้อมูลเราสามารถเตรียมตารางสำหรับวิเคราะห์ได้ดังนี้

ชั้นภูมิ (ขนาดครัวเรือน)	N_h	s_h	s_h^2	$N_h s_h$	$N_h s_h^2$	$N_h^2 s_h^2$	$\sqrt{c_h}$	$N_h s_h \sqrt{c_h}$	$N_h s_h / \sqrt{c_h}$
ขนาดเล็ก	600	20	400	12000	240000	144000000	5.00	60000	2400.00
ขนาดกลาง	300	30	900	9000	270000	81000000	5.48	43920	1642.33
ขนาดใหญ่	100	50	2500	5000	250000	25000000	6.63	33150	754.15
รวม	1000			26000	760000	250000000		136070	4796.48

$$D^2 = d_0^2/z^2_{1-\frac{\alpha_0}{2}} = (\pm 3)^2/(2.58)^2 = 1.3520822$$

$$N^2D^2 = 1,352,082.2$$

ดังนั้น $N^2D^2 + \sum_{h}^{3} N_h s_h^2 = 1,352,082.2 + 760,000 = 2,112,082.2$

(1) กรณีจัดสรรอย่างเท่าเทียมกัน

$$n_{eq} = L \sum_{h}^{L} N_h^2 s_h^2 / (N^2D^2 + \sum_{h}^{L} N_h s_h^2)$$

$$= \frac{3(250,000,000)}{2,112,082.2} = 355.09 = 355$$

ดังนั้น $n_h = n/L = 355/3 = 118.33$; $h = 1,2,3$ หรือ $n_1 = 119, n_2 = 118, n_3 = 118$

(2) กรณีจัดสรรให้ได้สัดส่วนกับขนาดชั้นภูมิ

$$n_{prop} = N \sum_{h}^{L} N_h s_h^2 / (N^2 D^2 + \sum_{h}^{L} N_h s_h^2)$$

$$= \frac{1{,}000(760{,}000)}{2{,}112{,}082.2} = 359.8 \cong 360$$

ดังนั้น $n_1 = N_1(n/N) = 600(360/1{,}000) \cong 216$

$n_2 = N_2(n/N) = 300(360/1{,}000) \cong 108$

$n_3 = N_3(n/N) = 100(360/1{,}000) \cong 36$

(3) กรณีจัดสรรแบบอุตมะ

$$n_{opt} = (\sum_{h}^{L} N_h s_h \sqrt{c_h})(\sum_{h}^{L} N_h s_h / \sqrt{c_h}) / (N^2 D^2 + \sum_{h}^{L} N_h s_h^2)$$

$$= \frac{(136{,}070)(4{,}796.48)}{2{,}112{,}082.2} \cong 309$$

ดังนั้น $n_1 = n(N_1 s_1/\sqrt{c_1})/(\sum_{h}^{3} N_h s_h/\sqrt{c_h} = 309(2{,}400/4{,}796.48)$

$\cong 155$

$n_2 = n(N_2 s_2/\sqrt{c_2})/(\sum_{h}^{3} N_h s_h/\sqrt{c_h}) = 309(1{,}642.33/4{,}796.48)$

$\cong 106$

$n_3 = n(N_3 s_3/\sqrt{c_3})/(\sum_{h}^{3} N_h s_h/\sqrt{c_h}) = 309(754.15/4{,}796.48)$

$\cong 48$

(4) กรณีจัดสรรแบบเนย์แมน

$$n_{ney} = \left(\sum_h^L N_h s_h\right)^2 / \left(N^2 D^2 + \sum_h^L N_h s_h^2\right)$$

$$= (26{,}000)^2 / 2{,}112{,}082.2$$

$$\cong 320$$

ดังนั้น $n_1 = n(N_1 s_1) / \sum_h^L N_h s_h = 320(12{,}000/26{,}000) \cong 148$

$$n_2 = n(N_2 s_2 / \sum_h^3 N_h s_h = 320(\ 9{,}000/26{,}000) \cong 111$$

$$n_3 = n(N_3 s_3) / \sum_h^L N_h s_h = 320(\ 5{,}000/26{,}000) \cong 61$$

ข. การกำหนดขนาดตัวอย่างเพื่อกะประมาณยอดรวม

ในที่นี้จะแสดงให้ดูเฉพาะกรณี n_{opt} เท่านั้น กรณีอื่นจะเว้นไว้เป็นแบบฝึกหัด

จาก $D^2 = \hat{V}(\hat{T}_{opt}) = \frac{1}{n}\left(\sum_h^L N_h s_h \sqrt{c_h}\right)\left(\sum_h^L N_h s_h/\sqrt{c_n}\right) - \sum_h^L N_h s_h^2$

$$1.3520822 = \frac{1}{n}(136{,}070)(4{,}796.48) - 760{,}000$$

$$n_{opt} = (136{,}070)(4{,}796.48)/(1.3520822 + 760{,}000)$$

$$\cong 859$$

ดังนั้น $n_1 = 859(2{,}400/4{,}796.48) \cong 430$

$$n_2 = 859(1{,}642.33/4{,}796.48) \cong 294$$

$$n_3 = 859(754.15/4{,}796.48) \cong 135$$

ข้อสังเกต จากตัวอย่างขอให้ข้อสังเกตเพิ่มเติมดังนี้

1. การสร้างสูตรสำเร็จสำหรับกำหนดขนาดตัวอย่างมิใช่สิ่งจำเป็นและการวิเคราะห์เพื่อคำนวณหาขนาดตัวอย่างโดยอาศัยสูตรสำเร็จมิได้ง่ายกว่าการวิเคราะห์โดยอาศัยความสัมพันธ์ $D^2 = V(\hat{\theta})$ แต่ประการใด แต่กลับพบว่าการคำนวณโดยอาศัยความสัมพันธ์ $D^2 = V(\hat{\theta})$ กระทำได้ง่ายกว่า อีกทั้งยังไม่ต้องเสียเวลาจดจำสูตรสำหรับกำหนดขนาดตัวอย่างให้สิ้นเปลืองสมองโดยไม่จำเป็นอีกด้วย

2. ความสัมพันธ์ $D^2 = V(\hat{\theta})$ เมื่อ $D^2 = d^2/Z^2_{1-\alpha/2}$ นั้น เราเรียกค่า D^2 ว่า Desired Variance หมายความว่าการกำหนดค่า D^2 (ซึ่งก็คือการกำหนดค่า d_0^2 และ α_0) มีผลเชื่อมโยงไปถึงการกำหนดค่าความแปรปรวนของการประมาณค่าด้วย ตามตัวอย่าง 3.4 เรากำหนดให้ $D^2 = (\pm 3)^2/(2.58)^2 = 1.3520822$ หมายความว่าเรากำหนดให้ $\hat{V}(\bar{x}) = 1.3520822$ และ $\hat{V}(\hat{T}) = 1.352082$ ซึ่งจะเป็นค่าที่ควบคุมผลของการวิเคราะห์ว่าเมื่อทำการสำรวจจริงแล้วความแปรปรวนของการประมาณค่าเฉลี่ยและยอดรวมจะต้องเท่ากับ 1.3520822 หรือคลาดเคลื่อนไปจากค่านี้ไม่มากนัก ถ้าคลาดเคลื่อนไปมากแสดงว่าความบกพร่องมิได้เกิดขึ้นจากการกำหนดขนาดตัวอย่าง แต่อาจเกิดขึ้น ณ ขั้นตอนใดขั้นตอนหนึ่ง เช่น อาจเกิดขึ้นในขั้นการปฏิบัติงานสนาม ขั้นจัดเตรียมข้อมูล ขั้นวิเคราะห์ข้อมูล หรืออื่น ๆ ดังนั้น D^2 จึงช่วยทำหน้าที่เป็นผู้ควบคุมงานไปในตัว เพราะถ้าจากผลการวิเคราะห์ข้อมูลพบว่า $\hat{V}(\bar{x})$ หรือ $\hat{V}(\hat{T})$ คลาดเคลื่อนไปจากค่า D^2 นักวิจัยจะได้ทราบว่าได้เกิดความบกพร่องขึ้นแล้วและสามารถแก้ไขได้ทันท่วงที นอกจากนี้ยังใช้เป็นเครื่องมือสำหรับเช็คว่าเราคำนวณหาขนาดตัวอย่างผิดพลาดหรือไม่อีกด้วย

อนึ่ง ความจริงประการดังกล่าวนี้จะเป็นจริงเสมอไม่ว่าจะใช้แผนจัดสรรตัวอย่างแบบใด ตามตัวอย่าง 3.4 จะพบว่า $V(\bar{x}_{eq}) \cong V(\bar{x}_{prop}) \cong V(\bar{x}_{opt}) \cong V(\bar{x}_{ney}) \cong 1.3520822$ เสมอ เช่นกันกับกรณี $V(\hat{T})$

ตัวอย่างเช่น ในกรณีของ $V(\bar{x}_{eq})$ จะพบว่า

$$\hat{V}(\bar{x}_{eq}) = \frac{L}{N^2} \cdot \frac{1}{n} \sum_h^L N_h^2 s_h^2 - \frac{1}{N^2} \sum_h^L N_h s_h^2$$

$$= \frac{1}{(1000)^2} \{3(250{,}000{,}000/355) - 760{,}000\}$$

$$= 1.352676$$

จะเห็นได้ว่า $\hat{V}(\bar{x}_{eq})$ มีค่าใกล้เคียงกับ D^2 ซึ่งเท่ากับ 1.3520822 มาก ความคลาดเคลื่อนอาจเกิดขึ้นจากการปัดเศษซึ่งก็เป็นสิ่งที่ปรากฏขึ้นได้เสมอ $\hat{V}(\bar{x}_{eq}) = 1.352676$ ช่วยชี้ให้เห็นว่าเราคำนวณหาขนาดตัวอย่างไว้ถูกต้องแล้ว

3.4 การประมาณค่าอัตราส่วน R

3.4.1 ข้อสังเกตที่สำคัญ

การประมาณค่า R นอกจากใช้ $\hat{R} = \bar{x}/\bar{y}$ หรือ $\sum_i^n x / \sum_i^n y$ แล้ว เรายังสามารถใช้ $\hat{R} = \hat{T}_x/\hat{R}_{T_y}$ ได้อีกด้วย เพราะพารามิเตอร์ $R = \sum_i^N x_i / \sum_i^N y_i$ หรือ $= \bar{X}/\bar{Y}$ นั้น $\hat{T}_x = N\bar{x}$ เป็นค่าประมาณของ $T_x = \sum_i^N x_i$ และ $\hat{T}_y = N\bar{y}$ เป็นค่าประมาณของ $T_y = \sum_i^N y_i$

ดังนั้นพารามิเตอร์ R จึงสามารถประมาณค่าได้ด้วยตัวประมาณค่าต่อไปนี้

$$\hat{R} = \bar{x}/\bar{y} \text{ หรือ } \hat{R} = \sum_i^n x_i / \sum_i^n y_i \text{ หรือ } \hat{R} = \hat{T}_x/\hat{T}_y = N\bar{x}/N\bar{y}$$

เหตุที่ต้องนำเรื่องนี้มากล่าวนำไว้ก็เพราะมีเหตุผลประการสำคัญคือ ในการศึกษาเรื่องแผนสำรวจแบบแบ่งชั้นภูมินั้นเรานิยมประมาณค่ายอดรวมของแต่ละชั้นภูมิก่อน คือประมาณ T_h ด้วย $\hat{T}_h = N_h\bar{x}_h$ จากนั้นจึงนำ $\hat{T}_h$ เหล่านี้มารวมกันตลอดในทุกชั้นภูมิ ค่าที่ได้ก็คือค่าประมาณของยอดรวมประชากร T และถ้าหารตลอดด้วยขนาดประชากร N จะได้ค่าประมาณของค่าเฉลี่ยประชากร $\bar{X}$ คือ

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

ในกรณีของการประมาณค่าอัตราส่วนประชากรตามแผนสำรวจแบบแบ่งชั้นภูมิก็เช่นกัน เราถือว่าค่าประมาณของ R คือ

$$\hat{R} = \frac{\hat{T}_x}{\hat{T}_y} = \sum_{h}^{L} N_h \bar{x}_h \Big/ \sum_{h}^{L} N_h \bar{y}_h$$

แต่ในการประมาณค่ายอดรวมประชากรซึ่งใช้สูตร $N\bar{x}$ หรือ $N\bar{y}$ สำหรับแผนสำรวจแบบ SRS และใช้สูตร $\sum_{h}^{L} N_h \bar{x}_h$ หรือ $\sum_{h}^{L} N_h \bar{y}_h$ สำหรับแผนสำรวจแบบแบ่งชั้นภูมินั้น มีข้อที่น่าสังเกตประการหนึ่งที่ยังมิได้กล่าวถึงคือ การประมาณค่าของยอดรวมนั้น ถ้านำเอาวิธีการของอัตราส่วนมาใช้จะมีประสิทธิภาพสูงกว่าการใช้ค่าเฉลี่ย ขอให้ดูตัวอย่างต่อไปนี้เพื่อความเข้าใจ

สมมุติว่าในท้องที่แห่งหนึ่งประกอบไปด้วยชาวนา 300 ครอบครัว เราต้องการประมาณดูว่าในท้องที่แห่งนี้ผลิตข้าวได้รวมทั้งสิ้นกี่ถัง โดยทำการสุ่มครอบครัวชาวนามาเป็นกลุ่มตัวอย่าง 20 ครอบครัว พบว่า $\sum_{i}^{20} x_i = 600$ ถัง ดังนี้ก็แสดงว่าโดยถัวเฉลี่ยแล้วชาวนาครอบครัวหนึ่งผลิตข้าวได้ประมาณ 600/20 = 30 ถัง จึงคาดว่าท้องที่แห่งนี้ผลิตข้าวได้รวมทั้งสิ้นประมาณ 300 × 30 = 9,000 ถัง

คราวนี้ถ้าลองย้อนมาพิจารณาครอบครัวชาวนาดูให้ถี่ถ้วนจะพบว่า ครอบครัวเหล่านั้นมีที่นาไม่เท่ากัน บางครอบครัวมี 20 ไร่ บางครอบครัวมี 200 ไร่ บางครอบครัวมี 40 ไร่ บางครอบครัวมี 80 ไร่ แตกต่างกันไปมิได้เท่าเทียมกัน ด้วยเหตุนี้ค่าเฉลี่ย $\bar{x} = 30$ ถัง/ครอบครัวจึงเป็นค่าที่ไม่น่าจะถูกต้อง อาจต่ำกว่าความเป็นจริง (Underestimate) หรือสูงกว่าความเป็นจริง (Overestimate) ซึ่งมีผลให้ค่าประมาณยอดรวม $\hat{T} = 9{,}000$ ถัง ต่ำกว่าความเป็นจริงหรือสูงกว่าความเป็นจริงได้

ด้วยเหตุนี้ ถ้าเราลองหันมาพิจารณาอัตราส่วนการผลิตข้าวต่อครอบครัวดูจะพบความจริงว่า โดยปกติแล้วครอบครัวที่มีที่นามากกว่าจะมีผลผลิตข้าวสูงกว่า ถ้าวัดผลผลิต

ออกมาในรูปผลผลิตต่อไร่ โดยให้ x_i เป็นตัวแปรแทนจำนวนผลผลิตข้าวของครอบครัวที่ i ให้ y_i เป็นตัวแปรแทนจำนวนที่นาของครอบครัวที่ i เราจะพบว่า $\hat{R}$ $\sum_i^n x_i / \sum_i^n y_i$ จะเป็นอัตราส่วนของผลผลิตต่อไร่ซึ่งค่านี้จะคงที่หรือมั่นคงกว่า $\bar{x}$ เพราะไม่ผันผวนไปตามจำนวนพื้นที่นาของแต่ละครอบครัว สมมุติว่าจากกลุ่มตัวอย่างครอบครัวชาวนา 20 ครอบครัว พบว่ามีผลผลิตรวมกันได้เท่ากับ 600 ถัง มีที่นารวมกัน ($\sum_i^{20} y_i$) เท่ากับ 200 ไร่ ดังนี้

$$\hat{R} = \sum_i^n x_i / \sum_i^n y_i = 600/200 = 3 \text{ ถังต่อไร่}$$

สมมุติทราบว่าในท้องที่นี้มีที่นารวมกันเป็นจำนวนทั้งสิ้น ($T_y = \sum_i^N y_i$) เท่ากับ 4,000 ไร่ เราสามารถกะประมาณยอดรวมผลผลิตข้าวในท้องที่ดังกล่าวได้ดังนี้

$$\hat{T}_x = 3 \times 4{,}000 = 12{,}000 \text{ ถัง} = \hat{R}T_y$$

นักศึกษาจะสังเกตเห็นได้ว่าการกะประมาณยอดรวมโดยอาศัยความรู้เรื่องอัตราส่วนเข้าช่วยนั้นบางครั้งจะให้ค่าประมาณที่ถูกต้องและน่าเชื่อถือกว่า เพราะเป็นวิธีที่ช่องโหว่น้อยกว่า

ดังนั้น จากการศึกษาในบทที่ 2 และข้อสังเกตข้างต้น เราจึงสามารถประมาณยอดรวมได้ถึง 3 วิธี แต่ละวิธีย่อมมีจุดอ่อนหรือจุดเด่นแตกต่างกัน เมื่อพิจารณาเห็นว่าวิธีใดน่าจะมีจุดบกพร่องอยู่ก็ให้เปลี่ยนไปใช้วิธีอื่นแทน ดังนี้

1. การประมาณค่ายอดรวมโดยอาศัยค่าเฉลี่ย คือ $\hat{T} = N\bar{x}$

วิธีนี้จะไม่ถูกต้องถ้ากำหนดหน่วยสำรวจไม่เหมาะสมหรือขาดความพินิจพิเคราะห์ในการใช้หน่วยที่เกี่ยวข้องดังตัวอย่างข้างต้น เราสามารถแก้ไขได้โดยการวางแผนก่อนสำรวจและออกแบบสำรวจให้รัดกุม เพราะในงานสำรวจโดยทั่วไปเรานิยมใช้ครัวเรือนเป็นหน่วยสำรวจ สำหรับกรณีของตัวอย่างนี้ เราควรเพิ่มข้อถามเกี่ยวกับพื้นที่นาเข้าไปในแบบสำรวจด้วย ก็จะแก้ปัญหาไปได้ ไม่ใช่จู่ ๆ ก็นำจำนวนครัวเรือนไปหารปริมาณผลผลิตรวม ซึ่งทำให้ค่าประมาณของ $\bar{X}$ คลาดเคลื่อนไปเนื่องจากมีความผันแปรในด้านพื้นที่นาระหว่างครัวเรือน

2. ประมาณโดยอาศัยอัตราส่วนคือ $\hat{T}_x = \hat{R}T_y$

วิธีนี้มีข้อบกพร่องที่เด่นชัดคือ การใช้ยอดรวมของประชากรจริง ๆ T_y มาร่วมวิเคราะห์หา $\hat{T}_x$ ทั้งนี้เพราะโดยปกติ $T_y = \sum_{i}^{N} y_i$ เป็นพารามิเตอร์ที่โดยปกติไม่อาจทราบค่าได้เว้นแต่จะทราบจากงานสำมะโนหรืออาศัยเทคนิคอื่น ๆ เช่น แผนที่ทางอากาศหรือดาวเทียม แต่เราก็อาจใช้ค่าประมาณของ T_y คือ $\hat{T}_y$ แทนกันได้คือใช้สูตร $\hat{T}_x = \hat{R}\hat{T}_y$ และขอให้สังเกตว่า ถ้านำ N หารสมการ $\hat{T}_x = \hat{R}T_y$ ตลอดจะพบว่า $\hat{\bar{X}} = \hat{R}\bar{Y}$ แปลว่าเรายังสามารถประมาณค่าเฉลี่ยโดยอาศัยอัตราส่วนได้อีกด้วย แต่วิธีนี้ก็ยังคงเป็นปัญหาอยู่ที่ $\bar{Y}$ และจะต้องมีเงื่อนไขหลายประการ (อ่าน Sampling Technique ของ Cochran หน้า 106)

3. ประมาณโดยอาศัยสัดส่วนคือ $\hat{T} = Np$ หรือ $N\hat{P}$

วิธีนี้มีจุดอ่อนเช่นเดียวกันกับข้อ 1 เพราะ p หรือ $\hat{P}$ คือ กรณีเฉพาะของ $\bar{x}$ เมื่อ $x = 0, 1$ ธรรมชาติของ $\bar{X}$ จึงไม่ต่างกัน การแก้ปัญหาให้ยึดถือแนวทางเดียวกัน

ด้วยเหตุผลประการต่าง ๆ เหล่านี้ ในบางครั้งเราอาจจำเป็นต้องประมาณอัตราส่วนของประชากรในกรณีของแผนสำรวจแบบแบ่งชั้นภูมิจากรูปทั่วไปคือ

$$\hat{R} = \hat{T}_x/\hat{T}_y = \sum_{h}^{L} N_h\bar{x}_h \Big/ \sum_{h}^{L} N_h\bar{y}_h$$

ซึ่งเรียกว่า Combined Ratio Estimate มาใช้สูตรต่อไปนี้แทนคือ

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

ซึ่งเรียกว่า Separate ratio estimate สูตรนี้พัฒนาขึ้นมาโดยง่ายจากข้อคิดที่กล่าวไว้ในตอนต้น กล่าวคือ

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

แต่ถือว่าเราทราบค่า T_y นำ T_y หารตลอด ดังนั้น

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

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

เมื่อ $T_y = \sum_h^L \sum_i^{N_h} y_{hi}$, $T_{yh} = \sum_i^{N_h} y_{hi}$; $h = 1,2,\ldots,L$

และ $\hat{R}_h = \sum_i^{n_h} x_{hi} / \sum_i^{n_h} y_{hi}$; $h = 1,2,\ldots,L$ หรือ $\hat{R}_h = \bar{x}_h/\bar{y}_h$; $h = 1,2,\ldots,L$

เราจะได้ศึกษาถึงการประมาณค่า R โดยละเอียดในตอนต่อไปนี้ ส่วนการใช้ประโยชน์นั้นเราสามารถเลือกใช้วิธีใดก็ได้ทั้งนี้สุดแล้วแต่สถานการณ์ทางปฏิบัติจะเอื้ออำนวยให้ใช้วิธีใดเกี่ยวกับเรื่องนี้ ขอให้นักศึกษาพิจารณาเหตุการณ์ในตัวอย่างที่ยกให้ดูแล้วเป็นอุธาหรณ์

3.4.2 การประมาณค่าของอัตราส่วนโดยวิธี Combined Ratio Estimate

Combined Ratio Estimate คือ การประมาณค่าอัตราส่วน R โดยที่ $\hat{R} = \sum_h^L N_h\bar{x}_h / \sum_h^L N_h\bar{y}_h$ คำว่า combined มาจากลักษณะการรวมตัวกันของ $\hat{T}_h$ ก่อนที่จะนำมาเปรียบเทียบกันเป็นอัตราส่วน ซึ่งหลักการรวมตัวก็ยึดหลักเกณฑ์ของงานแผนสำรวจแบบแบ่งชั้นภูมิที่ได้ศึกษามาแล้วในตอนก่อนและเพื่อให้เห็นความแตกต่างระหว่างวิธีประมาณค่าของ R ต่อไปนี้จะใช้ $\hat{R}_c$ หมายถึงค่าประมาณของอัตราส่วน R ที่ได้มาโดยวิธี Combined Estimation และใช้ $\hat{R}_s$ หมายถึงค่าประมาณของ R ที่ได้มาโดยวิธี Separated Estimation

อย่างไรก็ตาม ผู้เขียนขอแนะนำให้ใช้ $\hat{R}_c$ เพราะวิธีนี้มีวิธีการประมาณค่าเศษและส่วนของ $\hat{R}$ ในลักษณะที่เราคุ้นเคยดี แม้แต่สูตร $V(\hat{R})$ ก็มีสูตรโครงสร้างคล้าย $V(\bar{x}_{st})$ ที่เราคุ้นเคยดีแล้วเช่นกัน

สำหรับสัญลักษณ์บางตัวอาจดูแปลกตาไปบ้าง แต่มิได้มีความหมายต่างไปจากที่เคยนิยามไว้แล้วในตอนต้น เพียงแต่เพิ่มค่า x และ y ลงใน subscript เพื่อชี้ให้เห็นว่ากำลังพูดถึงตัวแปร x หรือตัว y เท่านั้น เช่น S^2_{xh} คือความแปรปรวนของตัวแปร x ในชั้นภูมิที่ h S^2_{yh} คือ ความแปรปรวนของตัวแปร y ในชั้นภูมิที่ h เป็นต้น ส่วน subscript h และ i ยังคงมีความหมายเช่นเดิม

ดังนั้น

1. $\hat{R}_c = \hat{T}_x/\hat{T}_y = \sum_h^L N_h\bar{x}_h \Big/ \sum_h^L N_h\bar{y}_h$

และ

2. $$V(\hat{R}_c) \cong \frac{1}{N^2}\cdot\frac{1}{\bar{Y}^2}\sum_h^L N_h^2\,\frac{N_h-n_h}{N_h}\cdot\frac{1}{n_h}\left(S^2_{xh} + R^2S^2_{yh} - 2R\rho_hS_{xh}S_{yh}\right)$$

โดยที่ $$S^2_{xh} = \frac{1}{N_h-1}\sum_i^{N_h}(x_{hi}-\bar{X}_h)^2,\quad S^2_{yh} = \frac{1}{N_h-1}\sum_i^{Nh}(y_{hi}-\bar{Y}_h)^2$$

$$\rho_h = \frac{S_{xyh}}{S_{xh}S_{yh}} = \frac{\sum_i^{N_h}(x_{hi}-\bar{X}_h)(y_{hi}-\bar{Y}_h)}{\sqrt{\sum_i^{N_h}(x_{hi}-\bar{X}_h)^2\sum_i^{N_h}(y_{hi}-\bar{Y}_h)^2}}$$

และตัวประมาณค่าของ $V(\hat{R}_c)$ คือ

3. $$\hat{V}(\hat{R}_c) \cong \frac{1}{N^2}\cdot\frac{1}{\bar{y}^2}\sum_h^L N_h^2\,\frac{N_h-n_h}{N_h}\,\frac{1}{n_h}\left(s^2_{xh} - \hat{R}_c^2s^2_{yh} - 2\hat{R}_cr_hs_{xh}s_{yh}\right)$$

โดยที่ $s_{xh}^2 = \frac{1}{n_h-1}\sum_{i}^{n_h}(x_{hi}-\bar{x}_h)^2,\ s_{yh}^2 = \frac{1}{n_h-1}\sum_{i}^{n_h}(y_{hi}-\bar{y}_h)^2$

$\hat{\varrho}_h = r_h =$ สหสัมพันธ์ระหว่างตัวแปร x และ y ในชั้นภูมิที่ h

$$= \frac{\sum^{n_h}(x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h)}{\sqrt{\sum_{i}^{n_h}(x_{hi}-\bar{x}_h)^2\sum_{i}^{n_h}(y_{hi}-\bar{y}_h)^2}}$$

พิสูจน์ จาก $\hat{R}_c = \hat{T}_{xst}/\hat{T}_{yst} = \sum_{h}^{L} N_h\bar{x}_h \Big/ \sum_{h}^{L} N_h\bar{y}_h = \dfrac{\sum_{h}^{L} N_h\bar{x}_h/N}{\sum_{h}^{L} N_h\bar{y}_h/N}$

$$\hat{R}_c = \bar{x}_{st}/\bar{y}_{st}$$

ดังนั้น
$$\hat{R}_c - R = \bar{x}_{st}/\bar{y}_{st} - R$$
$$= \frac{\bar{x}_{st}-R\bar{y}_{st}}{\bar{y}_{st}}$$
$$\cong (\bar{x}_{st}-R\bar{y}_{st})/\bar{Y} \qquad \because\ E(\bar{y}_{st}) = \bar{Y}$$

ดังนั้น $V(\hat{R}_c) = E(\hat{R}_c-R)^2 \cong \frac{1}{\bar{Y}^2}E(\bar{x}_{st}-R\bar{y}_{st})^2$

ให้ $u_{hi} = x_{hi} - Ry_{hi}$ แสดงว่า $\bar{u}_h = \frac{1}{n_h}\sum_{i}^{n_h} u_{hi} = \bar{x}_h - R\bar{y}_h$

$$\bar{u}_{st} = \frac{1}{N}\sum_{h}^{L} N_h\bar{u}_h = \frac{1}{N}\sum_{h}^{L} N_h(\bar{x}_h-R\bar{y}_h)$$
$$= \frac{1}{N}\sum_{h}^{L} N_h\bar{x}_h - \frac{R}{N}\sum_{h}^{L} N_h\bar{y}_h$$
$$= \bar{x}_{st} - R\bar{y}_{st}$$

นั่นคือ $V(\hat{R}_c) \cong \frac{1}{\bar{Y}^2}E\left(\frac{1}{N}\sum_{h}^{L} N_h\bar{u}_h\right)^2$

แต่ $\overline{U} = \frac{1}{N}\sum_{h}^{L}\sum_{i}^{N_h} u_{hi} = \frac{1}{N}\sum_{h}^{L}\sum_{i}^{N_h} x_{hi} - \frac{R}{N}\sum_{h}^{L}\sum_{i}^{N_h} y_{hi} = \overline{X} - R\overline{Y} = \overline{X} - \frac{\overline{X}}{\overline{Y}}\cdot\overline{Y} = 0$

$$\therefore \quad V(\hat{R}_c) \cong \frac{1}{\overline{Y}^2} E(\bar{u}_{st} - \overline{U})^2$$

$$\cong \frac{1}{\overline{Y}^2} V(\bar{u}_{st})$$

$$=> V(\hat{R}_c) \cong \frac{1}{\overline{Y}^2}\cdot\frac{1}{N^2}\sum_{h}^{L}\frac{N_h - n_h}{N_h}\,\frac{N_h^2 S_{uh}^2}{n_h}$$

โดยที่ $S_{uh}^2 = \frac{1}{N_h - 1}\sum_{i}^{N_h}(u_{hi} - \overline{U}_h)^2$

แต่ $\overline{U}_h = \frac{1}{N_h}\sum_{i}^{N_h} u_{hi} = \frac{1}{N_h}\sum_{i}^{N_h}(x_{hi} - Ry_{hi}) = \overline{X}_h - R\overline{Y}_h$

ดังนั้น $S_{uh}^2 = \frac{1}{N_h - 1}\sum_{i}^{N_h}\left\{(x_{hi} - Ry_{hi}) - (\overline{X}_h - R\overline{Y}_h)\right\}^2$

$$= \frac{1}{N_h - 1}\sum_{i}^{N_h}\left\{(x_{hi} - \overline{X}_h) - R(y_{hi} - \overline{Y})\right\}^2$$

$$= \frac{1}{N_h - 1}\left\{\sum_{i}^{N_h}(x_{hi} - \overline{X}_h)^2 + R^2\sum_{i}^{N_h}(y_{hi} - \overline{Y}_h)^2 - 2R\sum_{i}^{N_h}(x_{hi} - \overline{X}_h)(y_{hi} - \overline{Y}_h)\right\}$$

$$= \frac{1}{N_h - 1}\left\{(N_h - 1)S_{xh}^2 + (N_h - 1)R^2 S_{yh}^2 - 2R(N_h - 1)\varrho_h S_{xh} S_{yh}\right\}$$

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

นั่นคือ $V(\hat{R}_c) \cong \frac{1}{\overline{Y}^2}\cdot\frac{1}{N^2}\sum_{h}^{L}\frac{N_h - n_h}{N_h}\cdot N_h^2\cdot\frac{1}{n_h}(S_{xh}^2 + R^2 S_{yh}^2 - 2R\varrho_h S_{xh} S_{yh})$

สำหรับการประมาณค่ายอดรวม T_x เราจะใช้สูตรดังนี้คือ

4. $\hat{T}_x = \hat{R}_c T_y \;;\; T_y = N\overline{Y}$

และเราสามารถพิสูจน์โดยง่ายว่า

$$5.\ V(\hat{T}_x) \cong T_y^2 V(\hat{R}_c)$$

$$= \sum_h^L \frac{N_h - n_h}{N_h} N_h^2 \frac{1}{n_h} (S_{xh}^2 + R^2 S_{yh}^2 - 2R\varrho_h S_{xh} S_{yh})$$

$$6.\ \hat{V}(\hat{T}_x) \cong T_y^2 \hat{V}(\hat{R}_c) = \sum_h^L \frac{N_h - n_h}{N_h} \frac{N_h^2}{n_h} (s_{xh}^2 + \hat{R}_c s_{yh}^2 - 2\hat{R}_c \hat{\varrho}_h s_{xh} s_{yh})$$

ตัวอย่าง 3.5 ต้องการกะประมาณยอดขายของห้างสรรพสินค้าในช่วงเดือนเมษายนถึงพฤษภาคม ว่าเปลี่ยนแปลงไปอย่างไร การสำรวจครั้งนี้จำแนกห้างสรรพสินค้าออกเป็น 3 ชั้นภูมิตามขนาดของห้างคือ มีพื้นที่ไม่เกิน 500 ต.ร.ฟ. ซึ่งมีอยู่ทั้งสิ้น 1000 แห่ง จัดว่าเป็นห้างขนาดเล็ก ห้างที่มีพื้นที่ 500-999 ต.ร.ฟ. ซึ่งมีอยู่ทั้งสิ้น 500 แห่งจัดว่าเป็นขนาดกลาง และห้างที่มีพื้นที่ตั้งแต่ 1000 ต.ร.ฟ. ซึ่งมีอยู่ทั้งสิ้น 200 แห่ง จัดว่าเป็นห้างขนาดใหญ่

สุ่มตัวอย่างห้างสรรพสินค้ามาชั้นภูมิละ 50 แห่งเท่า ๆ กันโดยดำเนินการสำรวจในเดือนพฤษภาคม ข้อมูลที่ถามส่วนหนึ่งคือ x = ยอดขายในเดือนพฤษภาคม และ y = ยอดขายในเดือนเมษายน ดังตารางต่อไปนี้

ชั้นภูมิ (ห้างสรรพสินค้า)	N_h	n_h	$\bar{x}_h$ (พฤษภาคม)	$\bar{y}_h$ (เมษายน)
ขนาดเล็ก	1000	50	10	10
ขนาดกลาง	500	50	33	30
ขนาดใหญ่	200	50	60	50
รวม	1700	150		

และปรากฏข้อมูลอื่น ๆ ดังนี้

ชั้นภูมิ	s^2_{xh}	s^2_{yh}	$\hat{\varrho}_h$
ขนาดเล็ก	1	1	1
ขนาดกลาง	2	2	1
ขนาดใหญ่	3	3	1

จงคำนวณหา $\hat{R}_c$ และช่วงเชื่อมั่น 95% ของค่าจริง R

วิธีทำ ก. $\hat{R}_c = \sum_{h}^{3} N_h\bar{x}_h / \sum_{h}^{3} N_h\bar{y}_h$

$$= (1000\times 10+500\times 33+200\times 60)/(1000\times 10+500\times 3+200\times 50)$$

$$= 38{,}500/35{,}000$$

$$= 1.1$$

หมายความว่า ยอดขายในเดือนพฤษภาคมลดลงกว่าในเดือนเมษายนประมาณ 10%

หมายเหตุ ถ้า $\hat{R}_c = 1$ แสดงว่ายอดขายคงที่ไม่เปลี่ยนแปลง ถ้า $\hat{R}_c = 1.1$ แสดงว่ายอดขายของเดือนเมษายนมากกว่ายอดขายในเดือนพฤษภาคมอยู่ 0.1 ถึง 10%

$$\hat{V}(\hat{R}_c) \cong \frac{1}{N^2\bar{y}^2_{st}} \sum_{h}^{L} \frac{N_h-n_h}{N_h} \frac{N_h^2}{n_h} (s^2_{xh}+\hat{R}_c^2 s^2_{yh}-2\hat{R}_c\hat{\varrho}_h s_{xh}s_{yh})$$

ชั้นภูมิ	s_{xh}^2 (1)	$\hat{R}_c$ (2)	s_{yh}^2 (3)	$\hat{R}_c^2 s_{yh}^2$ (4)	$\hat{\varrho}_h$ (5)	$2\hat{R}_c\hat{\varrho}_h s_{xh} s_{yh}$ (6)	(1) + (4) − (6)
ขนาดเล็ก	1	1.1	1	1.21	1.0	2.2	0.01
ขนาดกลาง	2	1.1	2	2.42	1.0	4.4	0.02
ขนาดใหญ่	3	1.1	3	3.63	1.0	6.6	0.03

และ $\because \ (N\bar{y}_{st}) = (N \cdot \frac{1}{N} \sum_h^L N_h\bar{y}_h) = \sum_h^L N_h\bar{y}_h = \hat{T}_y = 35{,}000$

ดังนั้น

$$\hat{V}(\hat{R}_c) \cong (1/35{,}000)^2\{ \frac{1000-50}{1000} \frac{(1000)^2}{50} (0.01) + \frac{500-50}{500} \cdot \frac{(500)^2}{50} (.02) + \frac{200-50}{200} \frac{(200)^2}{50} (.03)\}$$

$$= (1/35{,}000)^2(190+90+18)$$

$$= 298/(35{,}000)^2$$

$$= .0000002$$

$$s(\hat{R}_c) = .00045$$

ช่วงเชื่อมั่น 95% ที่คาดว่าค่าจริงของ R จะปรากฏอยู่คือ

$$\{(1.1-(1.96)(.00045),\ 1.1+(1.96)(.00045)\} = (1.0991,\ 1.1009)$$