บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 การสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling)

เป็นวิธีการสุ่มตัวอย่าง โดยกำหนดปัจจัยที่ใช้เป็นเกณฑ์ในการแบ่งชั้นภูมิ จำนวนชั้นภูมิ ขนาดตัวอย่างทั้งหมด แล้วแบ่งประชากรออกเป็นชั้นภูมิต่างๆ โดยไม่ให้มีหน่วยซ้ำกัน คือ หน่วยทุกหน่วยในประชากร จะต้องอยู่ในชั้นภูมิใดชั้นภูมิหนึ่งเท่านั้น แล้วสุ่มตัวอย่างจากแต่ละชั้นภูมิอย่างเป็นอิสระต่อกัน

ประโยชน์ของการสุ่มตัวอย่างแบบชั้นภูมิ ไม่ได้มาจากเพียงความแม่นยำของตัวประมาณ แต่มาจากลักษณะที่ถือว่า แต่ละชั้นภูมิเป็นอิสระต่อกัน ทำให้การสำรวจได้สะดวก มีประสิทธิภาพยิ่งขึ้น

ตัวแปรที่ใช้เป็นเกณฑ์กำหนดชั้นภูมิที่ดีที่สุด คือ ตัวแปรที่กำลังศึกษาอยู่ ในทางปฏิบัติส่วนใหญ่จะใช้ตัวแปรที่มีความสัมพันธ์สูงกับตัวแปรที่สนใจ เป็นตัวกำหนดแบ่งชั้นภูมิ

การสุ่มตัวอย่างแบบชั้นภูมิให้มีความแม่นยำในการประมาณสูงนั้น จะต้องพยายามคัดหน่วยที่มีลักษณะคล้ายกันไว้ในชั้นภูมิเดียวกัน และให้หน่วยที่อยู่ต่างชั้นภูมิมีลักษณะต่างกัน เช่น ประชากร ประกอบด้วย บริษัทต่างๆที่มียอดขายต่างกันมาก เราอาจรวบรวมบริษัทที่มียอดขายใกล้เคียงกันไว้ในชั้นภูมิเดียวกัน ค่าความแปรปรวนของค่าเฉลี่ยตัวอย่างในแต่ละชั้นภูมิจะต่ำ แต่บางครั้งอาจไม่สามารถทำได้ เช่น ประชากรประกอบด้วย ครัวเรือนในกรุงเทพฯ ถ้าแบ่งชั้นภูมิตามลักษณะเพศของหัวหน้าครอบครัว และ อาชีพของหัวหน้าครอบครัว โดยที่ตัวแปรที่สนใจ คือ รายได้ครัวเรือน ซึ่งเป็นการยากที่รายได้ครัวเรือนมีค่าใกล้เคียงกันในชั้นภูมิเดียวกัน

2.11 สัญลักษณ์ที่ใช้

สมมุติให้ประชากรถูกแบ่งออกเป็น L ชั้นภูมิ ให้ h แทนชั้นภูมิ

N_h = จำนวนประชากรในชั้นภูมิที่ h

n_h = จำนวนตัวอย่างในชั้นภูมิที่ h

Y_{hi} = ค่าสังเกตประชากรในชั้นภูมิที่ h

y_{hi} = ค่าสังเกตตัวอย่างในชั้นภูมิที่ h

$\bar{Y}_h$ = ค่าเฉลี่ยประชากรในชั้นภูมิที่ h

= $(1 / N_h) \Sigma Y_{hi}$

$\bar{y}_h$ = ค่าเฉลี่ยตัวอย่างในชั้นภูมิที่ h

= $(1 / n_h) \Sigma y_{hi}$

S_h^2 = ค่าความแปรปรวนของประชากรในชั้นภูมิที่ h

= $\Sigma (Y_{hi} - \bar{Y}_h)^2 / (N_{hi} - 1)$

s_h^2 = ค่าความแปรปรวนของตัวอย่างในชั้นภูมิที่ h

= $\Sigma (y_{hi} - \bar{y}_h)^2 / (n_{hi} - 1)$

f_h = สัดส่วนในการเลือกตัวอย่าง(Sampling fraction) ของชั้นภูมิ h

= n_h / N_h

W_h = สัดส่วนของจำนวนหน่วยในชั้นภูมิ h

= N_h / N

$\bar{y}_s$ = ตัวประมาณค่าเฉลี่ยประชากร

= $\Sigma (N_h/N)\bar{y}_h = \Sigma W_h \bar{y}_h$

$\hat{Y}_s$ = ตัวประมาณค่ารวมประชากร = $N\bar{y}_s$

2.12 คุณสมบัติของ $\bar{y}_s$

ขึ้นอยู่กับลักษณะของค่า $\bar{y}_h$

1. ถ้า $\bar{y}_h$ เป็นตัวประมาณที่ไม่เอนเอียงของ $\bar{Y}_h$ แล้ว $\bar{y}_s$ จะเป็นตัวประมาณ ที่ไม่เอนเอียงของ $\bar{Y}$

$$E(\bar{y}_s) = E(\Sigma (N_h/N)\bar{y}_h)$$

$$= \Sigma (N_h/N) E(\bar{y}_h)$$

$$= \Sigma (N_h/N) \bar{Y}_h$$

$$= \Sigma Y_h/N = \bar{Y}$$

2. $V(\bar{y}_s) = V(W_h \bar{y}_h)$

$= \Sigma (W^2_h) V(\bar{y}_h) + 2 \Sigma\Sigma W_h W_j Cov(\bar{y}_h \bar{y}_j)$

y_h และ y_j เป็นอิสระต่อกัน ดังนั้น $Cov(\bar{y}_h \bar{y}_j) = 0$

$V(\bar{y}_s) = \Sigma (W_h^2) V(\bar{y}_h)$

โดยที่ $V(\bar{y}_h) = (1-f_h) S_h^2/n_h$

2.13 วิธีการสุ่มตัวอย่าง จากแต่ละชั้นภูมิ มีด้วยกันหลายวิธี เช่น

1. สุ่มแต่ละชั้นภูมิ เท่ากันหมด วิธีนี้ง่ายต่อการปฏิบัติ แต่ถ้าจำนวนประชากรในชั้นภูมิต่างๆมีค่าต่างกันมาก คือ ชั้นภูมิหนึ่ง มีจำนวนประชากรมาก อีกชั้นภูมิมีจำนวนประชากรน้อย วิธีนี้จะให้ความคลาดเคลื่อนมาก ไม่สามารถ ประมาณค่าประชากรได้ดี ในชั้นภูมิ กลุ่มใหญ่ได้

2. Proportional allocation โดยการกำหนดขนาดตัวอย่างของชั้นภูมิเป็นสัดส่วนกับ จำนวนหน่วยในชั้นภูมิ

นั่นคือ $n_h = nN_h/N = nW_h$

$n_h/n = N_h/N = W_h \qquad (f_h = n_h/N_h = n/N = f)$

$\bar{y}_s = \Sigma (N_h/N)\bar{y}_h$

$= \Sigma\Sigma y_{hi}/n$

นั่นคือ $\bar{y}_s$ คือเฉลี่ย $(\bar{y})$ ตัวอย่าง โดยไม่คำนึงถึงชั้นภูมิ

$V(\bar{y}_s) = \Sigma (W_h^2) V(\bar{y}_h)$

$= \Sigma W_h^2 (1-f_h) S_h^2/n_h$

$= (1-f_h)\Sigma W_h^2 S_h^2$

วิธีนี้ไม่ได้นำเอาความแตกต่างในระดับของความกระจายของข้อมูล และค่าใช้จ่ายต่อหน่วยในแต่ละชั้นภูมิมาพิจารณา ซึ่งในกรณีที่ชั้นภูมิต่างๆมีความกระจายในชั้นภูมิต่างกัน หรือมีค่าใช้จ่ายต่อหน่วยไม่เท่ากัน คุณภาพของตัวประมาณอาจไม่ดีเท่าที่ควร

3. Optimum Allocation โดยการกำหนดขนาดตัวอย่างของชั้นภูมิ โดยเน้นที่จะทำให้ความแปรปรวนต่ำที่สุด เมื่อทรัพยากรคงที่ หรือใช้ทรัพยากรน้อยที่สุด เมื่อกำหนดค่าความแปรปรวนไว้คงที่

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.2 การเลือกตัวอย่างแบบโควต้า (QUOTA SAMPLING)

เป็นวิธีการเลือกตัวอย่างที่ไม่คำนึงถึงความน่าจะเป็นในการคัดเลือกตัวอย่าง หน่วยตัวอย่างจะถูกคัดเลือกเป็นตัวอย่างอย่างอิสระ ผู้วิจัยจะคัดเลือกข้อมูลที่พบเป็นตัวอย่างทุกครั้งเมื่อข้อมูลมีลักษณะตรงตามที่สนใจจนกว่าจะได้ข้อมูลครบตามจำนวนที่ต้องการ

วิธีการเลือกตัวอย่างแบบโควต้า ใช้กันแพร่หลาย ในราวปี ค.ศ. 1930-1940 (จาก VARIOUS SAMPLING PROCEDURE :4) เพื่อใช้ทำนายผลการเลือกตั้ง และการวิจัยทางการตลาด วิธีนี้จะไม่สามารถวัดคุณภาพของตัวประมาณ (PRECISION OF THE ESTIMATE)ได้ เนื่องจากยังขึ้นอยู่กับความคิดเห็น และการตัดสินใจในการเลือกตัวอย่างของผู้สำรวจหรือผู้สัมภาษณ์ ดังนั้นตัวประมาณที่ได้จึงมีความลำเอียง (BIAS) และไม่สามารถวัดระดับของความลำเอียงได้

การเลือกตัวอย่างแบบโควต้า ให้มีประสิทธิภาพดีนั้น ผู้สำรวจต้องมีความรู้เกี่ยวกับลักษณะประชากรที่สนใจศึกษา เพื่อจะได้กำหนดจำนวนตัวอย่าง และปัจจัย เพื่อใช้เป็นเกณฑ์คัดเลือกตัวอย่างเข้ากลุ่ม ดังนั้นวิธีการเลือกตัวอย่างแบบโควต้า จึงมีลักษณะคล้ายกับวิธีการสุ่มตัวอย่างแบบภูมิ ต่างกันที่วิธีการเลือกตัวอย่างแบบโควต้านั้นพิจารณาถึงความสะดวก ความง่ายของผู้สำรวจเป็นหลัก และ ผู้สำรวจไม่ควรนำเอาเรื่องส่วนตัวเข้ามาเกี่ยวข้อง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย