การสุ่มตัวอย่างแบบแบ่งชั้น

ความหมาย

การสุ่มตัวอย่างแบบแบ่งชั้น(Stratified Random Sampling) หมายถึงการสุ่มที่แบ่งสมาชิกของประชากรออกเป็นชั้น ๆ ซึ่งมีลักษณะเป็นเอกพันธ์ โดยสมาชิกหนึ่งเมื่ออยู่ในชั้นใดจะไปปรากฏในชั้นอื่นๆ อีกไม่ได้ หลังจากนั้น ก็ทำการเลือกสมาชิกภายในแต่ละชั้นด้วยวิธีการสุ่มแบบง่าย หรือสุ่มแบบมีระบบภายในแต่ละชั้น

วิธีการสุ่มตัวอย่างแบบแบ่งชั้น

วิธีการสุ่มตัวอย่างแบบแบ่งชั้นเป็นวิธีการสุ่มตัวอย่างจากประชากรซึ่งมี N หน่วย โดยเริ่มต้นด้วยการแบ่งประชากรออกเป็น L ชั้น ในแต่ละชั้นมีสมาชิก $N_1, N_2, \ldots, N_L$ หน่วยตามลำดับ

นั่นคือ $N_1 + N_2 + \ldots + N_L = N$

ในการสุ่มแบบแบ่งชั้นนี้ ควรจะต้องทราบจำนวนสมาชิก N_i $(i=1,2,\ldots,L)$ ในแต่ละชั้น การสุ่มในแต่ละชั้น ทำการสุ่มอย่างอิสระ ซึ่งอาจใช้วิธีเดียวกันหรือแตกต่างกันก็ได้ โดยให้ขนาดตัวอย่างในแต่ละชั้นเป็น $n_1, n_2, \ldots, n_L$ ตามลำดับ ตัวอย่างเช่น ในการสุ่มจำนวนครู เราอาจแบ่งเป็นโรงเรียนขนาดใหญ่ กลาง และเล็ก เพื่อให้สอดคล้องกับขนาดของโรงเรียน ซึ่งถือว่ามีอิทธิพลต่อการสอนของครู ฯลฯ เป็นต้น

หลักการในการแบ่งประชากรออกเป็นชั้นๆ

ก. ประชากรที่จะสุ่มถูกแบ่งออกเป็นชั้น ๆ ตามคุณลักษณะที่เราทราบอย่างชัดเจน และจะกำหนดว่าให้แต่ละชั้นเป็นประชากรของตนเอง

ข. อาจแบ่งชั้นให้สอดคล้องกับระบบบริหารของหน่วยงาน

ค. การแบ่งประชากรออกเป็นชั้นๆ (subpopulation) สมาชิกภายในแต่ละชั้น มีลักษณะเป็นเอกพันธ์ (homogeneous) และในระหว่างแต่ละชั้นมีลักษณะเป็นวิวิธพันธ์ (heterogeneous)

ง. การสุ่มแบบแบ่งชั้น ช่วยให้การประมาณคุณลักษณะต่างๆ ของประชากรทั้งหมด ทำได้ถูกต้องแม่นยำ เนื่องจากความแปรปรวนระหว่างสมาชิกภายในชั้น จะมีน้อยกว่าความแปรปรวนระหว่างชั้น เมื่อนำค่าประมาณของประชากรย่อยแต่ละชั้นนี้มารวมกันก็ช่วยให้การประมาณค่าประชากรทั้งหมดทำได้ถูกต้องขึ้น

การประมาณค่าเฉลี่ยและผลรวมทั้งหมดของประชากร

ให้ $\bar{y}_i$ เป็นค่าเฉลี่ยของตัวอย่างที่สุ่มจากชั้นที่ i

μ_i เป็นค่าเฉลี่ยของประชากรในชั้นที่ i

T_i เป็นผลรวมของประชากรในชั้นที่ i

T เป็นผลรวมของประชากรทั้งหมด

$\therefore T = T_1 + T_2 + \ldots + T_L$

$\bar{y}_i$ เป็นค่าประมาณที่ไม่เอนเอียง ของ μ_i

และ $N_i\bar{y}_i$ เป็นค่าประมาณของ T_i เมื่อ $T_i = N_i\mu_i$ และ $\mu = \frac{T}{N} = \frac{\Sigma T_i}{N}$

ค่าประมาณเฉลี่ยของประชากร μ คือ

$$\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{L} N_i\bar{y}_i$$

เนื่องจาก $T = N\mu$

ดังนั้น $N\bar{y}_{st}$ เป็นค่าประมาณของ T

$$\therefore N\bar{y}_{st} = N_1\bar{y}_1 + N_2\bar{y}_2 + \ldots + N_L\bar{y}_L$$

$$= \sum_{i=1}^{L} N_i\bar{y}_i$$

อย่างไรก็ดี ควรหาขอบเขตของความเชื่อมั่นของค่าประมาณที่ได้ด้วย เพื่อจะช่วยให้ทราบว่าถ้าเรามีความเชื่อมั่น ๙๕ เปอร์เซ็นต์ ค่าประมาณที่ได้ควรจะอยู่ระหว่างค่าใด

ประโยชน์ของการสุ่มตัวอย่างแบบแบ่งชั้น

ก. ความแปรปรวนของการประมาณค่าเฉลี่ยของประชากรมักจะมีค่าน้อยกว่าการสุ่มตัวอย่างแบบง่าย เพราะ ความแปรปรวนของข้อมูลภายในแต่ละชั้นมีค่าน้อยกว่าความแปรปรวนของประชากรทั้งหมด

ข. การแบ่งประชากรออกเป็นชั้นๆ ทำให้การเก็บข้อมูล การวิเคราะห์ข้อมูล มีค่าใช้จ่ายต่ำกว่าการสุ่มแบบง่าย

ค. การประมาณค่าพารามิเตอร์แต่ละชั้น สามารถกระทำได้อย่างอิสระไม่ต้องขึ้นอยู่กับชั้นอื่น ๆ

ง. การสุ่มแบบแบ่งชั้น ช่วยให้การประมาณคุณลักษณะต่างๆ ของประชากรทั้งหมด ทำได้ถูกต้องแม่นยำมากกว่าการสุ่มวิธีอื่น

อนันต์ ศรีโสภา

บรรณานุกรม

อนันต์ ศรีโสภา **เทคนิคการสุ่มตัวอย่าง** คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ๒๕๒๔

Cochran, W.G. **Sampling Techniques.** 3 rd. ed. New York, John Wiley and Sons, Inc.,1963.

Kish,L. **Survey Sampling.** New York , John Wiley and Sons, Inc., 1965.