

การสุ่มตัวอย่างแบบง่าย

ความหมาย

การสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) หมายถึงการสุ่มตัวอย่างขนาด n จากประชากร N โดยที่ทำให้ตัวอย่างขนาด n แต่ละตัวอย่างมีความน่าจะเป็นที่จะถูกเลือกมาเท่า ๆ กัน

วิธีการสุ่มตัวอย่างแบบง่าย

การที่จะสุ่มตัวอย่างจากประชากรที่เราสนใจ โดยให้แต่ละตัวอย่างขนาด n มีโอกาสที่จะถูกคัดเลือกจากประชากรเท่า ๆ กันนั้น อาจจะคิดว่าการสุ่มที่ไม่ได้เตรียมการ หรือการสุ่มแบบตามบุญตามกรรมก็น่าจะได้ตัวแทนของประชากรได้ แต่บางครั้งการสุ่มแบบนี้ ทำให้เกิดความเอนเอียงได้ โดยทั่วไปการสุ่มแบบง่ายเราใช้ตารางเลขสุ่ม (table of random numbers) ซึ่งมีวิธีดำเนินการดังนี้

สมมุติว่า ประชากรมีขนาด N เท่ากับ ๑๐๐๐ ต้องการสุ่มตัวอย่างขนาด n ซึ่งเท่ากับ ๒๐ และต้องการให้ตัวอย่างขนาด n แต่ละตัวอย่างมีโอกาสที่จะได้รับการคัดเลือกเท่า ๆ กัน ก่อนอื่นจำเป็นจะต้องกำหนดตัวเลขตั้งแต่ ๐๐๐.๐๐๑.๐๐๒., ๙๙๙ ลงในสมาชิกแต่ละคนของประชากร ในกรณีที่ประชากรมีสมาชิก ๑.๐๐๐ คน ก็กำหนดโดย ๐๐๑ แทนสมาชิกคนที่หนึ่ง ๙๙๙ แทนสมาชิกคนที่ ๙๙๙ และ ๐๐๐ แทนสมาชิกคนที่ ๑.๐๐๐ เป็นต้น

จากตารางเลขสุ่มซึ่งตัวเลขต่างๆ ในตารางเลขสุ่มได้มาโดยการสุ่ม เขาจะพิมพ์ ไรท์ท้ายเล่มของหนังสือสถิติ หรือหนังสือเทคนิคการสุ่มตัวอย่าง สำหรับตัวอย่างข้างต้น เราใช้ตัวเลข ๒ ตัวในแถวแรก และอีกหนึ่งตัวในแถวที่สองเป็นตัวเลข ๓ ตัว จากตารางจะเห็นว่าคนที่ตกในตัวอย่างคนแรก คือคนที่ ๑๐๐ คนที่สองคือคนที่ ๓๗๕ คนที่สามคือคนที่ ๐๘๔ เช่นนี้เรื่อยไปจนครบคนที่ ๒๐ ดังปรากฏในตัวอย่างต่อไปนี้

๑๐๐	๖๖๐	๙๘๕	๖๕๔
๓๗๕	๓๑๐	๑๑๘	๘๐๑
๐๘๔	๘๕๒	๘๓๔	๗๔๓
๙๙๐	๖๕๓	๘๘๖	๖๙๙
๑๒๘	๗๓๗	๙๙๕	๐๙๘

อนึ่ง ถ้าหากตารางเลขสุ่มปรากฏว่ามีซ้ำครั้งที่สอง เราก็ข้ามไปนับเลขอื่นถัดไปแทน

การประมาณค่าเฉลี่ย และผลรวมทั้งหมดของประชากร

จากจุดมุ่งหมายของการสุ่มตัวอย่าง เพื่อหาข้อสรุปของประชากรจากข้อมูลต่าง ๆ ที่ปรากฏในตัวอย่างหรือเป็นการประมาณค่าพารามิเตอร์ของประชากรจากค่าสถิติของตัวอย่าง โดยทั่วไปกำหนดให้ \bar{Y} แทนค่าเฉลี่ยของประชากร และ Y แทนค่าผลรวมของประชากร สำหรับกลุ่มตัวอย่างกำหนด \bar{y} แทนค่าเฉลี่ยของกลุ่มตัวอย่าง และ \hat{y} แทนค่าผลรวมของตัวอย่าง ดังนี้

$$\text{ประชากร} \\ \text{ผลรวม} : Y = \sum_{i=1}^N Y_i = Y_1 + Y_2 + \dots + Y_N$$

$$\text{ตัวอย่าง} \\ \hat{y} = \sum_{i=1}^n y_i = y_1 + y_2 + \dots + y_n$$

$$\text{ประชากร} \\ \text{ค่าเฉลี่ย} : \bar{Y} = \left(\sum_{i=1}^N Y_i \right) / N = \frac{Y_1 + Y_2 + \dots + Y_N}{N}$$

ตัวอย่าง

$$\bar{y} = \left(\sum_{i=1}^n y_i \right) / n = \frac{y_1 + y_2 + \dots + y_n}{n}$$

อนึ่ง โปรดสังเกตว่าตัวอย่างจำนวน n หน่วยมิได้ประกอบด้วย n หน่วยแรกของประชากร n หน่วยในที่นี้หมายถึง จำนวนที่ได้จากการสุ่ม ซึ่งอาจจะเป็นหน่วยที่เท่าใดของประชากรก็ได้

เราต้องการจะประมาณค่าพารามิเตอร์ ของประชากรสองตัวคือ \bar{Y} และ Y จากค่าสถิติ \bar{y} และ \hat{y} นั้นคือ

การประมาณค่าเฉลี่ยของประชากร \bar{Y} จากค่าเฉลี่ยของตัวอย่าง \bar{y}

และการประมาณค่าผลรวมของประชากร Y จากค่า $\hat{y} = N\bar{y} = N(\sum_{i=1}^n y_i)/n$

อย่างไรก็ดี การประมาณจะไม่มี ความเอนเอียงต่อเมื่อค่าเฉลี่ยที่ได้จากตัวอย่างต่าง ๆ ขนาด n มีค่าเท่ากับค่าเฉลี่ยของประชากรหรือ \bar{y} เป็นค่าประมาณ

ที่ไม่ลำเอียงของ \bar{Y} เมื่อ $E(\bar{y})$ เท่ากับ \bar{Y} นั่นคือ

$$\begin{aligned} E(\bar{y}) &= \bar{Y} \\ \text{และ} \quad \hat{y} &= N\bar{y} \end{aligned}$$

อย่างไรก็ดี ค่าประมาณที่คำนวณได้นี้ เป็นแต่เพียงค่าคาดหมายที่ได้จากการประมาณเท่านั้น สำหรับค่าจริงของพารามิเตอร์ของประชากรเป็นเท่าไคนั้นยังไม่ทราบ ด้วยเหตุนี้จึงควรคำนวณหาขอบเขตความเชื่อมั่น กล่าวคือหาค่าขอบเขตต่ำและสูง (lower confidence limit และ upper confidence limit) ว่าค่าจริงของพารามิเตอร์ควรมีค่าอยู่ระหว่างค่าใด โดยกำหนดระดับความมีนัยสำคัญ .๐๑, .๐๕ หรือ .๑๐ แล้วแต่เราต้องการความถูกต้องในการประมาณค่าพารามิเตอร์ในระดับความเชื่อมั่นกี่เปอร์เซ็นต์

อนันต์ ศรีโสภาก

บรรณานุกรม

- อนันต์ ศรีโสภาก เทคนิคการสุ่มตัวอย่าง คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ๒๕๒๔
Cochran, W.G. Sampling Techniques. 2nd ed. New York, John Wiley and Sons, Inc., 1963.
Kish, L. Survey Sampling. New York, John Wiley and Sons, Inc., 1965.