

การสุ่มตัวอย่าง

ความหมายของการสุ่มตัวอย่าง

การสุ่มตัวอย่างเป็นกระบวนการคัดเลือกตัวอย่างจากประชากรที่ปราศจากการลำเอียง (unbiased) โดยตัวอย่างที่สุ่มได้นั้นสามารถใช้เป็นตัวแทนของประชากร เช่น ในการเลือกตัวอย่างของผู้ที่มีสิทธิออกเสียงลงคะแนน จากบัญชีรายชื่อผู้มีสิทธิออกเสียงเลือกตั้งทั้งหมดจำนวนหนึ่ง แล้วถามทัศนคติของแต่ละคนในตัวอย่างที่เลือกมานี้ว่าชอบผู้สมัครรับเลือกตั้งคนไหนมากที่สุด จากข้อมูลที่ได้มาได้จากตัวอย่างนี้ เราก็สามารถสรุปได้ว่าประชาชนทั้งหมดในหมู่บ้านนี้มีทัศนคติชอบใครมากที่สุด เป็นต้น

วิธีการสุ่มตัวอย่าง

ก่อนที่จะทำการสุ่มตัวอย่างจะต้องแบ่งประชากรออกเป็น ส่วน ๆ ซึ่งเรียกว่าหน่วยการสุ่ม (sampling units) หน่วยการสุ่มประกอบด้วยสมาชิกหนึ่งตัวหรือมากกว่าก็ได้ แต่ถ้าสมาชิกใดเมื่ออยู่ในหน่วยการสุ่มใดแล้วจะไปปรากฏที่หน่วยอื่นอีกไม่ได้ หน่วยการสุ่มอาจจะเป็นนักเรียนแต่ละคนหรือนักเรียนแต่ละห้อง หรือโรงเรียนแต่ละแห่งก็ได้ แล้วทำการคัดเลือกหน่วยการสุ่มให้มีจำนวนตามขนาดของตัวอย่าง (sample size) จากประชากร (population) โดยแต่ละหน่วยการสุ่มมีความน่าจะเป็นที่จะถูกเลือกมาเท่า ๆ กัน กลุ่มของหน่วยการสุ่มที่สุ่มมาจากประชากรนี้ เรียกว่าตัวอย่าง (sample) การสุ่มตัวอย่างมีหลายแบบ เช่น

การสุ่มแบบธรรมดา (simple random sampling) เป็นการเลือกตัวอย่างขนาด n จากประชากร N โดยที่ทำให้ตัวอย่างขนาด n แต่ละตัวอย่างมีโอกาสที่จะถูกเลือกมาเท่า ๆ กัน เช่น เราเลือกสุ่มประชาชนที่มีสิทธิออกเสียงลงคะแนนมาจำนวน n คน จากบัญชีรายชื่อผู้มีสิทธิออกเสียงลงคะแนนทั้งหมด โดยที่ทำให้ตัวอย่างที่มีขนาด n แต่ละตัวอย่างมีความน่าจะเป็นที่จะถูกเลือกเท่า ๆ กัน การเลือกตัวอย่างโดยวิธีนี้มักจะสมมติให้ผู้มีสิทธิออก

เสียงมีสถานภาพทางสังคมและเศรษฐกิจเหมือนกัน อย่างไรก็ตามทัศนคติของประชาชนที่มีต่อผู้สมัครเป็นสมาชิกสภาผู้แทนราษฎร อาจแตกต่างกันเนื่องจากประชากรแต่ละคนมีสถานภาพทางสังคมและเศรษฐกิจไม่เหมือนกัน ดังนั้นการสุ่มตัวอย่างแบบนี้จึงทำให้ไม่ได้ตัวแทนที่แท้จริงของประชากร

การสุ่มแบบแบ่งชั้น (stratified random sampling) สมมติว่า ประชาชนในหมู่บ้านนี้ประกอบด้วยคนที่มีความแตกต่างกัน 2 พวก คือ พวกที่มีการศึกษาสูง และพวกที่มีการศึกษาค่ำ พวกที่มีการศึกษาสูงอาจจะมีทัศนคติต่อผู้สมัครเป็นสมาชิกสภาผู้แทนราษฎรแตกต่างจากพวกที่มีการศึกษาค่ำ ดังนั้นเพื่อให้ได้ข้อมูลที่มีความถูกต้องเกี่ยวกับประชากร เราจำเป็นต้องสุ่มตัวอย่างจากประชากรแต่ละพวก การที่เราแบ่งประชากรออกเป็น 2 พวก หรือ 2 ชั้น (strata) แล้วทำการสุ่มแบบขววมตา จากแต่ละชั้น จึงเรียกการสุ่มแบบนี้ว่าการสุ่มแบบแบ่งชั้น

การสุ่มแบบการกะประมาณอัตราส่วน (ratio estimation) อนึ่งในการสุ่มตัวอย่าง บางครั้งยังมีตัวแปรอื่น ๆ บางตัวที่มีความสัมพันธ์กับค่าตอบที่เราต้องการจะทราบ ถ้าหากเราวัดตัวแปรที่เราต้องการและตัวแปรที่ใกล้เคียง (subsidiary variable) อื่น ๆ อีกก็จะช่วยให้ข้อมูลเพิ่มขึ้นในการกะประมาณค่าเฉลี่ยของประชากร วิธีการกะประมาณโดยอาศัยการใช้ตัวแปรใกล้เคียงเช่นนี้จึงเรียกรัตอนี้ว่า การกะประมาณอัตราส่วน วิธีการนี้จำเป็นต้องทำการวัดตัวแปร 2 ตัว คือ ตัวแปร y และตัวแปร x ของสมาชิกแต่ละสมาชิกในตัวอย่าง

การสุ่มแบบแบ่งกลุ่ม (cluster sampling)

ในการสุ่มโดยทั่วไปมักจะพบว่าบัญชีรายชื่อเกี่ยวกับประชากร ครัวเรือน หรือที่นาถือครองของชาวนา ฯลฯ มีการเปลี่ยน

แปลงอยู่เสมอ ไม่เป็นปัจจุบัน จึงทำให้หน่วยที่ใช้ในการสุ่มเป็นหน่วยที่ไม่เหมาะสม จะต้องใช้หน่วยการสุ่มที่ใหญ่กว่าซึ่งเรียกว่ากลุ่ม (cluster) หลังจากเราแบ่งประชากรออกเป็นกลุ่ม ซึ่งภายในแต่ละกลุ่มจะมีสมาชิกที่มีลักษณะคล้ายกันแล้วทำการสุ่มระหว่างกลุ่มเหล่านี้ด้วยวิธีการสุ่มแบบธรรมดา หลังจากนั้นก็ทำการวัดข้อมูลจากสมาชิกทั้งหมด ภายในกลุ่มที่สุ่มได้ วิธีการสุ่มแบบนี้เรียกว่าการสุ่มแบบแบ่งกลุ่ม แม้ว่าเราจะแบ่งประชากรออกเป็นชั้น หรือกลุ่ม แล้ววิธีการสุ่มทั้งสองนี้แตกต่างกันคือ การสุ่มแบบแบ่งชั้น เป็นการสุ่มแบบธรรมดา ภายในแต่ละชั้น ส่วนการสุ่มแบบแบ่งกลุ่มเป็นการสุ่มระหว่างกลุ่มต่าง ๆ แบบการสุ่มอย่างธรรมดา และทำการวัดข้อมูลจากสมาชิกทั้งหมดภายในกลุ่มที่สุ่มได้

การสุ่มแบบมีระบบ (systematic random)

บางครั้งชื่อของบุคคลในประชากรที่เราสนใจ ถูกจัดเรียงไว้ในบัญชีหรือในบัตรที่จัดเรียงอยู่ในดินสั๊กเพื่อเป็นการประหยัดและสะดวกในการสุ่มตัวอย่าง เราอาจสุ่มบัตรแต่ละใบจากการใช้ไม้บรรทัดวัดช่วงละหนึ่งนิ้วก็พอ ทุก ๆ ระยะห่างหนึ่งนิ้วจะหยิบบัตรขึ้นมาหนึ่งใบ วิธีการเช่นนี้เรียกว่าการสุ่มแบบมีระบบ การสุ่มแบบนี้ช่วยให้สุ่มตัวอย่างได้อย่างรวดเร็ว และไม่ทำให้เกิดความสับสนและผิดพลาดในการเก็บข้อมูล

การกำหนดขนาดของตัวอย่าง

เนื่องจากจุดมุ่งหมายในการสุ่มตัวอย่างเพื่อต้องการประมาณค่าพารามิเตอร์ของประชากร อาทิ ค่าเฉลี่ยของประชากร ผลรวมของประชากร และบางครั้งก็นำไปใช้ทดสอบสมมติฐานในการทำวิจัย โดยทั่วไปนักสุ่มตัวอย่างมักจะควบคุมปริมาณของข้อมูลในตัวอย่างโดยการจำกัดจำนวนหน่วยการสุ่มในตัวอย่างหรือขนาดของตัวอย่าง และเลือกวิธีในการสุ่มตัวอย่าง การเลือกวิธีในการสุ่มตัว

อย่างและการกำหนดจำนวนหน่วยการสุ่ม หรือขนาดตัวอย่างเพื่อให้ได้ตัวอย่างที่ตีนั้น ขึ้นอยู่กับจำนวนข้อมูลหรือข้อเท็จจริงที่เราต้องการทราบ ถ้าหากให้ θ เป็นค่าพารามิเตอร์ของประชากรที่เราสนใจ และ θ เป็นค่ากะประมาณของ θ ผู้สุ่มตัวอย่างจำเป็นจะต้องกำหนดขอบเขตความคลาดเคลื่อนในการกะประมาณว่าผลต่างระหว่างค่าทั้งสองนี้หรือค่าความคลาดเคลื่อนควรจะต้องน้อยกว่าค่าที่กำหนดไว้เท่าใด สมมติว่าน้อยกว่าค่า B ซึ่ง B เป็นค่าความคลาดเคลื่อนในการกะประมาณ เขียนเป็นสัญลักษณ์ก็จะได้ว่า ความคลาดเคลื่อนในการกะประมาณ $= |\theta - \theta| < B$

นอกจากนี้ยังจำเป็นต้องกำหนดความน่าจะเป็น (probability) ที่โอกาสจะเกิดความคลาดเคลื่อนในการกะประมาณซึ่งมีค่าน้อยกว่า B นี้ เมื่อทำการสุ่มตัวอย่างแบบนี้ซ้ำ ๆ อีกจะเท่ากับ $1 - \alpha$

นั่นคือ

$$P(\text{ความคลาดเคลื่อนในการกะประมาณ} < B) = 1 - \alpha$$

α ตามปกติเรามักจะเลือก $B = 2-5\%$ เมื่อ α เป็นความคลาดเคลื่อนมาตรฐานเนื่องจากการสุ่ม และให้ $1 - \alpha = .95$ เมื่อการแจกแจงของการสุ่มเป็นแบบปกติ (normal distribution)

หลังจากเราได้กำหนดขอบเขตความคลาดเคลื่อนในการกะประมาณโดยให้มีความน่าจะเป็นเท่ากับ $1 - \alpha$ แล้ว ขั้นตอนต่อไปก็ต้องพิจารณาว่าจะใช้ขนาดตัวอย่างเท่าใด และวิธีการสุ่มตัวอย่างแบบใดจึงจะทำให้ตัวอย่างที่ได้เป็นตัวแทนของประชากรและเป็นการสุ่มที่ประหยัดค่าใช้จ่าย

ประโยชน์ของการสุ่มตัวอย่าง

1. ประหยัดค่าใช้จ่ายแรงงานและเวลาในการเก็บรวบรวมข้อมูล
2. ช่วยให้การเก็บรวบรวมข้อมูลและการหาค่าสรุปเกี่ยวกับข้อมูลเป็นไปอย่างรวดเร็วทันต่อเหตุการณ์
3. สามารถขยายขอบเขตในการเก็บข้อมูลได้มากขึ้นในกรณีที่เรามีบุคลากรและเครื่องมือจำกัด เพราะ

ถ้าจะเก็บข้อมูลในประชากรทั้งหมดอาจจะต้องลดจำนวน ชนิดของข้อมูลลง และอาจจะเป็นไปไม่ได้ถ้าหากมีบุคลากร และเครื่องมือจำกัด

ข้อมูลบางชนิดต้องการบุคลากรและเครื่องมือที่มีคุณภาพ สูงมาทำการจัดเก็บ และยังช่วยให้ผู้จัดเก็บข้อมูลได้มีเวลา วิเคราะห์ตรวจทานอย่างละเอียดถี่ถ้วนก่อนที่จะนำไปใช้

4. สามารถเก็บข้อมูลได้ถูกต้องยิ่งขึ้น ทั้งนี้เพราะ

อนันต์ ศรีโสภา

บรรณานุกรม

อนันต์ ศรีโสภา เทคนิคการสุ่มตัวอย่าง คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
 ประสานมิตร พุทธศักราช 2524 300 หน้า

อนันต์ ศรีโสภา สถิติเบื้องต้น บริษัทสำนักพิมพ์ไทยวัฒนาพานิช จำกัด 599
 ถนนไมตรีจิต กรุงเทพมหานคร พ.ศ.2521 396 หน้า

Cochran, W.G., Sampling Techniques. 2d ed., New York :
 John Wiley and Sons, Inc., 1963.

Kish, L., Sourvey Sampling. New York : John Wiley and
 Sons, Inc., 1965.