

ความไม่เหมาะสมของการทดสอบสมมติฐาน กับตัวอย่างแบบไม่สุ่ม

อรุณ จิรวัดณ์กุล

การเลือกตัวอย่างสำหรับการวิจัยโดยปกติจะเลือกโดยวิธีสุ่ม เช่นการสุ่มแบบง่าย การสุ่มแบบเป็นระบบ เป็นต้น มีบางงานวิจัยที่ใช้การเลือกตัวอย่างแบบไม่ใช้ความน่าจะเป็น เช่น การเลือกแบบบังเอิญ การเลือกแบบเจาะจง หรือการเลือกแบบโควตา ซึ่งทำให้ได้ตัวอย่างแบบไม่สุ่ม (non-random sample) จึงมีคำถามว่าจะใช้การทดสอบสมมติฐานกับข้อมูลที่ได้จากตัวอย่างแบบไม่สุ่ม เพื่อสรุปผลการศึกษาระหว่างได้หรือไม่

ทฤษฎีทางสถิติที่ใช้ทดสอบสมมติฐาน ค่าสถิติของตัวอย่างให้เป็นข้อสรุปของประชากร จะมีข้อตกลงพื้นฐานเบื้องต้นว่า ข้อมูลที่นำมาทดสอบสมมติฐานต้องมาจากตัวอย่างแบบสุ่ม

ตัวอย่างแบบสุ่ม หมายถึงหน่วยตัวอย่างที่ใช้ในการศึกษาต้องสุ่มมาจากประชากร โดยที่แต่ละหน่วยศึกษามีโอกาสถูกเลือกเป็นตัวอย่างเท่ากัน หรือทราบความน่าจะเป็นของโอกาสที่ถูกเลือก ตัวแปรที่ได้จากตัวอย่างดังกล่าวเรียกว่าตัวแปรสุ่ม การเลือกตัวอย่างแบบสุ่มจะทำให้ได้ตัวอย่างที่ไม่เอนเอียง

ค่าสถิติ เช่นค่าเฉลี่ยที่คำนวณจากตัวอย่างแบบสุ่มจะเป็นค่าที่ไม่เอนเอียง (unbias) ทำให้การทดสอบสมมติฐานได้ข้อสรุปของประชากรที่ตรงกับความจริง

การทดสอบสมมติฐาน ไม่ว่าจะเป็นแบบ parametric เช่น t-test pair t-test ANOVA หรือแบบ non-parametric เช่น Mann-Whitney test Wilcoxon

signed-rank test Kruskal-Wallis test มีข้อตกลงเบื้องต้นว่า ข้อมูลจากตัวอย่างที่จะใช้ทดสอบสมมติฐานต้องมาจากตัวอย่างแบบสุ่ม⁽¹⁾

การจะเลือกใช้สถิติแบบ parametric หรือ non-parametric ขึ้นอยู่กับข้อตกลงเบื้องต้นของการแจกแจงของตัวแปรที่นำมาทดสอบสมมติฐาน ถ้าเป็น parametric ต้องทราบว่าหน่วยตัวอย่างถูกสุ่มมาจากประชากรที่มีการแจกแจงแบบใด หรือมีข้อตกลงเบื้องต้นว่าประชากรมีการแจกแจงแบบใด สำหรับ non-parametric ไม่มีข้อตกลงเบื้องต้นในเรื่องการแจกแจงของประชากร

ดังนั้นการทดสอบสมมติฐานไม่ว่าเป็น parametric และ non-parametric ข้อมูลที่จะนำมาทดสอบจะต้องได้มาจากตัวอย่างแบบสุ่ม ต่างกันเพียงเรื่องข้อตกลงเบื้องต้นของการแจกแจงของประชากรเท่านั้น

การเลือกตัวอย่างแบบบังเอิญ เป็นการเลือกตัวอย่างแบบไม่สุ่ม ทำให้ได้ตัวอย่างที่เอนเอียง (bias) เมื่อนำค่าสถิติที่คำนวณจากตัวอย่างที่เอนเอียงมาสรุปผลการศึกษา จะได้คำตอบที่เอนเอียงไม่ตรงกับความจริง ตัวอย่างเช่น ในการศึกษาเรื่อง ความเห็นของการทำแท้งเสรีของประชาชนที่อาศัยอยู่ในเขตเทศบาล นักวิจัยใช้วิธีการเลือกตัวอย่างแบบบังเอิญโดยเก็บข้อมูลจากผู้ที่มาซื้ออาหารที่ตลาดสดเทศบาลทั้งสองแห่ง ในวันจันทร์ พุธ และศุกร์ ระหว่างเวลา 6.00-12.00 น. เป็นเวลาหนึ่งสัปดาห์

ถ้าพิจารณาลักษณะคนที่จะมาซื้ออาหารที่ตลาดสดในวันธรรมดาอาจเป็นกลุ่มที่ทำงานบ้านเป็นหลัก ผู้ที่ทำงานสำนักงานอาจมาซื้อในวันซื้อเสาร์หรืออาทิตย์ นอกจากนี้บางคนไม่ได้ซื้ออาหารจากตลาดสด แต่ซื้อจากแหล่งอื่น เช่น ร้านค้าในชุมชน ตลาดเช้า ซูเปอร์มาร์เก็ต เป็นต้น แผนการเลือกตัวอย่างนี้ยังทำให้ได้ตัวอย่างที่เป็นผู้ชาย ผู้สูงอายุ หรือวัยรุ่นน้อย เพราะคนกลุ่มนี้จะไม่ค่อยมาตลาดในช่วงเวลาเก็บข้อมูล และผู้ที่ไม่ได้เป็นผู้ซื้ออาหารจะไม่มีโอกาสถูกเลือกเป็นตัวอย่าง ดังนั้นจะเห็นได้ว่าตัวอย่างที่เลือกได้จากแผนการสุ่มดังกล่าวมีความเอนเอียง ไม่สามารถใช้เป็นตัวแทนที่ดีของประชาชนที่อาศัยอยู่ในเขตเทศบาลได้ ถ้านักวิจัยนำข้อมูลที่เก็บได้ทดสอบสมมุติฐานทางสถิติจะมีความไม่เหมาะสมสองประการคือ

1. ตัวอย่างที่ได้เป็นตัวอย่างที่เอนเอียงจะทำให้ได้ข้อสรุปที่เอนเอียง
2. เมื่อตัวอย่างที่ศึกษาไม่ทราบความน่าจะเป็น

ของโอกาสที่ถูกเลือกเป็นตัวอย่างทำให้ไม่สามารถใช้ความน่าจะเป็นในการอนุมานผลการทดสอบของตัวอย่างให้เป็นข้อสรุปของประชากร

สรุป

นักวิจัยที่ใช้การเลือกตัวอย่างแบบบังเอิญ หรือวิธีเลือกตัวอย่างแบบไม่สุ่มอื่นๆ ตัวอย่างที่ได้จะมีความเอนเอียง ทำให้ได้คำตอบงานวิจัยที่เอนเอียงไม่ตรงกับความจริง ตามทฤษฎีค่าสถิติที่คำนวณจากข้อมูลของตัวอย่างดังกล่าวจะไม่สามารถนำมาทดสอบสมมุติฐานเพื่อใช้สรุปผลการศึกษาในประชากร

เอกสารอ้างอิง

1. อรุณ จิรวัดน์กุล. สถิติทางวิทยาศาสตร์สุขภาพเพื่อการวิจัยที่ใช้ในงานวิจัยทางวิทยาศาสตร์สุขภาพ. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร: วิทย์พัฒนา; 2552.