วารสารวิชาการสาธารณสุข
ปีที่ 24 ฉบับที่ 4 กรกฎาคม - สิงหาคม 2558

Journal of Health Science
Vol. 24 No. 4, July - August 2015

มุมสถิติ *Statistics Corner*

การวิเคราะห์ข้อมูลจากตัวอย่างที่สุ่มตามสะดวก

อรุณ จิรวัฒน์กุล วท.บ. (อาชีวอนามัย), วท.ม. (ชีวสถิติ), M.Sc. (Clinical Epidemiology)

การสุ่มตามสะดวกเก็บข้อมูลได้เร็ว (quick and dirty) แต่ข้อมูลจากตัวอย่างที่เก็บได้ ไม่สามารถใช้สรุปลักษณะประชากร เพราะเป็นการสุ่มแบบไม่ใช้ความน่าจะเป็น หน่วยตัวอย่างที่สุ่มได้แต่ละตัวมีโอกาสถูกสุ่มได้ไม่เท่ากัน ทำให้แต่ละตัวอย่างที่สุ่มได้มีลักษณะต่างกัน[1] ถ้านำค่าสถิติจากตัวอย่างที่สุ่มด้วยวิธีนี้มาอนุมานแบบจุด (point estimation) จะได้ค่าสรุปลักษณะประชากร (พารามิเตอร์) แตกต่างไปตามตัวอย่างที่สุ่มได้โดยไม่ทราบว่า ค่าพารามิเตอร์ที่อนุมานได้ถูกต้องหรือไม่ และการสุ่มแบบไม่ใช้ความน่าจะเป็นยังทำให้ไม่สามารถใช้ทฤษฎีทางสถิติอนุมานค่าสถิติจากตัวอย่างไปเป็นค่าพารามิเตอร์ของประชากร

ในกรณีที่ต้องการใช้ข้อมูล แต่มีข้อจำกัดเรื่องค่าใช้จ่ายและเวลา มีความจำเป็นต้องใช้วิธีสุ่มตามสะดวกในการสำรวจ เพื่อให้มีข้อมูลใช้ในการดำเนินงาน ข้อมูลที่ได้จากการสุ่มตามสะดวกแม้ไม่สามารถใช้เป็นตัวแทนของประชากรทั้งหมด แต่เป็นข้อมูลของประชากรศึกษาบางส่วน การมีข้อมูลบางส่วนของประชากรศึกษาย่อมจะดีกว่าไม่มีข้อมูลอะไรเลย ดังนั้นหากมีความจำเป็นต้องใช้การสุ่มตามสะดวกในการสำรวจ จะมีวิธีวิเคราะห์อย่างไรที่จะช่วยเพิ่มความถูกต้องของผลการสำรวจ

การสรุปผลจากตัวอย่างไปเป็นลักษณะของประชากรในกรณีที่ตัวอย่างสุ่มแบบไม่ใช้ความน่าจะเป็น ต้องใช้ตรรกะสรุปลักษณะที่ได้จากค่าสถิติของตัวอย่างไปเป็นค่าพารามิเตอร์ของประชากร แทนการใช้ทฤษฎีความน่าจะเป็น

วิธีการทางสถิติวิธีหนึ่งที่จะช่วยเพิ่มความถูกต้องของการประมาณค่าภาพรวมของประชากรจากตัวอย่าง คือ การแยกข้อมูลเป็นกลุ่ม (stratification) ตามปัจจัยที่มีอิทธิพลมากต่อตัวแปรผลหลัก (main outcome) ของเรื่องที่ศึกษา วิเคราะห์ผลแยกเป็นกลุ่มย่อย แล้วนำผลการวิเคราะห์ที่ได้จากแต่ละกลุ่มย่อยมารวมแบบถ่วงน้ำหนักตาม สัดส่วนของกลุ่มย่อยในประชากรจริง ผลรวมที่ได้จะเป็นค่าประมาณลักษณะประชากรได้ถูกต้องเพิ่มขึ้น

ในการดำเนินการดังกล่าว นักวิจัยต้องมีการทบทวนวรรณกรรมที่ดีเพื่อระบุให้ได้ว่าปัจจัยที่มีอิทธิพลมากเป็นตัวแปรใด และต้องมีข้อมูลสัดส่วนของกลุ่มย่อยที่แบ่งตามตัวแปรที่มีอิทธิพลดังกล่าวของประชากรที่ศึกษา

ตัวอย่างการสำรวจความคิดเห็นเรื่องสิทธิของการทำแท้งที่ไม่พึงประสงค์ในเขตเทศบาลเมือง ผู้ดำเนินการสำรวจเลือกใช้วิธีสุ่มตามสะดวกจากประชาชนที่มีอายุ 15 ปีขึ้นไปที่อาศัยอยู่ในเขตเทศบาลเมือง การทบทวนวรรณกรรมจากงานวิจัยที่ผ่านมาพบว่า ความคิดเห็นเรื่องดังกล่าวจะแตกต่างกันมากตามกลุ่มอายุและเพศ (กลุ่มอายุ และเพศ เป็นปัจจัยที่มีอิทธิพลมากต่อคำตอบ) และจำนวนประชากรในเขตเทศบาลเมืองจำแนกตามกลุ่มอายุและเพศได้มาจากฐานข้อมูลประชากรของทะเบียนราษฎร

ดำเนินการเก็บข้อมูลด้วยแบบสอบถาม โดยแบบสอบถามที่ใช้เก็บข้อมูลจะต้องมีตัวแปรปัจจัยที่มีอิทธิพลถามอยู่ในแบบสอบถามด้วย เพื่อจะได้ใช้ในการวิเคราะห์

แบ่งกลุ่มตามปัจจัยที่มีอิทธิพลดังกล่าว จากตัวอย่างการสำรวจความคิดเห็นเรื่องสิทธิของการทำแท้งที่ไม่พึงประสงค์ ตัวแปรผลหลักที่ใช้ตอบคำถามงานวิจัยนี้คือ ความเห็น (เห็นด้วย หรือไม่เห็นด้วย) ว่าผู้หญิงตั้งครรภ์ไม่พึงประสงค์ควรมีสิทธิทำแท้งได้โดยไม่ผิดกฎหมาย ส่วนตัวแปรปัจจัยที่มีอิทธิพลต่อคำตอบคือ ตัวแปรเพศและอายุ

ในขั้นตอนการวิเคราะห์ นักวิจัยใช้ตรรกะในการสรุปโดยกำหนดเป็นสมมุติฐานว่า ความคิดเห็นในเรื่องดังกล่าวจะแตกต่างตามเพศและอายุเช่นเดียวกันกับงานวิจัยที่ผ่านมา และประชาชนในกลุ่มอายุและเพศเดียวกันกับตัวอย่างมีความคิดเห็นเหมือนกัน จึงคำนวณอัตราผู้ที่เห็นด้วยว่าผู้หญิงตั้งครรภ์ไม่พึงประสงค์ควรมีสิทธิทำแท้งได้โดยไม่ผิดกฎหมาย จำแนกตามเพศและอายุของผู้ตอบ แล้วจึงนำอัตราดังกล่าวไปคำนวณหาจำนวนผู้เห็นด้วยจำแนกตามกลุ่มอายุและเพศ ตามจำนวนประชาชนในแต่ละกลุ่มจากฐานข้อมูลประชากรที่อาศัยอยู่ในเขตเทศบาลเมือง เช่น จากการสำรวจพบว่าในกลุ่มผู้หญิงอายุ 15-19 ปีมีผู้ที่เห็นด้วยร้อยละ 82.0 จากฐานข้อมูลประชากรของประชาชนที่อาศัยอยู่ในเขตเทศบาลเมืองมีผู้หญิงอายุ 15-19 ปีอยู่ 5,611 คน จึงนำมาคำนวณหาจำนวนผู้ที่เห็นด้วยจะได้เท่ากับ

$$\frac{5{,}611 \times 82}{100} = 4{,}601 \text{ คน}$$

คำนวณเช่นเดียวกันนี้ในทุกกลุ่มอายุและเพศ แล้วจึงนำจำนวนผู้เห็นด้วยที่คำนวณได้จากแต่ละกลุ่มอายุและเพศมารวมกันเป็นผู้ที่เห็นด้วยทั้งหมด คำนวณร้อยละของผู้เห็นด้วยโดยหารด้วยจำนวนประชากรอายุ 15 ปีขึ้นไปทั้งหมด คูณด้วยร้อย จะได้ร้อยละของผู้เห็นด้วยถูกต้องมากขึ้นกว่าผลที่คำนวณโดยตรงจากข้อมูลตัวอย่าง

วิธีการวิเคราะห์เป็นกลุ่มย่อยตามปัจจัยที่มีอิทธิพลแล้วนำมาปรับด้วยฐานประชากรนี้ ทำในลักษณะเดียวกันกับการปรับอิทธิพลของเพศอายุในการปรับฐานการตาย (adjusted death rate) หรือการเจ็บป่วย

สรุป

การวิเคราะห์ข้อมูลที่ได้จากการสุ่มตามสะดวกสามารถเพิ่มความถูกต้องของข้อสรุปภาพรวมของประชากรได้ โดยการทำการวิเคราะห์แบบแบ่งกลุ่มตามปัจจัยที่มีอิทธิพลมากต่อตัวแปลผลที่ใช้ตอบคำถามหลัก ซึ่งผู้วิเคราะห์ข้อมูลต้องทำการทบทวนวรรณกรรมให้มีความเข้าใจเรื่องที่ศึกษาดีพอที่สามารถระบุได้ว่าปัจจัยใดที่มีอิทธิมากต่อคำตอบที่ต้องการศึกษา และสามารถหาข้อมูลสัดส่วนของกลุ่มย่อยตามปัจจัยที่มีอิทธิพลดังกล่าวในประชากรได้

เอกสารอ้างอิง

1. Gohdes A. Different convenience samples, different stories: the case of Sierra Leone. Palo Alto: Human Rights Data Analysis Group at Benetech; 2010.