

มุขสक्ति

Statistics Corner

ความผิดพลาดที่พบบ่อยในการสรุปผลต่างของค่าเฉลี่ย

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

ในงานวิจัยเชิงวิเคราะห์ และเชิงทดลองเรื่องความรอบรู้ด้านสุขภาพ ผู้วิจัยกำหนดให้ตัวแปรความรู้เป็นตัวแปรหลักตัวแปรหนึ่งในการประเมินความรอบรู้ด้านสุขภาพ ในกรณีที่มีการเปรียบเทียบความต่างของคะแนนความรู้ระหว่างกลุ่ม นักวิจัยใช้การทดสอบสมมติฐานเปรียบเทียบค่าเฉลี่ยคะแนนความรู้ จากรายงานวิจัยที่ดีพิมพ์พบว่าการสรุปผลการทดสอบ จะพิจารณาจากความต่างทางสถิติอย่างเดียวโดยไม่ดูขนาดของความต่างศาสตร์ของการเรียนการสอน และระดับคะแนนความรู้ ทำให้การสรุปผลการทดสอบเกิดความผิดพลาด

ข้อผิดพลาดที่พบบ่อยในการสรุปผลต่างของค่าเฉลี่ยของคะแนนความรู้แบ่งได้ 4 กรณี

กรณีที่ 1 การสรุปผลความต่างโดยไม่นำขนาดความต่างมาร่วมพิจารณา ในการทดสอบสมมติฐานของค่าเฉลี่ยคะแนนความรู้ระหว่างกลุ่มในบางครั้งพบว่าค่าเฉลี่ยต่างกัน 2-3 คะแนนแต่ผลการทดสอบสมมติฐานพบความต่างอย่างมีนัยสำคัญแล้ว ทั้งนี้เพราะตัวแปรต่อเนื่องหน่วยการวัดมีความละเอียดสามารถวัดความต่างได้เป็นทศนิยม ความต่างเพียงไม่กี่คะแนนถ้ามากกว่าสองเท่าของความคลาดเคลื่อนมาตรฐาน ก็จะทำให้ผลต่างอย่างมีนัยสำคัญ เมื่อนำขนาดความต่างอย่างมีนัยสำคัญไปพิจารณากับงานที่จะนำไปใช้ อาจพบว่าขนาดความต่างนั้นไม่มีความหมาย เช่น เครื่องชั่ง 2 แบบให้ค่าเฉลี่ยน้ำหนักต่างกัน 20 กรัม ถ้าใช้เครื่องชั่งดังกล่าวชั่งน้ำหนักผู้ใหญ่ ความต่างที่พบไม่มีความหมาย จะเลือกใช้เครื่องชั่งแบบ

ใดก็ได้ แต่ถ้าเครื่องชั่งดังกล่าวใช้ชั่งน้ำหนักทารกความต่าง 20 กรัมจะมีความหมาย

ตารางที่ 1 การประเมินความสามารถในการอ่านเข้าใจเปรียบเทียบระหว่าง 2 กลุ่ม

กลุ่มที่	Mean	SD	N
1	78.5	3.2	50
2	76.1	3.6	50

95% CI=2.4(1.05-3.75), p-value=0.0006

ตารางที่ 1 เป็นเรื่องการประเมินความสามารถในการอ่านเข้าใจ ว่าทั้งสองกลุ่มมีความสามารถอ่านและเข้าใจคำถามในการประเมินความรอบรู้ด้านสุขภาพต่างกันหรือไม่ ถึงแม้ผลการทดสอบการอ่านพบมีความต่างอย่างมีนัยสำคัญทางสถิติ แต่ผู้วิจัยมีประสบการณ์ว่าคะแนนความสามารถในการอ่านต่างกัน 2.4 คะแนน สามารถอ่านแบบประเมินความแตกฉานด้านสุขภาพได้ไม่ต่างกันจึงสรุปผลว่า “พบว่าทั้งสองกลุ่มมีความสามารถในการอ่านเข้าใจต่างกันอย่างมีนัยสำคัญ แต่ขนาดความต่างที่พบไม่มีผลทำให้อาสาสมัครทั้งสองกลุ่มเข้าใจคำถามในการประเมินความแตกฉานด้านสุขภาพได้ต่างกัน”

ดังนั้นเมื่อพบว่าผลการทดสอบสมมติฐานมีความต่างอย่างมีนัยสำคัญ ผู้วิจัยต้องพิจารณาขนาดความต่างว่ามีความหมายเมื่อนำไปใช้งานหรือไม่

กรณีที่ 2 การพิจารณาความต่างของคะแนนความรู้ โดยไม่นำคะแนนเต็มของแบบประเมินมาร่วมพิจารณา เช่น ในการประเมินความรอบรู้ในเรื่องการควบคุมน้ำตาล ในเลือดของผู้ป่วยเบาหวาน ระหว่างอสม.กับผู้ป่วย ในตารางที่ 2 แบบประเมินมีคะแนนเต็ม 100 คะแนน

ตารางที่ 2 ความต่างของคะแนนความรู้ระหว่างอาสาสมัคร สาธารณสุขประจำหมู่บ้านและผู้ป่วย

กลุ่ม	Mean	SD	N
อสม.	38.1	4.6	44
ผู้ป่วย	48.6	4.2	44

หมายเหตุ: 95%CI=10.5 (8.6-12.4), p-value<0.001

ผลการทดสอบสมมติฐานค่าเฉลี่ยความรู้ทั้งสองกลุ่ม ต่างกันอย่างมีนัยสำคัญ โดยผู้ป่วยมีคะแนนเฉลี่ยสูงกว่า อสม. 10.5 คะแนน ถ้าพิจารณาจากคะแนนความรู้ที่มี กับความรู้ที่ควรมีทั้งหมด (คะแนนเต็ม) พบว่าทั้งสอง กลุ่มมีความรู้ไม่ถึงครึ่งของความรู้ที่ควรมี ผู้วิจัยต้องสรุปว่า “ถึงแม้ ผู้ป่วยจะมีความรู้สูงกว่าอสม. แต่ทั้งสองกลุ่ม ยังมีความรู้ไม่ถึงครึ่งของสิ่งที่ควรมี จึงจำเป็นต้องปรับปรุง วิธีการเพิ่มความรอบรู้ให้ทั้งสองกลุ่ม” เพื่อป้องกันความ ผิดพลาดกรณีนี้ ผู้วิจัยต้องพิจารณาระดับความรู้ของ อาสาสมัครร่วมกับผลการทดสอบสมมติฐาน

กรณีที่ 3 การเปรียบเทียบคะแนนรอบรู้ด้านสุขภาพ ไม่ควรนำผลต่างระหว่างการวัดก่อนและหลังการสอน ระหว่างสองกลุ่มมาเปรียบเทียบกับกลุ่มใดหลังสอน แล้วมีความรู้เพิ่มจากเดิมมากกว่ากัน เพราะหลักสูตรการ สอนความรอบรู้ด้านสุขภาพ จะออกแบบเนื้อหาเพื่อสอน ให้คนที่ไม่มีความรู้มาก่อน หรือมีความรู้อยู่บ้างแล้ว เมื่อ ได้รับการสอนทุกคนจะมีความรู้เพียงพอเหมือนกัน ดัง

นั้นการนำผลต่างความรู้ที่เพิ่มขึ้นมาเปรียบเทียบกับจึงไม่ สอดคล้องกับเนื้อหาของหลักสูตรการสอน และไม่สามารถสะท้อนว่าหลังการทดลองทั้งสองกลุ่มมีความรู้ ด้านสุขภาพต่างกันหรือไม่

กรณีที่ 4 ในการประเมินความรอบรู้ด้านสุขภาพที่มี การกำหนดเกณฑ์ เช่นกำหนดว่าผู้ที่มีความรอบรู้ด้าน สุขภาพเพียงพอต้องได้คะแนนความรู้ไม่น้อยกว่าร้อยละ 80 ในกรณีนี้ถ้านำค่าเฉลี่ยระหว่างกลุ่มมาเปรียบเทียบ อาจให้ผลสรุปที่ไม่ถูกต้อง เช่นในการประเมินความรอบรู้ เรื่อง HIV ระหว่างเพศชาย และหญิงมีค่าเฉลี่ยคะแนน ความรอบรู้ 84.8 และ 81.2 ตามลำดับ ถึงแม้ค่าเฉลี่ย คะแนนความรอบรู้จะเกินร้อยละ 80.0 แต่ไม่สามารถสรุป ได้ว่าทุกคนมีคะแนนความรอบรู้เกิน 80.0 เพราะค่าเฉลี่ย เป็นการนำเอาผู้ที่มีคะแนนสูงไปเฉลี่ยให้กับผู้ที่มีคะแนน ต่ำซึ่งเป็นการสรุปที่ไม่ถูกต้อง ในกรณีนี้ต้องนำคะแนน ความรอบรู้ของแต่ละคนมาพิจารณากับเกณฑ์ว่ามีความ รอบรู้ตามเกณฑ์หรือไม่ แล้วนำไปสรุปว่าเพศชาย/หญิง มีความรู้ตามเกณฑ์ร้อยละเท่าไร ในการเปรียบเทียบความ รอบรู้ระหว่างเพศจะทดสอบสมมติฐานสัดส่วนความ รอบรู้ตามเกณฑ์ระหว่างเพศ จึงจะให้ข้อสรุปที่ถูกต้อง

สรุป

การสรุปผลต่างค่าเฉลี่ยระหว่างกลุ่ม ควรนำผลการ ทดสอบสมมติฐาน ขนาดความต่างที่มีความหมายเมื่อนำ ไปใช้งาน และระดับคะแนนความรู้ มาร่วมพิจารณาสรุป ผลต่างของค่าเฉลี่ย

ไม่ควรนำผลการเปรียบเทียบความรู้ที่เพิ่มขึ้นระหว่าง ก่อนและหลังการทดลองมาสรุปผลความต่างของความ รอบรู้ และถ้ามีการกำหนดเกณฑ์ประเมินความรอบรู้ไม่ ควรใช้ค่าเฉลี่ยในการเปรียบเทียบ