

การพิจารณาขนาดความต่างที่มีความสำคัญ (1)

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

การทดสอบสมมติฐานงานวิจัย เมื่อพิจารณาค่า p value เทียบกับค่า α แล้ว ผลการทดสอบพบว่าต่างอย่างมีนัยสำคัญ นักวิจัยส่วนใหญ่จะสรุปว่า ความต่างนั้นมีความสำคัญมีประโยชน์ควรนำไปใช้งาน โดยไม่ได้พิจารณาขนาดความต่างว่ามีค่าน้อยเท่าไร ซึ่งเป็นการแปลผลที่ผิด

การแปลและอภิปรายผลการทดสอบสมมติฐานที่ถูกต้อง เมื่อผลการทดสอบพบว่าต่างอย่างมีนัยสำคัญ นักวิจัยต้องพิจารณาขนาดความต่างประกอบด้วย ในบทความนี้จะนำเสนอสถิติที่นิยมใช้ช่วยพิจารณาขนาดความต่างในงานวิจัยด้านการแพทย์และสาธารณสุข ทั้งที่เป็นงานวิจัยทางด้านพฤติกรรมศาสตร์/สังคมศาสตร์ และงานวิจัยด้านระบาดวิทยา/วิจัยทางคลินิก

ในงานวิจัยด้านพฤติกรรมศาสตร์/สังคมศาสตร์ การระบุขนาดความต่างนิยมพิจารณาจากค่าช่วงเชื่อมั่นและขนาดอิทธิพล (effect size)

ขนาดอิทธิพลที่ใช้ดูขนาดความต่างของค่าเฉลี่ยคำนวณจากผลต่างระหว่างค่าเฉลี่ยของกลุ่มทดลองกับกลุ่มควบคุมหารด้วยส่วนเบี่ยงเบนมาตรฐาน

$$\frac{\bar{X}_t - \bar{X}_c}{S}$$

ค่าขนาดอิทธิพลที่คำนวณได้จะไม่มีหน่วยเช่นเดียวกันกับคะแนนมาตรฐาน (standard score หรือ Z score) ทำให้สามารถใช้ในการระบุขนาดความต่างได้ดีกว่าความต่างระหว่างค่าเฉลี่ยอย่างเดียว การแปลความหมายของขนาดอิทธิพล Jacob Cohen⁽¹⁾ ได้นำเสนอเป็นเกณฑ์กว้างๆ ดังตารางที่ 1

ตารางที่ 1 การแปลความหมายของขนาดอิทธิพล

ค่าขนาดอิทธิพล	ขนาดความแตกต่าง
ประมาณ 0.2	น้อย
ประมาณ 0.5	ปานกลาง
มากกว่า 0.8	มาก

ผู้ใช้เกณฑ์นี้ต้องใช้ด้วยความระมัดระวังเพราะเกณฑ์กว้างๆนี้อาจแตกต่างกันไปตามตัวแปรผลและสาขาวิชา สูตรคำนวณค่าขนาดอิทธิพลจะแตกต่างกันไปสถิติที่ใช้ เช่น multiple regression ใช้ F test (Cohen's f^2), Chi square test (Cohen's w) เป็นต้น

ช่วงเชื่อมั่นเป็นการประมาณค่าขนาดความต่างของค่าเฉลี่ยระหว่างกลุ่มทดลองกับกลุ่มควบคุม เช่น ในการสอนสุขศึกษาผู้ป่วยเบาหวานที่สอนวิธีใหม่เทียบกับการสอนวิธีเดิมได้ค่า 95% ช่วงเชื่อมั่นผลต่างของค่าเฉลี่ยเท่ากับ 1.4 - 5.2 แสดงว่าคะแนนความรู้ที่สอนด้วยวิธีใหม่สูงกว่าวิธีเดิมอย่างน้อย 1.4 คะแนน และมีความเป็นไปได้ที่รู้เพิ่มมากที่สุดเพียง 5.2 คะแนน ถ้าคิดว่าขนาดความรู้ที่เพิ่มขึ้นไม่น่าจะมีผลต่อการเปลี่ยนแปลงพฤติกรรม จะสรุปว่าการสอนวิธีใหม่มีความรู้เพิ่มจากเดิมไม่มากพอที่จะนำไปใช้ควรไปพัฒนาหาวิธีการสอนใหม่ที่ดีกว่านี้ สูตรคำนวณค่าช่วงเชื่อมั่นจะแตกต่างกันไปตามสถิติที่ใช้

สำหรับงานวิจัยด้านระบาดวิทยา และวิจัยทางคลินิก สถิติที่นิยมใช้ในการระบุขนาดความต่าง ได้แก่ ช่วงเชื่อมั่น

RR, OR, ARR และ NNT

ตัวอย่างเรื่องการกิน aspirin เพื่อป้องกันการเกิด myocardial infarction (MI) ของชายอายุ 45-56 ปี โดยประเมินการเกิด MI ภายใน 2 ปี พบว่าในกลุ่มที่กิน aspirin เกิด MI 6.8% ($p_t = 0.680$) และกลุ่มที่ไม่กิน aspirin เกิด MI 10.08% ($p_c = 0.101$) คำนวณได้ค่า 95% ช่วงเชื่อมั่น = -1.10% to -5.46%, RR = 1.48, ARR = 3.28% และ NNT = 31

ช่วงเชื่อมั่นของ ($p_t - p_c$) จะใช้การพิจารณาว่า สัดส่วนหรือร้อยละของสองกลุ่มต่างกันเท่าไร เช่น จากตัวอย่าง ได้ค่า 95% ช่วงเชื่อมั่น -1.1% to -5.46% หมายความว่า การเกิด MI ในกลุ่มที่กิน aspirin น้อยกว่ากลุ่มที่ไม่ได้กินอย่างน้อย 1.10% และต่างกันสูงสุดไม่เกิน 5.46% ถ้าความต่างร้อยละ 1.10 มีความสำคัญต่อการป้องกันโรค (clinical importance) จะสรุปว่าการกิน aspirin มีประโยชน์ช่วยป้องกันการเกิด MI ค่า RR คำนวณจากค่า p_t/p_c เป็นค่าที่แสดงว่า การรักษาในกลุ่มทดลองได้ผลเป็นกี่เท่าของกลุ่มควบคุม (ควรให้ p ที่มีค่ามากเป็นพิเศษและค่าน้อยเป็นส่วนจะอ่านผลได้ง่าย) จากตัวอย่างได้ค่า RR = 1.48 หมายความว่า ผู้ที่ไม่ได้กิน aspirin จะมีโอกาสเกิด MI มากกว่าผู้ที่ได้กิน aspirin 1.48 เท่า

ส่วนค่า OR แปลความหมายได้เช่นเดียวกับ RR แต่ใช้กับวิธีการศึกษา cross-sectional และ case-control

ค่า absolute risk reduction (ARR) คำนวณจากค่า $|p_t - p_c| \times 100$ ใช้แสดงว่ากลุ่มทดลองได้ผลต่างจากกลุ่มควบคุมร้อยละเท่าไร เช่นจากตัวอย่างได้ค่า ARR = 3.28% หมายความว่าผู้ที่กิน aspirin มีโอกาสเกิด MI น้อยกว่าผู้ที่ไม่ได้กิน 3.28%

ค่า number needed to treat (NNT) คำนวณจากค่า $1/ARR$ ใช้แสดงว่าวิธีการรักษาใหม่จะต้องใช้กับผู้ป่วยกี่รายในช่วงเวลาที่กำหนดถึงจะได้ผลดีเพิ่มขึ้น 1 ราย จากตัวอย่าง NNT = 31 หมายความว่าถ้าให้กิน aspirin 31 คนเป็นเวลา 2 ปี จะสามารถลดการเกิด MI ได้เพิ่มหนึ่งคน

การแปลความหมายขนาดความต่างที่พบว่ามีมีความสำคัญหรือไม่ จะแตกต่างกันไปตามตัวแปลผลและเรื่องที่ศึกษา เช่น บางเรื่องตัวแปลผลเปลี่ยนแปลงยาก ความต่างร้อยละ 1 ก็มีความสำคัญ บางเรื่องต่างกันร้อยละ 10 ถือว่ายังไม่มากพอที่จะบอกว่ามีมีความสำคัญ ดังนั้นผู้ที่พิจารณาว่า ขนาดความต่างที่พบมีความสำคัญหรือไม่ จะต้องมีความรู้ในเรื่องดังกล่าวดี หรือเป็นผู้ทำงานเกี่ยวข้องกับการใช้ผลการวิจัยดังกล่าวเท่านั้น

มีความพยายามที่จะกำหนด minimal important difference (MID) เพื่อจะระบุว่า เรื่องใดควรพบความต่างเท่าไรจึงถือว่ามีความสำคัญ จะอธิบายในฉบับหน้า

สรุป

ในการแปลผลทดสอบสมมุติฐาน เมื่อพบว่าต่างอย่างมีนัยสำคัญแล้ว นักวิจัยต้องพิจารณาขนาดความต่างจากค่าช่วงเชื่อมั่น หรือสถิติที่ใช้แสดงขนาดความต่างอื่น การตัดสินใจขนาดความต่างที่พบมีความสำคัญหรือไม่ ผู้พิจารณาต้องมีความรู้ในเรื่องดังกล่าวดี หรือเป็นผู้ทำงานเกี่ยวข้องกับการใช้ผลการวิจัยดังกล่าว

เอกสารอ้างอิง

1. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.