วารสารวิชาการสาธารณสุข
ปีที่ 30 ฉบับที่ 4 กรกฎาคม - สิงหาคม 2564

Journal of Health Science
Vol. 30 No. 4, July - August 2021

มุมสถิติ

Statistics Corner

ข้อควรระวังในการแปลผล p value (1)

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

บทความวิจัยที่ตีพิมพ์ในปัจจุบันยังคงพบเห็นการแปลผล p value ที่ไม่ถูกต้องอยู่เสมอ การสรุปว่าต่างกันอย่างมีนัยสำคัญจาก p value มีการสรุปไม่ถูกต้องหลายลักษณะ นักวิจัย หรือผู้อ่านไม่เข้าใจความหมายของ p value ไม่ทราบว่าขนาดของ p value ที่คำนวณได้แปรตามขนาดของความต่าง หรือขนาดความสัมพันธ์ และขนาดตัวอย่าง ถ้านำเฉพาะ p value มาแปลความหมายของความต่างที่พบเพียงอย่างเดียว อาจทำให้เกิดการสรุปที่ผิดพลาด บทความนี้จะอธิบายความไม่ถูกต้องของการแปลผล p value ในลักษณะต่าง ๆ

เริ่มต้นจากการเข้าใจความหมาย p value ที่ถูกต้อง ในการทดสอบสมมุติฐานผลลัพธ์การทดลองโดยกำหนดค่า α = 0.05 เมื่อพบว่า p value = 0.031 ผู้วิจัยสรุปผลว่า “การทดสอบสมมุติฐาน พบว่า กลุ่มทดลองต่างจากกลุ่มควบคุมอย่างมีนัยสำคัญ โดยมีโอกาสสรุปผิด 3.1%” ข้อความดังกล่าวเป็นการสรุปที่ไม่ถูกต้อง

p value เป็นค่าความน่าจะเป็นที่เกิดจากความผิดพลาดของการสุ่ม (sampling error) เมื่อปฏิเสธสมมุติฐาน (null hypothesis) p value คำนวณจากการแจกแจงความน่าจะเป็นของค่าเฉลี่ยของตัวอย่าง ซึ่งเป็นการแจกแจงข้อมูลทฤษฎีที่สร้างจากตัวอย่างที่ทำซ้ำกันหลาย ๆ ครั้ง ดังนั้น p value ที่คำนวณได้ 0.031 จะสรุปว่า ถ้ามีการทดลองสุ่มตัวอย่างซ้ำหลาย ๆ ครั้งจะมีจำนวนตัวอย่างร้อยละ 3.1 ของการปฏิเสธสมมุติฐานอาจเกิดจากความผิดพลาดของการสุ่ม หรือเกิดจากความบังเอิญ

ดังนั้นเมื่อพิจารณา p value เทียบกับค่า α แล้วพบว่าปฏิเสธสมมุติฐาน จะสรุปว่าสองกลุ่มมีความแตกต่างอย่างมีนัยสำคัญ ข้อสรุปนี้อาจเป็นข้อสรุปที่ถูกหรือผิด โดย p value คือค่าที่แสดงว่าการสรุปว่าต่างมีโอกาสเป็นข้อสรุปผิดจากความบังเอิญเท่าไร

ส่วนการสรุปที่อธิบายว่า p value = 0.031 แสดงว่ากลุ่มทดลองต่างจากกลุ่มควบคุมโดยมีโอกาสสรุปผิด 3.1% ซึ่งความหมายว่าการทดสอบครั้งนี้ (หนึ่งครั้ง) มีโอกาสที่จะสรุปผิดร้อยละ 3.1 จึงเป็นการสรุปที่ไม่ถูกต้อง

การสรุปที่ถูกต้องคือ “การทดสอบสมมุติฐานพบว่ามีความต่างอย่างมีนัยสำคัญ แสดงว่ากลุ่มทดลองต่างจากกลุ่มควบคุม โดยความต่างที่พบมีโอกาสสรุปผิดจากความบังเอิญ (ความผิดพลาดจากการสุ่ม) ร้อยละ 3.1”

กรณีที่ 2 เมื่อพบว่า p value มากกว่าค่า α จะปฏิเสธสมมุติฐาน แล้วสรุปว่าทั้งสองกลุ่มไม่ต่างกัน อาจเป็นการสรุปที่ผิดเพราะ p value นอกจากแปรตามขนาดความต่างแล้วยังแปรตามขนาดตัวอย่างที่ศึกษาด้วย เช่นในการเปรียบเทียบผลการสอนวิธีการใช้ยาสูดพ่นผู้ป่วยโรคปอดอุดกั้นเรื้อรัง ระหว่างการสอนแบบเดิมเทียบกับการสอนแบบใหม่ มีผลการศึกษาแสดงในตาราง

วิธีสอน	n	ใช้ถูกต้อง	ความแตกต่าง [95% CI]	p value
เดิม	10	3 (30%)	40%[-10.2, 90.2]	0.09
ใหม่	10	7 (70%)		

จากตารางถึงแม้อัตราการใช้ถูกต้องเพิ่มขึ้นจากเดิมร้อยละ 40 (เดิมร้อยละ 30 เพิ่มเป็นร้อยละ 70) แต่ผลการทดสอบสมมุติฐานพบว่าค่า p value = 0.09 ซึ่งจะยอมรับสมมุติฐานว่าการสอนทั้งสองวิธีได้ผลไม่ต่างกันที่ α = 0.05 ถ้าพิจารณาอัตราการใช้ที่ถูกต้องที่เพิ่มขึ้นถึงร้อยละ 40 แต่ข้อสรุปที่ได้จากการทดสอบสมมุติฐานกลับระบุว่าไม่ต่างทำให้ผลการทดสอบสมมุติฐานให้ข้อสรุปที่ไม่สอดคล้องกับความเป็นจริง ที่เป็นเช่นนี้เพราะ p value นอกจากจะแปรตามขนาดความต่างแล้ว ยังแปรตามขนาดตัวอย่างด้วย ถ้าพิจารณาจาก 95% ช่วงเชื่อมั่นของความต่างระหว่างกลุ่ม พบว่า ทั้งสองวิธีต่างกัน 40% โดยความต่างมีค่าอยู่ระหว่าง -10.2 ถึง 90.2 การมีค่าศูนย์อยู่ภายในช่วงเชื่อมั่น จะสรุปว่ายอมรับสมมุติฐานการสอนทั้งสองวิธีได้ผลไม่ต่างกันที่ α = 0.05 เช่นเดียวกับการพิจารณาจากค่า p value เมื่อพิจารณาความกระชับของช่วงประมาณ พบว่าช่วงประมาณกว้างมาก (-10.2% ถึง 90.2%) แสดงว่าขนาดตัวอย่างที่ใช้ศึกษามีขนาดเล็กไม่เพียงพอที่จะระบุความต่างอย่างมีนัยสำคัญ

ถ้าเพิ่มขนาดตัวอย่างเป็นกลุ่มละ 20 คนโดยคงอัตราการใช้ถูกต้องของทั้งสองกลุ่มเท่าเดิมจะได้ p value = 0.01 ซึ่งจะให้ผลสรุปว่ามีความต่างอย่างมีนัยสำคัญ

ดังนั้นในกรณีที่พิจารณาจาก p value แล้วพบว่าต่างอย่างไม่มีนัยสำคัญ ผู้วิจัยควรนำเสนอค่าช่วงเชื่อมั่นด้วยเพื่อใช้พิจารณาว่าขนาดตัวอย่างที่ใช้เพียงพอหรือไม่ ถ้าขนาดตัวอย่างเพียงพอช่วงเชื่อมั่นจะมีความกระชับ ถ้าไม่เพียงพอช่วงเชื่อมั่นจะกว้าง ในกรณีนี้พบว่าช่วงเชื่อมั่นกว้างมากควรสรุปว่า “ผลการทดสอบสมมุติฐานพบว่าวิธีการสอนเดิมต่างจากวิธีการสอนใหม่อย่างไม่มีนัยสำคัญ ถึงแม้จะพบว่าวิธีสอนใหม่ทำให้ผู้เข้ารับการสอนใช้ได้ถูกต้องเพิ่มขึ้นร้อยละ 40 แต่การศึกษานี้ใช้ขนาดตัวอย่างไม่ใหญ่พอ ที่ระบุความต่างอย่างมีนัยสำคัญ”

กรณีที่ 3 การแปลความหมายว่า p value ยิ่งมีค่าน้อยผลการเปรียบเทียบยิ่งต่างกันมาก ข้อสรุปนี้ไม่ถูกต้องเพราะการที่ p value มีค่าน้อยอาจมาจากขนาดตัวอย่างที่ใหญ่ เช่นจากตัวอย่างเรื่องการสอนวิธีใช้ยาสูดพ่นผู้ป่วยโรคปอดอุดกั้นเรื้อรัง อัตราการใช้ถูกต้องของการสอนวิธีเดิมร้อยละ 30 และการสอนวิธีใหม่ร้อยละ 70 เท่าเดิม เมื่อขนาดตัวอย่างเท่ากับ 10 ได้ p value = 0.09 แต่ถ้าขนาดตัวอย่างเพิ่มขึ้นเป็น 20 p value = 0.01 จะเห็น p value มีค่าน้อยมาจากขนาดตัวอย่างที่เพิ่มขึ้น

กรณีที่พบว่า p value มีค่าน้อย ๆ การสรุปว่าต่างอย่างมีนัยสำคัญ โอกาสสรุปผิดจากความบังเอิญจะมีค่าน้อยตามขนาด p value

สรุป

เมื่อสรุปผลการทดสอบสมมุติฐานว่าต่าง ค่า p value คือความน่าจะเป็นของการสรุปผิดจากความบังเอิญ

การพบว่ามีความต่างอย่างไม่มีนัยสำคัญ อาจไม่ใช่ว่าสองกลุ่มไม่แตกต่างกัน

การทดสอบสมมุติฐานที่มี p value น้อย ๆ แสดงว่าการสรุปว่าต่างกันอย่างมีนัยสำคัญ มีโอกาสสรุปผิดจากความบังเอิญยิ่งมีค่าน้อย