

การแปลความหมาย p-value ที่ไม่ถูกต้อง

อรุณ จิรวินกุล

ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

ในการทดสอบสมมติฐานเพื่อเปรียบเทียบผลของ treatment ของงานวิจัยด้านการแพทย์และสาธารณสุข ส่วนใหญ่นิยมสรุปผลโดยใช้ p-value เช่น นักวิจัยต้องการสรุปว่าผู้ป่วยกลุ่มที่ได้รับความรู้เกี่ยวกับขั้นตอนต่าง ๆ ของการผ่าตัด จะช่วยทำให้มีความวิตกกังวลน้อยกว่ากลุ่มควบคุม ถ้าผลการศึกษาที่ทำในต่างประเทศพบว่า การเปรียบเทียบความวิตกกังวลระหว่าง ๒ กลุ่ม มี p-value ๐.๐๕ นักวิจัยสรุปว่าการได้รับความรู้ช่วยให้ผู้ป่วยลดความวิตกกังวลลงได้อย่างมีนัยสำคัญด้วยโอกาสผิดพลาดไม่เกินร้อยละ ๕

ผลการศึกษาที่ทำในประเทศไทยในเรื่องเดียวกันพบว่า มี p-value ๐.๐๐๑ นักวิจัยสรุปว่า การศึกษาในคนไทยพบว่าผู้ป่วยลดความวิตกกังวลมากกว่า (โดยดูจาก p-value ของการศึกษาของไทยมีค่าน้อยกว่าที่พบในต่างประเทศ) ข้อสรุปดังกล่าวเป็นการแปลความหมายของ p-value ที่ไม่ถูกต้อง

การแปลความหมายของ p-value ที่ไม่ถูกต้องที่พบบ่อยมีดังนี้

๑. มีการนำ p-value ไปเป็นค่าความน่าจะเป็นของความผิดพลาดชนิดที่ ๑ (α - error) เช่น ถ้าพบว่า p-value เท่ากับ ๐.๐๓ แล้วสรุปว่าผลการทดสอบปฏิเสธสมมติฐาน (H_0) มีโอกาสผิดพลาดไม่เกินร้อยละ ๓ หรือมีโอกาสสรุปถูกร้อยละ ๙๗ การแปลผลดังกล่าว

ไม่ถูกต้อง เพราะ p-value เป็นความน่าจะเป็นที่คำนวณจากข้อมูลที่ศึกษาจากเงื่อนไขว่าสมมติฐานถูกต้อง และขนาดของ p-value จะมีขนาดเล็กลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ส่วนความผิดพลาดชนิดที่ ๑ เป็นความผิดพลาดที่เกิดจากการตัดสินใจปฏิเสธสมมติฐาน ถ้า p-value มีค่าน้อย ๆ เช่น ๐.๐๐๑ เพียงแต่สรุปได้ว่าการตัดสินใจปฏิเสธสมมติฐานจะมีโอกาสเกิดความผิดพลาดชนิดที่ ๑ น้อย แต่จะระบุว่าความผิดพลาดชนิดที่ ๑ มีค่า ๐.๐๐๑ (เท่ากับ p-value) ไม่ได้

๒. การที่ได้ p-value เล็ก ๆ เช่น ๐.๐๐๑ แล้วสรุปว่ามีโอกาสแตกต่างกันมากกว่า p-value ขนาดใหญ่ เช่น ในตัวอย่างเรื่องการเปรียบเทียบความวิตกกังวลที่ได้กล่าวไว้ในตอนต้น ทั้งนี้เพราะการตัดสินใจปฏิเสธสมมติฐานเพียงแต่บอกว่าต่างกันด้วยระดับความผิดพลาดไม่เกินค่า α ที่ตั้งไว้ ดังนั้นการที่ p-value จะเล็กขนาดไหนก็สรุปได้ว่าต่างโดยไม่สามารถสรุปได้ว่าต่างมากกว่ากัน ถ้าจะดูว่าต่างมากกว่ากันจะต้องดูที่ค่าช่วงเชื่อมั่น OR RR หรือ absolute difference

๓. เมื่อพบว่าค่า p-value มีค่ามากกว่า ๐.๐๕ แล้วสรุปว่าค่าเฉลี่ยหรือค่าสัดส่วนที่เปรียบเทียบเท่ากัน ตัวอย่างเช่น ในการเปรียบเทียบค่าเฉลี่ยความวิตกกังวลในขณะรอผ่าตัดระหว่างหญิง (\bar{x} = ๖๒.๔) กับชาย (\bar{x} = ๔๕.๑) พบว่าค่า p-value = ๐.๐๖๒ ผู้วิจัยสรุปว่า

ผลการทดสอบสมมุติฐานแตกต่างกันอย่างไม่มีนัยสำคัญ จึงสรุปว่าค่าเฉลี่ยความวิตกกังวลในขณะรอผ่าตัดของชายหญิงเท่ากัน ซึ่งเป็นการสรุปที่ไม่ถูกต้อง เพราะค่า p-value จะเปลี่ยนไปตามขนาดตัวอย่าง โดยที่ p-value จะมีขนาดเล็กลงจนพบนัยสำคัญได้ ถ้าเพิ่มตัวอย่างให้ใหญ่พอ ควรสรุปเพียงว่าไม่แตกต่างกัน (ซึ่งหมายถึง อาจมีความต่างกันอยู่แต่ขนาดตัวอย่างไม่มากพอที่จะพิสูจน์ขนาดความแตกต่างนั้นได้ หรืออาจเท่ากันก็ได้)

สรุป

ค่า p มักจะมีการแปลความหมายผิดได้ง่าย ความเข้าใจวิธีการทางสถิติที่ถูกต้องจะช่วยให้การแปลความหมายค่า p ได้อย่างถูกต้อง การสรุปผลการ

เปรียบเทียบโดยใช้ช่วงเชื่อมั่นเป็นอีกวิธีหนึ่งที่จะช่วยลดความผิดพลาดในการแปล และสรุปการเปรียบเทียบ

บรรณานุกรม

๑. อรุณ จีรวัดน์กุล. ชีวิตสถิติที่ใช้ในงานวิจัยทางวิทยาศาสตร์ สุขภาพ. พิมพ์ครั้งที่ ๑. ขอนแก่น : คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น; ๒๕๕๐.
๒. Blume JD, Peipert JF. What your statistician never told you about p value. J Am Assoc Gynecol Laparosc 2003; 10(4) : 439-44.
๓. Sterne JAC. Teaching hypothesis tests - time for significant change. Statist Med 2002; 21 : 985-94.