

ขนาดผลกระทบ: ผลลัพธ์ที่สำคัญที่สุดของการทดสอบสมมติฐาน

EFFECT SIZE: THE MOST IMPORTANT PRODUCT OF HYPOTHESIS TESTING

สำเริง ไกยวงศ์

Samroeng Kaiyawong

คณะบริหารธุรกิจ มหาวิทยาลัยราชภัฏชัยภูมิ

Faculty of Business Administration, Chaiyaphum Rajabhat University

บทคัดย่อ

ในการวิจัยเชิงประจักษ์ด้านธุรกิจ เศรษฐกิจ และด้านอื่นๆ ของสังคมศาสตร์ ผลลัพธ์ที่สำคัญที่สุดของการทดสอบสมมติฐานไม่ใช่ค่า p-value แต่เป็นขนาดผลกระทบ เพราะว่าขนาดผลกระทบทำให้นักวิจัยสามารถสื่อสารการมีนัยสำคัญในทางปฏิบัติของผลการวิจัยและสามารถอธิบายเกี่ยวกับผลกระทบที่จะเกิดขึ้นในโลกแห่งความเป็นจริง ขนาดผลกระทบเป็นสารสนเทศที่มีความสำคัญสำหรับผู้ที่จะนำผลการวิจัยไปสู่การปฏิบัติซึ่งต้องการทราบว่าผลกระทบของตัวแปรอิสระมากกว่าศูนย์หรือไม่และผลกระทบมีขนาดมากเพียงใด การทดสอบนัยสำคัญทางสถิติไม่สามารถประเมินความสำคัญในทางปฏิบัติของผลการวิจัย และไม่ได้บอกสิ่งที่ผู้นำผลการวิจัยไปปฏิบัติต้องการจะรู้ การทดสอบนัยสำคัญทางสถิติจึงเป็นการวัดและรายงานผลการทดสอบสมมติฐานที่มีความหมายในทางทฤษฎีเท่านั้น ดังนั้นเมื่อใดก็ตามที่มีการทดสอบสมมติฐาน นักวิจัยควรวัดและรายงานขนาดผลกระทบในส่วนที่เป็นรายงานผลการวิจัยด้วย ไม่ใช่รายงานเฉพาะค่า p-value เท่านั้น

คำสำคัญ: ขนาดผลกระทบ การมีนัยสำคัญในทางปฏิบัติ โคเฮนส์ดี เอตัสแควร์ โอเมก้าสแควร์

Abstract

Effect sizes, not p-value, are the most important products of hypothesis testing in the empirical studies of business, economics, and other branches of social science because effect sizes allow researchers to communicate the practical significance of the research results and describe the observed effects. Effect sizes are the important information for practitioners who implement the research results and want to know whether each independent variable has an effect greater than zero and how big the effect is. Test of statistical significance cannot evaluate the practical importance of research results and cannot inform what practitioners want to know. Test of statistical significance only measure and report the theoretically meaningful results of hypothesis testing. Therefore, whenever hypothesis testing is performed, not only p-value should

be measured and reported in the research results, but also the effect sizes.

Keywords: Effect size, Practical significance, Cohen's d, Eta-squared, Omega-squared

บทนำ

ในช่วงหลายปีที่ผ่านมา มีวรรณกรรมเชิงวิชาการจำนวนมากที่ได้ตีพิมพ์เผยแพร่เพื่อแนะนำว่าเมื่อทำการทดสอบสมมติฐาน นักวิจัยควรรายงานการมีนัยสำคัญในทางปฏิบัติของผลการทดสอบด้วย ไม่ใช่รายงานเฉพาะการมีนัยสำคัญในทางสถิติเท่านั้น (Tomczak & Tomczak, 2014) เพราะว่าการทดสอบนัยสำคัญทางสถิติไม่สามารถประเมินความสำคัญของผลการทดสอบ (Vacha-Haase & Thompson, 2004) และไม่ได้บอกสิ่งที่ผู้นำผลการวิจัยไปปฏิบัติต้องการจะรู้ (Cohen, 1994) ในการทดสอบสมมติฐานนั้น ดัชนีที่ใช้วัดการมีนัยสำคัญทางสถิติ ได้แก่ p-value ส่วนดัชนีที่ใช้วัดการมีนัยสำคัญทางปฏิบัติ ได้แก่ ขนาดผลกระทบ (Effect size) อย่างไรก็ตาม หนังสือและตำราเรียนด้านสถิติ หนังสือและตำราเรียนด้านการวิจัยที่มีอยู่ในประเทศไทย รวมทั้งบทความวิจัยที่นำเสนอเกี่ยวกับการวัดและการตีความของขนาดผลกระทบยังมีจำนวนเพียงเล็กน้อย เพราะว่าการเรียนการสอนในวิชาสถิติและวิชาการระเบียบวิธีวิจัยเชิงปริมาณ อาจารย์ส่วนมากจะสอนวิธีคำนวณหาสถิติทดสอบ และประเมินการมีนัยสำคัญทางสถิติ ซึ่งเป็นวิธีวัดและรายงานผลการทดสอบสมมติฐานที่มีความหมายในทางทฤษฎีเท่านั้น แต่มีอยู่อย่างหนึ่งที่อาจารย์ส่วนใหญ่ไม่ค่อยสอนก็คือ วิธีการคำนวณและการตีความความหมายขนาดผลกระทบ ซึ่งเป็นวิธีวัดและรายงานผลการทดสอบสมมติฐานที่มีความหมายในโลกแห่งความเป็นจริง และเป็นค่าสถิติที่มีความสำคัญสำหรับบุคคลที่จะนำผลการวิจัยไปสู่การปฏิบัติ ซึ่งต้องการรู้ว่าตัวแปรอิสระมีผลกระทบต่อตัวแปรตามมากเพียงใดอย่างไรก็ตาม นักวิจัยส่วนใหญ่ไม่ค่อยได้แยกแยะระหว่างการมีนัยสำคัญทางสถิติกับการมีนัยสำคัญในทางปฏิบัติของผลการวิจัย ผลการทดสอบสมมติฐานที่พบว่า ไม่มี

นัยสำคัญทางสถิติจึงถูกตีความว่าไม่มีนัยสำคัญในทางปฏิบัติด้วย หรือผลการทดสอบสมมติฐานที่พบว่า มีนัยสำคัญทางสถิติถูกตีความแบบเหมารวมว่ามีนัยสำคัญในทางปฏิบัติด้วย แต่ความจริงไม่ได้เป็นเช่นนั้นเสมอไป ความแตกต่างระหว่างการมีนัยสำคัญในทางสถิติกับการมีนัยสำคัญในทางปฏิบัติสามารถอธิบายได้ด้วยตัวอย่างของ Kirk (1996) ซึ่งเป็นเรื่องราวเกี่ยวกับนักวิจัยคนหนึ่ง ที่เชื่อว่า การรับประทานยาที่ห่อหนึ่งซึ่งผลิตขึ้นมาใหม่สามารถจะทำให้ไอคิวของผู้ป่วยโรคอัลไซเมอร์เพิ่มขึ้น นักวิจัยได้ให้กลุ่มทดลองซึ่งป่วยเป็นโรคอัลไซเมอร์ จำนวน 6 คน รับประทานยาที่ผลิตขึ้นมาใหม่นั้น และให้กลุ่มควบคุมซึ่งป่วยเป็นโรคอัลไซเมอร์เหมือนกัน จำนวน 6 คน รับประทานยาปลอม (Placebo) เมื่อเวลาผ่านไประยะหนึ่งนักวิจัยได้วัดไอคิวของกลุ่มตัวอย่างทั้งสองกลุ่ม แล้วเปรียบเทียบไอคิวเฉลี่ยด้วยสถิติ Independent-samples t-test โดยใช้ระดับนัยสำคัญ .05 ผลการวิจัยพบว่า ไอคิวเฉลี่ยของกลุ่มทดลองสูงกว่าไอคิวเฉลี่ยของกลุ่มควบคุม 13 คะแนน แต่ผลการทดสอบสมมติฐานพบว่า ไอคิวเฉลี่ยไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($t = 1.61, p = .14$) ถ้าเราตัดสินใจบนพื้นฐานของการมีนัยสำคัญทางสถิติเท่านั้น ข้อสรุปของการทดสอบก็คือ ยาที่รับประทานไม่มีประสิทธิผลหรือยาใช้ไม่ได้ผล เมื่อผลการทดสอบสมมติฐานพบว่า ความแตกต่างของไอคิวของกลุ่มทดลองกับกลุ่มควบคุมไม่มีนัยสำคัญทางสถิติ จึงดูเหมือนว่าความแตกต่างของไอคิว 13 คะแนนที่เกิดขึ้นเป็นเพียงความคลาดเคลื่อนหรือความบังเอิญที่เกิดจากวิธีการสุ่มเลือกผู้ป่วยเข้ากลุ่มทดลองและกลุ่มควบคุม การไม่มีนัยสำคัญทางสถิติไม่ได้หมายความว่าไม่มีความแตกต่างระหว่างไอคิวของกลุ่มทดลองกับกลุ่มควบคุม จะเห็นได้ว่าการที่กลุ่มทดลองมีไอคิวแตกต่างจากกลุ่มควบคุมจำนวน 13 คะแนนนั้นเป็นความแตกต่าง

ที่เรียกว่ามาก (Substantive) ซึ่งเท่ากับว่ายาที่กลุ่มทดลองรับประทานได้ผลจริง การเพิ่มขึ้นของไอคิวจำนวน 13 คะแนน จึงถือว่ามากเพียงพอที่ทำให้กลุ่มทดลองมีความฉลาดและคุณภาพชีวิตดีขึ้นในระดับที่น่าพึงพอใจ และมากเพียงพอที่จูงใจให้มีการพิสูจน์เพิ่มเติมโดยการทดลองกับกลุ่มตัวอย่างที่มีขนาดใหญ่ขึ้น กรณีตัวอย่างดังกล่าวนี้ได้ชี้ให้เห็นว่า นักวิจัยไม่ควรตัดสินผลการวิจัยโดยพิจารณาจากการมีนัยสำคัญทางสถิติเท่านั้น แต่ควรตัดสินผลการวิจัยโดยพิจารณาจากขนาดผลกระทบด้วย ดังนั้น เพื่อให้ นักวิจัยทั้งเก่าและใหม่มีความรู้ความเข้าใจเกี่ยวกับขนาดผลกระทบและเพื่อแนะนำให้นักวิจัยมีการคำนวณ ตีความหมาย และรายงานขนาดผลกระทบในงานผลการวิจัย ผู้เขียนจึงได้ศึกษาและสรุปสาระสำคัญเป็น 4 หัวข้อ คือ (1) ความหมายของขนาดผลกระทบ (2) ความสำคัญของขนาดผลกระทบ (3) ประเภทของดัชนีวัดขนาดผลกระทบ และ (4) วิธีวัดและตีความหมายขนาดผลกระทบเมื่อทดสอบสมมติฐานด้วย t-test ผลจากการศึกษาจะนำเสนอตามลำดับหัวข้อดังนี้

1. ความหมายของขนาดผลกระทบ

จากการทบทวนวรรณกรรมเกี่ยวกับความหมายของขนาดผลกระทบ (Effect size) นักสถิติและนักวิจัยได้ให้ความหมายของคำว่า “ขนาดผลกระทบ” ไว้ดังนี้

Cohen (1988) กล่าวว่า ขนาดผลกระทบ หมายถึง ขนาดของผลลัพธ์ที่เกิดขึ้นในประชากร หรือหมายถึงระดับการเป็นเท็จของสมมติฐานว่าง

Kline (2004) กล่าวว่า ขนาดผลกระทบ หมายถึง ขนาดผลกระทบ (Impact) ของการแทรกแซง (Intervention) ที่มีต่อผลลัพธ์ของการทดลอง (Outcome)

Olejnik & Algina (2003) กล่าวว่า ขนาดผลกระทบ หมายถึง ดัชนีที่อยู่ในรูปของค่ามาตรฐานซึ่งใช้ในการประมาณค่าพารามิเตอร์ที่ไม่ได้ขึ้นอยู่กับขนาดตัวอย่าง โดยทำให้ทราบขนาดของความแตกต่างระหว่างค่าเฉลี่ยของประชากร หรือทำให้ทราบขนาดความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม

Livinge & Hullet (2002) กล่าวว่า ขนาดผลกระทบ หมายถึง ดัชนีที่บอกให้ทราบว่าตัวแปร 2 ตัวหรือมากกว่า 2 ตัวมีความสัมพันธ์กันมากเพียงใดหรือมีความแตกต่างระหว่างกลุ่มมากเพียงใด

Kotrlik & Williams (2003) กล่าวว่า ขนาดผลกระทบ หมายถึง ขนาดของผลต่างระหว่างค่าเฉลี่ยที่อยู่ในรูปของค่ามาตรฐาน หรือขนาดของความสัมพันธ์ที่อยู่ในรูปของค่ามาตรฐาน

ดังนั้น จึงสรุปได้ว่า ขนาดผลกระทบ หมายถึง ขนาดของผลต่างระหว่างค่าเฉลี่ย ขนาดของผลต่างระหว่างค่าสัดส่วน หรือขนาดความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม ซึ่งอยู่ในรูปค่ามาตรฐานและบอกให้ทราบขนาดผลกระทบในทางปฏิบัติที่จะเกิดขึ้นในโลกแห่งความเป็นจริง

2. ความสำคัญของขนาดผลกระทบ

Cohen (1990) กล่าวว่า ผลลัพธ์ที่สำคัญที่สุดที่ได้จากการทดสอบสมมติฐานไม่ใช่ค่า p-value แต่เป็นขนาดผลกระทบ (Effect size) ซึ่งนักวิจัยควรจะวัดและรายงานอย่างน้อย 1 ค่า เช่นเดียวกัน Lakens (2013) กล่าวว่า ขนาดผลกระทบเป็นผลลัพธ์ที่สำคัญที่สุดของการวิจัยเชิงประจักษ์ นักวิจัยโดยทั่วไปย่อมต้องการทราบว่า การเข้าไปแทรกแซงหรือการทดลองมีผลกระทบมากกว่าศูนย์หรือไม่ และผลกระทบมีมากน้อยเพียงใด อาจกล่าวได้ว่านับเป็นเวลากว่า 4 ทศวรรษแล้วที่นักสถิติได้ริเริ่มให้คำแนะนำแก่นักวิจัยด้านพฤติกรรมศาสตร์ว่า ในการรายงานผลการทดสอบสมมติฐานนั้นนอกจากจะรายงานค่า p-value แล้ว นักวิจัยควรจะรายงานดัชนีวัดขนาดผลกระทบด้วย การวัดขนาดของความแตกต่างระหว่างกลุ่มได้เริ่มต้นขึ้นประมาณปี ค.ศ. 1962 โดย Jacob Cohen ซึ่งเป็นนักสถิติและนักจิตวิทยาชาวอเมริกัน จึงกล่าวได้ว่า Cohen เป็นบุคคลสำคัญที่ได้วางรากฐานเกี่ยวกับการศึกษาขนาดผลกระทบอันเนื่องมาจากความแตกต่างของค่าเฉลี่ย ดัชนีวัดขนาดผลกระทบที่ Cohen ได้พัฒนาขึ้น ได้แก่ Cohen's d และ Cohen's f^2

ส่วนการวัดขนาดของความสัมพันธ์ระหว่างตัวแปรได้เริ่มต้นขึ้นประมาณปี ค.ศ. 1901 โดยนักสถิติซึ่งมีชื่อว่า Karl Pearson ดัชนีวัดขนาดผลกระทบของ Pearson ที่นักสถิติและนักวิจัยทั่วโลกรู้จักกันดี ได้แก่ ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) (Kirk, 1996) อย่างไรก็ตาม แม้ว่าการวัดและการรายงานขนาดผลกระทบจะไม่ใช่วิธีใหม่ แต่นักวิจัยส่วนมากก็ยังไม่รู้จักดัชนีวัดขนาดผลกระทบ ดังนั้น เมื่อทดสอบสมมติฐานโดยใช้ Z-test, One-sample t-test, Independent-samples t-test, Paired-samples t-test, ANOVA และ Chi-square เป็นต้น จึงไม่ได้วัดและไม่ได้รายงานดัชนีวัดขนาดผลกระทบในรายงานผลการวิจัย

สมาคมจิตวิทยาแห่งอเมริกัน (American Psychological Association หรือ APA) เป็นองค์กรหนึ่งที่มีบทบาทสำคัญในการวางต้นแบบหลายๆ เรื่องเกี่ยวกับการวิจัย เช่น ระบบการอ้างอิงแบบ APA และรูปแบบการเขียนรายงานผลการวิจัย องค์กรดังกล่าวได้เริ่มต้นรณรงค์ส่งเสริมสนับสนุนให้นักวิจัยรายงานขนาดผลกระทบโดยเขียนไว้ในหนังสือ Publication Manual of the American Psychological Association พิมพ์ครั้งที่ 5 ซึ่งสมาคมจิตวิทยาแห่งอเมริกาได้กล่าวถึงขนาดผลกระทบว่า “เป็นเรื่องที่สำคัญและจำเป็นเสมอที่นักวิจัยจะต้องรายงานดัชนีวัดขนาดผลกระทบหรือขนาดของความสัมพันธ์ในรายงานผลการวิจัย หลักการทั่วไปของการรายงานขนาดผลกระทบก็คือ นอกจากจะช่วยให้ผู้อ่านได้ทราบสารสนเทศเกี่ยวกับการมีนัยสำคัญทางสถิติแล้วยังช่วยให้ผู้อ่านได้รับสารสนเทศที่เพียงพอสำหรับใช้ในการประเมินขนาดของผลกระทบหรือขนาดของความสัมพันธ์ด้วย” (APA, 2001) ต่อมาปี ค.ศ. 2010 สมาคมจิตวิทยาแห่งอเมริกาได้ตีพิมพ์คู่มือดังกล่าวเป็นครั้งที่ 6 ซึ่งมีใจความตอนหนึ่งได้กล่าวถึงความสำคัญของขนาดผลกระทบว่า “เพื่อให้ผู้อ่านได้ทราบขนาดหรือความสำคัญของผลการวิจัย เป็นเรื่องจำเป็นเสมอที่ผู้วิจัยจะต้องเขียนรายงานดัชนีวัดขนาดผลกระทบไว้ในเนื้อหาส่วนที่เป็นผลการวิจัยด้วย และถ้าเป็นไปได้ควรจะรายงาน

ช่วงความเชื่อมั่นของขนาดผลกระทบเพื่อแสดงให้เห็นถึงความถูกต้องแม่นยำของขนาดผลกระทบที่คำนวณได้” (APA, 2010) นอกจากนี้ สมาคมวิจัยทางการศึกษาแห่งอเมริกา (American Educational Research Association หรือ AERA) ก็ได้ให้คำแนะนำไว้ว่า “ในการเขียนรายงานผลการทดสอบทางสถิติ นักวิจัยควรจะรายงานขนาดผลกระทบและดัชนีวัดความไม่แน่นอนของขนาดผลกระทบด้วย เช่น ค่าคลาดเคลื่อนมาตรฐานหรือช่วงความเชื่อมั่น พร้อมทั้งมีการตีความหมายของขนาดผลกระทบในเชิงคุณภาพโดยการอธิบายความหมายของผลกระทบเพื่อตอบคำถามวิจัย” (AERA, 2006) ถึงแม้ว่าองค์กรต่างๆ ดังกล่าวและบรรณาธิการของวารสารวิชาการด้านจิตวิทยาในอเมริกาจะส่งเสริมสนับสนุนให้นักวิจัยเขียนรายงานขนาดผลกระทบมาเป็นเวลากว่า 15 ปีแล้วก็ตาม แต่จะเห็นได้ว่าบทความวิจัยส่วนใหญ่ที่ตีพิมพ์เผยแพร่ในวารสารวิชาการไม่ได้วัดและรายงานขนาดผลกระทบ ดังจะเห็นได้จากผลการตรวจสอบซึ่งพบว่า มีบทความวิจัยเชิงปริมาณเพียง 148 เรื่องจากทั้งหมด 508 เรื่องที่มีการกล่าวถึงหรือรายงานขนาดผลกระทบ ซึ่งคิดเป็นร้อยละ 29 เท่านั้น (McMillan et al., 2002) จากการตรวจสอบวารสารวิชาการด้านจิตวิทยาที่มีความน่าเชื่อถือพบว่า ประเมินร้อยละ 50 ของบทความวิจัยที่ตีพิมพ์เผยแพร่ในช่วงปี 2009-2010 มีการรายงานผลการทดสอบสมมติฐานด้วย ANOVA แต่ไม่มีการรายงานขนาดผลกระทบใดๆ เลย และมีเพียง 1 ใน 4 ของบทความวิจัยที่รายงานผลการทดสอบสมมติฐานด้วย t-test แล้วมีการรายงานขนาดผลกระทบด้วย (Tomczak & Tomczak, 2014) ผลจากการสำรวจของ Morris & Fritz ในปี ค.ศ. 2011 พบว่า 2 ใน 5 ของบทความวิจัยที่ตีพิมพ์เผยแพร่ในปี ค.ศ. 2009 มีการเขียนรายงานขนาดผลกระทบ (Fritz, Morris & Richler, 2012) จึงเห็นได้ว่า นักวิจัยส่วนใหญ่ยังขาดความรู้ความเข้าใจเกี่ยวกับขนาดผลกระทบ เพราะว่าอาจารย์ในมหาวิทยาลัยส่วนมากไม่ได้สอนผู้เรียนเกี่ยวกับขนาดผลกระทบ ทำให้บัณฑิต มหาบัณฑิต และดุษฎีบัณฑิต กลายเป็นนักวิจัย

ที่ไม่มีความรู้ความเข้าใจเกี่ยวกับดัชนีวัดขนาดผลกระทบ นอกจากนี้การรณรงค์ให้นักวิจัยในประเทศไทยเกิดความรู้ความเข้าใจเกี่ยวกับดัชนีวัดขนาดผลกระทบยังไม่แพร่หลายเท่าที่ควร บทความวิชาการเรื่องนี้จึงเป็นจุดเริ่มต้นในเรื่องดังกล่าว

ในการตีความหมายผลการทดสอบสมมติฐานนั้น สิ่งที่นักวิจัยโดยทั่วไปไม่มีความเข้าใจผิดมาโดยตลอดมี 2 ประการคือ (1) การเข้าใจผิดว่าผลการทดสอบสมมติฐานที่ได้ค่า p-value ต่ำกว่า หมายถึง ผลกระทบของการทดลองมีขนาดมากกว่าผลการทดสอบสมมติฐานที่ได้ค่า p-value สูงกว่า และ (2) การเข้าใจผิดว่าการมีนัยสำคัญในทางสถิติ หมายถึง การมีนัยสำคัญในทางปฏิบัติ (Gliner, Leech & Morgan, 2002) ในกรณีแรก นักวิจัยส่วนมากจะตีความว่า p-value = .01 หมายถึง ผลกระทบจากการทดลองมากกว่าเมื่อเปรียบเทียบกับ p-value = .05 ความจริงก็คือ p-value หมายถึง ความน่าจะเป็นที่จะได้ค่าสถิติทดสอบเท่ากับหรือแตกต่าง (มากกว่าหรือน้อยกว่า) จากค่าสถิติทดสอบที่คำนวณได้จากกลุ่มตัวอย่างเมื่อสมมติฐานว่างเป็นจริง (Berenson, Levine & Krehbiel, 2012) ค่า p-value คือ ความน่าจะเป็นที่จะเกิดจากความบังเอิญจากการสุ่มเลือกตัวอย่างแล้วได้ค่าสถิติที่สนับสนุนสมมติฐานทางเลือก (Kotrlac & Williams, 2003) แต่จะเห็นได้ว่า p-value ไม่ได้บอกให้ทราบสารสนเทศใดๆ เกี่ยวกับขนาดของผลกระทบ โดยทั่วไปค่า p-value ที่ใช้ทดสอบการมีนัยสำคัญทางสถิตินั้นจะขึ้นอยู่กับค่าคลาดเคลื่อนมาตรฐาน (standard error) กล่าวคือ เมื่อขนาดตัวอย่างเพิ่มขึ้นจะทำให้ค่าคลาดเคลื่อนมาตรฐานลดลง และมีผลทำให้ค่า p-value ลดลง ผลการทดสอบจึงมีนัยสำคัญทางสถิติ บางครั้งการมีนัยสำคัญทางสถิติเกิดขึ้นเพราะว่าขนาดตัวอย่างมีจำนวนมาก อนึ่งในการทดสอบสมมติฐานนั้น แม้ว่าผลการทดสอบไม่มีนัยสำคัญทางสถิติ แต่อาจจะมีความสำคัญในทางปฏิบัติก็ได้ เช่น ในการเปรียบเทียบผลผลิตเฉลี่ยต่อไร่ระหว่างข้าวพันธุ์ A และผลผลิตเฉลี่ยต่อไร่ของข้าวพันธุ์ B ผู้วิจัยได้ทดสอบสมมติฐานแบบสองทิศทาง

(Two-tailed test) โดยใช้ $\alpha = .05$ เนื่องจากค่า p-value = .06 ดังนั้น ผู้วิจัยจึงไม่ปฏิเสธสมมติฐานว่างและสรุปผลการทดสอบว่า ไม่มีความแตกต่างอย่างมีนัยสำคัญระหว่างผลผลิตเฉลี่ยต่อไร่ของข้าวพันธุ์ A และข้าวพันธุ์ B เพียงเพราะเห็นว่า p-value = .06 มีค่ามากกว่า $\alpha = .05$ เพียง .01 เท่านั้น ในกรณีนี้ถ้า p-value = $\alpha = .05$ นักวิจัยก็อาจจะตัดสินใจปฏิเสธสมมติฐานว่างได้ และถ้านักวิจัยเปลี่ยนจากการทดสอบแบบสองทิศทาง (Two-tailed test) เป็นการทดสอบแบบทิศทางเดียว (One-tailed test) สมมติฐานว่างย่อมถูกปฏิเสธ เพราะว่า p-value สำหรับการทดสอบแบบทิศทางเดียวมีค่าเท่ากับ .03 เท่านั้น

ขนาดผลกระทบมีความสำคัญด้วยเหตุผล 3 ประการคือ (1) ทำให้นักวิจัยสามารถนำเสนอขนาดผลกระทบในรูปของค่ามาตรฐานซึ่งสามารถเข้าใจได้โดยไม่ต้องคำนึงถึงมาตรวัดที่ใช้วัดตัวแปรตาม ขนาดผลกระทบดังกล่าวทำให้นักวิจัยสามารถสื่อสารเกี่ยวกับการมีนัยสำคัญทางปฏิบัติ (Practical significance) ของผลการวิจัยแทนที่จะสื่อสารการมีนัยสำคัญทางสถิติ (Statistical significance) เพียงอย่างเดียว (2) ทำให้นักวิจัยสามารถเปรียบเทียบขนาดผลกระทบของงานวิจัยเรื่องหนึ่งกับงานวิจัยเรื่องอื่นๆ ที่ใช้ตัวแปรเดียวกัน และ (3) ขนาดผลกระทบที่ค้นพบจากงานวิจัยในอดีตสามารถนำมาใช้ในการวางแผนสำหรับงานวิจัยเรื่องใหม่ได้ (Lakens, 2013) การรายงานขนาดผลกระทบมีประโยชน์ดังนี้ (1) ขนาดผลกระทบทำให้ทราบขนาดของความสัมพันธ์ระหว่างตัวแปรและความสำคัญของความสัมพันธ์ที่ทำการประเมินนั้น ซึ่งครอบคลุมถึงความสัมพันธ์ในงานวิจัยที่มีจุดมุ่งหมายเพื่อตรวจสอบความสัมพันธ์และงานวิจัยที่มีจุดมุ่งหมายเพื่อประเมินขนาดของผลต่าง (2) ขนาดผลกระทบทำให้ผู้วิจัยสามารถเปรียบเทียบขนาดผลกระทบของตัวแปรอิสระตัวเดียวกันในงานวิจัยเรื่องหนึ่งกับงานวิจัยอีกเรื่องหนึ่ง (3) ขนาดผลกระทบสามารถนำไปคำนวณหาอำนาจทดสอบและขนาดตัวอย่างที่จำเป็นต้องใช้ในการวิจัย และ (4) ขนาดผลกระทบที่ได้จากการ

ศึกษานำร่องสามารถใช้ในการคาดคะเนผลการวิจัยที่จะเกิดขึ้นในอนาคตได้ (Tomczak & Tomczak, 2014)

โดยทั่วไปหากผลการทดสอบสมมติฐานไม่มีนัยสำคัญทางสถิติ นักวิจัยส่วนใหญ่จะตัดสินว่างานวิจัยนั้นไม่มีคุณค่า มีความเป็นไปได้ว่าผลการทดสอบที่ไม่มีนัยสำคัญทางสถิตินั้นอาจจะมีความสำคัญในทางปฏิบัติก็ได้ เมื่อเป็นเช่นนี้ ขนาดผลกระทบย่อมทำให้ผลการวิจัยที่ถูกตัดสินว่าไม่มีคุณค่ากลายเป็นงานวิจัยที่มีคุณค่าได้ ดังนั้น ไม่ว่าผลการทดสอบจะมีนัยสำคัญทางสถิติหรือไม่ นักวิจัยก็ควร

จะวัด ตีความหมาย และรายงานขนาดผลกระทบด้วย

3. ประเภทของดัชนีวัดขนาดผลกระทบ

ในอดีตที่ผ่านมาได้มีการนำเสนอวิธีการวัดขนาดผลกระทบหลายวิธี แต่วิธีที่คุ้นเคยกันมากที่สุดมี 3 วิธี ได้แก่ Cohen's d, Pearson's correlation coefficient (r) และ Odds Ratio (RR) ผู้ที่ศึกษาสถิติและนักวิจัยส่วนมากจะคุ้นเคยกับ Pearson's correlation coefficient ซึ่งเป็นดัชนีที่ใช้วัดขนาดความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณ 2 ตัวแปร และเกณฑ์ในการตีความหมายดังนี้

ตารางที่ 1 เกณฑ์ในการตีความหมายค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

ค่าสัมประสิทธิ์สหสัมพันธ์ (r)	ความหมาย	ร้อยละของ ความแปรปรวนร่วม (r^2)
.90-1.00	สูงมาก (Very high correlation)	ร้อยละ 81-100
.70-.90	สูง (High correlation)	ร้อยละ 49-81
.50-.70	ปานกลาง (Moderate correlation)	ร้อยละ 25-49
.30-.50	ต่ำ (Low correlation)	ร้อยละ 9-25
.00-.30	เล็กน้อย (Little correlation)	ร้อยละ 0-9

ที่มา: Kotrlík & Williams (2003)

หลักการที่สำคัญอีกอย่างหนึ่งของการวัดขนาดความสัมพันธ์ระหว่าง 2 ตัวแปรก็คือ การวัดค่าสัดส่วนของความแปรปรวนร่วม (Proportion of Shared Variance หรือ POV) ดัชนี POV มีค่าเทียบเท่ากับค่าสัมประสิทธิ์สหสัมพันธ์ยกกำลังสอง (r^2) ซึ่งเรียกว่าค่าสัมประสิทธิ์แห่งการกำหนด (Coefficient of determination) ถ้าตัวแปร X และตัวแปร Y มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -0.85 แล้วค่าสัมประสิทธิ์แห่งการกำหนดจะมีค่าเท่ากับ $-0.85 \times -0.85 = .72$ นั่นคือ $POV = r^2 = .72$ หรือ 72% ดังนั้น จึงตีความได้ว่า 72% ของความแปรปรวนทั้งหมดเป็นความแปรปรวนร่วมระหว่างตัวแปร X กับตัวแปร Y หากพิจารณาในเชิงเหตุ

และผลโดยกำหนดให้ X เป็นตัวแปรอิสระ และ Y เป็นตัวแปรตาม สามารถตีความได้ว่า 72% ของความแปรปรวนในตัวแปร Y อธิบายได้ด้วยความแปรปรวนของตัวแปร X กล่าวอีกนัยหนึ่ง ความแปรปรวนของตัวแปร Y ประมาณ 72% เกิดจากอิทธิพลของตัวแปร X ดัชนี POV มีค่าอยู่ระหว่าง 0 และ 1 อย่างไรก็ตาม ถ้า $POV = r^2 = 1$ นักสถิติจะถือว่าตัวแปร X และ Y เป็นตัวแปรเดียวกัน ดังนั้น ในการวัดขนาดความสัมพันธ์ระหว่างตัวแปร X และ Y เราจึงไม่คาดหวังที่จะเห็นดัชนี POV และ r^2 มีค่าเท่ากับ 1

ในกรณีที่ตัวแปรตามมีเพียง 1 ตัว แต่ตัวแปรอิสระมีมากกว่า 1 ตัว เราสามารถวัดขนาดผลกระทบของ

กลุ่มตัวแปรอิสระได้โดยการคำนวณหาค่าสัมประสิทธิ์แห่งการกำหนดพหุคูณ (Coefficient of multiple determination หรือ R^2 ซึ่งคำนวณได้เมื่อวิเคราะห์ข้อมูลด้วยสถิติ multiple regression analysis แต่ข้อจำกัดของดัชนี R^2 ก็คือ มีค่าสูงเกินจริง ซึ่งเกิดจากความคลาดเคลื่อนในการสุ่มเลือกกลุ่มตัวอย่าง (Sampling error) ได้แก่ ขนาดตัวอย่างและจำนวนตัวแปรอิสระที่มีอยู่ในสมการถดถอยพหุคูณ ซึ่งถือว่าเป็นแหล่งความแปรปรวนภายนอก (extraneous variation) เราสามารถปรับลดความแปรปรวนของตัวแปรตามที่เกิดจากความแปรปรวนภายนอกได้โดยการคำนวณหา Adjusted coefficient of multiple determination หรือ Adjusted R^2 จะเห็นได้ว่าดัชนี R^2 และ Adjusted R^2 เป็นค่าสถิติที่ใช้กันโดยทั่วไปในวารสารวิชาการด้านบริหารธุรกิจ และเป็นผลการวิเคราะห์ที่พบได้เสมอในการวิเคราะห์เศรษฐกิจ ดัชนีที่ใช้กันโดยทั่วไปในวารสารวิชาการด้านจิตวิทยา เรียกว่า Correlation ratio หรือ Eta-squared ซึ่งเขียนแทนด้วยสัญลักษณ์ η^2 เป็นดัชนีที่เกี่ยวข้องกับการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA) ดัชนี η^2 สะท้อนถึงสัดส่วนความแปรปรวนของตัวแปรตามที่เกิดขึ้นจากการเป็นสมาชิกกลุ่มในตัวแปรอิสระ อย่างไรก็ตาม η^2 เป็นดัชนีที่มีจุดอ่อนเหมือนกันกับ R^2 คือ มีความเอนเอียง กล่าวคือ ขนาดผลกระทบที่คำนวณได้จากดัชนี η^2 และ R^2 มีค่าสูงเกินความเป็นจริงด้วยเหตุนี้ ดัชนีวัดขนาดผลกระทบ 2 ดัชนี ได้แก่ Epsilon-squared (ϵ^2) และ Omega-squared (ω^2) จึงได้รับการพัฒนา ขึ้นโดยมีการปรับความเอนเอียงให้ลดน้อยลง (Snyder & Lawson, 1993) ดังนั้น ดัชนีทั้งสองนี้จึงเป็นดัชนีที่มีความเอนเอียงน้อยที่สุดและมีความน่าเชื่อถือมากที่สุด

โดยทั่วไป นักสถิติมีความสนใจในการเปรียบเทียบหรือวัดความสัมพันธ์ ดังนั้น ดัชนีวัดขนาดผลกระทบจึงสามารถจำแนกได้เป็น 2 กลุ่ม (Two families) คือ ดัชนีที่ใช้วัดความแตกต่างระหว่างกลุ่มซึ่งเรียกว่า d family และดัชนีที่ใช้วัดความสัมพันธ์ระหว่างตัวแปร ซึ่งเรียกว่า

r family (Ellis, 2010) อักษร d ย่อมาจากคำว่า Difference ซึ่งแปลว่า ความแตกต่างหรือผลต่าง ส่วนอักษร r ย่อมาจาก Relationship ซึ่งแปลว่า ความสัมพันธ์ นักสถิติคนอื่นๆ เช่น Rosenthal (1994), Kirk (1996) และ Richardson (1996) ก็ได้จำแนกดัชนีวัดขนาดผลกระทบออกเป็น 2 กลุ่ม ได้แก่ (1) การวัดผลต่างของค่าเฉลี่ยที่อยู่ในรูปค่ามาตรฐาน (standardized mean difference) ซึ่งเรียกว่า d family หรือ Difference family และ (2) การวัดค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ซึ่งเรียกว่า r family หรือ Correlation family โดยทำให้ทราบสัดส่วนความแปรปรวนของตัวแปรตามที่เกิดจากอิทธิพลของตัวแปรอิสระหลักและการแนวคิดเกี่ยวกับดัชนีวัดขนาดผลกระทบทั้ง 2 กลุ่มมีดังนี้ (Ellis, 2010)

กลุ่มที่ 1 ดัชนีที่ใช้ประเมินขนาดความแตกต่างระหว่างกลุ่ม (d family) ในการวิจัยเชิงทดลองนักวิจัยนิยมเปรียบเทียบระหว่างกลุ่มทดลอง (Experimental group) กับกลุ่มควบคุม (Control group) โดยที่กลุ่มทดลอง หมายถึง กลุ่มตัวอย่างที่ถูกแทรกแซงโดยนักวิจัย หรือได้รับการบำบัด (Treatment) กลุ่มทดลองอาจจะมีเพียง 1 กลุ่มหรือมากกว่า 1 กลุ่มก็ได้ ส่วนกลุ่มควบคุม หมายถึง กลุ่มตัวอย่างที่ไม่ได้ถูกแทรกแซง และไม่ได้รับการบำบัด แต่ถูกปล่อยให้ไปเป็นไปตามสภาพธรรมชาติ ดังนั้น การเปรียบเทียบระหว่างกลุ่มทดลองกับกลุ่มควบคุมจึงเป็นการวัดขนาดผลกระทบที่เกิดจากการแทรกแซง การกระทำ หรือการบำบัดนั่นเอง การเปรียบเทียบระหว่างกลุ่มอาจจะเป็นการเปรียบเทียบค่าเฉลี่ยที่แปลงให้อยู่ในรูปของค่ามาตรฐาน เช่น ดัชนี Cohen's d ซึ่งคำนวณได้จากการหารผลต่างระหว่างค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่ม ($\bar{x}_1 - \bar{x}_2$) ด้วยค่าเบี่ยงเบนมาตรฐานร่วม (Pooled standard deviation) หรือเป็นการเปรียบเทียบค่าความน่าจะเป็น (ร้อยละ) เช่น การเปรียบเทียบความน่าจะเป็นที่จะป่วยเป็นโรคมะเร็งปอดของกลุ่มบุคคลที่สูบบุหรี่ (p) กับความน่าจะเป็นที่จะป่วยเป็นโรคมะเร็งปอดของกลุ่มบุคคลที่ไม่ได้สูบบุหรี่ (q) การเปรียบเทียบ

ความน่าจะเป็นหรือโอกาสที่จะป่วยเป็นโรคมะเร็งปอดของบุคคลทั้ง 2 กลุ่ม สามารถคำนวณได้ 3 วิธี ดังนี้ (1) คำนวณหาผลต่างระหว่างความน่าจะเป็นทั้ง 2 ค่า โดยใช้สูตร $p - q$ (2) คำนวณหา Risk Ratio หรือ RR จากสูตร p/q และ (3) คำนวณหา Odds Ratio หรือ OR จากสูตร $[p/(1-p)] / [q/(1-q)]$ ค่าที่คำนวณได้คือ ดัชนีวัดขนาดผลกระทบซึ่งทำให้ทราบว่ากลุ่มบุคคลที่สูบบุหรี่มีโอกาสที่จะป่วยเป็นโรคมะเร็งปอดมากกว่ากลุ่มบุคคลที่ไม่ได้สูบบุหรี่ประมาณกี่เท่า

กลุ่มที่ 2 ดัชนีที่ใช้ประเมินขนาดความสัมพันธ์ระหว่างตัวแปร (r family) ได้แก่ ดัชนีที่ใช้วัดขนาดความสัมพันธ์ระหว่างตัวแปร 2 ตัวหรือมากกว่า 2 ตัว ดัชนีส่วนมากวัดขนาดผลกระทบด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (r) ซึ่งทำให้ทราบขนาดและทิศทางความสัมพันธ์ระหว่างตัวแปร 2 ตัว แต่บางคนอาจไม่ทราบว่าค่าสัมประสิทธิ์สหสัมพันธ์เป็นดัชนีวัดขนาดผลกระทบ ดัชนี Cohen's d ก็ถือว่าเป็นค่าสัมประสิทธิ์สหสัมพันธ์ที่อยู่ในรูปของค่ามาตรฐาน ขนาดของความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร เช่น X และ Y เมื่อวัดด้วยค่าสัมประสิทธิ์สหสัมพันธ์ย่อมสามารถนำไปเปรียบเทียบกับขนาดของความสัมพันธ์ระหว่าง 2 ตัวแปรใดๆ ที่วัดด้วยค่าสัมประสิทธิ์สหสัมพันธ์เหมือนกัน

Vacha-Haase & Thompson (2004) ได้จำแนกวิธีวัดขนาดผลกระทบออกเป็น 3 กลุ่ม ได้แก่ (1) ขนาดผลกระทบของผลต่างที่อยู่ในรูปค่ามาตรฐาน (standardized differences effect sizes) (2) ขนาดผลกระทบที่คำนวณจากค่าความแปรปรวน (variance-accounted-for effect sizes) และ (3) ขนาดผลกระทบที่มีการปรับลดความเอนเอียง (corrected effect sizes)

ส่วน Ferguson (2009) ได้จำแนกวิธีวัดขนาดผลกระทบออกเป็น 4 กลุ่ม ได้แก่ (1) ดัชนีวัดความแตกต่างระหว่างกลุ่ม (Group difference indices) ซึ่งเป็นการประมาณขนาดผลต่างระหว่างค่าเฉลี่ยของประชากร

2 กลุ่มหรือมากกว่า เช่น ดัชนี Cohen's d (2) ดัชนีวัดขนาดความสัมพันธ์ (Strength of association indices) ซึ่งเป็นการประมาณค่าสัดส่วนความแปรปรวนร่วมระหว่าง 2 ตัวแปรหรือมากกว่า 2 ตัวแปร เช่น ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน R^2 และ η^2 (3) ดัชนีวัดขนาดผลกระทบที่มีการปรับลดความเอนเอียง (Corrected estimates) เช่น Adjusted R^2 , ϵ^2 และ ω^2 และ (4) ดัชนีวัดขนาดความเสี่ยง (Risk estimates) ซึ่งเป็นการเปรียบเทียบความเสี่ยงที่จะเกิดผลลัพธ์อย่างหนึ่งกับประชากร 2 กลุ่มหรือมากกว่า

4. วิธีวัดและตีความหมายขนาดผลกระทบเมื่อทดสอบสมมติฐานด้วย t-test

สถิติทดสอบคือ t-test ซึ่งประกอบด้วย One-sample t-test, Independent-samples t-test และ Paired-samples t-test เป็นสถิติเชิงอนุมานที่นักวิจัยนำมาใช้บ่อยมากในทดสอบสมมติฐานเพื่อเปรียบเทียบค่าเฉลี่ย และนักวิจัยส่วนใหญ่ใช้ค่า p-value เป็นสารสนเทศในการตัดสินใจว่าผลการทดสอบมีนัยสำคัญทางสถิติหรือไม่ มีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานว่างหรือไม่ โดยมีการรายงานค่า p-value ในส่วนที่เป็นผลการวิจัยด้วย แต่เนื่องจากนักวิจัยส่วนใหญ่ไม่มีความรู้ความเข้าใจเกี่ยวกับขนาดผลกระทบ ดังนั้น จึงปรากฏว่ารายงานผลการวิจัยส่วนใหญ่ที่ตีพิมพ์เผยแพร่ในวารสารวิชาการทั้งระดับชาติและนานาชาติไม่มีการวัดตีความหมาย และรายงานขนาดผลกระทบในส่วนที่เป็นผลการวิจัย ความจริงแล้ว ไม่ว่าผลการทดสอบสมมติฐานจะมีนัยสำคัญทางสถิติหรือไม่ก็ตาม นักวิจัยก็ควรที่จะวัดและรายงานขนาดผลกระทบด้วย เมื่อทดสอบสมมติฐานด้วย t-test ดัชนีที่นิยมใช้วัดขนาดผลกระทบมี 3 ดัชนี ได้แก่ Cohen's d, Eta-squared (η^2) และ Omega-squared (ω^2) โดยมีเกณฑ์ในการตีความหมาย ดังนี้

ตารางที่ 2 ดัชนีวัดขนาดผลกระทบและเกณฑ์ในการตีความหมาย เมื่อทดสอบสมมติฐานด้วย t-test

ความหมายของ ขนาดผลกระทบ	Cohen's d Hedges's g	η^2	ω^2
น้อย (Small)	$d < .20$	$.01 < \eta^2 < .09$	$.01 < \omega^2 < .09$
ปานกลาง (Medium)	$.20 < d < .80$	$.10 < \eta^2 < .25$	$.10 < \omega^2 < .25$
มาก (Large)	$d > .80$	$\eta^2 > .25$	$\omega^2 > .25$

ที่มา: Privitera (2015)

ดัชนี Cohen's d อาจจะมีค่ามากกว่า 1 ก็ได้ ส่วนดัชนี η^2 และ ω^2 มีค่าสูงสุดเท่ากับ 1 นอกจากนี้ Hentschke & Stuttgen (2011) ได้กำหนดเกณฑ์ไว้ว่า ถ้าดัชนี η^2 มีค่าไม่เกิน .10 หมายถึง มีผลกระทบในระดับน้อย (Small) ถ้าดัชนี η^2 มีค่าตั้งแต่ .11 ถึง .29 หมายถึง มีผลกระทบในระดับปานกลาง (Medium) และถ้าดัชนี η^2 มีค่าตั้งแต่ .30 ขึ้นไป หมายถึง มีผลกระทบในระดับมาก (Large) ซึ่งเกณฑ์ดังกล่าวนี้จะใช้ในการตีความหมายสำหรับดัชนี ω^2 ด้วย

ในการทดสอบสมมติฐานด้วย One-sample t-test และ Paired-samples t-test ขนาดผลกระทบสามารถวัดได้โดยใช้ดัชนี Cohen's d ซึ่งมีสูตรคำนวณดังนี้ (Lakens, 2013; Tomczak & Tomczak, 2014)

$$d = \frac{\bar{D}}{\sqrt{\frac{\sum (D - \bar{D})^2}{n - 1}}}$$

ถ้าเป็นการทดสอบสมมติฐานโดยใช้สถิติ One-sample t-test ค่าของ D คือ ผลต่างระหว่างค่าสังเกตกับค่าที่ตั้งไว้ในสมมติฐาน $\bar{D}$ คือ ค่าเฉลี่ยของผลต่าง และ n คือ ขนาดตัวอย่าง ถ้าเป็นการทดสอบสมมติฐานโดยใช้ Paired-samples t-test ค่าของ D คือ ผลต่างระหว่างค่าสังเกตที่เป็นคู่กัน $\bar{D}$ คือ ค่าเฉลี่ยของผลต่าง และ n คือ จำนวนคู่ของผลต่าง

นอกจากนี้ เรายังสามารถคำนวณหาดัชนี Cohen's d ได้โดยตรงจากค่าสถิติ t และขนาดตัวอย่าง (จำนวน

คู่ของผลต่าง) โดยมีสูตรในการคำนวณดังนี้ (Lakens, 2013)

$$d = \frac{t}{\sqrt{n}}$$

ถ้าทดสอบสมมติฐานด้วย Independent-samples t-test ขนาดผลกระทบสามารถวัดได้ โดยใช้ดัชนี Cohen's d ซึ่งมีสูตรในการคำนวณ ดังนี้ (Lakens, 2013)

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

$$\text{หรือ } d = t \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

โดยที่ $\bar{x}_1$ และ $\bar{x}_2$ คือ ค่าเฉลี่ยของกลุ่มตัวอย่างที่ 1 และค่าเฉลี่ยของกลุ่มตัวอย่างที่ 2 ตามลำดับ n_1 และ n_2 คือ ขนาดตัวอย่างกลุ่มที่ 1 และขนาดตัวอย่างกลุ่มที่ 2 ตามลำดับ s_1^2 และ s_2^2 คือ ค่าความแปรปรวนของกลุ่มตัวอย่างที่ 1 และค่าความแปรปรวนของกลุ่มตัวอย่างที่ 2 ตามลำดับ และ t คือ Independent-samples t-test แต่เนื่องจากดัชนี Cohen's d คำนวณจากค่าเฉลี่ยของกลุ่มตัวอย่างจึงมีความเอนเอียงในการคำนวณขนาดผลกระทบของประชากร โดยเฉพาะในกรณีที่ขนาดตัวอย่างน้อยกว่า 20 ดังนั้น ดัชนี Hedges's g จึงได้รับการพัฒนาขึ้นโดยมีการปรับดัชนี Cohen's d

ให้สะท้อนความเป็นจริงและไม่มีความเอนเอียง โดยมีสูตรในการคำนวณดังนี้ (Lakens, 2013)

$$g = d \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

นอกจากดัชนี Cohen's d และ Hedges's g แล้ว ยังมีดัชนีที่สำคัญอีก 2 ดัชนีที่ได้รับความนิยมอย่างแพร่หลายในการวัดขนาดผลกระทบของตัวแปรอิสระเมื่อทดสอบสมมติฐานด้วย t-test ได้แก่ η^2 และ ω^2 อย่างไรก็ตาม ดัชนี η^2 ยังมีความเอนเอียง (Biased) เพราะการคำนวณดัชนี η^2 จะใช้ค่าความแปรปรวนที่ได้จากกลุ่มตัวอย่างและยังไม่ได้ปรับลดให้สะท้อนความเป็นจริง ค่าของดัชนี η^2 ที่คำนวณได้จึงมีค่าสูงเกินความจริง ดังนั้น การวัดขนาดผลกระทบโดยใช้ดัชนี ω^2 จึงมีความเหมาะสมมากกว่าดัชนี η^2 เพราะว่า ω^2 มีการปรับค่าให้สอดคล้องกับความเป็นจริงและเป็นดัชนีที่ใช้วัดขนาดผลกระทบของประชากร ดัชนี η^2 และ ω^2 ที่คำนวณได้สามารถตีความหมายได้เหมือนกันกับค่าสัมประสิทธิ์แห่งการกำหนด (Coefficient of determination) หรือ R^2 ใน Multiple regression สูตรที่ใช้ในการคำนวณดัชนี η^2 และ ω^2 มีดังนี้ (Privitera, 2015)

$$\eta^2 = \frac{t^2}{t^2 + df} \quad \text{และ} \quad \omega^2 = \frac{t^2 - 1}{t^2 + df}$$

จะเห็นได้ว่าดัชนี η^2 และดัชนี ω^2 สามารถคำนวณได้โดยตรงจากค่า t ที่คำนวณได้ และองศาความเป็นอิสระ สถิติทดสอบคือ One-sample t-test และ Paired-samples t-test มีองศาความเป็นอิสระเท่ากับ $df = n - 1$ ส่วน Independent-samples t-test มีองศาความเป็นอิสระเท่ากับ $df = n_1 + n_2 - 2$

ดัชนี η^2 มีค่าเทียบเท่ากับ r^2 ใน Simple regression และ R^2 ใน Multiple regression เราสามารถแปลง η^2 ให้เป็นค่า r โดยการใช้สูตรต่อไปนี้ (Field, 2009)

$$r = \eta = \sqrt{\frac{t^2}{t^2 + df}}$$

ดัชนี r มีชื่อเรียกว่า Point-biserial correlation coefficient ในการตีความขนาดผลกระทบโดยใช้ค่าสัมประสิทธิ์ Point biserial correlation coefficient นักสถิติทั้งสองท่าน ได้แก่ McGrath and Meyer ได้กำหนดเกณฑ์มาตรฐานไว้ดังนี้ ถ้า r มีค่าไม่เกิน .10 หมายถึง มีผลกระทบในระดับน้อย ถ้า r มีค่าตั้งแต่ .24 ถึง .36 หมายถึง มีผลกระทบในระดับปานกลาง และถ้า r มีค่าตั้งแต่ .37 ขึ้นไป หมายถึง มีผลกระทบในระดับมาก (Fritz, Morris & Richler, 2012)

บทสรุป

ผลจากการศึกษาสามารถสรุปได้ว่า ขนาดผลกระทบหมายถึง ดัชนีที่ใช้วัดขนาดผลต่างระหว่างค่าเฉลี่ยขนาดผลต่างระหว่างค่าสัดส่วน และขนาดความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามซึ่งอยู่ในรูปค่ามาตรฐานในการทดสอบสมมติฐานนั้น ผลการทดสอบจะมีนัยสำคัญทางสถิติหรือไม่ก็ตาม นักสถิติควรจะวัดและรายงานขนาดผลกระทบด้วย เพราะว่าผลการทดสอบที่ไม่มีนัยสำคัญทางสถิติอาจมีนัยสำคัญในทางปฏิบัติก็ได้ และผลการทดสอบที่มีนัยสำคัญทางสถิติอาจไม่มีนัยสำคัญทางปฏิบัติก็ได้ ดัชนีวัดขนาดผลกระทบเป็นค่าสถิติที่มีประโยชน์มากในทางปฏิบัติ ซึ่งทำให้สามารถตัดสินใจได้ว่าถ้านำผลการวิจัยไปสู่การปฏิบัติแล้วจะทำให้เกิดผลกระทบมากเพียงพอหรือไม่ ดัชนีที่นิยมใช้วัดขนาดผลกระทบเมื่อทดสอบสมมติฐานด้วยสถิติ t-test ได้แก่ Cohen's d, η^2 และ ω^2 การวัดขนาดผลกระทบโดยใช้ดัชนี ω^2 ได้รับการตัดสินว่ามีความเหมาะสมมากที่สุดเมื่อเปรียบเทียบกับดัชนี Cohen's d และ η^2 เพราะว่าดัชนี ω^2 มีการปรับค่าให้สอดคล้องกับความเป็นจริง ไม่มีความเอนเอียง และสามารถใช้วัดขนาดผลกระทบของประชากร

References

- American Educational Research Association. (2006). Standards for reporting on empirical social American science research in AERA publications. *Educational Researcher*, 35(6), 33-40.
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Berenson, M. L., Levine, D. M. & Krehbiel, T. C. (2012). *Basic business statistics: concepts and applications* (12th ed.). USA: Prentice Hall.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1990). Things I have learned (so far). *American psychologist*, 45, 1304-1312.
- Cohen, J. (1994). The earth is round ($p < .05$). *American psychologist*, 49, 997-1003.
- Ellis, P. D. (2010). *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*. United Kingdom: Cambridge University Press.
- Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532-538.
- Field, A. (2009). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publication.
- Fritz, C. O, Morris, P. E. & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2-18.
- Gliner, J. A., Leech, N. L. & Morgan, G. A. (2002). Problems with null hypothesis significance testing (NHST): What do the textbooks say? *The Journal of Experimental Education*, 71(1), 83-92.
- Hentschke, H. & Stuttgen, M. C. (2011). Computation of measures of effect size for neuroscience data sets. *European Journal of Neuroscience*, 34, 1887-1894.
- Kirk, R. E. (1996). Practical significance: a concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759.
- Kline, R. B. (2004). *Beyond significance testing: reforming data analysis methods in behavioral research*. Washington DC: American Psychological Association.
- Kotrlik, J. W. & Williams, H. A. (2003). The incorporation of effect size in information technology, learning, and performance research. *Information Technology, Learning, and Performance Journal*, 21(1), 1-7.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1-12.
- Levine, T. R. & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, 28(4), 612-625.

- McMillan, J. H., Lawson, S., Lewis, K. & Snyder, A. (2002). Reporting effect size: the road less traveled. *Paper presented at the 2002 annual meeting of the American Educational Research Association*, New Orleans.
- Olejnik, S. & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434-447.
- Privitera, G. J. (2015). *Statistics for the behavioral sciences* (2nd ed.). USA: Sage Publication.
- Richardson, J. T. (1996). Measures of effect size. *Behavioral Research Methods, Instruments & Computers*, 28, 12-22.
- Rosenthal, R. (1994). *Parametric measures of effect size*. In H. Cooper & L. V. Hedges (Eds.), *The handbook for research synthesis* (pp. 231–244). NY: Russell Sage Foundation.
- Rosnow, R. L. & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57(3), 221-237.
- Snyder, P. & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education*, 61(4), 334-349.
- Tomczak, M. & Tomczak, E. (2014). The need to report effect size estimates revisited: an overview of some recommended measures of effect size. *TRENDS in Sport Sciences*, 1(21), 19-25.
- Vacha-Haase, T. & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473-481.



Name and Surname: Samroeng Kaiyawong

Highest Education: Doctor of Management in Business Management, Sripatum University

University or Agency: Chaiyaphum Rajabhat University

Field of Expertise: Statistics, Economics, Business Management

Address: 167 Moo 1, Chaiyaphum-Tadtone Rd., Nafai, Mueang, Chaiyaphum 36000