9.1 ความนำ

ในการดำเนินงานวิจัยทางวิทยาศาสตร์สุขภาพ การกำหนดขนาดตัวอย่าง ถือเป็นขั้นตอนสำคัญของการออกแบบงานวิจัย เนื่องจากขนาดตัวอย่างที่เหมาะสม จะส่งผลโดยตรงต่อความสามารถของการวิเคราะห์ข้อมูลทางสถิติในการบ่งชี้ความแตกต่าง หรือ ผลกระทบของวิธีการ (treatment effect) ที่เกิดขึ้นจากการเปรียบเทียบระหว่างกลุ่ม หากขนาดตัวอย่างมีไม่เพียงพอ อาจทำให้อำนาจการทดสอบทางสถิติมีค่าต่ำและไม่สามารถบ่งชี้ความแตกต่างระหว่างกลุ่มที่มีอยู่จริงได้และส่งผลทำให้ข้อสรุปเกี่ยวกับประสิทธิภาพ หรือ ประสิทธิผลของสิ่งแทรกแซง คลาดเคลื่อนไปจากความเป็นจริงได้ ในทางตรงกันข้าม หากขนาดตัวอย่างมีมากเกินความจำเป็น อาจทำให้เกิดการใช้ทรัพยากรอย่างไม่เหมาะสม ทั้งในด้านเวลา งบประมาณและจำนวนผู้เข้าร่วมการวิจัย ดังนั้นการกำหนดขนาดตัวอย่างที่เหมาะสมจึงเป็นองค์ประกอบสำคัญในการออกแบบงานวิจัยเชิงทดลอง ซึ่งโดยทั่วไป การคำนวณขนาดตัวอย่างมีความสัมพันธ์กับองค์ประกอบสำคัญหลายประการ ได้แก่ ระดับนัยสำคัญทางสถิติ อำนาจการทดสอบ ขนาดผลกระทบและความแปรปรวนของข้อมูล ซึ่งองค์ประกอบเหล่านี้ ล้วนมีบทบาทต่อความสามารถของงานวิจัยในการบ่งชี้ความแตกต่างระหว่างกลุ่มได้อย่างถูกต้องและน่าเชื่อถือ โดยเฉพาะในบริบทของการวิเคราะห์ข้อมูลด้วยวิธี ANCOVA ภายใต้แบบแผนงานวิจัยสำหรับข้อมูลก่อน–หลังแบบวัดซ้ำสองกลุ่ม ทั้งงานวิจัยเชิงทดลองแบบสุ่มและแบบกึ่งทดลอง ซึ่งการนำตัวแปรพื้นฐาน หรือ ตัวแปรร่วมมาใช้ปรับค่าในการวิเคราะห์ข้อมูล จะสามารถช่วยลดความผันแปรของตัวแปรผลลัพธ์และเพิ่มประสิทธิภาพของการประมาณค่าผลกระทบของวิธีการได้และส่งผลทำให้อำนาจการทดสอบทางสถิติมีแนวโน้มเพิ่มสูงขึ้น เมื่อเปรียบเทียบกับการวิเคราะห์ที่ไม่ปรับค่าด้วยตัวแปรร่วม ***(Vickers and Altman 2001; Van Breukelen 2006)*** แต่อย่างไรก็ตาม แม้ว่าวิธี ANCOVA จะช่วยเพิ่มประสิทธิภาพของการวิเคราะห์ข้อมูลแต่ผลลัพธ์ที่ได้ยังคงขึ้นกับขนาดตัวอย่างที่นำมาใช้ ซึ่งหากขนาดตัวอย่างมีจำนวนจำกัด การวิเคราะห์ข้อมูลอาจไม่สามารถบ่งชี้ความแตกต่างที่แท้จริงของผลกระทบจากสิ่งแทรกแซงได้ แม้ว่า ค่าประมาณของผลกระทบจะแสดงแนวโน้มของความแตกต่างในเชิงทิศทางได้ก็ตามและปัญหานี้พบได้ค่อนข้างบ่อยในงานวิจัยทางคลินิก โดยเฉพาะเมื่อขนาดผลกระทบของวิธีการมีขนาดเล็ก หรือ เมื่อข้อมูลมีความผันแปรสูงมาก ***(Zhang, Paul et al. 2014)*** ประเด็นดังกล่าว สามารถสะท้อนให้เห็นได้จากกรณีศึกษาในบทก่อนหน้านี้ ซึ่งพบว่า แม้การใช้วิธี ANCOVA จะสามารถให้ค่าประมาณผลกระทบของวิธีการได้ แต่การทดสอบสมมติฐานทางสถิติ อาจยังไม่สามารถบ่งชี้ความแตกต่างระหว่างกลุ่มได้อย่างมีนัยสำคัญทางสถิติ ทั้งนี้อาจเกิดจากหลายสาเหตุ อาทิ ความผันแปรของข้อมูลที่สูง ขนาดผลกระทบที่ต่ำ หรือ ขนาดตัวอย่างที่ไม่เพียงพอ เป็นต้น ซึ่งจากสาเหตุดังกล่าว ล้วนส่งผลโดยตรงต่ออำนาจการทดสอบของการวิเคราะห์ข้อมูล ***(O'Connell, Dai et al. 2017)*** ดังนั้นในการนำวิธี ANCOVA มาใช้เพื่อให้ได้ข้อสรุปที่สามารถสะท้อนผลกระทบจากสิ่งแทรกแซงได้อย่างเหมาะสม การกำหนดขนาดตัวอย่างตั้งแต่ขั้นตอนการออกแบบงานวิจัย จึงมีความสำคัญอย่างยิ่ง เพราะจะช่วยเพิ่มความน่าเชื่อถือของผลลัพธ์ ลดความเสี่ยงของการเกิดข้อผิดพลาดชนิดที่สอง (Type II error) และเพิ่มโอกาสที่งานวิจัย จะสามารถบ่งชี้ผลกระทบที่แท้จริงของวิธีการ หรือสิ่งแทรกแซงได้อย่างมีประสิทธิภาพ ***(Shieh 2020)*** ด้วยเหตุนี้ บทนี้จึงมุ่งนำเสนอแนวคิดและแนวทางในการคำนวณขนาดตัวอย่างและอำนาจการทดสอบสำหรับวิธี ANCOVA ภายใต้แบบแผนงานวิจัยสำหรับข้อมูลก่อน–หลังแบบวัดซ้ำสองกลุ่ม โดยครอบคลุมเนื้อหาสำคัญ ได้แก่ แนวคิดพื้นฐานเกี่ยวกับการคำนวณขนาดตัวอย่างและอำนาจการทดสอบ การคำนวณขนาดตัวอย่างด้วยสูตรทางสถิติสำหรับวิธี ANCOVA และการคำนวณขนาดตัวอย่างด้วยโปรแกรมสำเร็จรูป ได้แก่ โปรแกรม STATA และโปรแกรม G*Power รวมถึงบทสรุปและเอกสารอ้างอิง

9.2 แนวคิดพื้นฐานเกี่ยวกับการคำนวณขนาดตัวอย่างและอำนาจการทดสอบ

ดังที่กล่าวไปข้างต้น การกำหนดขนาดตัวอย่าง ถือเป็นขั้นตอนสำคัญในการออกแบบการวิจัย เนื่องจากขนาดตัวอย่างที่เหมาะสม จะช่วยเพิ่มความสามารถของการวิเคราะห์ข้อมูลในการตรวจพบความแตกต่างของผลลัพธ์ระหว่างกลุ่ม แต่อย่างไรก็ตาม การคำนวณขนาดตัวอย่าง ไม่ได้เป็นเพียงกระบวนการเชิงเทคนิคที่อาศัยเฉพาะสูตรทางคณิตศาสตร์เท่านั้น หากแต่ยังมีพื้นฐานมาจากแนวคิดของการอนุมานทางสถิติ (statistical inference) และการทดสอบสมมติฐานทางสถิติ (statistical hypothesis testing) ซึ่งนำมาใช้เพื่อประเมินว่า “***ความแตกต่างที่พบจากข้อมูลตัวอย่างนั้น สามารถสะท้อนความแตกต่างที่แท้จริงในระดับประชากรได้ หรือไม่***” ภายใต้กรอบของการทดสอบสมมติฐานทางสถิติ นักวิจัยมักเริ่มต้นจากการกำหนด ***สมมติฐานหลัก*** (null hypothesis, H_0) ซึ่งโดยทั่วไปหมายถึง “*การไม่มีความแตกต่างระหว่างกลุ่ม*” และ ***สมมติฐานทางเลือก*** (alternative hypothesis, H_A) ซึ่งหมายถึง “*การมีความแตกต่างระหว่างกลุ่ม*” จากนั้นจึงใช้ข้อมูลตัวอย่างในการประเมินว่า หลักฐานที่ได้มีความเพียงพอที่จะปฏิเสธสมมติฐานหลัก หรือไม่ แต่อย่างไรก็ตาม การตัดสินใจจากข้อมูลตัวอย่างดังกล่าวย่อมมีโอกาสเกิดความคลาดเคลื่อนได้ เนื่องจากข้อมูลดังกล่าวเป็นเพียงตัวแทนส่วนหนึ่งของประชากรเท่านั้น ด้วยเหตุนี้ ในบริบทของการทดสอบสมมติฐานทางสถิติ ความคลาดเคลื่อนในการตัดสินใจ จึงถูกพิจารณาจำแนกออกได้เป็นสองลักษณะ ได้แก่ ความผิดพลาดชนิดที่หนึ่ง (Type I error) และความผิดพลาดชนิดที่สอง (Type II error) ซึ่งมีความสัมพันธ์โดยตรงกับระดับนัยสำคัญทางสถิติ (significance level; α) และอำนาจการทดสอบทางสถิติ (statistical power, $1-\beta$) ตามลำดับ ดังนั้นในการกำหนดขนาดตัวอย่าง นักวิจัยจึงจำเป็นต้องพิจารณาความสมดุลระหว่างองค์ประกอบเหล่านี้ เนื่องจากขนาดตัวอย่างที่ไม่เพียงพอ อาจทำให้การศึกษามีอำนาจการทดสอบต่ำและไม่สามารถตรวจพบความแตกต่างที่มีอยู่จริงระหว่างกลุ่มได้ *(Serdar, Cihan et al. 2021; Jansen, Groenwold et al. 2024)*

นอกจากระดับนัยสำคัญทางสถิติและอำนาจการทดสอบแล้ว ขนาดผลกระทบ (effect size) ยังถือเป็นอีกองค์ประกอบหนึ่งที่สำคัญในการคำนวณขนาดตัวอย่าง เนื่องจากถือเป็นตัวสะท้อนขนาดของผลกระทบที่คาดว่าจะเกิดขึ้นจากวิธีการ หรือ สิ่งแทรกแซงที่นำมาศึกษา ซึ่งหากขนาดผลกระทบมีขนาดเล็ก อาจจำเป็นต้องใช้ขนาดตัวอย่างมากขึ้น เพื่อให้สามารถตรวจพบความแตกต่างดังกล่าวได้ ในทางตรงกันข้าม หากขนาดผลกระทบมีขนาดใหญ่ งานวิจัยอาจใช้ขนาดตัวอย่างที่น้อยลงได้ *(Serdar, Cihan et al. 2021)* สำหรับการวิเคราะห์ข้อมูลด้วยวิธีการวิเคราะห์ความแปรปรวนร่วม หรือ ANCOVA ซึ่งถูกนำมาใช้อย่างแพร่หลายในงานวิจัยสำหรับข้อมูลก่อน–หลังแบบวัดซ้ำสองกลุ่ม การนำตัวแปรร่วม (covariate) มาปรับค่าในการวิเคราะห์ข้อมูลสามารถช่วยลดความผันแปรของตัวแปรผลลัพธ์และเพิ่มประสิทธิภาพของการวิเคราะห์ข้อมูลได้ จึงส่งผลทำให้งานวิจัยมีอำนาจการทดสอบเพิ่มขึ้น เมื่อเปรียบเทียบกับการวิเคราะห์ที่ไม่ปรับค่าด้วยตัวแปรร่วม *(Vickers and Altman 2001; Van Breukelen 2006)* จากเหตุดังกล่าว ในการวางแผนขนาดตัวอย่างสำหรับงานวิจัยที่ใช้วิธี ANCOVA จึงจำเป็นต้องคำนึงถึงบทบาทของตัวแปรร่วมร่วมกับองค์ประกอบพื้นฐานของการทดสอบสมมติฐานทางสถิติด้วย ดังนั้นเพื่อให้นักวิจัยสามารถเข้าใจหลักการที่เกี่ยวข้องกับการกำหนดขนาดตัวอย่างสำหรับการวิเคราะห์ข้อมูลด้วยวิธี ANCOVA ได้อย่างชัดเจน ในหัวข้อนี้ จึงได้นำเสนอแนวคิดพื้นฐานที่สำคัญ เพื่อเป็นพื้นฐานสำหรับการคำนวณขนาดตัวอย่างด้วยสูตรทางสถิติและการใช้โปรแกรมสำเร็จรูปทางสถิติ โดยจำแนกตามประเด็นที่เกี่ยวข้องได้ดังนี้

9.2.1 ความผิดพลาดในการทดสอบสมมติฐานทางสถิติ
9.2.2 อำนาจการทดสอบ
9.2.3 ขนาดผลกระทบ
9.2.4 บทบาทของตัวแปรร่วมในการเพิ่มประสิทธิภาพของงานวิจัย

9.2.1 ความผิดพลาดในการทดสอบสมมติฐานทางสถิติ

ในการทดสอบสมมติฐานทางสถิติ การตัดสินใจเกี่ยวกับความแตกต่างของผลลัพธ์ระหว่างกลุ่มจากข้อมูลตัวอย่าง มีโอกาสเกิดความคลาดเคลื่อนได้ เนื่องจากข้อมูลตัวอย่างเป็นเพียงตัวแทนบางส่วนของประชากรเท่านั้น และอาจมีความแปรปรวนที่เกิดขึ้นจากความบังเอิญ (random variation) ดังนั้นแม้ว่า นักวิจัยจะใช้วิธีการทางสถิติที่เหมาะสมในการวิเคราะห์ข้อมูลแล้วก็ตาม แต่ผลการตัดสินใจจากข้อมูลตัวอย่าง ยังอาจนำไปสู่ความผิดพลาดได้ ***(Serdar, Cihan et al. 2021)*** ภายใต้กรอบของการทดสอบสมมติฐานทางสถิติ ความผิดพลาดดังกล่าว จึงสามารถจำแนกออกได้เป็นสองลักษณะ ได้แก่ ความผิดพลาดชนิดที่หนึ่ง (Type I error) และความผิดพลาดชนิดที่สอง (Type II error) ซึ่งเป็นแนวคิดพื้นฐานที่เกี่ยวข้องโดยตรงกับการกำหนดขนาดตัวอย่างและอำนาจการทดสอบในงานวิจัย ***(Ranganathan, Deo et al. 2024)*** โดยมีรายละเอียดจำแนก ได้ดังนี้

(1) ความผิดพลาดชนิดที่หนึ่ง (Type I error)
(2) ความผิดพลาดชนิดที่สอง (Type II error)
(3) ความสัมพันธ์ระหว่าง Type II error และอำนาจการทดสอบ

(1) ความผิดพลาดชนิดที่หนึ่ง (Type I error)

ความผิดพลาดชนิดที่หนึ่ง หมายถึง ความผิดพลาดที่เกิดขึ้นเมื่อผลการวิเคราะห์ข้อมูลนำไปสู่ ***"การปฏิเสธสมมติฐานหลัก (H_0) ทั้งที่ในความเป็นจริงสมมติฐานหลักนั้นเป็นจริง"*** หรือ กล่าวอีกนัยหนึ่ง คือ การสรุปว่า *"มีความแตกต่างระหว่างกลุ่ม"* ทั้งที่ในระดับประชากรจริง ๆ แล้ว ไม่มีความแตกต่างดังกล่าวอยู่ ซึ่งความน่าจะเป็นของการเกิดความผิดพลาดชนิดนี้ สามารถแสดงได้ด้วยค่า α (alpha) ซึ่งเรียกว่า *"ระดับนัยสำคัญทางสถิติ (significance level)"* โดยในงานวิจัยทางวิทยาศาสตร์สุขภาพ ส่วนใหญ่มักกำหนดค่า α เท่ากับ 0.05 โดยหมายความว่า นักวิจัยยอมรับโอกาสในการเกิดความผิดพลาดชนิดที่หนึ่งได้ร้อยละ 5 หรือ กล่าวได้ว่า ***"มีโอกาสประมาณ 5% ที่จะสรุปว่า มีความแตกต่างระหว่างกลุ่ม ทั้งที่ในระดับประชากรไม่มีความแตกต่างนั้นอยู่จริง"*** ***(Ranganathan, Deo et al. 2024; Zrineh, Al-Usta et al. 2026)***

(2) ความผิดพลาดชนิดที่สอง (Type II error)

ความผิดพลาดชนิดที่สอง หมายถึง ความผิดพลาดที่เกิดขึ้นเมื่อผลการวิเคราะห์ข้อมูลนำไปสู่ ***"การไม่สามารถปฏิเสธสมมติฐานหลัก (H_0) ทั้งที่ในความเป็นจริงสมมติฐานหลักนั้นไม่เป็นจริง"*** หรือ สามารถกล่าวอีกนัยหนึ่ง คือ การสรุปว่า *"ไม่มีความแตกต่างระหว่างกลุ่ม"* ทั้งที่ในระดับประชากรมีความแตกต่างนั้นอยู่จริง ซึ่งความน่าจะเป็นของการเกิดความผิดพลาดชนิดนี้ สามารถแสดงได้ด้วยค่า β (beta) โดยในทางปฏิบัติค่า β นี้มักถูกกำหนดไว้ที่ 0.20 หรือ 0.10 ซึ่งสอดคล้องกับอำนาจการทดสอบทางสถิติ $(1-\beta)$ เท่ากับ 80% หรือ

90% ตามลำดับ ทั้งนี้การกำหนดค่า β ให้ต่ำลง จะช่วยลดความเสี่ยงของการไม่สามารถตรวจพบความแตกต่างที่มีอยู่จริงระหว่างกลุ่มได้ *(Zrineh, Al-Usta et al. 2026)*

(3) ความสัมพันธ์ระหว่าง Type II error กับอำนาจการทดสอบ

อำนาจการทดสอบทางสถิติ (statistical power) หมายถึง *"**ความน่าจะเป็นที่การทดสอบสมมติฐาน จะสามารถตรวจพบความแตกต่างที่แท้จริงระหว่างกลุ่มได้ เมื่อความแตกต่างนั้นมีอยู่จริงในระดับประชากร**" (Serdar, Cihan et al. 2021)* โดยสามารถเขียนความสัมพันธ์ระหว่างอำนาจการทดสอบและความผิดพลาดชนิดที่สอง ได้ดังนี้

$$\text{power} = 1-\beta$$

จากความสัมพันธ์ดังกล่าว สามารถพิจารณาได้ว่า หากค่า β มีค่าน้อย อำนาจการทดสอบของงานวิจัยจะมีค่าสูง ซึ่งหมายความว่า งานวิจัยนั้นมีความสามารถมากขึ้นในการตรวจพบผลกระทบของสิ่งแทรกแซงที่เกิดขึ้นจริงในระดับประชากร *(Zrineh, Al-Usta et al. 2026)* แต่อย่างไรก็ตาม หนึ่งในปัจจัยสำคัญที่ส่งผลต่อความผิดพลาดชนิดที่สอง นั่นคือ ขนาดตัวอย่างของงานวิจัย กล่าวคือ หากขนาดตัวอย่าง มีจำนวนไม่เพียงพอ งานวิจัยจะมีอำนาจการทดสอบต่ำและมีความเสี่ยงสูงที่จะไม่สามารถตรวจพบความแตกต่างที่มีอยู่จริงระหว่างกลุ่มในระดับประชากรได้ *(Darling 2022)*

ดังนั้น ในการออกแบบการวิจัย นักวิจัยจึงจำเป็นต้องกำหนดขนาดตัวอย่างที่เหมาะสม โดยคำนึงถึงความเสี่ยงของความผิดพลาดทั้งสองชนิด เพื่อช่วยเพิ่มโอกาสและความสามารถของงานวิจัยในการตรวจพบผลกระทบของวิธีการ หรือ สิ่งแทรกแซงที่นำมาศึกษาได้อย่างถูกต้องและน่าเชื่อถือมากขึ้น

9.2.2 อำนาจการทดสอบ

ดังที่กล่าวไว้ในหัวข้อก่อนหน้า ความผิดพลาดชนิดที่สอง (Type II error, β) มีความสัมพันธ์โดยตรงกับอำนาจการทดสอบของงานวิจัยในรูปสมการ power = $1-\beta$ ซึ่งหมายความว่า *"**หากค่า β ลดลง อำนาจการทดสอบของงานวิจัยจะเพิ่มขึ้น หรือ กล่าวอีกนัยหนึ่งได้ว่า งานวิจัยที่มีอำนาจการทดสอบสูงจะมีความสามารถมากขึ้นในการตรวจพบผลกระทบของวิธีการ หรือ สิ่งแทรกแซงที่เกิดขึ้นจริงในระดับประชากร**"* ซึ่งในงานวิจัยทางวิทยาศาสตร์สุขภาพโดยทั่วไป มักกำหนดค่าอำนาจการทดสอบที่ยอมรับได้ไว้ประมาณ 0.80 ซึ่งหมายความว่า *"งานวิจัยมีโอกาสประมาณ 80% ที่จะตรวจพบความแตกต่างที่แท้จริงระหว่างกลุ่ม หากความแตกต่างนั้นมีอยู่จริงในระดับประชากร"* แต่อย่างไรก็ตาม นักวิจัยควรระมัดระวังและหลีกเลี่ยงการแปลความหมายของค่าอำนาจการทดสอบในเชิงตัดสินว่า "ดี" หรือ "ไม่ดี" เนื่องจากการสื่อความหมายดังกล่าวอาจนำไปสู่ความเข้าใจที่คลาดเคลื่อนและชี้นำไปในทางที่ผิด (misleading) ได้ *(Higgs and Amrhein 2025)* ทั้งนี้เนื่องจากค่าอำนาจการทดสอบของงานวิจัยไม่ได้เป็นค่าคงที่ หากแต่ขึ้นกับองค์ประกอบสำคัญหลายประการ ได้แก่ ขนาดตัวอย่าง ขนาดผลกระทบ ความผันแปรของข้อมูลและระดับนัยสำคัญทางสถิติ (α) เป็นต้น ดังเช่น เมื่อขนาดตัวอย่างเพิ่มขึ้น อำนาจการทดสอบของงานวิจัยจะมีแนวโน้มเพิ่มขึ้นตามไปด้วย ในทางตรงกันข้าม หากขนาดตัวอย่าง

มีจำกัด งานวิจัยจะมีแนวโน้มมีอำนาจการทดสอบต่ำและมีความเสี่ยงที่จะไม่สามารถตรวจพบความแตกต่างที่มีระหว่างกลุ่มได้ แม้ว่า ความแตกต่างนั้นจะมีอยู่จริงในระดับประชากรก็ตาม *(Serdar, Cihan et al. 2021)*

ในบริบทของการวิเคราะห์ข้อมูลด้วยวิธีการวิเคราะห์ความแปรปรวนร่วม หรือ ANCOVA การนำตัวแปรร่วมมาปรับค่าในการวิเคราะห์ข้อมูลจะสามารถช่วยลดความผันแปรของตัวแปรผลลัพธ์และเพิ่มประสิทธิภาพของการวิเคราะห์ข้อมูลได้และส่งผลทำให้งานวิจัยมีอำนาจการทดสอบเพิ่มขึ้น เมื่อเปรียบเทียบกับการวิเคราะห์ที่ไม่ปรับค่าด้วยตัวแปรร่วม *(Vickers and Altman 2001; Van Breukelen 2006)* ด้วยเหตุนี้ การวางแผนขนาดตัวอย่างที่เหมาะสม จึงถือเป็นขั้นตอนสำคัญในการออกแบบงานวิจัย เนื่องจากขนาดตัวอย่างที่เพียงพอจะช่วยเพิ่มอำนาจการทดสอบของงานวิจัยและเพิ่มโอกาสในการตรวจพบผลกระทบของสิ่งแทรกแซงที่เกิดขึ้นจริงในระดับประชากร ดังนั้นการทำความเข้าใจแนวคิดเกี่ยวกับอำนาจการทดสอบ จึงถือเป็นพื้นฐานสำคัญสำหรับการคำนวณขนาดตัวอย่างในงานวิจัยทางวิทยาศาสตร์สุขภาพ โดยเฉพาะงานวิจัยที่นำเอาวิธี ANCOVA มาใช้ในการวิเคราะห์ข้อมูล ซึ่งในหัวข้อถัดไป ผู้เขียนจะได้กล่าวถึงองค์ประกอบสำคัญอีกประการหนึ่งของการคำนวณขนาดตัวอย่าง นั่นคือ ขนาดผลกระทบ หรือ effect size

9.2.3 ขนาดผลกระทบ (effect size)

ในการศึกษาวิจัยเชิงทดลองแบบสุ่ม หรือ แบบกึ่งทดลอง นักวิจัยส่วนใหญ่มักมีวัตถุประสงค์เพื่อประเมินว่า วิธีการ หรือ สิ่งแทรกแซงที่นำมาศึกษาก่อให้เกิดผลกระทบต่อผลลัพธ์ที่สนใจหรือไม่ โดยทั่วไป ***ผลกระทบของวิธีการ (treatment effect) หมายถึง "ความแตกต่างของผลลัพธ์ระหว่างกลุ่มที่เกิดจากวิธีการรักษา หรือสิ่งแทรกแซง"*** ซึ่งอาจรายงานในรูปของค่าประมาณทางสถิติ เช่น ความแตกต่างของค่าเฉลี่ย (mean difference) อัตราส่วนความเสี่ยง (risk ratio) อัตราส่วนโอกาส (odds ratio) หรือ hazard ratio เป็นต้น ทั้งนี้ขึ้นกับลักษณะของตัวแปรผลลัพธ์และวิธีการวิเคราะห์ข้อมูลที่ใช้เป็นสำคัญ *(Faraone 2008)* แต่อย่างไรก็ตาม ในการประเมินและเปรียบเทียบขนาดของความแตกต่างดังกล่าว นักวิจัยมักใช้ดัชนีทางสถิติที่เรียกว่า "***ขนาดผลกระทบ (effect size)***" ซึ่งถือเป็นตัวชี้วัดเชิงปริมาณเพื่อใช้สะท้อนขนาด หรือ ความแรงของความแตกต่างที่เกิดขึ้นระหว่างกลุ่ม *(Sullivan and Feinn 2012; Serdar, Cihan et al. 2021)* หรือ กล่าวอีกนัยหนึ่งได้ว่า ขนาด effect size เป็นตัวชี้วัดที่ช่วยบอกนักวิจัยว่า "***ผลกระทบของวิธีการนั้น มีขนาดมาก หรือ น้อยเพียงใด***" แม้ว่าในทางปฏิบัติ "ผลกระทบของวิธีการ (treatment effect)" และ "ขนาดผลกระทบ (effect size)" อาจมีการนำมาใช้สื่อความหมายในลักษณะที่สามารถเทียบเคียงกัน แต่ในเชิงสถิติทั้งสองแนวคิดดังกล่าว มีความหมายแตกต่างกันค่อนข้างชัดเจน ดังนี้

> *"**ผลกระทบของวิธีการ (treatment effect)" หมายถึง** ค่าประมาณของความแตกต่างระหว่างกลุ่มที่เกิดจากสิ่งแทรกแซง*
>
> ***"ขนาดผลกระทบ (effect size)"** หมายถึง ตัวชี้วัดที่ใช้สะท้อนขนาดของความแตกต่างดังกล่าว และมักถูกนำมาใช้เพื่อช่วยตีความผลการวิจัยและเปรียบเทียบผลการศึกษาระหว่างงานวิจัย หรือ ใช้ในการคำนวณอำนาจการทดสอบและขนาดตัวอย่าง (Sullivan and Feinn 2012)*

และเพื่อให้นักวิจัยได้เห็นความแตกต่างระหว่าง treatment effect และ effect size ได้อย่างชัดเจนและเป็นรูปธรรมมากยิ่งขึ้น ผู้เขียนจึงได้นำผลลัพธ์จากกรณีตัวอย่างที่ 8.2 ในบทที่ 8 ข้างต้นมาประกอบการอธิบาย โดยเป็นกรณีการใช้วิธี ANCOVA ที่นำตัวแปรพื้นฐานก่อนการทดลอง (pretest) มาปรับค่าตัวแปรผลลัพธ์หลังการทดลอง (posttest) ภายใต้เงื่อนไขที่ตัวแปรพื้นฐานระหว่างกลุ่มไม่สมดุลในแบบแผนงานวิจัย RCT เกี่ยวกับการศึกษาประสิทธิผลของโปรแกรมดูแลผู้ป่วยโรคไตเรื้อรังแบบเข้มข้นต่อการคงระดับสมรรถภาพไตของผู้ป่วยโรคไตเรื้อรัง **ดังแสดงในแผนภาพที่ 9.1**

```
. regress posttest i.group c.pretest

      Source |       SS           df       MS      Number of obs   =        40
-------------+----------------------------------   F(2, 37)        =   2727.76
       Model |  85.9560188         2  42.9780094   Prob > F        =    0.0000
    Residual |  .582963794        37  .015755778   R-squared       =    0.9933
-------------+----------------------------------   Adj R-squared   =    0.9929
       Total |  86.5389826        39  2.21894827   Root MSE        =    .12552

------------------------------------------------------------------------------
    posttest | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     1.group |   1.856206   .1344521    13.81   0.000      1.58378    2.128632
     pretest |   .7555151   .0218283    34.61   0.000     .7112868    .7997434
       _cons |   12.49617   1.291883     9.67   0.000     9.878568    15.11378
------------------------------------------------------------------------------

. margins group

Predictive margins                                      Number of obs = 40
Model VCE: OLS

Expression: Linear prediction, predict()
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       group |
           0 |    54.9769   .0700945   784.33   0.000     54.83487    55.11892
           1 |    56.8331   .0700945   810.81   0.000     56.69108    56.97513
------------------------------------------------------------------------------

. estat esize

Effect sizes for linear models

-------------------------------------------------------------
             Source | Eta-squared     df     [95% conf. interval]
--------------------+----------------------------------------
              Model |   .9932636      2     .9878219    .9951613
                    |
              group |   .8374323      1     .7226971    .8867271
            pretest |   .9700399      1     .9473789     .979042
-------------------------------------------------------------
Note: Eta-squared values for individual model terms are partial.
```

treatment effect

effect size (partial eta squared)

แผนภาพที่ 9.1 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์ผลกระทบของวิธีการ (treatment effect), ค่าเฉลี่ยที่ปรับค่าแล้ว (adjusted mean) และขนาดผลกระทบ (effect size) กรณีตัวแปรพื้นฐานระหว่างกลุ่มไม่สมดุล ภายใต้บริบทแบบแผนงานวิจัย RCT จากกรณีตัวอย่างที่ 8.2

จากแผนภาพที่ 9.1 จะเห็นได้ว่า ค่าประมาณของผลกระทบของวิธีการ (treatment effect) ที่ได้จากตัวแบบ ANCOVA มีค่าเท่ากับ 1.87 ซึ่งสะท้อนความแตกต่างของค่าเฉลี่ยของค่า eGFR หลังการติดตาม 12 สัปดาห์ (posttest) ระหว่างกลุ่มควบคุม (group=0) และกลุ่มทดลอง (group=1) หลังจากมีการปรับค่าตัวแปรพื้นฐาน (pretest) แล้ว โดยสามารถตีความได้ว่า "***ค่าเฉลี่ยของค่า eGFR หลังการติดตาม 12 สัปดาห์ของกลุ่มทดลองมีค่ามากกว่ากลุ่มควบคุมประมาณ 1.87 หน่วยหลังจากปรับค่าด้วยตัวแปรพื้นฐานแล้ว***"

แต่อย่างไรก็ตาม แม้ว่า ค่าผลต่างดังกล่าวจะสะท้อนขนาดของผลกระทบของวิธีการได้ในระดับหนึ่ง ***แต่ค่าประมาณของ treatment effect เพียงอย่างเดียว ยังไม่สามารถสะท้อนความแรงของผลกระทบในเชิงสถิติได้อย่างสมบูรณ์ เนื่องจากค่าดังกล่าวยังคงขึ้นกับหน่วยวัด หรือ สเกลของตัวแปรผลลัพธ์ที่นำมาใช้ในการวิเคราะห์*** ดังนั้นนักวิจัยจึงมักใช้ขนาดผลกระทบ (effect size) เป็นตัวชี้วัดเพิ่มเติม เพื่อสะท้อนความแรงของผลกระทบที่เกิดขึ้นจากสิ่งแทรกแซง โดย effect size ถือเป็นค่าที่มีการปรับมาตรฐาน (standardized measure) จึงช่วยทำให้สามารถนำไปเปรียบเทียบขนาดของผลกระทบระหว่างงานวิจัย หรือ ระหว่างการวิเคราะห์ด้วยวิธีการทางสถิติต่าง ๆ ได้อย่างเหมาะสมมากขึ้น ***(Sullivan and Feinn 2012)*** สำหรับการวิเคราะห์ข้อมูลด้วยวิธี ANCOVA ค่าขนาดผลกระทบที่ถูกนำมาใช้อย่างแพร่หลาย ประกอบด้วย ค่า eta squared (η^2) ค่า partial eta squared (η_p^2) และค่า Cohen's f เป็นต้น โดยจากผลลัพธ์ของโปรแกรม STATA ที่แสดงในแผนภาพที่ 9.1 พบว่าค่า partial eta squared (η_p^2) ของตัวแปรกลุ่ม (group) เท่ากับ 0.837 ซึ่งสามารถแปลความหมายได้ว่า

> ***"ภายใต้ตัวแบบ ANCOVA นี้ หลังจากปรับค่าด้วยตัวแปรพื้นฐาน (pretest) ในตัวแบบแล้ว ตัวแปรกลุ่ม (group) สามารถอธิบายความผันแปรของตัวแปรผลลัพธ์ (posttest) ได้ประมาณร้อยละ 83.7"***

ซึ่งในการคำนวณค่า partial eta squared (η_p^2) สามารถพิจารณาภายใต้รูปสมการดังนี้

$$\eta_p^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}} \qquad \text{---------- } ❶$$

เมื่อ SS_{effect} = ผลรวมกำลังสองของตัวแปรที่สนใจ (treat)

SS_{error} = ผลรวมกำลังสองของความคลาดเคลื่อน (residual)

ขณะที่ขนาดผลกระทบของค่า eta squared (η^2) ที่ถูกนิยมนำมาใช้ในกรณี one-way ANOVA สามารถพิจารณาภายใต้รูปสมการ ดังนี้

$$\eta^2 = \frac{SS_{effect}}{SS_{total}} \qquad \text{---------- } ❷$$

เมื่อ SS_{effect} = ผลรวมกำลังสองของตัวแปรที่สนใจ (treat)

SS_{total} = ผลรวมกำลังสองทั้งหมดของตัวแบบ

จากสมการดังกล่าว จะเห็นได้ว่า ***ค่า partial eta squared (η_p^2) เป็นตัวชี้วัดที่สะท้อนสัดส่วนของความแปรปรวนของผลลัพธ์ที่สามารถอธิบายได้จากปัจจัยที่สนใจ ภายใต้ตัวแบบทางสถิติที่กำหนด โดยพิจารณาร่วมกับความคลาดเคลื่อนที่เหลืออยู่ในตัวแบบ ขณะที่ค่า eta squared (η^2) จะสะท้อนสัดส่วนของความแปรปรวนทั้งหมดของผลลัพธ์ที่สามารถอธิบายได้จากปัจจัยที่สนใจ***

ในการตีความขนาดผลกระทบของค่า η^2 หรือ η_p^2 นักวิจัยส่วนใหญ่มักอ้างอิงเกณฑ์ที่ถูกนำเสนอโดย **Cohen** ซึ่งค่าประมาณ 0.01, 0.06 และ 0.14 สามารถตีความได้ว่า เป็นขนาดผลกระทบเล็ก ปานกลาง และใหญ่ตามลำดับ *(Cohen 1988)* แต่อย่างไรก็ตาม นักวิจัยควรตระหนักว่า ค่า η^2 และ η_p^2 มีแนวคิดและสูตรการคำนวณที่แตกต่างกัน ดังนั้นการนำเกณฑ์เดียวกันดังกล่าวมาใช้และแปลความหมาย จึงควรทำด้วยความระมัดระวัง *(Richardson 2011)* และในทางปฏิบัติสำหรับการรายงานขนาดผลกระทบดังกล่าว ควรรายงานควบคู่ไปกับช่วงความเชื่อมั่นที่ระดับ 95% เพื่อเพิ่มความชัดเจนและครอบคลุมของการรายงานผลการวิจัย *(Lakens 2013)*

จากผลการวิเคราะห์ในกรณีตัวอย่างที่ 8.2 พบว่า ค่า partial eta squared (η_p^2) มีค่าประมาณ 0.837 ซึ่งจัดอยู่ในระดับค่อนข้างใหญ่เมื่ออ้างอิงจากเกณฑ์ทั่วไป แต่อย่างไรก็ตาม เมื่อพิจารณาร่วมกันระหว่าง treatment effect และ effect size จะเห็นได้ว่า ทั้งสองค่ามีบทบาทในการแปลความหมายที่แตกต่างกัน กล่าวคือ

- **treatment effect** เท่ากับ 1.87 สะท้อนความแตกต่างของค่าเฉลี่ยที่ปรับค่าแล้วระหว่างกลุ่ม
- **effect size** หรือ η_p^2 เท่ากับ 0.837 สะท้อนความแรงของผลกระทบของวิธีการในเชิงสถิติภายใต้ตัวแบบที่ใช้วิเคราะห์

จากผลการวิเคราะห์ดังกล่าวข้างต้น จึงสามารถสรุปได้ว่า

> **ผลกระทบของวิธีการ (treatment effect)** *เป็นค่าที่มุ่งอธิบายความแตกต่างของค่าเฉลี่ย หรือ ค่ากลางของข้อมูลที่ถูกปรับค่าแล้วระหว่างกลุ่มที่นำมาเปรียบเทียบ โดยยังคงมีหน่วยวัดตามสเกลเดิมของตัวแปรผลลัพธ์*
> **ขณะที่ขนาดผลกระทบ (effect size)** *เป็นค่าที่มุ่งอธิบายขนาด หรือ ความแรงของผลกระทบในรูปของค่ามาตรฐาน ซึ่งช่วยให้สามารถเปรียบเทียบขนาดของผลกระทบระหว่างงานวิจัย หรือ ระหว่างบริบทที่ใกล้เคียงกันได้สะดวกมากขึ้น*

จากแนวคิดดังกล่าว แม้ว่า ขนาดผลกระทบ จะถูกนำมาใช้เพื่อสะท้อนความแรงของผลกระทบในเชิงสถิติ แต่ในบริบทของการออกแบบงานวิจัยและการคำนวณขนาดตัวอย่าง ค่าดังกล่าว ยังมีบทบาทสำคัญอย่างยิ่งต่อการกำหนดจำนวนตัวอย่างที่เหมาะสม เนื่องจากขนาดผลกระทบเป็นหนึ่งในองค์ประกอบหลักของการคำนวณอำนาจการทดสอบทางสถิติ ซึ่งมีความสัมพันธ์โดยตรงกับขนาดตัวอย่างที่จำเป็นต้องใช้ในงานวิจัย โดยทั่วไป หากขนาดผลกระทบของสิ่งแทรกแซงมีขนาดใหญ่ งานวิจัยจะสามารถตรวจพบความแตกต่างระหว่างกลุ่มได้ง่ายขึ้น และอาจใช้ขนาดตัวอย่างน้อยลงได้ ในทางตรงกันข้าม หากขนาดผลกระทบมีขนาดเล็ก การตรวจพบความแตกต่างดังกล่าวจะทำได้ยากขึ้นและจำเป็นต้องใช้ขนาดตัวอย่างมากขึ้นเพื่อเพิ่มอำนาจการทดสอบของงานวิจัย *(Cohen 1988; Serdar, Cihan et al. 2021)*