อำนาจของการทดสอบทางสถิติ : ข้อควรคำนึงสำหรับการกำหนดขนาดตัวอย่างเพื่อการวิจัย

ดร. องอาจ นัยพัฒน์ *

บทนำ

คำถามสำคัญประการหนึ่งที่ผู้เขียนมักได้รับฟังอยู่เสมอ ๆ จากบรรดานิสิต/นักศึกษาระดับบัณฑิตศึกษา ตลอดจนครูอาจารย์ทั่วไปที่กำลังทำวิจัยเชิงปริมาณ (quantitative research) คือ *"ขนาดกลุ่มตัวอย่างที่พอเพียงสำหรับการวิจัยควรเป็นเท่าไร"* แทนที่จะตอบคำถามนี้ทันทีด้วยการแนะนำให้ผู้ถามค้นหาคำตอบด้วยตนเองจากการเปิดตารางสำเร็จรูป (เช่น Krejcie & Morgan, 1970; Yamane, 1967) ซึ่งมักนิยมกระทำกันในการนำตารางสำเร็จรูปดังกล่าวมาใช้อ้างอิงในการกำหนดขนาดตัวอย่างเพื่อการวิจัย ผู้เขียนมักเริ่มด้วยการถามคำถามแรกกลับไปยังผู้ถามว่า *"ระดับนัยสำคัญทางสถิติ (α) ที่กำหนดไว้ในการวิจัยเป็นเท่าไร"* คำถามนี้ส่วนใหญ่ผู้เขียนมักได้รับคำตอบคือ "ระดับ .05 หรือ .01" แต่ครั้นพอผู้เขียนตั้งคำถามกลับไปใหม่ว่า *"ค่าประมาณความแปรปรวนของตัวแปรเกณฑ์ (criterion variable) และขนาดผลของการจัดกระทำทางการวิจัย (treatment effect size) ที่ศึกษามีค่าเท่าไร"* ผู้เขียนมักจะไม่ได้รับคำตอบของคำถามข้อนี้จากผู้ที่ต้องการทราบคำตอบว่าขนาดกลุ่มตัวอย่างที่พอเพียงควรเป็นเท่าไร ด้วยเหตุนี้จึงทำให้ผู้ต้องการทราบคำตอบเกี่ยวกับขนาดของกลุ่มตัวอย่างที่พอเพียงหลายคน (โดยเฉพาะอย่างยิ่งผู้ที่ใจร้อนต้องการทราบคำตอบอย่างรวดเร็ว) รู้สึกผิดหวังกับผู้เขียนที่ไม่สามารถตอบคำถามดังกล่าวได้ ดังนั้นจึงทำเสมือนสงวนท่าทีด้วยการถามคำถามย้อนกลับเพียงเพื่อหลีกเลี่ยงต่อการตอบคำถามดังกล่าวโดยตรง

ในประเด็นนี้ความจริงแล้วผู้เขียนมีคำตอบและข้อเสนอแนะที่ชัดเจนให้กับผู้

* อาจารย์ประจำ สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ

ต้องการทราบคำตอบ เพียงแต่ผู้เขียนเชื่อว่าการตอบคำถามนี้**ไม่สามารถกระทำได้โดยง่ายและอย่างรวดเร็ว**ด้วยการแค่เพียงเปิดตารางสำเร็จรูปที่กล่าวถึงข้างต้น แล้วหาขนาดกลุ่มตัวอย่างที่ระบุไว้ในตารางขนาดเล็กสุด ตามระดับนัยสำคัญ (α) และจำนวนของประชากร (N) ที่ทราบเท่านั้น ดังที่นักวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์หลายท่านเข้าใจ ในทางตรงข้ามการตอบคำถามนี้ต้องอาศัยข้อมูลสารสนเทศอื่น ๆ ประกอบการพิจารณา โดยเฉพาะอย่างยิ่ง**อำนาจของการทดสอบทางสถิติ** (power of statistical test) นอกเหนือไปจากระดับนัยสำคัญ (α) ค่าประมาณความแปรปรวนของตัวแปรเกณฑ์ และขนาดผลของการจัดกระทำทางการวิจัยต่ำสุดที่คาดว่ามีความสำคัญในทางปฏิบัติว่าเป็นเท่าไร

ผู้เขียนนำเสนอบทความนี้มีจุดมุ่งหมายสำคัญเพื่อต้องการอธิบายเกี่ยวกับอิทธิพลของขนาดกลุ่มตัวอย่างที่ส่งผลต่ออำนาจของการทดสอบทางสถิติและนำเสนอสูตรการคิดคำนวณและตารางสำเร็จรูปสำหรับช่วยในการกำหนดขนาดของกลุ่มตัวอย่างที่คำนึงถึงอำนาจของการทดสอบทางสถิติ นอกเหนือไปจากปัจจัยเกี่ยวข้องอื่น ๆ ซึ่งจะได้กล่าวถึงต่อไป

อำนาจและประสิทธิภาพของอำนาจการทดสอบทางสถิติ

อำนาจของการทดสอบทางสถิติ หมายถึง ความน่าจะเป็น (probability) ของการทดสอบสมมติฐานกลาง (null hypothesis; H_0) ที่นำไปสู่การปฏิเสธสมมติฐานกลาง (มีนัยสำคัญทางสถิติ) ได้อย่างถูกต้องตรงกับความเป็นจริงเมื่อสมมุติฐานกลางเป็นเท็จ หรือกล่าวอย่างง่าย ๆ ได้ว่าเป็นความสามารถของการทดสอบทางสถิติใด ๆ ที่จะปฏิเสธสมมุติฐานกลางที่เป็นเท็จ (false H_0) อำนาจของการทดสอบทางสถิติสามารถเขียนให้อยู่ในรูปทั่วไปได้ดังต่อไปนี้

อำนาจของการทดสอบทางสถิติ = 1 – (ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 2; type II error or β)

$$= 1-\beta \quad \text{.....(1)}$$

จากคำจำกัดความที่กล่าวมาข้างต้น จะพบว่าค่าอำนาจของการทดสอบทางสถิติใด ๆ สามารถคำนวณได้ในกรณีที่สมมติฐานกลางเป็นเท็จ ทั้งนี้เพราะเมื่อสมมติฐานกลางเป็นจริงแล้ว เราจะไม่มีโอกาสตัดสินใจคลาดเคลื่อนในการยอมรับ (ไม่ปฏิเสธ) สมมติฐานกลางที่เป็นเท็จ (β) ได้เลย อย่างไรก็ตามมี

โอกาสเป็นไปได้ที่อาจปฏิเสธสมมุติฐานกลางที่เป็นความจริงนั้น หรือเกิดความคลาดเคลื่อนในการตัดสินใจประเภทที่ 1 (type I error or α) ขึ้น ความสัมพันธ์เชิงผกผันระหว่างความคลาดเคลื่อนในการตัดสินใจปฏิเสธสมมติฐานกลางหรือไม่ สามารถสรุปได้ดังแผนภาพ 1 ดังนี้

สภาพความเป็นจริงเกี่ยวกับ H_0

การตัดสินใจ	H_0 เป็นจริง	H_0 เป็นเท็จ
ไม่ปฏิเสธ H_0	ตัดสินใจถูกต้อง ($p = 1-\alpha$)	$p = \beta$
ปฏิเสธ H_0	$p = \alpha$	ตัดสินใจถูกต้อง ($p=1-\beta$) **อำนาจของการทดสอบ**

แผนภาพ 1 ความสัมพันธ์เชิงผกผันระหว่างความคลาดเคลื่อนประเภทที่ 1 (α) และประเภทที่ 2 (β) ที่อาจเกิดขึ้นจากการตัดสินใจทางสถิติ

ในทางปฏิบัติโดยทั่วไป นักวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์ให้ความสำคัญต่อการควบคุมความคลาดเคลื่อนประเภทที่ 1 โดยพยายามควบคุมให้มีระดับต่ำ เช่น .05 หรือ .01 นั่นคือ นักวิจัยยอมเสี่ยงแค่เพียง 5% หรือ 1% เท่านั้นต่อการตัดสินใจผิดพลาดในการปฏิเสธสมมุติฐานกลางที่เป็นจริง ในบางสภาพการณ์ของการวิจัยเชิงทดลองบางเรื่องที่ศึกษากับกลุ่มตัวอย่างที่อยู่ในภาวะเสี่ยง อาทิ เด็กเล็ก กลุ่มเยาวชนติดยาเสพติด

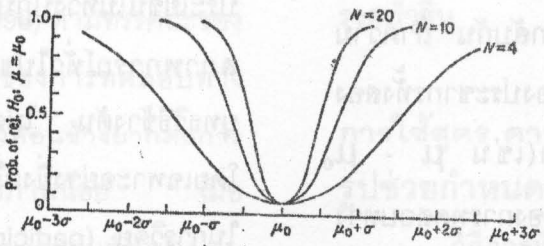
เด็กหรือผู้ใหญ่ที่มีความบกพร่องทางด้านสติปัญญาและการรับรู้ และคนชรา นักวิจัยอาจควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 เข้มงวดยิ่งขึ้นด้วยการลดระดับลงเหลือเพียง .001 ทั้งนี้เพื่อป้องกันความผิดพลาดอันเกิดจากการยอมรับวิธีการพัฒนา บำบัดรักษา หรือเพิ่มศักยภาพใด ๆ (ตามสมมติฐานการวิจัยว่าให้ผลดี) ทั้งที่ความจริงวิธีการดังกล่าวนั้นไร้ประสิทธิผล ซ้ำร้ายยังก่อให้เกิดผลข้างเคียง (size effect) เป็นอันตรายต่อกลุ่มตัวอย่าง ความเข้มงวดในการควบคุมระดับความคลาดเคลื่อนประเภทที่ 1 เช่นนี้ มีผลทำให้อำนาจของการทดสอบทางสถิติมีระดับต่ำลงค่อนข้างมาก เพราะมีโอกาสของการเกิดความคลาดเคลื่อนประเภทที่ 2 สูง (β มีระดับสูง ดังนั้น $1 - \beta$ มีระดับต่ำ) ดังนั้นในกรณีนี้ถ้านักวิจัยมีความต้องการออกแบบการวิจัยเชิงทดลองหรือกึ่งทดลองใด ๆ ให้มีความไว (sensitivity) ต่อการตรวจพบประสิทธิผลที่เกิดขึ้นจริงและมีความสำคัญในทางปฏิบัติ โดยพิจารณาจากเกณฑ์ดั้งเดิมคือความมีนัยสำคัญทางสถิติแล้ว ผู้วิจัยจะต้องกำหนดให้ความคลาดเคลื่อนประเภทที่ 2 มีระดับต่ำ นั่นคือ นักวิจัยจะต้องเพิ่มอำนาจของการทดสอบ ($1-\beta$) ให้สูงขึ้น

จากการให้ความสำคัญกับบทบาทของอำนาจการทดสอบทางสถิติในการออกแบบการวิจัย คำถามที่ตามมาคือ *"ระดับของอำนาจการทดสอบทางสถิติควรเป็นเท่าไรจึงจะพอเพียง"* (นั่นคือ "ระดับความเสี่ยงที่ยอมรับได้ของ β ควรเป็นเท่าไร") Cohen (1988) ศาสตราจารย์ของภาควิชาจิตวิทยาแห่งมหาวิทยาลัยนิวยอร์ค สหรัฐอเมริกา ผู้นำคนสำคัญที่กระตุ้นให้นักวิจัยทางพฤติกรรมศาสตร์ตระหนักถึงอำนาจของการทดสอบทางสถิติที่มีต่อการออกแบบการวิจัยได้แนะนำว่าระดับ β สำหรับใช้ในการวิจัยทั่วไปควรเท่ากับ .20 (หรืออำนาจของการทดสอบทางสถิติเท่ากับ .80) นี่แสดงว่าโดยทั่วไปนักวิจัยให้ความสำคัญกับความร้ายแรงของ α (.05) เป็นสี่เท่าของ β (.20) ในประเด็นนี้นักสถิติและผู้เชี่ยวชาญทางด้านแบบแผนการวิจัยบางท่าน (อาทิ Lipsey, 1998; Oakes & Feldman, 2001) เห็นว่านักวิจัยอาจสูญเสียความรู้ที่ได้รับจากผลการวิจัยซึ่งอาจมีคุณค่าในทางปฏิบัติเป็นจำนวนมากก็ได้ ทั้งนี้เพราะผลการจัดกระทำของการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์ส่วนใหญ่มีขนาดเล็ก ดังนั้นการออกแบบการวิจัยที่เข้มงวดกับการควบคุมระดับนัยสำคัญทางสถิติ (α) มากเกินไป จึงทำให้มีความเป็นไปได้น้อยต่อการตรวจพบผลการจัดกระทำของการวิจัยขนาดเล็ก (ไม่มีนัยสำคัญทางสถิติ) ในทางกลับกันถ้านักวิจัยประเมิน

แล้วว่าผลของการจัดกระทำทางการวิจัยขนาดเล็กดังกล่าวมีคุณค่าหรือประโยชน์ในทางปฏิบัติน้อย การออกแบบการวิจัยเพื่อให้มีอำนาจของการทดสอบทางสถิติสูง ๆ เพียงเพื่อให้ได้รับผลการวิจัยที่มีนัยสำคัญทางสถิติ (ซึ่งโดยทั่วไปนักวิจัยปรารถนา) ก็ควรลดความสำคัญลงมา (หรือกำหนด β มีระดับสูงขึ้น) ด้วยเหตุนี้การออกแบบการวิจัยเพื่อให้มี **อำนาจของการทดสอบทางสถิติอยู่ในระดับที่เหมาะสมและสอดคล้องกับสภาพการณ์และธรรมชาติของการทำวิจัย** จึงเป็นสิ่งที่นักวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์ควรให้ความสนใจ

อำนาจของการทดสอบทางสถิติกับการกำหนดขนาดกลุ่มตัวอย่าง

เป็นที่ทราบกันโดยทั่วไปในวงการวิจัยว่า เมื่อกำหนดให้ระดับนัยสำคัญทางสถิติสถิติเชิงอนุมานที่ใช้ทดสอบ และทิศทางการทดสอบเป็นแบบเดียวกันแล้ว การวิจัยใด ๆ ที่ทำการศึกษากับกลุ่มตัวอย่างที่มีขนาดใหญ่มีโอกาสสูงมาก (เกือบ 100%) ที่จะปฏิเสธสมมติฐานกลาง (พบผลการวิจัยมีนัยสำคัญทางสถิติ) ในทางตรงข้ามการวิจัยเรื่องเดียวกันถ้าศึกษากับกลุ่มตัวอย่างขนาดเล็ก มีความน่าจะเป็นของการไม่ปฏิเสธสมมติฐานกลางสูงมาก (พบผลการวิจัยไม่มีนัยสำคัญทางสถิติ) สาเหตุที่เป็นเช่นนี้เพราะว่าขนาดกลุ่มตัวอย่างเป็นปัจจัยสำคัญประการหนึ่งที่มีความสัมพันธ์ทางบวกกับอำนาจของการทดสอบทางสถิติ ซึ่งสามารถอธิบายด้วยแผนภาพฟังก์ชั่นของอำนาจของการทดสอบทางสถิติแบบสองด้าน (two – sided test) ที่ระดับนัยสำคัญทางสถิติ .05 ดังนี้



แผนภาพ 2 ฟังก์ชั่นอำนาจของการทดสอบทางสถิติกับขนาดกลุ่มตัวอย่างสามแบบ (N=20, 10, และ 4 คนตามลำดับ)(Dixon & Massey, 1957, p. 252 cited in Blalock, 1981, p. 249)

จากแผนภาพ 2 สมมติว่านักวิจัยต้องการทดสอบสมมุติฐาน $H_0 : \mu = \mu_0$ (หรือ $H_0 : \mu - \mu_0 = 0$) แล้ว จะพบว่าถ้าค่าเฉลี่ยที่แท้จริงของประชากร (μ) เป็นค่าที่เบี่ยงเบนออกไปจากค่าเฉลี่ยประชากรที่คาดหมายไว้ (μ_0) เท่ากับหนึ่งความคลาดเคลื่อนมาตรฐาน ($\mu - \mu_0 = \pm 1\sigma$) แล้ว อำนาจของการทดสอบหรือความน่าจะเป็นในการปฏิเสธ H_0 ที่เป็นเท็จจริง ๆ สำหรับการวิจัยซึ่งทำการศึกษากับกลุ่มตัวอย่างขนาดเท่ากับ 20 คน จะมีสูงกว่ากลุ่มตัวอย่างขนาดเท่ากับ 10 และ 4 คน ตามลำดับ อย่างไรก็ตามถ้าปรากฏว่าการเบี่ยงเบนดังกล่าวมีขนาดใหญ่มากแล้ว (เช่น $\mu - \mu_0 = \pm 2\sigma$) แล้ว จะพบว่าอำนาจของการทดสอบทางสถิติต่อการตรวจพบความแตกต่างขนาดสองความคลาดเคลื่อนมาตรฐานสำหรับกลุ่มตัวอย่างสามขนาดจะมีระดับสูงและใกล้เคียงกัน (แนวโน้มมีนัยสำคัญทางสถิติที่ระดับ .05) ในทางกลับกัน ถ้าความเบี่ยงเบนระหว่างค่าเฉลี่ยของประชากรทั้งสองประเภทมีขนาดเล็กมาก(เช่น $\mu - \mu_0 = \pm .25\sigma$) แล้ว อำนาจของการทดสอบทางสถิติต่อการตรวจพบความแตกต่างขนาดหนึ่งในสี่เท่าของความคลาดเคลื่อนมาตรฐานสำหรับกลุ่มตัวอย่างทั้งสามขนาดจะมีระดับต่ำและใกล้เคียงกัน (แนวโน้มไม่มีนัยสำคัญทางสถิติที่ระดับ .05) ดังนั้นในสภาพปกติทั่วไปถ้าความเบี่ยงเบนระหว่างค่าเฉลี่ยที่แท้จริงและที่คาดหมายของประชากรไม่มีขนาดใหญ่และเล็กมากเกินไป การเพิ่มขึ้นของขนาดกลุ่มตัวอย่างจะทำให้อำนาจของการทดสอบทางสถิติมีระดับสูงขึ้น

จากความสัมพันธ์ระหว่างขนาดกลุ่มตัวอย่างกับอำนาจของการทดสอบทางสถิติที่อธิบายมาแล้ว อาจทำให้คิดไปว่าในการออกแบบการวิจัยควรจะมีการเพิ่มขนาดกลุ่มตัวอย่างเพื่อให้อำนาจของการทดสอบทางสถิติมีระดับสูงขึ้น เป็นยุทธวิธีที่มีประสิทธิผลสำหรับแก้ไขข้อบกพร่องการทดสอบนัยสำคัญทางสถิติ (significance testing) ในกรณีที่การทดสอบดังกล่าวมีอำนาจหรือความไวต่อการตรวจพบผลของการทดลองหรือการจัดกระทำทางการวิจัยที่มีขนาดเล็กแต่มีคุณค่าหรือประโยชน์ในทางปฏิบัติ อย่างไรก็ตามในสภาพการณ์ทั่วไปของการวิจัยพบว่าการใช้ยุทธวิธีข้างต้น ไม่สามารถกระทำได้เสมอไป โดยเฉพาะอย่างยิ่งเมื่อทรัพยากรและผู้เข้าร่วมในการวิจัย (participants or subjects) มีจำนวนจำกัด ดังนั้นนักวิจัยอาจต้องอาศัยยุทธวิธีอื่น ๆ เช่น การผ่อนคลายความเข้มงวดของการควบคุมระดับนัยสำคัญทางสถิติให้มี

ระดับสูงขึ้น (เช่น จาก α = .01 เป็น .05 หรือ .10) วิธีการนี้กระทำได้ง่ายแต่ผู้เชี่ยวชาญทางสถิติและแบบแผนการวิจัยหลายท่าน (เช่น Lipsey, 1998; Murphy & Myors, 1998) เตือนให้กระทำด้วยความระมัดระวัง เพราะการเปลี่ยนแปลงระดับ α ไม่ได้มีความหมายสำคัญหรือเป็นประโยชน์ต่อการแปลความหมายผลของการวิจัยแต่อย่างใด ด้วยเหตุนี้ยุทธวิธีดังกล่าวแม้ว่าจะกระทำได้ง่ายกว่าการเพิ่มขนาดตัวอย่างแต่ก็มีประโยชน์ไม่มากเมื่อเปรียบเทียบกับการเพิ่มขนาดกลุ่มตัวอย่าง ซึ่งนอกจากจะช่วยเพิ่มอำนาจของการทดสอบแล้วยังช่วยเพิ่มเสถียรภาพของผลการวิจัยและช่วยลดโอกาสที่ความคลาดเคลื่อนของการสุ่มอันจะนำไปสู่การได้รับผลการวิจัยที่ไม่มีความหมายใด ๆ และซ้ำร้ายยังอาจก่อให้เกิดความเข้าใจผิดได้ สำหรับยุทธวิธีอื่นนอกเหนือจากการปรับ α ให้มีระดับสูงขึ้น ผู้สนใจสามารถศึกษาได้จาก Lipsey (1998) ตามทรรศนะของผู้เขียนการเพิ่มอำนาจของการทดสอบทางสถิติด้วยยุทธวิธีเหล่านั้นค่อนข้างยากต่อการปฏิบัติและอาจมีประสิทธิภาพน้อย เมื่อเปรียบเทียบกับการเพิ่มขนาดกลุ่มตัวอย่าง

ในปัจจุบันการกำหนดขนาดกลุ่มตัวอย่างสำหรับการวิจัยทางพฤติกรรมศาสตร์ที่ออกแบบการวิจัยเป็นแบบทดลองและกึ่งทดลอง นักวิจัยหันมาให้ความสำคัญกับอำนาจของการทดสอบทางสถิติ (หรือหันมาสนใจ β มากขึ้น) นอกเหนือไปจากระดับนัยสำคัญทางสถิติ (หรือสนใจ α แต่เพียงอย่างเดียว) ดังจะเห็นได้จากมีการตีพิมพ์เอกสารทางวิชาการจำนวนมาก (เช่น Cohen, 1988; Hinkel, Wiersma, & Jurs, 1994; Kramer & Thiemann, 1987; Lipsey, 1998; Olejnik, 1984) นำเสนอตารางสำเร็จรูปช่วยกำหนดขนาดกลุ่มตัวอย่างที่คำนึงถึงอำนาจของการทดสอบ ซึ่งผู้เขียนจะได้นำเสนอต่อไป ยิ่งไปกว่านั้นในปัจจุบันนี้ยังได้มีโปรแกรมคอมพิวเตอร์สำเร็จรูปรู (เช่น PASS 2000, nQuery 4, Stat – Power 2, และ Power and Precision) วางจำหน่ายในตลาดซอร์ฟแวร์ทางสถิติ เพื่อช่วยอำนวยความสะดวกในการคำนวณค่าอำนาจของการทดสอบทางสถิติและประมาณค่าขนาดกลุ่มตัวอย่างให้ง่ายและรวดเร็วขึ้น

การใช้สูตร ตาราง และแผนภูมิสำเร็จรูปช่วยกำหนดขนาดกลุ่มตัวอย่าง

วิธีการที่ได้รับความนิยมกันโดยทั่วไปในการกำหนดขนาดกลุ่มตัวอย่างเพื่อการทำวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์ได้แก่ การเปิดตารางสำเร็จรูปของ Krejcie &

Morgan (1970) และ Yamane (1967) เพราะสามารถกระทำได้โดยง่ายและสะดวก ดังที่กล่าวมาแล้วในตอนบทนำของบทความนี้ อย่างไรก็ตามผู้เขียนเชื่อว่าการกำหนดขนาดกลุ่มตัวอย่างโดยวิธีการเปิดตารางสำเร็จรูปดังกล่าว นักวิจัยควรกระทำด้วยความระมัดระวัง โดยเฉพาะอย่างยิ่งการกำหนดขนาดกลุ่มตัวอย่างเพื่อการทำวิจัยเชิงทดลองหรือกึ่งทดลอง (มีการทดสอบสมมุติฐานทางสถิติ) ทั้งนี้เพราะโดยพื้นฐานแล้วตารางสำเร็จรูปดังกล่าวพัฒนาขึ้นมาจากสูตรทั่วไป คือ (Cochran, 1977, p.75; Yamane, 1967, p. 886 – 887)

$$n = \frac{NZ^2pq}{d^2(N-1)+Z^2pq} \quad \text{.....(2)}$$

หรือ $= \frac{N}{(N-1)d^2+1}$ [เมื่อให้ $p = 0.5$ และ $z = 2$] ..(3)

[เมื่อ n แทน ขนาดกลุ่มตัวอย่าง, N แทนจำนวนประชากร, Z แทนคะแนนมาตรฐาน Z ที่ระดับนัยสำคัญทางสถิติใด ๆ ตามข้อกำหนดของการวิจัย, p แทนค่าประมาณสัดส่วนของประชากรที่นักวิจัยสนใจศึกษา, q แทน 1-p, และ d แทนระดับความแม่นตรงของการประมาณค่า]

จากสูตร (2) จะพบว่าการประมาณค่าขนาดกลุ่มตัวอย่างที่พอเพียงต้องทราบปัจจัยจำเป็น 4 ประการ ได้แก่ 1) จำนวนประชากร (N), 2) ระดับนัยสำคัญทางสถิติ (α), 3) ความแปรปรวนของสัดส่วนประชากร ($\sigma_p^2 = PQ/n$), และ 4) ระดับความแม่นตรงหรือขอบเขตความคลาดเคลื่อนของการประมาณค่าที่ยอมให้เกิดได้ (precision lelvel or margin of error; d) ความจริงค่าของ d ก็คือค่าของขนาดผลของการจัดกระทำทางการวิจัย (ES) ที่ผู้เขียนได้กล่าวมาแล้วข้างต้น นอกจากนี้การใช้สูตร (2) ต้องตั้งอยู่บนข้อตกลงเบื้องต้นที่ว่าข้อมูลของตัวแปรเกณฑ์ที่สนใจศึกษาต้องเป็นแบบแบ่งสอง (dichotomy) และจะต้องทราบจำนวนประชากรด้วยว่าเป็นเท่าไร ทั้งที่มีข้อจำกัดเหล่านี้แต่ดูเหมือนว่านิสิต/นักศึกษาระดับบัณฑิตศึกษาและนักวิจัยทั่วไปยังคงใช้ตารางสำเร็จรูปดังกล่าวต่อไป แม้ว่ามีบ่อยครั้งไม่เป็นไปตามข้อตกลงเบื้องต้นทั้งสองประการ นอกจากนี้ผู้เขียนเชื่อว่ากลุ่มตัวอย่างที่กำหนดมาจากสูตร (2) และ (3) รวมทั้งสูตรอื่น ๆ ที่นำเสนอด้วย Cochran (1977) ศาสตราจารย์ทางสถิติที่มีชื่อเสียงแห่งมหาวิทยาลัยฮาร์วาร์ด สหรัฐอเมริกา เหมาะสำหรับนำไปใช้กับการศึกษาวิจัยเชิงสำรวจตัวอย่าง (sample survey) ที่มีจุดมุ่งหมายสำคัญเพื่อกะประมาณค่าพารามิเตอร์ของประชากรที่ไม่ทราบค่า ณ ระดับความแม่นตรงหนึ่ง ๆ มากกว่าการศึกษาวิจัยเชิงทดลองหรือกึ่งทดลองที่มีจุดมุ่งหมายหลักเพื่อทดสอบสมมุติฐาน

เปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยหรือค่าสถิติอื่นใด (สำหรับเหตุผลจะได้กล่าวถึงต่อไปนี้)

ตามทรรศนะของผู้เขียนการกำหนดขนาดกลุ่มตัวอย่างด้วยสูตร (2) หรือตารางสำเร็จรูปของ Krejcie & Morgan (1970) และ Yamane (1967) นักวิจัยจะต้องเข้าใจว่าขนาดตัวอย่างที่ได้รับจากสูตร (2) (หรือตารางสำเร็จรูปทั้งสอง) ประมาณค่ามาจากการให้ความสำคัญกับระดับ**นัยสำคัญทางสถิติ (α) แต่เพียงอย่างเดียว** อย่างไรก็ตามยังมีระดับของ β **ที่นักวิจัยจะต้องให้ความสำคัญด้วยเหมือนกัน** หรือกล่าวอีกนัยหนึ่งคือ ในการกำหนดขนาดกลุ่มตัวอย่างนักวิจัยไม่ควรมองข้ามความสำคัญของอำนาจการทดสอบทางสถิติ ($1-\beta$) เพราะจะทำให้ผลการวิจัยที่พบภายหลังการทดสอบทางสถิติเชิงอนุมานมีทั้ง**นัยสำคัญทางสถิติและนัยสำคัญทางปฏิบัติ** สำหรับสูตรคำนวณขนาดกลุ่มตัวอย่างที่คำนึงถึงระดับ β (หรืออำนาจของการทดสอบทางสถิติ, $1-\beta$) คือ (Daniel, 1999, p.269; Guilford & Fruchter, 1978, p. 185; Lemeshow, Hosmer, Klar, & Lwanga, 1990, p. 37)

$$n = \frac{\sigma^2(Z_\alpha + Z_\beta)^2}{(ES)^2} \quad \text{.....(4)}$$

[เมื่อ n แทน ขนาดกลุ่มตัวอย่าง, σ^2 แทน ความแปรปรวนความคลาดเคลื่อนของประชากร, Z_α และ Z_β แทน คะแนนมาตรฐาน z ณ ระดับ α และ β ที่นักวิจัยกำหนดไว้ตามลำดับ และ ES แทนขนาดผลของการจัดกระทำทางการวิจัยซึ่งเทียบเคียงได้กับระดับความแม่นตรงที่นิยามในรูปของผลต่างระหว่างค่าเฉลี่ยของประชากรที่แท้จริงและที่คาดหมาย ($\mu-\mu_0$) อย่างไรก็ตามในปัจจุบันมักนิยมแสดงในรูปของหน่วยคะแนนมาตรฐาน ($ES = d\sigma$ เมื่อ $d = (\mu-\mu_0)/\sigma$]

การใช้สูตร (4) สำหรับคำนวณขนาดกลุ่มตัวอย่างตั้งอยู่บนข้อตกลงที่ว่าข้อมูลของตัวแปรเกณฑ์ที่ศึกษาต้องเป็นแบบต่อเนื่อง (continuous data) การทดสอบทางสถิติต้องเป็นแบบด้านเดียว (one-sided test) และประการสำคัญจะต้องใช้กับปัญหางานวิจัยที่ศึกษากับกลุ่มตัวอย่างกลุ่มเดียว (one-sample problem) ส่วนในกรณีที่นักวิจัยต้องการกำหนดขนาดกลุ่มตัวอย่างสำหรับการทดสอบแบบสองด้านกับงานวิจัยที่ศึกษากับกลุ่มตัวอย่างกลุ่มเดียว สามารถคำนวณได้จากสูตร (5) ดังนี้

$$n = \frac{\sigma^2(Z_{\alpha/2} + Z_\beta)^2}{(ES)^2} \quad \text{.....(5)}$$

หรือ

$$n = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$, เมื่อ ES = dσ ... (6)

ตัวอย่าง 1 : จากผลการวิจัยเชิงสำรวจในอดีตพบว่าน้ำหนักเฉลี่ยของผู้ชายที่มีอายุมากกว่า 55 ปีขึ้นไปและได้รับการวินิจฉัยเป็นโรคหัวใจเท่ากับ 90 กิโลกรัม โดยมีส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 20 กิโลกรัม อย่างไรก็ตามในปัจจุบันเชื่อว่าน้ำหนักเฉลี่ยของผู้ชายที่เป็นโรคหัวใจน่าจะต่ำกว่าผลการสำรวจที่ผ่านมา 4 กิโลกรัม นักวิจัยท่านหนึ่งต้องการสำรวจว่าความเชื่อดังกล่าวว่าเป็นความจริงหรือไม่ จึงทำการสุ่มตัวอย่างผู้ชายกลุ่มหนึ่งที่มีอายุ 55 ปีขึ้นไปสำหรับใช้ในการศึกษา ถ้านักวิจัยต้องการทดสอบ ณ ระดับนัยสำคัญที่ .05 ด้วยอำนาจของการทดสอบ .90 อยากทราบว่าขนาดกลุ่มตัวอย่างที่เพียงพอสำหรับการศึกษาควรเป็นเท่าไร

วิธีการ

จากสูตร (6)

$$n = \frac{\sigma^2(Z_{\alpha} + Z_{\beta})^2}{(ES)^2}$$

เมื่อ $\sigma=20; z_{\alpha}=z_{.05}=1.645, z_{\beta}=z_{.10}=1.282; ES=4$

แทนค่า $n = \frac{20^2(1.645+1.282)^2}{(4)^2}$

$= 214.18$

ดังนั้นกลุ่มตัวอย่างมีขนาดเท่ากับ 215 คน

จากตัวอย่างข้างต้นถ้าขนาดผลของการจัดกระทำแสดงอยู่ในรูปคะแนนมาตรฐาน (d) ซึ่งคำนวณจากสูตร $d=ES/\sigma$ หรือ $d=(\mu-\mu_0)/\sigma$ ดังนั้นจากเงื่อนไขที่กำหนดในตัวอย่าง ขนาดผลของการจัดกระทำมีค่าเท่ากับ 0.20 เท่าของความคลาดเคลื่อนมาตรฐาน [$ES=(4/20)\sigma=0.20\sigma$]

$$n = \frac{(Z_{\alpha} + Z_{\beta})^2}{d^2}$$

จากสูตร (6)

$$n = \frac{(1.645+1.282)^2}{(0.20)^2}$$

$= 214.18$

ทำนองเดียวกัน กลุ่มตัวอย่างที่เหมาะสมเท่ากับ 215 คน

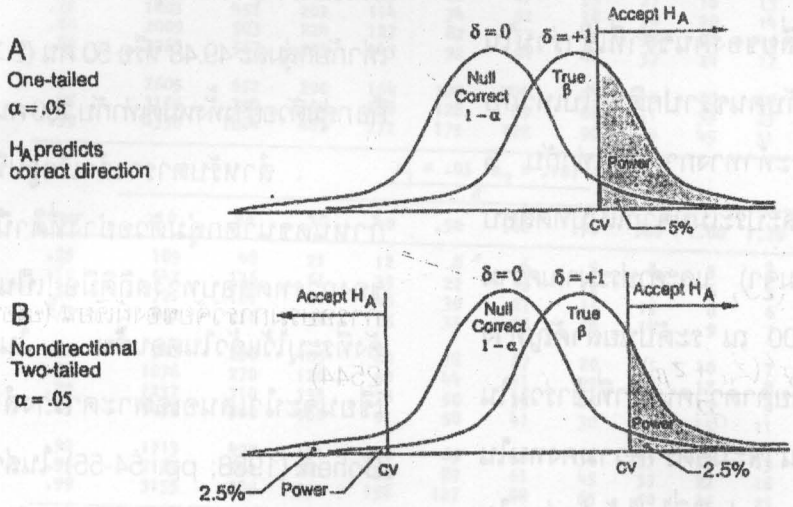
ข้อสังเกต : ถ้าผู้วิจัยคาดว่าน้ำหนักเฉลี่ยของผู้ชายในกลุ่มดังกล่าวอาจมีค่ามากกว่าหรือน้อยกว่า 90 กิโลกรัมก็ได้ ดังนั้นขนาดของกลุ่มตัวอย่างจะต้องประมาณจากสูตรที่ใช้สำหรับทดสอบสมมติฐานแบบสองด้าน (two-sided test) ดังนี้

$$n = \frac{(Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$

$$= \frac{(1.960 + 1.282)^2}{(.20)^2} = 262.76$$

ดังนั้นกลุ่มตัวอย่างที่เหมาะสมเท่ากับ 263 คน

จะพบว่ากลุ่มตัวอย่างที่ใช้สำหรับการทดสอบแบบสองด้านมีขนาดใหญ่กว่าการทดสอบแบบด้านเดียว หรือกล่าวอีกนัยหนึ่งคือ การทดสอบแบบสองด้านมีอำนาจของการทดสอบต่ำกว่าการทดสอบแบบด้านเดียว เมื่อปัจจัยที่เกี่ยวข้องอื่น ๆ มีระดับคงที่ กล่าวคือ มีขอบเขตของการไม่ปฏิเสธ H_0 ที่เป็นเท็จมากกว่าการทดสอบแบบด้านเดียว (ดูแผนภาพ 3 ประกอบคำอธิบาย)



แผนภาพ 3 : เปรียบเทียบอำนาจของการทดสอบทางสถิติระหว่างการทดสอบแบบด้านเดียว (A) และแบบสองด้าน (B) (หมายเหตุ : δ คือ $d = \mu - \mu_0$)

อนึ่ง ในกรณีปัญหาการวิจัยศึกษากับกลุ่มตัวอย่างจำนวน 2 กลุ่ม สามารถกระทำได้อย่างง่าย ๆ ด้วยการนำ 2 ไปคูณกับขนาดกลุ่มตัวอย่างที่คำนวณมาจากสูตร (4) หรือ (5) ขึ้นอยู่กับว่าการทดสอบเป็นแบบด้านเดียวหรือสองด้าน หรือกล่าวอย่างง่าย ๆ คือ กลุ่มตัวอย่างสำหรับปัญหาการวิจัยที่ศึกษากับกลุ่มตัวอย่างจำนวนสองกลุ่ม (two – sample problem) มีขนาดเป็นสองเท่าของกลุ่มตัว

อย่างที่ศึกษาปัญหาการวิจัยที่ศึกษากับกลุ่มตัวอย่างกลุ่มเดียว (one – sample problem)

ตัวอย่าง 2 : สมมุติว่านักวิจัยทางพฤติกรรมศาสตร์คนหนึ่งต้องการศึกษาเปรียบเทียบความคงทนในการจำข้อมูลสารสนเทศของคนชราอายุระหว่าง 65-75 ปีที่เข้าร่วมในโครงการฝึกสมาธิของบ้านพักคนชราแห่งหนึ่ง ถ้านักวิจัยต้องการให้มีอำนาจของการทดสอบเท่ากับ .80 สำหรับตรวจพบความแตกต่างระหว่างคะแนนเฉลี่ยของคนชราที่เข้าร่วมในโครงการฝึกสมาธิกับคนชราปกติทั่วไปเท่ากับ 5 คะแนน จัดกระทำทางการวิจัยเท่ากับ 5 คะแนน (โดยวัดและประเมินจากแบบทดสอบมาตรฐานด้านความจำ) และค่าประมาณความแปรปรวนเท่ากับ 100 ณ ระดับนัยสำคัญทางสถิติ .05 ถ้านักวิจัยคาดว่าคนชราที่เข้าร่วมในโครงการฝึกสมาธิน่าจะมีอัตราความคงทนในการจำสูงกว่าคนชราปกติที่ไม่ได้เข้าร่วมในโครงการดังกล่าว อยากทราบว่ากลุ่มตัวอย่างที่พอเพียงควรเป็นเท่าไร

วิธีการ

จากสูตร (4)

$$n = \frac{\sigma^2(z_\alpha + z_\beta)^2}{(ES)^2}$$

หรือ $$= \frac{\sigma^2(z_\alpha + z_\beta)^2}{d^2}$$

เมื่อ

$$\sigma^2 = 100;\ z_\alpha = z_{.05} = 1.645;$$
$$z_\beta = z_{.20} = 0.842,\ ES = 5$$

ดังนั้น $$n = \frac{100(1.645 + 0.842)^2}{(5)^2}$$

$$= 24.74$$

ดังนั้นกลุ่มตัวอย่างที่พอเพียงสำหรับปัญหาการวิจัยที่ต้องการศึกษากับกลุ่มตัวอย่างจำนวน 2 กลุ่ม คือ กลุ่มทดลองและกลุ่มควบคุม เท่ากับกลุ่มละ 49.48 หรือ 50 คน (2 × 24.74) นั่นคือ กลุ่มตัวอย่างทั้งหมดเท่ากับ 100 คน

สำหรับตารางสำเร็จรูปที่ช่วยในการกำหนดขนาดกลุ่มตัวอย่างที่คำนึงถึงอำนาจของการทดสอบทางสถิติมีอยู่เป็นจำนวนมาก ดังที่ระบุไว้แล้วในตอนต้น ในบทความนี้ผู้เขียนจะนำเสนอเฉพาะตารางสำเร็จรูปของ Cohen (1988; pp. 54-55) ในส่วนที่เกี่ยวกับการกำหนดขนาดกลุ่มตัวอย่างสำหรับการวิจัยจำนวน 2 กลุ่ม (กลุ่มทดลองและกลุ่มควบคุม) โดยใช้สถิติ t ทดสอบสมมุติฐานการวิจัย ส่วนการอธิบายการใช้ตารางสำเร็จรูปอื่น ๆ ผู้สนใจอาจศึกษาได้จากบทความของ ดร.ผจงจิต อินทสุวรรณ (2534) หรือจากเอกสารประกอบการอบรมการวิจัยของผู้เขียน (องอาจ นัยพัฒน์, 2544)

ตาราง 1 : ตารางสำเร็จรูปช่วยกำหนดขนาดกลุ่มตัวอย่างของ Cohen (1988, pp. 54-55) สำหรับการวิจัยที่ใช้สถิติ t ทดสอบ)

Table 2.4.1

n to detect d by t test

	a_1 = .01 (a_2 = .02)										
	d										
Power	.10	.20	.30	.40	.50	.60	.70	.80	1.00	1.20	1.40
.25	547	138	62	36	24	17	13	10	7	5	4
.50	1083	272	122	69	45	31	24	18	12	9	7
.60	1332	334	149	85	55	38	29	22	15	11	8
2/3	1552	382	170	97	62	44	33	25	17	12	9
.70	1627	408	182	103	66	47	35	27	18	13	10
.75	1803	452	202	114	74	52	38	30	20	14	11
.80	2009	503	224	127	82	57	42	33	22	15	12
.85	2263	567	253	143	92	64	48	37	24	17	13
.90	2605	652	290	164	105	74	55	42	27	20	15
.95	3155	790	352	198	128	89	66	51	33	23	18
.99	4330	1084	482	272	175	122	90	69	45	31	23

	a_1 = .05 (a_2 = .10)										
	d										
Power	.10	.20	.30	.40	.50	.60	.70	.80	1.00	1.20	1.40
.25	189	48	21	12	8	6	5	4	3	2	2
.50	542	136	61	35	22	16	12	9	6	5	4
.60	721	181	81	46	30	21	15	12	8	6	5
2/3	862	216	96	55	35	25	18	14	9	7	5
.70	942	236	105	60	38	27	20	15	10	7	6
.75	1076	270	120	68	44	31	23	18	11	8	6
.80	1237	310	138	78	50	35	26	20	13	9	7
.85	1438	360	160	91	58	41	30	23	15	11	8
.90	1713	429	191	108	69	48	36	27	18	13	10
.95	2165	542	241	136	87	61	45	35	22	16	12
.99	3155	789	351	198	127	88	65	50	32	23	17

	a_1 = .10 (a_2 = .20)										
	d										
Power	.10	.20	.30	.40	.50	.60	.70	.80	1.00	1.20	1.40
.25	74	19	9	5	3	3	2	2	2	2	2
.50	329	82	37	21	14	10	7	5	4	3	2
.60	471	118	53	30	19	14	10	8	5	4	3
2/3	586	147	65	37	24	17	12	10	6	4	3
.70	653	163	73	41	27	19	14	11	7	5	4
.75	766	192	85	48	31	22	16	13	8	6	4
.80	902	226	100	57	36	26	19	14	10	7	5
.85	1075	269	120	67	43	30	22	17	11	8	6
.90	1314	329	146	82	53	37	27	21	14	10	7
.95	1713	428	191	107	69	48	35	27	18	12	9
.99	2604	651	290	163	104	73	53	41	26	18	14

Table 2.4.1 *(continued)*

	a_2 = .01 (a_1 = .005)										
	d										
Power	.10	.20	.30	.40	.50	.60	.70	.80	1.00	1.20	1.40
.25	725	183	82	47	31	22	17	13	9	7	6
.50	1329	333	149	85	55	39	29	22	15	11	9
.60	1603	402	180	102	66	46	34	27	18	13	10
2/3	1810	454	203	115	74	52	39	30	20	14	11
.70	1924	482	215	122	79	55	41	32	21	15	12
.75	2108	528	236	134	86	60	45	35	23	17	13
.80	2338	586	259	148	95	67	49	38	25	18	14
.85	2611	654	292	165	106	74	55	43	28	20	15
.90	2978	746	332	188	120	84	62	48	31	22	17
.95	3564	892	398	224	144	101	74	57	37	26	20
.99	4808	1203	536	302	194	136	100	77	50	35	26
	a_2 = .05 (a_1 = .025)										
	d										
Power	.10	.20	.30	.40	.50	.60	.70	.80	1.00	1.20	1.40
.25	332	84	38	22	14	10	8	6	5	4	3
.50	769	193	86	49	32	22	17	13	9	7	5
.60	981	246	110	62	40	28	21	16	11	8	6
2/3	1144	287	128	73	47	33	24	19	12	9	7
.70	1235	310	138	78	50	35	26	20	13	10	7
.75	1389	348	155	88	57	40	29	23	15	11	8
.80	1571	393	175	99	64	45	33	26	17	12	9
.85	1797	450	201	113	73	51	38	29	19	14	10
.90	2102	526	234	132	85	59	44	34	22	16	12
.95	2600	651	290	163	105	73	54	42	27	19	14
.99	3675	920	409	231	148	103	76	58	38	27	20

ตัวอย่าง 3 จากสภาพการณ์เงื่อนไขของการวิจัยที่กล่าวในตัวอย่างที่ 2 จงกะประมาณขนาดกลุ่มตัวอย่างที่พอเพียงสำหรับการวิจัยโดยการใช้ตารางสำเร็จรูปของ Cohen (1998)

วิธีการ

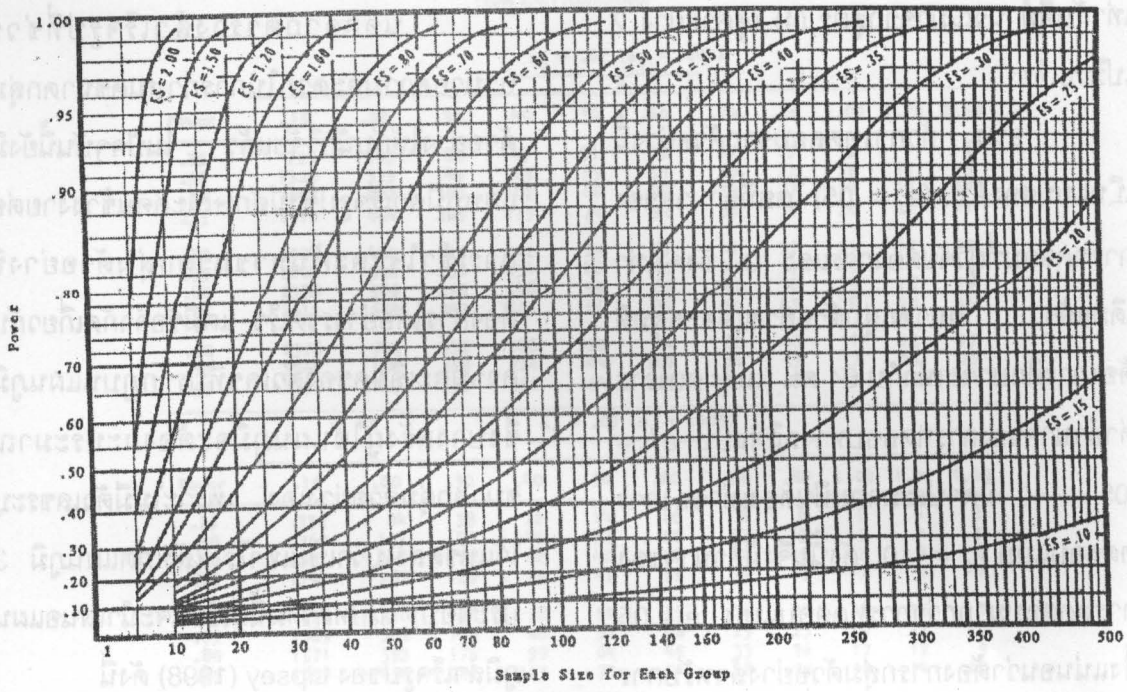
จากสภาพการณ์เงื่อนไขของการวิจัยในตวอย่างที่ 2 จะพบว่าอำนาจของการทดสอบ $(1-\beta) = .80$, $d = .50$ $(d=5/10)$ และระดับ α หรือ $a_1 = .05$ (หมายเหตุ : a_1 และ a_2 ที่แสดงไว้ในตารางของ Cohen คือ ระดับ α สำหรับการทดสอบแบบด้านเดียวและสองด้านตามลำดับ โดยที่อำนาจของการทดสอบแบบด้านเดียวมีค่าประมาณสองเท่าของการทดสอบแบบสองด้าน) ดังนั้นขนาดกลุ่มสำหรับการทดลองและควบคุม กลุ่มละ 50 คน (N = 100 คน)

ข้อสังเกต : จะพบว่าขนาดกลุ่มตัวอย่างที่กำหนดจากตารางของ Cohen (1988)

เท่ากับที่คำนวณมาจากสูตร (4) (ดูตัวอย่าง 2 เปรียบเทียบ)

อนึ่ง ถ้าการทดสอบจากตัวอย่างนี้เป็นแบบสองด้าน (a_2 = .05) โดยที่อำนาจของการทดสอบที่ผู้วิจัยต้องการและ α ยังคงเท่าเดิมแล้ว ขนาดกลุ่มตัวอย่างแต่ละกลุ่มที่ต้องการศึกษาจะเท่ากับ 64 คน (เหตุผลเพราะค่าอำนาจของการทดสอบทางสถิติแบบ a_1 = .05 มีค่าประมาณเป็นสองเท่าของการทดสอบแบบ a_2 = .10 ดังนั้นจึงมีอำนาจของการทดสอบมากกว่าการทดสอบแบบ a_2=.05 ซึ่งแน่นอนว่าต้องการกลุ่มตัวอย่างสำหรับการวิจัยมากกว่า – สังเกตตาราง 1 ประกอบ)

นอกจากตารางสำเร็จรูปที่ช่วยอำนวยความสะดวกในการกำหนดขนาดกลุ่มตัวอย่างให้กับนักวิจัยแล้ว ในปัจจุบันนี้ยังมีแผนภูมิสำเร็จรูปซึ่งมีลักษณะค่อนข้างง่ายต่อการเข้าใจและค้นหาขนาดกลุ่มตัวอย่างที่ต้องการได้อย่างรวดเร็ว แต่มีข้อจำกัดเกี่ยวกับความละเอียดของตัวเลขที่ปรากฏบนแผนภูมิ ซึ่งบางครั้งผู้ใช้แผนภูมิจะต้องกะประมาณขนาดกลุ่มตัวอย่างเอง เพราะไม่มีตัวเลขระบุไว้แน่ชัดลงไปว่าเป็นเท่าไร (สังเกตแผนภูมิ 3 ประกอบ) ในบทความนี้ผู้เขียนจะนำเสนอแผนภูมิสำเร็จรูปของ Lipsey (1998) ดังนี้



แผนภูมิ 3 : แผนภูมิสำเร็จรูปช่วยในการกำหนดขนาดกลุ่มตัวอย่างของ Lipsey (1998, p.44) สำหรับการวิจัยที่ใช้สถิติ t, one-way ANOVA หรือพาราเมตริกอื่นที่อยู่ในตระกูลเดียวกัน ทดสอบ ณ ระดับ α = .05 (แบบสองด้าน) หรือ α = .025 (แบบด้านเดียว)

ตัวอย่าง 4 : จากสภาพการณ์เงื่อนไขของการวิจัยที่กล่าวมาในตัวอย่างที่ 2 จงกะประมาณขนาดกลุ่มตัวอย่างโดยใช้แผนภูมิสำเร็จรูปของ Lipsey (1998)

วิธีการ เนื่องจากระดับอำนาจของการทดสอบทางสถิติที่ต้องการ = .80 d = .50 (ES = dσ) และ α ช .05 (แบบสองด้าน) จะได้ขนาดกลุ่มตัวอย่างประมาณกลุ่มละ 64 คน (อ่านค่า Power บนแกน Y = .80 จากนั้นลากเส้นขนานกับแกน X ไปตัดบนเส้นกราฟ ES = .50 แล้วลากเส้นขนานกับแกน Y มาตัดแกน X จะได้ขนาดกลุ่มตัวอย่างประมาณ 64 คน) ซึ่งเท่ากับจำนวนที่ได้รับจากตารางสำเร็จรูปของ Cohen (1988) ดังแสดงไว้ในตัวอย่าง 3

บทสรุป

กลุ่มตัวอย่างสำหรับวิจัยควรถูกกำหนดขึ้นโดยคำนึงถึงความคลาดเคลื่อนที่อาจเกิดขึ้นได้จากการตัดสินใจทางสถิติทั้งสองประเภท (α และ β) หรือกล่าวอีกนัยหนึ่งคือนักวิจัยจะต้องพิจารณาทั้งความมีนัยสำคัญทางสถิติและความสำคัญทางปฏิบัติให้สมดุลกันในการกำหนดขนาดกลุ่มตัวอย่างสำหรับการวิจัย การกำหนดขนาดกลุ่มตัวอย่างที่ใหญ่มากเกินไปจะทำให้มีโอกาสเป็นไปได้เสมอที่จะปฏิเสธสมมุติฐานกลางตลอดเวลา เนื่องจากกลุ่มตัวอย่างขนาดใหญ่มีอำนาจหรือความไวของการทดสอบทางสถิติต่อการตรวจพบขนาดผลของการจัดกระทำทางการวิจัยที่อาจจะไม่มีคุณค่าหรือประโยชน์ใด ๆ ในทางปฏิบัติ ดังนั้นนักวิจัยจะต้องหลีกเลี่ยงหรือระมัดระวังเป็นอย่างยิ่งในการรายงานผลการวิจัยที่มีนัยสำคัญทางสถิติสูงมาก (เช่น $p < .001$) ซึ่งได้รับจากการศึกษากับกลุ่มตัวอย่างขนาดใหญ่ ตามทรรศนะของผู้เขียนผลการดำเนินการจัดกระทำทางการวิจัย (effect sige) ที่มีขนาดใหญ่พอสมควรที่จะสามารถถูกตรวจพบว่ามีนัยสำคัญทางสถิติในการศึกษาวิจัยกับกลุ่มตัวอย่างขนาดเล็ก มีคุณค่าควรแก่การได้รับความสนใจมากกว่าผลการจัดกระทำทางการวิจัยที่มีขนาดเล็กซึ่งพบว่ามีนัยสำคัญทางสถิติอันเกิดจากการทำการศึกษากับกลุ่มตัวอย่างขนาดใหญ่

อำนาจของการทดสอบทางสถิติเป็นปัจจัยสำคัญที่นักวิจัยควรพิจารณาในการกำหนดขนาดกลุ่มตัวอย่าง การกำหนดกลุ่มตัวอย่างที่เหมาะสม หรือมีอำนาจของการทดสอบทางสถิติพอเหมาะต่อการตรวจพบขนาดผลของการจัดกระทำทางการวิจัยที่คาดว่ามีความสำคัญในทางปฏิบัติหรือความสนใจของนักวิจัย เป็นสิ่งที่มีความสำคัญเป็นอย่างยิ่งที่จะทำให้ผลของการวิจัยที่ค้นพบมีความหมายและมีความสำคัญไม่เพียงแต่ในทางสถิติแต่ในทางปฏิบัติอีกด้วย ผู้เขียนหวังว่าบทความนี้คงมีส่วนช่วยกระตุ้นให้นิสิต/นักศึกษาระดับบัณฑิตศึกษาและนักวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์ทั่วไป หันมาให้ความสำคัญกับ β และ $1-\beta$ บ้าง นอกเหนือไปจาก α และผลการวิจัยที่มีนัยสำคัญทางสถิติ (แต่หามีคุณค่าหรือความสำคัญในทางปฏิบัติไม่)

บรรณานุกรม

ผจงจิต อินทสุวรรณ.(2534). การใช้ตารางเพื่อช่วยในการกำหนดขนาดของกลุ่มตัวอย่าง. **วารสารวัดผลการศึกษา**, 38, 35 – 59.

องอาจ นัยพัฒน์.(2544). วิธีการสุ่มตัวอย่างและกำหนดขนาดตัวอย่างสำหรับการวิจัย ใน พ.ต.หญิง ดร.วาสนา นัยพัฒน์ (บรรณาธิการ), **การวิจัยทางการพยาบาล** (หน้า 90 – 136) เอกสารประกอบการอบรมวิจัยเชิงปฏิบัติการ (อัดสำเนา) วิทยาลัยพยาบาลกองทัพบก

Blalock, H. M., Jr. (1981). **Social statistics.** (2nd ed.). New York : McGraw-Hill.

Cocharn, W.G. (1977). **Sampling techniques.** (3rd ed.). New York : John Wiley & Sons.

Cohen, J. (1988). **Statistical power analysis for behavioral sciences.** (2nd ed.). Hillsdale, NJ : Lawrence Erlbaum.

Daniel, W.W. (1999). **Biostatistics : A foundation of analysis in health sciences.** (7th ed.). New York : John Wiley & Sons.

Guilford, J.P., & Frutcher, B. (1978). **Fundamental statistics in psychology and education.** (5th ed.). New York : McGraw-Hill.

Hinkel, D.E., Wiersma, W., & Jurs, S.G. (1994). **Applied statistics for the behavioral sciences.** (3rd ed.). Boston, MA : Houghton Mifflin.

Kramer, H.C., & Thiemann, S. (1987). **How many subjects: Statistical power analysis in research.** Newbury Park, CA : Sage.

Krejcie, R., & Morgan, D.W. (1970). Determining sample size for research activities. **Educational and Psychological Measurement. 30**, 607 – 610.

Lemeshow, S., Hosmer, D.W., Jr., Klar, J., & Lwanga, S. (1990). **Adequacy of sample size in health studies.** New York : John Wiley & Sons.

Lipsey, M.W. (1998). Design sensitivity : Statistical power for experimental research. In L. Bickman & D.J. Rog (Eds.), **Handbook of applied social research methods** (pp.39 – 68). Thousand Oaks, CA :Sage.

Murphy, K.R., & Myors, B. (1998). **Statistical power analysis : A simple and general model for traditional and modern hypothesis tests.** Mahwah, NJ : Lawrence Erlbaum.

Oakes, J.M., & Feldman, H.A. (2001). Statistical power for nonequivalent pretest – postest designs : The impact of change – score versus ANCOVA model. **Evaluation review, 25,** 3 – 28.

Olejnik, S.F. (1984). Planning educational research determining the necessary sample size. **Journal of Experimental Education, 53,** 40 – 48.

Yamane, T. (1967). **Statistics : An introductory analysis.** New York : Harper & Row.