ขนาดอิทธิพลและอำนาจทดสอบในการกำหนดขนาดตัวอย่าง สำหรับการวิจัยทางคลินิก

วัฒนา ชยธวัช[1*]

[1] หลักสูตรการแพทย์แผนไทยบัณฑิต คณะสาธารณสุขศาสตร์และสิ่งแวดล้อม มหาวิทยาลัยปทุมธานี

ผู้นิพนธ์ที่ให้การติดต่อ E-mail: vadhana.j@ptu.ac.th

บทคัดย่อ

ขนาดอิทธิพลและอำนาจทดสอบนำมาใช้ในการกำหนดขนาดตัวอย่างสำหรับการวิจัยทางคลินิกที่ใช้ขนาดตัวอย่างน้อยกว่าการวิจัยเชิงสำรวจ การกำหนดขนาดตัวอย่าง (n) ที่เหมาะสมสามารถศึกษาได้จากสื่อออนไลน์ เช่น G*Power Power and Sample Size .com Calculators และ Sample Size Calculators เป็นต้น การใช้โปรแกรมสำเร็จรูปเหล่านี้ ต้องมีการเลือกประเภทของการทดสอบสมมติฐานทางสถิติ การใส่ค่า ประชากร (N) ค่าระดับนัยสำคัญ (significance level α) หรือ ความคลาดเคลื่อนชนิดที่ 1 (α) หรือค่าช่วงความเชื่อมั่น (confidence interval 1-α) ค่าความคลาดเคลื่อนชนิดที่ 2 (β) หรือ อำนาจทดสอบ (1-β) และค่าขนาดอิทธิพล (effect size δ) ซึ่งล้วนเป็นเครื่องมือที่จำเป็นสำหรับนักวิจัย บทความนี้นำเสนอนิยามศัพท์และวิธีการทางสถิติที่ทำให้เกิดสูตรการกำหนดขนาดตัวอย่าง เพื่อให้สามารถนำไปใช้งานโปรแกรมสำเร็จรูปเหล่านั้นได้อย่างเหมาะสม

คำสำคัญ : การวิจัยทางคลินิก, ขนาดตัวอย่าง, ขนาดอิทธิพล, อำนาจทดสอบ

Effect Size and Power of a Test in Sample Size Determination for Clinical Research

Vadhana Jayathavaj [1*]

[1] Thai Traditional Medicine Curriculum, Faculty of Public Health and Environment, Pathumthani University

Corresponding Author E-mail: vadhana.j@ptu.ac.th

Abstract

Effect size and power of a test are used in sample size determination and in clinical research which the sample size is smaller than the survey research. The sample size calculators are available online i.e.; G*Power, Power and Sample Size .com Calculators, and Sample Size Calculators. The input parameters to these programs are the type of test statistics, number of population, significance level (confidence interval 1-α), type II error (β) or power of a test (1-β), and effect size (δ), these are very important tools for researchers. This article presented terms and definitions and derived the sample size formula for researchers to appropriate use of those sample size calculation programs.

Keyword: Clinical Research, Effect Size, Power of Test, Sample Size

บทนำ

การจัดประเภทการวิจัยขึ้นกับองค์ประกอบของการออกแบบการศึกษา (Study design elements) คือ การจัดกระทำ การควบคุม และกระบวนการสุ่ม (manipulation control and randomization)[1] การออกแบบงานวิจัยแบ่งออกเป็นการวิจัยเชิงทดลองและการวิจัยแบบไม่ทดลอง (experimental research and non-experimental research) การวิจัยเชิงทดลองเป็นการเปรียบเทียบกลุ่มควบคุมกับกลุ่มทดลองที่ได้รับการจัดกระทำ (การดำเนินการกับตัวแปรต่าง ๆ) ซึ่งใช้ในการวิจัยทางการแพทย์ ชีววิทยา และอื่น ๆ การวิจัยแบบไม่ทดลองมันเป็นการศึกษาเชิงพรรณนาหรือความสัมพันธ์ เป็นเพียงการอธิบายเหตุการณ์หรือปรากฏการณ์ที่เป็นอยู่โดยไม่มีการเข้าไปดำเนินการใด ๆ โดยนักวิจัย ซึ่งหมายถึงว่าไม่มีการจัดกระทำกับตัวแปรต่าง ๆ โดยนักวิจัย แต่เน้นที่ความเที่ยงตรงของการวัดมากกว่าความเที่ยงตรงของผลที่เกิดขึ้น (Statistics Solutions, 2018)[2] การวิจัยเชิงทดลองมีหลายประเภทเช่น ก่อน แท้จริง กึ่ง เป็นต้น (pre, true, quasi, etc.) ซึ่งขึ้นกับขนาดของการกำหนดควบคุมในวิธีการทดลอง[3] การวิจัยแบบไม่ทดลอง เช่น การศึกษาเชิงสำรวจ (survey study) การศึกษาความสัมพันธ์ (relationship study) การศึกษาเฉพาะกรณี (case study) และ การศึกษาติดตามผล (follow – up study) เป็นต้น การกำหนดขนาดตัวอย่าง (n) ที่เหมาะสมสามารถศึกษาได้จากสื่อออนไลน์มากมาย เช่น G*Power[4] Power and Sample Size .com Calculators[5] และ Sample Size Calculators[6] เป็นต้น การใช้โปรแกรมสำเร็จรูปเหล่านี้ ต้องมีการเลือกประเภทของการทดสอบสมมติฐานทางสถิติ การใส่ค่า ประชากร (N) ค่า ระดับนัยสำคัญ (significance level α) หรือ ความคลาดเคลื่อนชนิดที่ 1 (α) หรือค่าช่วงความเชื่อมั่น (confidence interval 1-α) หรือ ค่าความคลาดเคลื่อนชนิดที่ 2 (β) หรือ อำนาจทดสอบ (1-β) และค่าขนาดอิทธิพล (effect size) ซึ่งล้วนเป็นเครื่องมือที่จำเป็นสำหรับนักวิจัย

วัตถุประสงค์การวิจัย

บทความนี้นำเสนอนิยามศัพท์และวิธีการทางสถิติที่ทำให้เกิดสูตรการกำหนดขนาดตัวอย่าง เพื่อให้นักวิจัยเกิดความเข้าใจและสามารถนำไปใช้งานได้อย่างเหมาะสม

การคำนวณขนาดตัวอย่างโดยไม่รู้จำนวนประชากร

ผู้วิจัยเพียงแต่ทราบว่ามีประชากรจำนวนมาก และนำไปใช้กับการวิจัยแบบไม่ทดลอง โดยใช้สูตรตามสมการที่ 1[7,8]

$$n = \frac{P(1-P)Z^2}{e^2} \qquad (1)$$

n คือ จำนวนหน่วยตัวอย่าง
P คือ สัดส่วนของประชากรที่ผู้วิจัยกำหนดจะสุ่มเลือก
e คือ ความคลาดเคลื่อนจากการสุ่มตัวอย่าง
Z คือ ค่ามาตรฐานการแจกแจงปกติ ใช้ Z ณ ระดับความเชื่อมั่นที่กำหนด

เช่น 95% (1-α = 0.95, α = 0.05) $Z_{\pm 0.025} = \pm 1.96$

หรือ 99% (1-α = 0.99, α =0.01, $\frac{\alpha}{2}$ =0.005) $Z_{\pm 0.005} = \pm 2.58$

ตัวอย่างที่ 1 ถ้าสัดส่วนของประชากรที่สนใจจะศึกษาเท่ากับ 0.40 ต้องการความเชื่อมั่นร้อยละ 99 และยอมให้มีความคลาดเคลื่อนของการสุ่มตัวอย่างได้ร้อยละ 5

วิธีทำ P = 0.40 $Z_{\pm 0.005} = \pm 2.58$ e = 0.05

n = 0.40 (1-0.40) $2.58^2/0.05^2$ = 639.01 ประมาณ 640 หน่วย

จะพบว่า ที่ค่า P = 0.50 ค่า $P(1-P)$ มีค่าสูงสุด และจะทำให้ได้จำนวนหน่วยตัวอย่างมากที่สุด คือ n = 666 หน่วย

จากตัวอย่างข้างต้น P = 0.50 เมื่อใช้ความเชื่อมั่นร้อยละ 95 $Z_{\pm 0.025} = \pm 1.96$ e = 0.05 เช่นเดิมจะได้ n = 384.16 ตัวอย่าง ซึ่งเป็นตัวเลขที่คุ้นตา เนื่องจากหากไม่รู้จำนวนประชากรแล้วก็มักใช้ขนาดตัวอย่างประมาณ 400 ตัวอย่างกันเป็นจำนวนมาก

สูตรตามสมการที่ 1 นี้มีที่มาจากการแจกแจงแบบปกติมาตรฐาน (Standard normal distribution) และการแจกแจงแบบทวินาม (Binomial distribution) จากการกำหนดช่วงความเชื่อมั่น 95% ของสัดส่วนประชากร P โดยมี p เป็นสัดส่วนของตัวอย่างที่ใช้ประมาณ P ถ้า n มีขนาดใหญ่พอก็สามารถใช้ประมาณการแจกแจงแบบปกติประมาณการแจกแจงสัดส่วนของ P ได้

โดยเริ่มจากการแปลงค่า p เป็นค่ามาตรฐาน $Z_p = \frac{p-P}{\sigma_p}$ (เหมือนกับ $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$)

เมื่อ σ_p คือ ส่วนเบี่ยงเบนมาตรฐานของค่าสัดส่วนของตัวอย่าง

ที่ช่วงความเชื่อมั่นที่กำหนด เช่น 95%

ความน่าจะเป็นที่ Z_p อยู่ระหว่าง -1.96 < Z_p < +1.96 แทนค่า Z_p จะได้

-1.96 σ_p < $p-P$ < +1.96σ_p

ซึ่ง $p-P$ ก็คือ e ความแตกต่างของสัดส่วนของตัวอย่างกับสัดส่วนประชากร ซึ่งก็คือค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง

ดังนั้น $e^2 = 1.96^2\sigma_p^2$

เมื่อ σ_p คือ ส่วนเบี่ยงเบนมาตรฐานของสัดส่วนของตัวอย่าง

เนื่องจากความแปรปรวนของ p คือ $V(p) = \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$ ตามทฤษฎีของ Cochran[9] เมื่อ N มีขนาดใหญ่เมื่อเทียบกับ n แล้ว $\left(\frac{N-n}{N-1}\right)$ มีค่าเข้าใกล้ 1 และ $P(1-P)$ สามารถประมาณค่าด้วยสัดส่วนจากตัวอย่าง $p(1-p)$ ดังนั้น $V(p) = \frac{p(1-p)}{n}$ หรือ ในที่นี้ $V(p)$ ก็คือ σ_p^2

ดังนั้น $e^2 = \frac{1.96^2\, p(1-p)}{n}$

หรือ $$n = \frac{1.96^2 p(1-p)}{e^2}$$

หรือ สำหรับค่า α ใด ๆ $$n = \frac{Z^2_{\alpha/2} p(1-p)}{e^2} \quad (1)$$

การคำนวณขนาดตัวอย่าง เมื่อรู้จำนวนประชากร

จากสมการที่ 1 ได้มีการปรับปรุงเป็นสูตรของทาโร ยามาเน่[10] ที่ช่วงความเชื่อมั่น 95% ดังนี้

$$n = \frac{N}{1+Ne^2} \quad (2)$$

โดยการคูณทางขวาของสมการที่ 1 ด้วย finite population correction (fpc) เท่ากับ $(N-n)/N$ ซึ่งจะมีค่าใกล้กับ 1 เมื่อ N มีขนาดใหญ่เมื่อเทียบกับ n ที่ช่วงความเชื่อมั่น 95% $Z_{\alpha/2}$ =1.96 ซึ่งใกล้กับ 2 และ p = 0.50 ค่า $p(1-p)$ มีค่าสูงสุด = 0.25 (ก็คือ 2×2=4 และ 4÷0.25 = 1) ทำให้ได้สมการที่ 2 ดังกล่าว

ตัวอย่างที่ 2 ถ้ามีกลุ่มประชากรที่สนใจศึกษาจำนวน 3,000 คน ยอมให้เกิดความคลาดเคลื่อนจากการสุ่มตัวอย่างร้อยละ 5 หรือ 0.05 แล้ว ขนาดของกลุ่มตัวอย่างควรเป็นเท่าไร

วิธีทำ จากโจทย์ N = 3,000 e = 0.05

n = 3000/(1+3000x0.05^2) = 352.9412 หรือ 353 คน ที่ช่วงความเชื่อมั่น 95%

การคำนวณขนาดตัวอย่างในการวิจัยเชิงทดลอง

การกำหนดขนาดตัวอย่างในการวิจัยเชิงทดลอง โดยเฉพาะการวิจัยทางคลินิก ซึ่งกำหนดขนาดตัวอย่างจำนวนน้อยกว่าการวิจัยเชิงสำรวจที่แสดงในตัวอย่างที่ 1 และตัวอย่างที่ 2 เช่น การศึกษาผลการรักษาอาการท้องผูกด้วยการนวดไทยที่มีการเปรียบเทียบจำนวนครั้งที่ถ่ายอุจจาระต่อสัปดาห์ก่อนและหลังการรักษาโดยใช้อาสาสมัคร (ตัวอย่าง) เพียง 23 คน จากที่กำหนดไว้ 30 คน[11] ซึ่งสามารถอธิบายด้วยค่านิยามทางสถิติเพิ่มขึ้น คือ ขนาดอิทธิพล (effect size) และ อำนาจทดสอบ (power of test)

ขนาดอิทธิพล (effect size) เป็นสัดส่วนของความแตกต่างของค่าเฉลี่ยกับส่วนเบี่ยงเบนมาตรฐาน ในการทดลองเพื่อเปรียบเทียบกับค่าที่กำหนดหรือการเปรียบเทียบความแตกต่างการกำหนดขนาดอิทธิพลเป็นเรื่องยาก อาจต้องใช้ข้อมูลจากการศึกษาที่ผ่านมา หรือ ทำการทดลองนำร่อง (pilot study) หรือใช้ความเห็นของผู้เชี่ยวชาญ[12] โดยมีสูตรดังนี้

$$\delta = \frac{(\mu_1 - \mu_0)}{\sigma} \quad (3)$$

$\mu_1 - \mu_0$ คือ ความแตกต่างของค่าเฉลี่ย

σ คือ ความแปรปรวนของประชากร

ค่าของขนาดอิทธิพล[13,14] ที่กำหนดว่ามีความแตกต่างกันน้อย ปานกลาง และมาก สามารถนำไปใช้ในสูตรการคำนวณขนาดตัวอย่างในการวิจัยเชิงทดลองได้ กำหนดดังนี้

มีความแตกต่างกัน	น้อย	ปานกลาง	มาก
	0.20	0.50	0.80
	20%	50%	80%

เมื่อขนาดอิทธิพลมีค่าน้อยก็จะใช้จำนวนตัวอย่างมากขึ้น

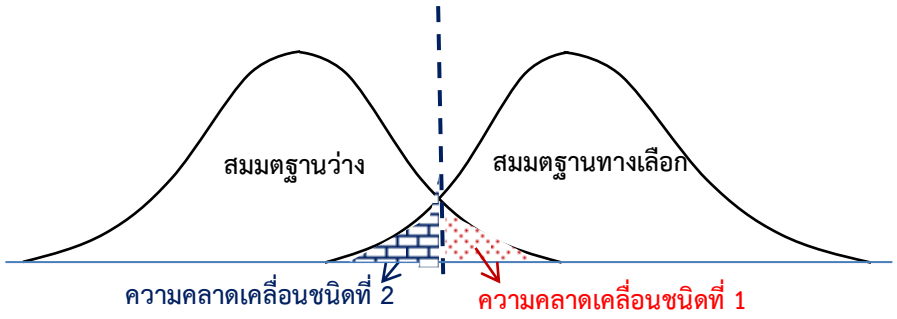
อำนาจทดสอบ (power of a test $1-\beta$)[14]

อำนาจทดสอบ คือ $1-\beta$ หรือ 1 ลบด้วยความคลาดเคลื่อนชนิดที่ 2 เมื่อความคลาดเคลื่อนชนิดที่ 2 คือการไม่ปฏิเสธสมมตฐานว่างเมื่อสมมตฐานว่างเป็นเท็จ ซึ่งในการการทดสอบสมมตฐาน ระดับนัยสำคัญ (α) ถ้า α = 0.05 คา Z จะมีค่า 1.96 (2-sided type หรือ 2-tailed type) และ 1.645 (1-sided type หรือ 1-tailed type) ความคลาดเคลื่อนที่ยอมรับได้ในสมมติฐานที่จะทดสอบหรือ type I error (α) (ค่า confidence level คือ 1 - α) และ อำนาจทดสอบ (power of a test $1-\beta$) ที่ 80% หรือ 0.80 หรือ β = 0.20 นั่นเอง

ความคลาดเคลื่อนชนิดที่ 1 ความคลาดเคลื่อนชนิดที่ 2 กับการตัดสินใจในการทดสอบสมมติฐานดังแสดงในตารางที่ 1 ส่วนความน่าจะเป็นของความคลาดเคลื่อนชนิดที่ 1 ความคลาดเคลื่อนชนิดที่ 2 แสดงในภาพที่ 1

ตารางที่ 1 ความคลาดเคลื่อนชนิดที่ 1 และชนิดที่ 2

ความเป็นจริง	การตัดสินใจกับสมมตฐาน	
	ปฏิเสธ H_0	ไม่ปฏิเสธ H_0
H_0 เป็นจริง	Type I error	การตัดสินใจถูกต้อง
	ความน่าจะเป็น α	ความน่าจะเป็น 1-α
H_0 เป็นเท็จ	การตัดสินใจถูกต้อง	Type II error
	ความน่าจะเป็น 1 - β	ความน่าจะเป็น β



ภาพที่ 1 ความคลาดเคลื่อนชนิดที่ 1 และความคลาดเคลื่อนชนิดที่ 2

การใช้ขนาดอิทธิพล (Effect Size ใช้สัญลักษณ์ δ) ในงานวิจัยเชิงทดลองถ้างานวิจัยนั้นเป็นเชิงทดลองเปรียบระหว่างกลุ่มทดลองกับกลุ่มควบคุมแล้ว ถ้าขนาดอิทธิพล (δ) มีค่ามากแล้ว จะทำให้ขนาดของกลุ่มตัวอย่างน้อย ในทำนองกลับกัน ถ้างานวิจัยนั้นมีขนาดอิทธิพล(δ) มีค่าน้อยแล้ว จะทำให้ขนาดของกลุ่มตัวอย่างมีจำนวนมากขึ้น โดยผู้วิจัยจะทราบว่างานวิจัยนั้นมีขนาดอิทธิพลมากน้อยเพียงใด ซึ่งต้องทบทวนจากวรรณกรรมที่เกี่ยวข้อง หรือ ทำการสำรวจเบื้องต้น หรือ ใช้ค่าขนาดอิทธิพล น้อย กลาง มาก ดังกล่าวข้างต้น

คาอำนาจทดสอบ (Power of a test) ซึ่งไดจากการคำนวณโดยใช้ค่าความคลาดเคลื่อนชนิดที่ 2 (type II error หรือ β) คาอำนาจทดสอบ = 1-β ถาอำนาจทดสอบสูงก็ต้องใช้ขนาดของกลุ่มตัวอย่างมากกว่าค่าอำนาจทดสอบที่น้อยกว่า คาอำนาจการทดสอบ (Power of a test) คือโอกาสการปฏิเสธสมมติฐานที่ไม่ถูกต้อง นิยมกำหนดค่า ความคลาดเคลื่อนชนิดที่ 2 β ร้อยละ 20 จึงไดคาอำนาจทดสอบร้อยละ 80 และ ระดับนัยสำคัญ (significant level) มักใช้ที่ 95% หรือ α = 0.05

สูตรที่ใช้สำหรับการทดสอบค่าเฉลี่ยตัวอย่างเดียวจากการกระจายแบบปกติเมื่อทดสอบทั้งสองทาง[15] เนื่องจากขั้นตอนการพิสูจน์ซับซ้อนมาก แต่ก็มีที่มาเช่นเดียวกับการแสดงในสมการที่ 1 ข้างต้น เป็นดังนี้

$$n = \frac{\left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}\right)^2 \sigma^2}{\left(\mu_0 - \mu_1\right)^2} \qquad (4)$$

หรือ

$$n = \frac{\left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}\right)^2}{\delta^2} \qquad (5)$$

เมื่อ $Z_{1-\frac{\alpha}{2}}$ คือ ค่า Z ที่ระดับความเชื่อมั่นที่กำหนด $(1-\alpha)$

$Z_{1-\beta}$ คือ ค่า Z ที่ระดับอำนาจทดสอบที่กำหนด β

δ คือ ขนาดอิทธิพล

สูตรจากสมการที่ 5 จะพบว่า

เมื่อ ขนาดอิทธิพล (δ) กำหนดให้มีค่าน้อย จะทำให้ต้องใช้ขนาดตัวอย่าง n เพิ่มขึ้นด้วย

เมื่อ อำนาจทดสอบ (1-β) กำหนดให้มีค่าเพิ่มขึ้น จะทำให้ต้องใช้ขนาดตัวอย่าง n เพิ่มขึ้นเช่นกัน

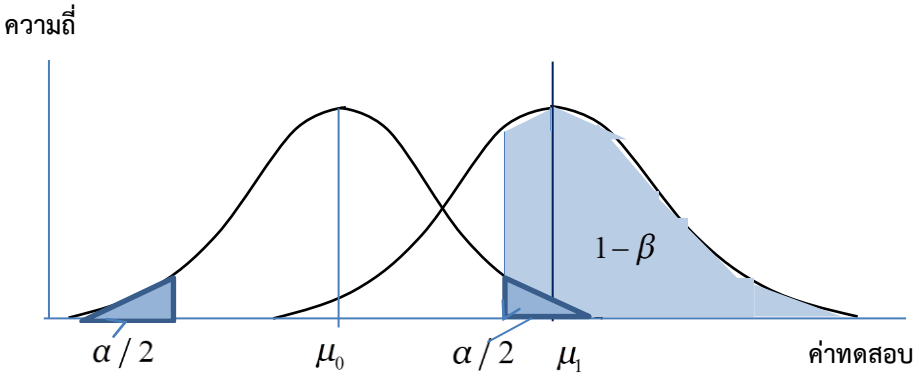
ตัวอย่างที่ 3 การศึกษาสมรรถนะของปอด ว่าเป็น 70% หรือไม่ การศึกษาที่ผ่านมามีค่าเฉลี่ย 75% ส่วนเบี่ยงเบนมาตรฐานเป็น 10% กำหนดให้อำนาจทดสอบเป็น 80%

$Z_{1-\frac{\alpha}{2}}$ = 1.96 เมื่อ α =0.05

$Z_{1-\beta}$ = 0.84 เมื่อ β = 0.20

n = $(1.96+0.84)^2/[(0.75-0.70)/0.10]^2$ = 31.36

จากตัวอย่างข้างต้นหากไม่มีผลการศึกษาในอดีตมาก่อน ก็จะไม่รู้ค่า $\mu_1 - \mu_0$ และ σ ก็สามารถใช้ค่าขนาดอิทธิพล น้อย ปานกลาง และมาก เท่ากับ 0.20 0.50 และ 0.80 ตามลำดับ กับ ความเชื่อมั่น 95% และ คาอำนาจทดสอบ 80% ก็จะใช้ขนาดตัวอย่างเพียง 40, 16 และ 10 ตัวอย่าง ตามลำดับ



ภาพที่ 2 แสดงการทดสอบค่าเฉลี่ยสองทางเมื่อค่าที่ทดสอบมากกว่าค่าเฉลี่ยของประชากร ($\mu_1 > \mu_0$)

จากภาพที่ 2 ถ้า Type I error หรือ α น้อย (แปลว่า มีโอกาสปฏิเสธสิ่งที่ถูกต้องน้อย) ช่วงความเชื่อมั่นกว้างขึ้น ค่า Z (ในที่นี้ คือ $Z_{1-\frac{\alpha}{2}}$) จะมากขึ้น ทำให้ n เพิ่มขึ้น และ ถ้า Type II error β น้อย (แปลว่า มีโอกาสยอมรับในสิ่งที่ผิดน้อย) อำนาจทดสอบ $1-\beta$ มาก (แปลว่า มีโอกาสปฏิเสธในสิ่งที่ผิดมาก) ค่า Z (ในที่นี้ คือ $Z_{1-\beta}$) จะมากขึ้น ทำให้ n เพิ่มขึ้นด้วย

การใช้โปรแกรมสำเร็จรูปในการคำนวณขนาดตัวอย่าง

เมื่อผู้วิจัยออกแบบการทดลองและกำหนดวิธีการทดสอบสมมติฐานแล้ว โปรแกรมสำเร็จรูป[4,5,6] จะให้เลือกรูปแบบการทดลองและวิธีการทดสอบสมมติฐานเพื่อคำนวณขนาดตัวอย่างให้ตรงตามวัตถุประสงค์ของผู้วิจัย โดยกำหนดให้ต้องใส่ ตัวเลขจำนวนประชากร (N) ค่าระดับนัยสำคัญ (significance level) หรือ ความคลาดเคลื่อนชนิดที่ 1 (α) หรือค่าช่วงความเชื่อมั่น (confidence interval 1-α) ค่าความคลาดเคลื่อนชนิดที่ 2 (β) หรืออำนาจทดสอบ (1-β) และค่าขนาดอิทธิพล (effect size δ) เพื่อให้โปรแกรมสามารถคำนวณผลที่ต้องการต่อไป

กิตติกรรมประกาศ

ขอขอบคุณ ดร. ชนากานต์ ยืนยง อธิการบดี มหาวิทยาลัยปทุมธานี และ ดร. นฤนาท ยืนยง คณบดีคณะสาธารณสุขศาสตร์และสิ่งแวดล้อม มหาวิทยาลัยปทุมธานี ที่ให้การสนับสนุนการผลิตบทความวิชาการครั้งนี้

เอกสารอ้างอิง

1. Thompson CB., and Panacek EA. (2006). Basics of Research Part 3 Research Study Designs: Experimental and Quasi-Experimental. **Air Medical Journal Associates**, 25(6), 242-246. doi:10.1016/j.amj.2006.09/001
2. Statistics Solutions. (2018). **Research Designs: Non-Experimental vs. Experimental.** Retrieved from https://www.statisticssolutions.com/research-designs-non-experimental-vs-experimental/
3. Walliman N. (2011). **Research Methods the basics**. Oxon: Routledge. p. 11.
4. Heinrich-Heine-Universität Düsseldorf. (2010-2019). **G*Power: Statistical Power Analyses for Windows and Mac.** Retrieved from http://www.gpower.hhu.de/
5. HyLown Consulting LLC. (2013-2018). **Power and Sample Size Home Calculators.** Retrieved from http://powerandsamplesize.com/Calculators/
6. UCSF Clinical & Translation Science Institute. (2018, August). **Sample Size calculators.** Retrieved from http://www.sample-size.net/
7. Cochran WG. (1963). **Sampling Techniques**, 2nd Ed.. New York: John Wiley and Sons, Inc.
8. Israel GD. (2003). **Determining Sample Size. PEOD6**. University of Florida IFAS Extension. Retrieved from https://www.tarleton.edu/academicassessment/documents/Samplesize .pdf
9. Cochran WG. (1997). **Sampling Techniques**, 3rd ed. New York: John Wiley & Sons. p. 51-53.
10. Yamane T. (1967). **Statistics, An Introductory Analysis**, 2nd Ed. New York: Harper and Row.
11. ขนิษฐา สร้อยมาลุน, จิราพร ระมาศ, วันเพ็ญ กำลังดี, และปริยาภัทร สิงห์ทอง. การศึกษาผลการรักษาอาการท้องผูกด้วยการนวดไทย. **วารสารหมอยาไทยวิจัย**, 5 เรื่องที่ 4.
12. Prajapati, B., Dunne, M.C., and Armstrong, R. (2011). Sample size estimation and statistical power analyses. **Optometry Today**, 16(07), 10-18.
13. Cohen, J. (1988). **Statistical power analysis for the behavioral sciences.** Lawrence Erlbaum Associates Inc.
14. Leppink J., O'Sullivan P., and KalWinston K. (2016). Effect size – large, medium, and small**.** **Perspect Med Educ**, 5:347–349. doi:10.1007/s40037-016-0308-y
15. Rosner B. (2016). **Fundamentals of Biostatistics.** 8th ed. Boston, MA: Cengage Learning. p. 267.