



บทความพิเศษ : การคำนวณขนาดตัวอย่าง

จุฬาลักษณ์ โกมลตรี, ประด.

การคำนวณขนาดตัวอย่างเป็นสิ่งที่จำเป็นในการวางแผนงานวิจัยซึ่งจะทำให้ผู้วิจัยทราบถึงความเป็นไปได้ ค่าใช้จ่ายและระยะเวลาของการศึกษา การคำนวณขนาดตัวอย่างต้องใช้ค่าประมาณต่างๆ เช่น ค่าเฉลี่ย, ค่าเบี่ยงเบนมาตรฐาน (SD) ซึ่งได้จากการศึกษาในอดีตที่เหมือนกัน, คล้ายกัน ในบางแง่, ประสบการณ์ และความคาดหวังของผู้วิจัย แต่อย่างไรก็ตามในบางครั้งไม่เคยมีการศึกษาลักษณะนี้ในอดีต ผู้วิจัยจึงไม่มีค่าต่างๆ ที่จำเป็นในการคำนวณขนาดตัวอย่าง ทำให้ไม่สามารถคำนวณขนาดตัวอย่างได้ ต้องทำการศึกษานำร่อง (pilot study, hypothesis generating study) ก่อนในผู้ปวยจำนวนหนึ่งที่มีลักษณะเช่นเดียวกับกลุ่มตัวอย่างที่ต้องการศึกษาจริง เพื่อให้ได้ค่าประมาณต่างๆ ที่จำเป็นสำหรับการคำนวณขนาดตัวอย่างในการศึกษาจริง โดยส่วนมากจำนวนตัวอย่างในการศึกษานำร่องจะไม่มากนัก เช่น ไม่เกิน 20 ราย ในโครงร่างของการศึกษาแบบนำร่องจึงไม่มีรายละเอียดของการคำนวณขนาดตัวอย่าง ผู้วิจัยเพียงแต่ระบุว่าเป็นการศึกษาแบบนำร่องด้วยเหตุผลใด และระบุจำนวนตัวอย่างที่จะทำการศึกษา เช่น ในการพัฒนาชุดโปรแกรมสำหรับผู้ปกครองเพื่อการสอนซ่อมเสริมในเด็กบกพร่องทางการเรียนรู้ ผู้วิจัยประเมินคะแนนความสามารถในการอ่านและเขียนสะกดคำของเด็กก่อนและหลังการสอน เนื่องจากเป็นโปรแกรมใหม่ จึงไม่มีการคำนวณขนาดตัวอย่าง ผู้วิจัยกำหนดไปเลยว่าการศึกษาในเด็ก 8 คน¹

การคำนวณขนาดตัวอย่างมักจะคำนวณตามวัตถุประสงค์หลักของการศึกษา แต่ถ้าสามารถคำนวณตามวัตถุประสงค์รองได้ด้วยก็ทำให้มั่นใจได้ว่ามีจำนวนตัวอย่างเพียงพอในการตอบคำถามวิจัยทุกคำถาม

สูตรการคำนวณขนาดตัวอย่างมีหลายสูตร ขึ้นกับ (1) วัตถุประสงค์ของการศึกษา (2) วิธีวิเคราะห์ทางสถิติที่จะใช้ และ (3) รูปแบบการศึกษา

ในด้านวัตถุประสงค์ของการศึกษา จะมีสูตรการคำนวณขนาดตัวอย่างหลายสูตร เช่น สูตรการคำนวณขนาดตัวอย่างเพื่อเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มตัวอย่าง 2 กลุ่ม, 3 กลุ่มที่เป็นอิสระต่อกัน สูตรการคำนวณขนาดตัวอย่างเพื่อเปรียบเทียบสัดส่วนระหว่างกลุ่มตัวอย่าง 2 กลุ่มที่เป็นอิสระต่อกัน แม้ในวัตถุประสงค์เดียวกันก็มีสูตรการคำนวณขนาดตัวอย่างหลายสูตรขึ้นกับวิธีวิเคราะห์ทางสถิติที่จะใช้ เช่น ในการเปรียบเทียบค่าเฉลี่ยจากกลุ่มตัวอย่าง 2 กลุ่ม อาจจะใช้สูตรการคำนวณขนาดตัวอย่างที่มาจาก 2-sample t-test ถ้าตัวแปรนั้นมีการแจกแจงแบบปกติ หรือใช้สูตรที่มาจาก Mann-Whitney U test ถ้าตัวแปรนั้นไม่มีการแจกแจงแบบปกติ ส่วนการเปรียบเทียบค่าสัดส่วน 2 ค่า อาจจะใช้สูตรการคำนวณขนาดตัวอย่างที่มาจาก Pearson's chi-square test หรือใช้สูตรที่มาจาก Yates' correction

ส่วนรูปแบบการศึกษาที่ต่างกันก็อาจจะนำไปสู่สูตรการคำนวณขนาดตัวอย่างที่แตกต่างกัน เช่น การหาความสัมพันธ์ระหว่างความเครียด (สูง, ไม่สูง) กับความคิดฆ่าตัวตาย (มี, ไม่มี) อาจจะมีรูปแบบการศึกษาเป็น (1) case-control หรือ (2) prospective cohort ก็ได้ ถ้าเป็นแบบ case-control กล่าวคือ case คือคนที่มีความคิดฆ่าตัวตายและ control คือคนที่ไม่มีความคิดฆ่าตัวตาย การคำนวณหาจำนวน case และ control (n_1 และ n_2 ตามลำดับ) จะต้องมีค่าประมาณของสัดส่วนการมีความเครียดสูงในกลุ่ม case และ control (p_1 และ p_2 ตามลำดับ) หรือ มีค่าประมาณของสัดส่วนการมีความเครียดสูงในกลุ่ม control (p_2) และค่า Odds ratio (OR) ของการมีความเครียดสูงก็ได้ การคำนวณจำนวน case, control จากการกำหนด p_1 , p_2 และ p_2 , OR จะมีรูปแบบของสูตรที่ต่างกัน แต่จะได้จำนวน case, control ที่เท่ากัน แต่ถ้ารูปแบบการศึกษาเป็น prospective cohort ซึ่งจะเริ่มจากคนที่ไม่มีความเครียดสูง (n_1 , n_2 ตามลำดับ) แล้วติดตามไประยะเวลาหนึ่งเพื่อดูการมีความคิดฆ่าตัวตาย ในการคำนวณขนาดตัวอย่างต้องประมาณค่าสัดส่วนการมีความคิดฆ่าตัวตายในกลุ่มที่ไม่มีและมีความเครียดสูง (p_1 และ p_2 ตามลำดับ) หรือ ประมาณค่าสัดส่วนการมีความคิดฆ่าตัวตายในกลุ่มที่ไม่มีมีความเครียดสูง (p_1) และ Relative risk (RR) การคำนวณจำนวนคนที่ไม่มีความเครียดสูงจาก p_1 , p_2 และจาก p_1 , RR จะมีสูตรต่างกัน แต่ได้จำนวนตัวอย่างเท่ากัน

การคำนวณขนาดตัวอย่างสามารถทำได้หลายทาง เช่น คำนวณมือจากสูตร, ใช้ตารางสำเร็จรูป, nomogram, ใช้โปรแกรมสำเร็จรูปต่างๆ เช่น nQuery Advisor, PASS (Power Analysis and

Sample Size), STATCAL, PS (Power and Sample Size) อื่นๆตาราง, nomogram และโปรแกรมต่างๆ อาจจะใช้สูตรต่างกัน ทำให้ได้จำนวนตัวอย่างแตกต่างกัน แต่จะไม่แตกต่างกันมาก

ปัญหาทั่วไปของนักวิจัยในการคำนวณขนาดตัวอย่าง คือความไม่ชัดเจนในวัตถุประสงค์ของการศึกษา การไม่มีความรู้เกี่ยวกับวิธีออกแบบการศึกษา การไม่เข้าใจศัพท์ทางสถิติ การไม่สามารถเปลี่ยนคำถามวิจัยเป็นคำถามทางสถิติ การเลือกสูตรการคำนวณ การหาตัวเลขที่จำเป็นในแต่ละสูตร การแทนค่าต่างๆ และการคำนวณ

บทความนี้จะกล่าวถึงสูตรการคำนวณขนาดตัวอย่างที่ใช้อยู่

ประเภทของการคำนวณขนาดตัวอย่าง

การคำนวณขนาดตัวอย่างแบ่งได้เป็น 2 ประเภทตามวัตถุประสงค์คือ

(1) การคำนวณขนาดตัวอย่างเพื่อการประมาณค่าพารามิเตอร์ในประชากร (parameter estimation) (คำว่าพารามิเตอร์ หมายถึงค่าที่แสดงลักษณะของประชากร เช่น ค่าเฉลี่ย (μ), ค่าสัดส่วน (π , P))

(2) การคำนวณขนาดตัวอย่างเพื่อการทดสอบสมมติฐาน (hypothesis testing)

การคำนวณขนาดตัวอย่างเพื่อการประมาณค่าพารามิเตอร์ในประชากรมีจุดประสงค์เพื่อให้ได้ค่าประมาณที่มีความแม่นยำ ไม่มีสมมติฐานทางสถิติที่ต้องทดสอบ การรายงานผลการศึกษาก็รายงานค่าที่ได้จากตัวอย่างและค่า $(1-\alpha)100\%$ confidence interval (CI) เช่นเมื่อ Type I error (α)=0.05 จะได้ 95%CI โดยไม่มีการรายงาน p-value เช่น ความ

ไวของ mammogram เท่ากับ 85% (95%CI=80%, 90%) ในสูตรการคำนวณขนาดตัวอย่างจึงกำหนดแต่เพียงขนาดของ α ไม่ต้องกำหนดขนาดของ β หรือ power (1- β) ส่วนการคำนวณขนาดตัวอย่างเพื่อทดสอบสมมติฐานต้องมีการเขียนสมมติฐานทางสถิติที่ต้องการทดสอบ (null hypothesis H_0 และ alternative hypothesis H_1) และในการรายงานผลการศึกษา

ต้องมี p-value และ/หรือ 95%CI สูตรการคำนวณขนาดตัวอย่างเพื่อทดสอบสมมติฐานจะต้องกำหนดทั้ง α และ β เนื่องจากในความเป็นจริงผู้วิจัยไม่ทราบว่าจะ H_0 หรือ H_1 จะถูกต้อง ตารางที่ 1 แสดงความหมายของ α , β และตารางที่ 2 แสดงค่า z_α , z_β ที่ใช้บ่อย

ตารางที่ 1 ความผิดพลาดในการทดสอบสมมติฐาน (α , β)

ผลการทดสอบสมมติฐาน	ความจริง	
	H_0 ถูกต้อง ($P_A = P_B$)	H_0 ผิด ($P_A \neq P_B$)
ไม่ปฏิเสธ H_0 (p-value $> \alpha$)	1- α (True negative, TN)	β (False negative, FN)
ปฏิเสธ H_0 (p-value $\leq \alpha$)	α (False positive, FP)	1- β (True positive, TP, Power)

ตารางที่ 2 ค่า z_α , z_β ที่ใช้บ่อย

2-sided Type I error		1-sided Type I error		Type II error	
2-sided α	$z_{\alpha/2}$	1-sided α	z_α	Power (1- β)	z_β
0.10	1.645	0.10	1.282	0.80	0.842
0.05	1.960	0.05	1.645	0.85	1.036
0.01	2.576	0.01	2.326	0.90	1.282

1. การคำนวณขนาดตัวอย่างเพื่อการประมาณค่าพารามิเตอร์ในประชากร

การคำนวณขนาดตัวอย่างเพื่อการประมาณค่าพารามิเตอร์ในประชากรที่ใช้บ่อยมีดังนี้

1.1 การประมาณค่าสัดส่วน 1 ค่า

การศึกษาเพื่อประมาณค่าสัดส่วน 1 ค่า เช่น การประมาณค่าความชุก (prevalence), อุบัติการณ์ (incidence) ของโรค และการประมาณค่าความไว (sensitivity), ความจำเพาะ (specificity) และความถูกต้อง (accuracy) ของวิธีการทดสอบ, เครื่องมือ เช่น ผู้วิจัยอยากทราบค่าความไวและความจำเพาะของ ultrasound, mammogram เทียบกับผลการตรวจทางพยาธิวิทยาของชิ้นเนื้อ (gold standard) ในการวินิจฉัยมะเร็งเต้านม ค่าต่างๆ เหล่านี้จะรายงานเป็นร้อยละเช่นความไว 90% แต่ในแง่ของสัดส่วนจะเขียนว่าสัดส่วนของความไว (p) เท่ากับ 0.90 ซึ่ง p จะมีค่าระหว่าง 0-1 สูตรการคำนวณขนาดตัวอย่างมาจากการกำหนด 95%CI ของค่าสัดส่วนจริงในประชากร (P) เช่น 95%CI ของค่าสัดส่วนความไวจริงเท่ากับ 0.90 ± 0.05 หรือผู้วิจัยเชื่อมั่น 95% ว่าความไวในประชากรจะมีค่าระหว่าง 0.85 ถึง 0.95 ค่า 0.05 ที่อยู่ข้างหลังเครื่องหมาย $\pm$ เป็นค่าความคลาดเคลื่อนที่ยอมรับได้ (margin of error, allowable error, e) ในการประมาณค่าสัดส่วน

สูตรทั่วไปของ $(1-\alpha)100\%CI$ คือ

$(1-\alpha)100\%CI$ ของค่าพารามิเตอร์ $P =$ ค่าสถิติที่ได้จากตัวอย่าง $\pm z_{\alpha/2} SE$ ของค่าสถิติจากตัวอย่าง

ดังนั้น เมื่อค่าสถิติคือค่าสัดส่วนจากตัวอย่าง (p) สูตรของ $(1-\alpha)100\%CI$ ของ P จึงคือ $(1-\alpha)100\%CI$ ของ $P = p \pm z_{\alpha/2} SE(p)$

$$= p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

เมื่อกำหนดให้ $(1-\alpha)100\%CI$ ของ $P = p \pm e$

$$\text{ดังนั้น } e = z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

$$\text{ซึ่งจะได้ว่า } n = \frac{z_{\alpha/2}^2 p(1-p)}{e^2} \quad \text{-----}(1)$$

สูตรที่ (1) คือสูตรการคำนวณขนาดตัวอย่างเพื่อการประมาณค่าสัดส่วน 1 ค่า

ตัวอย่างที่ 1 ในการศึกษาสุขภาพจิตคนไทยปี พ.ศ. ๒๕๕๐² ซึ่งจะรายงานผลเป็นความชุกของการมีสุขภาพจิตต่ำกว่าปกติ ถ้าผู้วิจัยประมาณว่าความชุกในปี พ.ศ. ๒๕๕๐ จะเท่ากับ 32% โดยมี 95%CI เท่ากับ $32\% \pm 3\%$ จะต้องทำการศึกษาในคน 929 คน ดังรายละเอียดการคำนวณด้วยสูตรที่ (1) คือ

$$n = \frac{1.96^2 (0.32)(0.68)}{0.03^2} = 928.8 = 929$$

จากสูตรที่ (1) จะเห็นได้ว่า

(1) ในสูตรมี e เป็นตัวหารดังนั้นเมื่อ e (error) มีค่าน้อยลง n จะมีค่ามากขึ้น กล่าวคือถ้าต้องการให้ค่าประมาณที่ได้จากตัวอย่างมีความผิดพลาดน้อยหรือมีค่าใกล้เคียงกับค่าจริงในประชากรมากต้องทำการศึกษาในตัวอย่างจำนวนมาก เช่น ถ้ากำหนดให้ 95%CI ของความชุกการมีสุขภาพจิตไม่ดีเท่ากับ $32\% \pm 2\%$ จะต้องทำการศึกษาในคน 2,090 คน

ในการกำหนดค่า e ต้องกำหนดให้ $p \pm e$ มีค่าไม่ต่ำกว่า 0 และไม่เกิน 1 เช่นเมื่อกำหนดให้ 95%CI ของ P เท่ากับ 0.95 ± 0.05 จะได้ $n=73$ ผู้วิจัยต้องการลดจำนวน n จึงเพิ่มค่า e เป็น 0.07 ซึ่งจะได้ $n=38$ แต่ 95%CI จะเท่ากับ 0.95 ± 0.07 หรือ 0.88, 1.02 หรือมี upper limit เกิน 1 หรือ 100% ซึ่งเป็นไปไม่ได้ ดังนั้นเมื่อ $p=0.95$ ค่า e จะต้องไม่ค่าไม่เกิน 0.05

การกำหนด e ทำได้ 2 แบบคือ

ก) e เป็นค่าคงที่ (absolute error) เช่น 0.1, 0.05

ข) e เป็นค่าที่สัมพันธ์กับ p (relative error) เช่น e เท่ากับ 20% ของ p ดังนั้นถ้า $p=0.1$, $e=(0.2 \times 0.1)=0.02$ และ 95%CI $=0.1 \pm 0.02$ หรือ 0.08 ถึง 0.12 ถ้า $p=0.7$, $e=(0.2 \times 0.7)=0.14$ และ 95%CI $=0.7 \pm 0.14$ หรือ 0.56 ถึง 0.84

(2) ถ้า α มีค่าน้อยลงเช่นลดลงจาก 0.05 เป็น 0.01 นั่นคือเปลี่ยนจาก 95%CI เป็น 99%CI หรือเปลี่ยนจากความเชื่อมั่น 95% เป็น 99% หรือมีความเชื่อมั่นมากขึ้นจะต้องใช้จำนวนขนาดตัวอย่างมากขึ้น ($z_{0.01/2} = 2.576$, $z_{0.05/2} = 1.96$) ในทางตรงข้าม ถ้าเปลี่ยนจาก 95%CI เป็น 90%CI จะใช้จำนวนขนาดตัวอย่างน้อยลง ($z_{0.10/2} = 1.645$) แต่ส่วนมากจะใช้ 95%CI

(3) สูตรที่ (1) มาจากการกำหนด 95%CI เป็นแบบ 2-sided คือเท่ากับ $p \pm e$ การเปลี่ยนเป็น 1-sided 95%CI ($Z_{0.05/2} = 1.96$, $Z_{0.05} = 1.645$) จะทำให้ n ลดลง แต่ 95%CI ที่ได้จะกลายเป็น $p+e$ (upper 95%CI) หรือ $p-e$ (lower 95%CI) อย่างใดอย่างหนึ่งซึ่งไม่ค่อยใช้ ดังนั้นจึงควรใช้เป็น 2-sided CI

(4) ในสูตรที่ (1) จะมี $p(1-p)$ เป็นตัวตั้ง ดังนั้นถ้าผลคูณระหว่าง $p(1-p)$ มีค่ามาก n จะมีค่ามากด้วย ซึ่งผลคูณระหว่าง $p(1-p)$ จะมีค่ามากที่สุดเมื่อ $p=0.5$ ดังนั้นในกรณีที่ผู้วิจัยไม่ทราบว่า p ควรจะเป็นเท่าไร การใช้ $p=0.5$ จะทำให้ได้จำนวนขนาดตัวอย่างมากที่สุด (เมื่อ e มีค่าเท่ากัน) เมื่อ $p=0.5$ สูตรการคำนวณขนาดตัวอย่างจะเป็น

$$n = \frac{1.96^2 (0.5)(0.5)}{e^2} \approx \frac{1}{e^2} \quad \text{-----}(2)$$

ในกรณีที่ประชากรมีจำนวนนับได้ (finite population) จะคำนวณ n ตามสูตรที่ (1) ก่อน แล้วปรับ n ที่คำนวณได้ตามขนาดของประชากร (N) จะได้ขนาดตัวอย่างใหม่คือ n' ซึ่งมีสูตรทั่วไปคือ

$$n' = \frac{n}{1 + \frac{n}{N}} \quad \text{-----}(3)$$

เมื่อ $n=1/e^2$ ตามสูตรที่ (2) สูตรที่ (3) จะกลายเป็นสูตรที่ (4) ซึ่งคือสูตรของ Yamane

$$n' = \frac{N}{1 + e^2 N} \quad \text{-----}(4)$$

เมื่อ $n = Z_{\alpha/2}^2 p(1-p)/e^2$ ตามสูตรทั่วไป สูตรที่ (1) สูตรที่ (3) จะมีสูตรทั่วไปคือ

$$n' = \frac{NZ_{\alpha/2}^2 p(1-p)}{e^2 N + Z_{\alpha/2}^2 p(1-p)} \quad \text{-----}(5)$$

n' จะมีค่าน้อยกว่า n เสมอ ถ้า N มีจำนวนมากๆ จะทำให้ n' และ n มีค่าใกล้เคียงกัน ในงานวิจัยทางการแพทย์ ส่วนมากผู้วิจัยจะไม่ทราบจำนวนผู้ป่วยทั้งหมด (N) ซึ่งอาจจะถือได้ว่าประชากรมี

จำนวนอนันต์ (infinity) จึงใช้เฉพาะสูตรที่ (1) เนื่องจาก $n' \cong n$

ตัวอย่างที่ 2 ในการศึกษาสุขภาพจิตและกลไกทางจิตของผู้ต้องขังและนักโทษไทย³ ประชากรคือผู้ต้องขังรวมทั่วประเทศในปี พ.ศ. ๒๕๔๕ จำนวน 254070 คน (N) เนื่องจากไม่เคยมีการศึกษามาก่อนจึงไม่ทราบค่าสัดส่วนการมีสุขภาพจิตต่ำกว่าเกณฑ์เฉลี่ย (p) ผู้วิจัยจึงคำนวณขนาดตัวอย่างจาก $p=0.5$ โดยให้มี 95%CI เท่ากับ 0.5 ± 0.02 ซึ่งจะได้ $n=2400$ เมื่อปรับตามจำนวน N จะได้ $n'=2378$ และเมื่อคาดว่าจะมีผู้ไม่ตอบแบบสอบถามจำนวนมากจึงเพิ่มขนาดตัวอย่างเป็น 2.2 เท่าคือเป็น 5,232 คน

สูตรที่ (3) ถึง (5) เป็นสูตรที่มีการใช้ผิดบ่อย สูตรที่ (3) และ (5) เป็นสูตรเมื่อทราบจำนวนประชากรในเรื่องนั้นๆ (N) ส่วนสูตรที่ (4) เป็นสูตรเฉพาะเมื่อทราบ N และกำหนดให้ $p=0.5$, $\alpha=0.05$ ในสูตรที่ (3), (4) และ (5) กลุ่มตัวอย่างได้จากการสุ่มแบบง่าย (simple random sampling, SRS) จากประชากร เช่น ในตัวอย่างที่ 2 จะจับสลากผู้ต้องขังจากเรือนจำต่างๆ มา 5,231 คน ซึ่งเป็นวิธีที่ง่าย ตัวอย่างที่ได้จะกระจายกันตามเรือนจำต่างๆ อาจจะไม่สอดคล้องกับความจริง เช่น เรือนจำบางแห่งมีผู้ต้องขังน้อยแต่สุ่มได้มาก หรือ เรือนจำบางแห่งมีผู้ต้องขังมากแต่สุ่มได้น้อย ในทางปฏิบัติจึงใช้วิธีการสุ่มแบบชั้นภูมิ (Stratified sampling) คือ ในแต่ละเรือนจำซึ่งคือชั้นภูมิ (stratum) จะสุ่มผู้ต้องขังตามสัดส่วนของจำนวนผู้ต้องขังทั้งหมดในเรือนจำนั้นๆ (proportionate allocation) กล่าวคือ เรือนจำที่มีผู้ต้อง

ขังจำนวนมากจะได้รับการสุ่มมามาก ส่วนเรือนจำที่มีผู้ต้องขังจำนวนน้อยจะได้รับการสุ่มมาน้อย

ตัวอย่างที่ 3 ในการสำรวจการติดเกมในเด็กนักเรียนโรงเรียนเอกชนชั้นประถมศึกษาปีที่ 5 และ 6 และมัธยมศึกษาที่ 1 ถึง 6 ถ้าเพื่อความสะดวกและเป็นไปได้ผู้วิจัยวางแผนว่าจะทำการศึกษาในโรงเรียนเอกชน 4 แห่ง โดยให้แต่ละโรงเรียนมีจำนวนนักเรียนที่จะสุ่มเท่ากัน และจะสุ่มตัวอย่างแบบ Stratified sampling โดยมี 3 ชั้นภูมิคือ ป.5-6, ม.1-3 และ ม.4-6 จึงจะคำนวณขนาดตัวอย่างในแต่ละชั้นภูมิ โดยประมาณว่าเด็ก ป.5-6, ม.1-3 และ ม.4-6 จะติดเกม 10%, 7% และ 2% ตามลำดับ โดยมี 95%CI เท่ากับ $10\% \pm 5\%$, $7\% \pm 3\%$ และ $2\% \pm 2\%$ ตามลำดับ จะต้องทำการศึกษาในเด็ก ป.5-6, ม.1-3 และ ม.4-6 จำนวน 139, 278 และ 189 คนตามลำดับ คือโรงเรียนละ 153 คน (ป.5-6, ม.1-3 และ ม.4-6 จำนวน 35, 70 และ 48 คนตามลำดับ) หรือรวม 612 คนใน 4 โรงเรียน

1.2 การประมาณค่าเฉลี่ย 1 ค่า

สูตรการคำนวณขนาดตัวอย่างเพื่อประมาณค่าเฉลี่ย 1 ค่าจะใช้หลักการเดียวกับการประมาณค่าสัดส่วน 1 ค่า กล่าวคือมาจากการกำหนด $(1-\alpha)100\%CI$ ของค่าเฉลี่ยจริงในประชากร (μ) คือ $(1-\alpha)100\%CI$ ของ μ

$$= \bar{x} \pm Z_{\alpha/2} SE(\bar{x}) = \bar{x} \pm Z_{\alpha/2} SD / \sqrt{n}$$

เมื่อกำหนดให้ $(1-\alpha)100\%CI$ ของ $\mu = \bar{x} \pm e$

$$\text{จะได้ว่า } e = \frac{Z_{\alpha/2} SD}{\sqrt{n}}$$

$$\text{และ } n = \left[\frac{Z_{\alpha/2} SD}{e} \right]^2 \quad \text{-----}(6)$$

สูตรที่ (6) คือสูตรการคำนวณขนาดตัวอย่างเพื่อการประมาณค่าเฉลี่ย 1 ค่า ซึ่งเป็นสูตรที่มีการใช้น้อย

ตัวอย่างที่ 4 ถ้าในการศึกษาสุขภาพจิตคนไทยปี พ.ศ. ๒๕๕๐ จะรายงานผลการศึกษาเป็นค่าเฉลี่ยจากเครื่องมือดัชนีชี้วัดสุขภาพจิตคนไทยฉบับสมบูรณ์ (ปี 2007) ซึ่งมี 55 ข้อ ๆ ละ 1-4 คะแนน หรือมีคะแนนเต็มระหว่าง 55-220 คะแนน โดยประมาณว่าคนไทยน่าจะมีค่าเฉลี่ยประมาณ 170, SD=30 และมี 95%CI ของ $\mu=170\pm2$ จะต้องทำการศึกษาในคน 865 คนดังการคำนวณข้างล่าง

$$n = \left[\frac{1.96 * 30}{2} \right]^2 = 864.4 = 865$$

ในสูตรที่ (6) มี e เป็นตัวหารและ SD เป็นตัวตั้ง ดังนั้นถ้ากำหนดให้ e มีค่าน้อยหรือ 95%CI มีช่วงแคบจะต้องใช้จำนวนขนาดตัวอย่างมากขึ้น และเมื่อ SD มีค่ามาก n จะมีค่ามากด้วย (ตารางที่ 3)

ตารางที่ 3 ขนาดตัวอย่างในการประมาณค่าเฉลี่ยของคะแนนสุขภาพจิตคนไทย

SD	e	95%CI ของค่าเฉลี่ย (μ)	n
30	2	170 ± 2	865
	1	170 ± 1	3458
35	2	170 ± 2	1177
40	2	170 ± 2	1537

ตัวอย่างที่ 5 ในการสำรวจระดับสติปัญญา (IQ) เด็กไทย ถ้าประมาณว่าเด็กไทยน่าจะมีค่าเฉลี่ยของ IQ ประมาณ 105, SD=17 และมี 95%CI ของ $\mu=105\pm0.5$ จะต้องทำการศึกษาในคน 4,441 คน แต่เนื่องจากในการศึกษานี้จะมีการสุ่มตัวอย่างแบบ Stratified multi-stage cluster sampling โดยจะศึกษาใน 15 จังหวัดทั่วประเทศจึงต้องมีการเพิ่มจำนวนตัวอย่างขึ้นอีกตาม Design effect (Deff) เช่น ถ้าประมาณว่า Deff=2 จะต้องใช้ $n=2*4441 = 8882$

สรุป

การคำนวณขนาดตัวอย่างเพื่อการประมาณค่าพารามิเตอร์ในประชากร จะกำหนดเฉพาะ α ไม่ต้องกำหนด β หรือ Power ($1-\beta$) เนื่องจากไม่มีสมมติฐานทางสถิติที่ต้องทดสอบและไม่มีการรายงาน p-value ในผลการศึกษา เมื่อ $\alpha=0.05$ สูตรการคำนวณขนาดตัวอย่างมาจากการกำหนด 95%CI ของค่าสัดส่วน (π , P) หรือของค่าเฉลี่ย (μ) ในประชากร

เอกสารอ้างอิง

1. Sansanee S. Prototype development: learning disable children remedial teaching programs for parents. J Ment Health Thailand 2008;16:69-77.
2. Apichai M, Yongyud W, Tavee T, Watchanee H, Emhathai S, Priwan R, et al. Thai mental health 2007. J Ment Health Thailand 2009;17:104-17.
3. Duangta G. Mental health and defenses mechanism of Thai inmates and prisoners. J Ment Health Thailand 2007;15:79-92.