

สถิติ: นัยสำคัญ อำนาจทดสอบ ขนาดตัวอย่าง และ ขนาดอิทธิพล Statistics: Significance, Power, Sample Size & Effect Size

วัฒนา ชยรัช

สาขาวิชาการแพทย์แผนไทย คณะสหเวชศาสตร์ มหาวิทยาลัยปทุมธานี

Vadhana Jayathavaj

*Department of Thai Traditional Medicine, Faculty of Allied Health Sciences,
Pathumthani University*

Corresponding Author: vadhana.j@ptu.ac.th

บทคัดย่อ

วิชาสถิติมีคำนิยาม ข้อมูล ตัวเลข สัญลักษณ์ สูตร และรูปภาพ ซึ่งเป็นสิ่งที่ทำให้วิชาสถิติเป็นวิชาที่ยาก ผู้เรียนมีทัศนคติเชิงลบ และปิดกั้นตนเองที่จะเรียนรู้ สถิติเป็นวิชาสำคัญในการทำวิจัย การวิเคราะห์อำนาจทดสอบซึ่งเป็นองค์ประกอบของ ระดับนัยสำคัญ อำนาจทดสอบ ขนาดอิทธิพล และขนาดตัวอย่าง ความเข้าใจในใจสถิติพื้นฐาน ข้อมูล การแจกแจงความถี่ ร้อยละ การวัดแนวโน้มเข้าสู่ส่วนกลาง ความน่าจะเป็น การแปลงเป็นค่ามาตรฐาน การกระจายของการสุ่มตัวอย่าง และการทดสอบสมมติฐาน ตามลำดับ เป็นส่วนสำคัญในการเรียนรู้การวิเคราะห์อำนาจทดสอบ บทความนี้ประกอบด้วย การบรรยายอย่างง่าย พร้อมตัวอย่าง สัญลักษณ์ และรูปภาพตามความจำเป็นเพื่อให้นักศึกษาหรือนักวิจัยเข้าใจ และนำไปใช้ได้อย่างเหมาะสม

คำสำคัญ: นัยสำคัญ, อำนาจทดสอบ, ขนาดตัวอย่าง, ขนาดอิทธิพล

Abstract

Statistics has definitions, data, numbers, symbols, formulas, and figures, that made this subject is too difficult to learn and the learners have negative attitudes and blocked their learning. Statistics is an essential part in doing research. The power analysis is the combination of the four related factors: the significance level, the power of a test, the effect size, and the sample size. The basic knowledge including basic statistics (the frequency distribution, percentages, the measure of central tendency, the probability density function, the conversion

Received : 24 January 2021

Revised : 20 April 2021

Accepted : 4 May 2021

Online publication date : 30 June 2021

to the standard normal distribution, and testing the hypothesis respectively) is an essential part to learn the power analysis. This article composed in a simple language with examples, symbols, and pictures as needed for students or researchers to comprehend and apply appropriately.

Keywords: Significance level, Power of a test, Sample Size, Effect Size

บทนำ

วิชาสถิติเป็นศาสตร์และศิลป์ในการรวบรวม จัดระบบ วิเคราะห์ และสรุปผลจากข้อมูล การมีความเข้าใจพื้นฐานเกี่ยวกับวิธีการทางสถิติ ก็จะสามารถใช้วิจารณ์ญาณและประเมินสิ่งต่าง ๆ จากข้อมูลที่พบในชีวิตประจำวันได้ (Annenberg Learner, 2020) ความรอบรู้วิชาสถิติเป็นส่วนประกอบสำคัญของการศึกษาระดับปริญญาตรีและบัณฑิตศึกษาในทุกสาขา ในหลักสูตรอื่นที่ไม่ใช่สาขาสถิติจะใช้เวลาในการสอนสถิติที่จำกัด การใช้สถิติในโลกของความเป็นจริงนั้นต้องมีความรู้เกี่ยวกับแนวคิดพื้นฐานทางสถิติ มีความสามารถในการเชื่อมโยงสถิติกับสถานการณ์จริง และสามารถสังเคราะห์องค์ประกอบของการศึกษาทางสถิติและการสื่อสารผลลัพธ์ในลักษณะที่ชัดเจนได้ (Mustafa, 1996) วิชาสถิติเป็นการรวบรวมข้อมูล การใช้ข้อมูลและการประยุกต์ผลการวิเคราะห์ทางสถิตินำมาใช้ประกอบการตัดสินใจในชีวิตประจำวัน ไม่ใช่แค่การคิดเลขในวิชาคณิตศาสตร์ ความวิตกกังวลในวิชาสถิติจึงมีมากกว่าความวิตกกังวลในวิชาคณิตศาสตร์ (Cruise, Cash, Bolton, 1985). นักศึกษาหลักสูตรจิตวิทยาในระดับปริญญาตรีมีความวิตกกังวลและทัศนคติเชิงลบวิชาสถิติในระดับที่มีนัยสำคัญ (Ruggeri, Dempster, Hanna, Cleary, 2008) นักศึกษารัฐศาสตร์ก็มีความวิตกกังวลกับวิชาสถิติ การกลัวเกินไปที่จะเรียนรู้ ทัศนคติเชิงลบที่มีต่อสถิติเป็นอุปสรรคต่อการเรียนรู้สถิติเพื่อให้ได้มาซึ่งทักษะเชิงปริมาณ (Slootmaeckers, Kerremans, Adriaensen, 2014) เนื้อหาของวิชาสถิติเริ่มจากการกำหนดคำนิยาม จากนั้นก็จะใช้สูตรและสัญลักษณ์กับตัวแปร (ที่อาจเทียบได้กับตัวอักษรที่อยู่ในผืนผ้ายันต์ ที่ผู้รู้เท่านั้นจึงจะอ่านออกและเข้าใจ) เพื่อลดการใช้คำอธิบายยืดยาว ยังมีข้อมูลตัวเลข การแทนข้อมูลด้วยรูปภาพการกระจายที่เป็นกราฟแท่ง หรือเส้นโค้งรูประฆังคว่ำ ทั้งสมมาตร มีความเบ้ และความโด่ง หากการเรียนการสอนในรายวิชาสถิติเบื้องต้นทำให้นักศึกษามองว่าเป็นวิชาที่ยาก มีสูตรที่ต้องท่องจำมากมาย เนื้อหาหน้าเบื่อแล้ว เมื่อนักศึกษามีทัศนคติความพึงพอใจเชิงลบแล้วก็จะไม่มีความตั้งใจที่จะเรียนรู้ ส่งผลกระทบต่อผลสัมฤทธิ์การเรียนสถิติเบื้องต้น การจัดการเรียนรู้ที่ผู้เรียนเรียนไม่รู้เรื่องหรือไม่เข้าใจหรือเข้าใจผิดอาจส่งผลให้นำสถิติไปใช้ในทางที่ไม่เหมาะสม ทำให้เกิดการตัดสินใจผิดพลาดมากมาย ดังนั้นจึงต้องมีการแก้ปัญหาด้านความพึงพอใจและผลสัมฤทธิ์ทางการเรียนในรายวิชาสถิติเบื้องต้น เนื่องจากวิชาสถิติเบื้องต้นเป็นพื้นฐานสำคัญสำหรับการศึกษาและทำงานวิจัยในศาสตร์อื่น ๆ (บรรทม สุระพร, 2558) ในประเทศไทย มีการศึกษาทัศนคติต่อรายวิชาสถิติของกลุ่มผู้เรียนระดับพื้นฐาน ระดับกลาง และระดับใช้งานกับนักศึกษาระดับปริญญาบัณฑิต และระดับบัณฑิตศึกษา ครู อาจารย์ และนักวิจัยที่เป็นเพื่อนกับผู้วิจัยในเครือข่ายสังคมออนไลน์ เฟซบุ๊กจำนวน 163 คน พบว่า กลุ่มผู้เรียนที่รู้สึกดีต่อวิชาสถิติก็มีแนวโน้มที่จะแสดงออกทางความสามารถทาง

ปัญหา และเห็นคุณค่าต่อวิชาสถิติในระดับดีด้วย รวมทั้งกลุ่มผู้เรียนที่เห็นว่าวิชาสถิติเป็นวิชายากแต่มีความน่าสนใจก็มีแนวโน้มที่จะมุ่งมั่นหรือทุ่มเทในการเรียนตามไปด้วย (อาฟีฟี ลาเต๊ะ, 2561)

ความเข้าใจสถิติเกี่ยวกับ นัยสำคัญ (significance) อำนาจทดสอบ (power of a test) ขนาดตัวอย่าง (sample size) และ ขนาดอิทธิพล (effect size) ที่ใช้ในการคำนวณขนาดตัวอย่างได้แก่ โปรแกรม G*Power (Buchner, Erdfelder, Faul, Lang, 2020) จะทำให้นักวิจัยสามารถใช้ผลจากโปรแกรมสำเร็จรูปนี้ได้อย่างมั่นใจมากขึ้น (Mysiak, 2020) นอกเหนือไปจากการคำนวณขนาดตัวอย่างจากสูตรของทาโร่ ยามาเน่ (Yamane, 1970) ที่ใช้เพียงจำนวนประชากร (population N) ค่าความคลาดเคลื่อนของการสุ่มตัวอย่าง (sampling error) หรือ สูตรของเคจซีและมอร์แกน (Krejcie, Morgan, 1970) ที่สูตรการคำนวณขนาดตัวอย่างเพิ่มค่าไคสแคว์ที่องศาอิสระ (degrees of freedom) เท่ากับ 1 ระดับความเชื่อมั่น 95% ($\chi^2 = 3.841$) และสัดส่วนของลักษณะประชากรที่สนใจ (ถ้าไม่ทราบกำหนดให้ $p=0.50$) ในบทความวิชาการมักจะพบการอ้างถึงการคำนวณขนาดตัวอย่างโดยใช้สูตรของทาโร่ ยามาเน่ หรือ สูตรของเคจซีและมอร์แกนอยู่เสมอ โดยผู้วิจัยไม่ให้รายละเอียดเกี่ยวกับค่า นัยสำคัญ และค่าสัดส่วน (p) เลย นอกจากนี้ ยังมีการกำหนดขนาดตัวอย่างโดยอ้างอิงงานวิจัยอื่นด้วย การสร้างความรู้ความเข้าใจจะทำให้ นักวิจัยสามารถใช้สถิติตามหลักวิชาการที่ควรจะเป็นได้ต่อไป

บทความนี้ได้ใช้การเล่าเรื่องอย่างง่าย การยกตัวอย่างเท่าที่จำเป็น การใช้สูตรสัญลักษณ์ให้น้อยลง เพื่อลดความเป็นวิชาการสถิติเต็มรูปแบบ เพื่อการเข้าใจในพื้นฐานวิชาสถิติและการกำหนดขนาดตัวอย่างในงานวิจัย โดยนำเสนอเริ่มต้นจากข้อมูล ตัวแปร การแจกแจงความถี่ ร้อยละ และความน่าจะเป็น และนำไปสู่การกระจายของการสุ่มตัวอย่าง (sampling distribution) ซึ่งเป็นที่มาของสูตรการคำนวณขนาดตัวอย่าง และการทดสอบสมมติฐาน การนำเสนอโดยลำดับนี้น่าจะมีส่วนทำให้เกิดความเข้าใจเกี่ยวกับนัยสำคัญ (significance) อำนาจทดสอบ (power of a test) ขนาดตัวอย่าง (sample size) และ ขนาดอิทธิพล (effect size) ได้ตามสมควร โดยเน้นความเข้าใจมากกว่าวิธีทำซึ่งในอดีตใช้การคำนวณด้วยมือหรือใช้คอมพิวเตอร์เขียนโปรแกรมซึ่งทำให้สถิติเป็นเรื่องยุ่งยากมาก แต่ในปัจจุบันมีโปรแกรมสำเร็จรูปวิเคราะห์ทางสถิติให้ใช้มากมาย เพียงแต่บันทึกข้อมูลตัวแปรต่าง ๆ เป็นข้อมูลนำเข้าตามรูปแบบที่โปรแกรมเหล่านั้นกำหนด ก็สามารถเรียกคำสั่งให้ทำการวิเคราะห์ผลได้แล้ว

สถิติเบื้องต้น

การแจกแจงความถี่ ร้อยละ และความน่าจะเป็น

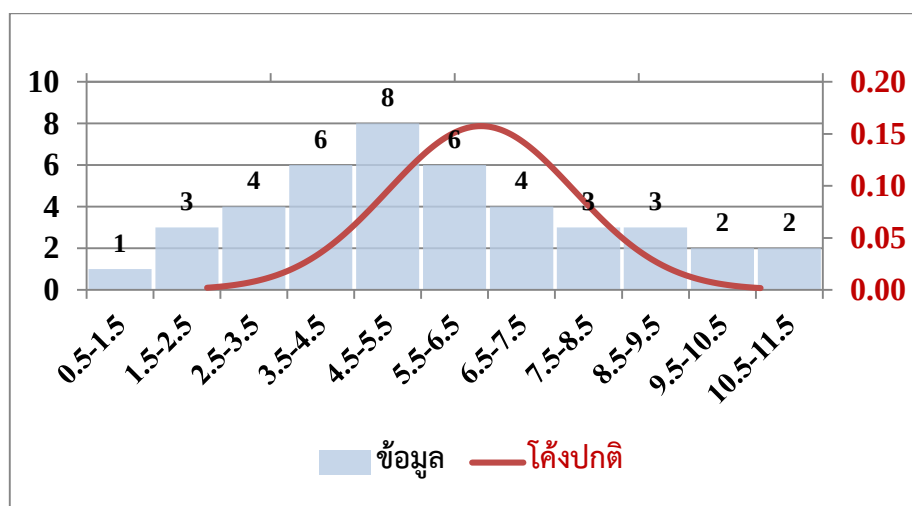
เริ่มต้นจากการเก็บรวบรวมข้อมูล จากนั้นนำข้อมูลมาดูการกระจายด้วยการแจกแจงความถี่ (frequency distribution) ก็คือ ผลที่ได้จากการแจกนับ (tally หรือ frequency count) การนับว่าสิ่งที่ศึกษามีค่าอยู่ในค่าที่กำหนดหรือช่วงค่าที่กำหนดก็รายการ ดังแสดงในตารางที่ 1 สดมภ์ความถี่ [a] เมื่อรวมกันเป็น ผลรวมความถี่ [b] ซึ่งมาจากข้อมูลที่ศึกษาจำนวนหนึ่ง แต่การจะไปเปรียบเทียบกับอีกกลุ่มหนึ่งซึ่งมีจำนวนรวมไม่เท่ากัน การแปลงเป็นสัดส่วนร้อยละในสดมภ์ [c] ก็จะทำให้การเปรียบเทียบทำได้ง่ายขึ้น เช่น สัดส่วนร้อยละระหว่างเพศ ส่วนการแปลงเป็นสัดส่วนที่มีผลรวมเป็น 1 ในสดมภ์ [d] ก็คือโอกาสความน่าจะเป็น (probability) นั่นเอง อาจแสดงในรูปกราฟแท่งว่าแต่ละค่าของข้อมูลมีข้อมูลจำนวนกี่รายการที่มีค่าอยู่ในช่วงค่านั้น (ใช้คำว่าช่วงค่านั้น ในงานแต่ละเรื่องจะเป็นค่าของสิ่งที่ศึกษา เช่น น้ำหนัก ส่วนสูง) ดังแสดงในภาพที่ 1 แกนตั้งด้านขวาคือ ความถี่ ส่วนแกนตั้งด้านซ้ายของภาพ ก็คือ ร้อยละ หรือ ความน่าจะเป็น ซึ่งการนับจำนวน กับร้อยละใช้สเกลต่างกันแต่ซ้อนบนภาพเดียวกันได้

ตารางที่ 1 ตัวอย่างการแจกแจงความถี่ แปลงเป็นสัดส่วนร้อยละ และความน่าจะเป็น

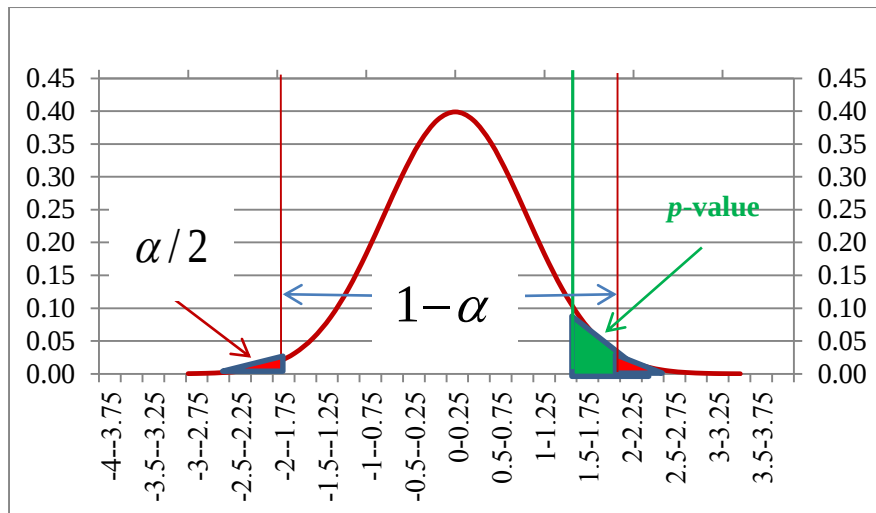
ค่าที่กำหนด	ช่วงค่าที่กำหนด	ความถี่ [a]	ร้อยละ [c] $=100 \times [a]/[b]$	ความน่าจะเป็น $[d]=[a]/[b]$
1	0.5-1.5	1	2.38%	0.02
2	1.5-2.5	3	7.14%	0.07
3	2.5-3.5	4	9.52%	0.10
4	3.5-4.5	6	14.29%	0.14
5	4.5-5.5	8	19.05%	0.19
6	5.5-6.5	6	14.29%	0.14
7	6.5-7.5	4	9.52%	0.10
8	7.5-8.5	3	7.14%	0.07
9	8.5-9.5	3	7.14%	0.07
10	9.5-10.5	2	4.76%	0.05
11	10.5-11.5	2	4.76%	0.05
รวม [b]		42	100.00%	1.00
ค่าต่ำสุด			1	
ค่าสูงสุด			11	
ค่าเฉลี่ย			5.71	
ค่าส่วนเบี่ยงเบนมาตรฐาน			2.53	

การวัดแนวโน้มเข้าสู่ส่วนกลาง

นอกจากนี้ยังมีการคำนวณแนวโน้มเข้าสู่ส่วนกลางเป็นการดูว่าข้อมูลแตกต่างกันอย่างไร ก็คือการกระจายของข้อมูล คือ ค่าต่ำสุด (minimum) ค่าสูงสุด (maximum) เมื่อนำค่าสูงสุดลบออกด้วยค่าต่ำสุดก็จะได้ค่าพิสัย (range) ซึ่งบอกว่าข้อมูลสูงสุดต่างไปจากข้อมูลต่ำสุดเท่าใด และข้อมูลมีศูนย์กลางของข้อมูลอยู่ที่ไหน โดยดูจากค่าเฉลี่ย (mean หรือ average) คือ ค่าข้อมูลทั้งหมดนำมารวมกันแล้วหารด้วยจำนวนข้อมูล (นอกจากนี้ยังมีค่ามัธยฐาน (median) คือ ค่าที่มีตำแหน่งอยู่กึ่งกลางของข้อมูลทั้งหมด เมื่อได้เรียงข้อมูลตามลำดับ ไม่ว่าจะจากน้อยไปมาก หรือจากมากไปน้อย ถ้าข้อมูลเป็นจำนวนคู่ก็นำเอาค่าของคู่ตรงกลางบวกกันแล้วค่าแล้วหารสอง กับฐานนิยม (mode) คือจำนวนข้อมูลที่มีค่าเดียวกันเป็นจำนวนมาก) และส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ที่เป็นการรวมข้อมูลรายการแต่ละรายการที่ลบด้วยค่าเฉลี่ยแล้วยกกำลังสองหารด้วยจำนวนข้อมูลแล้วจึงนำผลที่ได้ไปหาค่า (ถอด) สแควร์รูท (square root) เพื่อให้ได้ค่าเฉลี่ยขนาดของความแตกต่างไปจากค่าเฉลี่ย เพื่อใช้เป็นตัวแทนค่าของข้อมูลกลุ่มดังกล่าวที่สามารถนำไปเปรียบเทียบกับข้อมูลกลุ่มอื่น ๆ ที่เกี่ยวข้องกัน ถ้าค่าเฉลี่ยสองกลุ่มไม่เท่ากัน ก็สามารถบอกได้โดยว่าค่าเฉลี่ยต่างกันเท่าไร แต่ถ้าค่าเฉลี่ยสองกลุ่มเท่ากันแล้ว ข้อมูลกลุ่มไหนเกาะกลุ่มดีกว่ากัน กลุ่มที่ข้อมูลเกาะกลุ่มกันดีก็จะมีค่าส่วนเบี่ยงเบนมาตรฐานต่ำกว่ากลุ่มอื่น ตัวอย่างของการกระจายความถี่ของข้อมูลตารางที่ 1 มีค่าเฉลี่ยเท่ากับ 5.71 และ ส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 2.53 ดังแสดงในภาพที่ 1 ตัวเลขบนแกนตั้งด้านขวาของภาพคือจำนวนนับความถี่ (frequency) ของสิ่งที่สนใจซึ่งอยู่ในช่วงบนแกนนอน เมื่อแปลงค่าเป็นเส้นโค้งปกติตามทฤษฎีด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเดิมแล้ว (ขอละสูตรการกระจายความน่าจะเป็นของการกระจายแบบปกติไว้ (the standard normal probability distribution)) ก็จะได้พบความแตกต่างระหว่างข้อมูลที่เก็บมากับค่าเส้นโค้งปกติตามทฤษฎี ซึ่งมีค่าความน่าจะเป็นตามแกนตั้งด้านซ้ายของภาพ



ภาพที่ 1 รูปแสดงการกระจายของข้อมูลและค่าความน่าจะเป็นโค้งปกติ



ภาพที่ 2 รูปแสดงการกระจายของข้อมูลและโค้งปกติ

การกระจายของการสุ่มตัวอย่าง

ประชากร (population) คือ ทุก ๆ หน่วยของสิ่งที่นักวิจัยสนใจศึกษา การศึกษาจากประชากรที่มีจำนวนมากทุกรายการ เรียกว่า สำมะโน (census) ซึ่งมีค่าใช้จ่ายสูง ใช้เวลาและอาจเป็นไปได้ยาก การศึกษาจากตัวอย่าง (sample) โดยการเลือกตัวอย่าง (sampling) ก็จะประหยัดงบประมาณ เวลา และทันต่อการนำไปใช้ตัดสินใจ การศึกษาจากประชากรทั้งหมดที่มีรูปแบบการกระจายของข้อมูลที่เป็นการกระจายปกติตามที่เห็นเป็นรูปประจักษ์ว่าที่สมมาตรแล้ว ดังแสดงในภาพที่ 2

ถ้าเก็บข้อมูลจากประชากรที่มีการกระจายแบบโค้งปกติด้วยขนาดตัวอย่างเท่า ๆ กัน ซ้ำ ๆ หลายหลายครั้ง แล้วนำค่าเฉลี่ยแต่ละครั้งนั้นมาศึกษาการกระจาย จะเรียกว่า การกระจายของการสุ่มตัวอย่าง (sampling distribution) ก็จะมีการกระจายเป็นโค้งปกติ และค่าเฉลี่ยของค่าเฉลี่ย (the mean of the sample mean) นี้สามารถใช้ประมาณการค่าเฉลี่ยของประชากรได้ และส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยของตัวอย่างที่เก็บมาซ้ำ ๆ (the standard error of the mean) นี้จะเท่ากับ ส่วนเบี่ยงเบนมาตรฐานของประชากรหารด้วยสแควร์รูทของขนาดตัวอย่างที่ใช้ (Groebner, Shannon, Fry, Smith, 2001; Bluman, 1988; Hingle, Wiersma, Jurs, 1979) โดยปกติก็จะแสดงด้วยสัญลักษณ์ ดังนี้

ตารางที่ 2 สัญลักษณ์ที่ใช้ในสถิติ

รายการ	ประชากร	ตัวอย่าง	การสุ่มตัวอย่าง
จำนวนข้อมูล	N	n	n
จำนวนครั้ง	1	1	m
ค่าเฉลี่ย	μ	$\bar{x}$	$\bar{x}$
ส่วนเบี่ยงเบนมาตรฐาน	σ	s	$s_{\bar{x}}$

ซึ่ง $s_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ และสูตรนี้เองที่นำไปสู่การหาขนาดตัวอย่าง

สัญลักษณ์เหล่านี้ เป็นที่ยอมรับกันทั่วไป

การทดสอบสมมติฐาน

การกระจายของค่าเฉลี่ยของการเก็บตัวอย่างหลาย ๆ ครั้งนี้เอง นำไปสู่คำถามว่าในการเก็บตัวอย่างครั้งหนึ่งหรือทำการทดลองครั้งหนึ่งแล้วโอกาสที่พบค่าเฉลี่ยนั้นเป็นพวกเดียวกับประชากรกลุ่มนั้นหรือไม่ ซึ่งถ้าเป็นการทดลองที่มีการจัดกระทำ (intervention) ก็ย่อมต้องการเห็นผลแตกต่างไปจากประชากรปกติที่ไม่ได้มีการจัดกระทำตามบริบทของงานวิจัยนั้น ๆ ซึ่งเป็นการนำไปสู่การทดสอบสมมติฐานนั่นเอง (hypothesis testing) ที่นักวิจัยมักจะใช้คำศัพท์เฉพาะกับผลการทดสอบสมมติฐานเป็นคำสั้น ๆ ว่า Sig. หรือ ไม่ Sig. มาจากคำว่า significance หรือ non significance ซึ่งหมายถึง แตกต่างหรือไม่อย่างมีนัยสำคัญทางสถิติ

การกระจายของการสุ่มตัวอย่างจากประชากรที่มีการกระจายปกติ (normal distribution) การทดลองแต่ละครั้งก็คือ การเลือกตัวอย่างจำนวนหนึ่งมาศึกษา เมื่อแปลงเป็นค่ามาตรฐานแล้วอยู่ในส่วนใดโค้งปกติมาตรฐานดังแสดงในภาพที่ 2 ซึ่งถ้าอยู่ระหว่าง -3 ถึง +3 ก็อยู่ในร้อยละ 99.74 ที่ใกล้ค่าเฉลี่ยของประชากร ตามปกติมักจะตั้งเกณฑ์ให้ค่ามาตรฐานอยู่ที่ -1.96 ถึง +1.96 หรือ มีโอกาสที่จะไม่แตกต่างไปจากค่าเฉลี่ยร้อยละ 95 หรือจะเรียก ช่วงที่กำหนดนี้ว่า ช่วงความเชื่อมั่น (confidence interval $1-\alpha$) ส่วนที่เกินออกไปจาก -1.96 และ +1.96 ก็จะมีค่าด้านละร้อยละ 2.5 ($\alpha/2$) เมื่อรวมกัน 2 ข้างก็คือ ร้อยละ 5 (α) (สามเหลี่ยมฐานโค้งสีแดงในภาพที่ 2) ซึ่งจะเรียกว่า ความคลาดเคลื่อนชนิดที่ 1 (Type I error) หรือ α error (alpha error)

สมมติฐานหลัก (Null Hypothesis) คือ สมมติฐานที่ตั้งขึ้นซึ่งระบุว่าค่าเฉลี่ยของตัวอย่างเมื่อแปลงเป็นค่ามาตรฐานแล้ว ไม่แตกต่างไปจากค่าเฉลี่ยของประชากร

สมมติฐานทางเลือกหรือสมมติฐานการวิจัย (Alternative or Research Hypothesis) คือ สมมติฐานที่ตั้งขึ้นซึ่งระบุว่าค่าเฉลี่ยของตัวอย่างเมื่อแปลงเป็นค่ามาตรฐานแล้ว แตกต่างไปจากค่าเฉลี่ยของประชากร

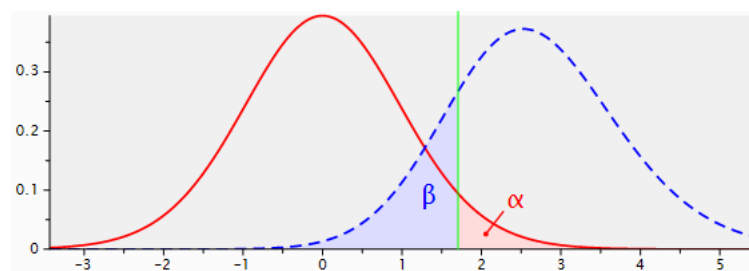
ความคลาดเคลื่อนชนิดที่ 1 (Type I error) หรือ ค่าอัลฟา (α) ซึ่งมักจะกำหนดไว้ที่ 0.05 ซึ่งถือเป็นค่ามาตรฐานที่ใช้กันทั่วไปในงานวิจัยหลาย ๆ สาขา เมื่อ $1 - \alpha$ คือ ช่วงความเชื่อมั่น (confidence interval) α คือ โอกาสที่จะปฏิเสธว่าค่ามาตรฐานที่แปลงมานี้แตกต่างไปจากค่าเฉลี่ยของประชากร แต่ถ้าสมมติฐานที่ตั้งขึ้นเป็นจริงแล้วปฏิเสธก็จะเป็นความคลาดเคลื่อนชนิดที่ 1 (Type I error) การกำหนดค่า α เป็น 0.05 และ 0.01 ก็หมายถึงว่ามีโอกาสร้อยละ 5 หรือร้อยละ 1 ที่จะปฏิเสธสมมติฐานหลัก ทั้ง ๆ ที่สมมติฐานหลักนั้นเป็นจริงนั้น การกำหนดค่า α ให้น้อยลงก็ต้องใช้ขนาดตัวอย่างมากขึ้นหรือในทางกลับกัน ถ้ากำหนดให้มีความคลาดเคลื่อนได้มากก็เพราะใช้จำนวนตัวอย่างน้อย (Prajapati, Dunne, Armstrong, 2009)

ขนาดตัวอย่างหรือจำนวนผู้เข้าร่วมการศึกษา (อาสาสมัคร) มีบทบาทสำคัญที่จะทำให้ผลลัพธ์มีนัยสำคัญหรือไม่ เมื่อต้องการให้ความแตกต่างระหว่างกลุ่มมีมากโดยใช้จำนวนตัวอย่างจำนวนน้อยเท่าที่ระดับนัยสำคัญกำหนด เช่น α น้อยกว่า 0.05 ตามทฤษฎีแล้วจะพบความแตกต่างอย่างมีนัยสำคัญเมื่อทำการศึกษาในตัวอย่างจำนวนมากพอ ซึ่งก็จะทำให้มีค่าใช้จ่ายสูงและก็มีความยากเพิ่มมากขึ้น (Mysiak, 2020)

ผู้ที่เสนอให้ใช้ค่า α เป็น 0.05 คือนักสถิติชื่อ Fisher ในปี พ.ศ. 2468 (ค.ศ. 1925) โดยถือเป็นระดับการตัดสินใจที่สะดวก (a convenient cutoff level) โดย 0.05 (Quinn, Keough, 2002) ก็คือ ในการทดลอง 20 ครั้งจะยอมให้ตัดสินใจผิดพลาดได้ 1 ครั้ง นั่นเอง

ค่าพี (p-value) เป็นค่าความน่าจะเป็นที่คำนวณได้จากตำแหน่งค่ามาตรฐานที่แปลงมาจากเฉลี่ยของกลุ่มตัวอย่าง (สามเหลี่ยมฐานโค้งสีเขียวที่ครอบคลุมสามเหลี่ยมฐานโค้งสีแดงด้านซ้ายของภาพที่ 2) ในการทดสอบสมมติฐานโปรแกรมสำเร็จรูปจะคำนวณค่า p-value มาให้ด้วย ในภาพที่ 2 ค่าพีมากกว่าค่าแอลฟาก็จะไม่ปฏิเสธสมมติฐานหลัก ภาษานักวิจัยจะพูดสั้น ๆ ว่า ไม่ Sig. แต่ถ้าค่าพีน้อยกว่าค่าแอลฟาก็จะปฏิเสธสมมติฐานหลัก ไปยอมรับสมมติฐานรองว่าค่าเฉลี่ยของการสุ่มตัวอย่างครั้งนี้แตกต่างไปจากค่าเฉลี่ยของประชากร

ความคลาดเคลื่อนชนิดที่สอง (Type II error) หรือ ค่าบีต้า (β) ก็คือโอกาสที่ยอมรับสมมติฐานว่าง ทั้ง ๆ ที่สมมติฐานว่าไม่เป็นจริง หรือมีโอกาสผิดพลาดที่จะยอมรับซึ่งสมมติฐานว่างเท่ากับ β ดังนั้นค่า $1 - \beta$ จึงเป็นโอกาสที่จะปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่าไม่เป็นจริง ค่า $1 - \beta$ นี้จึงถูกเรียกว่า อำนาจทดสอบ (Power of a test) ดังแสดงในภาพที่ 3 ซึ่งโดยทั่วไปจะใช้อำนาจทดสอบที่ร้อยละ 80 หรือ 0.80 (Araujo, Froyland, 2007)



ภาพที่ 3 ภาพจากโปรแกรม G*Power

การวิจัยเชิงทดลองไม่ได้ใช้จำนวนประชากรคำนวณขนาดตัวอย่าง เนื่องจากประชากรมีจำนวนมากในสูตรการคำนวณก็จะมีค่าอีกค่าหนึ่ง คือ ขนาดอิทธิพล (effect size) โดยความหมายคือ ขนาดของตัวแปรหรือสิ่งที่ศึกษาต้องต่างกันเท่าไรจึงจะยอมรับหรือปฏิเสธสมมติฐานว่าง เนื่องจากสมมติฐานว่าง กำหนดให้สิ่งที่ศึกษาไม่แตกต่างกันและสมมติฐานว่างเป็นจริงขนาดอิทธิพลก็จะเป็นศูนย์ แต่เมื่อสมมติฐานหลักไม่เป็นจริงแล้ว ในกรณีนี้ขนาดอิทธิพลย่อมมีค่ามากกว่าศูนย์กล่าวคือสิ่งที่ศึกษาอยู่นั้นต้องต่างกันเท่าไร จึงจะตัดสินใจได้ว่ามันต่างกัน ขนาดอิทธิพลจึงเป็นดัชนีของการปฏิเสธสมมติฐานว่าง การกำหนดขนาดอิทธิพลกำหนดโดยใช้ผลการวิจัยในรูปแบบเดียวกันจากที่มีมาก่อน แต่ถ้าไม่มีหรือหางานวิจัยที่ไม่สามารถเทียบเคียงได้แล้ว จะกำหนดขนาดอิทธิพลอย่างไร อย่างไรก็ตาม การวิเคราะห์อำนาจทดสอบ มีปัจจัยที่เกี่ยวข้อง 4 ประการ คือ อำนาจทดสอบ ($1 - \beta$) ระดับนัยสำคัญ (significance criterion α) ขนาดตัวอย่าง (sample size n) และ ขนาดอิทธิพล (effect size δ) ซึ่งเมื่อจะหาค่าของปัจจัยใดปัจจัยหนึ่งจะขึ้นกับการกำหนดค่าของปัจจัยสามประการที่เหลือ (Cohen, 1988) ซึ่งสามารถใช้โปรแกรม G*Power คำนวณได้

การกำหนดขนาดอิทธิพลในการทดสอบค่าเฉลี่ยสองกลุ่มที่ศึกษาว่าแตกต่างกันหรือไม่ ขนาดอิทธิพลคือ ความแตกต่างของค่าเฉลี่ยสองกลุ่มที่ศึกษาหารด้วยส่วนเบี่ยงเบนรวมของทั้งสองกลุ่ม ถ้าขนาดอิทธิพลเท่ากับ 1 แปลว่าค่าเฉลี่ยสองกลุ่มต่างกัน 1 ส่วนเบี่ยงเบนมาตรฐาน ถ้ามีค่าเป็น 2 ก็แปลว่าค่าเฉลี่ยสองกลุ่มต่างกันสองเท่าของส่วนเบี่ยงเบนมาตรฐาน และก็เป็นอย่างนี้ไปเรื่อย ๆ ซึ่งค่าส่วนเบี่ยงเบนมาตรฐานก็เหมือนกับค่ามาตรฐาน (ค่า z) 1 ส่วนเบี่ยงเบนมาตรฐานก็คือ ค่า z ที่เท่ากับ 1 กรณีนี้สูตรการคำนวณขนาดอิทธิพลก็เหมือนกับสูตรการแปลงค่าเป็นค่ามาตรฐาน (ค่า z) (McLeod, 2019) (ที่ $z = (x - \mu) / \sigma$ กับ $\delta = (\bar{x} - \mu) / \sigma$)

ขนาดอิทธิพลเป็นการวัดเชิงปริมาณถึงขนาดของสิ่งที่วัดที่ส่งผลของการทดลอง ขนาดอิทธิพลมากสัมพันธ์กับความสัมพันธ์ของสองสิ่งก็มากตามไปด้วย การกำหนดขนาดอิทธิพลเป็นงานที่ยุ้งยาก ในบางกรณีก็ใช้ค่าขนาดอิทธิพลจากการศึกษาที่ผ่านมา หรือทำการศึกษานำร่องหรือการใช้ผู้เชี่ยวชาญกำหนด การกำหนดขนาดอิทธิพลน้อยก็ต้องใช้ตัวอย่างจำนวนมากขึ้น (Prajapati, Dunne, Armstrong, 2009) อย่างไรก็ตาม Cohen (1988) ได้เสนอแนะมาตรฐานขนาดอิทธิพลเป็น เล็ก กลาง และใหญ่ ไว้ดังแสดงในตารางที่ 3

ตารางที่ 3 ค่าขนาดอิทธิพล

การทดสอบ	เล็ก	กลาง	ใหญ่
การทดสอบค่าเฉลี่ยสองกลุ่มแตกต่างกัน	0.20	0.50	0.80
การทดสอบค่าเฉลี่ยหลายกลุ่มแตกต่างกัน	0.10	0.25	0.40
การทดสอบไคสแควร์	0.10	0.30	0.50
ค่าสหสัมพันธ์เพียร์สัน	0.10	0.30	0.50

ที่มา: Cohen (1988)

ตัวอย่างการใช้โปรแกรม G*Power

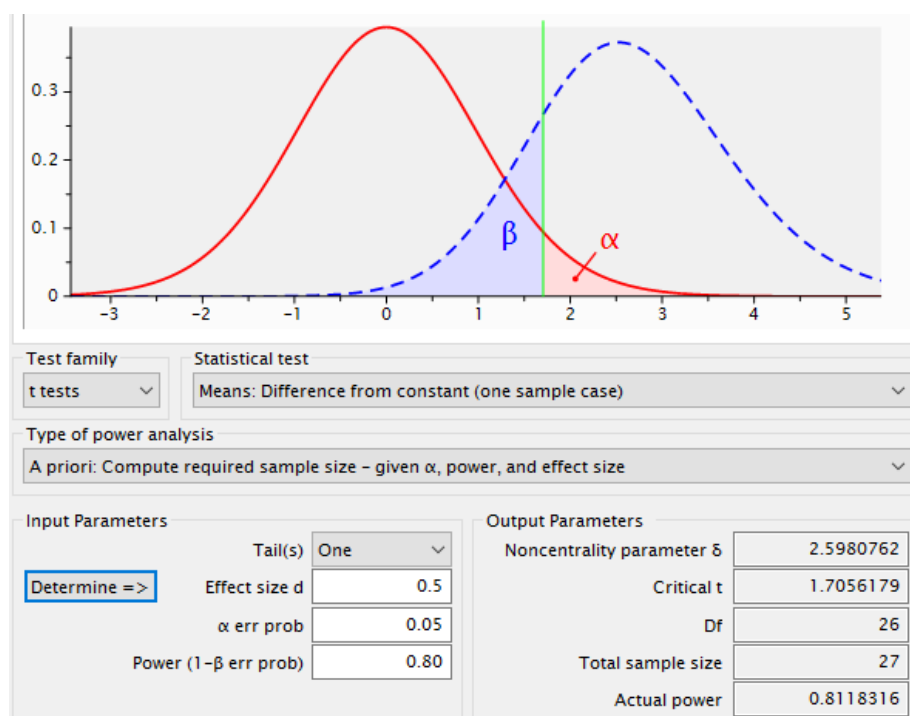
ความสัมพันธ์ของระดับนัยสำคัญ (significance level) อำนาจทดสอบ (Power of a test) ขนาดตัวอย่าง (sample size) ขนาดอิทธิพล (effect size) ในการทดสอบทางสถิติว่าค่าเฉลี่ยจากกลุ่มที่ทำการวิจัยจะแตกต่างไปจากค่าคงที่ค่าหนึ่งหรือไม่ พร้อมทั้งแสดงตัวอย่างการใช้โปรแกรม G*Power คำนวณขนาดตัวอย่างในการทดสอบสมมติฐานว่าค่าเฉลี่ยแตกต่างไปจากค่าที่กำหนดหรือไม่ โดยกำหนดเป็นตัวอย่าง ดังนี้

ระดับนัยสำคัญ (significance level) = 0.05

อำนาจทดสอบ (Power of a test) = 0.80

ขนาดอิทธิพล (effect size) = 0.50

โปรแกรม G*Power คำนวณขนาดตัวอย่าง (sample size) ได้เท่ากับ 27 ตัวอย่าง โดยแสดงค่า Actual power ด้วย เนื่องจากจำนวนตัวอย่างเป็นตัวเลขนับจำนวนเต็ม ผลการคำนวณที่แตกต่างไปจากค่าเริ่มต้นที่นำเข้าจึงนำมาแสดงไว้ด้วย ดังแสดงในภาพที่ 4 ในการวิเคราะห์อำนาจทดสอบ ผู้วิจัยสามารถเลือกใน Type of power analysis ได้ว่าจะให้คำนวณค่าใด เมื่อกำหนด 3 ค่าให้ เช่น ถ้าไปสำรวจแล้วแล้วเก็บตัวอย่างมาได้มากกว่าน้อยกว่าที่มากกว่า 27 ตัวอย่างแล้ว ถ้ายังกำหนดค่านัยสำคัญคงเดิม อำนาจทดสอบคงเดิม ค่าขนาดอิทธิพลจะเป็นเท่าใด เป็นต้น ขณะเดียวกัน ถ้าจะคำนวณหาอำนาจทดสอบโดยกำหนดขนาดอิทธิพลคงเดิมแล้ว ก็ต้องบันทึกข้อมูลเข้าเป็น นัยสำคัญ ขนาดตัวอย่าง และขนาดอิทธิพล



ภาพที่ 4 การคำนวณขนาดตัวอย่างด้วยโปรแกรม G*Power

ความรอบรู้สถิติเพื่องานวิจัย

การเริ่มต้นจากสถิติพื้นฐานถึงการวิเคราะห์อำนาจทดสอบที่มีความเชื่อมโยงกันของปัจจัยสี่ประการ คือ นัยสำคัญ อำนาจทดสอบ ขนาดอิทธิพล และขนาดตัวอย่าง และการยกตัวอย่างการใช้โปรแกรม G*Power ผู้วิจัยสามารถใช้โปรแกรมสำเร็จรูปนี้ในการคำนวณทั้งก่อนและหลังการทำงานวิจัย เมื่อจำนวนตัวอย่างแตกต่างไปจากที่ได้คำนวณและเสนอไว้ตามแผนงาน ซึ่งก็หวังว่าความรู้ความเข้าใจจากบทความนี้จะช่วยให้ผู้อ่านสามารถเขียนวิธีการวิเคราะห์อำนาจทดสอบและสรุปผลในรายงานการวิจัยหรือบทความวิจัยได้อย่างเหมาะสมต่อไป

บรรณานุกรม

- บรรทม สุระพร. (2558). ผลสัมฤทธิ์ทางการเรียนในวิชาสถิติเบื้องต้นและความพึงพอใจของผู้เรียนเมื่อใช้การจัดกลุ่มเรียนรู้เป็นทีมเทคนิค STAD. วารสารวิทยาศาสตร์ มข. ปีที่ 43 ฉบับที่ 3 : หน้า 552-563.
- อาฟีฟี ลาเต๊ะ. (2561). ทักษะคิดต่อรายวิชาสถิติของกลุ่มผู้เรียนระดับพื้นฐาน ระดับกลาง และระดับใช้งาน. Veridian E-Journal, Silpakorn University ฉบับภาษาไทย สาขามนุษยศาสตร์สังคมศาสตร์ และศิลปะ. ปีที่ 11 ฉบับที่ 1 เดือนมกราคม – เมษายน 2561 : หน้า 2591-2604.
- Annenberg Learner. [Online]. (2020). Against All Odds: Inside Statistics: **What Is Statistics?** Retrieved from <https://www.learner.org/series/against-all-odds-inside-statistics/what-is-statistics/>
- Araujo, P., Froyland, L. (2007), **Statistical power and analytical quantification**. Journal of Chromatography B. Vol. 847: p. 305-308.
- Bluman, A.G. (1998). **Elementary Statistics**. 3rd ed. Boston: WCB McGraw-Hill.
- Buchner, A., Erdfelder, E., Faul, F., Lang, A.-G. [Online]. (2020, October 15). **G*Power3.1 manual October 15, 2020**. Retrieved from https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/MathematischNaturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf
- Cohen J. (1988). **Statistical power analysis for the behavioral sciences**. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers; 1988.
- Groebner, D.F., Shannon, P.W., Fry, P.C., Smith, K.D. (2001). **Business Statistics A Decision-Making Approach**. New Jersey: Prentice-Hall.
- Hingle, D.E., Wiersma, W., Jurs, S.G. (1979). **Applied Statistics for the Behavioral Sciences**. Chicago: Rand McNally College.
- Krejcie, R. V., and Morgan, D. W. (1970). **Determining Sample Size for Research Activities**. Educational and Psychological Measurement. Vol. 30: p. 607 – 610.

- McLeod, S. [Online]. (2019). **SimplyPsychology. What does effect size tell you?** Retrieved from <https://www.simplypsychology.org/effect-size.html>
- Mysiak, K. [Online]. (2020, Aug 29). **Towards Data Science. The Relationship between Significance, Power, Sample Size & Effect Size.** Retrieved from <https://towardsdatascience.com/the-relationship-between-significance-power-sample-size-effect-size-899fcf95a76d>
- Mustafa, R.Y. (1996). **The Challenge of Teaching Statistics to Non-Specialists.** Journal of Statistics Education. Vol. 4 (No. 1): p. 1-10. DOI: 10.1080/10691898.1996.11910504
- Prajapati, B., Dunne, M., Armstrong, R. [Online]. (2009, November). **Sample size estimation and statistical power analyses.** Ot Peer Reviewed. 2009, November. Retrieved from https://www.researchgate.net/publication/2653_99772
- Quinn, G.P., Keough, M.J. (2002). **Experimental Design and Data Analysis for Biologists** (1st ed.). Cambridge, UK: Cambridge University Press. p. 46–69.
- Ruggeri, K., Dempster, M., Hanna, D., Cleary, C. (2008). **Experiences and expectations: The real reason nobody likes stats.** Psychology Teaching Review. Vol.14 (No.2): p. 75–83.
- Slootmaeckers, K., Kerremans, B., Adriaensen, J. (2014). **Too Afraid to Learn: Attitudes towards Statistics as a Barrier to Learning Statistics and to Acquiring Quantitative Skills.** Politics. Vol.34 (No. 2): p.191–200. doi:10.1111/1467-9256.12042
- Yamane, T. (1970). **Statistics: an introductory analysis.** New York: Harper and Row.