Continuing Education

การใช้สถิติสำหรับงานวิจัยวิทยาศาสตร์สุขภาพ: การออกแบบงานวิจัยและมาตรการวัด

Basic statistics for health science research: the study design and measurement scale

ยสนันท์ จันทรเวคิน

Yosananda Chantravekin

คณะทันตแพทยศาสตร์ มหาวิทยาลัยกรุงเทพธนบุรี

Faculty of Dentistry, Bangkokthonburi University

บทคัดย่อ ในการเรียนในระดับหลังปริญญา ปัญหาหนึ่งที่นักศึกษาพบคือความไม่เข้าใจเกี่ยวกับการใช้สถิติ ส่งผลต่อการแปลผลและการใช้ความคิดเชิงวิพากษ์ในขั้นตอนการทบทวนวรรณกรรม ไปจนถึงขั้นตอนการดำเนินงานวิจัย การเขียนรายงานวิจัย และการตีพิมพ์ผลงานวิจัย ในบทความนี้จะอธิบายเรื่องการออกแบบงานวิจัยและมาตรการวัดของตัวแปรแต่ละชนิด ซึ่งเป็นพื้นฐานสำคัญของการใช้สถิติที่เหมาะสม

คำสำคัญ ชีวสถิติ มาตรการวัด การออกแบบงานวิจัย

Abstract In postgraduate studying, some students did not understand and could not use the statistics well, leading to inability to interpret results and perform critical thinking in the literature review, as well as the problems in the research, and writing the report and manuscript steps. In this CE article, the classification of research types, and measurement scale, which are the basic issues for selecting the proper statistics would be reviewed.

***Keywords*:** biostatistics, measurement scale, research design

Correspondence to: Yosananda Chantravekin
Faculty of Dentistry
Bangkokthonburi University
E-mail: thaijomfs@gmail.com

J Med Glob 2025 Sep; 4(3)
Website: https://he01.tci-thaijo.org/index.php/JMedGlob
ISSN: 2821-918X (Online)

How to site this article: Yosananda Chantravekin. Basic statistics for health science research: the study design and measurement scale. J Med Glob. 2025 Sep; 4(3): 87 – 91.

บทนำ

สถิติเป็นพื้นฐานสำคัญของงานวิทยาศาสตร์สุขภาพ การสร้างองค์ความรู้ใหม่หรือการทำวิจัยล้วนแล้วแต่ต้องใช้สถิติในการทดสอบสมมติฐาน หรือแม้แต่การอ่านและประเมินความน่าเชื่อถือของบทความวิชาการก็ต้องอาศัยความรู้ความเข้าใจเกี่ยวกับสถิติเช่นเดียวกัน ในการเรียนการสอนระดับหลังปริญญา ไม่ว่าจะเป็นการทำวิทยานิพนธ์ในหลักสูตรสายวิชาการรวมทั้งการฝึกอบรมหลักสูตรแพทย์ประจำบ้าน หรือการทำสาระนิพนธ์ในหลักสูตรสายวิชาชีพ ผู้เรียนจำเป็นต้องมีความรู้พื้นฐานเกี่ยวกับสถิติเป็นอย่างดี สิ่งสำคัญในขั้นตอนการทบทวนวรรณกรรมคือการคิดเชิงวิพากษ์ภายหลังการสืบค้นบทความวิจัย ในการอ่าน วิพากษ์และการประเมินความน่าเชื่อถือ จำเป็นต้องมีความเข้าใจในสถิติพื้นฐาน เช่นเดียวกับการพัฒนาโครงร่างงานวิจัย (proposal developing) หลังจากผู้เรียนกำหนดระเบียบวิธีวิจัยเรียบร้อยแล้ว สิ่งสำคัญคือต้องระบุได้ว่าตัวแปรที่วัดมีมาตรการวัด (measurement scale) อยู่ในระดับใด เพื่อจะได้เลือกใช้สถิติที่เหมาะสม นำไปสู่การแปลผลที่ถูกต้อง สามารถเขียนรายงานการวิจัยและจัดเตรียมต้นฉบับเพื่อตีพิมพ์ได้เป็นอย่างดี จึงจะเห็นได้ว่าความชัดเจนในเรื่องชนิดของงานวิจัย มาตรการวัด และการเลือกใช้สถิติ ล้วนมีความเกี่ยวข้องสัมพันธ์กันเป็นอย่างมาก จึงเป็นที่มาของประเด็นต่างๆ ที่จะอธิบายในบทความศึกษาต่อเนื่องชุดนี้

ชนิดของงานวิจัย

งานวิจัยโดยเฉพาะงานวิจัยทางวิทยาศาสตร์สุขภาพ สามารถจำแนกได้หลายแบบด้วยกัน การจำแนกแบบง่ายที่สุดน่าจะเป็นการแบ่งเป็น 2 ประเภท คือ งานวิจัยเชิงพรรณนา (descriptive research) และงานวิจัยเชิงทดลอง (experimental research)

งานวิจัยเชิงพรรณนา หมายถึงงานวิจัยที่เป็นการสังเกตหรือวัดสิ่งที่เกิดขึ้นแบบ “as it is” โดยผู้วิจัยไม่ได้เข้าไปดัดแปลงแก้ไขสถานการณ์ใดๆ งานวิจัยเชิงพรรณนาอาจแบ่งย่อยได้เป็น 4 ประเภท ได้แก่ งานวิจัยเชิงสังเกต (observational study) งานวิจัยเชิงความสัมพันธ์ (correlation study) งานวิจัยที่มีการออกแบบเชิงพัฒนาการ (developmental design) และงานวิจัยเชิงสำรวจ (survey study) ซึ่งเป็นการใช้แบบสอบถาม

งานวิจัยเชิงสังเกต เป็นการเก็บข้อมูลจากตัวอย่างที่ศึกษา อาจเป็นลักษณะข้อมูลเชิงปริมาณ (quantitative) หรือข้อมูลเชิงคุณภาพ (qualitative) อาจเป็นข้อมูลในลักษณะความถี่ (frequency) หรือเป็นข้อมูลที่แสดงเป็นตัวเลข ซึ่งลักษณะข้อมูลที่ได้จะส่งผลต่อสถิติที่เลือกใช้

งานวิจัยเชิงความสัมพันธ์เป็นการหาความสัมพันธ์ระหว่างตัวแปรสองตัว ซึ่งสถิติที่เลือกใช้ขึ้นอยู่กับมาตรการวัดของตัวแปรนั้นๆ นอกจากนี้ยังมีสถิติอีกชนิดหนึ่งที่เกี่ยวข้องกับการหาความสัมพันธ์ แต่เป็นลักษณะการทำนายค่า กล่าวคือเมื่อทราบค่าตัวแปรหนึ่งจะทำนายค่าของอีกตัวแปรหนึ่ง สถิติดังกล่าวเรียกว่าการวิเคราะห์การถดถอย (regression analysis)

สำหรับการออกแบบเชิงพัฒนาการ ปกติจะแบ่งได้เป็น 2 ลักษณะ ลักษณะแรกคือ cross sectional design เป็นการเก็บข้อมูลกลุ่มตัวอย่างที่มีความหลากหลาย ณ เวลาหนึ่งเวลาใด อีกลักษณะคือ longitudinal study ซึ่งเป็นการเก็บข้อมูลกลุ่มตัวอย่างที่มีคุณสมบัติเหมือนกันไปเป็นช่วงระยะเวลา ตัวอย่างเช่น หากเราเก็บข้อมูลประวัติการตีพิมพ์ผลงานวิชาการของทันตแพทย์ ณ คลินิกทันตกรรม ชั้น 3 อาคารคณะทันตแพทยศาสตร์ ในวันเสาร์ที่ 6 กันยายน 2568 สมมติว่าในคลินิกตอนนั้นมีอาจารย์อาวุโส 3 คน อาจารย์รุ่นใหม่ 3 คน และอาจารย์พิเศษ 2 คน มีบัณฑิตที่จบปริญญาโทมาทำงานต่อเนื่อง 4 คน และมีนักศึกษาปริญญาโทปี 2 และปี 1 จำนวน 23 และ 27 คนตามลำดับ จะเห็นได้ว่าประสบการณ์การทำงานของกลุ่มตัวอย่างแต่ละคนมีความแตกต่างกันอย่างมาก อาจารย์อาวุโสอาจมีผลงานตีพิมพ์ทั้งในและต่างประเทศคนละหลายสิบเรื่อง อาจารย์รุ่นใหม่และอาจารย์พิเศษอาจมีผลงานตีพิมพ์ในวารสารในประเทศคนละ 4-5 เรื่อง มีผลงานในวารสารต่างประเทศคนละ 1-2 เรื่อง บัณฑิตและนักศึกษาปริญญาโทอาจมีผลงานตีพิมพ์คนละ 1-2 เรื่อง ในขณะที่บางคนอาจไม่เคยตีพิมพ์เลย การเก็บข้อมูลในลักษณะนี้ถือเป็น cross sectional study

ส่วนการเก็บข้อมูลแบบ longitudinal กลุ่มตัวอย่างจะมีคุณสมบัติตอนเริ่มต้นเหมือนกัน เช่น เก็บข้อมูลเฉพาะ

นักศึกษาปริญญาโทรหัส 68 (ที่จะจบปีการศึกษาในปี พ.ศ. 2570) และติดตามผลงานตีพิมพ์ในช่วงปี 2571 – 2575 ในกรณีนี้ผลงานตีพิมพ์อาจมีความแตกต่างกันตามที่ทำงานของบัณฑิต บัณฑิตบางคนที่ทำงานในคลินิกอาจมีผลงานตีพิมพ์เพียง 1 – 2 เรื่อง โดยอาจเป็นผลงานที่สืบเนื่องจากการเรียนปริญญาโท บัณฑิตที่รับราชการอาจมีผลงานตีพิมพ์ลงวารสารในประเทศ 3 – 4 เรื่อง เนื่องจากอาจต้องใช้ประกอบการเลื่อนขั้น ในขณะที่บัณฑิตที่ทำงานเป็นอาจารย์หรือศึกษาต่อในระดับที่สูงขึ้น อาจมีผลงานตีพิมพ์ในวารสารทั้งในและต่างประเทศจำนวนหนึ่ง ซึ่งจะสังเกตเห็นว่าการเก็บข้อมูลแบบ cross sectional กับ longitudinal กลุ่มตัวอย่างจะมีพัฒนาการที่แตกต่างกัน

ส่วนการเก็บข้อมูลด้วยแบบสอบถาม โดยทั่วไปจะมีการออกแบบแบบสอบถาม ปรึกษาผู้เชี่ยวชาญ ทดลองหาความเที่ยงและความตรง จากนั้นแจกแบบสอบถามไปยังกลุ่มเป้าหมาย เมื่อได้รับการตอบกลับ ผู้วิจัยก็จะทำการ tabulate ข้อมูล และใช้สถิติชนิดต่างๆ ในการวิเคราะห์ ซึ่งก็จะเป็นการเก็บและวิเคราะห์ข้อมูลในอีกลักษณะหนึ่ง

ส่วนงานวิจัยเชิงทดลองนั้น ผู้วิจัยมีการดัดแปลงหรือปรับสถานการณ์ให้กับผู้เข้าร่วมงานวิจัย โดยทั่วไปงานวิจัยในลักษณะนี้มักจะมีกลุ่มควบคุม และมีการปิดบัง ตัวอย่างเช่น การศึกษาผลของคาเฟอีนต่ออารมณ์ของผู้ดื่มกาแฟ หากเราออกแบบให้บาริสต้าทำการสุ่มว่าจะชงกาแฟแบบที่มีหรือไม่มีคาเฟอีนให้ลูกค้า หลังจากลูกค้าดื่มกาแฟไปประมาณสิบนาที ผู้วิจัยจะเข้าไปเก็บข้อมูลเกี่ยวกับอารมณ์ภายหลังการดื่มกาแฟ โดยกลุ่มทดลองเป็นกาแฟที่มีคาเฟอีน กลุ่มควบคุมไม่มีคาเฟอีน หากเราออกแบบให้ทั้งบาริสต้า ลูกค้า และผู้เก็บข้อมูลไม่ทราบว่าใครได้กาแฟแบบไหน ลักษณะ true experimental นี้อาจเทียบได้กับ triple blinded randomized control trial ซึ่งถือว่ามีความน่าเชื่อถือสูงสุด เพราะตัดความลำเอียงออกไปได้ทั้งหมด

ส่วนงานวิจัยคล้ายๆ กัน แต่ไม่ได้มีการสุ่มว่าจะชงกาแฟแบบใด ลูกค้าสามารถเลือกได้ว่าจะรับกาแฟที่มีหรือไม่มีคาเฟอีน บาริสต้าก็ทราบว่าลูกค้าได้รับกาแฟแบบใด อาจมีแต่ผู้เก็บข้อมูลที่ไม่ทราบ งานวิจัยนี้อาจเรียกว่าเป็น quasi-experimental ซึ่งอาจมีความลำเอียงมากกว่าเมื่อเทียบกับ true experimental แต่จะเห็นได้ว่ามีกลุ่มควบคุมเหมือนกัน และส่วนใหญ่จะเก็บข้อมูลเป็นตัวเลข ทำให้ส่วนใหญ่มักเลือกใช้สถิติ independent sample t-test ในการเปรียบเทียบค่าเฉลี่ยของทั้งสองกลุ่ม ทั้งงานวิจัยที่เป็น true experimental และ quasi-experimental

ส่วนงานวิจัยอีกแบบหนึ่งที่ผู้วิจัยไม่ได้ใส่ intervention และเป็นลักษณะการเก็บข้อมูลย้อนหลัง (retrospective study) โดยทราบผลที่เกิดขึ้นแล้ว จากนั้นทำการสืบค้นปัจจัยเสี่ยง เช่น การศึกษาผู้ป่วยที่ทราบแล้วว่าเป็นมะเร็งปอดหรือไม่ แล้วสืบค้นว่าผู้ที่เป็นมะเร็งปอดสูบบุหรี่หรือไม่ สูบมากน้อยเท่าใด กลุ่มที่ไม่เป็นมะเร็งได้สูบหรือไม่ มากน้อยเท่าใด แล้วนำมาเปรียบเทียบกัน งานวิจัย case control ในลักษณะนี้ตำราส่วนใหญ่จัดเป็น observational study แต่ก็มีตำราบางเล่มอนุโลมให้อยู่ในกลุ่มงานวิจัยเชิงทดลอง เพราะถือว่ามี intervention แม้ผู้วิจัยจะไม่ได้เป็นผู้กำหนดให้ก็ตาม

มาตรการวัด

มาตรการวัดของตัวแปรเป็นสิ่งสำคัญที่ผู้วิจัยต้องให้ความสำคัญ เพราะมีความเกี่ยวข้องกับสถิติโดยตรง โดยปกติมาตรการวัดแบ่งเป็น 4 ระดับ ได้แก่ มาตรานามบัญญัติ (nominal scale) มาตราเรียงลำดับ (ordinal scale) มาตราอันตรภาค (interval scale) และมาตราอัตราส่วน (ratio scale)

มาตรานามบัญญัติ คือการแบ่งกลุ่มตัวอย่างตามลักษณะเฉพาะ เช่น เพศ แบ่งเป็นชาย หญิง สีที่ชอบ แบ่งเป็นแดง เหลือง เขียว น้ำเงิน ในบางกรณีอาจมีการกำหนดตัวเลข เช่น 0 หมายถึงเพศชาย 1 หมายถึงเพศหญิง แต่ตัวเลขนี้ไม่ได้มีความหมายว่า 1 มากกว่า 0 และไม่สามารถนำมาคำนวณทางคณิตศาสตร์ได้

มาตราเรียงลำดับ มีการเพิ่มคุณสมบัติของการจัดอันดับหรือการมากกว่าน้อยกว่าเพิ่มขึ้นมา แต่ระยะระหว่างแต่ละลำดับอาจไม่เท่ากัน ในตำราส่วนใหญ่มักยกตัวอย่างลำดับการวิ่งเข้าเส้นชัยของนักวิ่ง โดยลำดับที่ 1 เข้าก่อนลำดับที่ 2 ลำดับที่ 2 เข้าก่อนลำดับที่ 3 โดยความห่างระหว่างลำดับ 1-2 และ 2-3 อาจไม่เท่ากัน เช่น ลำดับที่ 1 เข้าก่อนลำดับที่ 2 เป็นเวลา 2 วินาที ลำดับที่ 2 เข้าก่อน 3 เป็นเวลา 5 วินาที

เป็นต้น แต่จากประสบการณ์การสอน ลักษณะมาตราเรียงลำดับที่เราพบบ่อยกว่า น่าจะเป็น rubrics score ที่เราใช้ในการให้คะแนน เช่น การให้คะแนนในองค์ประกอบการรับและการเตรียมความพร้อมนักศึกษาในงานประกันคุณภาพการศึกษาดังที่เห็นในตาราง

ในการรายงานการดำเนินงานตามตัวบ่งชี้นี้ ให้อธิบายกระบวนการหรือแสดงผลการดำเนินงานอย่างน้อยให้ครอบคลุมประเด็นต่อไปนี้

- การรับนักศึกษา
- การเตรียมความพร้อมก่อนเข้าศึกษา

ในการประเมินเพื่อให้ทราบว่าอยู่ในระดับคะแนนใด ให้พิจารณาในภาพรวมของผลการดำเนินงานทั้งมด ที่ทำให้ได้นักศึกษาที่มีความพร้อมที่จะเรียนในหลักสูตร

เกณฑ์การประเมิน :

0	1	2	3	4	5
• ไม่มีระบบ • ไม่มีกลไก • ไม่มีแนวคิดในการกำกับติดตามและปรับปรุง • ไม่มีข้อมูลหลักฐาน	• มีระบบมีกลไก • ไม่มีการนำระบบกลไกไปสู่การปฏิบัติ/ดำเนินงาน	• มีระบบ มีกลไก • มีการนำระบบกลไกไปสู่การปฏิบัติ/ดำเนินงาน • มีการประเมินกระบวนการ • ไม่มีการปรับปรุง/พัฒนา/บูรณาการกระบวน การ	• มีระบบ มีกลไก • มีการนำระบบกลไกไปสู่การปฏิบัติ/ดำเนินงาน • มีการประเมินกระบวนการ • มีการปรับปรุง/พัฒนา/บูรณาการกระบวนการจากผลการประเมิน	• มีระบบ มีกลไก • มีการนำระบบกลไกไปสู่การปฏิบัติ/ดำเนินงาน • มีการประเมินกระบวนการ • มีการปรับปรุง/พัฒนากระบวนการจากผลการประเมิน • มีผลจากการปรับปรุงเห็นชัดเจนเป็นรูปธรรม	• มีระบบ มีกลไก • มีการนำระบบกลไกไปสู่การปฏิบัติ/ดำเนินงาน • มีการประเมินกระบวนการ • มีการปรับปรุง/พัฒนา/บูรณาการ กระบวน การจากผลการประเมิน • มีผลจากการปรับปรุงเห็นชัดเจนเป็นรูปธรรม • มีแนวทางปฏิบัติที่ดี โดยมีหลักฐานเชิงประจักษ์ยืนยัน และกรรมการผู้ตรวจประเมินสามารถให้เหตุผลอธิบายการเป็นแนวปฏิบัติที่ดีได้ชัดเจน

ตารางที่ 1 เกณฑ์การประเมินระดับคะแนนของกระบวนการรับและเตรียมความพร้อมนักศึกษา (สำนักงานคณะกรรมการการอุดมศึกษา. คู่มือการประกันคุณภาพการศึกษาภายใน พ.ศ. 2557)

จากตารางที่เป็น rubrics score จะเห็นว่าระดับคะแนนเริ่มจากระดับ 0 ถึง 5 โดยคะแนน 0 หมายถึงไม่มีระบบกลไกใดๆ คะแนน 1 หมายถึงมีระบบกลไก แต่ไม่สามารถแสดงหลักฐานว่าได้นำระบบกลไกดังกล่าวไปปฏิบัติ และเพิ่มขึ้นเรื่อย ๆ จนถึงคะแนน 5 หมายถึงมีระบบกลไก มีการปฏิบัติ มีการประเมินและปรับปรุงกระบวนการ จนสามารถใช้เป็นแนวทางปฏิบัติที่ดี หรือ Best Practice ได้ ซึ่งจะเห็นว่าความห่างระหว่างระดับคะแนนนั้นจะมีรายละเอียดที่ไม่เท่ากัน ทำให้ระยะห่างระหว่าง 0 กับ 1 อาจไม่เท่ากับระยะห่างระหว่าง 4 กับ 5 เนื่องจากความยากง่ายในการจัดเตรียมหลักฐานเอกสารนั้นมีความแตกต่างกัน

โดยทั่วไประดับคะแนนของมาตราเรียงลำดับ ไม่ควรนำมาใช้คำนวณทางคณิตศาสตร์ เช่น หาค่าเฉลี่ย (mean) หรือนำมาบวกลบคูณหาร แต่บางครั้งก็อาจอนุโลมให้นำมาคำนวณได้ เช่น การคำนวณค่าเฉลี่ยขององค์ประกอบต่างๆ ในงานประกันคุณภาพ

ส่วนมาตราอันตรภาคนั้น ระยะห่างระหว่างแต่ละตัวเลขจะเท่ากัน และมาตราอัตราส่วนจะเพิ่มการมีศูนย์แท้ ตัวอย่างเช่น องศาเซลเซียส ถือเป็นมาตราอันตรภาค เพราะระยะระหว่าง 10 องศากับ 11 องศา จะเท่ากับระยะระหว่าง 30 องศากับ 31 องศา แต่ 0 ขององศาเซลเซียสยังไม่ใช่ศูนย์แท้ เพราะศูนย์องศาเซลเซียสไม่ได้หมายความว่ามีอุณหภูมิและยังมีตัวเลขติดลบ ต่างจากศูนย์ขององศาเคลวิน ซึ่งถือว่าเป็นศูนย์แท้ องศาเคลวินจึงถือเป็นมาตราอัตราส่วน ใน

ชีวิตประจำวันตัวอย่างของมาตราอัตราส่วน เช่น ระยะทาง เงินเดือน ระยะทาง 0 ซม. คือไม่มีระยะทางเลย เช่นเดียวกับเงินเดือน 0 บาท คือไม่มีเงินเข้าเลย เป็นต้น

สรุปสถิติสำหรับการวิจัย

เมื่อเราทำความเข้าใจระเบียบวิธีวิจัยและมาตรการวัดแล้ว ในบทความฉบับนี้จะขอสรุปสถิติพื้นฐานที่เกี่ยวข้องกับลักษณะของงานวิจัยและมาตรการวัดแต่ละประเภทในตารางที่ 2 ส่วนรายละเอียดแนวคิดของสถิติแต่ละชนิดจะนำเสนอในบทความฉบับต่อไป

ชนิดของงานวิจัย	**มาตราอันตรภาค หรือ อัตราส่วน**	**มาตราเรียงลำดับ (หรือกรณีที่เป็นอันตภาค/อัตราส่วน แต่การกระจายข้อมูล/ความแปรปรวนไม่ผ่าน assumption)**	**มาตรานามบัญญัติ**
หาความสัมพันธ์ระหว่างตัวแปร	Pearson's r	Spearman's rho	Chi-square test
เปรียบเทียบค่าเฉลี่ยระหว่างสองกลุ่มที่เป็นอิสระต่อกัน	Independent t-test	Mann–Whitney U test	-
เปรียบเทียบค่าเฉลี่ยระหว่างสองกลุ่มที่เกี่ยวข้องกัน	Paired t-test	Wilcoxon signed-rank test	-
เปรียบเทียบค่าเฉลี่ยตั้งแต่สามกลุ่มขึ้นไปและเป็นอิสระต่อกัน*	One-way ANOVA	Kruskal–Wallis H test	-
เปรียบเทียบค่าเฉลี่ยตั้งแต่สามกลุ่มขึ้นไปและเกี่ยวข้องกัน*	Repeated measures ANOVA	Friedman test	-

ตารางที่ 2 สถิติพื้นฐานสำหรับงานวิจัยและมาตรการวัดแต่ละประเภท

* การทำ ANOVA ระบุได้เพียงมีอย่างน้อยหนึ่งกลุ่มที่มีค่าเฉลี่ยแตกต่างจากกลุ่มอื่น แต่จะเป็นคู่ใดต้องทำการทดสอบ Post hoc multiple comparison