สถิติเพื่อการแพทย์และการวิจัย101

ศุภมาส พันธ์เชย

บทคัดย่อ

สถิติและความรู้เรื่องการวิจัยทางระบาดวิทยา ทั้งหลักการ เครื่องมือ และการทดสอบต่างๆ เป็นเครื่องมือสำคัญสำหรับบุคลากรทางการแพทย์ที่จะใช้ในการตัดสินใจรักษา และใช้ในการพิจารณาความน่าเชื่อถือของบทความวิจัยทางการแพทย์ ในบทความนี้จะนำเสนอความรู้พื้นฐานอย่างคร่าวๆ เพื่อให้เข้าใจวัตถุประสงค์ ข้อดี และข้อจำกัดของรูปแบบการวิจัยแบบต่างๆ เพื่อจะได้นำไปปรับใช้ในการวิจัยและเวชปฏิบัติต่อไป

คำสำคัญ : สถิติ การวิจัย.

* นายแพทย์เชี่ยวชาญ, กลุ่มงานกุมารเวชกรรม โรงพยาบาลกระบี่, E-mail: suppamaspuncheuy@yahoo.com

Statistics for Medical Practice and Research 101

*Suppamas puncheuy**

Abstract

Statistics and epidemiological principles, measure, and tests are important tools that can used to facilitate clinical decision-making and are key in implementing and evaluating clinical research. This review provides an overview of basic knowledge to understand the purpose, advantages, and limitation of various study designs. To apply these concepts to clinical research and practice.

Keywords : Statistics, Research.

*Specialist doctor, Pediatrics work group Krabi hospital ,E-mail: suppamaspuncheuy@yahoo.com

บทนำ

สถิติเป็นส่วนสำคัญสำหรับบุคลากรทางการแพทย์ สำหรับการทำวิจัยและการอ่านบทความวิจัยต่างๆ และวิเคราะห์ถึงความน่าเชื่อถือของบทความนั้นๆ แต่ขณะเดียวกันสถิติเองก็เป็นเรื่องที่เข้าใจยาก สำหรับคนทั่วไป เพราะมีทั้งคำศัพท์เฉพาะที่ต้องแปลความหมาย ชื่อการวิเคราะห์แบบต่างๆ ที่ต้องทำความเข้าใจ เพื่อพิจารณาใช้งานได้อย่างถูกต้องกับรูปแบบการวิจัยและการวิเคราะห์ที่ต้องการ ดังนั้นในบทความนี้จะเน้นไปที่สถิติรูปแบบต่างๆอย่างคร่าวๆ ไม่ลงในรายละเอียดปลีกย่อยที่ผู้วิจัยสามารถไปหาอ่านในหนังสือสถิติต่างๆที่วางขายในท้องตลาดได้ เพราะในหนังสือส่วนใหญ่จะมีการยกตัวอย่าง แสดงวิธีทำทีละขั้นตอน ทำให้อ่านเข้าใจง่าย รวมถึงสอนวิธีการใช้คู่กับโปรแกรมสถิติสำเร็จรูป เหมาะสำหรับผู้ที่กำลังทำวิจัย ที่จะได้เรียนรู้ควบคู่ไปกับการใช้งานจริง และแก้ไขปัญหาไปทีละขั้นตอนจนสำเร็จ

สิ่งแรกสำหรับการวิจัยก็คือการคิดหัวข้อการวิจัย เราจะศึกษาอะไร เพื่อจุดประสงค์อะไร และวิธีการวิจัยที่เหมาะสมควรใช้วิธีอะไร เพื่อจะได้คำตอบสำหรับคำถามของเรา และสถิติที่เหมาะสมควรใช้ตัวไหน[1] ต่อด้วยรายละเอียดปลีกย่อย อันได้แก่ การเลือกกลุ่มประชากร เครื่องมือที่จะใช้ในการวิจัย มีความถูกต้อง เที่ยงตรง กระบวนการวิจัยปราศจากอคติและปัจจัยรบกวน สถิติมีความน่าเชื่อถือ ทั้งหมดนำไปสู่กระบวนการวิเคราะห์ข้อมูล และหาบทความอ้างอิงมาประกอบการวิเคราะห์ผลการวิจัย

สุดท้ายแต่ไม่ท้ายสุด ก็คือการเขียนบทความวิจัยและหาที่ตีพิมพ์เผยแพร่ผลงาน ดังตารางที่ 1

จากขั้นตอนเหล่านี้ สถิติ เป็นแค่ขั้นตอนหนึ่ง แต่ก็เป็นขั้นสำคัญสำหรับการวิเคราะห์ผลการวิจัย และตอบคำถามการวิจัยของเราให้มีความน่าเชื่อถือเชิงประจักษ์ได้

ตารางที่ 1 การออกแบบวิธีวิจัย[2]

ขั้นตอน	กระบวนการ
หัวข้อวิจัย	- คำถามงานวิจัย - วัตถุประสงค์งานวิจัย - ค้นคว้าหาข้อมูล - สร้างกรอบแนวคิด
กรอบแนวคิดการวิจัย	- เลือกวิธีวิจัย - เลือกกลุ่มตัวอย่างวิจัย - ขอจริยธรรมในมนุษย์ - เลือกพื้นที่ในการวิจัย - สร้างเครื่องมือในการวิจัย - การเก็บข้อมูล - การวิเคราะห์ข้อมูล
เผยแพร่ผลงานวิจัย	- ดูข้อกำหนดของแต่ละวารสารในการลงตีพิมพ์ - ปรับแก้ไขผลงาน ตามคำวิจารณ์ของผู้ทรงคุณวุฒิ

การเลือกใช้สถิติแบบไหนขึ้นอยู่กับวิธีวิจัย ซึ่งมีหลากหลายวิธี ดังตารางที่ 2 ที่ผู้วิจัยจำเป็นต้องเลือกให้เหมาะกับวัตถุประสงค์การวิจัยของตัวเอง

ตารางที่ 2 แสดงประเภทวิธีวิจัยแยกตามวิธีวิจัย และวิธีการวิจัยพื้นฐาน[2,3]

ประเภทวิธีวิจัย (Study design)			
การแบ่งกลุ่มตามการศึกษาเชิงสังเกต (Observational)		การศึกษาเชิงทดลอง (Experimental study)	
การศึกษาเชิงพรรณนา (Descriptive)	**การศึกษาเชิงวิเคราะห์ (Analytical)**	**การศึกษาเชิงทดลอง (Experimental)**	**การศึกษาเชิงไม่ทดลอง (Non- Experimental)**
- การศึกษาแบบสำรวจ (Survey study/ prevalence study) - การศึกษาระดับกลุ่มประชากร (Ecological study/ correlational study) - การศึกษาการเฝ้าระวังโรค (Surveillance) - การศึกษารายงานผู้ป่วย (Case report study) - การศึกษารายงานชุดผู้ป่วย (Case series study)	- การศึกษาแบบภาคตัดขวาง (Cross-sectional study/ prevalence study) - การศึกษาภาคตัดขวางแบบซ้ำ (Repeated- Cross-sectional study) - การศึกษาแบบติดตาม (Longitudinal study) - การศึกษาแบบโคฮอร์ท (Cohort study) - การศึกษาแบบผู้ป่วย-ไม่ป่วย (Case control study /case reference) - การศึกษาเชิงการจับคู่แบบผู้ป่วย-ไม่ป่วย (Matching case control)	- การศึกษาเชิงการทดลอง (Experimental study) - การศึกษาแบบสุ่มที่มีกลุ่มควบคุม (Randomized control trail: RCT)	- การศึกษาแบบกึ่งทดลอง (Quasi-experimental study)
การทบทวนวรรณกรรมอย่างเป็นระบบ (Systematic reviews/Meta-analysis)			

การเลือกวิธีวิจัยพิจารณาจากคำถามในการวิจัย ได้แก่ การทดสอบประสิทธิภาพของชุดตรวจ หาสาเหตุของโรค เปรียบเทียบผลการรักษา (treatment or intervention) หรือติดตามผลการรักษาระยะยาว เป็นต้น ดังตารางที่ 3 ซึ่งวิธีวิจัยแบบต่างๆ ก็มีข้อดีและข้อเสียแตกต่างกัน ผู้วิจัยควรศึกษาและเลือกให้เหมาะสมกับการตอบคำถามวิจัย เวลาที่ใช้ในการวิจัย รวมถึงทุนในการวิจัยนั้นๆด้วย ดังรายละเอียดในตารางที่ 4

ตารางที่ 3 วิธีวิจัยที่นิยมใช้สำหรับคำถามวิจัยประเภทต่างๆ[3]

วัตถุประสงค์การวิจัย			
ทดสอบประสิทธิภาพชุดตรวจ	**หาสาเหตุของโรค**	**เปรียบเทียบผลการรักษา (Treatment or Intervention)**	**ติดตามผลการรักษาระยะยาว (Long- term outcome)**
- การศึกษาแบบภาคตัดขวาง (Cross-sectional study) - การศึกษาแบบสุ่มที่มีกลุ่มควบคุม (Randomized controlled trials)	-การศึกษาแบบโคฮอร์ท (Cohort study): (E⟶O) การศึกษารายงานผู้ป่วย (Case-control study): (E⟵O) -การศึกษาแบบภาคตัดขวาง (Cross-sectional study): (E/O)	-การศึกษาแบบสุ่มที่มีกลุ่มควบคุม (Randomized controlled trials)	-การศึกษาแบบโคฮอร์ท (Cohort study) -การศึกษาแบบสุ่มที่มีกลุ่มควบคุม (Randomized controlled trials)

ตารางที่ 4 ข้อดี ข้อเสียของการวิธีวิจัยแบบต่างๆ[1,2]

การศึกษา	ข้อดี	ข้อเสีย
การศึกษาเชิงพรรณนา (Descriptive studies)		
-การศึกษาระดับกลุ่มประชากร (Ecological study/ correlational study)	ทำได้เร็ว, ใช้ทุนวิจัยไม่มาก ใช้ข้อมูลที่มีอยู่แล้ว (ข้อมูลทุติยภูมิ) สามารที่จะใช้ข้อมูลของคนเป็นจำนวนมาก สามารถใช้เป็นตัวแทนของประชากรทั้งหมดได้	ไม่สามารถทราบได้ว่าอันไหนเกิดก่อนเกิดหลัง ไม่สามารถควบคุมผลของปัจจัยรบกวนได้ ข้อมูลเป็นตัวแทนของค่าเฉลี่ยในประชากรไม่ใช่ค่าจริงๆของแต่ละคน ใช้ข้อมูลทุติยภูมิ มักจะไม่ครบถ้วน
-การศึกษารายงานผู้ป่วย (Case report study) -การศึกษารายงานชุดผู้ป่วย (Case series study)	อาจนำไปสู่ข้อสันนิษฐานใหม่เกี่ยวกับปัจจัยเสี่ยงของโรคได้ ซึ่งสามารถใช้เป็นสื่อข้อมูลได้ ได้เรียนรู้เกี่ยวกับวิธีการศึกษาใหม่ๆ ได้ความละเอียดของข้อมูล	ใช้เวลานาน ผลการรักษาที่ได้อาจจะยังไม่เหมาะสมกับประชากรอื่นๆ เสี่ยงต่อความเอนเอียงในการคัดเลือก (selection bias) การศึกษารายงานชุดผู้ป่วยปราศจากกลุ่มเปรียบเทียบที่เหมาะสม อาจจะนำไปสู่การสรุปผลที่ผิดพลาดได้
-การศึกษาแบบสำรวจ (Cross-sectional survey study/PREVALENCE STUDY)	หาสมมุติฐานใหม่ หาความถี่ของโรค ทำง่ายและใช้เวลาน้อย เหมาะกับการวางแผนงานด้านสาธารณสุข สามารถติดตามแนวโน้มของโรคหรือผลได้	ไม่ทราบว่าปัจจัยอะไรเกิดก่อนเกิดหลัง มีปัญหาเรื่องการให้ข้อมูลที่มีอคติจากการถามย้อนหลัง แยกแยะปัจจัยเสี่ยงกับปัจจัยที่ทำให้ผู้ป่วยอยู่รอดหรือตายได้ยาก ไม่เหมาะกับโรคหายากที่เกิดขึ้นไม่บ่อย โรคที่รุนแรงมีการตายอย่างรวดเร็ว หรือมีระยะฟักตัวสั้น

ตารางที่ 4 ข้อดี ข้อเสียของการวิธีวิจัยแบบต่างๆ[1,2] (ต่อ)

การศึกษา	ข้อดี	ข้อเสีย
การศึกษาเชิงวิเคราะห์ (Analytic studies)		
-การศึกษาแบบผู้ป่วย-ไม่ป่วย (Case control study)	ทำได้เร็วและใช้ทุนวิจัยไม่มาก เหมาะกับการศึกษาโรคที่มีระยะฟักตัวนาน เหมาะกับการศึกษาโรคหายาก เหมาะกับการศึกษาหลายๆปัจจัยที่ทำให้เกิดโรค(โรคเดียว)	ง่ายต่อการเกิดความเอียงเอนในการคัดเลือกและการนึกย้อนความทรงจำ (selection and recall bias) ทำให้การแปลผลการรักษาอาจจะเบี่ยงเบนไปจากความจริงได้ ศึกษาผลของปัจจัยสัมผัสที่พบไม่บ่อยนั้น ไม่ค่อยดีเท่าไร ไม่เหมาะต่อการศึกษาที่ไม่ทราบผลตั้งแต่เริ่มการรักษา หรือจำเป็นต้องวัดผลเป็นแบบตัวแปรต่อเนื่อง
- การศึกษาแบบโคฮอร์ท (Cohort study) : Prospective/Retrospective	มีปัญหาเรื่องอะไรเกิดก่อนเกิดหลังน้อย (temporal ambiguity) เพราะเหตุเกิดก่อนผล ภาวะการเป็นโรคไม่มีผลกระทบในการวัดปัจจัยสัมผัส สามารถศึกษาผลได้หลายอย่างในแต่ละปัจจัยเสี่ยงที่ศึกษา	สิ้นเปลืองค่าใช้จ่ายและเสียเวลามาก (prospective studies) เป็นไปได้ยากในแง่การปฏิบัติในกรณีที่เป็นโรคหายากหรือไม่ค่อยเกิดขึ้นบ่อย เกิดอคติในการวัดผลได้ กรณีผู้เข้าร่วมการศึกษาขาดการติดต่อหรือถูกตัดออกจากการศึกษาเป็นจำนวนมาก ส่งผลต่อการแปลผลได้ มีอคติในการเลือกกลุ่มตัวอย่าง จนส่งผลกระทบต่อการศึกษา ในการศึกษาย้อนหลัง ผลการศึกษาขึ้นอยู่กับความถูกต้องในแบบบันทึกข้อมูล
การศึกษาเชิงทดลอง (Experimental studies)		
- การศึกษาแบบสุ่มที่มีกลุ่มควบคุม (Randomized control trail: RCT)	ได้องค์ความรู้ใหม่ (strongest evidence of cause-effect relationship) ผลการวิเคราะห์มีความน่าเชื่อถือ (Gold standard ในการวิจัยระบาดวิทยา) การสุ่มช่วยลดการเกิดอคติและปัจจัยรบกวน ลักษณะของประชากรถูกระบุอย่างชัดเจน	ใช้เวลายาวนานและมีค่าใช้จ่ายสูง อาจมีปัญหาในทางปฏิบัติในการทดลอง เมื่อต้องปกปิดวิธีการรักษาต่อผู้เข้าร่วมการทดลอง ตัวอย่างที่มีคุณลักษณะต่างกัน กระจายตัวให้กลุ่มทดลองแต่ละกลุ่มไม่เท่ากัน ซึ่งจะมีอิทธิพลต่อการแปลผล อาจจะหากลุ่มตัวอย่างได้ไม่เพียงพอต่อการศึกษา ต้องพิจารณาปัจจัยทางด้านจริยธรรมการทดลองในมนุษย์อย่างเข้มงวด

ซึ่งการวิจัยวิธีต่างๆก็มีคะแนนความน่าเชื่อถือแตกต่างกัน เรียงจากน้อยไปมาก ได้แก่ รายงานผู้ป่วย (Case report) การศึกษาแบบภาคตัดขวาง (Cross-sectional study) การศึกษารายงานผู้ป่วย (Case-control study) การศึกษาแบบโคฮอร์ท (Cohort study) การศึกษาแบบสุ่มที่มีกลุ่มควบคุม (Randomized controlled trials) และมีที่มีความน่าเชื่อถือสูงสุดในพิรามิดของการวิจัยก็คือ การทบทวนวรรณกรรมอย่างเป็นระบบ (Systematic reviews/Meta-analysis)

เมื่อได้วิธีการวิจัยแล้ว การเลือกสถิติที่ใช้ก็ขึ้นอยู่กับชุดข้อมูลที่เก็บในการวิจัย หรือตัวแปรต่างๆที่ต้องการนำมาวิเคราะห์ ดังตารางที่ 5

ตารางที่ 5 สถิติที่ใช้กับตัวแปรต่างๆ[3]

Measurement scale	Two treatment groups, different individuals	More than two treatment groups, different individuals	Before&after single treatment, same individuals	Multiple treatment, same individuals	Association between two variations
Continuous (normal distributed Population*)	T test	Analysis of variance (ANOVA)	Pair t test	Repeated measure ANOVA	Linear regression/ correlation coefficient
Categorical	Chi-square		McNemar's test	Chi-square for trend	Relative risk/ Odds ratio
Ordinal scale	Mean-Whitney test	Kruskal-Wallis test	Wilcoxon Ranksum test	Freidman test	Spearman rank correlation
Time to event	Log rank test				

*ถ้าไม่สามารถยืนยันได้ว่ากลุ่มประชากรมีการกระจายตัวอย่างปกติ ให้จัดกลุ่มข้อมูลเป็นช่วงๆ ใช้ค่า median และใช้สถิติสำหรับ ordinal scale

ซึ่งสถิติเหล่านี้สามารถเลือกได้จากโปรแกรมสถิติสำเร็จรูป แต่สิ่งสำคัญก็คือการกรอกข้อมูลตัวแปรให้ถูกต้องตามลักษณะของข้อมูล ดังตารางที่ 6 ในการกรอกข้อมูลลงในตารางเก็บข้อมูลโปรแกรมสถิติต้องมีการกำหนดค่า เช่น 1 แทนค่าอายุ 0-5 ปี, 2 แทนค่าอายุ 6-10 ปี, 3 แทนค่าอายุ 11-15 ปี เป็นต้น และต้องกรอกความหมายของแต่ละตัวแปรเอาไว้ในช่องสำหรับกรอกความหมาย เพื่อเวลาคำนวณสถิติ จะได้แปลผลได้สะดวก[4]

ตารางที่ 6 จำแนกตัวแปรตามหน่วยการวัดตัวแปร ได้ 2 กลุ่ม[2]

ประเภท		**ตัวแปรตามหน่วยการวัดกลุ่ม 1**	
		ตัวแปรจัดกลุ่ม (Categorical)	**ตัวแปรต่อเนื่อง (Continuous)**
ตัวแปรตามหน่วยการวัดกลุ่ม 2	**ตัวแปรเชิงคุณภาพ (Qualitative)**	- มาตรานามบัญญัติ (Nominal scale) - มาตราอันดับ(Ordinal scale)	
	ตัวแปรเชิงปริมาณ (Quantitative)		- มาตราอันตรภาค (Interval scale) - มาตราอัตราส่วน (Ratio scale)

การพิจารณาตัวแปรจากระดับมาตราวัดของตัวแปร แบ่งได้เป็น 4 มาตราวัด ได้แก่

1.มาตรานามบัญญัติ (Nominal scale) เป็นการวัดเพื่อจัดกลุ่มหรือแยกประเภทตามลักษณะที่แตกต่างกัน เช่น เพศ ศาสนา อาชีพ ในการวิจัย ผู้วิจัยจะต้องกำหนดตัวแปรในระดับนี้ให้เป็นตัวเลข เช่น เพศชาย เป็นตัวเลข 1 เพศหญิงเป็นตัวเลข 2 เป็นต้น ซึ่งค่าตัวเลขดังกล่าวจะไม่มีความหมายในเชิงปริมาณ แต่มีความหมายเพียงเพื่อจำแนกกลุ่มเท่านั้น ตัวเลขในระดับนี้ไม่สามารถนำมาบวก ลบ คูณ หารได้ แต่สามารถนำมาจำแนกความถี่ ว่าแต่ละกลุ่มมีจำนวนเท่าใดได้

2.มาตราอันดับ/ลำดับ (Ordinal scale) การวัดในระดับนี้เป็นการวัดที่สูงกว่าระดับนามบัญญัติ สามารถบอกถึงลำดับของข้อมูลหรือตัวแปรได้ว่า มากหรือน้อยกว่า สูงหรือต่ำกว่า ก่อนหรือหลัง เช่น ระดับการศึกษา เป็นต้น แต่ไม่สามารถระบุได้ว่าความห่างของแต่ละลำดับมีปริมาณเท่ากันหรือไม่ ตัวเลขในระดับนี้จึงไม่สามารถนำมา บวก ลบ คูณ หารได้

3.มาตราอันตรภาค (Interval scale) การวัดในระดับนี้เป็นการวัดที่สูงกว่ามาตราอันดับ คือ นอกจากจะสามารถบอกถึงความแตกต่างแล้ว ยังสามารถบอกถึงปริมาณความแตกต่างได้ว่า ที่แตกต่างกันนั้นแตกต่างกันอยู่เท่าไร ตัวเลขในระดับนี้จึงสามารถนำมาบวก ลบกันได้ แต่ไม่สามารถ คูณ หารได้ เนื่องจากในระดับนี้ไม่มีจุดศูนย์แท้ (absolute zero) แต่เป็นเพียงค่าศูนย์สมมุติ (arbitrary zero) เช่น คะแนนการสอบ การสอบได้คะแนน 0 คะแนน ไม่ได้หมายความว่าไม่มีคะแนนเลย เพียงแต่ทำคะแนนไม่ได้เท่านั้น ซึ่งต่างจากการไม่ได้เข้าสอบซึ่งถือว่าไม่มีคะแนนเลย นอกจากนี้การไม่มีศูนย์แท้ยังทำให้การวัดในมาตรานี้ไม่สามารถบอกถึงอัตราส่วนหรือสัดส่วนของการมากน้อยกว่ากันได้ เช่น เมื่อเปรียบเทียบอุณหภูมิ 42 องศาเซลเซียส กับอุณหภูมิ 25 องศาเซลเซียส สามารถบอกได้ว่า ร้อนกว่ากัน 17 องศาเซลเซียส แต่ไม่สามารถบอกได้ว่าร้อนกว่ากันเป็นกี่เท่า

4.มาตราอัตราส่วน (Ratio scale) การวัดในระดับนี้เป็นการวัดที่สมบูรณ์ที่สุด คือ สามารถนำมาเรียงความสำคัญบอกความสูงต่ำ มากน้อยได้ บอกปริมาณความแตกต่างได้ และสามารถบอกถึงอัตราส่วนของความแตกต่างได้ด้วย ทั้งนี้ถือว่าระดับการวัดที่มีศูนย์แท้ ดังนั้นตัวเลขในระดับนี้สามารถนำมา บวก ลบ คูณ หารกันได้ เช่น น้ำหนัก ความสูง ความเร็ว รายได้ เป็นต้น[4,5]

ซึ่งตัวแปรทั้ง 4 ระดับนี้ อาจแบ่งเป็นสองกลุ่มตามลักษณะการจัดกลุ่ม เช่น ตัวแปรเชิงคุณภาพและตัวแปรเชิงปริมาณ หรือตัวแปรจัดกลุ่ม และตัวแปรต่อเนื่อง ดังตารางที่ 6 และเปรียบเทียบคุณสมบัติของการวัดระดับต่างๆดังตารางที่ 7

ตารางที่ 7 เปรียบเทียบคุณสมบัติของการวัดระดับต่างๆ[5,6]

	นามบัญญัติ	มาตราอันดับ	มาตราอันตรภาค	มาตราอัตราส่วน
จำแนกประเภท	/	/	/	/
กำหนดความแตกต่าง	/	/	/	/
บอกลำดับความแตกต่าง		/	/	/
บอกปริมาณความแตกต่าง			/ บวก ลบ	/ บวก ลบ คูณ หาร
ค่าศูนย์			มีศูนย์สมมุติ	มีศูนย์แท้

สถิติแบ่งออกเป็น 2 ประเภทคร่าวๆ ได้แก่สถิติเชิงพรรณนา ใช้สำหรับอธิบายคุณลักษณะทางประชากรของกลุ่มตัวอย่าง และสถิติเชิงอนุมาน ใช้สำหรับทดสอบสมมุติฐานการวิจัย การจะเลือกใช้สถิติใดในการวิเคราะห์ข้อมูลจะต้องรู้ว่าตัวแปรที่ศึกษาอยู่ในมาตราวัดใด สถิติที่ใช้ในการวิเคราะห์ข้อมูลจำแนกออกเป็น 3 ระดับ ได้แก่

1) Univariate analysis ใช้สำหรับวิเคราะห์ตัวแปรทีละตัว เช่น เพศ อายุ ระดับการศึกษา เป็นต้น สถิติที่ใช้ได้แก่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตราฐาน ฯลฯ

2) Bivariate analysis ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรทีละคู่ (ตัวแปรอิสระ 1 ตัว กับตัวแปรตาม 1 ตัว โดยไม่คำนึงถึงตัวแปรที่ 3) สถิติที่ใช้ได้แก่ Chi-square, t-test, One-way ANOVA, Correlation, Simple linear regression เป็นต้น

3) Multivariate analysis ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป กับตัวแปรตาม 1 ตัว สถิติที่ใช้ได้แก่ Multiple linear regression, Path analysis, Factor analysis, MANOVA เป็นต้น[5]

ค่าพารามิเตอร์ (Parameter) คือค่าต่างๆที่รวบรวมมาจากประชากรหรือคำนวณได้จากประชากร ใช้อักษรกรีกเป็นสัญลักษณ์

ส่วนค่าสถิติ (Statistic) คือค่าต่างๆที่รวบรวมมาจากกลุ่มตัวอย่างหรือคำนวณได้จากกลุ่มตัวอย่าง ใช้ตัวอักษรภาษาอังกฤษเป็นสัญลักษณ์ ดังตารางที่ 8

ตารางที่ 8 แสดงสัญลักษณ์ทางสถิติที่ใช้บ่อย[4,5]

	ค่าพารามิเตอร์	ค่าสถิติ
ค่าเฉลี่ยเลขคณิต	μ (มิว)	$\bar{x}$
ส่วนเบี่ยงเบนมาตรฐาน	σ (ซิกม่า)	S.D. / S
ความแปรปรวน	σ^2	$S.D.^2$ / S^2
สัมประสิทธิ์สหสัมพันธ์	ρ (โร)	r
ค่าสัดส่วน	P	p
ขนาด	N (ประชากร)	n (กลุ่มตัวอย่าง)

ในสถิติเชิงพรรณนา สถิติในการนับจำนวน ที่ใช้บ่อยได้แก่ ร้อยละ การแจกแจงความถี่ ค่ากลาง และค่าการกระจาย ซึ่งค่ากลางที่ใช้ได้แก่ ค่าเฉลี่ยเลขคณิต (Mean) ค่ามัธยฐาน (Median) ฐานนิยม (Mode) ซึ่งใช้ควบคู่กับค่าการกระจายตัวของข้อมูล ได้แก่ ค่าพิสัย (Range : R) ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation : S.D.) ซึ่งจะวัดการกระจายของข้อมูลเพียงชุดเดียว ถ้ามีข้อมูลมากกว่าหนึ่งชุด จะใช้ค่า สัมประสิทธิ์ของพิสัย (Coefficiency of range : C.R.) สัมประสิทธิ์ความผันแปร (Coefficient of variation : C.V.)[5] นอกจากนี้ยังค่าความคลาดเคลื่อนมาตรฐาน (Standard Error of Mean : SEM) ความแปรปรวน (Variance) ความเบี่ยงเบนควอไทล์ (Interquartile มีค่าเท่ากับค่า percentile ที่ 75 ลบด้วยค่า percentile ที่ 25)[1,4]

แต่ละวิธีก็มีข้อดี ข้อเสียแตกต่างกัน ดังตารางที่ 9 ค่าการวัดแนวโน้มเข้าสู่ส่วนกลางควรใช้ควบคู่กับค่าสถิติที่บอกการกระจายตัวของข้อมูล เพื่อจะบอกถึงลักษณะของข้อมูลได้ชัดเจนขึ้น

ความคลาดเคลื่อนมาตรฐาน (Standard Error of Mean) หรือส่วนเบี่ยงเบนมาตรฐานของการแจกแจงของค่าเฉลี่ยของตัวอย่าง หมายถึง ค่าที่แสดงว่าโดยเฉลี่ยแล้วค่าเฉลี่ยของตัวอย่างแต่ละตัวแตกต่างจากค่าเฉลี่ยของประชากรมากน้อยเพียงใด โดยคำนวณจากค่าส่วนเบี่ยงเบนมาตรฐานหารด้วย รากที่สองของขนาดตัวอย่าง

$$SE\ (\overline{X}) = \frac{SD}{\sqrt{n}}$$

ซึ่งค่า SEM นี้แตกต่างจากค่า Standard error (SE)

$$SE = \sqrt{p(1-p)/n}$$)

ซึ่งจะนำไปใช้คิดคำนวณค่าความเชื่อมั่นต่อ (95%Confidence Interval)

$$95\%\ CI = p \pm 1.96\sqrt{p(1-p)/n}$$

95% CI = p ±1.96

โดย p คือ probability of event

1-p คือ probability of no event

n คือ sample size

และ 1.96 มาจากปลายหางสองด้านของ Z-score แปลผลได้ว่า ถ้าการสังเกตนี้เกิดขึ้น 100 ครั้ง ในกลุ่มประชากรที่คล้ายๆกัน โอกาส 95 ครั้งที่จะเกิดเหตุการณ์นี้ จะอยู่ในช่วง p-1.96SE ถึง p+1.96SE[1]

ตารางที่ 9 ข้อดี ข้อเสียของการวัดแนวโน้มเข้าสู่ส่วนกลางและการกระจายตัวของข้อมูลแบบต่างๆ[1,2,7]

การวัด	ความหมาย	ข้อดี	ข้อเสีย
การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central tendency)			
ค่าเฉลี่ยเลขคณิต (Mean)	ค่าเฉลี่ยของผลรวมของข้อมูลทั้งหมด หาได้จากการนำผลรวมของข้อมูลทั้งหมดหารด้วยจำนวนข้อมูลทั้งหมด	เป็นการทดสอบสถิติพื้นฐานเพื่อวัดค่าแนวโน้มเข้าสู่ส่วนกลางอย่างละเอียดและนำไปใช้ในสถิติชั้นสูงต่อไป	ไวต่อค่าสุดโต่งหรือค่าผิดปกติ (outliers)
ค่ามัธยฐาน (Median)	ค่ากึ่งกลางของชุดข้อมูลหรือค่าที่อยู่ในตำแหน่งกึ่งกลาง	ช่วยหาค่ากลางได้อย่างรวดเร็ว แต่ไม่ละเอียด ไม่ได้รับผลกระทบจากค่าผิดปกติ ถ้าการกระจายตัวของข้อมูลไม่เป็นโค้งปกติ ค่ามัธยฐานนี้จะให้ข้อมูลได้มากกว่าค่าเฉลี่ย	ไม่เหมาะกับการทดสอบทางสถิติชั้นสูง
ค่าฐานนิยม (Mode)	ค่าข้อมูลที่มีความถี่สูงสุด	ช่วยบอกค่าที่พบบ่อย สามารถช่วยให้เห็นสาเหตุที่เป็นไปได้ของโรค	ไม่เหมาะกับการทดสอบทางสถิติชั้นสูง ในกราฟการกระจายตัวที่เป็นบวก (เบ้ขวา) อาจทำให้มีการแปลผลผิดเหมือนกราฟทวิฐานนิยม(Bimodal) ได้

ตารางที่ 9 ข้อดี ข้อเสียของการวัดแนวโน้มเข้าสู่ส่วนกลางและการกระจายตัวของข้อมูลแบบต่างๆ[1,2,7] (ต่อ)

การวัด	ความหมาย	ข้อดี	ข้อเสีย
การวัดการกระจายตัวของข้อมูล (Variability / spread / dispersion)			
ค่าพิสัย (Range) R = Xmax - Xmin	ความแตกต่างของค่าสูงสุดและค่าต่ำสุดของชุดข้อมูล	คำนวณไม่ยาก ง่ายต่อการเข้าใจ	พิสัยเหมาะกับชุดข้อมูลขนาดเล็กมากกว่าขนาดใหญ่ที่แนวโน้มค่าพิสัยจะสูงขึ้น ไม่เหมาะสำหรับการคำนวณสถิติชั้นสูงเพราะใช้ค่าคำนวณแค่2ตัวบอกผลอย่างคร่าวๆ ไวต่อค่าผิดปกติ
ความแปรปรวน (Variance (V) $V = SD^2 = \frac{\sum (x-\bar{x})^2}{n-1}$	ค่าเฉลี่ยของความแตกต่างระหว่างค่าแต่ละค่าในชุดข้อมูลกับค่าเฉลี่ยเลขคณิตของชุดข้อมูลนั้นยกกำลังสอง	บอกผลรวมความแตกต่างของแต่ละค่าข้อมูลกับค่าเฉลี่ย	ไวต่อค่าผิดปกติ
ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) $SD = \sqrt{V} = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$	ค่าที่บ่งบอกถึงการกระจายของข้อมูล เมื่อพิจารณาว่าค่าข้อมูลแต่ละค่าจะต่างจากค่ากลางมากน้อยเพียงใด หรือคือรากที่สองของความแปรปรวน	เมื่อการกระจายตัวของข้อมูลปกติ ค่าเฉลี่ยและส่วนเบี่ยงเบนจะบอกถึงการกระจายตัวของข้อมูลทั้งหมดได้	ไม่สามารถเปรียบเทียบค่าเบี่ยงเบนของกลุ่มตัวอย่างนี้กับค่าเฉลี่ยของกลุ่มตัวอย่างที่ขนาดแตกต่างกันตรงๆได้
สัมประสิทธิ์ของพิสัย (Coefficiency of range) $C.R. = \frac{Xmax-Xmin}{Xmax+Xmin}$	อัตราส่วนระหว่างผลต่างของค่าสูงสุด และค่าต่ำสุด หารด้วยผลบวกของค่าสูงสุดและต่ำสุดของข้อมูลชุดนั้น มันแสดงเป็นค่าร้อยละ	คำนวณไม่ยาก	ไม่เหมาะสำหรับการคำนวณสถิติชั้นสูงเพราะใช้ค่าคำนวณแค่2ตัวบอกผลอย่างคร่าวๆ
สัมประสิทธิ์ความผันแปร (Coefficient of variation) $C.V. = \frac{S.D.}{\bar{X}}$	อัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานหารด้วยค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น	ใช้เพื่อเปรียบเทียบ 2 กลุ่มตัวอย่างหรือมากกว่า ที่มีค่าเฉลี่ยแตกต่างกัน	ไม่ได้รับผลกระทบจากขนาดที่แตกต่างกันของค่าเฉลี่ย
ค่าความคลาดเคลื่อนมาตรฐาน (Standard Error of Mean) $SEM = \frac{SD}{\sqrt{n}}$)	ส่วนเบี่ยงเบนมาตรฐานของการแจกแจงของค่าเฉลี่ยของตัวอย่าง คำนวณจากค่าส่วนเบี่ยงเบนมาตรฐานหารด้วย รากที่สองของขนาดตัวอย่าง	ใช้เปรียบเทียบค่าเฉลี่ยของกลุ่มประชากรที่แตกต่างกัน บอกถึงช่วงค่าเฉลี่ยของแต่ละตัวอย่างอยู่ห่างจากค่าเฉลี่ยของประชากรมากน้อยเพียงใด	มักใช้ผิดกับค่าส่วนเบี่ยงเบน(SD) เพื่อเพิ่มค่าความเที่ยงตรง

ความแตกต่างระหว่างส่วนเบี่ยงเบนมาตรฐานกับความคลาดเคลื่อนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation : S.D) โดยเฉลี่ยแล้วข้อมูลแต่ละตัวแตกต่าง/อยู่ห่าง จากค่าเฉลี่ยมากน้อยเพียงใด

ความคลาดเคลื่อนมาตรฐาน (Standard Error of Mean : SEM) โดยเฉลี่ยแล้วค่าเฉลี่ยของตัวอย่างแต่ละตัวแตกต่าง/อยู่ห่างจากค่าเฉลี่ยของประชากรอยู่มากน้อยเพียงใด SEM ใช้ในกรณีบรรยายอธิบายหรือสรุปลักษณะของประชากร[1]

การทดสอบประสิทธิภาพชุดตรวจ/หาสาเหตุของโรค

ในการเข้าใจสาเหตุของโรคหรือการหาชุดตรวจโรคที่เหมาะสม ต้องการข้อมูลที่เหมาะสมจากการทดสอบในกลุ่มตัวอย่าง ซึ่งต้องผ่านการประเมินคุณภาพและความถูกต้อง ก่อนจะนำไปใช้ในประชากรทั้งหมด โดยการประเมินความถูกต้องจะประเมินผ่านค่าความไว (sensitivity) และค่าความจำเพาะ (specificity) ประเมินค่าการทำนาย ผ่านค่า positive และ negative predictive value การคิดคำนวณจะทำผ่านตาราง 2x2

ตารางที่ 10 ตาราง 2x2 ทดสอบคุณลักษณะ[1-3]

		Disease/Outcome		
		+	-	
Test/ Exposure risk	+	a True Positive (TP)	b False Positive (FP)	PPV $= \frac{TP}{(TP+FP)}$
	-	c False Negative (FN)	d True Negative (TN)	NPV $= \frac{TN}{(TN+FN)}$
OR $= \frac{a\ X\ d}{b\ X\ c}$		Sensitivity $= \frac{TP}{(TP+FN)}$	Specificity $= \frac{TN}{(TN+FP)}$	RR $= \frac{\frac{a}{a+b}}{\frac{d}{c+d}}$

- Sensitivity (Pr (T+/D+) หรือ “true positive rate” คือ สัดส่วนของผู้ป่วยที่ให้ผลการทดสอบเป็นบวกต่อผู้ป่วยทั้งหมด ชุดตรวจที่มีค่าความไวสูงเหมาะสำหรับการคัดกรองโรคที่มีความรุนแรงมากแต่สามารถรักษาได้ เพื่อให้ผู้ป่วยได้รับการวินิจฉัยที่เร็ว

- Specificity (Pr (T-/D-) หรือ “true negative rate” สูง หมายความว่าผู้ป่วยที่ให้ผลการตรวจเป็นบวกมีโอกาสที่จะเป็นโรคจริงสูง ชุดตรวจที่มีค่าความจำเพาะสูง เหมาะสำหรับการยืนยันการวินิจฉัย โดยเฉพาะที่ก่อให้เกิดผลเสียต่อผู้ป่วยทั้งด้านร่างกายและจิตใจสูง เช่น โรคมะเร็ง เป็นต้น

โดยทั่วไปเรามักจะคาดหวังให้วิธีการตรวจวินิจฉัยที่พัฒนาขึ้นมีทั้ง Sensitivity และ Specificity สูงสุด แต่มักเป็นไปไม่ได้ เมื่อเพิ่ม Sensitivity ให้สูงขึ้น การตรวจนั้นมักมี Specificity ลดลง ในทางตรงกันข้ามการตรวจที่มี Specificity สูงขึ้นมักมี Sensitivity ต่ำ วิธีการที่สามารถนำมาใช้เพื่อเลือกจุดตัดที่เหมาะสมได้ คือการสร้าง Receiver Operator Characteristic (ROC) curve คือสร้างกราฟความสัมพันธ์ระหว่าง true positive rate (Sensitivity) กับ false positive rate (1–Specificity) โดยการแปรค่าจุดตัด (cut – off point) ที่ใช้ต่างๆ กัน

- Positive predictive value (Pr(D+/T+) หมายถึงความน่าจะเป็นที่ผู้ป่วยจะเป็นโรคนั้นจริงเมื่อการตรวจให้ผลบวก การตรวจที่มีความจำเพาะสูงมักจะมีค่า PPV สูง

- Negative predictive value (Pr(D-/T-) หมายถึงความน่าจะเป็นที่ผู้ป่วยไม่ได้เป็นโรคนั้นเมื่อการตรวจให้ผลเป็นลบ การตรวจที่มีความไวสูงมักจะมีค่า NPV สูง

ในขณะที่ค่า Sensitivity และSpecificity ของการตรวจใดๆ มีค่าคงที่ ค่า predictive value กลับขึ้นอยู่กับความชุกของโรคนั้นในประชากรที่มีการนำการตรวจวินิจฉัยนั้นๆไปใช้ด้วย ดังนั้นแม้ว่าในงานวิจัยส่วนใหญ่มักคำนวณค่า predictive value ไว้ให้ด้วย ผู้วิจัยพึงระลึกไว้ว่าค่าเหล่านั้นมักนำไปใช้ในเวชปฏิบัติไม่ได้ เนื่องจากความชุกของโรคนั้นในประชากรทั่วไปมักมีความแตกต่างจากกลุ่มประชากรที่ใช้เป็นกลุ่มตัวอย่างในงานวิจัย และนอกจากนั้นความชุกของโรคยังอาจผันแปรตามกลุ่มอายุ เพศ ถิ่นที่อยู่ และปัจจัยอื่นๆ

- Likelihood ratios หมายถึงอัตราส่วนของความน่าจะเป็นของผลการตรวจในผู้ที่เป็นโรคเปรียบเทียบกับผู้ที่ไม่เป็นโรค ค่า Likelihood ratios มีความแปรตามความชุกของโรคในกลุ่มตัวอย่างน้อยกว่าค่า Predictive values

- ค่า Prevalence อาจเรียกอีกอย่างหนึ่งว่า Pretest probability ค่านี้สามารถนำมาแปลงเป็นค่า Pretest odds ได้ และจากนั้นก็สามารถนำไปหาค่า Posttest odds ได้โดยอาศัยสูตรในตารางที่ 11 หลังจากนั้นก็สามารถนำเอาค่าที่ได้มาแปลงเป็น Posttest probability เพื่อบอกความน่าจะเป็นที่ผู้ป่วยรายนั้นจะเป็นโรคที่ทำการตรวจวินิจฉัยได้ การคำนวณหาค่า posttest probability ตามวิธีที่แสดงไว้ข้างบนเป็นเรื่องค่อนข้างยุ่งยาก จึงมีการจัดสร้าง “likelihood ratio nomogram” หรือ “Fagan nomogram” เพื่อให้การใช้งานง่ายมากขึ้น[8]

- Odds ratio คือ การหาสัดส่วนของความสัมพันธ์ของปัจจัยเสี่ยง คิดคำนวณจาก กลุ่มตัวอย่างที่ได้รับปัจจัยเสี่ยง หารด้วยกลุ่มควบคุมที่ได้รับปัจจัยเสี่ยง ดังตารางที่ 10 และ 11เปรียบเทียบ prevalence ของปัจจัยเสี่ยง (exposure) ระหว่างสองกลุ่มของผลลัพธ์ (outcome) ที่แตกต่างกัน

- Relative risk คือการคำนวณหาความเสี่ยงของการเกิดโรคหรือความเสี่ยงสัมพันธ์ โดยคิดจากอุบัติการณ์ที่ได้รับความเสี่ยง หารด้วยอุบัติการณ์ที่ไม่ได้รับความเสี่ยง ดังตารางที่ 10 และ 11 เปรียบเทียบ incidence ของโรคระหว่างสองกลุ่มที่มีปัจจัยเสี่ยง (exposure) แตกต่างกัน[1,3]

ตารางที่ 11 วิธีการคำนวณค่าต่างๆในตาราง 2x2 [1,3,8]

ค่า	สูตรคำนวณ
Sensitivity	a / a + c
Specificity	d / b + d
Positive Predictive value	a / a + b
Negative Predictive value	d / c + d
Odds ratio	ad / bc = LR+/LR-
Relative risk : (I E+)/ (IE-)	[a / a + b] / [d / c + d]
Accuracy	[a + d] / [a + b + c + d]
Likelihood Ratios for positive test : LR+	[a / a + c] / [b / b + d] = True positive rate/ False positive rate = Sensitivity/ (1-specificity)

ตารางที่ 11 วิธีการคำนวณค่าต่างๆในตาราง 2x2 [1,3,8] (ต่อ)

ค่า	สูตรคำนวณ
Likelihood Ratios for negative test : LR-	[c / a + c] / [d / b + d] = False negative rate/ True negative rate = (1-sensitivity)/ Specificity
Prevalence or Pre-test probability	[a + c] / [a + b + c + d]
Pre-test odds	Pre-test probability / (1-pre-test probability)
Post-test odds	Pre-test odds X Likelihood ratio
Post-test probability	Post-test odds / (post-test odds + 1)
Absolute risk reduction (ARR)	[a / a + b] – [c / c + d] = (IE+) - (IE-)
Number needed to treat (NNT)	1/ARR

การสร้างและทดสอบสมมุติฐาน (Hypothesis generation and testing)

การศึกษาหรือวิจัยในทางการแพทย์จะมีการตั้งสมมติฐานเสมอ โดยทั่วไปจะตั้งไว้ 2 สมมติฐานดังนี้

1. สมมุติฐานหลัก Null hypothesis (H_0) คือสมมติฐานของความไม่แตกต่าง โดยระบุไว้ว่า ตัวแปร 2 ตัวนั้น ไม่มีความแตกต่างระหว่างกลุ่ม หรือไม่มีความสัมพันธ์กัน

2. สมมุติฐานทางเลือก Alternative hypothesis (H_1) หรือสมมุติฐานรอง คือตั้งไว้ว่า มีความแตกต่างระหว่างกลุ่ม หรือมีความสัมพันธ์ระหว่างตัวแปรทางใดทางหนึ่ง (บวกหรือลบ) ทั้งนี้สมมุติฐานหลักและรองต้องมีความหมายตรงข้ามกันเสมอ

เมื่อตั้งสมมติฐานแล้ว จึงไปเก็บข้อมูลกลุ่มตัวอย่าง นำผลไปวิเคราะห์ทางสถิติเพื่อทดสอบสมมติฐานว่า จะรับหรือปฏิเสธสมมติฐาน ความหมายก็คือ เนื่องจากผลที่ได้จากการศึกษาเป็นผลในตัวอย่าง (sample) หากพบว่ามีความแตกต่างระหว่างกลุ่มก็ยังไม่ทราบว่าในประชากร (population) ที่สุ่มตัวอย่างมา จะมีความแตกต่างระหว่างกลุ่มหรือไม่ เมื่อทำการทดสอบสมมติฐาน จะคำนวณได้ค่าที่บอกว่ามีนัยสำคัญทางสถิติหรือไม่ จากคำกล่าวที่ว่า การเกิดเหตุการณ์ที่มีโอกาสเกิดขึ้นได้ยาก ถือว่ามีนัยสำคัญทางสถิติ โดยทั่วไปมักกำหนดว่า หากเหตุการณ์นั้นเกิดน้อยกว่าร้อยละ 5 ถือว่ามีนัยสำคัญทางสถิติ ในภาษาทางสถิติคือ p-value < .05 (p = probability value) ในโปรแกรมสถิติทั่วไป เมื่อใส่ข้อมูลดิบเข้าไป เครื่องคอมพิวเตอร์จะคิด p-value ให้ ซึ่งสามารถแปลผลได้ว่ามีนัยสำคัญทางสถิติหรือไม่ เช่น ถ้ากำหนดนัยสำคัญทางสถิติไว้ที่ .05 เมื่อเครื่องคำนวณได้ค่า p-value เป็น .02 ก็หมายความว่ามีนัยสำคัญทางสถิติ หากได้ค่าเท่ากับ .05 หรือมากกว่าจะถือว่าไม่มีนัยสำคัญทางสถิติ อย่างไรก็ตามอาจกำหนดนัยสำคัญที่ระดับอื่น ๆ ได้ เช่น .01 เป็นต้น[7]

วิธีการทดสอบสมมติฐานที่ใช้บ่อยในเวชปฏิบัติได้แก่ t-test, chi-square และ correlation

t-test เป็นการทดสอบว่าค่าเฉลี่ยระหว่างกลุ่ม 2 กลุ่มแตกต่างกันหรือไม่ ตัวแปรเป็นแบบต่อเนื่อง (มาตราอันตรภาค/มาตราอัตราส่วน) จึงใช้ t-test ในการทดสอบสมมติฐาน t-test เป็นค่าสถิติที่นิยมใช้เปรียบเทียบระหว่างกลุ่มตัวอย่างเล็ก ๆ น้อยกว่า 30 คนและข้อมูลของกลุ่มตัวอย่างมาจากกลุ่มประชากรที่มีการแจกแจงเป็นโค้งปกติ ส่วนการเปรียบเทียบระหว่างกลุ่มตัวอย่างที่มีขนาดใหญ่มากกว่า 30 ขึ้นไป และไม่คำนึงถึงการแจกแจงของประชากร ใช้ Z - test สูตรการคำนวณ และวิธีใช้

ลักษณะเดียวกันกับ t - test ทุกประการ สถิติที่ใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยมากกว่าสองกลุ่มขึ้นไป สถิติที่นิยมใช้คือ F - test หรือ การวิเคราะห์ความแปรปรวน (ANOVA) การวิเคราะห์ความแปรปรวนเป็นวิธีการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มตัวอย่าง 3 กลุ่มขึ้นไป

Chi-square test เป็นสถิติที่วิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเชิงคุณภาพแบบกลุ่ม ที่มีมาตราวัดนามบัญญัติ/มาตราอันดับ การทดสอบไคสแควร์แบ่งออกได้ 3 ลักษณะ ได้แก่ การทดสอบความกลมกลืน การทดสอบความสัมพันธ์ระหว่างตัวแปร/ทดสอบความเป็นอิสระ และการทดสอบความเป็นเอกภาพ การทดสอบไคสแควร์มีหลักการในการแปลผลวิเคราะห์ดังนี้ เมื่อผลการวิเคราะห์พบค่าคาดหวัง (E) < 5 เกินกว่า 20 % ของจำนวนเซลล์ ให้อ่านผลการวิเคราะห์จากค่าสถิติ Fisher's Exact Test และเมื่อผลการวิเคราะห์พบค่าการคาดหวัง (E) < 5 ไม่ถึง 20% ของจำนวนเซลล์ให้อ่านผลการวิเคราะห์จากค่าสถิติ Pearson Chi-Square การทดสอบไคสแควร์จะได้ผลดีเมื่อกลุ่มตัวอย่างมีขนาดใหญ่

Correlation สหสัมพันธ์ เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณที่เป็นแบบต่อเนื่องทั้งคู่ (มาตราอันตรภาค/มาตราอัตราส่วน) ในการพิจารณาความสัมพันธ์ระหว่างตัวแปรว่ามีมากน้อยเพียงใดนั้นจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient : r) ของเพียร์สัน เป็นค่าวัดความสัมพันธ์ ค่า r จะมีค่าได้ระหว่าง -1 ถึง +1 ถ้าค่า r มีค่าบวก สูงใกล้ 1 แสดงว่าสัมพันธ์กันดี แต่ถ้าเป็นลบ แสดงว่าสัมพันธ์ผกผันกัน ถ้าค่า r มีค่าเท่ากับศูนย์แปลว่าตัวแปรสองตัวไม่มีความสัมพันธ์กัน การตีความหมายของค่าสัมประสิทธิ์สหสัมพันธ์ จะต้องพิจารณาทั้งขนาดและทิศทางของความสัมพันธ์ และต้องดูที่ p-value ด้วยว่ามีนัยสำคัญทางสถิติหรือไม่

การเลือกสถิติที่ใช้ขึ้นอยู่กับหลายปัจจัยทั้งชนิดของตัวแปร รูปแบบของการวิจัย กลุ่มเปรียบเทียบที่ใช้ ความสัมพันธ์ที่ต้องการทดสอบ โดยสามารถดูได้คร่าวๆจากตารางที่ 5 หรืออ่านในหนังสือสถิติต่างๆถึงเงื่อนไขอย่างละเอียดของสถิติชนิดต่างๆ[4-7,9]

ค่าความเชื่อมั่น (Confidence interval)

ในการศึกษาหรือวิจัยเรามักจะทำในกลุ่มตัวอย่าง ดังนั้น ค่าที่ได้จึงเป็นค่าในกลุ่มตัวอย่าง แต่เราอยากรู้ค่านั้นในประชากรที่สุ่มตัวอย่างมา จึงต้องหาค่าดังกล่าวซึ่งค่านี้ก็คือ confidence interval ค่าในกลุ่มตัวอย่างเรียกว่า point estimate ค่าในประชากรเรียกว่า population estimate

ตัวอย่าง : อัตราชุกของโรคความดันโลหิตสูงในกลุ่มตัวอย่าง = 20% (สำรวจ1000 คน พบ 200 คน) คำนวณค่า 95 C.I. ได้เท่ากับ 18.2% – 22.1% แปลผลว่า อัตราชุกของโรคเบาหวานในประชากรที่ถูกสุ่มตัวอย่างมาน่าจะมีค่าอยู่ระหว่าง 18.2% ถึง 22.1% ด้วยความเชื่อมั่น 95%

การแปลผล C.I. ของการศึกษาอื่นๆ ไม่ว่าจะเป็น relative risk, odds ratio, correlation coefficient, RRR, ARR, NNT ฯลฯ ก็เป็นไปในทำนองเดียวกัน

การศึกษาที่หาค่า relative เช่น R.R., O.R. ถ้า 95% C.I. คร่อม 1 แสดงว่า p-value > .05 ไม่มีนัยสำคัญทางสถิติ (non-significant)

การศึกษาที่หาค่า relative เช่น R.R., O.R. ถ้า 95% C.I. ไม่คร่อม 1 แสดงว่า p-value < .05 มีนัยยะสำคัญทางสถิติ (significant)

การศึกษาที่หาค่า absolute เช่น ARR., difference ถ้า 95% C.I. คร่อม 0 แสดงว่า p-value > .05 ไม่มีนัยยะสำคัญทางสถิติ

การศึกษาที่หาค่า absolute เช่น ARR., difference ถ้า 95% C.I. ไม่คร่อม 0 แสดงว่า p-value < .05 มีนัยสำคัญทางสถิติ[7]

ความคลาดเคลื่อน (Type I (alpha) and Type II (beta) errors)

ผลการศึกษาวิจัยย่อมมีความคลาดเคลื่อนได้ ความคลาดเคลื่อนมี 2 แบบได้แก่

แบบที่ 1 สรุปผลว่ามีความแตกต่างในขณะที่ความจริงไม่มี เรียกว่า เกิด Type I (alpha) error คือ การปฏิเสธสมมุติฐานหลัก เมื่อเป็นจริง

แบบที่ 2 สรุปผลว่าไม่มีความแตกต่างในขณะที่ความจริงมี เรียกว่า เกิด Type II (beta) error คือ การยอมรับสมมุติฐานหลัก เมื่อเป็นเท็จ ดังตารางที่ 12

ก่อนการศึกษาวิจัย มักจะกำหนด alpha error ไว้ที่ 0.05 และกำหนด beta error ไว้ที่ 0.2 นั่นคือยอมรับ

ความคลาดเคลื่อนแบบ Type I ร้อยละ 5 และยอมรับความคลาดเคลื่อนแบบ Type II ร้อยละ 20 โดยค่าความคลาดเคลื่อนสองประเภทนี้มักจะสวนทางกัน ในการวิจัยต้องพยายามควบคุมความคลาดเคลื่อนสองประเภทนี้ให้เกิดน้อยที่สุด โดยการลดความคลาดเคลื่อนด้วยการเพิ่มขนาดกลุ่มตัวอย่าง โดยนิยมคำนวณขนาดกลุ่มตัวอย่างโดยใช้โปรแกรม เช่น G*power ที่สามารถใช้ได้กับสถิติหลายรูปแบบ

ค่าอำนาจการทดสอบ (Power) ได้จากการคำนวณโดยใช้ค่า β จะมีค่าเท่ากับ 1 – beta error หมายความว่าเมื่อกำหนด beta error ไว้ที่ 0.2 แสดงว่ามี power เท่ากับ 0.8 หรือร้อยละ 80 นั่นเอง (งานวิจัยโดยทั่วไปควรกำหนด power ไว้ประมาณร้อยละ 70-90 ยิ่งค่าสูง ยิ่งต้องใช้ทุนวิจัยมาก ใช้กลุ่มตัวอย่างมาก)

ในการทดสอบทางสถิติทุกครั้งต้องมีการกำหนดค่านัยสำคัญทางสถิติ ซึ่งได้แก่ ค่า α น้อยละ 0.05 เพื่อนำมาใช้เป็นเกณฑ์ในการตัดสินผลการทดสอบสมมุติฐานว่าจะยอมรับหรือปฏิเสธสมมุติฐานหลัก[1,2,7,9-10]

ตารางที่ 12 แสดงความคลาดเคลื่อนในการทดสอบสมมุติฐาน[2,5,9-10]

ข้อสรุปจากการศึกษา	**ความเป็นจริง**	
	H_0 เป็นจริง	**H_0 เป็นเท็จ**
ปฏิเสธ H_0	Type I error (α)	ตัดสินใจถูกต้อง 1- β
ยอมรับ H_0	ตัดสินใจถูกต้อง 1- α	Type II error (β)

นอกจากนี้ความคลาดเคลื่อนในการศึกษาวิจัยยังแบ่งเป็นสองกลุ่ม คือ random error และ systemic error

Random error คือ ความคลาดเคลื่อนเกิดขึ้นโดยบังเอิญ ทำให้ค่าวัดได้จากกลุ่มประชากรแตกต่างจากค่าเป็นจริง ส่งผลให้ความสัมพันธ์ของปัจจัยที่ศึกษากับผลลัพธ์ เกิดความบิดเบือนได้ เช่น ความแตกต่างทางชีวภาพของกลุ่มประชากร (Biological variation) การสุ่มตัวอย่างผิดพลาด (Sampling error) การวัดผิดพลาด (Measurement error) อาจแก้ไขได้โดยการเพิ่มจำนวนประชากร มีการควบคุมมาตรฐานของเครื่องมือและอบรมผู้ใช้เครื่องมือ เป็นต้น การศึกษาที่ควบคุมความคลาดเคลื่อนนี้ดีพอ จะสามารถนำผลไปใช้เป็นสถิติอ้างอิงหรือ การอนุมานเชิงสถิติได้ (Statistical inference)

Systemic error (bias) คือ ความคลาดเคลื่อนเกิดจากขั้นตอนของวิธีการศึกษาวิจัย ความคลาดเคลื่อนของข้อมูลปัจจัยเสี่ยงต่างๆ ส่งผลให้ผลลัพธ์การศึกษาไม่ถูกต้อง เช่น การเลือกกลุ่มตัวอย่างที่ไม่คล้ายกัน (Selection bias) ซึ่งเป็นอคติที่พบบ่อยที่สุด โดยเฉพาะใน case control study ผู้วิจัยพึงระวังว่ากลุ่มตัวอย่างที่คัดเลือกในโรงพยาบาลอาจไม่ใช่ตัวแทนของประชากรทั้งหมดได้ กลุ่มประชากรหายไประหว่างการศึกษา (Lost of follow up bias) คุณภาพของแบบสอบถามที่ไม่ชัดเจน อคติของผู้ให้คำตอบ (Information bias) ผู้วิจัยวิเคราะห์ข้อมูลแยกย่อยเพื่อหาความสัมพันธ์ทางสถิติทั้ง ๆที่ไม่ได้ตั้งสมมุติฐานเกี่ยวกับตัวแปรนี้มาก่อน เรียกว่า “ fishing expedition” (Data analysis bias) ผู้วิจัยสรุปผลการวิเคราะห์ในระดับประชากรแต่ไปอธิบายสิ่งที่เกิดในระดับบุคคลหรือ ecologic fallacy (Conclusion bias) งานวิจัยที่มีนัยสำคัญทางสถิติมักได้รับการตีพิมพ์มากกว่าที่ไม่มี ทำให้มีการสรุปผลผิดพลาด (Publication bias)

อีกปัจจัยที่อาจส่งผลต่อการผลลัพธ์การวิจัยก็คือ ปัจจัยรบกวน (Confounding factor) คือปัจจัยที่อาจส่งผลต่อตัวแปรต้นและตัวแปรตามและจะส่งผลต่อผลลัพธ์การวิจัยให้มีความคลาดเคลื่อนบิดเบือนไปจากความเป็นจริง อาจจะส่งผลทำให้เกิดผลลัพธ์ที่มากขึ้น น้อยลง หรือตรงกันข้ามก็ได้ การควบคุมปัจจัยรบกวนและอคติ เริ่มตั้งแต่การเลือกจำนวนกลุ่มประชากรให้เหมาะสม ใช้ Randomization protocol เคร่งครัดต่อการใช้ protocol และหลังเก็บข้อมูลอาจจะควบคุมปัจจัยรบกวนในขั้นตอนของการวิเคราะห์ โดยการทำ Stratified analysis หรือ Multivariate analysis ตามหลัง[2,7,10]

สรุป

ในการเลือกใช้สถิติเพื่อการวิจัย มักจะพบปัญหาที่เกิดขึ้นเสมอๆ ดังนั้นผู้วิจัยจึงควรพิจารณาให้รอบคอบและมีความระมัดระวังให้มากขึ้น ข้อควรระวังในการเลือกใช้สถิติมีดังต่อไปนี้

1. หลีกเลี่ยงการใช้สถิติที่สูงเกินความจำเป็น ซึ่งเป็นอุปสรรคต่อผู้อ่านที่ไม่มีพื้นฐานทางสถิติ เช่น ถ้าต้องการเปรียบเทียบผลสัมฤทธิ์ 2 วิธี สถิติที่เหมาะสมคือ t - test ไม่จำเป็นต้องใช้ ANOVA

2. ควรเลือกใช้สถิติให้สอดคล้องกับปัญหาการวิจัย

3. ควรเลือกใช้สถิติโดยพิจารณาองค์ประกอบอื่น ๆ เช่น ตัวแปร ระดับของการวัด การสุ่มกลุ่มตัวอย่างหรือใช้ประชากรทั้งหมด ฯลฯ องค์ประกอบเหล่านี้จำเป็นต้องนำมาร่วมพิจารณาในการตัดสินใจเลือกใช้สถิติด้วยทุกครั้ง เพื่อให้ได้ผลการวิเคราะห์ที่ถูกต้องแม่นยำ และน่าเชื่อถืออันจะนำไปสู่ผลการวิจัยที่มีคุณภาพ[6]

เอกสารอ้างอิง

1.Marie A. Krousel-Wood, Richard B. Chambers, Paul Muntner. Clinicians' Guide to Statistics for Medical Practice and Research: Part I. Ochsner J. 2006 Winter; 6(2): 68–83.

2.เจษฎา คุณโน. วิธีการวิจัยพื้นฐาน สำหรับวิทยาศาสตร์สุขภาพ. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย; 2565.

3.ชูศักดิ์ โอกาศเจริญ. Evidence Based Medicine. เอกสารประกอบการประชุม รามาธิบดีกุมารเวชศาสตร์ปฏิทัศน์ครั้งที่17 (The 17th Ramathibodi Annual Review in Pediatrics 2023); 22-27 พฤษภาคม 2566; ภาควิชากุมารเวชศาสตร์ คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล. กรุงเทพฯ:คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล ; 2556.

4.ปาริชาติ โรจน์พลากร-กู๊ช, ยุวดี ฤาชา. สถิติสำหรับงานวิจัยทางการพยาบาลและการใช้โปรแกรม SPSS for Windows. พิมพ์ครั้งที่ 3. กรุงเทพฯ: บริษัท จุดทอง จำกัด; 2555.

5.ปัทมา สุพรรณกุล. สถิติประยุกต์ สำหรับงานวิจัยด้านสาธารณสุข. พิมพ์ครั้งที่ 3. พิษณุโลก: สำนักพิมพ์มหาวิทยาลัยนเรศวร; 2565.

6.ภัทรธิรา ผลงาม. การเลือกใช้สถิติในการวิเคราะห์ข้อมูล. [อินเทอร์เน็ต]. [เข้าถึงเมื่อ 28 สิงหาคม 2566]; เข้าถึงได้จาก: URL : https://www.tuct.ac.th/km/article/stat_1.pdf

7.ชำแก้ว หวานวารี. สถิติในเวชปฏิบัติ. เอกสารประกอบการสอนรายวิชาเวชศาสตร์ครอบครัวและชุมชน 3 (RAID511) หลักสูตรแพทยศาสตรบัณฑิต คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี. [อินเทอร์เน็ต].2551 [เข้าถึงเมื่อ 28 สิงหาคม 2566]; เข้าถึงได้จาก: URL : https://www.rama.mahidol.ac.th/fammed/sites/default/files/public/pdf/Stat.pdf

8.อติพร อิงค์สาธิต. เอกสารประกอบการสอน หลักการพิจารณานำงานวิจัยเกี่ยวกับการตรวจวินิจฉัยมาประยุกต์ในเวชปฏิบัติ. [อินเทอร์เน็ต]. [เข้าถึงเมื่อ 28 สิงหาคม 2566]; เข้าถึงได้จาก: URL : https://www.rama.mahidol.ac.th/fammed/sites/default/files/public/pdf/EBM_Diagnostic_study.pdf

9.กัลยา วานิชย์บัญชา. สถิติสำหรับงานวิจัย. พิมพ์ครั้งที่ 14. กรุงเทพมหานคร: หจก.สามลดา; 2566.

10.Marie A. Krousel-Wood, Richard B. Chambers, Paul Muntner. Clinicians' Guide to Statistics for Medical Practice and Research: Part II. Ochsner J. 2007 Spring; 7(1): 3–7.