

การเลือกใช้สติวิเคราะห์และปัญหาที่พบบ่อย

นพวรรณ เปี้ยชื่อ¹ PhD, RN, อพย. (การพยาบาลเวชปฏิบัติชุมชน)

บทคัดย่อ: การวิเคราะห์ข้อมูลมีผลต่อความตรงในการสรุปทางสถิติ การเลือกชนิดของสติที่ใช้ในการวิเคราะห์ข้อมูลจึงควรพิจารณาจากคำตามหรือวัดถูกประสงค์การวิจัย แบบวิจัย และระดับการวัดของตัวแปรสิ่งที่คุกคามต่อความตรงในการสรุปทางสถิติ ได้แก่ 1) พลังในการทดสอบต่า เนื่องจากขนาดตัวอย่างไม่เพียงพอ ความไม่เหมาะสมในการกำหนดระดับนัยสำคัญ การกำหนดสมมติฐานหรือการเลือกชนิดของสติ 2) การละเมิดข้อตกลงเบื้องต้น และ 3) ความคลาดเคลื่อนที่เกิดจากการวิเคราะห์ข้อมูลโดยเลือกชนิดของสติไม่เหมาะสม ปัญหาที่พบบ่อยในการวิเคราะห์ข้อมูลคือ ปัญหาจากขั้นตอนของการเก็บรวบรวมข้อมูล การจัดการข้อมูล วิธีการวิเคราะห์ข้อมูล การแปลผล การนำเสนอหรือการเผยแพร่ผลการวิจัย

คำสำคัญ: ปัญหาที่พบบ่อย สติวิเคราะห์ การวิจัย

¹รองศาสตราจารย์ โรงเรียนพยาบาลรามาธิบดี คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล

บทนำ

การวิเคราะห์ข้อมูล (data analysis) เป็นส่วนหนึ่งของการออกแบบการวิเคราะห์ (analysis design) ซึ่งเป็นองค์ประกอบของการออกแบบวิจัย (research design) และเป็นขั้นตอนสำคัญของกระบวนการวิจัย (research process) ที่ผู้วิจัยกำหนดไว้ล่วงหน้าในโครงร่างการวิจัย โดยคำนึงถึงค่าตามวิจัย แบบวิจัย ระดับการวัดของตัวแปรและข้อตกลงเบื้องต้นแต่อาจมีการปรับเปลี่ยนได้ หากไม่เป็นไปตามแผน เช่น มีการละเมิดข้อตกลงเบื้องต้น เป็นต้น การเลือกใช้สถิติในการวิเคราะห์ข้อมูลที่ไม่เหมาะสม มีผลให้เกิดความคลาดเคลื่อนของความตรงในการวิจัย (research validity) โดยเฉพาะอย่างยิ่ง ในส่วนของความตรงในการสรุปทางสถิติ (statistical conclusion validity)¹

การวิเคราะห์ข้อมูลเป็นการดำเนินการลดข้อมูลอย่างเป็นระบบ ในรูปแบบของการจัดกลุ่ม จัดลำดับ จัดกระทำ แปลผลและสรุปผลเพื่อให้ได้ค่าตอบของค่าตามวิจัยวัตถุประสงค์หรือสมมติฐาน การวิจัยโดยลักษณะของค่าตามวิจัย อาจเป็นการศึกษาสถานการณ์ ชนิด จำนวนลักษณะ การเปรียบเทียบความแตกต่าง การศึกษาความสัมพันธ์ การศึกษาอิทธิพล การพยากรณ์หรือการทำนายโดยข้อมูลที่ทำการวิเคราะห์ได้จากการวัดตัวแปรระดับการวัดของตัวแปรแบ่งเป็น 4 ระดับคือ 1) นามสกุล (nominal scale) 2) อันดับสเกล (ordinal scale) 3) อันตรภาคสเกล (interval scale) และ 4) เรซิสเกล

(ratio scale) ในกรณีวิเคราะห์ข้อมูลที่ศึกษาจากประชากร ค่าที่วิเคราะห์ได้เรียกว่าค่าพารามิเตอร์ สำหรับข้อมูลที่ศึกษาจากตัวอย่าง ค่าที่วิเคราะห์ได้เรียกว่าค่าสถิติ

การวิเคราะห์ข้อมูลเชิงปริมาณ ประกอบด้วยสถิติบรรยาย (descriptive statistics) และสถิติอ้างอิง (inferential statistics) สถิติบรรยายเป็นสถิติที่ใช้ในการสรุปข้อมูลและบรรยายลักษณะที่ศึกษาได้แก่การนับการวัด (การแจกแจงความถี่ ร้อยละ สัดส่วน) การวัดแนวโน้มเข้าสู่ส่วนกลาง (ค่าเฉลี่ย มัธยฐาน ฐานนิยม) และการวัดการกระจาย (พิสัย ค่าเบี่ยงเบนมาตรฐาน ความแปรปรวน สัมประสิทธิ์ความผันแปร) โดยสามารถศึกษาข้อมูลจากประชากรหรือข้อมูลจากตัวอย่าง การวิเคราะห์ข้อมูลไม่ต้องทดสอบสมมติฐานทางสถิติ เมื่อศึกษาข้อมูลจากตัวอย่างจึงอ้างอิงไปยังประชากรไม่ได้ สำหรับการวิเคราะห์ข้อมูลด้วยสถิติ อ้างอิง ต้องมีการทดสอบสมมติฐานโดยแบ่งเป็นการทดสอบชนิดพารามิตริก (parametric test) ซึ่งคำนึงถึงลักษณะการแจกแจงของประชากรที่ตัวอย่างได้รับการสุ่มมาต้องมีการแจกแจงแบบปกติ (normal distribution) และการทดสอบชนิดนันพารามิตริก (non parametric test) ซึ่งไม่คำนึงถึงลักษณะการแจกแจงของประชากร (distribution free)

การเลือกชนิดของสถิติที่ใช้ในการวิเคราะห์ข้อมูลพิจารณาจากค่าตามหรือวัตถุประสงค์การวิจัย แบบวิจัย และระดับการวัดของตัวแปร ดังตารางที่ 1

ตารางที่ 1 แนวทางการเลือกชนิดของการวิเคราะห์ข้อมูล

วัตถุประสงค์	จำนวนกลุ่ม/ตัวแปร	ระดับการวัด	สถิติที่ใช้
ศึกษาลักษณะกลุ่ม	1 ตัวแปร	-	Nominal Frequency (n), percentage, mode
			Ordinal Median, percentage, quartile
			Interval/ratio Mean
ทดสอบการกระจาย	1 ตัวแปร	-	Nominal Range
			Ordinal Range
			Interval/ratio Range, SD, S^2 , Coefficient of variation (CV)
วิเคราะห์	2 ตัวแปร	-	Nominal X^2
ความสัมพันธ์			Ordinal Spearman's rank correlation
			Interval/ratio Pearson's Product moment correlation
เปรียบเทียบ	2 กลุ่ม	อิสระ	Nominal X^2
		ไม่อิสระ	
		อิสระ	Ordinal Rank sum tes
		ไม่อิสระ	
		อิสระ	Interval/ratio Independent t-test
		ไม่อิสระ	
	≥ 3 กลุ่ม	อิสระ	Interval/ratio ANOVA
		ไม่อิสระ	

สิ่งคุกคามต่อความตรงในการสรุปทางสถิติ

สิ่งที่คุกคามหรือบกวนความตรงในการสรุปทางสถิติมีดังนี้

1. พลังในการทดสอบต่ำ (low statistical power)¹ นั่นคือผลการวิเคราะห์ข้อมูล ไม่พบความแตกต่างหรือความสัมพันธ์ทั้งที่มีความแตกต่างกัน หรือความสัมพันธ์อยู่จริง ทำให้ผลการวิเคราะห์ไม่ตรงกับความจริง (low precision) อาจเนื่องมาจากการสั่นสะเทือนปัจจัยดังนี้

1) ขนาดตัวอย่างไม่เพียงพอ ทำให้เกิดความคลาดเคลื่อนจากการสุ่ม (random error) และความคลาดเคลื่อนมาตรฐานของความแปรปรวน (standard error of variance) เพิ่มขึ้น โอกาสปฏิเสธ

สมมติฐานนั้นลังจึงยกขึ้นแนวทางการแก้ปัญหาคือ กำหนดขนาดตัวอย่างที่เพียงพอ โดยอ้างอิงจากขนาด อิทธิพล (effect size) ของศึกษาที่ผ่านมาตามหลัก การวิเคราะห์อำนาจการทดสอบ (power analysis)² หลักการกำหนดตัวอย่างเมื่อทราบประชากร³ หรือ หลักการทางระบาดวิทยา⁴

2) การกำหนดระดับนัยสำคัญ (type I error: α) ที่น้อยมาก ถึงแม้ว่าจะทำให้การวิเคราะห์ข้อมูลมีความเชื่อมั่นสูง ($1 - \alpha$) แต่มีผลให้เกิดความคลาดเคลื่อนชนิดที่ 2 (type II error: β) สูง และ พลัง (power: $1 - \beta$) ใน การทดสอบต่ำ การวิจัยทางการพยาบาลโดยทั่วไปจึงกำหนดระดับนัยสำคัญที่ .05

3) การกำหนดสมมติฐานแบบไม่มีทิศทาง (non directional hypothesis) ไม่โอกาสปฏิเสธ

สมมติฐานนั้ลได้ยากกว่าสมมติฐานแบบมีทิศทาง (directional hypothesis) อย่างไรก็ตาม การพิจารณา กำหนดสมมติฐานแบบมีทิศทางหรือไม่มีทิศทางขึ้น อยู่กับเหตุผลเชิงทฤษฎีหรือหลักฐานเชิงประจักษ์ เช่น การศึกษาผลของการนวดต่อความวิตกกังวลของ ผู้ป่วย ควรตั้งสมมติฐานแบบมีทิศทาง เนื่องจาก สามารถอธิบายได้ตามหลักการเชิงทฤษฎีว่า การนวด มีกลไกที่ส่งผลให้ความวิตกกังวลของผู้ป่วยลดลง

4) การเลือกชนิดของสถิติที่ไม่เหมาะสมใน การวิเคราะห์ข้อมูลทำให้เกิดความคลาดเคลื่อนใน การสรุปผลซึ่งจะส่งผลต่อความถูกต้องในการตอบ คำถามการวิจัย ผู้วิจัยจึงต้องศึกษาและเข้าใจอย่าง ถ่องแท้เกี่ยวกับหลักการและแนวทางในการวิเคราะห์ ข้อมูลรวมทั้งพิจารณาความสอดคล้องกับคำถามวิจัย แบบวิจัยและระดับการวัดของตัวแปร

2. การละเมิดข้อตกลงเบื้องต้น (violation of assumption)^{1,5} สถิติวิเคราะห์แต่ละชนิดมีข้อตกลง เบื้องต้นที่แตกต่างกัน ปัญหาที่พบได้บ่อยคือ การ ละเมิดข้อตกลงเบื้องต้นเรื่องการแจกแจงของข้อมูล เป็นแบบปกติ ซึ่งส่วนใหญ่ผู้วิจัยไม่ได้รายงานผลการ ทดสอบข้อตกลงเบื้องต้นไว้ในรายงานวิจัย อย่างไร ก็ตามสามารถสังเกตได้จากการนำเสนอข้อมูลค่า เฉลี่ยและค่าเบี่ยงเบนมาตรฐาน ซึ่งหากพบว่า ค่าเบี่ยง เบนมาตรฐานใกล้เคียงกับครึ่งหนึ่งหรือมากกว่าครึ่ง หนึ่งของค่าเฉลี่ย บ่งชี้ว่าสัมประสิทธิ์ความผันแปร (coefficient of variation: CV) มีค่ามาก แสดงว่า ข้อมูลน่าจะมีความเบี้ยว (skewness) หรือความโด่ง (kurtosis) ไม่เหมาะสมกับการใช้สถิติชนิดพารา เมตريค หากผู้วิจัยยังคงวิเคราะห์ด้วยสถิติชนิดพารา เมตريค จะทำให้เกิดความคลาดเคลื่อนในการสรุป ดังนั้นผู้วิจัยควรนำเสนอผลการทดสอบข้อตกลง เบื้องต้นโดยสรุป ก่อนนำเสนอผลการวิเคราะห์เพื่อ ตอบวัตถุประสงค์หรือทดสอบสมมติฐาน

3. Fishing the error rate problem ทำให้เกิด ความคลาดเคลื่อนชนิดที่ 1 (type I error)¹ เมื่อผู้วิจัย ทำการเปรียบเทียบค่าเฉลี่ยหลาย ๆ ครั้ง หรือวิเคราะห์ ความสัมพันธ์จากข้อมูลที่มีตัวแปรจำนวนมาก ตัวอย่างเช่น การวิจัยมีวัตถุประสงค์เพื่อเปรียบเทียบ ค่าเฉลี่ยของตัวแปรในตัวอย่าง 3 กลุ่ม ซึ่งควร วิเคราะห์ด้วยสถิติ ANOVA แต่หากผู้วิจัยทำการ วิเคราะห์ข้อมูลด้วยสถิติ Independent t-test สามครั้ง เพื่อเปรียบเทียบค่าเฉลี่ยของตัวแปรระหว่างตัวอย่าง กลุ่มที่ 1 และ 2, กลุ่มที่ 1 และ 3, กลุ่มที่ 2 และ 3 จะ ทำให้เกิดความคลาดเคลื่อนมากขึ้น จึงเรียกว่า Fishing the error rate

ปัญหาที่พบบ่อยในการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเป็นการจัดการข้อมูลอย่าง เป็นระบบภายหลังเสร็จสิ้นการเก็บรวบรวมข้อมูล เพื่อจัดเตรียมให้พร้อมเข้าสู่การวิเคราะห์ การแปลผล และการนำเสนอหรือเผยแพร่ผลงานวิจัย ปัญหาที่พบ บ่อยในการวิเคราะห์ข้อมูลมีดังนี้

1) การเก็บรวบรวมข้อมูล (data collection) ปัญหาที่พบบ่อยในขั้นตอนการเก็บรวบรวมข้อมูลคือ อัตราการไม่ตอบกลับ (non response rate) สูง ในการนี้ ที่เป็นการศึกษาแบบตัดขวาง (cross-sectional study) หรือมีการสูญหาย (drop out) ของตัวอย่าง ในกรณีที่มีการติดตามเก็บรวบรวมข้อมูลมากกว่าหนึ่ง ครั้งในการศึกษาระยะยาว (longitudinal study) หรือ การศึกษาเชิงทดลอง นอกจานนี้ยังพบว่า การเก็บ รวบรวมข้อมูลไม่สอดคล้องกับประชากรที่ศึกษาหรือ ไม่เป็นไปตามเกณฑ์คัดเข้า ตัวอย่าง เช่น ประชากรที่ ศึกษาเป็นผู้สูงอายุ แต่ผู้วิจัยเก็บรวบรวมข้อมูลใน ผู้ใหญ่ด้วย เป็นต้น

2) การจัดการข้อมูล (data management) ภายหลังจากการเก็บรวบรวมข้อมูลสิ้นสุดลง โดยปกติ ผู้วิจัยจะเริ่มจัดการข้อมูลดิบที่เก็บรวบรวมมาทำการ ประมวลข้อมูล (data processing) ขั้นตอนที่สำคัญใน การประมวลข้อมูลคือ บรรณาธิกร (editing) เป็นการ ตรวจสอบความครบถ้วนสมบูรณ์ ความถูกต้อง และ เป็นไปตามข้อตกลง รวมทั้งการกำหนดรหัสและการ ลงทะเบียน โดยมีประเด็นปัญหาที่พบบ่อยคือ ความ สับสนในการลงทะเบียนและ การกลับค่าแบบ ในกรณีที่ เป็นค่าตามเชิงลบที่ผ่านมาพบว่ามีผู้วิจัยที่ทำการกลับ ค่าแบบในข้อมูลดิบจากแบบสอบถามที่ได้ทำการ เก็บรวบรวมข้อมูล และจึงลงข้อมูลในฐานข้อมูลใน คอมพิวเตอร์ ซึ่งบ่อยครั้งที่พบความผิดพลาดจาก การกลับค่าแบบตามวิธีการดังกล่าว แนวทางในการ แก้ปัญหาคือ ลงข้อมูลในฐานข้อมูลตามรหัสที่ผู้วิจัย กำหนดแล้วจึงใช้คำสั่ง recode ในโปรแกรมวิเคราะห์ ข้อมูลเพื่อทำการกลับค่าแบบ

นอกจากนี้ แนวทางหนึ่งในการตรวจสอบความ ถูกต้องของการลงข้อมูลก่อนทำการวิเคราะห์สามารถ ทำได้โดยการใช้โปรแกรมวิเคราะห์ข้อมูลสำเร็จรูป วิเคราะห์ค่าความถี่ ค่าต่ำสุดและค่าสูงสุดจะช่วยให้ ผู้วิจัยทราบว่าการลงข้อมูลดังกล่าวถูกต้องตามข้อมูล จริงหรือเป็นไปตามรหัสที่ผู้วิจัยกำหนดหรือไม่

3) การวิเคราะห์ข้อมูล (data analysis) การ วิเคราะห์ข้อมูลต้องให้ครบตามวัตถุประสงค์หรือ สมมติฐานที่กำหนดไว้ และในการเลือกใช้สถิติ วิเคราะห์ต้องคำนึงถึง ค่าตามวิจัย ระดับการวัดของ ตัวแปร และแบบวิจัยดังตัวอย่างต่อไปนี้

การวิเคราะห์ข้อมูลโดยใช้สถิติบรรยายในเรื่อง การนับการวัด เช่น การจัดกลุ่มข้อมูลรายได้ของ ตัวอย่าง หากผู้วิจัยแบ่งรายได้เป็น 2 กลุ่มคือ มากกว่า 2 หมื่นบาทต่อเดือน และ น้อยกว่า 2 หมื่นบาทต่อเดือน ผู้ที่มีรายได้เท่ากับ 2 หมื่นบาทจะไม่อยู่ในกลุ่มใด

ปัญหาที่พบอีกประเด็นหนึ่งคือ ค่าร้อยละ 1) หากจำนวนตัวอย่างน้อยมาก อาจไม่จำเป็นต้อง วิเคราะห์ค่าร้อยละ 2) การนำเสนอค่าร้อยละ ผลรวม ต้องมีค่าไม่เกิน 100.0 และไม่น้อยกว่า 100.0 3) สำหรับข้อคิดเห็นที่เปิดโอกาสให้ตอบได้มากกว่าหนึ่ง คำตอบ ควรใช้จำนวนผู้ตอบทั้งหมดเป็นค่าเทียบ 100 4) ในการเปรียบเทียบค่าร้อยละของข้อมูลที่มี การจำแนกสองทาง จำนวนรวมที่ใช้เทียบ 100 ควร เป็นจำนวนรวมในแนวตัวแปรอิสระ

การวิเคราะห์แนวโน้มเข้าสู่ส่วนกลางหรือ ค่าตัวกลาง เป็นการบอกลักษณะเบื้องต้นของข้อมูล การคำนวณค่าเฉลี่ยใช้ค่าสัมเกตทุกค่า จึงเป็นตัวกลาง ที่มีความเสถียรมากที่สุด อย่างไรก็ตามในการนำเสนอ ผลในเชิงเปรียบเทียบค่าเฉลี่ยต้องอยู่บนฐานค่าแบบ เดิมเดียวกัน อีกประการหนึ่งคือ ลักษณะธรรมชาติ ของข้อมูลบางชนิด ไม่มีการแจกแจงเป็นแบบปกติ เช่น รายได้ และหากการทดสอบยืนยันว่าข้อมูลไม่มี การแจกแจงแบบปกติ การวิเคราะห์ค่าตัวกลางที่ เหมาะสมคือ มัธยฐาน

การวิเคราะห์ข้อมูลโดยใช้สถิติอ้างอิงที่เป็นการ ทดสอบแบบพารามิตริก ต้องมีการตรวจสอบ ข้อตกลงเบื้องต้น ในกรณีการทดสอบการแจกแจง ของข้อมูลว่า เป็นแบบปกติหรือไม่ หากจำนวน ตัวอย่าง 50 ขึ้นไป สามารถใช้วิธี Kolmogorov Smirnov test แต่หากจำนวนตัวอย่างน้อยกว่า 50 ควรใช้วิธี Shapiro Wilk test ในกรณีที่พบว่าข้อมูลไม่มี การแจกแจงเป็นแบบปกติ ในทางปฏิบัติหากพบว่ามี ค่า outlier ที่สามารถพิจารณาตัดออก แล้วทำการ ทดสอบใหม่พบว่า ข้อมูลมีการแจกแจงเป็นแบบปกติ และอาจพิจารณาร่วมกับ histogram จะเป็นข้อมูล ที่ช่วยให้ผู้วิจัยสามารถตัดสินใจเลือกใช้สถิติแบบ พารามิตริกหรือนันพารามิตริกได้อย่างเหมาะสม

ในบางครั้งผู้วิจัยต้องปรับเปลี่ยนแผนการวิเคราะห์ข้อมูล ซึ่งส่วนใหญ่เนื่องมาจากการละเมิดข้อตกลงเบื้องต้นของการวิเคราะห์สถิตินั้น ๆ หรือพบว่าข้อมูลในระยะก่อนทดลองของกลุ่มควบคุมและกลุ่มทดลองแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ผู้วิจัยต้องใช้การวิเคราะห์ความแปรปรวนร่วม (ANCOVA) แทนการทดสอบที (t-test) อย่างไรก็ตาม ผู้วิจัยต้องทบทวนและปรับวัตถุประสงค์ของการศึกษาให้สอดคล้องกับการวิเคราะห์ข้อมูล

การวิเคราะห์เชิงสาเหตุ (causal analysis) เป็นอีกข้อจำกัดที่พบบ่อยในการศึกษาเชิงสำรวจ หรือการศึกษาเชิงระบาดวิทยาแบบ Retrospective Case Control ซึ่งผู้วิจัยควรระบุในรายงานวิจัยว่า เป็นข้อจำกัดของแบบวิจัยที่หากพิจารณาตามหลักการแล้ว จะไม่สามารถสรุปความสัมพันธ์เชิงเหตุผลได้เนื่องจากเป็นการเก็บรวบรวมข้อมูลเพียงครั้งเดียว หรือเป็นการศึกษาจากข้อมูลย้อนหลัง

การทดสอบแบบนันพารามեตริกที่ใช้บ่อยในการวิจัยทางการพยาบาลคือ การทดสอบไคสแควร์ (Chi-square test) หากมีการละเมิดข้อตกลงเบื้องต้น ผู้วิจัยต้องทำการยุบรวมกลุ่มย่อยที่มีความหมายไปในทางเดียวกันเข้าด้วยกันหรืออ้างอิงตามทฤษฎีไม่ใช่ยุบรวมกลุ่มย่อยที่มีค่าความถี่คาดหวัง (expected frequency) น้อยเข้าด้วยกัน สำหรับการวิเคราะห์ Fisher's exact probability test เป็นการทดสอบโดยการเปรียบเทียบค่าความน่าจะเป็น ไม่มีการคำนวณค่าไคสแควร์ การนำเสนอผลการวิเคราะห์จึงไม่มีค่าสถิติไคสแควร์ปรากฏอยู่

4) การแปลผลการวิเคราะห์ข้อมูล (interpretation of data) ในการแปลผลการวิเคราะห์ข้อมูล ผู้วิจัยควรพิจารณาทั้งความมีนัยสำคัญทางสถิติ (statistical significance) และความมีนัยสำคัญทาง

คลินิก (clinical significance) กล่าวคือ ความมีนัยสำคัญทางสถิติที่พบจากผลการวิเคราะห์ข้อมูล อาจไม่ได้สะท้อนถึงความมีนัยสำคัญทางคลินิก เช่น ผลการวิเคราะห์ข้อมูลพบว่า กลุ่มทดลองมีความดันโลหิตภายนอกต่ำกว่า กลุ่มควบคุม แต่หากการเปลี่ยนแปลงระดับความดันโลหิตที่ลดลงอย่างมีนัยสำคัญทางสถิตินั้นยังอยู่ในเกณฑ์สูงอาจไม่แสดงถึงความมีนัยสำคัญทางคลินิก นอกจากนี้ผู้วิจัยต้องแปลผลการวิเคราะห์ข้อมูลให้เหมาะสมตามหลักการเชิงสถิติของสถิติแต่ละชนิด และสอดคล้องกับคำมั่นวิจัย วัตถุประสงค์หรือสมมติฐานของการวิจัย

5) การนำเสนอหรือการเผยแพร่ผลการวิจัย (presentation and dissemination of results)

การนำเสนอค่าของโอกาสในการตัดสินใจสรุปผลอ้างอิง (probability value: p-value) ในปัจจุบันนิยมรายงานผล p-value จากค่าจริง (มักใช้ทศนิยม 3 ตำแหน่ง) ที่เป็นผลจากการวิเคราะห์ข้อมูล เช่น $p = .034$ หากกว่าการรายงานค่าที่แสดงระดับคือ $p < .05$ สำหรับค่า p ที่เท่ากับ $.000$ ให้รายงานผลเป็น $p < .001$ เนื่องจากพื้นที่ในการปฏิเสธสมมติฐานนั้นน้อยมาก

ความคลาดเคลื่อนที่พบบ่อยในการนำเสนอผลการวิเคราะห์ข้อมูล มีทั้งความคลาดเคลื่อนการละกัดคำ การใช้สัญลักษณ์ทางสถิติที่ไม่สอดคล้องกัน การนำเสนอข้อมูลตัวเลขและค่าสถิติที่ไม่สอดคล้องกัน ขาดข้อมูลสถิติที่จำเป็นในตารางนำเสนอผลการวิจัย หากมีการใช้อักษรย่อหรือรหัส ต้องระบุความหมายหรือคำเติมท้ายตาราง หมายเหตุท้ายตารางโดยทั่วไปเป็นการเสนอความหมาย การกำหนดรหัส ตัวแปรทุน หรือระดับนัยสำคัญ ตัวอย่างความคลาดเคลื่อนในการเผยแพร่ผลการวิจัย เช่น ผู้วิจัยต้องการ

นำเสนอ “ค่าสถิติ t” แต่ระบุว่า “t-test” ซึ่งมีความหมายแตกต่างกัน ดังนั้นจึงจำเป็นต้องมีความละเอียด

รอบคอบในทุกขั้นตอนของการวิเคราะห์ข้อมูล โดยสรุปผู้วิจัยควรเลือกชนิดของสถิติที่เหมาะสมในการวิเคราะห์ข้อมูล โดยพิจารณาจากค่าตาม หรือตัตุประสงค์การวิจัย แบบวิจัย และระดับการวัด ของตัวแปร และคำนึงถึงลิ่งคุกคามต่อความตรงใน การสรุปทางสถิติ ที่สำคัญคือ พลังในการทดสอบต่าง และการละเมิดข้อตกลงเบื้องต้น รวมทั้งจัดการปัญหา ที่พบบ่อยในขั้นตอนของการเก็บรวบรวมข้อมูล การจัดการข้อมูล วิธีการวิเคราะห์ข้อมูล การแปลผล การนำเสนอหรือการเผยแพร่ผลการวิจัย

เอกสารอ้างอิง

1. Burns N, Grove SK. *The Practice of nursing research: Appraisal, synthesis, and generation of evidence*. 7th ed. St. Louis: Saunders; 2013.
2. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. New Jersey: Lawrence Erlbaum Associates, Publishers; 1988.
3. Cochran WG. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons; 1977.
4. Kelsey JL. *Methods in Observational Epidemiology and Biostatistics*. New York: Oxford University Press; 1986.
5. Plichta SB, Kelvin EA, Munro BH. *Statistical methods for health care research*. 6th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2013.

Statistical Analysis and Common Problems

Noppawan Piaseu¹ PhD, RN, Dip. ACNP

Abstract: Data analysis results in statistical conclusion validity. Types of statistical analysis should be selected as appropriate for research questions or research objectives, research design, and levels of measurement. Threats to statistical conclusion validity include: 1) low statistical power due to inadequate sample size, inappropriate level of significance, hypothesis or type of statistical analysis, 2) violation of assumption, and 3) fishing the error rate problem. Common problems in data analysis are in the process of data collection, data management, data analysis method, interpretation, presentation or dissemination of research.

Keywords: common problems, statistical analysis, research

¹Associate Professor, Ramathibodi School of Nursing, Faculty of Medicine, Ramathibodi Hospital, Mahidol University