

แนะนำสถิติในการวิจัยทางคลินิก

ภาณุวัฒน์ เสาศิทธิชัย

บทนำ

วัตถุประสงค์ของการวิจัยทางการแพทย์ ซึ่งมีการนำวิชาสถิติเข้ามาช่วย ก็เพื่ออนุมานข้อสรุปที่ได้จากกลุ่มตัวอย่าง ไปยังประชากรและผู้ป่วยคนอื่นๆ ทั้งในปัจจุบันและอนาคต โดยอาศัยแบบจำลองทางสถิติเป็นตัวเชื่อม สำหรับ เนื้อหาของวิชาสถิติ จะเกี่ยวข้องกับความแปรปรวนของข้อมูล โดยมีคำถามสำคัญคือ ความแปรปรวนที่พบเห็นมีความสัมพันธ์กับปัจจัยบางอย่างที่วัดได้และทำนายได้หรือเกิดจากความบังเอิญเป็นหลัก¹ ถ้าข้อมูลมีความแปรปรวนสูง กลุ่มตัวอย่างต้องใหญ่มาก จึงจะอนุมานข้อสรุปไปยังประชากรได้อย่างมั่นใจ

งานวิจัยทางคลินิกมักมีเนื้อหาเกี่ยวกับสาเหตุหรือปัจจัยการเกิดโรค การวินิจฉัยโรค วิธีการรักษา การพยากรณ์ผลการรักษาหรือการพยากรณ์โรค และวิธีป้องกันโรค เป็นต้น โดยคำถามวิจัยมักเป็นคำถามเชิงเปรียบเทียบ (หรือแปลงให้เป็นคำถามเชิงเปรียบเทียบได้)² ระหว่างวิธีการ 2 วิธีหรือมากกว่า หรือเปรียบเทียบปัจจัย 2 ตัวหรือ 2 ระดับหรือมากกว่า หรืออีกนัยหนึ่งเป็นการเปรียบเทียบระหว่างกลุ่มผู้ป่วย หรือข้อมูลทางคลินิก ตั้งแต่ 2 กลุ่มหรือ 2 ชุดขึ้นไป ยกตัวอย่างเช่น การศึกษาประสิทธิภาพของยา 3 ตัว ต้องเปรียบเทียบกลุ่มผู้ป่วย 3 กลุ่มที่ได้รับยาต่างกัน ในแง่ของผลการรักษาบางอย่าง (ว่ากลุ่มใดมีผลการรักษาดีกว่ากัน) หรือการศึกษาความแม่นยำของอุปกรณ์วินิจฉัยโรคในช่องท้อง ต้องเปรียบเทียบผลการวินิจฉัยด้วยอุปกรณ์ดังกล่าวกับผลการวินิจฉัยด้วยวิธีมาตรฐานที่เป็น “reference standard” ว่าตรงกันหรือต่างกันอย่างไร หรือ การศึกษาปัจจัยเสี่ยงต่อการเกิดโรค ที่ปัจจัยแต่ละตัวมีหลายระดับค่า ต้องเปรียบเทียบว่าบางระดับ ของปัจจัยหนึ่งๆ จะเพิ่ม (หรือลด) ความเสี่ยงต่อการเกิดโรค เมื่อเทียบกับระดับพื้นฐาน (baseline level) หรือไม่ เป็นต้น

บทความนี้มีวัตถุประสงค์ให้ผู้อ่าน ซึ่งเป็นคัลยแพทย์หรือแพทย์ประจำบ้านที่มีความสนใจเกี่ยวกับงานวิจัยและสถิติวิเคราะห์ แต่ยังขาดความคุ้นเคยกับแนวความคิดสำคัญของวิชาสถิติ ได้รับคำแนะนำเบื้องต้นเกี่ยวกับแนวความคิดดังกล่าว เพื่อเป็นพื้นฐานในการศึกษาวิชาสถิติในการวิจัยต่อไป หรือเพื่อการอ่านบทความทางวิชาการให้เกิดความเข้าใจมากยิ่งขึ้น ดังนั้นในบทความนี้จะไม่มีการวิเคราะห์ข้อมูลในรายละเอียดให้ผู้อ่านรู้สึกหนักใจแต่อย่างใด

นิยามศัพท์ และแนะนำแนวความคิดพื้นฐาน

ประชากร (population) คือกลุ่มคน สิ่งของ หรืออื่นๆ (เรียกโดยรวมว่าสมาชิก) ที่มีโอกาสจะได้รับเลือกเข้าร่วมการศึกษาหนึ่งๆ โดยประชากรอาจมีอยู่แล้ว ณ เวลาใดก็ตาม เช่นประชากรในประเทศหนึ่งๆ (สมาชิกคือคน) หรือประชากรอาจเกิดจากการกระทำหรือการทดลอง เช่นการโยนเหรียญเสี่ยงโชค (สมาชิกคือการโยนเหรียญในแต่ละครั้ง)ซึ่งจะมีจำนวนประชากรตามจำนวนครั้งที่โยนเหรียญ ประชากรอาจมีจำนวนจำกัด เช่นประชากรในประเทศหนึ่งๆ หรือ อาจมีจำนวนเป็นอนันต์ เช่นประชากรของการสุ่มเลือกจุดบนแผ่นกระดาษ ประชากรที่เกิดจากการทดลองก็อาจมีจำนวนจำกัด เช่นการกำหนดให้โยนเหรียญจำนวน 100,000 ครั้ง หรืออาจมีจำนวนเป็นอนันต์ เช่น การโยนเหรียญอย่างไม่สิ้นสุด ซึ่งในกรณีหลังนี้ ย่อมต้องใช้เวลาดทดลองเป็นอนันต์ จึงจะได้จำนวนประชากรเป็นอนันต์ ซึ่งเป็นไปไม่ได้ในทางปฏิบัติ เป็นต้น

กลุ่มตัวอย่าง (sample) คือกลุ่มที่เกิดจากการเลือกตัวแทนจากประชากร ซึ่งอาจเป็นคนจำนวนหนึ่ง หรือการโยนเหรียญจำนวนหนึ่ง ตัวแทนที่ดีของประชากร อันมีคุณสมบัติใกล้เคียงกับประชากรมากที่สุด ต้องได้จากการสุ่มเลือก (random sampling)

ความแปรปรวน (variability)³ คือความแตกต่างหรือความหลากหลายของค่าวัดที่แปรไปตามสมาชิกและปัจจัยต่างๆ อาจมีหรือไม่มีสาเหตุชัดเจน อาจจะทำนายได้หรือทำนายไม่ได้ก็ตาม เช่น “อายุ” เป็นค่าวัด หรือตัวแปร ที่มีค่าหลากหลายและแตกต่างกันบุคคล สามารถทำนายค่าในอนาคตได้หากรู้ค่าเริ่มต้น หรือ “อาชีพ” ก็เป็นตัวแปรที่แปรตามบุคคล แต่ไม่สามารถทำนายล่วงหน้าได้อย่างแม่นยำ หรือ “ความสูง” ก็แปรตาม

บุคคล แต่มีปัจจัยทางอายุ เพศ เชื้อชาติ สภาวะโภชนาการ เป็นต้น ที่อาจช่วยทำนาย ความสูงของแต่ละบุคคลได้ระดับหนึ่ง

ความแปรปรวนร่วม (covariability) ของตัวแปรตั้งแต่สองตัวขึ้นไปหรือระหว่าง ปัจจัยสองปัจจัย หรือระหว่างปัจจัยกับผลลัพธ์ หรือระหว่างตัวแปรอิสระกับตัวแปรตาม หมายถึงการที่ตัวแปรหลายตัวมีค่าแตกต่างกันระหว่างบุคคล (หรือหน่วยวิเคราะห์ใดๆ ก็ตาม) และตัวแปรต่างๆ เหล่านี้ มีการเปลี่ยนแปลงร่วมกันอย่างมีแบบแผนและทำนาย ได้ในระดับหนึ่ง เช่น ความสูงอาจแปรตามอายุของบุคคล ซึ่งเปลี่ยนแปลงอย่างมีแบบแผน กล่าวคือความสูงจะเพิ่มขึ้นตามอายุที่เพิ่มขึ้นใน 20 ปีแรกของชีวิต ต่อมาจะคงที่ และจะ ค่อยๆ ลดลงตามอายุเมื่อบุคคลมีอายุเกิน 60 ปี เป็นต้น ความแปรปรวนร่วมกันที่มี แบบแผนชัดเจนเช่นนี้ เรียกว่า **ความสัมพันธ์ระหว่างตัวแปร (association between variables)** โปรดสังเกตว่าความแปรปรวนร่วม หรือความสัมพันธ์ระหว่างตัวแปร มิได้บ่ง บอกว่าความสัมพันธ์ดังกล่าวจะต้องเป็นเหตุและผลซึ่งกันและกันแต่อย่างใด เช่น อายุที่ เพิ่มขึ้นอาจไม่ใช่สาเหตุของความสูงที่เพิ่มขึ้น เป็นต้น

ตัวอย่างความสัมพันธ์ในกรณีอื่นๆ ได้แก่ ความรุนแรงของความเจ็บปวดของ แต่ละบุคคลที่แปรตามวิธีผ่าตัด 2 วิธี ค่าความสูงที่แปรตามไม้บรรทัด วิถีวัด และคนที่วัด (อย่างหลังนี้ถือเป็น measurement bias และ errors)

ความสัมพันธ์อาจมีจริง ("true" association) กล่าวคือมีความสัมพันธ์ในระดับ ประชากรจริง หรือความสัมพันธ์นั้นอาจพบเพียงแค่นี้ในระดับตัวอย่างจากการศึกษา อัน เป็นตัวแทนของประชากร โดยที่ความสัมพันธ์ที่พบในระดับตัวอย่าง อาจไม่เป็นจริงใน ระดับประชากรก็ได้ ("chance" association)

ดังนั้น ความแปรปรวนจึงแบ่งเป็น 2 ประเภทหลัก กล่าวคือ ประเภทแรก ความ แปรปรวนที่มีแบบแผน สัมพันธ์กับปัจจัยหรือตัวแปรอื่นๆ จริง ในระดับประชากร อันจะ ทำนายได้หากทราบปัจจัยที่เกี่ยวข้องดังกล่าว และประเภทที่สอง คือความแปรปรวนที่ไม่ ทราบปัจจัยที่เกี่ยวข้อง และทำนายไม่ได้แน่ชัด ในวิชาสถิติจึงจำลองให้ความแปรปรวน อย่างหลัง มีลักษณะเฉพาะที่เรียกว่า **ความแปรปรวนโดยบังเอิญ หรือความแปรปรวน แบบ สุ่ม (chance or random variation)**

คำว่าสุ่มในกรณีดังต่อไปนี้มี 2 ความหมาย กล่าวคือ ในความหมายแรกหมายถึง

การเลือกสมาชิกจากประชากรโดยที่สมาชิกทุกรายมีโอกาสได้รับเลือก เรียกว่าการเลือกอย่างสุ่ม อีกความหมายหนึ่ง หมายถึงการเกิดเหตุการณ์บางอย่าง เช่นจากการทดลอง ที่ไม่สามารถทำนายได้แน่นอน หรือไม่มีแบบแผน และมักไม่มีสาเหตุใดๆ หรือไม่มีสาเหตุที่ทราบในขณะนั้น ที่จะทำให้เกิดเหตุการณ์ดังกล่าว (อันที่จริง ในระดับนามธรรม ทั้งสองความหมายของคำว่าสุ่ม คือความหมายเดียวกัน) อย่างไรก็ตาม จะต้องมีการกำหนดเพิ่มเติมบางประการ (ทางคณิตศาสตร์) เพื่อให้คำนวณตัวชี้วัดของการเกิดเหตุการณ์ต่างๆ ได้ (ตัวชี้วัดอันได้แก่ ความน่าจะเป็น (probability) หรือโอกาส (chance) ที่จะเกิดเหตุการณ์นั้นๆ) หากคำนวณไม่ได้ ก็จะไม่มีความวิเศษ หรือการคำนวณค่า p ให้เรียน

ตัวอย่างของความแปรปรวนสังเกตได้ทั่วไปในผลการทดลองในห้องปฏิบัติการ ผลการทดลองทางคลินิก หรือการศึกษาทางคลินิกต่างๆ ประเด็นของการศึกษาหรือทดลองเหล่านี้ ก็เพื่อแยกแยะความแปรปรวนที่มีแบบแผนและทำนายได้ หรือที่เรียกว่า **ความแปรปรวนอย่างเป็นระบบ (systemic variation)** จากความแปรปรวนอย่างสุ่ม หรืออีกนัยหนึ่ง เพื่อตอบคำถามว่า (ความแปรปรวนใน) ผลการศึกษานั้น อธิบายได้จากการแปรปรวนอย่างเป็นระบบอย่างน้อยบางส่วน หรือจะอธิบายได้จากการแปรปรวนอย่างสุ่มหรือโดยบังเอิญเพียงอย่างเดียว

ตัวแปร (variable) คือค่าวัดคุณสมบัติบางประการ (ของสมาชิกในประชากร) เป็นตัวเลขที่มีความแปรปรวน แต่ตัวแปรหลายๆ ตัว ที่วัดจากกลุ่มตัวอย่าง นิยมเรียกว่า **ข้อมูล (data)** ดังนั้นตัวแปรจำนวนเท่าใดก็ตาม ที่วัดในกลุ่มตัวอย่างหนึ่งๆ เรียกว่าข้อมูลทั้งหมด

การกระจายของตัวแปรหรือข้อมูล เป็นการมองความแปรปรวนของข้อมูลอีกมุมหนึ่ง กล่าวคือจะเน้นความ “กว้าง” ของความแปรปรวน อันที่จริงมีตัวสถิติตัวหนึ่ง ที่มีชื่อเรียกว่า “ความแปรปรวน” (variance) เหมือนกัน ก็เป็นตัววัดการกระจายหรือความกว้างของความแปรปรวนของข้อมูลโดยตรง ผู้วิจัยอาจนำเสนอการกระจายของข้อมูลในรูปแบบของตัวสถิติบางตัว หรือนำเสนอในรูปแบบแผนภาพ เช่น ฮิสโตแกรม (รูปที่ 1x) ก็ได้

ตัวสถิติ (statistic) มีได้หมายถึงวิเศษสถิติ แต่หมายถึงตัวเลขที่ได้จากข้อมูลที่ใช้ประมาณตัววัดในระดับประชากร หรืออาจเป็นตัวเลขใดๆ ก็ตาม ที่คำนวณจากข้อมูลใน

ระดับตัวอย่าง การสรุปข้อมูลเป็นตัวเลข 2 ถึง 4 ตัว (summary statistics) ก็เป็นสถิติประเภทหนึ่ง ข้อมูลหรือตัวแปรที่เก็บหรือวัดได้ในกลุ่มตัวอย่าง จะไม่มีความหมายใดๆ ถ้าไม่มีการสรุปข้อมูลในรูปแบบที่ตีความได้หรือในรูปแบบที่ทำการวิเคราะห์ทางสถิติได้ อันที่จริง การตีความข้อมูล ก็หมายถึงการมองหาแบบแผนระหว่างความแปรปรวนของข้อมูลแต่ละชิ้น หรือมองหาความสัมพันธ์ระหว่างตัวแปรนั่นเอง

การสรุปข้อมูล⁴ ขึ้นกับคุณสมบัติเฉพาะของข้อมูลแต่ละประเภท ซึ่งแบ่งได้อย่างคร่าวๆ เป็น 3 ประเภท

(1) ข้อมูลหรือตัวแปรประเภท มาตรฐานวัดอันตรภาค (interval scale) เป็นข้อมูลแบบต่อเนื่อง (continuous) หรือข้อมูลนับ (counts) ซึ่งอย่างน้อยที่สุด จะมีคุณสมบัติเชิงอันตรภาคและเรียงลำดับได้ กล่าวคือ ความแตกต่างของค่าตัวแปรที่มีช่วงเท่ากัน จะมีความหมายเดียวกัน ยกตัวอย่างเช่น “อายุ” เป็นตัวแปรมาตรฐานวัดอันตรภาค เพราะความแตกต่างระหว่าง อายุ 26 กับ 24 ปี มีความหมายเดียวกับความแตกต่างระหว่างอายุ 32 กับ 30 ปี กล่าวคือ มีความแตกต่างคิดเป็น 2 ปีเท่ากัน มักนิยมที่จะสรุปตัวแปรประเภทนี้ เป็นค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน หรือ ค่ามัธยฐานและพิสัย

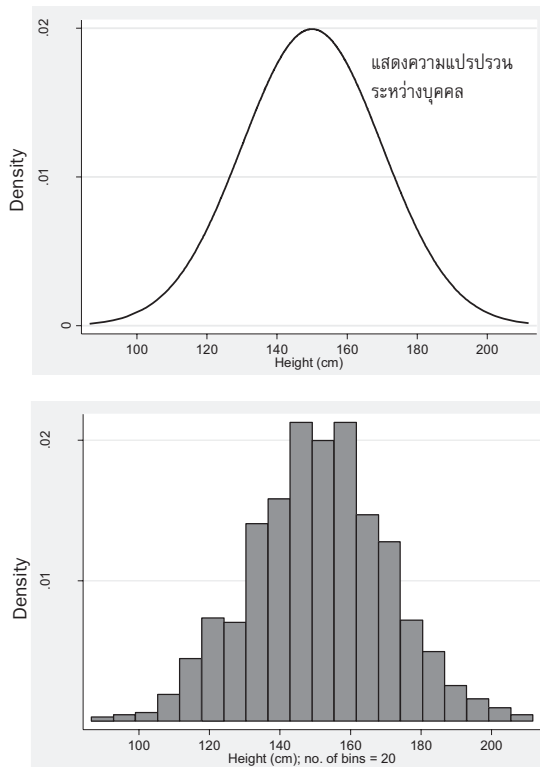
(2) ข้อมูลหรือตัวแปรประเภท มาตรฐานวัดค่านาม (nominal scale) เป็นข้อมูลไม่ต่อเนื่อง เรียกว่ามีหมวดหมู่ (category) และมีจำนวนหมวดหมู่ที่จำกัด ข้อมูลประเภทนี้ไม่มีคุณสมบัติเรียงลำดับได้ ไม่สามารถนำมาบวกกลบกันได้ แต่จะนับจำนวนในแต่ละหมวดหมู่และคิดเป็นสัดส่วนได้ จึงนิยมที่จะสรุปตัวแปรประเภทนี้เป็นจำนวนนับและสัดส่วน ตัวอย่างข้อมูลประเภทนี้ได้แก่ “เพศ” (หญิง/ชาย) และ “สาเหตุการเสียชีวิต” (มีหลายสาเหตุ) เป็นต้น

(3) ข้อมูลหรือตัวแปรประเภท มาตรฐานวัดแสดงลำดับ (ordinal scale) จะจัดเป็นหมวดหมู่ได้เหมือนตัวแปรค่านาม แต่จะมีคุณสมบัติเพิ่มคือ การเรียงลำดับได้โดยธรรมชาติเหมือนตัวแปรอันตรภาค จึงมักสรุปตัวแปรประเภทนี้เป็นจำนวนนับและสัดส่วน หรือเป็นค่ามัธยฐานและพิสัย ตัวอย่างข้อมูลประเภทนี้ได้แก่ “ระดับความพึงพอใจ” และ “ระยะของมะเร็ง” เป็นต้น

การแจกแจงของข้อมูล และการแจกแจงของกลุ่มตัวอย่าง

การกระจายของข้อมูล หรือตัวแปร ในระดับประชากร แสดงได้จาก การแจกแจงของข้อมูล (data distribution) ในระดับประชากร และการกระจายของข้อมูลในระดับกลุ่มตัวอย่าง แสดงได้จากการแจกแจงระดับตัวอย่าง (รูปที่ 1ก และ 1ข ตามลำดับ)

การแจกแจงของกลุ่มตัวอย่าง⁵ (sampling distribution) มักเป็นการแจกแจงของตัวสถิติที่ผู้วิจัยสนใจ เช่น ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน หรือ ตัวสถิติทดสอบ เป็นต้น



รูปที่ 1 ก (บน) การแจกแจงของข้อมูลในระดับประชากร (เป็นการแจกแจงความสูงในประชากร) ซึ่งมีลักษณะเป็นนอร์มอล (Normal) ส่วน รูปที่ 1ข (ล่าง) การกระจายของข้อมูลในระดับตัวอย่าง 1,000 ราย ในรูปแบบฮิสโตแกรม

การแจกแจงที่กล่าวถึงนี้ ต้องสร้างจากกลุ่มตัวอย่างที่เป็นไปได้ทั้งหมด จึงเป็นการแจกแจงในระดับประชากร และขึ้นกับขนาดของกลุ่มตัวอย่างโดยตรง ยกตัวอย่างเช่น ถ้าข้อมูลที่ต้องการศึกษาคือความปวดหลังผ่าตัดที่วัดเป็นมาตรา visual analog scale (VAS) และผู้วิจัยสนใจการแจกแจงของค่าเฉลี่ยของความปวด ก็ต้องระบุว่าเป็นค่าเฉลี่ยในกลุ่มตัวอย่างขนาดเท่าใด เช่น 40 ราย เป็นต้น แล้วดึงกลุ่มตัวอย่าง 40 ราย จากประชากรให้ครบทุกกลุ่มที่เป็นไปได้ (เช่นจากการสุ่มแบบไม่แทนที่ หรือ sampling without replacement จากประชากรนั้นๆ แต่เมื่อสุ่มเสร็จแล้วใส่คืนหมด แล้วสุ่มใหม่) จากนั้นนำข้อมูลของแต่ละกลุ่มตัวอย่างมาคำนวณค่าเฉลี่ยความปวด แล้วนำค่าเฉลี่ยทั้งหมดมาสร้างกราฟแสดงการแจกแจงของค่าเฉลี่ยดังกล่าว

ความแตกต่างที่สำคัญที่สุด ระหว่างการแจกแจงของข้อมูล กับการแจกแจงของตัวสถิติหรือการแจกแจงของกลุ่มตัวอย่าง คือ การแจกแจงของข้อมูล เช่นข้อมูลความปวดของคนๆ หนึ่ง หวังแสดงให้เห็นความแปรปรวนระหว่างบุคคล แต่การแจกแจงจากของตัวสถิติเป็นการแจกแจงเพื่อแสดงความแปรปรวนระหว่างกลุ่มตัวอย่าง หรือระหว่างค่าสถิติที่คำนวณจากกลุ่มตัวอย่าง เช่นค่าเฉลี่ยหรือค่าเบี่ยงเบนมาตรฐานของความปวด เป็นต้น และการแจกแจงอย่างหลังนี้ต้องแสดงในระดับประชากรเท่านั้นจึงจะเป็นประโยชน์ (จะสังเกตได้ว่า การแจกแจงของข้อมูลในระดับประชากร ก็เป็นกรณีพิเศษของการแจกแจงของกลุ่มตัวอย่าง กล่าวคือในกรณีที่กลุ่มตัวอย่างมีขนาดเท่ากับ 1 คน โปรดดูตารางที่ 1 ประกอบ)

การแสดงการแจกแจงของตัวสถิติโดยไม่อาศัยทฤษฎี ไม่น่าจะทำได้ในทางปฏิบัติ เพราะต้องอาศัยการทดลอง (ทำการวิจัย) ซ้ำแล้วซ้ำอีก โดยในแต่ละครั้งต้องทดลองในกลุ่มตัวอย่างที่แตกต่างกัน และที่เป็นไปได้ทั้งหมดจากประชากรนั้นๆ ซึ่งในชีวิตจริงผู้วิจัยมักทำการศึกษาเพียงครั้งเดียว (ก็จะได้กลุ่มตัวอย่าง 1 กลุ่มเท่านั้น) และอาจมีผู้วิจัยอื่นๆ ทำการศึกษาแบบเดียวกันซ้ำบ้าง (ก็จะได้จำนวนกลุ่มตัวอย่างมากขึ้น) แต่ก็ไม่เพียงพอที่จะสร้างการแจกแจงของตัวสถิติในระดับประชากรได้ ดังนั้นการแจกแจงดังกล่าว จึงต้องคำนวณโดยอาศัยแบบจำลองทางสถิติ (ตัวแบบทางคณิตศาสตร์) และอาจใช้ข้อมูลของกลุ่มตัวอย่างที่มีอยู่ ประกอบการกำหนดแบบจำลอง จึงจะได้มาซึ่งการแจกแจงในระดับประชากร ของค่าเฉลี่ย หรือค่าเบี่ยงเบนมาตรฐาน หรือตัวสถิติอื่นๆ ที่สนใจ ด้วย

ตารางที่ 1 ความแตกต่างระหว่างการแจกแจงข้อมูล กับการแจกแจงกลุ่มตัวอย่าง

คุณลักษณะ	การแจกแจงข้อมูล/ตัวแปร	การแจกแจงกลุ่มตัวอย่าง
1. ระดับของการแจกแจง	ระดับประชากรหรือกลุ่มตัวอย่างก็ได้	ระดับประชากรเท่านั้น
2. แสดงความแปรปรวนระหว่าง	สมาชิก (บุคคล)	กลุ่มตัวอย่าง (กลุ่มบุคคล)
3. ตัวแปรที่ศึกษา	ตัวแปรที่วัดในสมาชิกแต่ละคน เช่น ความปวดของคนที่หนึ่ง	ตัวสถิติที่วัดในแต่ละกลุ่ม เช่น ความปวดเฉลี่ยในกลุ่มตัวอย่างกลุ่มหนึ่ง
4. รูปร่างการแจกแจง (Shape of the distribution)	- มีรูปร่างคงที่ในระดับประชากร - รูปร่างคงที่โดยประมาณในกลุ่มตัวอย่างที่มีขนาดใหญ่พอสมควร - ความแปรปรวนเป็นค่าคงที่ประชากร ⁵	- มีรูปร่างที่แคบลงหรือเปลี่ยนไปเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น - ความแปรปรวนลู่เข้า 0 เมื่อขนาดตัวอย่างเข้าใกล้ขนาด
5. การคำนวณฟังก์ชันการแจกแจง (Distribution function)	เป็นการคำนวณตามข้อมูล หรือยึดข้อมูลเป็นหลัก	เป็นการคำนวณตามทฤษฎี (อาศัยข้อมูลตามจำเป็น)
6. การพึ่งแบบจำลองหรือตัวแบบทางสถิติ	จะอาศัยหรือไม่อาศัยตัวแบบก็ได้	ต้องอาศัยตัวแบบเป็นหลัก
7. ประโยชน์ที่นำไปใช้ในทางสถิติ	ใช้บรรยายข้อมูล หรือสร้างแบบจำลอง	ใช้ทดสอบสมมติฐานทางสถิติ ฯลฯ
8. ชื่อที่ใช้เรียกตัววัดความแปรปรวน	ความเบี่ยงเบนมาตรฐาน (standard deviation, sd)	ความคลาดเคลื่อนมาตรฐาน (standard error, se)

เหตุนี้ความถูกต้องของการแจกแจงของตัวสถิติ จึงขึ้นกับความถูกต้องแบบจำลองและความแม่นยำของข้อมูลที่ใช้กำหนดการแจกแจงนั้นเป็นสำคัญ

การทดสอบสมมติฐานทางสถิติ

ประโยชน์ในทางคลินิก ของการแจกแจงข้อมูลหรือตัวแปรในระดับประชากร จะสังเกตได้จากตัวอย่างดังต่อไปนี้ กล่าวคือ หากแพทย์ได้ผ่าตัดผู้ป่วยคนหนึ่ง แล้วพบว่า

ผู้ป่วยมีอาการปวดหลังผ่าตัดเป็นอย่างมากเมื่อวัดด้วยมาตร VAS และแพทย์คนนั้นอยากทราบว่า ความปวดนี้มากเกินความปวดของคนอื่นๆ ไปหรือไม่ หากแพทย์คนนั้นทราบการแจกแจงของความปวดหลังการผ่าตัดด้วยวิธีดังกล่าวในประชากรที่ตนดูแลอยู่ (ไม่ต้องสนใจว่าจะทราบได้อย่างไร!) ก็จะตอบคำถามได้โดยเปรียบเทียบระดับความปวดที่วัดได้ กับการแจกแจงนั้น

แต่ถ้าแพทย์อยากทราบว่าผู้ป่วยคนเดียวคนนี้ มีโอกาสป่วยเป็นโรคขาดสารอินโดฟินมากน้อยเพียงใด (ซึ่งจะทำให้ผู้ป่วยมีความรู้สึกปวดหลังผ่าตัดมากกว่าปกติ) แพทย์คนนั้นก็ควรนำความปวดที่วัดได้ในผู้ป่วยนั้น ไปเปรียบเทียบกับ การแจกแจงความปวดในประชากรที่ป่วยเป็นโรคขาดสารอินโดฟิน ที่ได้รับการผ่าตัดด้วยวิธีเดียวกัน

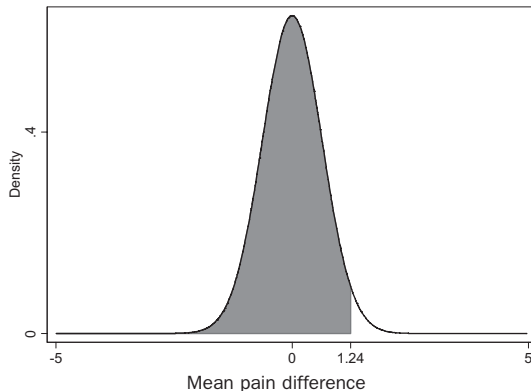
และถ้าหากแพทย์อยากทราบว่าผู้ป่วยมีอาการปวดที่สอดคล้องกับความปวดของคนปกติทุกๆ ไป หรือกับคนที่ป่วยเป็นโรคขาดสารอินโดฟินมากกว่ากัน (เท่ากับว่าแพทย์ต้องการวินิจฉัยว่าผู้ป่วยที่ได้รับการผ่าตัดนั้น มีโอกาสป่วยเป็นโรคขาดสารอินโดฟินร่วมด้วย หรือเป็นผู้ป่วยที่ไม่มีโรคดังกล่าวมากกว่ากัน โดยใช้ความปวดหลังผ่าตัดเป็นตัวช่วยวินิจฉัย) ก็ต้องนำความปวดที่วัดได้ไปเทียบกับการแจกแจงความปวดทั้งสองกรณี กล่าวคือกรณีประชากรปกติและกรณีผู้ป่วยเป็นโรคขาดสารอินโดฟิน ที่ได้รับการผ่าตัดด้วยวิธีเดียวกันทั้งหมด แล้วหาตัวชี้วัดที่บ่งบอกว่าความปวดดังกล่าวสอดคล้องกับความปวดที่มาจากประชากรกลุ่มใดมากกว่ากัน

หลักการ **ทดสอบสมมติฐานทางสถิติ** ก็เหมือนกับกรณีวินิจฉัยโรคที่กล่าวถึงแล้ว แต่จะใช้ทดสอบกลุ่มตัวอย่าง ไม่ใช่กับผู้ป่วยเพียงหนึ่งคน ยกตัวอย่างเช่น ถ้าหากผู้วิจัยทำการศึกษาสุ่มเปรียบเทียบการผ่าตัด 2 วิธี (วิธี ก กับ ข) ในผู้ป่วย 40 ราย แบ่งเป็น 2 กลุ่ม (20 คนต่อกลุ่ม) เพื่อสังเกตความแตกต่างในแง่ความปวดหลังผ่าตัดที่วัดโดย VAS (มีค่า 0 ถึง 10) แล้วพบว่าความปวดโดยเฉลี่ยในกลุ่ม ก สูงกว่าในกลุ่ม ข คำถามที่อาจเกิดขึ้นคือ ความแตกต่าง (โดยเฉลี่ย) ดังกล่าวที่สังเกตได้จากการศึกษา จะเป็นจริงในระดับประชากร หรือเป็นความแตกต่างโดยบังเอิญกันแน่

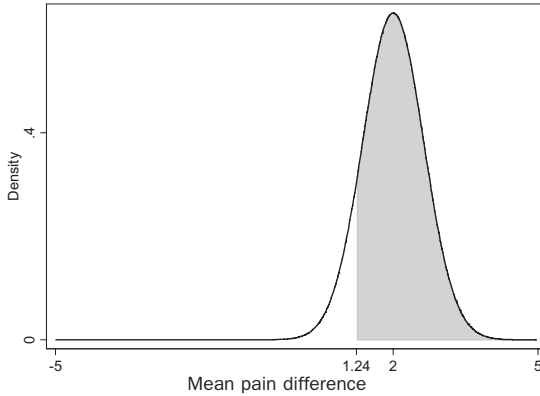
ในการตอบคำถาม ผู้วิจัยต้องกำหนดแบบจำลอง หรือตัวแบบทางสถิติให้กับตัวแปรความปวด แล้วสร้างการแจกแจงให้แก่ตัวสถิติที่ต้องการศึกษา (ตัวสถิติในตัวอย่างนี้คือ ค่าเฉลี่ยของความแตกต่างในความปวดหลังการผ่าตัด ระหว่างวิธี ก กับ ข)

ซึ่งก็คือการแจกแจงของกลุ่มตัวอย่างนั่นเอง โดยมีการแจกแจง 2 กรณี คือกรณีที่ได้จากการกำหนด **สมมติฐานว่าง** (ไม่มีความแตกต่างโดยเฉลี่ยในระดับประชากร หรืออีกนัยหนึ่งคือ กลุ่มตัวอย่างทั้งสอง ได้รับการคัดเลือกหรือสุ่ม มาจากประชากรเดียวกัน; รูปที่ 2) และกรณีที่ได้จาก **สมมติฐานทางเลือก** (มีความแตกต่างจริงในระดับประชากร; รูปที่ 3) แล้วนำผลการวิจัย มาเปรียบเทียบกับ การแจกแจงของกลุ่มตัวอย่างทั้ง 2 กรณีว่าจะสอดคล้องกับการแจกแจงกรณีใด (รูปที่ 4) หรืออีกนัยหนึ่ง จะสอดคล้องกับสมมติฐานใดมากกว่ากัน (โปรดสังเกตว่าในทางทฤษฎี การแจกแจงทั้ง 2 กรณี ได้จากการเลือกกลุ่มตัวอย่างขนาด 40 ราย ที่เป็นไปได้ทั้งหมด จากประชากรที่ศึกษา)

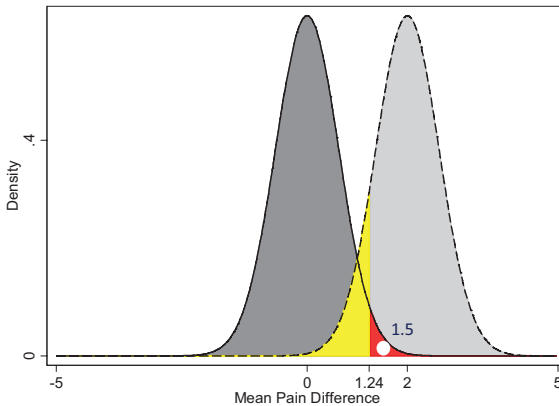
การมีสองสมมติฐานที่ตรงข้ามกัน กล่าวคือสมมติฐานว่างและสมมติฐานทางเลือก จะช่วยให้การตอบคำถามทางวิจัยมีความชัดเจนยิ่งขึ้น และการปรับสมมติฐานว่างและสมมติฐานทางเลือกตามความเหมาะสม จะทำให้คำถามทางวิทยาศาสตร์มีความหลากหลายมากขึ้นอีกด้วย นอกจากนี้ ในบางกรณีการเลือกสมมติฐานทางเลือกที่แคบมากขึ้น หรือมีความจำเพาะสูงขึ้น อาจทำให้ความไว (เพาเวอร์) ของการทดสอบเพิ่มขึ้นได้



รูปที่ 2 การแจกแจงของกลุ่มตัวอย่าง (sampling distribution) ของตัวสถิติ “ค่าความแตกต่างของความปวดโดยเฉลี่ย” (mean pain difference) ภายใต้สมมติฐานว่าง สังเกตว่าค่าเฉลี่ยของการแจกแจงเป็น 0 (ศูนย์) นอกจากนี้ได้แสดงพื้นที่สีขาว คิดเป็นร้อยละ 5 ของพื้นที่ทั้งหมด และแสดงค่า “cut-off” 1.24 ที่เป็นขอบเขตของพื้นที่ดังกล่าว ซึ่งถือเป็นพื้นที่ “ปฏิเสธสมมติฐานว่าง”



รูปที่ 3 การแจกแจงของตัวสถิติ “ค่าความแตกต่างของความปวดโดยเฉลี่ย” ภายใต้สมมติฐานทางเลือกตัวหนึ่ง (ที่มีค่าความแตกต่างเป็น 2 หน่วยโดยเฉลี่ย) ซึ่งมีพื้นที่สีขาวคิดเป็นร้อยละ 20 ของพื้นที่ทั้งหมด ถือเป็นพื้นที่ “ปฏิเสธสมมติฐานทางเลือก” ตัวนี้โดยมีค่า “cut-off” เดียวกับในรูปที่ 2 ส่วนพื้นที่สีเทาถือเป็นพื้นที่ “ยอมรับสมมติฐานทางเลือก”



รูปที่ 4 ผลการศึกษาสมมติ ที่ได้ค่าความแตกต่างเฉลี่ยคือ 1.5 (วงกลม) เปรียบเทียบกับการแจกแจงภายใต้สมมติฐานว่าง (เส้นโค้งทึบ) และภายใต้สมมติฐานทางเลือก (เส้นโค้งโปร่ง) ว่าน่าจะสนับสนุนสมมติฐานใดมากกว่ากัน ในกรณีนี้น่าจะสนับสนุนสมมติฐานทางเลือก เพราะค่า 1.5 ตกอยู่ในพื้นที่ “ปฏิเสธสมมติฐานว่าง” (รูปที่ 2) และอยู่ในพื้นที่ “ยอมรับสมมติฐานทางเลือก” (รูปที่ 3)

ผลการทดสอบทางสถิติที่สรุปยอมรับสมมติฐานทางเลือก (สรุปว่ามีความแตกต่างระหว่าง 2 กลุ่มเปรียบเทียบ) เปรียบได้กับการวินิจฉัยว่าผู้ป่วยเป็นโรค ส่วนการทดสอบที่สรุปยอมรับสมมติฐานว่าง (สรุปว่าไม่มีความแตกต่าง) เปรียบเสมือนวินิจฉัยว่าไม่เป็นโรค หากสรุปจากการทดสอบว่ายอมรับสมมติฐานทางเลือกทั้งๆ ที่สมมติฐานว่างเป็นจริง ที่เรียกว่า **ความผิดพลาด หรือความคลาดเคลื่อน แบบที่ 1 (type I error)** ก็เสมือนกับการวินิจฉัยว่าคนๆ หนึ่งเป็นโรคทั้งๆ ที่ไม่ได้เป็น ก็เรียกว่าผลบวกเทียม (false positive diagnosis) หรือการทดสอบที่ยอมรับสมมติฐานว่างทั้งที่สมมติฐานทางเลือกเป็นจริง ที่นักสถิติเรียกว่า **ความผิดพลาดแบบที่ 2 (type II error)** ก็เปรียบได้กับการวินิจฉัยว่าไม่เป็นโรคทั้งๆ ที่เป็น (ผลลบเทียมหรือ false negative diagnosis)

นักสถิติเน้นให้มีความผิดพลาดแบบที่ 1 น้อยๆ (ไม่เกินร้อยละ 5) กล่าวคือให้มีโอกาสสูงที่จะสรุปยอมรับสมมติฐานว่างเมื่อสมมติฐานว่างเป็นจริง (ไม่น้อยกว่าร้อยละ 95) ก็เสมือนกับทำให้แพทย์ยอมรับโอกาสการวินิจฉัยว่าไม่เป็นโรคในขณะที่ไม่ได้เป็นโรคจริงให้มาก ในทางการแพทย์เรียกว่าให้ยอมรับการวินิจฉัยที่มีความจำเพาะสูง (high specificity) นั้นเอง แต่นักสถิติจะยอมรับความผิดพลาดแบบที่ 2 ได้มากกว่า เช่น ไม่เกินร้อยละ 20 กล่าวคือ ให้มีโอกาสรสรุปยอมรับสมมติฐานทางเลือกโดยที่สมมติฐานทางเลือกเป็นจริงเพียงร้อยละ 80 หรือมากกว่า เสมือนให้แพทย์ยอมรับการวินิจฉัยถูกว่าเป็นโรคเมื่อเป็นโรคจริงเพียงร้อยละ 80 หรืออีกนัยหนึ่งให้ยอมรับความไวในการวินิจฉัยโรค (sensitivity) ไม่มากเท่าความจำเพาะ (รูปที่ 5 ประกอบ)

การทดสอบทางสถิติ ถือได้ว่าเป็นการทดสอบความสัมพันธ์ (test of association) ยกตัวอย่างการทดสอบความแตกต่างระหว่างการผ่าตัด 2 วิธีในแง่ของความปวดที่วัดด้วยมาตร VAS ว่าน่ามีความแตกต่างจริงหรือไม่ ก็เป็นการทดสอบความสัมพันธ์ระหว่างวิธีผ่าตัดกับความปวด (กล่าวคือ เมื่อการผ่าตัดเปลี่ยนไป ความปวดก็เปลี่ยนไปด้วย และเปลี่ยนอย่างมีแบบแผน คือการผ่าตัด ก จะปวดมากกว่าการผ่าตัด ข โดยเฉลี่ย) ซึ่งความสัมพันธ์ทางสถิติเป็นความสัมพันธ์โดยเฉลี่ยเท่านั้น อันหมายความว่า สำหรับผู้ป่วยคนหนึ่งๆ ไม่มีใครบอกได้แน่นอนว่าหากได้รับการผ่าตัดด้วยวิธี ก ผู้ป่วยดังกล่าวจะรู้สึกปวดมากกว่าการได้รับผ่าตัดด้วยวิธี ข แต่ถ้ามีผู้ป่วยหลายๆ คน โดยแต่ละคนได้รับการสุ่มผ่าตัดวิธีใดวิธีหนึ่งในสองวิธีนั้น ความปวดเฉลี่ยในกลุ่ม ก จะสูงกว่าในกลุ่ม ข ก่อน

“2×2 Table for Research Studies”

	Conclude: Different	Conclude: Not Different
No True Difference	Type I error (α)	“Specificity”
True Difference	Power = $1 - \beta$ “Sensitivity”	Type II error (β)

รูปที่ 5 ตารางความคลาดเคลื่อนทางการวิจัย (research errors) ประเภท (types) ที่ 1 และ 2 และการนิยาม “ความไว” (sensitivity หรือ power) และ “ความจำเพาะ” (specificity) ของงานวิจัย

ข้างแน่นอน

ปัญหาหลักของวิธีทดสอบสมมติฐานที่กล่าวถึงแล้วคือ ผู้วิจัยต้องทราบรายละเอียดของการแจกแจงของข้อมูลหรือตัวแปรผลลัพธ์ที่ตนสนใจภายใต้สมมติฐานต่างๆ ในระดับประชากร ซึ่งในทางปฏิบัติไม่มีผู้ใดทราบได้ ยกตัวอย่างเช่น ในกรณีที่ผลลัพธ์มีการแจกแจงแบบนอร์มอล ผู้วิจัยต้องทราบค่าเบี่ยงเบนมาตรฐานของการแจกแจงดังกล่าวด้วย ซึ่งไม่มีใครทราบค่าที่แท้จริง คัพท์ทางสถิติคือไม่มีผู้ใดทราบค่า **พารามิเตอร์ (parameter)** เช่นค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานในระดับประชากร ดังนั้นนักสถิติจึงต้องเลี่ยงปัญหานี้ โดยสร้างการแจกแจงของตัวสถิติทดสอบอื่นๆ ในระดับประชากร ที่ไม่พึ่งพารามิเตอร์ที่ไม่มีใครทราบค่า ซึ่งมีสองวิธีหลัก

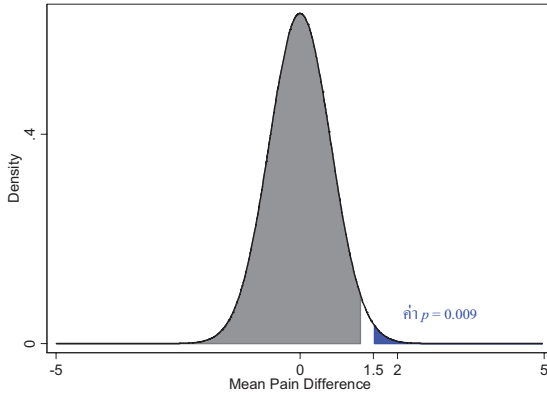
วิธีแรกจะใช้ในกรณีที่กำหนดได้ว่าตัวแปรเชิงปริมาณที่จะทดสอบมีการแจกแจงในทางทฤษฎีอย่างไร ซึ่งส่วนมากจะหมายถึงการแจกแจงแบบนอร์มอล และให้ประมาณพารามิเตอร์ของการแจกแจงดังกล่าวจากข้อมูลที่มีอยู่ (เช่น ในกรณีการแจกแจงแบบนอร์มอล ให้ประมาณค่าเฉลี่ยหรือค่าเบี่ยงเบนมาตรฐานหรือทั้งสองค่า) แล้วสร้างตัวสถิติทดสอบจากค่าประมาณเหล่านี้ ที่มีคุณสมบัติเหมาะสมที่จะใช้ทดสอบสมมติฐานต่างๆ และมีการแจกแจง (ของตัวสถิตินั้น) ในระดับประชากร ที่ปลอดจากพารามิเตอร์ที่ไม่ทราบค่าแท้จริง

ตัวสถิติดังกล่าวเรียกว่า “ตัวสถิติหลัก” หรือ “pivotal statistic” และมีชื่อที่นักวิจัยส่วนใหญ่คุ้นกันดี อันได้แก่ t-statistic (ที่ใช้ในการทดสอบ t-test; รูปที่ 7) และ F-statistic (ใช้ในการทดสอบ F-test หรือ ANOVA) เป็นต้น สถิติการทดสอบเหล่านี้มีชื่อเรียกรวมๆ ว่า **สถิติทดสอบที่พึ่งการแจกแจง** อันหมายถึงการแจกแจงแบบนอร์มอลเสียส่วนใหญ่ (distribution-dependent statistical test) หรืออาจเรียกอีกชื่อหนึ่งที่ว่าอาจคุ้นมากกว่า (แต่ไม่ใช่ชื่อที่เหมาะสม) คือ **สถิติทดสอบแบบพารามิเตอร์ (parametric statistical test)**

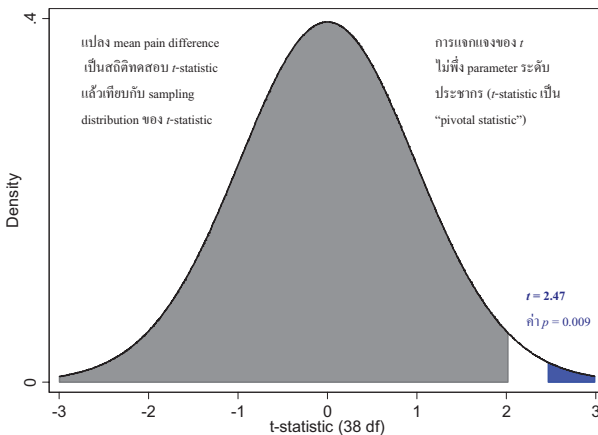
วิธีที่สอง จะใช้ในกรณีที่ไม่สามารถกำหนดได้ว่า ตัวแปรเชิงปริมาณจะมีการแจกแจงอย่างไรในทางทฤษฎี หรือทราบเพียงว่าไม่น่าเป็นการแจกแจงแบบนอร์มอล ก็จะใช้สถิติทดสอบที่พึ่งเฉพาะการเรียงลำดับข้อมูล (rank-based statistics หรือ “order statistics”) ซึ่งมักมีชื่อเรียกรวมๆ ว่า **สถิติทดสอบที่ไม่พึ่งการแจกแจง (distribution-free statistical test)** หรืออีกชื่อหนึ่งคือ **สถิติทดสอบแบบไม่พารามิเตอร์ (non-parametric statistical test)** การทดสอบเหล่านี้มีชื่อเฉพาะที่คุ้นกันดีเช่นกัน ได้แก่ Wilcoxon rank-sum test หรือ Mann-Whitney U test หรือ Kruskal-Wallis test เป็นต้น

ในกรณีที่ตัวแปรผลลัพธ์เป็นประเภทมาตรวัดค่านาม (nominal variable) และผู้วิจัยต้องการทดสอบความแตกต่างระหว่างกลุ่มเปรียบเทียบ ตัวสถิติที่ใช้ทดสอบโดยทั่วไปจะมีการแจกแจงแบบโคสแคร์ ซึ่งเป็นการแจกแจงที่ไม่พารามิเตอร์ที่ไม่ทราบค่าอยู่แล้ว และสำหรับกรณีตัวแปรผลลัพธ์ประเภทมาตรวัดแสดงลำดับ (ordinal variable) สถิติที่ใช้ทดสอบความแตกต่างที่นิยม มักเป็นสถิติทดสอบที่พึ่งการเรียงลำดับข้อมูล ซึ่งก็ไม่ต้องทราบค่าพารามิเตอร์ระดับประชากรเช่นกัน

ในทางปฏิบัติ การทดสอบทางสถิติมักหมายถึงการทดสอบว่ากลุ่มต่างๆ ในการศึกษา มีความแตกต่างกันจริงหรือไม่ ดังที่กล่าวแล้ว ดังนั้นตัวสถิติทดสอบที่นิยม จึงเป็นตัวชี้วัดว่าข้อมูลที่ได้นั้น แสดงถึงความแตกต่างจริงหรือไม่ เช่นกัน ตัวสถิติไม่ว่าจะเป็น t-statistic หรือโคสแคร์ จะมีค่ารอบๆ 0 (ศูนย์) หากข้อมูลแสดงว่าไม่น่ามีความแตกต่างจริง แต่จะมีค่าห่างจาก 0 หากมีแนวโน้มจะแตกต่าง การแจกแจงของสถิติทดสอบจึงสร้างขึ้นภายใต้สมมติฐานว่างเป็นหลัก เพราะหากตัวสถิติทดสอบจากการศึกษาหนึ่งๆ ภายใต้สมมติฐานว่าง มีค่าห่างจาก 0 มากจนตกไปอยู่ในส่วนปลายของการแจกแจง อัน



รูปที่ 6 ความหมายของค่า p ซึ่งเป็นพื้นที่ใต้เส้นโค้งของการแจกแจงตัวสถิติทดสอบภายใต้สมมติฐานว่าง ที่วัดจากผลการศึกษาซึ่งในตัวอย่างคือ 1.5 (ดูรูปที่ 4) พื้นที่นี้ หรือค่า p มีค่าเท่ากับ 0.009 (ในขณะที่พื้นที่ใต้เส้นโค้งทั้งเส้นมีค่าเท่ากับ 1.00) อันบ่งบอกถึงความแตกต่างที่วัดได้มีนัยสำคัญทางสถิติ



รูปที่ 7 การแจกแจงของตัวสถิติหลัก หรือ pivotal statistic ซึ่งในกรณีนี้เป็นตัวสถิติ t เป็นการแจกแจงที่ได้จากการดัดแปลงรูปที่ 6 หลังจากที่เปลี่ยนความแตกต่างเฉลี่ยไปเป็นตัวสถิติ t แล้ว จะสังเกตเห็นว่า ข้อสรุปที่ได้และค่า p จะเหมือนกับในรูปที่ 6 ทุกประการ (หมายเหตุ: degree of freedom มีค่าเท่ากับ 38 เพราะขนาดตัวอย่างโดยรวมมี 40 ราย)

แสดงถึงโอกาสที่เกิดขึ้นได้น้อย ตัวสถิติจากการศึกษาดังกล่าว ย่อมไม่สนับสนุนสมมติฐานว่าง แต่จะสนับสนุนสมมติฐานทางเลือก กล่าวคือ สนับสนุนว่า น่าจะมีความแตกต่างระหว่างกลุ่มจริง มากกว่าไม่แตกต่าง นั่นเอง

นิยามค่า p และช่วงความเชื่อมั่นร้อยละ 95

ค่า p (p -value) คือ ค่าความน่าจะเป็นหรือโอกาส (probability or chance) ที่จะเกิดผลลัพธ์ที่เท่ากับ หรือสุดโต่งหรือสุดขีดมากกว่าผลที่ได้จากการศึกษาหนึ่งๆ ภายใต้สมมติฐานว่าง นิยมใช้เป็นตัววัดตัวหนึ่งของความ “ห่าง” ของผลการศึกษานั้น จากสมมติฐานว่าง กล่าวคือ ถ้าค่า p ยิ่งน้อย ผลการศึกษาก็ยิ่ง “ห่าง” จากสมมติฐานว่าง ก็ยิ่งบ่งบอกว่าสมมติฐานว่างไม่น่าเป็นจริง (ไม่เป็นจริงในระดับประชากร) จึงควรปฏิเสธสมมติฐานว่าง แล้วยอมรับสมมติฐานทางเลือกแทน ผู้วิจัยสามารถคำนวณ ค่า p ดังกล่าวได้จากพื้นที่ใต้กราฟการแจกแจงของตัวสถิติทดสอบ (รูปที่ 6 และ 7)

ค่า p ที่มีนัยสำคัญ มักหมายถึงค่า p ที่เท่ากับหรือน้อยกว่าค่าความผิดพลาดแบบที่ 1 (type I or alpha error) ซึ่งนิยามกำหนดให้มีค่า 0.05 หรือระดับที่ร้อยละ 5 ของจำนวนการศึกษาทั้งหมดที่อาจทำได้ในทางทฤษฎี (hypothetical 5% error rate) ดังนั้นในทางปฏิบัติ ถ้าการศึกษาหนึ่งๆ ให้ค่าสถิติทดสอบที่ตกลงในบริเวณของการแจกแจงของตัวสถิติดังกล่าวภายใต้สมมติฐานว่าง ที่มีพื้นที่น้อยกว่า 0.05 (ตาม alpha error ที่กำหนดไว้) ก็ให้ปฏิเสธสมมติฐานว่าง (ไม่มีความแตกต่างจริงระหว่างกลุ่มเปรียบเทียบ; รูปที่ 2 และ 4) แล้วยอมรับสมมติฐานทางเลือก (มีความแตกต่างระหว่างกลุ่มจริง; รูปที่ 3 และ 4) แทน โดยมีโอกาสสรุปหรือตัดสินใจผิดได้ ไม่เกินร้อยละ 5 ในระยะยาว

ช่วงความเชื่อมั่น ที่ร้อยละ 95 (95% confidence interval) เป็นตัวชี้วัดความไม่แน่นอนแบบหนึ่ง ของการประมาณค่าพารามิเตอร์ ที่ประมาณจากข้อมูลของกลุ่มตัวอย่าง ยกตัวอย่างเช่น หากประมาณค่าเฉลี่ยของความปวดหลังผ่าตัดจากกลุ่มตัวอย่างผู้ป่วยกลุ่มหนึ่ง ได้ 5.6 หน่วย (จาก VAS ที่มีค่าระหว่าง 0 ถึง 10) และคำนวณช่วงความเชื่อมั่นร้อยละ 95 ได้เป็น 4.1 ถึง 7.3 หมายความว่า ถึงแม้ค่าเฉลี่ยความปวดหลังผ่าตัดในประชากรนั้นๆ ประมาณค่าได้ 5.6 แต่มีความ “เชื่อมั่น” ที่ร้อยละ 95 ว่าค่าจริงในระดับประชากร (พารามิเตอร์) น่าจะอยู่ระหว่าง 4.1 ถึง 7.3 หน่วย

อีกตัวอย่างหนึ่งได้แก่การศึกษาเปรียบเทียบความเสี่ยงของการเกิดไข้ (febrile morbidity) หลังผ่าตัดคลอด (cesarean section) ระหว่างผู้ป่วยที่ได้รับการทำความสะอาดและไม่ได้รับการทำความสะอาดช่องคลอดก่อนผ่าตัด ซึ่งมีผู้ป่วยทั้งหมด 600 ราย แล้วพบว่าผู้ป่วยที่ไม่ได้รับการทำความสะอาดช่องคลอดมีความเสี่ยงที่จะเป็นไข้เพิ่มขึ้น 2.7 เท่า และมีช่วงความเชื่อมั่นร้อยละ 95 ระหว่าง 1.9 และ 3.9 เท่า ก็หมายความว่าในระดับประชากร ค่าอัตราส่วนความเสี่ยง (พารามิเตอร์) นี้ ถึงจะประมาณได้ว่ามีค่าคร่าวๆ คือ 2.7 เท่า แต่ก็น่าจะมีค่าจริงอยู่ในช่วง 1.9 ถึง 3.9 เท่า ด้วยความเชื่อมั่นร้อยละ 95

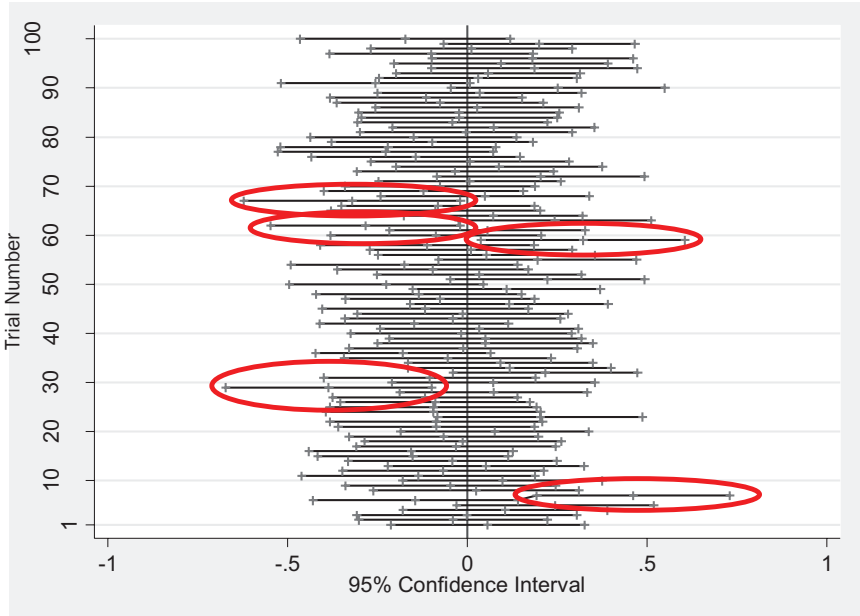
สังเกตว่า “ความเชื่อมั่นร้อยละ 95” ที่คำนวณได้จากการศึกษาหนึ่งๆ ไม่ใช่ “ความน่าจะเป็นร้อยละ 95” จึงกล่าวไม่ได้ว่าค่าพารามิเตอร์ที่จะประมาณ จะมีโอกาสเท่ากับ 0.95 (หรือร้อยละ 95) ที่จะมีค่าจริงอยู่ภายในช่วงดังกล่าว ในทางสถิติ ความหมายที่ถูกต้องที่สุดของช่วงความเชื่อมั่นร้อยละ 95 คือ “ในจำนวนการศึกษาที่เป็นไปได้ในทางทฤษฎี (hypothetical long run of experiments) ค่าพารามิเตอร์จริง จะตกอยู่ในช่วงความเชื่อมั่นที่คำนวณได้จากการศึกษาแต่ละการศึกษา คิดเป็นสัดส่วนร้อยละ 95 ของจำนวนการศึกษาทั้งหมด” (รูปที่ 8 ประกอบ)

ตัวอย่างการทดสอบสมมติฐานทางสถิติ

ขอยกตัวอย่างการทดสอบสมมติฐานทางสถิติ โดยอาศัยการศึกษาเปรียบเทียบความเสี่ยงของการเกิดไข้หลังผ่าตัดคลอด ซึ่งจากการศึกษาดังกล่าว พบการเกิดไข้ 93 ราย ในผู้ป่วยผ่าตัดคลอด 300 รายที่ไม่ได้รับการทำความสะอาดช่องคลอดก่อนผ่าตัด คิดเป็นสัดส่วนร้อยละ 31 และเกิดไข้ 34 รายในผู้ป่วย 300 รายที่ได้รับการทำความสะอาด คิดเป็นสัดส่วนร้อยละ 11 คำถามทางคลินิกคือ สัดส่วนการเกิดไข้แตกต่างกันจริงระหว่างการทำหรือไม่ทำความสะอาดช่องคลอดหรือไม่

ข้อมูลที่ได้ สามารถนำเสนอเป็นตาราง 2×2 เรียกว่าตารางสัญจร (contingency table) ดังในตารางที่ 2

การทดสอบทางสถิติ เริ่มที่การกำหนดสมมติฐานว่าง คือไม่มีความแตกต่างจริงระหว่างกลุ่ม และสมมติฐานทางเลือก คือมีความแตกต่างจริง แล้วอาศัยตัวชี้วัดที่สร้างจากข้อมูล เป็นตัวสถิติทดสอบในการตัดสินว่าควรสนับสนุนหรือปฏิเสธสมมติฐานใด ซึ่ง



รูปที่ 8 “ช่วงความเชื่อมั่นที่ร้อยละ 95” ทั้งหมด 100 ช่วง แต่ละช่วงเสนอเป็นเส้นนอนที่มีปลายเส้นเป็นเครื่องหมาย + (หมายถึงขอบเขตหรือ limits ของช่วง) และคำนวณจากการศึกษาจำลอง (simulated trials) ทั้งหมด 100 ครั้ง ในการจำลองเหตุการณ์นี้ ค่าความแตกต่างที่แท้จริงคือ 0 (ศูนย์) ดังนั้นความหมายที่ถูกต้องของ “ช่วงความเชื่อมั่นที่ร้อยละ 95” คือค่าเฉลี่ยที่แท้จริง (ศูนย์) จะตกอยู่ในช่วงดังกล่าว เป็นสัดส่วน 95 ใน 100 ดังในรูปนี้ และ 5 ใน 100 ช่วง (ที่แสดงอยู่ในวงรี) จะไม่มีค่าที่แท้จริงตกผ่าน

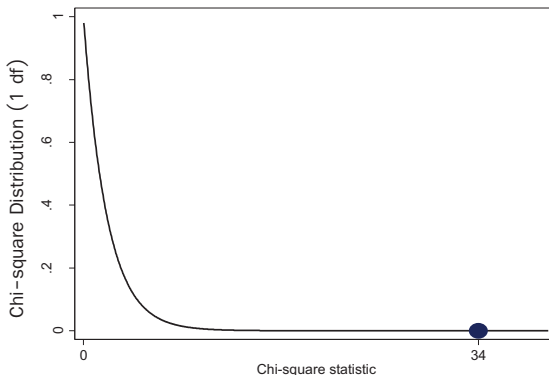
ตารางที่ 2 การเกิดใช้หลังผ่าตัดคลอดในแต่ละกลุ่ม

ผลลัพธ์	ท่าความสะอาดช่องคลอด (n = 300)	ไม่ได้ทำความสะอาด (n = 300)
มีใช้หลังผ่าตัด	34 (11%)	93 (31%)
ไม่มีใช้	266 (89%)	207 (69%)

คือตัวสถิติไคสแควร์ ที่มีการแจกแจงของตัวสถิติแบบไม่พึ่งพารามิเตอร์และมีชื่อไคสแควร์เช่นกัน

ตัวสถิติไคสแควร์ภายใต้สมมติฐานว่าง สำหรับตารางสัญจร 2×2 จะมีการแจกแจงดังในรูปที่ 9 และสำหรับข้อมูลตัวอย่างนี้ ค่าสถิติไคสแควร์ มีค่าเท่ากับ 34.8 ซึ่งต่างจาก 0 (ศูนย์) มากจนตกไปอยู่ในบริเวณส่วนปลายของการแจกแจง ทำให้ค่า p น้อยกว่า 0.001 อันแสดงถึงการสนับสนุนสมมติฐานทางเลือกมากกว่าสมมติฐานว่าง กล่าวคือในการศึกษาตัวอย่าง สัดส่วนของการเกิดไข้หลังผ่าตัด น่าจะแตกต่างระหว่างกลุ่มจริงมิได้เกิดขึ้นโดยบังเอิญ โดยมีแนวโน้มสูงกว่ามากในกลุ่มที่ไม่ทำความสะอาดช่องคลอด (2.7 เท่า)

ถ้าหากสถิติทดสอบมีค่าน้อย และค่า p ก็มากกว่า 0.05 ย่อมแสดงว่าข้อมูลที่ไม่สนับสนุนว่ามีความแตกต่างระหว่างกลุ่มเปรียบเทียบจริง จึงควรยอมรับสมมติฐานว่างคือ “ไม่มีความแตกต่าง” อย่างไรก็ตาม การตีความผลการวิเคราะห์นี้อย่างถูกต้อง กลับมิใช่ว่าไม่มีความแตกต่างระหว่างกลุ่ม แต่ความแตกต่างที่มี *อธิบายได้จากความแปรปรวนสุ่ม (random variation) ทั้งหมด* เนื่องจากความแปรปรวนสุ่มจะมีอิทธิพลมากในการ



รูปที่ 9 เส้นโค้งการแจกแจงของตัวสถิติไคสแควร์ภายใต้สมมติฐานว่างสำหรับตารางสัญจร 2×2 (ซึ่งมี 1 degree of freedom) จะสังเกตได้ว่าค่าตัวสถิติจากข้อมูลในตารางที่ 2 (มีค่า 34.8) ตกอยู่ส่วนปลายของการแจกแจง (วงกลมทึบ) อันบ่งบอกว่า น่าจะปฏิเสธสมมติฐานว่างและยอมรับสมมติฐานทางเลือก

ศึกษาที่มีจำนวนกลุ่มตัวอย่างน้อย หรือความแตกต่างจริงมีค่าน้อย (เมื่อเทียบกับความแปรปรวนของข้อมูล) การสรุปที่เหมาะสมสำหรับผลการศึกษาที่ไม่พบความแตกต่างระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติ คือ “ไม่พบหลักฐานว่ามีความแตกต่างอย่างมีนัยสำคัญ” กล่าวคือ อาจมีความแตกต่างจริงในระดับประชากรก็ได้ แต่การศึกษานี้ไม่สามารถตรวจพบ (detect) หรือไม่ไวพอที่จะพบความแตกต่างดังกล่าว เช่นเดียวกับ การไม่ตรวจพบว่าเป็นโรค อาจมิได้หมายความว่าไม่เป็นโรคเสมอไป

ในปัจจุบันการคำนวณตัวสถิติ หรือการเปรียบเทียบกับค่าแจกแจงของตัวสถิติ ภายใต้สมมติฐานว่าง หรือการคำนวณค่า p หรือการทดสอบทางสถิติอื่นๆ ล้วนอาศัยโปรแกรมสถิติและเครื่องคอมพิวเตอร์เป็นตัวหลักในการคำนวณ จึงทำให้การวิเคราะห์ข้อมูลในปัจจุบันนั้นง่ายขึ้น แต่ก็เป็นตัวส่งเสริมให้มีการวิเคราะห์ข้อมูลที่ไม่เหมาะสมเกิดขึ้นอย่างมากมายเช่นกัน ผู้วิเคราะห์ข้อมูลจึงต้องมีความรู้ที่ถูกต้อง เข้าใจในหลักการทางสถิติที่สำคัญเป็นอย่างดี ก่อนจะวิเคราะห์ข้อมูลจริง มิเช่นนั้นก็ควรให้นักสถิติเป็นผู้วิเคราะห์ข้อมูลให้ แต่ผู้วิจัยก็ต้องเข้าใจหลักการวิเคราะห์ด้วย จึงจะหลีกเลี่ยงการถูกผลวิเคราะห์ชักจูงไปในทางที่มิชอบ

สรุป

บทความนี้ได้เสนอและแนะนำแนวความคิดพื้นฐานทางสถิติวิเคราะห์บางประการ ไม่ว่าจะเป็นเรื่องของประชากร กลุ่มตัวอย่าง ความแปรปรวนอย่างเป็นระบบและแบบสุ่ม และข้อมูลประเภทต่างๆ นอกจากนี้ ยังได้แนะนำการทดสอบทางสถิติ สมมติฐานทางสถิติ ตัวสถิติทดสอบ การแจกแจงของตัวสถิติ รวมไปถึงการนิยาม ค่า p และช่วงความเชื่อมั่นร้อยละ 95 โดยปิดท้ายด้วยตัวอย่างการทดสอบสถิติอย่างง่าย ผู้เขียนหวังว่าบทความนี้จะมีประโยชน์สำหรับผู้เริ่มทำงานวิจัยไม่มากนัก

เอกสารแนะนำและอ้างอิง

1. Glantz SA. Primer of biostatistics, 7th ed. New York: McGraw-Hill, 2012.
2. Amitage P, Berry G, Matthews JNS. Statistical methods in medical research, 4th ed. Oxford: Blackwell Science Ltd, 2002.

3. Lertsithichai P. Variability and randomness in medical research. Thai J Surg 2010;31:1-6.
4. Lertsithichai P. Principles of statistics for surgeons II: summary or descriptive statistics. Thai J Surg 2011;32:1-8.
5. Lertsithichai P. Principles of statistics for surgeons III: factors influencing sample size estimation. Thai J Surg 2012;33:25-32.