บทที่ 4

วิชาสถิติเบื้องต้นสำหรับนักระบาดวิทยาภาคสนาม

(Introduction to Statistics for Field Epidemiologists)

นายแพทย์ปณิธี ธัมมวิจยะ
แพทย์หญิงดารินทร์ อารีย์โชคชัย

หัวข้อสำคัญ

วิชาสถิติกับงานสาธารณสุขและระบาดวิทยา

- ความหมายของสถิติและชีวสถิติ
- ความสำคัญของวิชาสถิติในงานสาธารณสุขและระบาดวิทยา
- แนวคิดเรื่องประชากรกับกลุ่มตัวอย่าง

สถิติเชิงพรรณนา

- ความหมายและชนิดของตัวแปร
  - ตัวแปรจัดกลุ่ม (categorical variable)
  - ตัวแปรตัวเลข (numerical variable)
- การสรุปและนำเสนอข้อมูลด้วยตารางและรูป
- การสรุปด้วยตัวเลข
  - ค่ากลาง (measures of central location)
  - ค่าการกระจาย (measures of spreading)

สถิติเชิงอนุมาน

- ลักษณะและหลักการพื้นฐานของสถิติเชิงอนุมาน
- การทดสอบสมมติฐาน และ p-value
- การประมาณค่าและ confidence interval
- t-test
- Chi-squared test
- แนวคิดของการวิเคราะห์การถดถอย
- การวิเคราะห์การถดถอยเชิงเส้น (linear regression analysis)
- การวิเคราะห์การถดถอยโลจิสติก (logistic regression analysis)

วัตถุประสงค์

ภารกิจสำคัญอย่างหนึ่งของนักระบาดวิทยาภาคสนามคือ การตัดสินใจเลือก และดำเนินการควบคุมป้องกันโรคอย่างเหมาะสม และทันเวลาโดยอาศัยข้อมูลหลักฐานต่างๆ มาประกอบการพิจารณา ซึ่งเรียกกันว่า evidence-based prevention practice อย่างไรก็ตาม ข้อมูลส่วนใหญ่ที่นักระบาดวิทยาได้มาในตอนแรกมักจะอยู่ในสภาพที่ไม่เรียบร้อยยังไม่พร้อมใช้งาน ซึ่งบางคนเรียกว่าข้อมูลยังดิบอยู่ การที่จะทำให้ข้อมูลที่มีในมือ “สุก” หรือเรียบเรียงให้เป็นระบบระเบียบ เกิดความชัดเจน และสามารถนำมาใช้ประโยชน์ได้นั้น เป็นทักษะที่ต้องเรียนรู้ฝึกฝน

เนื้อหาในบทนี้มีเป้าหมายเพื่อแนะนำให้นักระบาดวิทยาภาคสนามรู้จักเครื่องมือสำคัญชิ้นหนึ่งที่ใช้ในการทำให้ข้อมูลเชิงปริมาณที่นักระบาดวิทยาได้รวบรวมมาสามารถนำมาใช้ประโยชน์ได้ นั่นก็คือ สถิติ (หรือชีวสถิติ) เนื้อหาในบทนี้จะครอบคลุมวิธีการวิเคราะห์ทางสถิติขั้นพื้นฐานที่สำคัญ และใช้บ่อยในทางการแพทย์และสาธารณสุข แม้ว่าในตำราสถิติทั่วไปนิยมที่จะแสดงที่มาหรือหลักการของวิธีการวิเคราะห์ทางสถิติรูปแบบต่างๆ ด้วยสัญลักษณ์หรือความสัมพันธ์ทางคณิตศาสตร์ เนื่องจากมีความถูกต้องในเชิงเนื้อหา กระชับและเข้าใจได้ง่ายสำหรับผู้ที่เป็นนักสถิติหรือผู้ที่มีโอกาสเรียนสถิติแบบจริงจัง แต่เนื่องจากวิธีการนำเสนอในลักษณะดังกล่าวอาจทำให้ผู้อ่านบางท่านที่ไม่ใช่นักสถิติ และไม่คุ้นเคยกับสัญลักษณ์หรือสมการทางคณิตศาสตร์เกิดความรู้สึกว่าสถิติเป็นวิชาที่เข้าถึงยากและไม่สนุก ประกอบกับผู้ที่ทำงานระบาดวิทยาภาคสนามมักมีความต้องการเรียนรู้เพื่อให้รู้จักเครื่องมือทางสถิติ เข้าใจแนวทางในการเลือกใช้และแปลผลได้อย่างถูกต้องเป็นหลัก มากกว่าที่จะเน้นความเข้าใจที่มาหรือทฤษฎีทางคณิตศาสตร์เชิงลึก เนื้อหาในบทนี้จึงมุ่งเน้นที่จะอธิบายเครื่องมือหรือวิธีการวิเคราะห์ทางสถิติขั้นพื้นฐานแต่ละชนิดว่ามีแนวคิดหรือหลักการทำงานอย่างไร เมื่อใดจะนำไปใช้ และจะแปลผลได้อย่างไร โดยอาศัยการบรรยายหลักการหรือแนวคิดที่สำคัญแทนการใช้สัญลักษณ์หรือสมการทางคณิตศาสตร์ ยกเว้นกรณีที่จำเป็น หากผู้อ่านสนใจที่จะศึกษารายละเอียดของวิธีการวิเคราะห์ (ต้องการลงมือทำด้วยตนเอง) หรือต้องการทราบหลักการทางคณิตศาสตร์ที่ลึกซึ้งขึ้น ก็สามารถค้นคว้าเพิ่มเติมได้จากตำราทางสถิติ รวมถึงเอกสารและแหล่งข้อมูลต่างๆ ซึ่งส่วนหนึ่งได้แนะนำไว้ท้ายบท

หลังจากศึกษาเนื้อหาในบทนี้แล้ว สิ่งที่ผู้อ่านควรได้รับ คือ

- เข้าใจความสำคัญของเครื่องมือทางสถิติที่มีต่องานระบาดวิทยา
- ทราบวิธีการนำเสนอข้อมูลเชิงพรรณนาอย่างเหมาะสมด้วยตาราง รูปภาพ และตัวเลข
- ทราบวิธีการเลือกและแปลผลการทดสอบทางสถิติที่พบบ่อย
- ทราบข้อจำกัดหรือข้อผิดพลาดที่พบบ่อยของการใช้เครื่องมือทางสถิติในงานระบาดวิทยา

วิชาสถิติกับงานสาธารณสุขและระบาดวิทยา

ความหมายของสถิติและชีวสถิติ

เมื่อพูดถึงคำว่า *สถิติ* โดยทั่วไปมักจะถูกใช้ใน 2 ความหมายซึ่งมีความเกี่ยวข้องกัน ความหมายแรกซึ่งคนส่วนใหญ่รู้จักและใช้กันทั่วไปคือ ค่าสถิติ (statistic) หมายถึงค่าที่ได้จากการรวบรวมข้อมูล (ซึ่งเรียกว่าข้อมูลดิบ) และนำมาทำการสรุปให้เป็นตัวเลขตัวเดียวหรือหลายตัว เพื่อบอกให้เราทราบถึงสถานการณ์หรือลักษณะด้านใดด้านหนึ่งของเรื่องนั้นๆ เช่น สถิติชีพ สถิติการเกิดอุบัติเหตุจราจร สถิติการบริโภคยาสูบ ฯลฯ

ส่วนในความหมายที่ 2 ซึ่งเป็นความหมายที่เรากำลังพูดถึงในเนื้อหาบทนี้คือ วิชาสถิติ (statistics) ในฐานะที่เป็นวิชาหรือศาสตร์สาขาหนึ่ง ซึ่งเป็นแขนงย่อยของวิชาคณิตศาสตร์ โดยมีเนื้อหาเกี่ยวข้องกับวิธีการหรือขั้นตอนในการรวบรวม แยกแยะ สรุปย่อ และวิเคราะห์ข้อมูล รวมถึงการนำผลที่ได้ไปสู่การอนุมาน (หรือการสร้างข้อสรุป) อย่างเป็นวิทยาศาสตร์ เราสามารถกล่าวได้ว่าสถิติในความหมายแรกเป็นผลผลิตของสถิติในความหมายที่ 2 ซึ่งอาจถือว่าเป็นกระบวนการผลิต

ส่วนคำว่า ชีวสถิติ (biostatistics) หมายถึง วิชาสถิติในส่วนที่เป็นการนำไปประยุกต์ใช้กับข้อมูลทางด้านชีวการแพทย์ ซึ่งรวมถึงข้อมูลด้านสาธารณสุข ตัวอย่างของการประยุกต์วิชาสถิติไปใช้กับสาขาอื่นๆ ได้แก่ ธรณีสถิติ (geostatistics) เศรษฐมิติ (econometrics) จิตมิติ (psychometrics) ฯลฯ วิชาเหล่านี้ล้วนมีรากฐานมาจากวิชาสถิติ และคณิตศาสตร์ทั้งสิ้น แต่มีความแตกต่างกันไปในรายละเอียด และเทคนิควิธีการเฉพาะด้านอันเป็นผลมาจากการที่แต่ละวิชามีจุดมุ่งหมาย และลักษณะของข้อมูลที่แตกต่างกัน เนื้อหาในบทนี้แม้จะเป็นเรื่องของชีวสถิติแต่จะขอใช้คำว่าวิชาสถิติเพื่อความสะดวก

วิชาสถิติสามารถแบ่งย่อยลงไปได้อีกเป็น 2 แขนง ได้แก่ สถิติเชิงพรรณา (descriptive statistics) และสถิติเชิงอนุมาน (inferential statistics) ซึ่งจะได้กล่าวในรายละเอียดต่อไป ผู้อ่านไม่ควรนำทั้ง 2 ชื่อนี้ไปสับสนกับชื่อที่ใช้เรียกรูปแบบการศึกษาทางระบาดวิทยาซึ่งแบ่งออกเป็น ระบาดเชิงพรรณนา (descriptive epidemiology) และระบาดวิทยาเชิงวิเคราะห์ (analytic epidemiology) ซึ่งมีรายละเอียดอยู่ในบทที่ 3

ความสำคัญของวิชาสถิติในงานสาธารณสุขและระบาดวิทยา

นักระบาดวิทยารวมถึงเจ้าหน้าที่สาธารณสุขเกือบทุกแขนงต้องทำงานอยู่กับข้อมูลด้านสุขภาพที่เป็นข้อมูลเชิงปริมาณ (หมายความว่าเป็นข้อมูลที่สามารถนับ วัด และสามารถนำมาแจกแจง หรือเปรียบเทียบด้วยวิธีการทางคณิตศาตร์ เช่น บวก ลบ คูณ หาร ฯลฯ ได้) และส่วนมากเป็นข้อมูลขนาดใหญ่ เนื่องจากลักษณะการทำงานด้านสาธารณสุขเกี่ยวข้องกับคนจำนวนมากหรือประชากร ตัวอย่างของข้อมูลเหล่านี้ ได้แก่ ข้อมูลการรักษาพยาบาลของผู้ป่วยในโรงพยาบาลซึ่งอาจเป็นสมุดเวชระเบียน หรือฐานข้อมูลอิเล็กทรอนิกส์ ข้อมูลการเฝ้าระวังทางระบาดวิทยา ฯลฯ ซึ่งมีความครบถ้วนถูกต้องแตกต่างกันไปขึ้นอยู่กับชนิดปัญหาสุขภาพและแหล่งข้อมูล ข้อมูลจากการศึกษาวิจัยที่ได้ออกแบบวางแผนมาอย่างดี หรือข้อมูลการสอบสวนการระบาดที่ต้องรีบรวบรวมภายในเวลาจำกัด ฯลฯ ข้อมูลเหล่านี้เมื่อได้รวบรวมมาในตอนแรกมักจะ “กอง” อยู่ในสภาพที่ไม่เป็นระเบียบ ยากต่อการเข้าใจ และมีความคลาดเคลื่อนปะปนอยู่ด้วยไม่มากก็น้อย ทั้งที่เป็นความคลาดเคลื่อนแบบสุ่มหรือความบังเอิญ (random error หรือ chance) ซึ่งเกิดจากการที่ทำการศึกษารวบรวมข้อมูลมาจากกลุ่มตัวอย่างแทนที่จะเป็นประชากรทั้งหมดที่ตั้งใจจะศึกษา และความคลาดเคลื่อนแบบเป็นระบบหรือความลำเอียง (systematic error หรือ bias) ซึ่งเกิดได้หลายสาเหตุ และพบได้ในทุกขั้นตอนของการศึกษา (ดูรายละเอียดในบทที่ 5) ผู้ที่จะนำข้อมูลเหล่านี้ไปใช้ประโยชน์จึงจำเป็นต้องเข้าใจธรรมชาติ หรือข้อจำกัดของตัวข้อมูล และสามารถเลือกใช้วิธีการที่ถูกต้องในการจัดการ วิเคราะห์ และแปลผลข้อมูลดังกล่าว สถิติหรือชีวสถิติจึงนับเป็นเครื่องมือสำคัญที่ใช้ในการวิเคราะห์ข้อมูลเชิงปริมาณดังกล่าว ซึ่งจะช่วยให้นักระบาดวิทยามีความเข้าใจในข้อมูลที่กำลังศึกษา และช่วยให้มองเห็นว่าความคลาดเคลื่อนที่เกิดขึ้นมีมากน้อยเพียงใด

แนวคิดเรื่องประชากรกับกลุ่มตัวอย่าง

กลุ่มเป้าหมายในการทำงานของนักระบาดวิทยา คือ “ประชากร” สิ่งที่เราอยากรู้ก็คือว่ามีอะไรเกิดขึ้นกับประชากรที่เราดูแลอยู่หรือให้ความสนใจ

โดยทั่วไปเมื่อพูดถึงคำว่าประชากร เช่น ฐานข้อมูลประชากรไทยของกระทรวงมหาดไทย พ.ศ. 2554 ประชากรแรงงานต่างด้าว หรือคำว่าประชากรในวิชาประชากรศาสตร์ ชีววิทยา นิเวศน์วิทยา ฯลฯ คำเหล่านี้มักจะหมายถึงคน (หรือสิ่งมีชีวิตชนิดอื่นๆ เช่น พืช สัตว์ ฯลฯ) ที่อยู่รวมกันในสถานที่ และช่วงเวลาหนึ่งตามที่ผู้ศึกษาของงานชิ้นนั้นๆ ให้ความสนใจ และได้ให้คำจำกัดความไว้

สำหรับในทางสถิติและการวิจัย คำว่าประชากร (population) มีความหมายพิเศษซึ่งแตกต่างออกไปจากความหมายข้างต้น ประชากรในทางสถิติ หมายถึง กลุ่ม (หรือการรวมกัน) ของสิ่งที่เราสนใจต้องการศึกษาหรืออยากได้ความรู้ โดยสิ่งที่เราสนใจนี้สามารถเป็นอะไรก็ได้ แม้ว่าในการศึกษาวิจัยทางการแพทย์และสาธารณสุขจะพบว่าประชากรคนเป็นสิ่งที่เรานิยมทำการศึกษา แต่ในทางสถิติแล้วประชากรที่ทำการศึกษาไม่จำเป็นที่จะต้องเป็นคน (หรือสิ่งมีชีวิต) แต่อาจจะเป็นวัตถุสิ่งของ สถานที่ หรือเหตุการณ์ใดๆ ก็ตามที่ผู้ศึกษาให้ความสนใจ ตัวอย่างของหน่วยประชากรแบบอื่นๆ ที่อาจพบ ได้แก่

- *ชุมชน* เช่น นักระบาดวิทยาคนหนึ่งต้องการศึกษาความสัมพันธ์ระหว่างปริมาณน้ำฝนกับการเกิดโรคไข้เลือดออกในรอบ 3 เดือน โดยอาจกำหนดให้ประชากรในการศึกษาแต่ละหน่วยเป็นตำบล เช่น ผู้ทำการศึกษาอาจทำการวัดปริมาณน้ำฝนในเวลา 3 เดือนของตำบล 100 แห่ง และนำไปหาความสัมพันธ์กับอัตราป่วยโรคไข้เลือดออกของแต่ละตำบลในช่วงเวลาเดียวกัน ซึ่งในการศึกษาลักษณะนี้อาจมีผู้เข้าใจว่าประชากรที่ศึกษาคือคนที่อยู่ใน 100 ตำบลดังกล่าวในช่วงเวลาที่ทำการศึกษา แต่ที่ถูกต้องแล้วในทางสถิติถือว่าประชากรในการศึกษานี้คือตำบล 100 แห่ง ไม่ใช่ตัวบุคคล
- *เหตุการณ์* ตัวอย่างเช่น การศึกษาผู้ป่วยที่มารักษาด้วยอุจจาระร่วงในรอบ 1 ปี ผู้ทำการศึกษาแต่ละคนอาจจะกำหนดประชากรที่ตัวเองสนใจไว้แตกต่างกัน เช่น นักระบาดวิทยาคนหนึ่งอาจจะศึกษาในลักษณะที่ประชากรเป็นคน คือ ดูว่ามีคนมารักษากี่คนใน 1 ปีด้วยโรคดังกล่าว และทำการวิเคราะห์ข้อมูลตามจำนวนคน แม้ว่าอาจมีคนที่ป่วยหรือมารักษามากกว่า 1 ครั้งในปีที่ศึกษา แต่ก็นับเป็นคนเดียวกัน (โดยที่อาจเก็บข้อมูลด้วยว่าแต่ละคนป่วยกี่ครั้ง) ในขณะที่นักระบาดวิทยาอีกคน อาจจะศึกษาในลักษณะที่ประชากรแต่ละหน่วยเป็นเหตุการณ์การป่วยก็ได้ คือ ดูว่ามีการป่วยเกิดขึ้นกี่ครั้งใน 1 ปี และทำการวิเคราะห์โดยอาศัยฐานข้อมูลจำนวนการป่วยเป็นตัวตั้ง ซึ่งจะทำให้จำนวนประชากรที่ศึกษามากกว่าจำนวนคนที่ป่วย เนื่องจากคนคนหนึ่งหากป่วย 2 ครั้งก็จะถูกนับเป็น 2 เหตุการณ์ หรือ 2 หน่วยประชากร

หน่วยของประชากรที่ศึกษาควรจะเป็นอะไรขึ้นอยู่กับคำถามของการศึกษาและความเป็นไปได้ในทางปฏิบัติของการเก็บรวบรวมข้อมูล การทำความเข้าใจให้ถูกต้องว่าหน่วยของประชากรที่ศึกษาคืออะไร มีความสำคัญมากต่อการเลือกวิธีทางสถิติที่จะใช้วิเคราะห์ข้อมูล การแปลผล และการนำผลการศึกษาไปใช้ประโยชน์

แม้ว่าประชากรจะเป็นเป้าหมายของการศึกษาหาความรู้ของนักระบาดวิทยาอย่างที่กล่าวแล้วข้างต้น แต่ในทางปฏิบัติแล้วการศึกษาโดยการเก็บข้อมูลจากประชากรทั้งหมดอาจจะเป็นไปไม่ได้ เป็นไปได้แต่ยากที่จะทำหรือไม่คุ้มค่าที่จะทำเนื่องจากค่าใช้จ่ายสูงหรือต้องใช้เวลามาก นักระบาดวิทยาสามารถที่จะเลือกเพียงบางส่วนของประชากรมาทำการศึกษา ซึ่งเรียกว่าการเลือกตัวอย่าง (sampling) โดยส่วนย่อยของประชากรที่ถูกเลือกมาเพื่อทำการศึกษา และเก็บข้อมูลนี้เรียกว่า ตัวอย่าง (sample) ซึ่งนอกจากจะมีความเป็นไปได้สูง ประหยัดใช้เวลาสั้นกว่าการเก็บข้อมูลจากประชากรทั้งหมดแล้ว ข้อดีอีกประการหนึ่งคือ ข้อมูลที่ได้จากกลุ่มตัวอย่างบางครั้งมีความถูกต้องแม่นยำมากกว่าการรวบรวมข้อมูลจากประชากร เนื่องจากกลุ่มตัวอย่างมีขนาดเล็กกว่า

ทำให้สามารถควบคุมคุณภาพของการเก็บข้อมูลได้ดีกว่าภายใต้ทรัพยากร (เวลา คน งบประมาณ ฯลฯ) ที่เท่ากัน เช่น การฝึกอบรมเจ้าหน้าที่ให้ทำการเก็บข้อมูลด้วยแบบสอบถาม หรือการทบทวนตรวจสอบความถูกต้องภายหลังจากที่ได้ข้อมูลมาแล้วย่อมทำได้ดีกว่า

กลุ่มตัวอย่างที่ได้รับการยอมรับว่าสามารถสะท้อนให้เห็นภาพของประชากรได้ดีหรือที่เรียกว่า มีความ*เป็นตัวแทน* (representativeness) ในทางสถิติ ควรมีลักษณะโดยทั่วไปอย่างน้อย 2 ประการ ได้แก่

1) *ขนาดตัวอย่าง* (sample size) ต้องเพียงพอ ซึ่งในการศึกษาแต่ละรูปแบบสามารถคำนวณได้ว่าต้องการกลุ่มตัวอย่างอย่างน้อยจำนวนเท่าไรจึงจะให้คำตอบที่ทำให้เรามีความเชื่อมั่นในระดับที่เราต้องการ

2) *เป็นกลุ่มตัวอย่างที่มาจากการเลือกโดยอาศัยกลไกการสุ่ม* (randomness) ซึ่งหมายถึง อาศัยความน่าจะเป็น (probability) เป็นกลไกในการกำหนดว่าตัวอย่างใดจะถูกเลือก เช่น อาจทำการเลือกโดยวิธีการเขียนชื่อให้กับประชากรทุกหน่วยลงในสลาก หรือกำหนดหมายเลขให้กับประชากรทุกหน่วยแล้วทำการเลือกกลุ่มตัวอย่างโดยการจับสลาก หรือเปิดตารางเลขสุ่ม (random number table) จนได้จำนวนตัวอย่างตามที่ต้องการ ซึ่งวิธีการข้างต้นนี้เรียกว่า *การเลือกตัวอย่างแบบสุ่มอย่างง่าย* (simple random sampling) หากทำการเลือกมาด้วยวิธีอื่นที่ไม่ได้อาศัยกลไกความน่าจะเป็นก็อาจจะได้ข้อมูลที่ไม่มีความเป็นตัวแทน (ในทางสถิติ) ของประชากร หรือหากอาศัยกลไกความน่าจะเป็นในการเลือกแต่ใช้รูปแบบที่มีความซับซ้อนก็จำเป็นต้องทำการวิเคราะห์ด้วยวิธีที่ซับซ้อนยุ่งยากกว่าเช่นกัน จึงจะสามารถใช้ผลการศึกษานั้นเป็นค่าประมาณของประชากรได้

การทดสอบทางสถิติทั้งหมดซึ่งจะได้กล่าวถึงต่อไปในบทนี้ มี*ข้อสมมติ* (assumption) หรือตั้งอยู่บนเงื่อนไขเบื้องต้นว่า ตัวอย่างที่ถูกเลือกมาศึกษาแต่ละหน่วยไม่เกี่ยวข้องกันในทางสถิติต้องมี*ความเป็นอิสระต่อกัน* (independence) ซึ่งมีความหมายว่าโอกาสที่ตัวอย่างหนึ่งๆ จะถูกเลือกมาทำการศึกษา ไม่ได้ขึ้นอยู่กับการที่ว่าตัวอย่างอื่นใดในการศึกษานั้นถูกเลือกหรือไม่ นั่นคือลักษณะหรือคุณสมบัติของตัวอย่างใดๆ ก็ตามไม่ได้มีความสัมพันธ์กับ (หรือส่งผลต่อ) โอกาสในการได้มาซึ่งตัวอย่างอื่นๆ ในการศึกษานั้น กลุ่มตัวอย่างที่มีความเป็นอิสระต่อกันนี้ก็ได้มาจากวิธีการเลือกโดยอาศัยกลไกการสุ่มเช่นเดียวกัน

อย่างไรก็ตาม วิธีการวิเคราะห์ทางสถิติต่างๆ ส่วนใหญ่แล้วจะมีสมมติฐานที่จำเป็นอย่างอื่นๆ อีก (นอกเหนือจาก independence หรือ random sampling) ซึ่งแตกต่างกันไปในแต่ละวิธี รายละเอียดเหล่านั้นอยู่นอกเหนือขอบเขตของเนื้อหาในบทนี้ ผู้ที่จะทำการวิเคราะห์และแปลผลข้อมูลจำเป็นต้องศึกษาเพิ่มเติมจากตำราทางสถิติ

สถิติเชิงพรรณนา: การสรุปย่อข้อมูลของกลุ่มตัวอย่าง

เมื่อได้รวบรวมข้อมูลของกลุ่มตัวอย่างที่ผ่านการตรวจสอบความถูกต้องมาแล้ว สิ่งแรกที่นักระบาดวิทยาควรทำ คือ การทำความเข้าใจข้อมูลที่มีอยู่ในมือเหล่านี้ สถิติเชิงพรรณนา (descriptive statistics) คือเครื่องมือที่ใช้ในการทำให้เห็นภาพลักษณะเชิงปริมาณของตัวอย่างที่มีอยู่ในมืออย่างเป็นระเบียบเรียบร้อยชัดเจนขึ้น วัตถุประสงค์หลักของสถิติเชิงพรรณนาก็คือ *การสรุปย่อ* (summary) ข้อมูลที่ยุ่งเหยิง และมีจำนวนมากให้กลายเป็นข้อมูลที่มีคุณภาพมากขึ้น (คือทำให้เห็นภาพของข้อมูลและสามารถใช้สื่อสารให้คนอื่นเข้าใจได้ง่าย) และมีประสิทธิภาพสูงขึ้น (แสดงข้อมูลด้วยตัวเลข ตาราง หรือรูปจำนวนไม่มากและใช้เวลาไม่นาน แต่สามารถสะท้อนให้เห็นภาพของฐานข้อมูลขนาดใหญ่ได้ในระดับที่เพียงพอต่อการใช้ประโยชน์)

ความหมายและชนิดของตัวแปร

ข้อมูลเชิงปริมาณที่รวบรวมมาจากกลุ่มตัวอย่างมักจะเก็บมาด้วยวิธีการวัดชนิดต่างๆ เช่น การสังเกต การตอบแบบสอบถาม การทดสอบทางห้องปฏิบัติการ ฯลฯ คุณสมบัติหรือลักษณะด้านใดด้านหนึ่งของแต่ละตัวอย่างที่ได้วัดหรือสังเกตมานี้เราเรียกว่า *ตัวแปร* (variable) เช่น อายุ เพศ น้ำหนัก ส่วนสูง ความดันโลหิต ฯลฯ

วิธีการแบ่งประเภทของตัวแปรมีหลายรูปแบบ ในที่นี้ขอเริ่มต้นด้วยการแบ่งออกเป็น 2 กลุ่มใหญ่ ได้แก่

1. ตัวแปรจัดกลุ่ม (categorical variable)

หมายถึง ตัวแปรแสดงคุณลักษณะเชิงคุณภาพ หรือแยกประเภท โดยที่ไม่สามารถระบุความแตกต่างของคุณสมบัติ หรือลักษณะนั้นๆ ออกมาเป็นตัวเลข หรือปริมาณที่แน่ชัดได้ เช่น เพศ การศึกษา ระยะของโรค สถานภาพทางสังคม ฯลฯ ซึ่งตัวแปรในกลุ่มนี้สามารถแบ่งได้อีกเป็น 2 ประเภทย่อย คือ

- **ตัวแปรนามบัญญัติ** (nominal variable) หมายถึง ตัวแปรจัดกลุ่มที่ไม่สามารถบอกระดับความมากน้อยของค่าแต่ละค่า (หรือไม่สามารถจัดอันดับได้) เช่น เพศ เชื้อชาติ การป่วย (ป่วย/ไม่ป่วย) ฯลฯ
- **ตัวแปรอันดับ** (ordinal variable) หมายถึง ตัวแปรจัดกลุ่มที่สามารถบอกได้ว่าค่าแต่ละค่ามีความแตกต่างกันในเชิงปริมาณ (ไล่เรียงมากน้อยได้) แต่ไม่สามารถระบุเป็นตัวเลขได้ชัดเจนว่าต่างกันเพียงใด (แม้ว่าอาจจะเขียนแสดงชื่อประเภทด้วยตัวเลขก็ตาม) โดยทั่วไปค่าของตัวแปรกลุ่มนี้มักจะมีตั้งแต่ 3 ระดับขึ้นไป เช่น ระดับสถานภาพทางสังคม (สูง-กลาง-ต่ำ) ระยะโรคมะเร็งรังไข่ (เกรด 1-4) ฯลฯ

2. ตัวแปรตัวเลข (numerical variable)

หมายถึง ตัวแปรที่แสดงคุณลักษณะเชิงปริมาณที่สามารถแสดงผลด้วยตัวเลขซึ่งแสดงความแตกต่างเชิงปริมาณที่แน่ชัดได้ โดยอาจได้มาจากการวัด เช่น น้ำหนัก ระดับน้ำตาลในเลือด อุณหภูมิ ฯลฯ หรือการนับได้แก่ อายุ จำนวนบุตร ฯลฯ ตัวแปรในกลุ่มนี้สามารถแบ่งได้อีกเป็น 2 ประเภทย่อย คือ

- **ตัวแปรช่วง** (interval variable) หมายถึง ตัวแปรตัวเลขที่ไม่มีค่าศูนย์ที่แท้จริง (คือค่าศูนย์ของตัวแปรไม่ได้หมายความว่าไม่มีสิ่งนั้นอยู่เลย) เช่น อุณหภูมิในหน่วยองศาเซลเซียส (0 องศาเซลเซียสก็ยังคงมีความร้อนอยู่ในระดับหนึ่ง) วันที่ตามปฏิทิน ฯลฯ ค่าของตัวแปรประเภทนี้สามารถนำมาเปรียบเทียบกันโดยการนำมาลบเพื่อดูผลต่างได้ แต่ไม่สามารถนำมาเปรียบเทียบกันโดยการหารเพื่อดูเป็นอัตราส่วนได้
- **ตัวแปรอัตราส่วน** (ratio variable) หมายถึง ตัวแปรตัวเลขที่มีค่าศูนย์ที่แท้จริง (คือการที่ค่าศูนย์ของตัวแปรมีความหมายว่า ปริมาณหรือระดับของลักษณะดังกล่าวเท่ากับศูนย์จริงๆ คือไม่มีปริมาณนั้นอยู่นั่นเอง) เช่น จำนวนบุตร (บุตร 0 คน หมายความว่าไม่มีบุตร) อุณหภูมิในหน่วยองศาเคลวิน (อุณหภูมิ 0 องศาเคลวิน หมายความว่าไม่มีปริมาณความร้อนอยู่) ฯลฯ ค่าของตัวแปรประเภทนี้สามารถนำมาเปรียบเทียบกันได้ทั้งแบบการนำมาลบเพื่อดูผลต่างและการหารเพื่อดูเป็นอัตราส่วน

ลักษณะสำคัญของตัวแปรแต่ละประเภทสรุปไว้ในตารางที่ 1 สำหรับเครื่องมือทางสถิติที่จะกล่าวถึงในบทนี้อาศัยเพียงการแบ่งตัวแปรออกเป็น 2 กลุ่มใหญ่ข้างต้น (คือ ตัวแปรจัดกลุ่ม และตัวแปรตัวเลข) ก็เพียงพอสำหรับการใช้งาน แต่ในการใช้เทคนิคการวิเคราะห์ทางสถิติระดับสูงขึ้นไป จำเป็นที่จะต้องเข้าใจการแบ่งประเภทของตัวแปรที่ละเอียดมากยิ่งขึ้น

ตารางที่ 1 ลักษณะตัวแปรประเภทต่างๆ

ประเภทของตัวแปร		ลักษณะ	ตัวอย่าง
ตัวแปรจัดกลุ่ม	ตัวแปรนามบัญญัติ	แยกประเภท แต่จัดอันดับไม่ได้	เพศ เชื้อชาติ
	ตัวแปรอันดับ	จัดอันดับมากน้อยได้	ระยะโรคมะเร็ง
ตัวแปรตัวเลข	ตัวแปรช่วง	บอกปริมาณความแตกต่างได้ แต่ไม่มีค่าศูนย์จริง	อุณหภูมิ (เซลเซียส) วันที่ตามปฏิทิน
	ตัวแปรอัตราส่วน	บอกปริมาณความต่างได้ และมีค่าศูนย์จริง	อุณหภูมิ (เคลวิน) จำนวนบุตร

การสรุปและนำเสนอข้อมูลด้วยตารางและรูป

หลังจากได้ข้อมูลมาแล้ว การใช้ตารางและรูปเป็นเครื่องมือสำคัญที่ใช้ในการสรุปย่อข้อมูล และนำเสนอข้อมูล โดยช่วยทำให้เห็นภาพของข้อมูลชัดเจนขึ้น เข้าใจง่าย ซึ่งเป็นที่นิยมในหมู่นักวิชาการทุกสาขา แต่หลายครั้งพบว่ายังมีการใช้ตารางและรูปในการนำเสนอข้อมูลอย่างไม่ถูกต้อง หรือใช้ไม่เต็มประสิทธิภาพ

ไม่มีกฎตายตัวว่าข้อมูลแบบใดควรนำเสนอด้วยตาราง แบบใดควรนำเสนอด้วยรูป คำแนะนำกว้างๆ ที่อาจใช้ประกอบการพิจารณาการเลือกรูปแบบในการนำเสนอได้แสดงไว้ในตารางที่ 2

ตารางที่ 2 ลักษณะเด่นของการนำเสนอข้อมูลด้วยตารางเปรียบเทียบกับการนำเสนอด้วยรูป

การนำเสนอด้วยตาราง	การนำเสนอด้วยรูป
▪ นำเสนอข้อมูลที่มีรายละเอียดซับซ้อนได้มากกว่า และแสดงข้อมูลได้แม่นยำกว่า ▪ จัดเตรียมได้ง่าย อาศัยทักษะน้อยกว่า ▪ ใช้พื้นที่น้อยกว่ารูปในการแสดงปริมาณข้อมูลที่เท่ากัน	▪ ดูแล้วเข้าใจง่าย สื่อความหมายได้ชัดเจนกว่าการนำเสนอด้วยตาราง ▪ จดจำได้ง่ายกว่า ▪ สามารถสื่อให้เห็นความสัมพันธ์ระหว่างตัวแปรได้ดีกว่าตาราง

มีคำแนะนำและเทคนิคในรายละเอียดมากมายที่เกี่ยวข้องกับการนำเสนอข้อมูลด้วยรูป และตาราง ผู้ที่สนใจสามารถศึกษาเพิ่มเติมได้จากเอกสารแนะนำท้ายบท ในที่นี้ขอสรุปหลักการสำคัญที่ถือเป็นหัวใจของการนำเสนอด้วยตารางและรูปมี 2 ประการ ดังนี้

1. ครบถ้วน (complete)

ตารางหรือรูปต้องให้ข้อมูลสำคัญครบถ้วน เพียงพอที่จะทำให้ผู้อ่านเข้าใจสิ่งที่กำลังนำเสนอ และสามารถแปลผลเบื้องต้นได้โดยไม่จำเป็นต้องกลับไปอ่านที่ตัวเนื้อหา เช่น

- ชื่อตาราง หรือรูปต้องระบุให้ชัดเจนว่านำเสนอข้อมูลอะไร ของใคร (กลุ่มตัวอย่าง หรือประชากรกลุ่มใด) และจำนวนที่แสดงมีทั้งหมดเท่าใด สถานที่ใด ช่วงเวลาใด และมีการแบ่งแยกประเภทด้วยตัวแปรใดบ้าง (ถ้ามี) รวมถึงแหล่งข้อมูลที่นำมาอ้างอิง ถ้าหากไม่ได้เป็นข้อมูลที่ผู้นำเสนอเป็นผู้รวบรวมมาด้วยตนเอง
- กรณีตาราง ระบุให้ชัดเจนว่าแต่ละแถว (row) แต่ละสดมภ์ (column) นำเสนอข้อมูลอะไร หน่วยเป็นอะไร
- กรณีรูป ระบุให้ชัดเจนว่าแต่ละแกน (กรณีกราฟชนิดต่างๆ) รูปร่าง สัญลักษณ์ สี หรือลวดลายต่างๆ ที่ใช้ นำเสนอข้อมูลอะไร หน่วยเป็นอะไร
- หากมีการใช้ตัวย่อหรือเครื่องหมายแทนข้อความ ต้องระบุให้ชัดเจนว่าหมายถึงอะไร

2. เรียบง่าย (simple)

ในแต่ละการนำเสนอตารางหรือรูป ข้อมูลที่ต้องการสื่อและรูปแบบการสื่อไม่ควรซับซ้อน หรืออาจทำให้เกิดความสับสนโดยไม่จำเป็น เช่น

- แต่ละตารางควรนำเสนอข้อมูลที่แสดงการกระจายของข้อมูลเพียง 1 หรือ 2 ตัวแปร (one-variable หรือ two-variable table) ในเวลาเดียวกัน หรืออย่างมากไม่ควรเกิน 3 ตัวแปร (three-variable table)
- ไม่ควรนำเสนอรูปภาพ เช่น กราฟ หรือแผนภูมิต่างๆ ในลักษณะรูป 3 มิติ
- ควรแสดงความแตกต่างของข้อมูลต่างๆ โดยอาศัยรูปร่างหรือลวดลายที่ต่างกัน และหลีกเลี่ยงการแสดงข้อมูลด้วยสีต่างๆ ซึ่งอาจเกิดปัญหาในกรณีที่มีการพิมพ์หรือถ่ายสำเนาเป็นภาพขาว-ดำ
- การนำเสนอข้อมูลด้วยกราฟชนิดต่างๆ จุดเริ่มต้นของแกนที่แสดงข้อมูลเชิงปริมาณ เช่น จำนวนผู้ป่วย หรือร้อยละของประชากร (ซึ่งมักจะเป็นแกนตั้ง) ควรตั้งต้นที่ค่าศูนย์ และไม่ควรใช้ scale break (คือมีการตัดค่าของแกนตั้งบางช่วงออกไป) นอกจากนี้ หากแกนที่แสดงปริมาณไม่ได้ใช้มาตราส่วนเลขคณิต (arithmetic scale) ซึ่งหมายความว่าแต่ละช่วงบนแกนแสดงความแตกต่างของปริมาณที่ต่างกัน ต้องระบุให้ชัดเจน เช่น ในกรณีที่ใช้มาตราส่วนลอการิทึม (logarithmic scale)

ต่อไปนี้เป็นตัวอย่างของการนำเสนอข้อมูลด้วยตาราง และรูปชนิดต่างๆ ที่พบได้บ่อย

ตาราง (table) ใช้สำหรับการนำเสนอข้อมูลเชิงปริมาณซึ่งมีอยู่ 2 ลักษณะ โดยลักษณะแรกเป็นการแสดงค่าต่างๆ ที่ได้จากการคำนวณด้วยวิธีการทางสถิติ (เช่น ค่าเฉลี่ย มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ช่วงความเชื่อมั่น ซึ่งจะได้กล่าวถึงต่อไป) และอีกลักษณะหนึ่งเป็นการแสดงการแจกแจงความถี่ (เช่น จำนวนหรือร้อยละของกลุ่มตัวอย่างที่มีลักษณะต่างๆ) ซึ่งมักจะแสดงด้วยตารางที่มีลักษณะดังต่อไปนี้

- ตาราง 1 ตัวแปร (one-variable table) คือตารางที่แสดงค่า หรือแจกแจงความถี่ครั้งละ 1 ตัวแปร ในเวลาเดียวกัน (ตารางที่ 3)

ตารางที่ 3 จำนวนและร้อยละผู้ป่วยโรคคอตีบแยกตามกลุ่มอายุ

กลุ่มอายุ	จำนวน	ร้อยละ	ร้อยละสะสม
0-5	14	23	23
6-11	24	40	63
12-20	13	22	85
มากกว่า 21	9	15	100
รวม	60	100	-

- ตาราง 2 ตัวแปร (two-variable table) ตารางที่แสดงค่า หรือแจกแจงความถี่ครั้งละ 2 ตัวแปร ในเวลาเดียวกัน (ตารางที่ 4)

ตารางที่ 4 จำนวนและร้อยละผู้ป่วยโรคคอตีบแยกตามกลุ่มอายุและเพศ

กลุ่มอายุ	จำนวนผู้ป่วย (ร้อยละ)	
	ชาย	หญิง
0-5	8 (27)	6 (20)
6-11	12 (40)	12 (40)
12-20	6 (20)	7 (23)
มากกว่า 21	4 (13)	5 (17)
รวม	30 (100)	30 (100)

- ตาราง 3 ตัวแปร (three-variable table) ตารางที่แสดงค่าหรือแจกแจงความถี่ครั้งละ 3 ตัวแปร ในเวลาเดียวกัน (ตารางที่ 5)

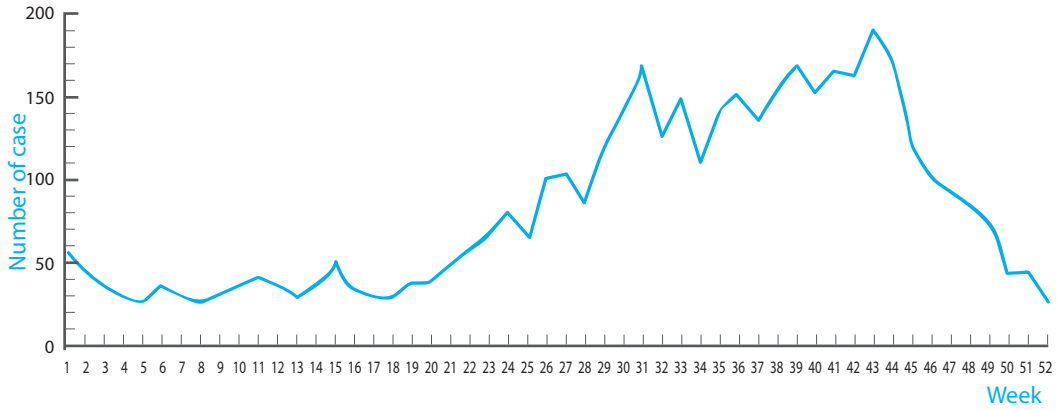
ตารางที่ 5 จำนวนผู้ป่วยโรคคอตีบแยกตามกลุ่มอายุ เพศ และเชื้อชาติ

กลุ่มอายุ	จำนวนผู้ป่วย			
	ชาย		หญิง	
	ไทย	ต่างชาติ	ไทย	ต่างชาติ
0-5	3	5	3	3
6-11	4	8	5	7
12-20	4	2	2	5
มากกว่า 21	1	3	2	3
รวม	12	18	12	18

รูป (figure) เราสามารถนำเสนอข้อมูลด้วยรูปได้หลายลักษณะ โดยที่พบบ่อยในงานด้านระบาดวิทยา ได้แก่

- กราฟ (graph) ใช้นำเสนอข้อมูลของตัวแปรตัวเลขที่เป็นค่าต่อเนื่อง เช่น เวลา วันเริ่มป่วย อายุ ฯลฯ

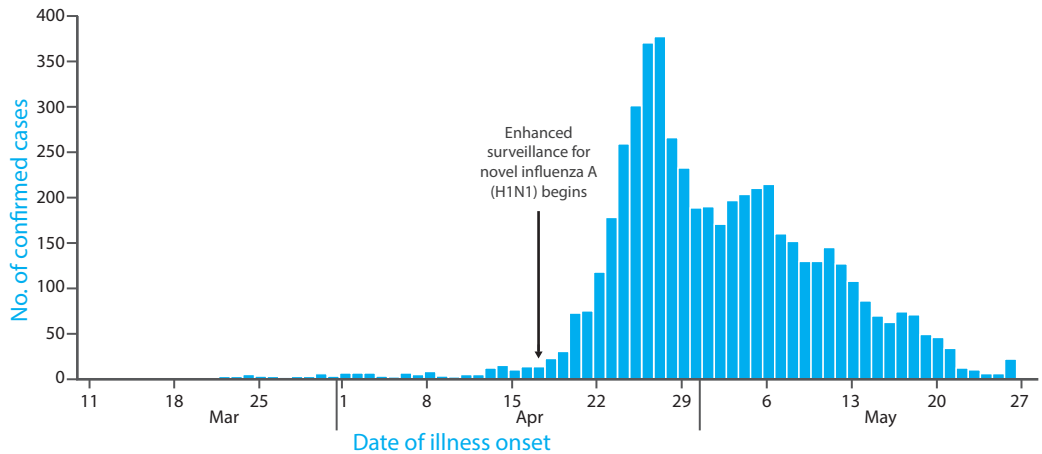
กราฟเส้น (line graph) เป็นการนำเสนอข้อมูลโดยใช้เส้นลากเชื่อมต่อจุดต่างๆ เพื่อแสดงให้เห็นถึงการเปลี่ยนแปลงขึ้นลง (fluctuation) หรือแนวโน้ม (trend) ของความถี่ (จำนวนหรืออัตรา) ตามช่วงเวลาหรืออายุที่เปลี่ยนแปลงไป (ซึ่งเป็นตัวแปรตัวเลขที่เป็นค่าต่อเนื่อง) โดยแกนนอนจะแสดงค่าของเวลา หรืออายุ ส่วนแกนตั้งแสดงความถี่ (รูปที่ 1)



รูปที่ 1 จำนวนผู้ป่วยโรคเลปโตสไปโรซิสแยกตามสัปดาห์เริ่มป่วย ประเทศไทย พ.ศ. 2554 (จำนวนผู้ป่วย 4,261 ราย)

ที่มา: สำนักระบาดวิทยา กรมควบคุมโรค

ฮิสโตแกรมหรือภาพแท่งความถี่ (histogram) จัดเป็นกราฟชนิดหนึ่งที่ใช้แสดงการแจกแจงความถี่ตัวแปรตัวเลขที่มีลักษณะเป็นค่าต่อเนื่อง เช่น อายุ น้ำหนัก วันเริ่มป่วย ฯลฯ โดยแกนนอนจะแสดงค่าของตัวแปรที่เป็นค่าต่อเนื่องซึ่งอาจถูกจัดกลุ่มออกเป็นช่วงต่างๆ ในขณะที่แกนตั้งจะแสดงความถี่ (จำนวนหรือร้อยละ) โดยความถี่ของค่าแต่ละค่า (หรือช่วงของค่าแต่ละช่วง) จะถูกแสดงด้วยความสูงของแท่งสี่เหลี่ยม เนื่องจากตัวแปรในแกนนอนเป็นค่าต่อเนื่อง ดังนั้นจึงไม่มีการเว้นช่องว่างระหว่างแท่งสี่เหลี่ยมที่แสดงความถี่ของข้อมูลที่มีค่าติดกัน ตัวอย่างของฮิสโตแกรมที่พบบ่อย และสำคัญที่สุดในงานระบาดวิทยาภาคสนาม คือ epidemic curve ซึ่งบางท่านเรียกเป็นภาษาไทยว่าเส้นโค้งการระบาด (รูปที่ 2)

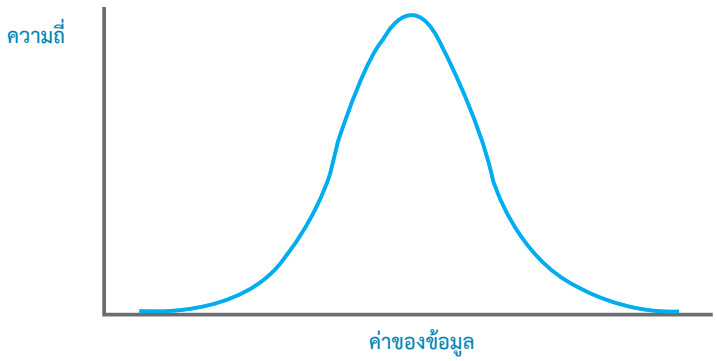


รูปที่ 2 จำนวนผู้ป่วยโรคไข้หวัดใหญ่ 2009 ที่มีผลยืนยันทางห้องปฏิบัติการแยกตามวันเริ่มป่วย ประเทศเม็กซิโก มีนาคม-พฤษภาคม 2552 (จำนวนผู้ป่วย 5,305 ราย)

ที่มา: The US Centers for Disease Control and Prevention

เมื่อเราทำการแจกแจงความถี่ของข้อมูลตัวแปรตัวเลขที่มีค่าต่อเนื่อง จะพบว่ามีรูปแบบการแจกแจงที่เป็นไปได้หลากหลาย แต่มีการแจกแจงในรูปแบบหนึ่งซึ่งมีความสำคัญ และมีประโยชน์มาก ซึ่งผู้ที่ใช้วิชาสถิติเป็นเครื่องมือทุกคนจำเป็นต้องรู้จักคือ *การแจกแจงปกติ* (normal distribution) ซึ่งมีรูปร่างคล้ายระฆังคว่ำ (รูปที่ 3) หากพบว่าข้อมูลที่เรารวบรวมมามีลักษณะใกล้เคียงหรือคล้ายกับรูปแบบการแจกแจงดังกล่าว เราก็สามารถนำคุณสมบัติของการแจกแจงดังกล่าวไปประยุกต์ใช้กับข้อมูลของเราได้ หนึ่งในคุณสมบัติที่สำคัญของการแจกแจงแบบปกติที่ต้องทราบในเบื้องต้นนี้คือ ระฆังต้องมีลักษณะสมมาตร ซึ่งหมายถึง

- ค่าของตัวแปรที่มีความถี่สูงสุด (หรือยอดระฆัง) อยู่ประมาณกึ่งกลาง และมีเพียงยอดเดียว (unimodal)
- ค่าเฉลี่ยเลขคณิตและค่ามัธยฐานมีค่าใกล้เคียงกัน
- ไม่มีข้อมูลที่มีค่าต่ำ หรือสูงผิดปกติมากๆ หรือฐานระฆังต้องไม่เบ้ไปข้างใดข้างหนึ่ง

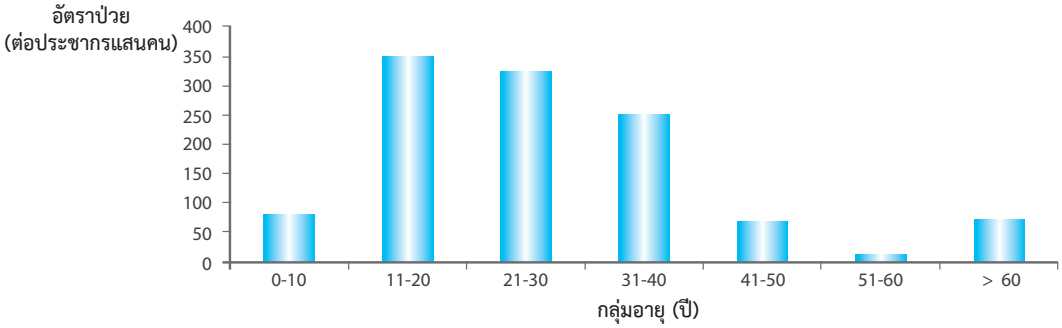


รูปที่ 3 การแจกแจงความถี่แบบปกติหรือรูประฆังคว่ำ

แผนภูมิ (chart) ใช้นำเสนอข้อมูลของตัวแปรจัดกลุ่ม (เช่น เชื้อชาติ อาชีพ) หรือตัวแปรตัวเลขที่ไม่ได้เป็นค่าต่อเนื่อง หรือไม่ต้องการเน้นที่ความต่อเนื่องของข้อมูล (เช่น กลุ่มอายุ)

แผนภูมิแท่ง (bar chart) ได้แก่

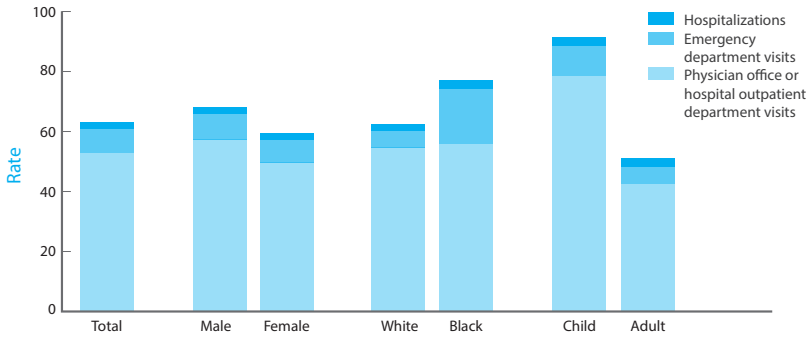
1) simple bar chart ใช้แสดงการแจกแจงความถี่ของข้อมูลครั้งละ 1 ตัวแปร (รูปที่ 4)



รูปที่ 4 อัตราป่วยโรคไวรัสตับอักเสบเอแยกตามกลุ่มอายุ ในการระบาดครั้งหนึ่งที่จังหวัดเชียงรายและลำปาง พ.ศ. 2548 (จำนวนผู้ป่วย 871 ราย)

ที่มา: สำนักระบาดวิทยา กรมควบคุมโรค

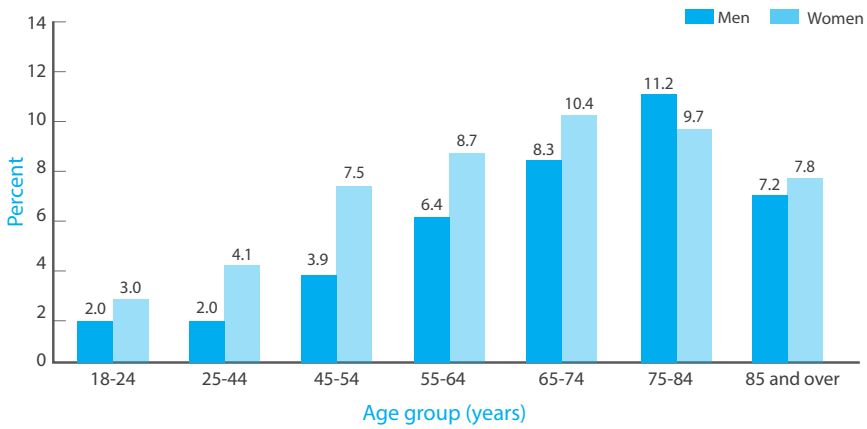
2) Compound bar chart (component or stacked bar chart) ใช้แสดงการแจกแจงความถี่ของข้อมูลครั้งละ 2 ตัวแปร โดยแต่ละแท่งของแผนภูมิแสดงความถี่ของตัวแปรที่ 1 (ตัวแปรหลัก) จากนั้นแบ่งแท่งความถี่ออกเป็นส่วนย่อยๆ ตามค่าตัวแปรที่ 2 (ตัวแปรรอง) โดยแสดงด้วยสีหรือลวดลายที่แตกต่างกัน (รูปที่ 5 ตัวแปรหลัก ได้แก่ เพศ เชื้อชาติ และกลุ่มอายุ ส่วนตัวแปรรองคือ ประเภทของการรับบริการสุขภาพ)



รูปที่ 5 อัตราการรับบริการที่สถานบริการสุขภาพต่อผู้ป่วยโรคหอบหืด 100 ราย จำแนกตามเพศ เชื้อชาติ และกลุ่มอายุ ประเทศสหรัฐอเมริกา ปี ค.ศ. 2001-2009

ที่มา: The US Centers for Disease Control and Prevention

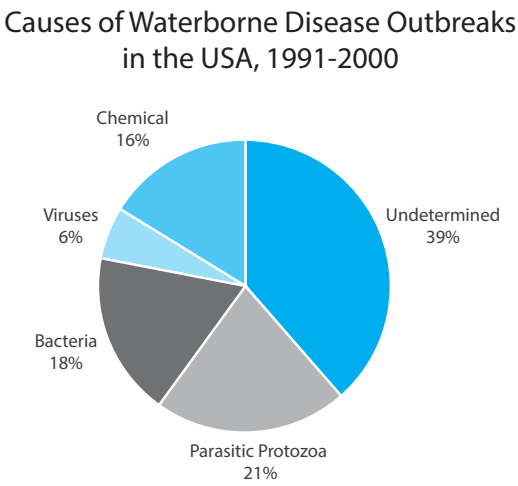
3) Multiple bar chart (grouped bar chart) ใช้แสดงการแจกแจงความถี่ของข้อมูลครั้งละ 2 ตัวแปร โดยแต่ละแท่งของแผนภูมิแสดงความถี่แยกย่อยตามค่าของตัวแปรที่ 1 (ตัวแปรหลัก) และค่าตัวแปรที่ 2 (ตัวแปรรอง) ในเวลาเดียวกัน อย่างไรก็ตาม แท่งที่แสดงค่าเดียวกันของตัวแปรหลักถูกวางเรียงไว้ด้วยกัน (รูปที่ 6 ตัวแปรหลักคือกลุ่มอายุ ส่วนตัวแปรรองคือเพศ)



รูปที่ 6 ร้อยละของความชุกโรคปอดอุดกั้นเรื้อรัง จำแนกตามเพศ และกลุ่มอายุในผู้ที่มีอายุตั้งแต่ 18 ปีขึ้นไป ประเทศสหรัฐอเมริกา ปี ค.ศ. 2007-2009

ที่มา: The US Centers for Disease Control and Prevention

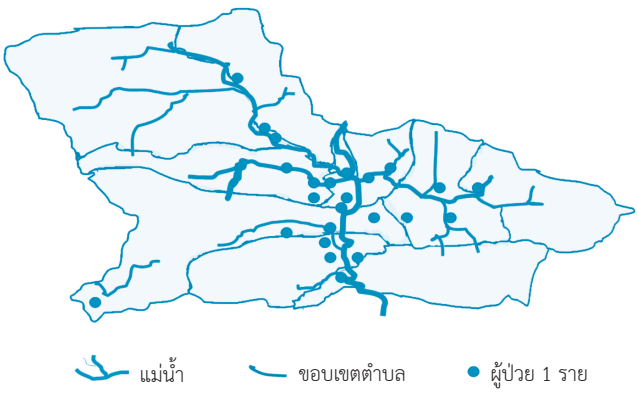
แผนภูมิวงกลม (pie chart) ใช้แสดงการแจกแจงความถี่ของข้อมูลครั้งละ 1 ตัวแปร โดยแต่ละค่าของตัวแปรควรแสดงค่าความถี่เป็นร้อยละกำกับไว้ด้วยเสมอ (รูปที่ 7)



รูปที่ 7 ร้อยละของสาเหตุของการระบาดของโรคที่มีน้ำเป็นสื่อ ประเทศสหรัฐอเมริกา ปี ค.ศ. 1991-2000
ที่มา: The American Chemistry Council

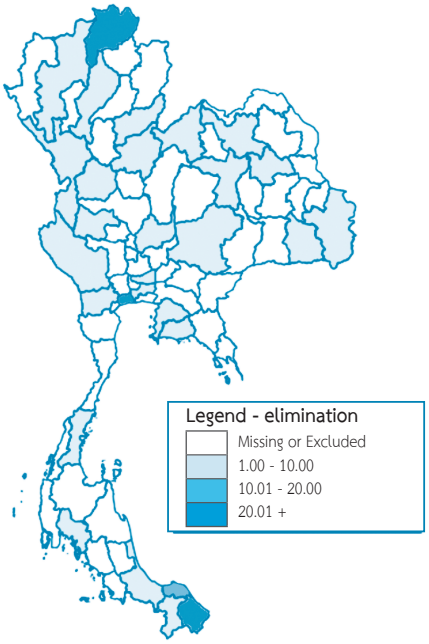
แผนที่ (map) ที่พบบ่อยในงานระบาดวิทยาภาคสนาม ได้แก่

1) แผนที่แสดงพิกัดทางภูมิศาสตร์ของสิ่งที่เราสนใจ โดยแสดงเป็น spotted map (รูปที่ 8)



รูปที่ 8 แผนที่แสดงที่อยู่ของผู้ป่วยโรค facial palsy ที่อำเภอแห่งหนึ่ง ประเทศไทย มกราคม-กันยายน 2542 (จำนวนผู้ป่วย 24 ราย)
ที่มา: สำนักระบาดวิทยา กรมควบคุมโรค

2) แผนที่แสดงปริมาณ ความถี่ หรือสัดส่วนของสิ่งที่เราสนใจในแต่ละพื้นที่ที่ทำการศึกษา (โดยไม่ได้มุ่งเน้นพิกัดที่แน่นอน) ซึ่งนิยมแสดงเป็น choropleth map (รูปที่ 9)



รูปที่ 9 แผนที่แสดงอัตราการรายงานผู้ป่วยกล้ามเนื้ออ่อนปวกเปียกเฉียบพลัน (acute flaccidparalysis) ในเด็กอายุต่ำกว่า 15 ปี (ต่อประชากรแสนคน) ประเทศไทย มกราคม-มีนาคม 2555
ที่มา: สำนักระบาดวิทยา กรมควบคุมโรค

การสรุปด้วยตัวเลข

ในกรณีที่เป็นตัวแปรแยกประเภทซึ่งไม่สามารถนำค่าของตัวแปรเหล่านี้มาบวกลบคูณหารได้อย่างมีความหมาย การสรุปข้อมูลของตัวแปรเหล่านี้จึงทำได้เพียงการแจกแจงความถี่โดยการแสดงเป็นจำนวนหรือร้อยละตามแต่ละค่าของตัวแปร และอาจนำผลที่ได้ไปใส่ไว้ในตาราง หรือทำเป็นแผนภูมิชนิดต่างๆ ดังที่ได้กล่าวแล้วข้างต้น

ส่วนในกรณีที่เป็นตัวแปรตัวเลขซึ่งค่าของตัวแปรมีความหมายทางคณิตศาสตร์นั้น เริ่มแรกให้นำข้อมูลมาแจกแจงความถี่เช่นเดียวกัน โดยแสดงการแจกแจงด้วยฮิสโตแกรม จากนั้นจึงค่อยทำการสรุปข้อมูลด้วยตัวเลข ซึ่งจะแสดงให้เห็นถึงคุณสมบัติของข้อมูลใน 2 ลักษณะด้วยกัน คือ ค่ากลางและการกระจายของค่าของตัวแปรนั้นๆ ในกลุ่มตัวอย่าง

ค่ากลาง (measures of central location)

หมายถึง ตัวเลขตัวหนึ่งที่ถูกเลือกมาเป็นตัวแทนเพื่อให้สะท้อนภาพของข้อมูลในกลุ่มตัวอย่างทั้งหมด โดยตัวเลขนี้ควรแสดงให้เห็นว่าในภาพรวมแล้วข้อมูลที่ได้เก็บมามีค่าอยู่ตรงไหน หรือระดับไหน ค่ากลางที่นิยมใช้มีหลายค่า ที่สำคัญ ได้แก่

ค่าเฉลี่ยเลขคณิต (arithmetic mean)

ซึ่งนิยมเรียกสั้นๆ ว่าค่าเฉลี่ย (mean หรือ average) คำนวณโดยการนำผลรวมของค่าตัวแปรของกลุ่มตัวอย่างทั้งหมดหารด้วยจำนวนตัวอย่าง หรือเขียนแสดงเป็นสมการดังนี้

$$mean = (X_1 + X_2 + X_3 + \ldots + X_n) / n$$

โดยที่ $X_1, X_2, X_3, \ldots, X_n$ คือ ค่าตัวแปรของตัวอย่างลำดับที่ 1, 2, 3,...,n ตามลำดับ

n คือ จำนวนตัวอย่างทั้งหมด

ค่าเฉลี่ย หรือที่นิยมเรียกอีกอย่างว่า $\bar{x}$ (อ่านว่า x-bar) ถือเป็นค่ากลางที่นิยมใช้มากที่สุด เนื่องจากมีคุณสมบัติทางคณิตศาสตร์ที่เหนือกว่าค่ากลางตัวอื่นๆ ทำให้สามารถนำไปใช้ประโยชน์ในการคำนวณค่าทางสถิติต่างๆ อย่างอื่นได้ นอกจากนี้ยังสอดคล้องกับสามัญสำนึก เนื่องจากเป็นค่ากลางที่ได้มาจากการใช้ค่าของข้อมูลทั้งหมดในกลุ่มตัวอย่างมาคำนวณ

ปัญหาที่สำคัญของค่าเฉลี่ยก็คือ หากมีข้อมูลที่ไม่ได้แจกแจงแบบปกติ แต่มีลักษณะเบ้ (skewed) คือมีตัวอย่างจำนวนน้อยที่มีค่าสูงผิดปกติแตกต่างจากกลุ่มมากๆ ซึ่งเรียกการกระจายลักษณะนี้ว่า เบ้ขวา (right skewed หรือ positive skewed) หรือมีตัวอย่างจำนวนน้อยที่มีค่าต่ำผิดปกติแตกต่างจากกลุ่มมากๆ ซึ่งเรียกว่าเบ้ซ้าย (left skewed หรือ negative skewed) ส่วนตัวอย่างที่มีค่าสูง หรือต่ำผิดปกติแตกต่างจากกลุ่มนั้นเรียกว่าเป็น outlier หากพบว่าข้อมูลมีการกระจายในลักษณะที่เบ้ก็จะส่งผลให้ค่าเฉลี่ยไม่เป็นตัวแทนที่ดีของข้อมูลส่วนใหญ่เนื่องจากในการคำนวณค่าเฉลี่ยจะถูกดึงให้ขยับเข้าไปใกล้ค่าที่สูงหรือต่ำผิดปกตินั้นมากขึ้น

ค่ามัธยฐาน (median)

เป็นค่ากลางที่ได้มาจากการเรียงลำดับตัวอย่างตามค่าของตัวแปร (จากต่ำสุดไปสูงสุด หรือกลับกันก็ได้) หลังจากนั้นก็นำเอาค่าของตัวแปรของตัวอย่างที่อยู่ ณ จุดกึ่งกลางของการเรียงลำดับมาเป็นค่ากลางของกลุ่มตัวอย่าง ในกรณีที่จำนวนตัวอย่างเป็นเลขคี่จะสามารถหาตัวอย่างที่อยู่ ณ จุดกึ่งกลางนั้นได้ แต่ในกรณีที่จำนวนตัวอย่างเป็นเลขคู่จะไม่มีตัวอย่างที่อยู่ ณ จุดกึ่งกลางพอดี ก็ให้นำค่าตัวแปรของ 2 ตัวอย่างที่อยู่ใกล้จุดกึ่งกลางมากที่สุดมาหาค่าเฉลี่ยแล้วนำมาเป็นค่ากลางของกลุ่มตัวอย่าง

จะเห็นได้ว่าค่ามัธยฐานมีข้อดีคือ ใช้ค่าของข้อมูลที่อยู่ ณ จุดกึ่งกลางมาเป็นตัวแทนของกลุ่มตัวอย่างโดยไม่ได้อาศัยค่าของข้อมูลอื่นๆ เลย จึงทำให้ไม่ได้รับผลกระทบจากข้อมูลที่สูงหรือต่ำผิดปกติมากๆ

จุดอ่อนของมัธยฐานคือ การนำไปใช้ประโยชน์ในการคำนวณทางสถิติขั้นที่สูงขึ้น เช่น การประมาณค่าในประชากร หรือการทดสอบหาความสัมพันธ์มีที่ใช้จำกัดกว่าค่าเฉลี่ย เนื่องจากไม่ได้เกิดจากการใช้วิธีการคำนวณทางคณิตศาสตร์โดยตรงกับข้อมูลของกลุ่มตัวอย่างทั้งหมด แต่เป็นการเลือกใช้เพียง 1 หรือ 2 ตัวอย่างโดยตัดค่าของตัวอย่างที่เหลือออกไป

ฐานนิยม (mode)

ฐานนิยมของข้อมูลกลุ่มตัวอย่างใดๆ หมายถึง ค่าของตัวแปรที่มีความถี่สูงสุด (หรือมีจำนวนตัวอย่างที่มีค่าซ้ำกันมากที่สุดนั่นเอง) ข้อดีของฐานนิยม คือ หาค่าได้ง่าย โดยเพียงดูการแจกแจงความถี่ก็สามารถบอกได้ทันที แต่ข้อด้อยของฐานนิยมก็คือ ข้อมูลบางชุดอาจไม่มีฐานนิยม (ไม่มีข้อมูลที่มีค่าซ้ำกันเลย) หรือมีฐานนิยมมากกว่า 1 ค่า (มีค่าที่ข้อมูลซ้ำกันด้วยความถี่เท่ากันมากกว่า 1 ค่า) ซึ่งจะเกิดปัญหาทันทีว่าควรจะรายงานค่าใดเป็นค่ากลาง นอกจากนี้ ยังมีข้อจำกัดในการใช้ประโยชน์ในการใช้งานทางสถิติขั้นที่สูงขึ้นเช่นเดียวกับมัธยฐาน

ค่าเฉลี่ยเรขาคณิต (geometric mean)

มีแนวคิดคล้ายคลึงกับค่าเฉลี่ยเลขคณิต แต่ใช้การคูณแทนการบวก คำนวณโดยการนำค่าตัวแปรของตัวอย่างทั้งหมดมาคูณกัน และนำผลคูณนี้มาถอดรากที่ n (n คือ จำนวนตัวอย่างทั้งหมด) หรือเขียนแสดงเป็นสมการดังนี้

geometric mean = $(X_1 * X_2 * X_3 * ... * X_n)^{1/n}$

โดยที่ X_1, X_2, X_3, ..., X_n คือ ค่าตัวแปรของตัวอย่างลำดับที่ 1, 2, 3,...,n ตามลำดับ

n คือ จำนวนตัวอย่างทั้งหมด

ค่าเฉลี่ยเรขาคณิตนี้ นิยมใช้ในกรณีที่เป็นข้อมูลที่มีลักษณะเบ้ขวา (มีข้อมูลจำนวนน้อยที่มีค่าสูงผิดปกติมากๆ) เช่น ข้อมูลผลตรวจทางห้องปฏิบัติการต่างๆ โดยเฉพาะกรณีที่รายงานผลเป็นระดับความเจือจางสูงสุดที่ยังตรวจพบ (titer) หรือระยะเวลาในการเกิดเหตุการณ์ที่เราสนใจศึกษา (ซึ่งอาจจะมีบางตัวอย่างที่เกิดเหตุการณ์ช้ากว่าคนอื่นมากๆ) ในกรณีเหล่านี้ค่าเฉลี่ยเรขาคณิตจะมีค่าใกล้เคียงกับค่ามัธยฐาน (เมื่อเปรียบเทียบกับค่าเฉลี่ยเลขคณิต) แต่มีข้อดีกว่าค่ามัธยฐานตรงที่อาศัยค่าของข้อมูลทุกตัวมาใช้ในการคำนวณ อย่างไรก็ตามค่าเฉลี่ยเรขาคณิตยังมีข้อจำกัดทางคณิตศาสตร์มากกว่าค่าเฉลี่ยเลขคณิต

ดูตัวอย่างการคำนวณค่ากลางชนิดต่างๆ ได้ในตารางที่ 6

ตารางที่ 6 ตัวอย่างการคำนวณค่ากลางของข้อมูลความสูงของนักเรียนจำนวน 10 คน

ข้อมูลความสูง (ซม.)	120, 135, 140, 125, 131, 142, 127, 132, 131, 122
ค่าเฉลี่ยเลขคณิต	= (120 + 135 + 140 + 125 + 131 + 142 + 127 + 132 + 131 + 122)/10 = 130.5
มัธยฐาน	เรียงข้อมูลจากค่าน้อยไปมาก: 120 122 125 127 131 131 132 138 140 142 จะได้ค่าเฉลี่ยของข้อมูล 2 ตัวที่อยู่ตรงกลาง (ลำดับที่ 5 และ 6) = (131+131)/2 = 131
ฐานนิยม	= 131 (เนื่องจากมีนักเรียน 2 คนที่สูง 131 เท่ากัน)
ค่าเฉลี่ยเรขาคณิต	= $(120 * 135 * 140 * 125 * 131 * 142 * 127 * 132 * 131 * 122)^{1/10}$ = 131.3

ค่าการกระจาย (measures of spreading)

หมายถึง ตัวเลขที่ใช้แสดงให้เห็นถึงการกระจาย หรือความหลากหลายของข้อมูลที่เรามี

พิสัย (range)

พิสัยของข้อมูล หมายถึง ความแตกต่างระหว่างค่าสูงสุด (maximum) กับค่าต่ำสุด (minimum) ของข้อมูลชุดนั้น วิธีการคำนวณพิสัยก็เพียงแต่นำค่าสูงสุดตั้งแล้วลบด้วยค่าต่ำสุด ในทางสถิติแล้วจะรายงานตัวเลขซึ่งเป็นความแตกต่างนี้เพียงตัวเลขเดียวเป็นค่าพิสัย แต่ในงานด้านระบาดวิทยาหรือการวิจัยสาขาอื่นๆ เวลาพูดถึงพิสัยนิยมที่จะรายงานโดยบอกเป็นค่าสูงสุดกับค่าต่ำสุด (ตัวเลข 2 ค่า) แทน เช่น รายงานเป็น "...(ค่าต่ำสุด)...ถึง...(ค่าสูงสุด)..."

พิสัยมีประโยชน์ในการช่วยแสดงให้เห็นภาพว่าข้อมูลมีการกระจายไปไกลมากน้อยเพียงใด ซึ่งมีประโยชน์มากในทางระบาดวิทยา เช่น ในรายงานการระบาดของโรคส่วนใหญ่แล้วจำเป็นจะต้องแสดงให้เห็นว่าผู้ป่วยที่พบมีอายุต่ำสุดและสูงสุดเท่าใด อย่างไรก็ตามพิสัยมีข้อจำกัด คือ อาจไม่ได้แสดงให้เห็นว่าข้อมูลส่วนใหญ่ (หรือโดยภาพรวม) มีการกระจายอย่างไรหากเป็นกรณีที่ข้อมูลไม่ได้แจกแจงแบบปกติหรือมีลักษณะเบ้นั่นเอง

พิสัยระหว่างควอร์ไทล์ (interquartile range)

เนื่องจากพิสัยแบบที่ได้กล่าวถึงข้างต้นมีข้อจำกัดในกรณีที่ข้อมูลแจกแจงไม่ปกติ นักสถิติจึงนิยมใช้ค่าพิสัยอีกรูปแบบหนึ่งในการบอกการกระจาย ซึ่งเรียกว่า พิสัยระหว่างควอร์ไทล์ แต่ก่อนที่จะพูดถึงพิสัยในรูปแบบนี้ มีคำที่นักระบาดวิทยาจำเป็นต้องรู้จักเป็นพื้นฐานก่อน 2 คำ ได้แก่

1) เปอร์เซ็นต์ไทล์ (percentile) หมายถึง ตัวเลขบอกค่าของตัวแปรที่อยู่ในตำแหน่งต่างๆ

เมื่อตัวอย่างที่เราศึกษาถูกเรียงลำดับตามค่าของตัวแปรจากน้อยไปมาก และกลุ่มตัวอย่างนั้นๆ ถูกแบ่งออกเป็น 100 ส่วนเท่าๆ กัน โดยเปอร์เซ็นต์ไทล์ที่ P (P^{th} percentile) หมายถึง ค่าของตัวแปรที่มีจำนวนตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับค่าดังกล่าวร้อยละ P เช่น

เปอร์เซ็นต์ไทล์ที่ 10 (10^{th} percentile) หมายถึง ค่าของตัวแปรที่มีจำนวนตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับค่าดังกล่าวอยู่ร้อยละ 10

เปอร์เซ็นต์ไทล์ที่ 50 (50^{th} percentile) หมายถึง ค่าของตัวแปรที่มีจำนวนตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับค่าดังกล่าวอยู่ร้อยละ 50 ซึ่งก็คือค่ามัธยฐาน (median) นั่นเอง

เปอร์เซ็นต์ไทล์ที่ 100 (100^{th} percentile) หมายถึง ค่าของตัวแปรที่มีจำนวนตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับค่าดังกล่าวอยู่ร้อยละ 100 ซึ่งก็คือค่าสูงสุด (maximum) นั่นเอง

2) ควอร์ไทล์ (quartile) หมายถึง ตัวเลขบอกค่าของตัวแปรที่อยู่ในตำแหน่งต่างๆ เมื่อตัวอย่างที่เราศึกษาถูกเรียงลำดับตามค่าของตัวแปรจากน้อยไปมาก และกลุ่มตัวอย่างนั้นๆ ถูกแบ่งออกเป็น 4 ส่วนเท่าๆ กัน โดยที่

ควอร์ไทล์ที่ 1 (1^{st} quartile หรือ Q_1) หมายถึง ค่าของตัวแปรที่มีจำนวนตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับค่าดังกล่าว 1 ใน 4 หรือร้อยละ 25 ซึ่งก็คือเปอร์เซ็นต์ไทล์ที่ 25 นั่นเอง

ควอร์ไทล์ที่ 2 (2^{nd} quartile หรือ Q_2) หมายถึง ค่าของตัวแปรที่มีจำนวนตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับค่าดังกล่าว 2 ใน 4 หรือร้อยละ 50 ซึ่งก็คือเปอร์เซ็นต์ไทล์ที่ 50 หรือมัธยฐานนั่นเอง

ควอร์ไทล์ที่ 3 (3^{rd} quartile หรือ Q_3) หมายถึง ค่าของตัวแปรที่มีจำนวนตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับค่าดังกล่าว 3 ใน 4 หรือร้อยละ 75 ซึ่งก็คือเปอร์เซ็นต์ไทล์ที่ 75

ควอร์ไทล์ที่ 4 (4^{th} quartile หรือ Q_4) หมายถึง ค่าของตัวแปรที่มีจำนวนตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับค่าดังกล่าว 4 ใน 4 หรือร้อยละ 100 ซึ่งก็คือเปอร์เซ็นต์ไทล์ที่ 100 หรือค่าสูงสุดนั่นเอง

สำหรับค่า*พิสัยระหว่างควอร์ไทล์* (interquartile range: IQR) หมายถึง พิสัยของข้อมูลร้อยละ 50 ที่อยู่ตรงกลาง หรือพูดอีกแบบหนึ่งก็คือว่า ข้อมูลร้อยละ 50 ที่อยู่ตรงกลางมีการกระจายตัวมากน้อยเพียงใด โดยวิธีการคำนวณก็ได้มาจากการนำค่าควอร์ไทล์ที่ 3 (หรือเปอร์เซ็นต์ไทล์ที่ 75) ลบด้วย ควอร์ไทล์ที่ 1 (หรือเปอร์เซ็นต์ไทล์ที่ 25) ซึ่งเขียนแสดงด้วยสมการได้ดังนี้

$$IQR = Q_3 - Q_1$$

สำหรับวิธีการคำนวณ เนื่องจากในทางปฏิบัติ โปรแกรมทางสถิติส่วนใหญ่จะแสดงค่าควอร์ไทล์ต่างๆ ให้อยู่แล้ว การคำนวณค่าควอร์ไทล์ด้วยมือจึงไม่ได้นำมากล่าวถึงในเนื้อหาบทนี้ สิ่งซึ่งนักระบาดวิทยาต้องทำจึงเหลือแต่เพียงการคำนวณค่าพิสัยระหว่างควอร์ไทล์

เช่นเดียวกับการรายงานค่าพิสัยที่ได้กล่าวแล้วข้างต้น การรายงานค่าพิสัยระหว่างควอร์ไทล์ ก็สามารถรายงานแบบที่นักสถิตินิยม (คือรายงานผลต่างระหว่าง Q_3 กับ Q_1 เป็นตัวเลขจำนวนเดียว) หรือแบบที่นักระบาดวิทยาหรือนักวิจัยทั่วไปนิยม (คือรายงานทั้งค่า Q_3 และ Q_1 เป็นตัวเลขสองจำนวน) ก็ได้

ข้อดีของการใช้พิสัยระหว่างควอร์ไทล์เป็นตัววัดการกระจายของข้อมูลก็ คือ ค่าที่ได้จะไม่ได้รับผลกระทบจากการที่ข้อมูลมีลักษณะเบ้ จึงมักจะพบว่าถูกรายงานควบคู่กับค่ามัธยฐาน อย่างไรก็ตามต้องระลึกเสมอว่าพิสัยที่ได้เป็นของข้อมูลร้อยละ 50 ที่อยู่ตรงกลาง จึงไม่ได้แสดงการกระจายของข้อมูลทั้งหมดในกลุ่มตัวอย่าง

มีนักสถิติบางท่านแนะนำว่า หากข้อมูลเบ้ก็อาจรายงานทั้งมัธยฐาน พิสัยระหว่างควอร์ไทล์ และพิสัยร่วมกัน ซึ่งเรียกว่าการสรุปด้วย 5 ตัวเลข (five-numbers summary) ซึ่งจะทำให้เห็นภาพข้อมูลที่ชัดเจน แต่ควรเลือกใช้รายงานเฉพาะกับข้อมูลหรือตัวแปรที่มีความสำคัญ

ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

ส่วนเบี่ยงเบนมาตรฐานเป็นค่าที่ใช้วัดการกระจายที่ใช้บ่อยที่สุดในข้อมูลที่มีการแจกแจงแบบปกติ (หรือใกล้เคียง) จึงมักพบรายงานคู่กับค่าเฉลี่ยอยู่เสมอ

แนวคิดเริ่มต้นในการหาค่าที่ช่วยให้เราเห็นการกระจายของข้อมูลก็คือ การนำค่าของตัวอย่างแต่ละตัวมาลบออกด้วยค่าเฉลี่ย ซึ่งพูดอีกแบบก็คือการหาระยะห่าง หรือผลต่างระหว่างข้อมูลของแต่ละตัวอย่างกับค่าเฉลี่ยนั่นเอง ข้อมูลใดที่มีค่าสูงกว่าค่าเฉลี่ยก็จะมีผลต่างเป็นค่าบวก ส่วนข้อมูลใดที่มีค่าต่ำกว่าค่าเฉลี่ย ก็จะมีผลต่างเป็นค่าลบ ถ้านำผลต่างเหล่านี้ทั้งหมดมารวมกันในทันที ค่าที่ได้ย่อมเป็นศูนย์เสมอ ซึ่งไม่มีประโยชน์ในการบอกการกระจายของข้อมูล วิธีในการจัดการกับปัญหานี้ในอดีตมีทางเลือกที่สำคัญอยู่ 2 แบบ แบบที่ 1 คือ การเอาผลต่างมารวมกันโดยไม่สนใจเครื่องหมายลบ (ซึ่งก็คือการใช้ค่าสัมบูรณ์ หรือ absolute value นั่นเอง) ส่วนแบบที่ 2 คือการเอาผลต่างเหล่านี้มายกกำลังด้วยจำนวนคู่อะไรก็ได้ (ยกเว้นศูนย์) ซึ่งก็ทำให้ได้ค่าที่เป็นบวกทั้งหมดเช่นกันแล้วจึงนำมารวมกัน แต่ในทางปฏิบัติก็ใช้การยกกำลัง 2 (square) เพราะซับซ้อนน้อยที่สุด

เนื่องจากในอดีตการคำนวณค่าเหล่านี้ล้วนต้องทำด้วยมือทั้งสิ้น ทางเลือกแบบที่ 1 (ค่าสัมบูรณ์) จึงไม่ได้รับความนิยมเนื่องจากสามารถเขียนแสดงความสัมพันธ์ทางคณิตศาสตร์ได้ยากกว่าเมื่อเปรียบเทียบทางเลือกแบบที่ 2 (ยกกำลังสอง) นอกจากนี้การคำนวณในลักษณะนี้ยังสามารถนำมาใช้ดูว่าข้อมูลมีการกระจายตัวอย่างไร (ดังจะได้กล่าวถึงต่อไปในตอนท้ายของหัวข้อส่วนเบี่ยงเบนมาตรฐาน)

เมื่อเราได้ค่าผลรวมของค่ายกกำลังสองของผลต่าง หรือระยะห่างระหว่างค่าของแต่ละตัวอย่างกับค่าเฉลี่ยแล้ว เราก็นำค่าผลรวมดังกล่าวมาหารด้วยจำนวนตัวอย่างทั้งหมดอีกครั้ง เพื่อให้ได้เป็นค่าเฉลี่ยของค่ายกกำลังสอง ของผลต่าง ซึ่งถูกเรียกว่า *ความแปรปรวน* (variance)

โดยหลักการ เราสามารถเขียนแสดงการคำนวณค่าความแปรปรวนเป็นสมการได้ดังนี้

$$\text{variance} = [(X_1\text{-mean})^2 + (X_2\text{-mean})^2 + (X_3\text{-mean})^2 + \ldots + X_n\text{-mean})^2)] / n$$

โดยที่ $X_1, X_2, X_3, \ldots, X_n$ คือ ค่าตัวแปรของตัวอย่างลำดับที่ 1, 2, 3,...,n ตามลำดับ

mean คือ ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่าง

n คือ จำนวนตัวอย่างทั้งหมด

ในทางทฤษฎี หากเราสามารถเก็บข้อมูลได้จากประชากรทั้งหมดที่เราสนใจศึกษา เราก็สามารถใช้สูตรข้างต้นในการหาความแปรปรวนในประชากรได้ทันที อย่างไรก็ตาม ในทางปฏิบัติข้อมูลที่เราได้มาจากกลุ่มตัวอย่างที่เราเลือกมาจากประชากรที่เราสนใจ หากเราถือว่ากลุ่มตัวอย่างนั้นเป็นตัวแทนที่ดีของประชากรได้ เราก็สามารถใช้ข้อมูลจากกลุ่มตัวอย่างไปคำนวณหาค่าความแปรปรวนที่มีอยู่ในประชากรนั้นได้เช่นกัน โดยในทางสถิติสามารถพิสูจน์ได้ว่าค่าความแปรปรวนที่คำนวณได้จากข้อมูลของกลุ่มตัวอย่างด้วยสมการข้างต้นจะได้ค่าที่ต่ำกว่า (underestimate) ค่าความแปรปรวนจริงในประชากรเสมอ ดังนั้น วิธีการที่ใช้ในการคำนวณค่าความแปรปรวน

ของกลุ่มตัวอย่างเพื่อให้ได้ค่าที่แสดงความแปรปรวนของประชากรที่เราสนใจศึกษา จึงมีการปรับเปลี่ยนเล็กน้อยจากสมการข้างต้น ดังนี้

$$\text{variance} = [(X_1\text{-mean})^2 + (X_2\text{-mean})^2 + (X_3\text{-mean})^2 + \ldots + (X_n\text{-mean})^2)] / (n-1)$$

โดยที่ $X_1, X_2, X_3, \ldots, X_n$ คือ ค่าตัวแปรของตัวอย่างลำดับที่ 1, 2, 3,...,n ตามลำดับ
mean คือ ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่าง
n คือ จำนวนตัวอย่างทั้งหมด

จะสังเกตได้ว่าตัวหารในสมการนี้เปลี่ยนจาก n เป็น n–1 ซึ่งทำให้ผลลัพธ์ที่ได้มีค่าสูงขึ้นเนื่องจากตัวหารมีค่าน้อยลง โดยในทางสถิติสามารถพิสูจน์ให้เห็นได้ว่าการหารด้วยค่า n-1 นี้ ทำให้ค่าความแปรปรวนที่ได้จากสมการใหม่นี้เป็น*ค่าประมาณที่ไม่ลำเอียง* (unbiased estimate) ของค่าความแปรปรวนในประชากร (ค่า n-1 ในกรณีนี้ ทางสถิติเรียกว่า degree of freedom) วิธีการคำนวณความแปรปรวนแบบที่ 2 นี้ จึงเป็นวิธีการคำนวณที่ใช้ทั่วไปในทางปฏิบัติเนื่องจากถือได้ว่าเราไม่เคยได้ข้อมูลที่ครบถ้วนจากประชากรทั้งหมดที่เราสนใจ

อย่างไรก็ตาม เนื่องจากความแปรปรวนเกิดจากการนำค่าผลต่างไปยกกำลังสอง ดังนั้นหน่วยของความแปรปรวนจึงเป็นยกกำลังสองของหน่วยดั้งเดิมของข้อมูล เช่น หากเป็นข้อมูลอายุซึ่งหน่วยดั้งเดิมเป็นปี ความแปรปรวนก็จะมีหน่วยเป็นปียกกำลังสอง (ปี^{2}) ซึ่งในทางปฏิบัติจะเห็นได้ว่าความแปรปรวนนั้นแปลความหมายได้ยาก ทั้งในส่วนที่เป็นตัวเลขค่าความแปรปรวนนั้นเอง และในส่วนของหน่วยของความแปรปรวน ด้วยเหตุผลดังกล่าว นักสถิติจึงได้นำค่าความแปรปรวนมาถอดรากที่สอง (square root) เพื่อให้ได้ค่าที่มีหน่วยเหมือนกับหน่วยดั้งเดิมของข้อมูล ซึ่งค่าดังกล่าวถูกเรียกว่า *ส่วนเบี่ยงเบนมาตรฐาน* (standard deviation: SD) ซึ่งสามารถเขียนแสดงเป็นสมการได้ดังนี้

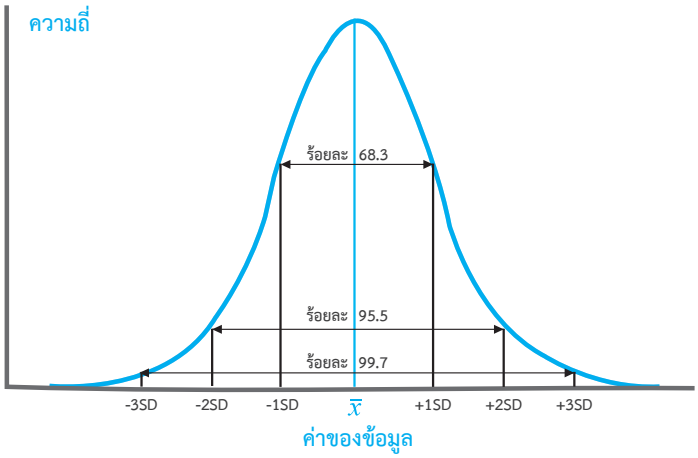
$$\text{SD} = (\text{variance})^{1/2}$$
$$= \{[(X_1\text{-mean})^2 + (X_2\text{-mean})^2 + (X_3\text{-mean})^2 + \ldots + (X_n\text{-mean})^2)] / (n-1)\}^{1/2}$$

โดยที่ $X_1, X_2, X_3, \ldots, X_n$ คือ ค่าตัวแปรของตัวอย่างลำดับที่ 1, 2, 3,...,n ตามลำดับ
mean คือ ค่าเฉลี่ยเลขคณิตของกลุ่มตัวอย่าง
n คือ จำนวนตัวอย่างทั้งหมด

ในส่วนของการแปลผล แม้ว่าส่วนเบี่ยงเบนมาตรฐานจะมีหน่วยเช่นเดียวกับหน่วยของข้อมูล แต่หากพิจารณาในทางสถิติแล้ว เราไม่อาจกล่าวได้ว่าส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับค่าเฉลี่ยของระยะห่างระหว่างข้อมูลแต่ละตัวอย่างกับค่าเฉลี่ยของกลุ่มตัวอย่าง (ซึ่งได้มาจากการหาค่าเฉลี่ยของค่าสัมบูรณ์ของระยะห่าง) เนื่องจากมีวิธีการคำนวณที่แตกต่างกัน ในทางปฏิบัติเราจึงนิยมใช้ค่าส่วนเบี่ยงเบนมาตรฐานสำหรับการเปรียบเทียบระหว่างข้อมูลตั้งแต่ 2 ชุดหรือ 2 กลุ่มขึ้นไป ว่ากลุ่มใดการกระจายมากกว่าหรือน้อยกว่าเมื่อเปรียบเทียบกับกลุ่มอื่นๆ เท่านั้น

อย่างไรก็ตาม ส่วนเบี่ยงเบนมาตรฐานไม่ได้มีประโยชน์แค่เพียงบอกการกระจายของข้อมูลในภาพรวมเท่านั้น ในกรณีที่ข้อมูลมีการแจกแจงเป็นปกติ ค่าดังกล่าวยังมีประโยชน์ในการบอกรายละเอียดว่าข้อมูลมีการกระจายตัวอย่างไรในแต่ละช่วงของค่าส่วนเบี่ยงเบนมาตรฐาน (รูปที่ 10) โดยเราสามารถระบุได้ว่าหากข้อมูลมีการแจกแจงปกติ

- ตัวอย่างประมาณร้อยละ 68.3 จะมีค่าอยู่ระหว่าง ค่าเฉลี่ย ± 1 SD
- ตัวอย่างประมาณร้อยละ 95 จะมีค่าอยู่ระหว่าง ค่าเฉลี่ย ± 1.96 SD
- ตัวอย่างประมาณร้อยละ 95.5 จะมีค่าอยู่ระหว่าง ค่าเฉลี่ย ± 2 SD
- ตัวอย่างประมาณร้อยละ 99.7 จะมีค่าอยู่ระหว่าง ค่าเฉลี่ย ± 3 SD



รูปที่ 10 การกระจายความถี่ของข้อมูลตามค่าส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่มีการแจกแจงเป็นปกติ

เช่นเดียวกับการใช้ค่าเฉลี่ย ข้อจำกัดของการใช้ส่วนเบี่ยงเบนมาตรฐานในการบอกการกระจายของข้อมูลก็เป็นผลมาจากการใช้ข้อมูลจากตัวอย่างทั้งหมด หากข้อมูลดังกล่าวมีลักษณะเบ้ก็จะทำให้ส่วนเบี่ยงเบนมาตรฐานมีค่าสูงขึ้นกว่าที่ควรจะเป็นเนื่องจากอิทธิพลของข้อมูลที่สูงหรือต่ำผิดปกติจากกลุ่มมากๆ

หลักเกณฑ์อย่างง่ายในการเลือกใช้ค่ากลางและการกระจายสามารถสรุปได้ ดังที่แสดงในตารางที่ 7

ตารางที่ 7 สรุปแนวทางการเลือกค่ากลางและการกระจายให้เหมาะสมกับชนิดของข้อมูล

ชนิดของข้อมูล	ค่ากลาง	การกระจาย
แจกแจงแบบปกติ	ค่าเฉลี่ยเลขคณิต	ส่วนเบี่ยงเบนมาตรฐาน
ไม่ได้แจกแจงแบบปกติ	มัธยฐาน หรือ ค่าเฉลี่ยเรขาคณิต (กรณีข้อมูลเบ้ขวา)	พิสัยระหว่างควอร์ไทล์ และ/หรือ พิสัย

สถิติเชิงอนุมาน: การสะท้อนภาพข้อมูลของประชากร

ลักษณะและหลักการพื้นฐานของสถิติเชิงอนุมาน

สถิติเชิงอนุมาน เป็นกระบวนการในการนำค่าของข้อมูลที่ได้จากกลุ่มตัวอย่าง (เรียกว่าค่าสถิติ: statistic) ซึ่งเป็นเพียงส่วนหนึ่งของประชากร มาใช้ในการตัดสินใจหรือประมาณค่าต่างๆ ที่เราสนใจของประชากรทั้งหมด (เรียกว่าค่าพารามิเตอร์: parameter) ซึ่งเป็นจุดมุ่งหมายหลักของการศึกษาทางระบาดวิทยา

โดยส่วนใหญ่ สามารถพูดแบบกว้างๆ ได้ว่าสถิติเชิงอนุมาน คือ เครื่องมือที่ใช้ในการประมาณ หรือทำนายค่าต่างๆ ในประชากรโดยอาศัยค่าที่ได้จากกลุ่มตัวอย่างที่เราเก็บรวบรวมมา ซึ่งสามารถระบุถึงความแม่นยำของการทำนาย (หรือโอกาสที่จะทำนายถูก) เป็นปริมาณได้อย่างชัดเจน อย่างไรก็ตามการทำนายในลักษณะดังกล่าวจะไม่มีทางทำได้หากเราไม่ทราบถึงวิธีในการได้มาซึ่งกลุ่มตัวอย่าง และความน่าจะเป็นในการที่จะได้กลุ่มตัวอย่างที่มีลักษณะดังกล่าว ดังนั้นในตำราสถิติขั้นพื้นฐานส่วนใหญ่จะระบุว่า หากต้องการใช้วิธีการทางสถิติเชิงอนุมาน กลุ่มตัวอย่างที่ได้มาจะต้องมาจาก*การเลือกตัวอย่างแบบสุ่มอย่างง่าย* (simple random sampling) ซึ่งหมายความว่า

- ทุกหน่วยของประชากรที่เราสนใจต้องมีโอกาสถูกเลือก คือไม่มีหน่วยใดที่มีโอกาสถูกเลือกเป็นศูนย์
- โอกาส หรือความน่าจะเป็นที่จะถูกเลือกของแต่ละหน่วยเท่ากัน ซึ่งมีค่าเท่ากับจำนวนกลุ่มตัวอย่างที่จะเลือกหารด้วยจำนวนประชากรทั้งหมด
- มีโอกาสได้กลุ่มตัวอย่างในทุกรูปแบบที่เป็นไปได้ภายใต้ขนาดกลุ่มตัวอย่างที่กำหนด แม้ว่ากลุ่มตัวอย่างแต่ละรูปแบบอาจจะมีโอกาสเกิดขึ้นมากน้อยแตกต่างกันไป

กล่าวโดยสรุป หากเราทราบความน่าจะเป็นของการได้มาซึ่งกลุ่มตัวอย่างจากประชากรที่เราสนใจ เราก็สามารถทำการวิเคราะห์ด้วยสถิติเชิงอนุมานโดยอาศัยหลักความน่าจะเป็นเพื่อย้อนกลับไปทำการตัดสินใจหรือประมาณค่าของประชากรกลุ่มดังกล่าวได้ (รูปที่ 11)

รูปที่ 11 ความสัมพันธ์ระหว่างข้อมูลที่ได้จากกลุ่มตัวอย่างและประชากร

ตัวอย่าง หากเรามีประชากรที่สนใจจะศึกษาความสูง 100 คน แต่เราเลือกตัวอย่างโดยการสุ่มอย่างง่ายมาเพียง 10 คน เพื่อวัดส่วนสูง โดยทั้ง 10 คน มีความสูง (เซนติเมตร) ดังนี้ 150, 170, 168, 142, 157, 163, 188, 175, 179 และ 166 หากเราคำนวณค่าเฉลี่ยของความสูงของกลุ่มตัวอย่างนี้ก็จะได้

$$\bar{x} = (150 + 170 + 168 + 142 + 157 + 163 + 188 + 175 + 179 + 166)/10 = 165.8 \text{ เซนติเมตร}$$

อย่างไรก็ตาม หากเราทำการเลือกตัวอย่างจากประชากรทั้ง 100 คนใหม่อีกครั้ง โดยสุ่มออกมาจำนวน 10 คน เท่าเดิม เราก็ย่อมมีโอกาสที่จะได้กลุ่มตัวอย่างที่แตกต่างจากข้างต้น และได้ความสูงเฉลี่ยแตกต่างไปจากเดิม (ไม่จำเป็นต้องเท่ากับ 165.8) ซึ่งความแตกต่างของค่าที่ได้จากการสุ่มแต่ละครั้งเรียกว่าความคลาดเคลื่อนจากการเลือกตัวอย่าง (sampling error หรือ random error)

สมมติว่าหากเราสามารถทำการเลือกตัวอย่างขนาด 10 คน ออกมาจากประชากรกลุ่มนี้ได้ โดยทำซ้ำแล้วซ้ำอีกเป็นจำนวนนับครั้งไม่ถ้วน เราก็จะได้ค่าเฉลี่ยของแต่ละกลุ่มตัวอย่างที่มีความแตกต่างกันมากมาย โดยอาจมีบางกลุ่มตัวอย่างที่มีค่าเฉลี่ยเท่ากันหรือใกล้เคียงกัน หากการสุ่มดังกล่าวได้มาตามหลักการความน่าจะเป็นโดยไม่มีอคติ หรือกลไกอื่นใดมาบิดเบือนให้ได้มาซึ่งกลุ่มตัวอย่างที่ไม่สอดคล้องกับหลักการดังกล่าว เราสามารถสรุปได้ว่า ค่าเฉลี่ยของค่าเฉลี่ยที่ได้จากกลุ่มตัวอย่างที่สุ่มได้แต่ละครั้ง (mean of sample means) จะเท่ากับค่าเฉลี่ยของประชากร (population mean) และเราสามารถถือว่าค่าเฉลี่ยของตัวอย่าง (sample mean) ที่เราได้มา (ซึ่งในทางปฏิบัติมักจะทำการเลือกตัวอย่างเพียงแค่ครั้งเดียว) เป็นค่าประมาณที่ไม่ลำเอียง (unbiased estimate) ของค่าเฉลี่ยของประชากร ซึ่งหมายความว่าโดยเฉลี่ยแล้วค่าเฉลี่ยของกลุ่มตัวอย่างที่ได้จะเท่ากับค่าเฉลี่ยของประชากร

อย่างไรก็ตาม แม้ว่าค่าที่ได้จากการเลือกตัวอย่างแต่ละครั้งจะไม่มีอคติ แต่ก็ไม่จำเป็นว่าจะต้องเท่ากับค่าของประชากร เนื่องจากยังมีความคลาดเคลื่อนอันเนื่องมาจากการเลือกตัวอย่าง นักสถิติได้อธิบายความคลาดเคลื่อนดังกล่าว โดยอาศัยหลักการว่าความคลาดเคลื่อนของค่าที่ได้จากกลุ่มตัวอย่างแต่ละครั้งเท่ากับส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยของตัวอย่าง (standard deviation of sample mean) ที่ได้จากการสุ่มนับครั้งไม่ถ้วนนั่นเอง และนักสถิติเรียกความคลาดเคลื่อนดังกล่าวว่า *ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย* (standard error of the mean หรือบางครั้งเรียกเพียงสั้นๆ ว่า standard error) ซึ่งสามารถคำนวณ ได้ดังนี้

standard error of the mean = standard deviation of sample mean
= standard deviation of sample $/\sqrt{n}$

โดยที่ n คือ ขนาดของกลุ่มตัวอย่าง

จะเห็นได้ว่า โดยหลักการแล้วเราควรจะต้องทำการเลือกตัวอย่างเป็นจำนวนนับครั้งไม่ถ้วนเพื่อที่จะคำนวณค่า standard error of the mean แต่ในทางปฏิบัติหากเรามีข้อมูลของกลุ่มตัวอย่างและสามารถคำนวณส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่างได้ (โดยวิธีการที่ได้กล่าวแล้วก่อนหน้านี้) เราก็สามารถคำนวณค่า standard error of the mean ได้แม้ว่าเราจะทำการเลือกตัวอย่างออกมาเพียงครั้งเดียว ข้อสังเกตประการหนึ่งที่สำคัญเกี่ยวกับค่า standard error ก็คือ ค่าดังกล่าวมีความสัมพันธ์แบบผกผันกับรากที่สองของขนาดกลุ่มตัวอย่าง ดังนั้นหากเราเพิ่มขนาดกลุ่มตัวอย่างเป็น 4 เท่า (โดยที่สมมติว่าค่าส่วนเบี่ยงเบนมาตรฐานยังคงเท่าเดิม) ก็จะได้ค่า standard error ที่เล็กลง คือเหลือเพียงครึ่งหนึ่งของค่าเดิม

นอกเหนือจากการประมาณค่าเฉลี่ยของประชากรแล้ว การประมาณสัดส่วน (proportion) ของการมีลักษณะที่เราสนใจในประชากร เราก็สามารถหาค่า standard error of the proportion โดยอาศัยหลักการเช่นเดียวกับที่ใช้กับค่าเฉลี่ย เพียงแต่มีวิธีการคำนวณในรายละเอียดที่แตกต่างกันซึ่งสามารถศึกษาเพิ่มเติมได้จากเอกสารที่แนะนำไว้ท้ายบท

การทดสอบสมมติฐาน และ p-value

ลักษณะแรกของการใช้สถิติเชิงอนุมาน คือ การทดสอบสมมติฐาน (hypothesis testing) บ่อยครั้งที่เราเก็บรวบรวมข้อมูลและคำนวณค่าสถิติจากกลุ่มตัวอย่าง เพื่อที่จะตัดสินใจเกี่ยวกับค่าพารามิเตอร์ในประชากร โดยที่เราต้องการทราบว่าข้อมูลที่เรามีอยู่สอดคล้องหรือไม่กับสมมติฐานที่เราได้ตั้งไว้ กระบวนการในลักษณะนี้เรียกว่า การทดสอบสมมติฐาน เช่น หากเราต้องการทราบว่าประชากรผู้ป่วยโรคเบาหวานที่ได้ยาชนิดใหม่

มีระดับน้ำตาลในเลือดแตกต่างจากผู้ป่วยที่ได้รับยาชนิดเดิมหรือไม่ เราก็สามารถตั้งสมมติฐานว่าระดับน้ำตาลในเลือดของผู้ป่วยเบาหวานที่ได้ยาทั้ง 2 ชนิด ไม่ได้มีความแตกต่างกัน (หรือบางคนอาจจะตั้งสมมติฐานในทางกลับกันว่า ระดับน้ำตาลของทั้ง 2 กลุ่มมีความแตกต่างกันก็ได้) จากนั้นเราก็จะทำการเก็บรวบรวมข้อมูลเพื่อดูว่ามีความสอดคล้องหรือไม่กับสมมติฐานที่เราได้ตั้งไว้

อย่างไรก็ตามในทางสถิติเรานิยมที่จะตั้งสมมติฐานหลักในลักษณะที่บอกว่าประชากร 2 กลุ่มที่เราต้องการเปรียบเทียบไม่มีความแตกต่างกัน (เรียกว่าสมมติฐานว่าง หรือ null hypothesis : H_0) และให้ตั้งสมมติฐานในทางกลับกันซึ่งมักจะเป็นสิ่งที่เราสนใจ (คือ 2 กลุ่มมีความแตกต่างกัน) เป็นสมมติฐานที่เราถือว่าเป็นคู่แข่งของสมมติฐานหลัก (เรียกว่า สมมติฐานทางเลือก หรือ alternative hypothesis : H_1 หรือ H_A) ซึ่งหากเราพบว่าข้อมูลของเราเข้าได้หรือสอดคล้องกับสมมติฐานหลักที่ตั้งไว้ ก็จะสรุปว่าเราไม่มีหลักฐานเพียงพอที่จะหักล้างสมมติฐานหลัก หรือพูดอีกนัยหนึ่งก็คือสมมติฐานหลักยังคงมีความเป็นไปได้สูงและยังไม่มีน้ำหนักเพียงพอที่จะสนับสนุนสมมติฐานที่เป็นคู่แข่ง ในทางกลับกันหากเราพบว่าข้อมูลของกลุ่มตัวอย่างที่เรามีไม่สอดคล้องกับสมมติฐานหลัก ก็จะสรุปว่าเรามีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานหลักและสมมติฐานที่เป็นคู่แข่งมีความเป็นได้สูง

จะสังเกตได้ว่าสมมติฐานที่เราตั้งไว้เป็นเรื่องของประชากร แต่ข้อมูลในมือที่เรามีเป็นข้อมูลของกลุ่มตัวอย่าง ดังนั้นการที่เราจะสรุปว่าข้อมูลของเราสอดคล้องหรือไม่กับสมมติฐาน จึงเป็นเรื่องที่มีโอกาสผิดพลาดได้เนื่องมาจากการที่เราเลือกกลุ่มตัวอย่างแทนที่จะศึกษาในประชากรทั้งหมดซึ่งเรียกว่า *ความคลาดเคลื่อนจากการเลือกตัวอย่าง* (sampling error) ด้วยเหตุดังกล่าวการพิจารณาความสอดคล้องระหว่างข้อมูลกับสมมติฐานจึงจำเป็นต้องอาศัยหลักความน่าจะเป็น โดยที่หากเราพบว่าความน่าจะเป็นที่ข้อมูลของเราจะสอดคล้องกับสมมติฐานต่ำมาก (เช่น ต่ำกว่าร้อยละ 5 หรือร้อยละ 1) เราก็จะสามารถสรุปได้ว่าข้อมูลของเราไม่สอดคล้องกับสมมติฐานด้วยความมั่นใจที่สูงว่าเราน่าจะสรุปได้ถูกต้อง (เช่น สูงกว่าร้อยละ 95 หรือร้อยละ 99) สิ่งที่บอกให้เราทราบว่าข้อมูลที่เรามีมีความสอดคล้องกับสมมติฐานหลักมากน้อยเพียงใดคือ p-value ซึ่งในทางสถิติแล้ว p-value มีความหมายว่า หากสมมติฐานหลักนั้นถูกต้อง การที่เราจะสุ่มได้ตัวอย่างและได้ข้อมูลลักษณะเหมือนกับที่เรามี (หรือห่างไกลจากสมมติฐานหลักยิ่งกว่าข้อมูลที่เรามี) มีโอกาสมากน้อยเพียงใด ดังนั้นถ้าเราได้ค่า p-value ที่ต่ำแสดงว่าข้อมูลที่เรามีไม่สอดคล้องกับสมมติฐานหลัก (ยิ่งต่ำมากแสดงว่าข้อมูลที่เรามียิ่งไม่สอดคล้องอย่างมากกับสมมติฐานหลัก) เราจึงสามารถปฏิเสธสมมติฐานหลักดังกล่าว ด้วยความเชื่อมั่นในระดับสูงและเชื่อว่าสมมติฐานที่เป็นคู่แข่งมีความเป็นไปได้สูงที่จะถูกต้อง

ค่า p-value ที่คนส่วนใหญ่ถือว่าต่ำ คือ น้อยกว่า 0.05 แต่ไม่จำเป็นเสมอไปที่จะต้องเป็นตัวเลขดังกล่าว ในความเป็นจริงแล้ว การที่จะกำหนดว่าค่า p-value เท่าใดจึงจะถือว่าต่ำในขนาดที่ยอมรับได้ ควรพิจารณาจากผลดีผลเสียที่จะเกิดขึ้นหากข้อสรุปของเราผิด โดยความผิดพลาดที่เกิดจากการที่เราปฏิเสธสมมติฐานหลักว่าไม่ถูกต้องโดยที่ในความเป็นจริงแล้วสมมติฐานดังกล่าวถูกต้อง เรียกว่า type I error และแสดงความน่าจะเป็นที่จะเกิดความผิดพลาดชนิดนี้ด้วย α (อ่านว่า alpha) โดยความน่าจะเป็นที่จะเกิดความผิดพลาดลักษณะดังกล่าวมีค่าเท่ากับ p-value อย่างไรก็ตาม วารสารทางวิชาการส่วนใหญ่ยังนิยมที่จะกำหนดจุดตัดของค่า p-value ที่ถือว่าต่ำไว้ที่ค่าตัวเลขหนึ่ง (เช่น ต่ำกว่า 0.05) โดยเรียกค่าดังกล่าวว่าระดับนัยสำคัญ (significant level หรือ α level) และหากพบว่าค่า p-value ที่ได้จากการวิเคราะห์ใดๆ ต่ำกว่าค่าดังกล่าวก็จะถือว่ามีนัยสำคัญทางสถิติ (statistical significance) อย่างไรก็ตามนักระบาดวิทยาในปัจจุบันมีแนวโน้มให้ความสำคัญกับ significant level น้อยลง (ยกเว้นเพื่อประโยชน์ในการตีพิมพ์เผยแพร่) และให้ความสนใจที่จะรายงานตัวเลขค่า p-value ที่แท้จริงมากกว่า อีกประเด็นสำคัญที่ต้องตระหนักคือการมีนัยสำคัญทางสถิติหรือไม่ ไม่จำเป็นจะต้องสอดคล้องกับการมีนัยสำคัญทางคลินิคหรือทางสาธารณสุขของผลการศึกษาดังกล่าว

ในทางกลับกันหากสมมติฐานหลักนั้นไม่เป็นความจริงแต่เราไม่มีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานดังกล่าว เรียกว่า type II error และแสดงความน่าจะเป็นที่จะเกิดความผิดพลาดชนิดนี้ด้วย β (อ่านว่า beta) (ดูตารางที่ 7) หากเราสนใจโอกาสหรือความน่าจะเป็นในการที่จะปฏิเสธสมมติฐานหลัก โดยที่ในความเป็นจริงแล้วสมมติฐานดังกล่าวไม่ถูกต้อง เราก็สามารถคำนวณโอกาสดังกล่าวได้โดยมีค่าเท่ากับ 1-β และเรียกค่าดังกล่าวว่า power ของการศึกษา โดยทั่วไปนิยมที่จะกำหนดให้ค่า power สูงกว่าร้อยละ 80

ตารางที่ 8 ความสัมพันธ์ระหว่างการทดสอบสมมติฐานกับความผิดพลาดในการสรุปผล

ข้อสรุปจากการศึกษา	ความเป็นจริง	
	H_0 เป็นความจริง	H_0 ไม่เป็นความจริง
ไม่ปฏิเสธ H_0	สรุปถูกต้อง	สรุปผิดพลาด (type II หรือ β error)
ปฏิเสธ H_0	สรุปผิดพลาด (type II หรือ α error)	สรุปถูกต้อง

อย่างไรก็ตาม ความผิดพลาดในการสรุปผลของการทดสอบสมมติฐานไม่ได้เกิดจาก sampling error เพียงอย่างเดียว แต่ยังสามารถเกิดความลำเอียง (bias) ประเภทต่างๆ รวมถึงวิธีการวิเคราะห์ที่ไม่ถูกต้อง (ประเด็นเหล่านี้อยู่นอกเหนือขอบเขตของเนื้อหาในบทนี้) ดังนั้นความถูกต้องของโอกาสหรือความเป็นไปได้ที่จะเกิดความคลาดเคลื่อนของสรุปผลการทดสอบสมมติฐานจึงตั้งอยู่บนเงื่อนไขว่าการศึกษาต้องไม่มีอคติใดๆ รวมถึงการใช้วิธีการทางสถิติที่ถูกต้องในการวิเคราะห์

การประมาณค่า และ confidence interval

อีกลักษณะหนึ่งของการใช้สถิติเชิงอนุมาน คือ การประมาณค่า (estimation) โดยเราทำการเก็บรวบรวมข้อมูลและคำนวณค่าสถิติจากกลุ่มตัวอย่าง แล้วนำค่าดังกล่าวไปใช้ประมาณค่าพารามิเตอร์ของประชากร อย่างไรก็ตามด้วยเหตุผลที่เราไม่สามารถหลีกหนีปัญหาเรื่อง sampling error การประมาณค่าของประชากรจึงไม่สามารถระบุเป็นค่าตัวเลขค่าใดค่าหนึ่งด้วยความเชื่อมั่น 100% ได้ แต่ต้องอาศัยการระบุเป็นช่วงระหว่างตัวเลข 2 ตัวและความเป็นไปได้ (หรือความเชื่อมั่น) ที่ค่าของประชากรจะถูกครอบคลุมโดยวิธีการศึกษาและสุ่มเลือกตัวอย่างในลักษณะดังกล่าว ช่วงของการประมาณค่าในประชากรที่ได้จากการเลือกตัวอย่าง เรียกว่า *ช่วงความเชื่อมั่น* (confidence interval) โดยค่าตัวเลขที่น้อยกว่าเรียกว่า lower confidence limit ส่วนค่าตัวเลขที่มากกว่าเรียกว่า upper confidence limit ซึ่งหากเรากำหนดว่าต้องการช่วงที่มีระดับความเชื่อมั่น (confidence level) ร้อยละ 95 เราก็จะเรียกช่วงดังกล่าวว่า 95% confidence interval ในทางสถิติหมายความว่า หากเราได้ทำการศึกษาประชากรกลุ่มดังกล่าวด้วยวิธีการศึกษาและการเลือกตัวอย่างแบบเดิมโดยทำซ้ำแล้วซ้ำอีกเป็นจำนวนนับครั้งไม่ถ้วน เราจะพบว่าร้อยละ 95 ของช่วงความเชื่อมั่นที่ได้ (ซึ่งแต่ละครั้งจะมีค่า lower และ upper confidence limit และความกว้างของช่วงความเชื่อมั่นที่แตกต่างกันไป ตามลักษณะตัวอย่างที่สุ่มได้) จะครอบคลุมค่าจริงในประชากร

อย่างไรก็ตาม ความหมายเชิงทฤษฎีของช่วงความเชื่อมั่นในลักษณะดังกล่าวไม่สอดคล้องกับวิธีการปฏิบัติ ซึ่งส่วนใหญ่แล้วเราทำการศึกษาและเลือกตัวอย่างเพียงครั้งเดียว จึงมีผู้นิยมที่จะแปลผลของช่วงความเชื่อมั่นว่าโอกาสที่ค่าของประชากรจะตกอยู่ในช่วงดังกล่าวเท่ากับร้อยละ 95 ซึ่งในทางสถิติแล้วเป็นการแปลที่ไม่ถูกต้อง

แม้ว่าจะสอดคล้องกับสามัญสำนึกของคนทั่วไป (ที่ไม่ใช่นักสถิติ) ก็ตาม เนื่องจากในทางสถิติถือว่าค่าพารามิเตอร์ของประชากรเป็นค่าคงที่ ความน่าจะเป็นที่ช่วงความเชื่อมั่นที่เราได้คำนวณจากกลุ่มตัวอย่างจึงขึ้นอยู่กับว่าตัวเลข 2 ตัวดังกล่าวครอบคลุมค่าประชากรหรือไม่ หากครอบคลุมค่าดังกล่าวก็ถือว่าความน่าจะเป็นเท่ากับร้อย หรือหากไม่ครอบคลุมค่าดังกล่าวก็ถือว่าความน่าจะเป็นเท่ากับศูนย์ ซึ่งพอจะเปรียบได้กับการที่เราพูดว่าความน่าจะเป็นที่ตัวเลข 8 (ค่าประชากร) จะอยู่ระหว่างค่า 5 ถึง 10 (ช่วงความเชื่อมั่น) เท่ากับร้อยละร้อย ในขณะที่ความน่าจะเป็นที่ตัวเลข 12 (ค่าประชากร) จะอยู่ระหว่างค่า 5 ถึง 10 เท่ากับร้อยละศูนย์ คงจะเป็นเรื่องแปลก หากเราพูดว่าโอกาสที่ตัวเลข 8 (หรือ 12) จะอยู่ระหว่างค่า 5 ถึง 10 เท่ากับร้อยละ 95 ในทางปฏิบัติ หากต้องการแปลผลช่วงความเชื่อมั่นให้ถูกต้องและเกิดประโยชน์ คงจะต้องมองว่าช่วงความเชื่อมั่นให้ข้อมูลที่เป็นประโยชน์ต่อเราใน 2 ลักษณะ (ซึ่งมีประโยชน์มากกว่าค่า p-value ที่ได้จากการทดสอบสมมติฐาน) ได้แก่

1) ช่วงระหว่างตัวเลข lower และ upper confidence limit นอกจากจะทำให้เราเห็นว่าค่าพารามิเตอร์น่าจะมีค่าอยู่ในช่วงใดหรือมากน้อยเพียงใดแล้ว (แม้จะไม่สามารถระบุความน่าจะเป็นได้ชัดเจนดังได้กล่าวแล้วข้างต้น) ช่วงดังกล่าวสามารถมองได้ว่าเป็นเซต (set) หรือช่วงของค่าต่างๆ ที่หากค่าพารามิเตอร์ (ซึ่งเราไม่ทราบค่า) ของประชากรตรงกับค่าใดค่าหนึ่งในช่วงดังกล่าวแล้ว ก็จะถือว่ามีความสอดคล้องกับข้อมูลที่เรามีอยู่ในมือ โดยหากตัดสินใจสรุปว่ามีความสอดคล้องกันก็จะมีโอกาสผิดพลาดไม่มากไปกว่าร้อยละ1-confidence level (เช่น น้อยกว่าร้อยละ 5 หากเป็นค่า 95% confidence interval) อย่างไรก็ตาม ค่าต่างๆ ที่อยู่ในช่วงดังกล่าวจะมีโอกาสที่จะสรุปผิดไม่เท่ากัน (คือ p-value ของแต่ละตัวเลขในช่วงดังกล่าวมีค่าไม่เท่ากัน แม้ทุกตัวจะมีค่าน้อยกว่าร้อยละ 5) โดยจุดที่มีโอกาสที่จะสรุปผิดมากที่สุดก็คือกรณีที่ค่าพารามิเตอร์มีค่าเท่ากับค่า lower หรือ upper confidence limit ซึ่งความน่าจะเป็นที่จะสรุปผิดในกรณีดังกล่าว ก็คือค่า α ที่เราได้กล่าวถึงมาแล้วในเรื่องการทดสอบสมมติฐานนั่นเอง ความเข้าใจในลักษณะเช่นนี้ของช่วงความเชื่อมั่นนำมาซึ่งวิธีการแปลผลที่เป็นที่นิยมโดยคนทั่วไป นั่นก็คือหากเราต้องการดูว่าตัวแปรปัจจัยหนึ่งที่เราสนใจมีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติ (ที่ระดับน้อยกว่า 0.05) กับตัวแปรผลลัพธ์หากเราพบว่า 95% confidence interval ของค่า risk ratio (RR) หรือ odds ratio (OR) ไม่ครอบคลุมค่า 1 (ซึ่งเรียกว่า null value ของ RR และ OR หรือแปลว่าไม่มีความสัมพันธ์ระหว่าง 2 ตัวแปร) เราก็สามารถสรุปได้ในทำนองเดียวกับการที่เราทำการทดสอบสมมติฐานแล้วได้ค่า p-value น้อยกว่า 0.05 คือปัจจัยทั้งสองมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับดังกล่าวนั่นเอง

2) ความกว้างของช่วงความเชื่อมั่นเป็นสิ่งที่บอกความแม่นยำ (precision) ของการประมาณค่า หากช่วงความเชื่อมั่นแคบก็ถือได้ว่ามีความแม่นยำสูง โดยทั่วไปแล้วหากทำการศึกษาลักษณะเดียวกันในประชากรกลุ่มเดียวกัน หากเพิ่มขนาดตัวอย่างก็จะได้ช่วงความเชื่อมั่นที่แคบลงหรือมีความแม่นยำสูงขึ้น ดังจะเห็นได้จากวิธีในการคำนวณช่วงความเชื่อมั่น ดังนี้

confidence interval ที่ระดับความเชื่อมั่น $(1-\alpha)$

$$= \bar{x} + [Z_{1-\alpha/2} \times (\text{standard deviation of sample} / \sqrt{n})]$$

โดยที่ $\bar{x}$ = sample mean

$Z_{1-\alpha/2}$ = standard score (Z-score) ที่ $1-\alpha/2$ ซึ่งสามารถหาค่าได้จากตาราง Z-score ในตำราทางสถิติมาตรฐานทั่วไป เช่น หากเป็น 95% confidence interval (หรือค่า α เท่ากับ 0.05) จะมีค่า $Z_{1-\alpha/2}$ เท่ากับ 1.96

n = sample size

ส่วนเครื่องหมาย + หรือ – ใช้ในการคำนวณ upper หรือ lower confidence limit ตามลำดับ

จากการคำนวณช่วงความเชื่อมั่นที่ได้แสดงข้างต้น จะเห็นได้ว่านอกจากขนาดตัวอย่างจะส่งผลต่อความกว้างของช่วงความเชื่อมั่นแล้ว หากเราต้องการให้ได้ช่วงความเชื่อมั่นที่มีระดับความเชื่อมั่นสูงขึ้น (โดยไม่เพิ่มขนาดกลุ่มตัวอย่าง) ก็จะพบว่าช่วงความเชื่อมั่นที่ได้จะมีความกว้างขึ้นเช่นเดียวกัน เช่น ช่วงความเชื่อมั่นที่ระดับร้อยละ 95 จะกว้างกว่าช่วงความเชื่อมั่นที่ระดับร้อยละ 90

โดยสรุป ช่วงความเชื่อมั่นแสดงให้เห็นถึงความคลาดเคลื่อนที่เกิดจาก sampling error ในการใช้ข้อมูลจากกลุ่มตัวอย่างเพื่อประมาณค่าพารามิเตอร์ของประชากร อย่างไรก็ตามปัญหาความคลาดเคลื่อนในการแปลผลช่วงความเชื่อมั่นก็มีลักษณะเช่นเดียวกับการทดสอบสมมติฐาน นั่นก็คือความคลาดเคลื่อนดังกล่าวสามารถเกิดจากความลำเอียง (bias) ประเภทต่างๆ รวมถึงการใช้วิธีการทางสถิติที่ไม่ถูกต้องในการวิเคราะห์เช่นเดียวกัน

เนื้อหาในส่วนต่อไปจะเป็นการแนะนำวิธีการทางสถิติที่ใช้บ่อยในการทดสอบสมมติฐานหรือ การประมาณค่าของประชากร

t-test

หากนักระบาดวิทยาสนใจที่จะเปรียบเทียบค่าของตัวแปรที่เป็นตัวเลข (numerical variable) โดยต้องการดูว่าค่าเฉลี่ยของตัวแปรดังกล่าวในประชากร 2 กลุ่มว่ามีความแตกต่างกันหรือไม่ (เช่น อยากทราบว่าความสูงเฉลี่ยของเด็กไทยอายุ 5 ปี มีความแตกต่างกันหรือไม่ระหว่างเพศชายและหญิง) เมื่อได้รวบรวมข้อมูลจากกลุ่มตัวอย่างซึ่งได้มาจากวิธีการสุ่มอย่างง่ายของทั้ง 2 เพศแล้ว และทำการคำนวณความสูงเฉลี่ยของกลุ่มตัวอย่างทั้ง 2 เพศ (ซึ่งเป็นเรื่องปกติที่จะพบว่าค่าเฉลี่ยของ 2 กลุ่มตัวอย่างไม่เท่ากัน) การทดสอบทางสถิติที่เรียกว่า *t*-test จะช่วยให้เราทราบว่าความสูงเฉลี่ยของทั้ง 2 เพศที่ได้มาจากข้อมูลกลุ่มตัวอย่างมีความแตกต่างกันมากพอหรือไม่ที่จะสรุปว่าความสูงเฉลี่ยของประชากรทั้ง 2 เพศมีความแตกต่างกันจริง ซึ่งหากลองนำมาเขียนเป็นสมมติฐานที่จะทำการทดสอบ ก็อาจจะเขียนได้เป็นอย่างน้อย 3 กรณี

<u>กรณีที่ 1</u>

H_0 : ความสูงเฉลี่ยของประชากรไทยอายุ 5 ปีเพศชาย <u>เท่ากับ</u> เพศหญิง

H_1 : ความสูงเฉลี่ยของประชากรไทยอายุ 5 ปีเพศชาย <u>ไม่เท่ากับ</u> เพศหญิง

<u>กรณีที่ 2</u>

H_0 : ความสูงเฉลี่ยของประชากรไทยอายุ 5 ปีเพศชาย <u>เท่ากับหรือน้อยกว่า</u> เพศหญิง

H_1 : ความสูงเฉลี่ยของประชากรไทยอายุ 5 ปีเพศชาย <u>มากกว่า</u> เพศหญิง

<u>กรณีที่ 3</u>

H_0 : ความสูงเฉลี่ยของประชากรไทยอายุห้าปีเพศชาย <u>เท่ากับหรือมากกว่า</u> เพศหญิง

H_1 : ความสูงเฉลี่ยของประชากรไทยอายุห้าปีเพศชาย <u>น้อยกว่า</u> เพศหญิง

ดังที่ได้กล่าวก่อนหน้านี้ว่าในความเป็นจริงแล้วสิ่งที่เราให้ความสนใจหรือคำตอบที่คาดไว้มักจะเป็นสมมติฐานทางเลือก (alternative hypothesis) ดังนั้นในทางปฏิบัติเราจึงมักจะได้สมมติฐานแบบนี้ขึ้นมาก่อน (เช่น หากคาดว่าเพศชายโดยเฉลี่ยสูงกว่าเพศหญิงก็จะได้สมมติฐาน H_1 ดังในกรณีที่ 2 ส่วนสมมติฐานว่าง (null hypothesis) ที่ถือว่าเป็นสมมติฐานหลักที่จะทำการทดสอบทางสถิติจะได้ตามมา โดยจะระบุเป็นผลลัพธ์แบบอื่นที่อาจเกิดขึ้นนอกเหนือจากผลลัพธ์ที่ตั้งไว้ใน alternative hypothesis (ดังตัวอย่างในกรณี 2 ก็คือว่าเพศชายต้องไม่สูงกว่าเพศหญิง หรือพูดอีกแบบว่าเพศชายมีความสูงเฉลี่ยเท่ากับหรือน้อยกว่าเพศหญิงนั่นเอง)

ประเด็นสำคัญอีกอันหนึ่งที่ควรทราบคือ หากเราตั้งสมมติฐานที่เป็นคู่แข่งในลักษณะที่ระบุว่ากลุ่มหนึ่งจะมีค่าเฉลี่ยมากกว่าหรือน้อยกว่าอีกกลุ่มหนึ่งเพียงด้านใดด้านหนึ่ง (ดังเช่นกรณีที่ 2 และ 3) เราจะเรียกการทดสอบในลักษณะนี้ว่าการทดสอบสมมติฐานแบบด้านเดียว (one-sided test) แต่หากว่าเราไม่แน่ใจว่ากลุ่มใดจะมากกว่าหรือน้อยกว่าอีกกลุ่มหนึ่งแต่คาดว่าน่าจะไม่เท่ากัน เราก็สามารถตั้งสมมติฐานที่เป็นคู่แข่งดังในกรณีที่ 1 ได้ (คือ ค่าเฉลี่ยใน 2 กลุ่มไม่เท่ากัน โดยที่กลุ่มหนึ่งอาจมากกว่าหรือน้อยกว่าอีกกลุ่มหนึ่งก็ได้) และเรียกการทดสอบในลักษณะนี้ว่าการทดสอบสมมติฐานแบบ 2 ด้าน (two-sided test)

โดยปกติหากเราทำการวิเคราะห์ข้อมูลด้วยโปรแกรมสำเร็จรูปทางสถิติ เมื่อได้สั่งให้ทำการทดสอบด้วย *t*-test โปรแกรมส่วนใหญ่จะรายงานผลของการทดสอบให้เราทั้ง 3 กรณีข้างต้นในคราวเดียวกัน (รายงาน *p*-value มาให้ 3 ค่าพร้อมกัน) หน้าที่ของผู้ที่ทำการวิเคราะห์คือการเลือกนำเฉพาะค่าที่ตรงกับกรณีที่ตนสนใจจะทดสอบไปรายงานและแปลความหมาย อย่างที่กล่าวแล้วข้างต้น ข้อมูลที่สำคัญที่ใช้ในการตัดสินผลการทดสอบสมมติฐานก็คือ *p*-value ดังนั้นหากเราพบว่าค่าดังกล่าวมีค่าต่ำมาก ซึ่งหมายความว่าข้อมูลของกลุ่มตัวอย่างไม่สอดคล้องกับสมมติฐานว่างอย่างมาก เราก็จะสรุปว่าสามารถปฏิเสธสมมติฐานดังกล่าวได้ (โดยมีโอกาสผิดพลาดไม่เกินค่า *p*-value) และมีความเป็นไปได้ที่สมมติฐานที่เป็นคู่แข่งจะถูกต้อง เช่น หากในกรณีที่ 1 เราพบว่าได้ค่า *p*-value เท่ากับ 0.03 ก็สามารถสรุปได้ว่าเรามีความมั่นใจที่จะปฏิเสธสมมติฐานว่างที่ว่าทั้ง 2 เพศมีความสูงเฉลี่ยเท่ากัน (คือเมื่อดูผลการศึกษาแล้วเราเชื่อว่า 2 เพศน่าจะมีความสูงเฉลี่ยไม่เท่ากันนั่นเอง) โดยมีโอกาสสรุปผิดไม่เกินร้อยละ 3

หากเราทำการเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างที่มีความเกี่ยวข้องกัน (dependent sample) เช่น ต้องการเปรียบเทียบระดับไขมันในเลือดของกลุ่มตัวอย่างก่อนและหลังการได้รับยาลดไขมัน (ทุกคนในกลุ่มตัวอย่างได้รับยาดังกล่าว) หรือเปรียบเทียบความสูงของกลุ่มตัวอย่างนักเรียนกลุ่มเดียวกันตอนอยู่ชั้น ป.3 กับ ตอนอยู่ชั้น ป.6 (วัดซ้ำในนักเรียนคนเดิม) จะถือว่ากลุ่มตัวอย่างดังกล่าวเกี่ยวข้องกันหรือไม่ได้เป็นอิสระต่อกัน วิธีการทดสอบค่าเฉลี่ยที่ใช้จะแตกต่างเล็กน้อยจาก *t*-test ที่ได้กล่าวมาข้างต้น โดยเรียกการทดสอบค่าเฉลี่ยของกลุ่มตัวอย่างที่เกี่ยวข้องกันนี้ว่า dependent *t*-test หรือ paired *t*-test (ส่วนการทดสอบแบบที่กล่าวแล้วข้างต้นก็สามารถเรียกแบบเต็มได้ว่า independent *t*-test หรือ unpaired *t*-test)

เงื่อนไขหรือข้อสมมติ (assumption) ที่จำเป็นต้องมีหากจะทำการทดสอบ *t*-test ก็คือ

- กลุ่มตัวอย่างที่จะนำมาเปรียบเทียบได้ จะต้องมาโดยการเลือกตัวอย่างแบบสุ่มอย่างง่าย (simple random sampling) จากประชากร
- ค่าที่สนใจจะเปรียบเทียบของแต่ละตัวอย่างที่นำมาศึกษาเป็นอิสระต่อกัน (ยกเว้นกรณี paired *t*-test) หรือพูดอีกแบบว่าค่าของตัวอย่างหนึ่งๆ ไม่ได้ส่งผลต่อค่าของตัวอย่างอื่นๆ
- ค่าตัวแปรเชิงปริมาณที่จะศึกษาของแต่ละกลุ่มมาจากประชากรที่มีการแจกแจงปกติ (normal distribution) และมีความแปรปรวน (หรือกระจายตัว) เท่ากันหรือใกล้เคียงกัน (equal variance)

นอกจากนี้ข้อจำกัดที่สำคัญประการหนึ่งของ *t*-test ก็คือใช้ได้เฉพาะในกรณีเปรียบเทียบข้อมูลของ2 กลุ่มตัวอย่าง หากเรามีกลุ่มตัวอย่างตั้งแต่ 3 กลุ่มขึ้นไปที่ต้องการเปรียบเทียบจะต้องใช้วิธีการอื่น เช่น analysis of variance (ANOVA) หรือ linear regression และหากข้อสมมติเรื่องการกระจายแบบปกติและความแปรปรวนที่เท่ากันไม่เป็นจริงในข้อมูลของเรา ก็จะต้องใช้การทดสอบที่เรียกว่า Wilcoxon rank sum test แทน ซึ่งถือเป็นการทดสอบในกลุ่มที่ไม่ต้องอาศัยสมมติฐานเรื่องการกระจายของข้อมูล และเราเรียกการทดสอบที่อยู่ในกลุ่มนี้ว่า nonparametric tests

Chi-squared test

หากเราสนใจที่จะทำการเปรียบเทียบสัดส่วนของการมีลักษณะอย่างใดอย่างหนึ่งในประชากร เช่น หากเราต้องการเปรียบเทียบว่าระหว่างประชากรเพศชายและหญิงที่อายุมากกว่า 60 ปี มีสัดส่วนของผู้เป็นความดันโลหิตสูงต่างกันหรือไม่ (สนใจว่ามีหรือไม่มีภาวะดังกล่าว โดยไม่ได้สนใจว่ามีค่าความดันโลหิตเป็นเท่าใด) เมื่อเราทำการสุ่มอย่างง่ายเพื่อให้ได้กลุ่มตัวอย่างของแต่ละเพศและทำการประเมินว่ามีผู้เป็นความดันโลหิตสูงในสัดส่วนเท่าใด ซึ่งเราสามารถพูดอีกแบบได้ว่าเราต้องการศึกษาความสัมพันธ์ระหว่างเพศกับการเป็นความดันโลหิตสูงในประชากรกลุ่มดังกล่าว ว่ามีอยู่จริงหรือไม่ หากเราทำการเก็บข้อมูลและได้ตัวเลข ดังนี้

- กลุ่มตัวอย่างผู้ชาย 100 คน มีผู้เป็นความดันโลหิตสูง 35 คน
- กลุ่มตัวอย่างผู้หญิง 100 คน มีผู้เป็นความดันโลหิตสูง 28 คน

เราสามารถแสดงข้อมูลดังกล่าวในรูปแบบของตาราง 2 คูณ 2 (2 x 2 table) (เนื่องจากสามารถแบ่งข้อมูลกลุ่มเปรียบเทียบ (หรือตัวแปรปัจจัย) และตัวแปรผลลัพธ์ที่เราสนใจได้อย่างละ 2 กลุ่ม และเราแสดงการกระจายของข้อมูลด้วยตารางที่มี 2 แถว และ 2 สดมภ์) ดังแสดงในตารางที่ 9

ตารางที่ 9 การเป็นความดันโลหิตสูงแยกตามเพศ

เพศ	เป็นความดันโลหิตสูง	ไม่เป็นความดันโลหิตสูง	รวม
ชาย	35	65	100
หญิง	28	72	100
รวม	63	137	200

เราจะเห็นได้ว่าสัดส่วนของการเป็นความดันโลหิตสูงแตกต่างกันในกลุ่มตัวอย่าง (ร้อยละ 35 กับร้อยละ 28) แต่สิ่งที่เราสนใจคือ อยากทราบว่าความแตกต่างของสัดส่วนการเป็นความดันโลหิตสูงระหว่างทั้ง 2 เพศ (หรือความสัมพันธ์ระหว่างเพศกับการเป็นความดันโลหิตสูง) มีอยู่จริงในประชากรหรือไม่ ซึ่งเราจะใช้การทดสอบที่เรียกว่า chi-squared test เป็นเครื่องมือในการตอบคำถามดังกล่าว โดยสมมติฐานที่ต้องการทดสอบมีดังนี้

H_0 : สัดส่วนของผู้เป็นความดันโลหิตสูงในประชากรอายุมากกว่า 60 ปี เพศชาย<u>เท่ากับ</u>เพศหญิง (หรือเพศกับการเป็นความดันโลหิตสูงไม่มีความสัมพันธ์กัน)

H_1 : สัดส่วนของผู้เป็นความดันโลหิตสูงในประชากรอายุมากกว่า 60 ปี <u>แตกต่างกัน</u>ระหว่าง 2 เพศ (หรือเพศกับการเป็นความดันโลหิตสูงมีความสัมพันธ์กัน)

ในกรณีตัวอย่างที่ได้ยกมานี้ หากเราได้ค่า *p*-value เท่ากับ 0.18 ซึ่งเป็นค่าที่ค่อนข้างสูงและเราไม่สามารถที่จะปฏิเสธ H_0 ได้อย่างมั่นใจ (เนื่องจากหากเราปฏิเสธก็จะมีโอกาสผิดพลาดได้ถึงร้อยละ 18) เราก็จะสรุปว่าเราไม่มีหลักฐานที่จะปฏิเสธสมมติฐานว่าง ที่ว่าเพศกับการเป็นความดันโลหิตสูงไม่มีความสัมพันธ์กันในประชากรอายุมากกว่า 60 ปี ดังนั้น ด้วยข้อมูลที่เรามีในมือขณะนี้จึงยังไม่เพียงพอที่จะสรุปว่าเพศมีความสัมพันธ์กับการเป็นความดันโลหิตสูง

การทดสอบ chi-squared test นี้สามารถใช้ในการเปรียบเทียบสัดส่วนของกลุ่มตัวอย่างมากกว่า 2 กลุ่มได้ และลักษณะตัวแปรที่ต้องการดูสัดส่วนไม่จำเป็นที่จะต้องมีค่าได้แค่ 2 อย่าง (ป่วย/ไม่ป่วย หรือตาย/ไม่ตาย) แต่สามารถมีได้หลายค่า ซึ่งเราสามารถแสดงข้อมูลดังกล่าวในรูปตารางได้เช่นกัน โดยเรียกว่า ตาราง r x c

ซึ่งมาจากจำนวนแถว (row) คูณกับจำนวนสดมภ์ (column) เช่น ไม่ป่วย/ป่วยไม่รุนแรง/ป่วยรุนแรง ยกตัวอย่างเช่น เราอาจต้องการศึกษาว่าในคนสามกลุ่มอายุมีสัดส่วนของกลุ่มดัชนีมวลกาย (Body Mass Index, BMI) แตกต่างกันหรือไม่ ดังแสดงผลในตารางที่ 10

ตารางที่ 10 ระดับดัชนีมวลกายแยกตามกลุ่มอายุ

ระดับดัชนีมวลกาย	อายุ 15-45 ปี	อายุ 46-60 ปี	อายุ 61 ปีขึ้นไป	รวม
อ้วน	27	46	22	95
น้ำหนักมากเกินไป	44	58	38	140
ปกติ	38	55	44	137
น้ำหนักน้อยเกินไป	16	32	40	88
รวม	125	191	144	460

ในกรณีตัวอย่างนี้ เราเรียกว่าเป็นตาราง 4x3 เนื่องจากมี 4 แถวและ 3 สดมภ์ และเราจะได้สมมติฐานที่ต้องการทดสอบดังนี้

H_0 : สัดส่วนของระดับดัชนีมวลกายในประชากรทั้ง 3 กลุ่มอายุในตารางไม่มีความแตกต่างกัน (หรืออายุที่แบ่งออกเป็น 3 กลุ่มดังตารางกับระดับดัชนีมวลกายไม่มีความสัมพันธ์กัน)

H_1 : สัดส่วนของระดับดัชนีมวลกายในประชากรทั้ง 3 กลุ่มอายุในตาราง มีความแตกต่างกันบางระดับในระหว่างกลุ่มอายุอย่างน้อย 1 คู่ (หรืออายุที่แบ่งออกเป็น 3 กลุ่มดังตารางกับระดับดัชนีมวลกายมีความสัมพันธ์กัน)

จากข้อมูลดังกล่าวหากเราได้ p-value เท่ากับ 0.032 ซึ่งหมายความว่าหากเราปฏิเสธสมมติฐานว่างที่ระบุว่าอายุที่แบ่งออกเป็น 3 กลุ่มดังตารางกับระดับดัชนีมวลกายไม่มีความสัมพันธ์กัน เราจะมีโอกาสผิดพลาดไม่เกินร้อยละ 3.2 ซึ่งถือว่าต่ำมาก ดังนั้นเราจึงสรุปว่า ในประชากรที่เราศึกษา อายุ (ที่แบ่งออกเป็น 3 กลุ่มดังตาราง) กับระดับดัชนีมวลกายมีความสัมพันธ์กัน หรือมีความแตกต่างกันของการกระจายของระดับดัชนีมวลกายในแต่ละกลุ่มอายุ

กรณีตัวอย่างข้างต้นจะเห็นได้ว่า chi-squared test บอกได้แต่เพียงว่ามีความสัมพันธ์ระหว่าง 2 ปัจจัยที่เราทำการศึกษาในประชากร (ในที่นี้คืออายุกับระดับดัชนีมวลกาย) ซึ่งหมายความว่ามีความแตกต่างของสัดส่วนอย่างน้อย 1 ระดับในอย่างน้อย 1 คู่เปรียบเทียบ แต่ผลการทดสอบนี้ไม่ได้บอกว่าข้อมูลมีความแตกต่างเกิดขึ้นกี่ระดับ หรือกี่กลุ่มที่แตกต่างกัน และเป็นที่ระดับหรือกลุ่มใดที่มีความแตกต่างกัน หากต้องการทราบว่าความแตกต่างเกิดขึ้นที่ระดับใดหรือกลุ่มเปรียบเทียบใดจะต้องใช้วิธีการประยุกต์โดยปรับการวิเคราะห์ให้ค่าของตัวแปรที่ต้องการดูสัดส่วน (ในที่นี้คือระดับดัชนีมวลกาย) เหลือเพียง 2 ระดับ เช่น อ้วนเทียบกับไม่อ้วน (นำกลุ่มน้ำหนักมากเกินไป กลุ่มน้ำหนักปกติ และกลุ่มน้ำหนักน้อยเกินไปมารวมกัน) และเปรียบเทียบระหว่างกลุ่มอายุทีละคู่ หรืออาจใช้วิธีอื่น เช่น logistic regression

ข้อจำกัดที่สำคัญอย่างหนึ่งของ chi-squared test คือ การทดสอบนี้จะให้ผลที่ไม่น่าเชื่อถือหากจำนวนตัวอย่างที่ศึกษามีจำนวนไม่มากร่วมกับสัดส่วนของปัจจัยทั้งสองที่จะศึกษาความสัมพันธ์กันมีค่าต่ำในบางระดับหรือบางกลุ่ม (ผู้อ่านที่สนใจสามารถดูรายละเอียดของประเด็นนี้ในตำราสถิติมาตรฐานทั่วไป) หากพบปัญหา

ดังกล่าวในกรณีที่เป็นตาราง 2x2 ก็สามารถใช้การทดสอบที่มีชื่อว่า fisher's exact-test ทดแทนได้ ซึ่งการทดสอบแบบนี้มีการตั้งสมมติฐานและวิธีการแปลผลเช่นเดียวกันกับ chi-squared test ในโปรแกรมสถิติสำเร็จรูปบางโปรแกรมจะบอกให้เราทราบโดยอัตโนมัติในกรณีที่พบว่าตัวอย่างที่มีไม่สามารถใช้ผลการวิเคราะห์ chi-squared test ได้ โดยโปรแกรมจะแนะนำให้ใช้ค่าจาก fisher's exact-test แทน

แนวคิดของการวิเคราะห์การถดถอย (regression analysis)

การวิเคราะห์การถดถอยถือเป็นเครื่องมือที่สำคัญมากทั้งในการศึกษาทางระบาดวิทยาและการวิจัยในสาขาอื่นๆ โดยวัตถุประสงค์ของการวิเคราะห์การถดถอย คือเพื่อหาความสัมพันธ์เชิงปริมาณระหว่างตัวแปรที่เป็นผลลัพธ์ที่เราสนใจหรือตัวแปรตาม (outcome หรือ dependent variable) เช่น ระดับน้ำตาลในเลือด การเป็นหรือไม่เป็นไข้หวัดใหญ่ หรือระยะเวลาในการรอดชีวิตภายหลังจากเป็นโรคมะเร็ง ฯลฯ กับตัวแปรที่ใช้เป็นตัวทำนายหรือตัวแปรต้น (predictor หรือ independent variable) เช่น เพศ อายุ ชนิดของการรักษาหรือยาที่ใช้ ฯลฯ การวิเคราะห์การถดถอยอาศัยแบบจำลองคณิตศาสตร์ (mathematical model) ที่เขียนแสดงในรูปแบบสมการ เพื่อจำลองหรือเลียนแบบปรากฏการณ์การเกิดโรค (หรือผลลัพธ์อื่นๆ ที่เราสนใจ) ในธรรมชาติ ซึ่งเป็นผลมาจากปัจจัยต่างๆ มากมาย โดยที่ปัจจัยต่างๆ เหล่านี้หลายปัจจัยมีความสัมพันธ์ซึ่งกันและกันด้วย นอกเหนือไปจากการที่มีความสัมพันธ์กับตัวโรคหรือผลลัพธ์ที่เราสนใจ หากปัจจัยใดไม่ได้ถูกนำมาใส่ไว้ในแบบจำลองก็หมายความว่าผู้ทำการวิเคราะห์ถือว่าปัจจัยดังกล่าวไม่ได้มีความสัมพันธ์กับผลลัพธ์ที่เราสนใจ การวิเคราะห์ในลักษณะนี้จะช่วยให้เราสามารถศึกษาความสัมพันธ์ระหว่างตัวแปรปัจจัยที่เราสนใจกับการเกิดผลลัพธ์ โดยที่สามารถควบคุมอิทธิพลของปัจจัยอื่นๆ ที่มีอิทธิพลต่อผลลัพธ์ที่ได้ใส่ไว้ในแบบจำลองได้หลายปัจจัยในเวลาเดียวกัน (ดูเรื่อง confounding ในบทที่ 5) รูปแบบของการวิเคราะห์การถดถอยมีหลายชนิดขึ้นอยู่กับลักษณะของตัวแปรผลลัพธ์ที่เราวัด เช่น

- Linear regression ใช้ในกรณีที่ตัวแปรผลลัพธ์เป็นตัวแปรตัวเลข เช่น ระดับไขมันในเลือด ความดันโลหิต ดัชนีมวลกาย ฯลฯ
- Logistic regression ใช้ในกรณีที่ตัวแปรผลลัพธ์เป็นตัวแปรแยกประเภท เช่น เป็นโรค/ไม่เป็นโรค ตาย/ไม่ตาย ไม่ป่วย/ป่วยไม่รุนแรง/ป่วยรุนแรง ฯลฯ
- Cox proportional hazards regression ใช้ในกรณีที่ตัวแปรผลลัพธ์เป็นระยะเวลาในการเกิดเหตุการณ์หรือผลลัพธ์ที่เราสนใจ เช่น ระยะเวลาในการรอดชีพหลังเป็นมะเร็งต่อมลูกหมาก ระยะเวลาในการเกิดไตวายหลังเป็นโรคเบาหวาน ฯลฯ
- Poisson regression ใช้ในกรณีที่ตัวแปรผลลัพธ์เป็นอัตราเกิดโรค (incidence rate) หรือจำนวนการเกิดเหตุการณ์ เช่น incidence rate ของการเกิดโรคกล้ามเนื้อหัวใจขาดเลือดเฉียบพลัน จำนวนคนป่วยที่เกิดขึ้นในแต่ละหมู่บ้าน ฯลฯ

เนื้อหาในบทนี้จะกล่าวถึงเฉพาะหลักการเบื้องต้นของ linear regression analysis และ logistic regression analysis ซึ่งถือเป็นวิธีวิเคราะห์การถดถอยขั้นพื้นฐานและใช้บ่อยในการศึกษาทางระบาดวิทยา โดยทั้ง 2 วิธีการนี้ล้วนมีข้อสมมติอยู่หลายประการ ซึ่งจำเป็นที่จะต้องทำการทดสอบด้วยว่าข้อสมมติเหล่านั้นเป็นจริงหรือไม่หากจะทำการวิเคราะห์ด้วยวิธีการเหล่านี้ ผู้ที่สนใจสามารถศึกษารายละเอียดได้จากตำราสถิติที่แนะนำไว้ท้ายบท

การวิเคราะห์การถดถอยเชิงเส้น (linear regression analysis)

ในกรณีที่ตัวแปรผลลัพธ์เป็นตัวแปรตัวเลข (numerical variable) ที่ไม่ใช่อัตราเกิดโรคหรือจำนวนการเกิดเหตุการณ์ เช่น น้ำหนักตัว และเราต้องการศึกษาความสัมพันธ์ระหว่างปัจจัยต่างๆ หลายตัว (ซึ่งอาจเป็นตัวแปรตัวเลขและตัวแปรจัดกลุ่มก็ได้) ที่อาจจะมีความสัมพันธ์กับตัวแปรผลลัพธ์ดังกล่าว เช่น ส่วนสูง อายุ เพศ เชื้อชาติ ฯลฯ เราก็สามารถใช้แบบจำลองการถดถอยเชิงเส้น (linear regression model) ในการอธิบายความสัมพันธ์ดังกล่าว

โดยแบบจำลอง (หรือสมการ) ดังกล่าวมีลักษณะทั่วไปดังต่อไปนี้

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ...+ \beta_k x_k + \varepsilon$$

โดยที่ Y = ตัวแปรผลลัพธ์ ในที่นี้คือน้ำหนักตัว (กิโลกรัม)

β_0 = จุดตัดแกน y (intercept) หรือค่าของตัวแปรผลลัพธ์ เมื่อค่าของตัวแปรปัจจัย (x) ทุกตัวที่อยู่ในแบบจำลองมีค่าเท่ากับศูนย์

β_i = ค่าสัมประสิทธิ์การถดถอย (Regression coefficient) ของตัวแปรปัจจัยแต่ละตัว

x_i = ค่าของตัวแปรปัจจัยแต่ละตัว (x) ในที่นี้ ได้แก่ ส่วนสูง (เซนติเมตร) อายุ (ปี) เพศ (หญิง/ชาย) เชื้อชาติ (ไทย/ลาว/กัมพูชา)

ε = ความคลาดเคลื่อน (error) ที่ไม่สามารถอธิบายได้ด้วยตัวแปรที่อยู่ในแบบจำลอง

โดยหลักการแล้ว หากแบบจำลองใดมีค่าความคลาดเคลื่อนต่ำ ก็แสดงว่าตัวแปรปัจจัยต่างๆ ที่เราใส่ไว้ในแบบจำลองนั้นมีความเหมาะสมและถือว่าเป็นแบบจำลองที่อธิบายความสัมพันธ์ที่เราสนใจจะศึกษาได้ดีกว่าเมื่อเปรียบเทียบกับแบบจำลองที่มีค่าความคลาดเคลื่อนสูง

ในทางปฏิบัติเราไม่ได้ใช้ค่าความคลาดเคลื่อนนี้โดยตรงในการพิจารณาว่าแบบจำลองมีความเหมาะสมหรือไม่ แต่เราจะดูว่าตัวแปรปัจจัยต่างๆ ที่เราได้ใส่ไว้ในแบบจำลองสามารถอธิบายความแปรเปลี่ยนของค่าตัวแปรผลลัพธ์ได้ดีมากน้อยเพียงใด โดยให้ดูจากค่าที่เรียกว่า R^2 (ซึ่งจะได้ออกมาโดยอัตโนมัติจากโปรแกรมสำเร็จรูปทางสถิติส่วนใหญ่) โดยค่าดังกล่าวนี้จะแสดงสัดส่วนของความแปรปรวนของตัวแปรผลลัพธ์ที่อธิบายได้ด้วยตัวแปรปัจจัยต่างๆ (ในภาพรวม) ที่อยู่ในแบบจำลอง ต่อความแปรปรวนของตัวแปรผลลัพธ์นั้นทั้งหมด หรือ

$$R^2 = \frac{\text{ความแปรปรวนของของตัวแปรผลลัพธ์ที่อธิบายได้ด้วยตัวแปรปัจจัยต่างๆ}}{\text{ความแปรปรวนทั้งหมดของตัวแปรผลลัพธ์}}$$

ซึ่งค่าดังกล่าวนี้จะอยู่ในช่วงระหว่าง 0 ถึง 1 หากแบบจำลองใดมีค่า R^2 สูง (เข้าใกล้ 1) ก็แสดงว่าแบบจำลองของเราสามารถอธิบายความสัมพันธ์ได้ดีกว่าแบบจำลองที่มีค่า R^2 ต่ำกว่า

อย่างไรก็ตาม นักสถิติหลายท่านแนะนำให้ใช้ R^2 กับแบบจำลองที่มีตัวแปรปัจจัยเพียงตัวแปรเดียว หากในแบบจำลองของเรามีตัวแปรปัจจัยตั้งแต่ 2 ตัวขึ้นไป (ดังเช่นกรณีตัวอย่างข้างต้น) ก็ให้ใช้ค่า adjusted R^2 แทน ซึ่งมีหลักในการแปลผลเช่นเดียวกับค่า R^2

เมื่อได้ประเมินแล้วว่าแบบจำลองมีความเหมาะสม (หรือยอมรับได้) สิ่งที่ต้องพิจารณาต่อไปคือการดูว่าตัวแปรปัจจัยแต่ละตัวมีความสัมพันธ์อย่างไรกับตัวแปรผลลัพธ์ ซึ่งต้องอาศัยจากการดูค่า β_i (สัมประสิทธิ์การถดถอย) ของตัวแปรผลลัพธ์แต่ละตัว ซึ่งค่าดังกล่าวนี้เปรียบได้กับค่าความชันของสมการเส้นตรง (ซึ่งก็คือการเปลี่ยนแปลงของค่าตัวแปร y เมื่อค่าตัวแปร x เพิ่มขึ้น 1 หน่วย) ดังนั้นเราจึงสามารถแปลผลกว้างๆ ได้ว่าค่า β_i ของตัวแปรปัจจัยหนึ่ง คือ "ค่าการเปลี่ยนแปลงของตัวแปรผลลัพธ์ เมื่อค่าของตัวแปรปัจจัยนั้นเพิ่มขึ้น 1 หน่วย โดยที่ตัวแปรปัจจัยตัวอื่นๆ ที่เหลือถูกกำหนดให้คงที่" ซึ่งหากค่า β เป็นค่าบวกก็หมายความว่าตัวแปรผลลัพธ์กับตัวแปรปัจจัยเปลี่ยนแปลงไปในทิศทางเดียวกัน แต่หากเป็นค่าลบก็หมายความว่าตัวแปรทั้งสองเปลี่ยนแปลงในทิศตรงข้ามกัน โดยตัวแปรปัจจัยแต่ละชนิดมีความแตกต่างของรายละเอียดในการแปลผล ดังนี้

- กรณีตัวแปรปัจจัยเป็นตัวแปรตัวเลข (เช่น ส่วนสูง หรือ อายุ)

กรณีนี้จะตรงไปตรงมา นั่นก็คือ หากค่า β ของตัวแปรส่วนสูง เท่ากับ 0.2 จะหมายความว่าหากประชากรที่เราศึกษามีส่วนสูงเพิ่มขึ้น 1 เซนติเมตร น้ำหนักโดยเฉลี่ยจะเพิ่มขึ้น 0.2 กิโลกรัม เมื่อปัจจัยอื่นๆ ในแบบจำลองมีค่าคงที่หรือเหมือนกัน

- กรณีตัวแปรปัจจัยเป็นตัวแปรจัดกลุ่มที่มีเพียง 2 กลุ่ม (เช่น เพศ ซึ่งแบ่งเป็นหญิงกับชาย)

กรณีนี้จะขึ้นอยู่กับว่าเรากำหนดค่าให้กับเพศหญิงกับเพศชายในลักษณะใด หากกำหนดค่าเพศหญิงเป็น 0 และเพศชายเป็น 1 หากพบว่า ค่า β ของตัวแปรเพศเท่ากับ 5 ก็หมายความว่า หากประชากรที่เราศึกษาเป็นเพศชายจะมีน้ำหนักโดยเฉลี่ยมากกว่าเพศหญิง 5 กิโลกรัม เมื่อปัจจัยอื่นๆ ในแบบจำลองมีค่าคงที่หรือเหมือนกัน แต่ถ้าข้อมูลชุดเดียวกันนี้ถูกปรับค่าตัวแปรเพศใหม่โดยกำหนดว่าเพศหญิงเป็น 1 และเพศชายเป็น 0 (โดยที่ตัวแปรอื่นๆ ยังคงเหมือนเดิม) เมื่อทำการวิเคราะห์ใหม่เราก็จะพบว่า ค่า β ของตัวแปรเพศจะได้เท่ากับ - 5 (ซึ่งหมายความว่าเพศหญิงจะมีน้ำหนักโดยเฉลี่ยน้อยกว่าเพศชาย 5 กิโลกรัม เมื่อปัจจัยอื่นๆ ในแบบจำลองมีค่าคงที่)

- กรณีตัวแปรปัจจัยเป็นตัวแปรจัดกลุ่มที่มากกว่า 2 กลุ่ม (เช่น เชื้อชาติ ซึ่งสมมติว่าแบ่ง 3 กลุ่ม ได้แก่ ไทย ลาว กัมพูชา)

กรณีที่ตัวแปรปัจจัยมีค่าที่เป็นไปได้ตั้งแต่ 3 กลุ่มขึ้นไปจำเป็นจะต้องสร้างตัวแปรใหม่ขึ้นมา (เรียกว่า dummy variable) เพื่อใช้แทนตัวแปรเดิม โดยจะให้ตัวแปรใหม่นี้เป็นตัวแทนของแต่ละกลุ่ม แต่เนื่องจากผู้วิเคราะห์จะต้องกำหนดว่าจะให้กลุ่มใดกลุ่มหนึ่งเป็นกลุ่มอ้างอิงก่อน (เพื่อที่จะให้กลุ่มที่เหลือมาเป็นกลุ่มเปรียบเทียบ) ดังนั้นตัวแปรที่จะสร้างใหม่ขึ้นมาจะมีจำนวนเท่ากับ จำนวนกลุ่มลบด้วย 1 เช่น ในตัวแปรเชื้อชาติเราแบ่งเป็น 3 กลุ่ม ดังนั้นจะต้องสร้างตัวแปรใหม่ขึ้นมาแทนตัวแปรเชื้อชาติเดิมนี้ เท่ากับ 3-1 หรือ 2 ตัว หากเรากำหนดให้คนไทยเป็นกลุ่มอ้างอิง ตัวแปรใหม่ที่สร้างขึ้นมาก็อาจมีลักษณะดังแสดงในตารางที่ 11

ตารางที่ 11 ตัวอย่างการสร้างตัวแปรใหม่ (dummy variable) สำหรับตัวแปรเชื้อชาติ

เชื้อชาติ (ตัวแปรเดิม)	ลาว (ตัวแปรใหม่ 1)	กัมพูชา (ตัวแปรใหม่ 2)
ไทย	0	0
ลาว	1	0
กัมพูชา	0	1

จะเห็นได้ว่าเมื่อเราให้คนไทยเป็นกลุ่มอ้างอิง เราก็สามารถสร้างตัวแปรใหม่ขึ้นมาอีก 2 ตัว (โดยให้ชื่อว่าลาวและกัมพูชาตามลำดับ) หากตัวอย่างรายใดเป็นคนไทยก็จะมีค่าตัวแปรทั้ง 2 ตัวนี้เป็น 0 หากเป็นคนลาวก็จะมีค่าตัวแปรลาวเป็น 1 และค่าตัวแปรกัมพูชาเป็น 0 ส่วนคนกัมพูชาก็จะมีค่าตัวแปรทั้ง 2 ตรงข้ามกับคนลาว

เมื่อสร้างตัวแปรใหม่ทั้ง 2 ตัวดังกล่าวได้แล้ว เราก็จะนำตัวแปรทั้ง 2 ตัวนี้ไปใส่ในแบบจำลองแทน (ไม่ใส่ตัวแปรเชื้อชาติเดิม) และค่า β ของแต่ละตัวแปรใหม่แต่ละตัวก็คือค่าของตัวแปรกลุ่มนั้นๆ เทียบกับกลุ่มอ้างอิง เช่น หากค่า β ของตัวแปรลาวเท่ากับ -3 ก็หมายความว่า คนลาวจะมีน้ำหนักโดยเฉลี่ยน้อยกว่าคนไทย 3 กิโลกรัม เมื่อปัจจัยอื่นๆ ในแบบจำลองมีค่าคงที่หรือเหมือนกัน เป็นต้น

จะเห็นได้ว่าด้วยวิธีการสร้างตัวแปรในลักษณะนี้เราสามารถที่จะเปรียบเทียบได้เฉพาะระหว่างคนลาวกับคนไทย และระหว่างคนกัมพูชากับคนไทย แต่จะไม่สามารถเปรียบเทียบระหว่างคนลาวกับคนกัมพูชาได้โดยตรง หากต้องการทำการเปรียบเทียบดังกล่าวก็จะต้องสร้างตัวแปรใหม่ในลักษณะที่ให้กลุ่มคนลาวหรือคนกัมพูชาเป็นกลุ่มอ้างอิงแทนและทำการวิเคราะห์ใหม่อีกครั้ง

อย่างไรก็ตาม ค่า β นี้เป็นค่าแสดงความสัมพันธ์ที่ได้จากการศึกษาข้อมูลของกลุ่มตัวอย่าง หากเราต้องการทราบว่าค่าความสัมพันธ์ที่แท้จริงในประชากรเป็นอย่างไร ก็จำเป็นต้องทำการประมาณโดยอาศัยค่าช่วงความเชื่อมั่นหรือทำการทดสอบสมมติฐานและหาค่า p-value ตามที่ได้กล่าวแล้วข้างต้น

การวิเคราะห์การถดถอยโลจิสติก (logistic regression analysis)

หากตัวแปรผลลัพธ์ที่เราสนใจเป็นตัวแปรจัดกลุ่มที่แบ่งเป็น 2 กลุ่ม (binary categorical variable) เช่น ใช่หรือไม่ใช่ ป่วยหรือไม่ป่วย รอดหรือเสียชีวิต ฯลฯ เราสามารถศึกษาความสัมพันธ์ระหว่างตัวแปรปัจจัยต่างๆ กับตัวแปรผลลัพธ์ดังกล่าวโดยใช้วิธีการวิเคราะห์ที่เรียกว่า binary logistic regression (หากแบ่งเป็นมากกว่า 2 กลุ่มก็สามารถใช้ logistic regression ได้เช่นกัน แต่เป็นรูปแบบวิธีที่มีความซับซ้อนขึ้นและจะไม่พูดถึงในที่นี้) ยกตัวอย่างเช่น เราอาจจะสนใจศึกษาความสัมพันธ์ระหว่างการเกิดโรคเบาหวาน (เป็นหรือไม่เป็น) กับ ตัวแปรปัจจัยต่างๆ ได้แก่ เพศ อายุ อาชีพ การมีพ่อแม่เป็นโรคเบาหวาน ฯลฯ

ลักษณะสำคัญของตัวแปรผลลัพธ์ (หรือตัวแปร Y) ในกรณีนี้ก็คือมีค่าที่เป็นไปได้เพียง 2 ค่า คือ เป็นหรือ ไม่เป็น (ซึ่งนิยมแทนด้วยตัวเลข 1 หรือ 0 ตามลำดับ) ซึ่งไม่ยืดหยุ่นเพียงพอที่จะนำมาใส่ไว้ในสมการเส้นตรงในลักษณะเดียวกันกับที่เราใช้ใน linear regression เมื่อจะนำตัวแปรดังกล่าวมาใส่ในแบบจำลองจึงต้องมีการปรับให้ตัวแปรดังกล่าวอยู่ในรูปของค่า natural logarithm ของ odds ของการเป็นโรคแทน โดยที่ odds ของการเป็นโรคมีค่าดังนี้

$$\text{Odds ของการเป็นโรค} = P/(1-P) = \frac{\text{Probability ที่จะเป็นโรค}}{\text{Probability ที่จะไม่เป็นโรค}}$$

จะเห็นได้ว่า odds เป็นรูปแบบหนึ่งของการแสดงความเป็นไปได้ที่จะเป็นโรคเช่นเดียวกับโอกาสหรือความน่าจะเป็น (probability) แต่มีความหมายแตกต่างกัน เช่น ถ้าโอกาสในการเป็นโรคในประชากรกลุ่มหนึ่งเท่ากับ 0.3 (หรือแปลง่ายๆ ว่าหากสุ่มประชากรมา 100 คน น่าจะพบคนเป็นโรค 30 คน) เราก็จะคำนวณ odds ของการเป็นโรคในประชากรกลุ่มดังกล่าวได้เท่ากับ 0.3/(1-0.3) = 0.43 ซึ่งหมายความว่า โอกาสในการเป็นโรคคิดเป็น 0.43 เท่าของโอกาสที่จะไม่เป็นโรค (หรืออาจแปลว่าหากเลือกตัวอย่างมาจำนวนหนึ่งโดยการสุ่มจากประชากรกลุ่มนี้ จะพบว่าอัตราส่วนของจำนวนผู้ที่เป็นโรคต่อจำนวนผู้ที่ไม่เป็นโรคในกลุ่มตัวอย่างที่ได้น่าจะมีค่าประมาณ 0.43)

เมื่อทำการปรับค่าตัวแปรผลลัพธ์ให้เป็น natural logarithm ของ odds หรือ ln(Odds) ของการเป็นโรคแล้ว จะทำให้เราสามารถสร้างแบบจำลองที่มีลักษณะคล้ายคลึงกับแบบจำลองของ linear regression ได้ การนำ ln(Odds) ของการเป็นโรคมาใส่ในแบบจำลอง (เรียกว่า logit model) สามารถแสดงได้ดังสมการต่อไปนี้

$$\ln(\text{Odds}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ...+ \beta_k x_k + \varepsilon$$

โดย ln(Odds) คือ Natural logarithm ของ Odds ของการเป็นโรค

ถ้ามองจากแง่มุมของผู้ปฏิบัติ (ซึ่งเป็นผู้กำหนดตัวแปรปัจจัยที่จะใส่ในแบบจำลองและใช้โปรแกรมสำเร็จรูปในการวิเคราะห์) วิธีการได้มาซึ่งค่า β ต่างๆ ในแบบจำลองนี้ดูคล้ายคลึงกับการวิเคราะห์การถดถอยเชิงเส้น แต่อันที่จริงแล้ววิธีการทางคณิตศาสตร์ที่อยู่เบื้องหลังของการวิเคราะห์ 2 แบบนี้มีความแตกต่างกันอย่างมาก

ความแตกต่างที่สำคัญในการแปลผลการวิเคราะห์ 2 อย่างที่ผู้ปฏิบัติต้องทราบคือ ในการวิเคราะห์การถดถอยโลจิสติกส์นั้น การใช้ค่า β ตรงๆ ของตัวแปรปัจจัยแต่ละตัวในการแปลผลจะมีประโยชน์น้อย (แปลตรงๆ ได้ว่า คือ ค่า ln(Odds) ที่เปลี่ยนไปหากค่าตัวแปรปัจจัยนั้นๆ เพิ่มขึ้น 1 หน่วย โดยที่ปัจจัยอื่นๆ ในแบบจำลองไม่เปลี่ยนแปลง) แต่ด้วยรูปแบบของแบบจำลองในลักษณะดังกล่าวนี้ หากเราทำการใส่ anti-logarithm ให้กับค่า β แล้ว ค่าที่ได้ออกมาจะเป็น ค่า Odds Ratio (OR) ของตัวแปรปัจจัยนั้นๆ ทันที (ดูรายละเอียดได้ในตำราสถิติทั่วไป) หรือสามารถเขียนแสดงได้ดังนี้

$$e^{\beta \text{ ของตัวแปรปัจจัยใดๆ}} = \text{Odds Ratio ของตัวแปรปัจจัยนั้น}$$

ดังนั้นโปรแกรมสำเร็จรูปหลายโปรแกรมจึงนิยมที่จะแสดงผลการวิเคราะห์ออกมาในรูปของ odds ratio แทนที่จะแสดงเป็นค่า β (หรืออาจจะเลือกให้แสดงทั้ง 2 ค่าก็ได้)

ตัวอย่างการแปลผลค่า Odds Ratio ที่ได้จากกรณีตัวอย่างข้างต้น เช่น

- กรณีตัวแปรปัจจัยเป็นตัวแปรตัวเลข เช่น อายุ (ปี) หากได้ค่า OR ของตัวแปรอายุเท่ากับ 1.2 ก็หมายความว่า odds ของการเป็นโรคเบาหวานในคนที่อายุหนึ่งๆ โดยเฉลี่ยแล้วจะมีค่าเป็น 1.2 เท่าของ odds ของการเป็นโรคในคนที่มีอายุน้อยกว่านั้น 1 ปี (โอกาสเป็นโรคสูงขึ้นเล็กน้อยเมื่ออายุมากขึ้น 1 ปี) โดยที่ถือว่าปัจจัยอื่นๆ ในแบบจำลองมีค่าคงที่หรือเหมือนกัน
- กรณีตัวแปรปัจจัยเป็นตัวแปรจัดกลุ่ม เช่น เพศ (ชาย/หญิง) โดยสมมติให้ ชาย = 1 และ หญิง = 0 หากได้ค่า OR ของตัวแปร เพศเท่ากับ 0.85 ก็หมายความว่า Odds ของการเป็นโรคเบาหวานในผู้ชายโดยเฉลี่ยแล้วจะมีค่าเป็น 0.85 เท่าของ Odds ของการเป็นโรคในผู้หญิง (โอกาสเป็นโรคในผู้ชายต่ำกว่าในผู้หญิง) โดยที่ถือว่าปัจจัยอื่นๆ ในแบบจำลองมีค่าคงที่หรือเหมือนกัน

สรุป

แม้จะรู้ว่าสถิติเป็นเครื่องมือที่มีประโยชน์อย่างมากในการทำงาน แต่นักระบาดวิทยาหลายท่านก็ยังรู้สึกว่าวิชาสถิติเป็นเรื่องลึกลับที่ยากเกินกว่าจะทำความเข้าใจได้ ซึ่งก็เป็นความจริงในกรณีที่เกี่ยวข้องกับการใช้วิธีการวิเคราะห์ทางสถิติขั้นสูง อย่างไรก็ตาม เทคนิคทางสถิติขั้นพื้นฐานส่วนใหญ่ที่ได้กล่าวถึงในเนื้อหาบทนี้ เป็นสิ่งที่ไม่ยากเกินกว่าจะทำความเข้าใจ และหากนักระบาดวิทยานำไปใช้ได้อย่างถูกต้องก็จะเป็นประโยชน์อย่างมาก เนื้อหาในบทนี้ไม่ได้ตั้งใจที่จะทำให้ผู้อ่านที่ไม่เคยผ่านการศึกษาวิชาสถิติอย่างเป็นระบบสามารถใช้เทคนิคหรือเครื่องมือเหล่านี้ได้เหมือนกับผู้เชี่ยวชาญ แต่หวังว่าจะเป็นจุดเริ่มต้นที่จะทำให้พอเห็นภาพว่าแนวคิดกว้างๆ ที่อยู่เบื้องหลังการทำงานของเครื่องมือทางสถิติ (อย่างน้อยก็ชนิดที่ได้กล่าวถึงในเนื้อหาบทนี้) เป็นอย่างไร หรือเกิดความสนใจและศึกษาค้นคว้าเพิ่มเติมจากผู้เชี่ยวชาญ ตำรา หรือสื่อการเรียนรู้ต่างๆ

สิ่งหนึ่งที่ต้องคำนึงถึงอยู่เสมอในการใช้เครื่องมือทางสถิติก็คือ วิชาสถิติไม่ใช่สิ่งมหัศจรรย์ที่จะช่วยเปลี่ยนให้ข้อมูลที่ถูกเก็บมาอย่างไม่มีคุณภาพให้กลายเป็นผลการวิเคราะห์ที่ดีเยี่ยมอย่างที่เราต้องการได้ ข้อมูลที่มาจากวิธีการศึกษาที่เหมาะสม เก็บรวมรวมมาอย่างดีและมีอคติน้อยเท่านั้นจึงจะนำมาซึ่งผลการวิเคราะห์ที่มีคุณภาพ อย่างไรก็ตามแม้ว่าข้อมูลที่มีอยู่จะมีคุณภาพแล้ว แต่ผลการวิเคราะห์อาจจะไม่ถูกต้องก็ได้อันเนื่องมาจากใช้วิธีการทางสถิติที่ไม่เหมาะสม ดังที่เราจะเห็นได้ว่าเครื่องมือทางสถิติที่ได้กล่าวถึงในบทนี้แต่ละชนิด หากจะนำไปใช้ได้อย่างถูกต้องนั้น จำเป็นจะต้องอาศัยสมมติฐานหรือเงื่อนไขหลายประการที่จะต้องมีขึ้นก่อน และถึงแม้ว่าเงื่อนไขดังกล่าวจะเกิดขึ้นครบถ้วนหรือยอมรับได้แล้วก็ตาม เครื่องมือทางสถิติแต่ละชิ้นก็ล้วนมีจุดดีจุดด้อยแตกต่างกันไป ผู้ที่จะใช้เครื่องมือเหล่านี้จึงต้องมีความเข้าใจในเทคนิคการวิเคราะห์ที่ตนใช้เป็นอย่างดี ในปัจจุบันที่การวิเคราะห์ทางสถิติทั้งขั้นพื้นฐานและขั้นสูงสามารถทำได้อย่างสะดวกรวดเร็วด้วยโปรแกรมสำเร็จรูป โดยผู้วิเคราะห์ไม่มีความจำเป็นที่จะต้องเข้าใจวิธีการคำนวณทางคณิตศาสตร์ในรายละเอียดลึกๆ อีกต่อไป จึงมีความจำเป็นอย่างยิ่งที่นักระบาดวิทยาจะต้องทำความเข้าใจถึงเงื่อนไขหรือข้อกำหนดของเทคนิคแต่ละอย่างและวิธีการแปลผลอย่างถูกต้องดังที่ได้กล่าวมาแล้วข้างต้น ทั้งในฐานะของผู้ผลิตและผู้บริโภคผลการวิเคราะห์ทางสถิติ หากขาดซึ่งความเข้าใจและไม่สามารถเลือกใช้เครื่องมือทางสถิติได้อย่างมีวิจารณญาณ (หรือเรียกว่าไม่มี statistical literacy) แล้ว นักระบาดวิทยาก็อาจจะนำผลการวิเคราะห์ไปใช้อย่างไม่ถูกต้องโดยไม่รู้ตัว และผู้รับผลเสียนั้นก็คือประชาชนหรือชุมชนที่เราตั้งใจจะนำผลการศึกษานั้นไปใช้ประโยชน์

บรรณานุกรม

1. ศิริชัย วงศ์วัฒนไพบูลย์. หลุมพรางในการนำเสนอข้อมูลข่าวสารทางระบาดวิทยา. นนทบุรี: สำนักระบาดวิทยา; 2550.
2. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. Oxford: Blackwell; 2002.
3. Bland M. An introduction to medical statistics. 3rd ed. Oxford: Oxford University Press; 2002.
4 Kirkwood B, Sterne J. Essential of medical statistics. 2nd ed. Oxford: Wiley-Blackwell; 2001.
5. Lang TA, Secic M. How to report statistics in medicine. 2nd ed. Philadelphia: American College of Physicians; 2006.
6. Rosner B. Fundamentals of biostatistics. 7th ed. Boston: Brooks/cole, Cengage Learning; 2010.