

สรุปสอนเสริมครั้งที่ 2

วิชา 50103 สถิติและการวิจัยสำหรับวิทยาศาสตร์สุขภาพ หน่วยที่ 6 - 10

วิเคราะห์แบบประเมินก่อนเรียน สอนเสริมครั้งที่ 2

1. เครื่องมือที่ใช้วัดแล้วได้ผลการวัดตรงกันทุกครั้ง เรียกว่าเครื่องมือที่มีคุณสมบัติอย่างไร

ตอบ ความเที่ยงหรือความเชื่อมั่น ข้อนี้จะมีคำตอบใกล้เคียง ให้ตัดสินใจ 2 ข้อ คือ ความเที่ยงกับความตรง ข้อแตกต่างคือ เปรียบเหมือน เราปลูกดอกไม้ใส่เป้า ความเที่ยงคือ ปลูกดอกไม้กี่ครั้งโดนเป้าทุกครั้ง หรือตำแหน่งเดียวกันทุกครั้ง แต่ไม่ตรงจุดที่เราต้องการคือจุดตรงกลางเป้า

ส่วนความตรงปากี่ครั้งก็ตรง กลางเป้า เพราะฉะนั้น ความเที่ยงอาจจะไม่มีความตรง แต่ความตรงต้องมีความเที่ยงแน่ๆ

2. การหาความเที่ยงด้วยสูตรคูเดอร์- ริชาร์ดสัน เหมาะกับเครื่องมือที่มีลักษณะอย่างไร

ตอบ แบบสอบถาม แบบเลือกตอบ 1 ข้อ ที่มีการให้คะแนนเป็น 0 และ 1

ข้อนี้ เราต้องใช้ความจำครับ ว่า ความเที่ยงมีวิธีการหาที่ประเภท และคุณลักษณะเด่นของวิธีนั้นเป็นอย่างไร

3. การจัดการข้อมูลที่ดีทำให้เกิดผลได้อย่างไร

ตอบ วิเคราะห์ได้รวดเร็ว ถูกต้อง ครบถ้วน ประหยัด

ข้อนี้ให้ดูในหัวข้อประโยชน์ของการจัดการข้อมูล

4. สื่อที่ใช้กับคอมพิวเตอร์ในการบันทึกข้อมูลมีดังต่อไปนี้ ยกเว้นข้อใด

ตอบ แบบบันทึกข้อมูล ข้อนี้ ต้องดูว่ามีอะไรที่ไม่เกี่ยวกับคอมพิวเตอร์ในการบันทึกข้อมูล หรืออุปกรณ์บันทึกข้อมูล เกี่ยวข้องกับคอมพิวเตอร์มีอะไรบ้าง

5. จากข้อมูลต่อไปนี้ จงหาพิสัยของข้อมูล 100,121,99,84,111,118,120,109,130

ตอบ 48

ข้อนี้ ไม่ยากแต่เราต้องจำให้ขึ้นใจเลยว่า พิสัยหาแบบใด

สูตรหาพิสัย = ค่าสูงสุด - ค่าต่ำสุด $130-82 = 48$

6. การหาความสัมพันธ์โดยใช้ Pearson's product moment correlation เป็นการหาความสัมพันธ์ของข้อมูลที่มีลักษณะแบบใด

ตอบ ตัวแปรอิสระเป็นข้อมูลปริมาณ ตัวแปรตามเป็นข้อมูลปริมาณ

ตัวแปรอิสระคือตัวแปรที่ส่งผลต่อตัวแปรตาม เช่น ค่าใช้จ่ายต่อการโฆษณา กับยอดขาย เมื่อมีการโฆษณามาก ก็จะส่งผลต่อยอดขายสินค้าสูงขึ้น ตัวแปรอิสระคือค่าใช้จ่ายโฆษณา ตัวแปรตามคือยอดขาย

7. ทอดลูกเต๋าเที่ยงตรง 2 ลูก ความน่าจะเป็นที่ได้แต้มมีค่ารวมกัน 7 มีค่าเท่ากับเท่าใด

ตอบ 1/16 ข้อสอบแบบนี้ต้องพบแน่ครับในการสอบครั้งนี้

วิธีทำ สูตรคือเหตุการณ์ที่เราสนใจหารด้วยเหตุการณ์ที่จะเกิดขึ้นหรือเป็นไปได้ทั้งหมด

สิ่งที่เราสนใจหรือต้องการคือการจับคู่ของลูกเต๋า 2 ลูกที่มีแต้มรวมกันได้ 7 ได้แก่ (1,6) (2,5) (3,4) (4,3) (5,2) (6,1) มีทั้งหมด 6 คู่ เหตุการณ์ที่เป็นไปได้ ลูกเต๋ามี 6 หน้า $6 \times 6 = 36$ ครั้ง แทนที่ในสูตร $6/36$ ตัดกันให้เหลือตัวเลขน้อยเท่ากับ $1/6$

8. จากสถิติการคลอดที่โรงพยาบาลแห่งหนึ่งพบว่าใน 1 ปี มีผู้มาคลอด 12,000 ราย โดยเฉลี่ยพบว่าเป็นเด็กตายคลอดร้อยละ 0.1 จงหาความแปรปรวนของจำนวนเด็กตายคลอดต่อเดือน

ตอบ 1

เป็นการแจกแจงปิงซง ค่าเฉลี่ย = ความแปรปรวน = np

ผู้มาคลอดต่อเดือน = $12000/12 = 1000$ คน

$$= 1000 \times (0.1/100) = 1$$

9. การประมาณค่าแบ่งได้ 2 ประเภทคือ

ตอบ การประมาณค่าเป็นจุด และการประมาณค่าเป็นช่วง

10. ถ้าสมมุติฐานการวิจัยสงสัยว่าจะมีความมากกว่ากันจะเป็นการทดสอบตามข้อใด

ตอบ การทดสอบข้างเดียว

ความหมายและตัวอย่างของเครื่องมือวิจัย

เครื่องมือวิจัยเป็นอุปกรณ์หรือเทคนิคที่นักวิจัยใช้ในการเก็บรวบรวมข้อมูล หรือวัดตัวแปรที่ต้องการศึกษา

ชนิดของเครื่องมือวิจัย	ตัวอย่าง
1. เครื่องมือวัดทางสรีรวิทยาและเครื่องมือวัดทางวิทยาศาสตร์อื่นๆ	1. เครื่องมือวัดทางกายภาพ เช่น เครื่องมือวัดความดัน 2. เครื่องมือวัดทางเคมี เช่น เครื่องมือวัดระดับน้ำตาลในเลือด 3. เครื่องมือวัดจุลชีววะ เช่น เครื่องมือตรวจแบทที่เรีย
2. แบบสอบถาม	1. แบบสอบถามชนิดเลือกตอบ 2. แบบประมาณค่า 3. แบบสอบถามชนิดปลายเปิด
3. แบบสัมภาษณ์	1. แบบสัมภาษณ์ชนิดมีโครงสร้างแน่นอน 2. แบบสัมภาษณ์ชนิดไม่มีโครงสร้างไม่แน่นอน
4. แบบสังเกต	1. แบบสังเกตชนิดมีโครงสร้าง 2. แบบสังเกตชนิดไม่มีโครงสร้าง
5. แบบทดสอบมาตรฐาน	1. แบบวัดความถนัดทางคณิตศาสตร์ 2. แบบวัดบุคลิกภาพ 3. แบบทดสอบความสามารถทางสติปัญญา

ดูข้อดี ข้อจำกัดของเครื่องมือชนิดต่างๆ หน้า 272-273

คุณภาพของเครื่องมือวิจัย

1. **ความตรง หรือ ความสมเหตุสมผล (validity)** เครื่องมือที่ดีจะต้องตรงในการวัด หมายถึง เครื่องมือนั้นจะต้องวัดในสิ่งที่ตรงกับที่ผู้วิจัยต้องการวัด ได้อย่างถูกต้องและครบถ้วนความตรงของเครื่องมือวัดแบ่งได้เป็น 3 ประเภท คือ

1.1 **ความตรงตามเนื้อหา (content validity)** หมายถึงเครื่องมือสามารถวัดได้ตรงตามเนื้อหาที่ต้องการวัด เช่น เครื่องมือวัดความพึงพอใจของผู้ป่วยความพึงพอใจของผู้ป่วยต่อการบริการรักษาพยาบาลต้องสามารถวัดความพึงพอใจของผู้ป่วยจริง

1.2 ความตรงเชิงสัมพันธ์กับเกณฑ์ (criterion-related validity) เป็นการหาความตรงของเครื่องมือโดยการหาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากการนำเครื่องมือกับเกณฑ์มาตรฐานไปทดลองใช้ แบ่งเป็น 2 ประเภท

- ความตรงตามสภาพการณ์ เป็นการหาความสัมพันธ์ระหว่างเครื่องมือกับเกณฑ์มาตรฐาน โดยทดลองใช้ในเวลาเดียวกัน เช่น เครื่องมือวัดความวิตกกังวลสูง ในขณะที่เดียวกันเมื่อใช้เครื่องมือมาตรฐานอื่นวัดก็ให้ผลมีความวิตกกังวลสูงเช่นกัน แสดงว่าเครื่องมือมีความตรงตามสภาพการณ์

- ความตรงเชิงพยากรณ์ เป็นการหาความสัมพันธ์ระหว่างเครื่องมือกับเกณฑ์มาตรฐานภายหลังจากนำเครื่องมือไปทดลองใช้ เช่น เครื่องมือที่ทำนายอาการซึมเศร้าที่สร้างขึ้นกำหนดว่าถ้าได้คะแนนสูงกว่า 20 คะแนนถือว่า มีโอกาสเกิดอาการซึมเศร้า เมื่อนำเครื่องมือไปวัดสตรีหลังคลอด 1 วันคนหนึ่งได้ 22 คะแนน และปรากฏว่าสตรีดังกล่าวมีอาการซึมเศร้าจริงเมื่อหลังคลอด 6 สัปดาห์ แสดงว่าเครื่องมือนี้มีความตรงเชิงพยากรณ์

1.3 ความตรงตามโครงสร้าง (construct validity) เป็นลักษณะของเครื่องมือที่มีรูปแบบหรือ โครงสร้างตามทฤษฎีที่ควรจะเป็น ในการวัด มักใช้กับเครื่องมือที่มีลักษณะเป็นแบบวัดทางจิตวิทยา หรือแบบวัดที่มีหลายมิติ เช่น วัดคุณภาพชีวิต แบบสอบถาม พฤติกรรมการส่งเสริมสุขภาพ แบบสอบถามการสนับสนุนทางสังคม ที่นิยมได้แก่

- การวิเคราะห์โดยยึดทฤษฎีเป็นเกณฑ์
- การหาความตรงตามโครงสร้างโดยใช้การวิเคราะห์ทางสถิติขั้นสูง
- การนำเครื่องมือไปหาความสัมพันธ์กับเครื่องมือมาตรฐานที่มีความตรงตามโครงสร้าง

2. ความเที่ยงหรือความเชื่อมั่น (reliability)

1. การหาความเที่ยงจากความคงที่ (stability) เป็นความเที่ยงจากความคงที่ในการวัดในเวลาต่างกัน หากนำเครื่องมือไปวัดซ้ำ ในเวลาที่ต่างกัน ผลที่ได้จะใกล้เคียงกันหรือคงที่เพียงใด โดยการหาสัมประสิทธิ์ความเที่ยง หรือเรียกอีกชื่อหนึ่งว่าวิธีการ ทดสอบซ้ำ วิธีโดยการนำเครื่องมือไปวัดในกลุ่มตัวอย่างเดียวกันสองครั้งในเวลาแตกต่างกัน แล้วหาสัมประสิทธิ์สหสัมพันธ์ แบบเพียร์สัน

ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีค่าระหว่าง -1.00 ถึง 1.00 ยิ่งค่าใกล้ 1 เท่าใดแสดงว่ายิ่งมีความสัมพันธ์กันสูง โดยทั่วไปเครื่องมือเครื่องมือมีความเที่ยงในระดับที่พอใจเมื่อค่าความเที่ยงสูงกว่า 0.70

2. การหาความเที่ยงจากความสม่ำเสมอภายใน (internal consistency) เป็นความเที่ยงที่แสดงถึงความสม่ำเสมอของข้อความ ทั้งหมดในแบบวัดว่าสามารถวัดในเรื่องเดียวกันได้มากน้อยเพียงใด วิธีการหาความเที่ยงชนิดนี้ทำได้โดยการนำเครื่องมือไป ทดสอบวัดเพียงครั้งเดียว แล้วนำมาวิเคราะห์ค่าความเที่ยง วิธีที่นิยมใช้ 3 วิธีคือ

- การคำนวณแบบแบ่งครึ่ง (split – half method) แบ่งครึ่งแบบวัดเป็น 2 ส่วนแล้วหาความสัมพันธ์ (ตัวอย่างหน้า 279)
- การคำนวณแบบสัมประสิทธิ์แอลฟา (Alpha Coefficient) ใช้ได้ทั้งแบบเลือกตอบและมาตราส่วนประมาณค่า
- การคำนวณโดยใช้สูตรของคูเดอร์ ริชาร์ดสัน (Kuder- Richardson) ใช้กับแบบสอบถามที่มีการให้คะแนนเป็น 0 และ 1 แบบ KR-20 จะให้ความเที่ยงสูงกว่า KR-21
- การหาความเที่ยงจากผู้สังเกตมากกว่า 1 คน ใช้กับเครื่องมือที่เป็นแบบสังเกตนำไปใช้หลายคน

ความเที่ยงของการสังเกต = $\frac{\text{จำนวนที่สังเกตเหมือนกัน}}{\text{จำนวนที่สังเกตเหมือนกัน} + \text{จำนวนที่สังเกตต่างกัน}}$

- การหาความเที่ยงของความคล้ายคลึงกัน เป็นการหาความสัมพันธ์ของแบบวัดคู่ขนาน

3. ความยากง่าย (difficulties) เครื่องมือรวบรวมข้อมูลโดยเฉพาะแบบทดสอบจะต้องมีความยากง่ายพอเหมาะกับผู้ตอบ เพราะถ้ามีความยากมากเกินไป ผู้ตอบตอบผิดหมด เครื่องมือชิ้นนั้นจะไม่สามารถวัดสิ่งที่ต้องการได้เลย

วิธีคำนวณความยากง่ายมักใช้กับแบบทดสอบ โดยคำนวณจากสัดส่วนของผู้ตอบถูกต้องกับจำนวนผู้ตอบทั้งหมดในข้อคำถามข้อนั้น เช่น มีคนตอบถูก 10 คน จากจำนวนผู้ตอบทั้งหมด 20 คน ข้อคำถามข้อนั้นจะมีค่าความยากง่ายเท่ากับ .50 **ถ้ามีค่าเข้าใกล้ 1 แปลว่าง่าย ถ้าเข้าใกล้ 0 แปลว่ายาก**

4. อำนาจจำแนก (discrimination power) เครื่องมือที่ดีจะต้องมีอำนาจจำแนกสูงในการแบ่งแยกสิ่งที่ต้องการวัดออกมา เช่น แยกผู้มีความรู้เรื่องการออกกำลังกาย ออกจากผู้ที่ไม่มีความรู้เรื่องการออกกำลังกาย คนตอบถูกจะต้องมีความรู้เรื่องนั้นจริงๆ และคนที่ไม่มีความรู้ก็ควรจะต้องผิด

การหาอำนาจจำแนกทำได้โดยการแบ่งกลุ่มผู้ตอบออกเป็น 2 กลุ่ม คือกลุ่มที่ได้คะแนนสูง และกลุ่มที่ได้คะแนนต่ำ ค่าอำนาจจำแนกจะแทนด้วยตัวอักษร r และมีค่าระหว่าง -1.0 จนถึง +1.0 **ถ้ามีค่าเข้าใกล้ 0 แปลว่ามีอำนาจจำแนกน้อย ถ้าเข้าใกล้ 1 แสดงว่ามีอำนาจจำแนกมาก ข้อคำถามที่ดีต้องมีอำนาจจำแนก 0.2 ขึ้นไป และยิ่งค่ามากยิ่งดี**

5. ความเป็นปรนัย (objectivity) เครื่องมือรวบรวมข้อมูลที่ดีจะต้องมีความเป็นปรนัยสูง กล่าวคือทุกคนเมื่ออ่านหรือใช้เครื่องมือชิ้นนั้นแล้วจะเข้าใจความหมายได้ตรงกันเสมอไม่ว่าจะอ่านเวลาใดก็ตาม รวมทั้งการแปลผลออกมาเป็นคะแนนจะต้องมีเกณฑ์ที่แน่นอนไม่ว่าใครจะเป็นผู้แปลผลจะต้องได้ค่าคะแนนตรงกันเสมอ

6. ความหมายในการวัด (meaningfulness) เครื่องมือที่ดีจะต้องถามหรือวัดในสิ่งที่สอดคล้องกับความเป็นจริงมากที่สุด เช่น ในการถามระดับการศึกษา แล้วให้ผู้ตอบเลือกว่า ประถมศึกษา มัธยมศึกษา ซึ่งมีความหมายแตกต่างกันในระดับเดียวกัน ถ้าต้องการวิเคราะห์ให้ละเอียด อาจมีการแยกชั้นปีที่ศึกษาด้วย

7. ความสามารถในการนำไปใช้ (usability) เครื่องมือที่ดีจะต้องสามารถนำไปใช้ได้โดยง่าย ไม่ยุ่งยากซับซ้อน ใช้เวลาเหมาะสมในการนำไปใช้ การแปลผลง่ายไม่ยุ่งยาก และยุติธรรม คุ่มค่ากับแรงงานและงบประมาณที่เสียไป

การพัฒนาเครื่องมือเพื่อใช้ในการวิจัย

ในขั้นตอนของการเก็บรวบรวมข้อมูลเพื่อการวิจัย ผู้วิจัยจะต้องพิจารณาว่าในการวิจัยของตนเองต้องการวัดตัวแปรอะไรบ้าง และจะใช้เครื่องมือชนิดใดเพื่อเก็บรวบรวมข้อมูลนั้น เครื่องมือที่ใช้ในการวิจัยอาจได้มาจากคัดเลือกเครื่องมือที่มีอยู่แล้วมาใช้ หรือพัฒนาขึ้นใหม่ ดังนี้

การเลือกเครื่องมือที่มีอยู่แล้วมาใช้ในการวิจัย ผู้วิจัยจำเป็นต้องพิจารณาในประเด็นต่อไปนี้

- 1. เครื่องมือชิ้นนั้นสร้างขึ้นเพื่อวัดอะไร** มีความสอดคล้องกับเรื่องที่จะวิจัยหรือไม่
- 2. เครื่องมือชิ้นนั้นสร้างขึ้นภายใต้กรอบแนวคิดหรือทฤษฎีอะไร** ทฤษฎีแต่ละทฤษฎีอาจให้แนวคิดที่แตกต่างกันดังนั้นในการทบทวนวรรณกรรมผู้วิจัยจึงต้องศึกษาที่ผู้สร้างเครื่องมือใช้
- 3. เครื่องมือชิ้นนั้นสร้างขึ้นเพื่อใช้กับคนกลุ่มใด** เพราะคนแต่ละกลุ่มอาจมีความแตกต่างกันทั้งในเรื่องของวัยสภาวะด้านร่างกายและจิตใจ อารมณ์ สังคม
- 4. เนื้อหาหรือคำถามที่อยู่ในเครื่องมือชิ้นนั้นยังคงทันสมัยอยู่หรือไม่** และข้อคำถามย่อยภายในสอดคล้องกันเพียงพอหรือไม่ เครื่องมือบางฉบับสร้างขึ้นเป็นเวลานานเกินไปอาจล้าสมัยหรือไม่สอดคล้องกับสถานการณ์
- 5. การนำเครื่องมือไปใช้ในการเก็บรวบรวมข้อมูล** เครื่องมือมีอบางชนิดอาจต้องอาศัยผู้ที่ได้รับการฝึกฝนมาโดยเฉพาะในการเก็บข้อมูล หรือการแปลผล เช่น เครื่องมือวัดทางจิตวิทยาหรือพัฒนาการ

6. คุณภาพของเครื่องมือ ได้แก่ ความตรง ความเที่ยง และคุณสมบัติอื่นสำหรับเครื่องมือที่ดีควรอยู่ในระดับที่ยอมรับได้ เพราะการนำเครื่องมือที่ด้อยคุณภาพมาใช้จะมีผลเสียต่องานวิจัยทั้งหมด

7. ความเหมาะสมในด้านของระยะเวลาและค่าใช้จ่ายในการนำเครื่องมือมาใช้ในการเก็บข้อมูล เครื่องมือมาตรฐานบางฉบับมีคุณภาพดี แต่ต้องเสียค่าใช้จ่ายแก่เจ้าของลิขสิทธิ์ในราคาที่แพงมาก เครื่องมือที่มีความยุ่งยากซับซ้อนในการนำไปใช้ อาจไม่คุ้มค่าแก่การนำไปใช้

8. ลิขสิทธิ์ของเครื่องมือ การพัฒนาเครื่องมือแต่ละฉบับถือเป็นทรัพย์สินทางปัญญา การละเมิดลิขสิทธิ์นอกจากมีความผิดตามกฎหมายแล้ว ยังอาจมีผลทำให้ไม่สามารถเผยแพร่ผลงานวิจัยได้ โดยเฉพาะกรณีที่เป็นเครื่องมือที่พัฒนาในต่างประเทศ

การสร้างเครื่องมือใหม่ มักใช้ในกรณีที่มีแนวคิดที่ต้องการวัดไม่เคยมีผู้ศึกษามาก่อน หรือไม่มีเครื่องมือที่เหมาะสมที่จะใช้ในการวิจัย การสร้างเครื่องมือใหม่ มีขั้นตอนดังนี้

1. การระบุแนวคิดหรือสิ่งที่ต้องการวัดให้ชัดเจน โดยการวิเคราะห์ห้แนวคิดที่จะศึกษา (concept analysis) ซึ่งจะช่วยให้ทราบว่ามีแนวคิดนั้นมีขอบเขตและองค์ประกอบภายในอย่างไร และนำมาใช้นิยามเชิงแนวคิด (conceptual definition) และนิยามเชิงปฏิบัติการ (operational definition) เพื่อให้สามารถวัดได้ในทางปฏิบัติด้วย เช่น ระดับการศึกษา หมายถึง ระดับการศึกษาสูงสุด โดยนับเป็นจำนวนปีที่ได้รับการศึกษาทั้งหมดในระบบ

2. การสร้างข้อคำถาม การสร้างข้อคำถามจะเกิดขึ้นได้ก็ต่อเมื่อมีการวิเคราะห์แนวคิดจนได้ความหมายขอบเขต และองค์ประกอบที่ชัดเจน ในการสร้างข้อคำถามที่ดีผู้วิจัยจะต้องยึดหลักที่สำคัญ ดังนี้

- คำถามควรมีลักษณะสั้น กระชับ ชัดเจน ใช้ภาษาเข้าใจง่าย
- คำถามต้องมีเนื้อหาเกี่ยวข้องกับที่จะศึกษา ไม่นอกประเด็น
- หลีกเลี่ยงการสร้างคำถามที่มีเนื้อหาหลายๆ คำถาม
- หลีกเลี่ยงคำถามที่ก่อให้เกิดความอับอาย เสื่อมศักดิ์ศรี
- หลีกเลี่ยงคำถามนำ ทำให้ผู้ตอบมีใจเอนเอียง
- คำตอบของคำถามปลายปิดจะต้องครอบคลุมทุกคำตอบ และในบางข้ออาจต้องเพิ่มคำตอบปลายเปิด เช่น อื่นๆ โปรดระบุไว้

3. การกำหนดรูปแบบของการวัด โดยพิจารณาจากจากคำนิยามเชิงปฏิบัติการของมโนคติหรือตัวแปรนั้น ๆ จึงตัดสินใจเลือกหลายรูปแบบมี ดังนี้

1. แบบลิเคิร์ต (Likert scale) มีสเกล 5 ระดับ
2. แบบออสกูด มีสเกล 7 ระดับ
3. แบบตอบถูกได้ 1
4. การจัดเกลาและปรับปรุงแก้ไขแบบวัด
5. การนำแบบวัดไปให้ผู้เชี่ยวชาญพิจารณา
6. การนำเครื่องมือไปทดลองใช้
7. การแก้ไขและจัดทำเครื่องมือฉบับสมบูรณ์

การจัดการข้อมูลการวิจัย

การจัดการข้อมูล หมายถึง การนำข้อมูลในแบบสอบถามหรือแบบทดสอบ หรือแบบบันทึกข้อมูลที่รวบรวมมาได้ มาสร้างรหัสเป็นตัวเลขตามความต้องการเพื่อการวิเคราะห์ด้วยมือหรือด้วยคอมพิวเตอร์

ความสำคัญของการจัดการข้อมูล คือ การจัดการข้อมูลที่ช่วยให้การวิเคราะห์ข้อมูลเกิดความรวดเร็ว ถูกต้อง ครบถ้วน ประหยัดเวลาและงบประมาณ

1. การเตรียมข้อมูล (data preparation) มีขั้นตอนดังนี้

1. การตรวจสอบแก้ไขข้อมูลเบื้องต้น หรือการบรรณาธิการ
2. การให้รหัสข้อมูล
3. การบันทึกข้อมูล
4. การตรวจสอบข้อมูลแก้ไขรหัสข้อมูลครั้งสุดท้าย

2. การตรวจสอบแก้ไขข้อมูลเบื้องต้น

1. ความครบถ้วนสมบูรณ์ (completeness) ตรวจสอบความเรียบร้อยสมบูรณ์ของคำถามทั้งหมดที่ได้รับมา ข้อใดไม่ตอบต้องมีเหตุผลว่าทำไมไม่ตอบ เช่น ท่านเคยรับประทานอาหารที่โรงพยาบาลแห่งนี้หรือไม่ ถ้าตอบว่าไม่เคย ต้องไม่ตอบข้อต่อไปว่าอาหารที่นี้รสชาติเป็นอย่างไร

2. ความสอดคล้องของคำตอบ (consistency) คือการตรวจสอบคำถามและคำตอบที่ได้รับมาต้องมีความสอดคล้องกัน เช่น คำถามเกี่ยวกับความสูง แต่คำตอบที่ได้รับเป็นเรื่องเกี่ยวกับน้ำหนัก

3. ความสอดคล้องกับวัตถุประสงค์ คำตอบสอดคล้องกับวัตถุประสงค์หรือไม่ การวิจัย

การให้รหัสข้อมูล การบันทึกข้อมูล และการตรวจสอบแก้ไขรหัสข้อมูลครั้งสุดท้าย

1. การให้รหัสข้อมูล

ข้อมูลที่ได้จากการเก็บรวบรวมเรียกว่าข้อมูลดิบ (raw data) เมื่อผ่านการตรวจสอบในขั้นแรกมาแล้ว จะนำมาให้รหัสตามความเหมาะสมของข้อมูล เป็นการสร้างคู่มือการลงรหัส แยกเป็นพวกได้ดังนี้

1. ข้อมูลที่ใช้จำแนก (identification – ID) เป็นข้อมูลที่ใช้รหัสเป็นตัวแทนหมายเลขในการกำหนดคุณลักษณะของสิ่งตัวอย่างในการวิจัย

2. ข้อมูลเชิงปริมาณ (quantitative data) เป็นข้อมูลที่มีระดับการวัดเป็นช่วงมาตรา หรืออัตราส่วนมาตรา ได้มาจากการ ชั่ง ตวง วัด เช่น น้ำหนัก ปริมาตร อายุ ความสูง ความเร็วของแสง ปริมาณคลอรีนในน้ำดื่ม พลังงานของแสง

3. ข้อมูลเชิงคุณภาพ (qualitative data) เช่นเพศ อาชีพ ศาสนา สี หรืออันดับมาตรา เช่น มาก ปานกลาง น้อย ได้มาจากการนับ หรือการกำหนดน้ำหนักของคะแนน (weighting score)

2. การบันทึกข้อมูล

การบันทึกข้อมูลในการวิเคราะห์ด้วยมือ อาจทำได้ด้วยการทำรอยคะแนน (Tally mark) ส่วนการวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ มีให้เลือกหลายโปรแกรมคือ

Excel เป็นโปรแกรมสำหรับ กรอกข้อความตัวเลข และรูปภาพลงในช่องตารางเรียกว่า Work sheet

EPI INFO ใช้วิเคราะห์ข้อมูลทางด้านวิทยาการระบาด

ACCESS เป็นโปรแกรมพัฒนาระบบจัดเก็บข้อมูล สามารถค้นหาข้อมูลได้ตามเงื่อนไขที่ต้องการ

SPASS ใช้วิเคราะห์ข้อมูลทางด้านสังคมศาสตร์

3. การตรวจสอบแก้ไขรหัสข้อมูลครั้งสุดท้าย

1. เปรียบเทียบรหัสข้อมูลจากคู่มือลงรหัสกับแผ่นพิมพ์
2. พิมพ์ข้อมูลชุดเดียวกัน 2 คน เขียนโปรแกรมจับคู่ข้อมูลทั้ง 2 ชุด

การตรวจสอบข้อมูลเป็นสิ่งที่จำเป็นและสำคัญอย่างยิ่ง เพราะถ้าข้อมูลไม่ถูกต้อง ทำให้ผลการวิเคราะห์ผิดพลาด การแปรผลและสรุปผลจะผิดพลาดไปด้วย

ตารางจำลอง (dummy table) คือ ตารางที่นักวิจัยสร้างขึ้นเสมือนเป็นตารางที่จะนำเสนอจริงในผลการวิจัย เพียงแต่ขาดตัวเลขที่จะกรอกไปเท่านั้น ช่วยเป็นกรอบในการวิเคราะห์ แปรผล ช่วยให้การศึกษามีข้อผิดพลาดได้น้อย ประหยัดค่าใช้จ่ายให้นักวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ การสร้างตารางจำลองที่ดี ทำให้นักวิจัยทราบว่าข้อมูลอะไรขาดหายไป อาจต้องตามไปเก็บข้อมูลเพิ่มเติมถ้าข้อมูลนั้นมีความสำคัญ

องค์ประกอบของตาราง ตารางโดยทั่วไปจะประกอบด้วย

1. **หมายเลขของตาราง** เป็นการระบุลำดับที่ของตารางว่าเป็นตารางที่เท่าใด โดยทั่วไปมีมากกว่า 1 ตาราง กำหนดให้อยู่ซ้ายมือ
2. **ชื่อของตาราง** เป็นข้อความที่อยู่ถัดจากหมายเลขของตารางและอยู่แถวเดียวกัน ชื่อตารางจะต้องเป็นข้อความสั้น ๆ แต่ได้ความกระจ่างชัด เพื่อให้ทราบว่าข้อมูลที่แสดงในตารางเป็นเรื่องอะไร ที่ไหนและเมื่อใด
3. **หัวเรื่องแถว** เป็นส่วนของตารางที่อธิบายถึงข้อมูลที่ปรากฏอยู่ในแนวนอน
4. **หัวเรื่องสดมภ์** เป็นส่วนของตารางที่อธิบายถึงข้อมูลที่ปรากฏอยู่ในแนวตั้ง
5. **ตัวเรื่อง** เป็นส่วนของตารางที่เป็นตัวเลข และตัวเลขเหล่านี้จะจำแนกตามรายละเอียดของหัวเรื่องแถวและหัวเรื่องสดมภ์

ชนิดของตารางจำลอง

1. **ตาราง 1 ตัวแปร (univariate tabulation or one-dimensional)** เป็นตารางอย่างง่ายที่ประกอบไปด้วยตารางที่มีตัวแปรเดียว ในหัวตารางที่เป็นแถวตั้งจะมีเพียงจำนวนและร้อยละ มักจะนำเสนอคุณลักษณะของตัวอย่างไม่ได้แสดงความสัมพันธ์ระหว่างตัวแปร ดังตัวอย่างที่นำเสนอ เพศ (นามมาตรา) และกลุ่มอายุ (เป็นข้อมูลเชิงปริมาณชนิดอัตราส่วนมาตรา ที่นำมาจัดกลุ่มใหม่) ในกรณีข้อมูลเชิงปริมาณ เมื่อนำมาเสนอในรูปของการจัดกลุ่มนักวิจัยต้องนำเสนอค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน รวมทั้งค่าต่ำสุดค่าสูงสุดของตัวแปรนั้นด้วย ดังตัวอย่าง

ตาราง จำนวนและร้อยละของผู้ตอบแบบสอบถามเรื่องการวางแผนครอบครัว จำแนกตามเพศและอายุ			
ตัวแปร		จำนวน	ร้อยละ
เพศ	ชาย	300	37.5
	หญิง	500	62.5
	รวม	800	100
กลุ่มอายุ (ปี)	20 - 29	200	25.5
	30 - 49	250	31.3
	50 ขึ้นไป	350	43.7
	รวม	800	100
ค่าเฉลี่ย \pm ส่วนเบี่ยงเบนมาตรฐาน		41.2 \pm 12.7	
ค่าต่ำสุด-ค่าสูงสุด		20 - 60	

2. **ตาราง 2 ตัวแปร (bivariate tabulation or two-dimensional)** เป็นตารางที่มีตัวแปร 2 ตัว ในหัวตารางที่เป็นแถวบนและแถวตั้งจะเป็นตัวแปรคนละกลุ่ม เช่น ตัวแปรอิสระหรือตัวแปรตาม มักจะเป็นการวิเคราะห์แบบตารางไขว้ การจะจัดตัวแปรตามหรือตัวแปรอิสระเป็นแถวตั้งหรือแถวบนแล้วแต่ความเหมาะสมของหน้ากระดาษ ตัวอย่าง

ตาราง จำนวนและร้อยละของผู้ป่วยด้วยโรคซาร์สจำแนกตามประวัติการไปต่างประเทศ

ประวัติการไป ต่างประเทศ	การเจ็บป่วยด้วยโรค ซาร์ส		จำนวน	รวม ร้อยละ
	ป่วย	ไม่ป่วย		
เคยไป	80	20	100	100.0
ไม่เคยไป	0	100	100	100.0
รวม	80	120	200	100.0

ตามตาราง ที่ 7.6 มี 2 ตัวแปรคือ การเจ็บป่วยด้วยโรคซาร์สและประวัติการไปต่างประเทศ ตัวแปรทั้งสองแยกชั้นย่อยเป็น 2 ชั้น คือ การเจ็บป่วยแยกเป็นป่วยและไม่ป่วย ประวัติการไปต่างประเทศแยกเป็นเคยไปและไม่เคยไป ตามตารางนี้ถือว่าประวัติการไปต่างประเทศเป็นตัวแปรอิสระ การคำนวณร้อยละใช้ประวัติการไปต่างประเทศเป็นฐานในการคำนวณ แต่การแปรผลจะแปลผลตามตัวแปรตามคือ การเจ็บป่วยด้วยโรคซาร์ส

3. ตารางที่มีตัวแปรมากกว่า 2 ตัวแปร

ตาราง ความสัมพันธ์ระหว่างอาชีพ ระดับการศึกษาและความพึงพอใจต่อการรับบริการใน โรงพยาบาลของผู้มารับบริการ

การศึกษา	เกษตรกร		ข้าราชการ		รวม
	พอใจ	ไม่พอใจ	พอใจ	ไม่พอใจ	
ต่ำกว่าปริญญาตรี	80	20	60	40	200
ปริญญาตรีหรือสูงกว่า	20	80	40	60	200
รวม	100	100	100	100	400

ความสำคัญของคอมพิวเตอร์ในการวิเคราะห์ข้อมูล

คอมพิวเตอร์ช่วยทำให้ข้อมูลที่มีจำนวนมากและต้องการเวลามากในการวิเคราะห์สำเร็จลงได้อย่างรวดเร็วและแม่นยำ ง่ายขึ้น

สื่อที่ใช้ในการบันทึกข้อมูลด้วยคอมพิวเตอร์ ได้แก่ การ์ด แผ่นดิสเก็ตต์ เทปแม่เหล็ก ดิสก์หรือจานแม่เหล็ก แผ่นซีดี คอมแพคแฟลช การ์ดมัลติมีเดีย สื่ออื่นๆ เช่น removable disk

วิธีการบันทึกข้อมูล หลังจากรวบรวมข้อมูล ผ่านกระบวนการตรวจสอบข้อมูลเบื้องต้น ได้มีการสร้างคู่มือลงรหัส และลงรหัส ในแบบฟอร์มที่สร้างขึ้นแล้ว และทำความเข้าใจกับคอมพิวเตอร์ อุปกรณ์หรือสื่อต่างๆ ที่ใช้กับคอมพิวเตอร์ ทราบโปรแกรมคอมพิวเตอร์ที่จะเลือกใช้พร้อมทั้งโปรแกรมอื่นๆ ในเครื่องคอมพิวเตอร์ ก็สามารถที่จะบันทึกหรือคีย์ข้อมูลเข้าเครื่องคอมพิวเตอร์ได้ ทั้งนี้ในการบันทึกข้อมูลสำหรับงานวิจัยเรื่องหนึ่งควรใช้โปรแกรมคอมพิวเตอร์ในการบันทึกเพียงโปรแกรมเดียว ไม่ควรใช้หลายโปรแกรม หรือแบ่งการบันทึกข้อมูลก็ควรใช้เป็นโปรแกรมเดียวกัน

การตรวจสอบข้อมูลขั้นสุดท้าย (final editing)

ประโยชน์จากการใช้คอมพิวเตอร์ช่วยในการตรวจสอบข้อมูลขั้นสุดท้าย ก่อนทำการวิเคราะห์ คือ

1. ตรวจสอบรหัสข้อมูลจากแผ่นพิมพ์ เป็นการตรวจสอบข้อมูลที่พิมพ์ออกมากับรหัสที่ลงไว้ในแบบพิมพ์
2. ตรวจสอบรหัสข้อมูลโดยการพิมพ์ซ้ำ เป็นการตรวจสอบโดยให้พิมพ์ซ้ำโดยคนเดียวกันหรือคนอื่น
3. ตรวจสอบรหัสข้อมูลโดยดูตัวเลขที่เป็นไปไม่ได้

การตรวจสอบข้อมูลขั้นสุดท้ายนี้เป็นสิ่งจำเป็น เพราะถ้าข้อมูลไม่ชัดเจนมีข้อผิดพลาดเกิดขึ้น จะทำให้การวิเคราะห์ข้อมูล สะดุดเป็นระยะ ต้องเสียเวลาไปค้นหาข้อมูลที่ถูกต้อง หรือเมื่อวิเคราะห์ข้อมูลไปแล้วเพิ่งมาพบข้อผิดพลาดหรือความ เป็นไปไม่ได้ ก็ต้องย้อนกลับมาวิเคราะห์ข้อมูลใหม่อีกครั้ง ทำให้เสียเวลา งบประมาณ และความรู้สึก

การวิเคราะห์ข้อมูลและการนำเสนอข้อมูล

การวิเคราะห์ข้อมูล คือการนำข้อมูลที่รวบรวมมาได้ นำมาสู่วิธีทางสถิติ แบ่งออกเป็น 2 ส่วน คือ

1. สถิติพรรณนา (descriptive statistics) เป็นการวิเคราะห์ข้อมูลชนิดที่ไม่อาศัยความน่าจะเป็น
 2. การอนุมานเชิงสถิติ (inference statistics) เป็นการวิเคราะห์ข้อมูลชนิดที่อาศัยความน่าจะเป็น
- นอกจากนี้การวิเคราะห์ข้อมูลยังต้องมีความรู้เกี่ยวกับประเภทของข้อมูล แบ่งเป็น 2 ประเภทคือ

1. **ข้อมูลเชิงคุณภาพ** (qualitative data) เช่น เพศ ศาสนา การเกิดโรค
2. **ข้อมูลปริมาณ** (quantitative data) เช่น ข้อมูล ตัวแปร ปริมาณ ตัวเลข วัดได้ น้ำหนัก ส่วนสูง อายุ

ข้อมูลทั้ง 2 ประเภท ยังมีระดับการวัดตัวแปรหรือข้อมูลเป็น 4 ระดับคือ

1. **ระดับนามมาตรา** (nominal scale) ชื่อ เปรียบเทียบความมากน้อยไม่ได้
2. **ระดับอันดับมาตรา** (ordinal scale) บอกอันดับได้แต่บอกไม่ได้ว่ามากน้อยกว่ากันเท่าไร เช่น นางงาม
3. **ระดับช่วงมาตรา** (interval scale)
4. **ระดับอัตราส่วนมาตรา** (ratio scale) มี 0 แท้

การแจกแจงความถี่

1. การแจกแจงแบบเรียงตัวเลข ได้แก่การนำตัวเลขมาเรียงลำดับ จากมากไปน้อยหรือน้อยไปมาก
2. การจัดเรียงเป็นหมวดหมู่แล้วแจกแจงความถี่
 - ตารางแจกแจงความถี่แบบธรรมดา
 - ตารางแจกแจงความถี่แบบมีอันตรภาคชั้น

หลักการสร้างตารางแจกแจงความถี่แบบมีอันตรภาคชั้น

1. หาค่าสูงสุดของข้อมูล
2. หาค่าต่ำสุดของข้อมูล
3. หาพิสัย พิสัย = ค่าสูงสุด - ค่าต่ำสุด
4. กำหนดจำนวนชั้น
5. หาอันตรภาคชั้น = $\text{พิสัย} / \text{จำนวนชั้น}$
6. การจัดเรียงชั้น อาจเริ่มเรียงให้ชั้นที่มีค่าน้อยอยู่ชั้นบนก่อน หรืออาจเรียงให้ชั้นที่มีค่ามากอยู่ชั้นบนก่อนก็ได้ แต่การเรียงให้ชั้นค่าน้อยอยู่บนก่อน แล้วชั้นค่ามากเรียงชั้นล่างลงไปตามลำดับ จะทำให้ดูง่ายกว่า
7. ในตารางที่เรียงค่าน้อยไว้ชั้นบน ค่าของขีดล่างในชั้นนี้จะต้องมีค่าน้อยกว่าหรือเท่ากับค่าต่ำสุดของข้อมูลและค่าของขีดบนของชั้นสูงสุด จะต้องมีความมากกว่าหรือเท่ากับค่าสูงสุดของข้อมูล
8. แจกแจงความถี่ของค่าของข้อมูลที่มีค่าอยู่ในชั้นต่างๆ ด้วยเครื่องหมาย tally mark แล้วรวมลงเป็นค่าความถี่

การวัดแนวโน้มเข้าสู่ส่วนกลาง

- มัชฌิมมาเลขคณิต
- มัชฌิมฐาน
- ฐานนิยม
-

การวัดการกระจาย

- พิสัย
- ส่วนเบี่ยงเบนควอร์ไทล์
- ส่วนเบี่ยงเบนมาตรฐาน

มัชฌิมเลขคณิต (arithmetic mean) เป็นค่าเฉลี่ยที่ได้จากวิธีเลขคณิตธรรมดา คือ การรวมข้อมูลแต่ละตัวเข้าด้วยกันหารด้วยจำนวนข้อมูลทั้งหมด สัญลักษณ์ที่ใช้แทนมี 2 ตัวคือ กรณีเป็นค่าเฉลี่ยประชากรใช้ μ (มิว) และกรณีเป็นค่าเฉลี่ยของตัวอย่างใช้ \bar{X} (เอ็กซ์บาร์)

1. การคำนวณหามัชฌิมเลขคณิตจากข้อมูลที่ไม่ได้แจกแจงความถี่ หมายถึง การหาค่าเฉลี่ยของข้อมูลที่ยังไม่ได้มีการแจกแจงความถี่ ยังอยู่ในสภาพกระจกระยะหรือเป็นข้อมูลดิบ มักใช้กับข้อมูลที่มีจำนวนไม่มาก

สูตร $\bar{X} = \sum x / n$ เช่น

ตัวอย่าง สมมติว่ามีนักเรียนอนุบาลจำนวน 10 คน ได้รับเงินค่าขนมสำหรับไปโรงเรียนจากผู้ปกครอง ดังนี้ 3,3,4,5,5,6,7,7 และ 8 บาท จงคำนวณหามัชฌิมเลขคณิตของข้อมูลชุดนี้

วิธีทำ

$$\bar{X} = \frac{3+3+4+5+5+6+7+7+8}{10} = \frac{53}{10} = 5.3$$

นั่นคือ ค่าเฉลี่ยของค่าขนมของเด็กนักเรียนอนุบาลเท่ากับ 5.3 บาท

2. การคำนวณหามัชฌิมเลขคณิตจากข้อมูลที่มีการแจกแจงความถี่ สูตร $\bar{X} = \sum fx/n$

ข้อดีของมัชฌิมเลขคณิต

1. เข้าใจง่ายและคำนวณง่าย
2. เป็นค่าที่ได้มาจากค่าของข้อมูลทุกตัว
3. เป็นค่าที่แน่นอน
4. เป็นค่าตัวกลางที่เหมาะสมกับการกระจายปกติ หรือเกือบปกติ
5. เหมาะสำหรับการคำนวณขั้นต่อไป

ข้อเสีย

1. มีค่าตรงกับความเป็นจริงของข้อมูลไม่กี่ราย หรืออาจจะไม่มีเลย
2. เป็นค่าที่ถูกกระทบกระเทือน โดยค่าของข้อมูลที่สูงหรือต่ำมากอย่างผิดปกติ หรือข้อมูลที่ผิดปกติ

มัธยฐาน (median) คือ ค่าของข้อมูลที่มีตำแหน่งตรงกลางของข้อมูลชุดหนึ่ง ซึ่งได้จัดเรียงลำดับค่า หรือจัดหมู่ไว้เรียบร้อยแล้ว นั่นคือการเรียงข้อมูลจากน้อยไปหามาก การคำนวณมัธยฐานมี 2 วิธี

1. การคำนวณมัธยฐานจากข้อมูลที่ไม่ได้แจกแจงความถี่

มัธยฐานของข้อมูล 1 4 5 7 8 9 คือ $(5 + 7) / 2 = 6$

2. การคำนวณมัธยฐานจากข้อมูลที่มีการแจกแจงความถี่ (สูตรคำนวณ หน้า 17)

ข้อดีของมัธยฐาน

1. เข้าใจง่าย
2. ค่ามัธยฐานที่หาได้จากข้อมูลที่เรียงลำดับค่าแล้ว จะเป็นค่าที่ตรงกับค่าจริงของข้อมูลนั้น
3. เป็นค่าที่ไม่ถูกรบกวนกระทบกระเทือน โดยค่าของข้อมูลที่มีค่าสูงหรือต่ำมากผิดปกติ หรือข้อมูลที่ผิดปกติ
4. ถ้าทราบค่าข้อมูลในตำแหน่งกลาง ๆ ก็พอจะคำนวณหาค่ามัธยฐานได้
5. ไม่ได้ใช้ทุกค่ามาคำนวณ เพราะในการแจกแจงความถี่ที่อยู่ในรูปของตารางแจกแจงความถี่ที่มีช่วงเปิด เช่น น้อยกว่า หรือมากกว่าก็สามารถคำนวณได้

ข้อเสียของมัธยฐาน

1. ค่ามัธยฐานไม่แน่นอน ถ้าการแจกแจงข้อมูลไม่สมมาตร
2. ค่ามัธยฐานที่ได้จากข้อมูลที่จัดหมวดหมู่ จะเป็นค่าที่ไม่ถูกต้องตามความเป็นจริง
3. จะหาค่ามัธยฐานไม่ได้ หากไม่ได้จัดเรียงลำดับค่าของข้อมูลให้เรียบร้อยก่อน

ฐานนิยม (mode) ฐานนิยมคือค่าที่มีความถี่มากที่สุด หรือค่าที่ซ้ำกันมากที่สุด (ตัวอย่างคำนวณ หน้า 20)

เช่น ฐานนิยมของข้อมูล 3 8 5 6 4 5 คือ 5

ข้อดีของฐานนิยม

1. เข้าใจง่าย
2. หาได้ง่ายจากข้อมูลที่เรียงลำดับค่าแล้ว
3. เป็นค่าที่ไม่ถูกรบกวนกระทบกระเทือน โดยค่าของข้อมูลที่มีค่าสูง หรือค่าต่ำมากผิดปกติ หรือข้อมูลที่ผิดปกติ
4. อาศัยเสียงข้างมากเป็นเกณฑ์

ข้อเสียของฐานนิยม

1. ฐานนิยมอาจมีหลายค่า
2. ฐานนิยมบางค่าอยู่ก่อนไปทางค่าสูงมาก บางค่าก่อนไปทางค่าที่ต่ำมาก
3. หาค่าที่แน่นอนไม่ได้จากข้อมูลที่จัดเป็นหมู่
4. ไม่เหมาะที่จะคำนวณขั้นต่อไป

การวัดการกระจาย นิยมมี 3 ชนิด คือ 1. พิสัย 2. ส่วนเบี่ยงเบนควอร์ไทล์ 3. ส่วนเบี่ยงเบนมาตรฐาน

1. **พิสัย** คือ ผลต่างระหว่างค่าของข้อมูลที่สูงที่สุด และค่าที่ต่ำสุดของข้อมูลชุดหนึ่ง ๆ = ค่าสูงสุด - ค่าต่ำสุด
2. **ส่วนเบี่ยงเบนควอร์ไทล์** (quartile deviation) คือ ค่า ครึ่งหนึ่งของผลต่างระหว่างควอร์ไทล์ที่ 3 กับควอร์ไทล์ที่ 1 ของข้อมูลชุดหนึ่ง ๆ $Q.D. = \frac{Q_3 - Q_1}{2}$

2 (ตัวอย่าง คำนวณหน้า 28)

3. **ส่วนเบี่ยงเบนมาตรฐาน** (standard deviation) ใช้สัญลักษณ์ σ (sigma อ่านว่าซิกมา) แทนส่วนเบี่ยงเบนมาตรฐานของประชากร และ s แทนส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง เป็นค่าที่ได้จากการถอดกรณฑ์ที่สอง (square root) ของค่าเฉลี่ยของกำลังสองของผลต่างระหว่างค่าแต่ละค่าของข้อมูลกับมัชฌิมเลขคณิต เป็นการวัดการกระจายที่ดีที่สุดในงานวิจัย (ตัวอย่าง หน้า 29)

ถ้าส่วนเบี่ยงเบนมาตรฐานมีค่ามาก หมายความว่าค่าต่าง ๆ ของข้อมูลชุดนั้นมีการกระจายมากรูปโค้งจะกว้างแบน ถ้าส่วนเบี่ยงเบนมาตรฐานมีค่าน้อย แสดงว่าค่าต่าง ๆ ของข้อมูลชุดนั้นมีการกระจายน้อยรูปโค้งจะแคบและสูง

สัมประสิทธิ์แห่งการกระจาย

สัมประสิทธิ์แห่งการกระจายเป็นวิธีการที่นำมาใช้ในการเปรียบเทียบการกระจายของข้อมูลต่างชนิดกันซึ่งข้อมูลต่างชุดกันที่มีค่ากระจายเท่ากัน อาจมีขนาดของข้อมูลต่างกันมากก็ได้

สัมประสิทธิ์แห่งการกระจาย = $\frac{S}{x} \times 100$

x

ค่านี้อาจเรียกได้อีกชื่อหนึ่งว่า “สัมประสิทธิ์ของความแปรปรวน” (coefficient of variation, C.V.)

ห้องเรียน	มัธยมศึกษาคนใด	ส่วนเบี่ยงเบนมาตรฐาน
ห้อง ก	20	5
ห้อง ข	24	3

จะได้ $CV_g = (5/20) \times 100 = 25$

จะได้ $CV_b = (3/24) \times 100 = 12.5$

ดังนั้น ห้อง ก มีการกระจายมากกว่า ห้อง ข

การเปรียบเทียบค่าเฉลี่ยและการเปรียบเทียบค่าสัดส่วน

การเปรียบเทียบค่าเฉลี่ย

1. การเปรียบเทียบค่าเฉลี่ยของประชากร 1 กลุ่ม

1.1 ทราบค่าความแปรปรวนของประชากร (σ^2 known) สถิติที่ใช้คือ z-test

1.2 ไม่ทราบค่าความแปรปรวนของประชากร (σ^2 unknown) สถิติที่ใช้คือ t-test

2. การเปรียบเทียบความแตกต่างของค่าเฉลี่ยของประชากร 2 กลุ่ม

2.1 ประชากร 2 กลุ่มเป็นอิสระต่อกัน

2.2.1 ทราบค่าความแปรปรวนของประชากร (σ_1^2, σ_2^2 known) สถิติที่ใช้คือ z-test

2.2.2 ไม่ทราบค่าความแปรปรวนของประชากร (σ_1^2, σ_2^2 unknown) สถิติที่ใช้คือ t-test

2.2 ประชากร 2 กลุ่มไม่เป็นอิสระต่อกัน สถิติที่ใช้คือ paired t – test

การเปรียบเทียบค่าสัดส่วน

1. การเปรียบเทียบค่าสัดส่วนของประชากร 1 กลุ่ม สถิติที่ใช้ทดสอบคือ z-test

2. การเปรียบเทียบความแตกต่างของค่าสัดส่วนของประชากร 2 กลุ่ม

2.1 ประชากร 2 กลุ่ม ที่เป็นอิสระต่อกัน สถิติที่ใช้ทดสอบคือ z-test

2.2 ประชากร 2 กลุ่มที่ไม่เป็นอิสระต่อกัน สถิติที่ใช้ทดสอบคือ McNemar test

3. การเปรียบเทียบความแตกต่างของค่าสัดส่วนของประชากรมากกว่า 2 กลุ่ม

3.1 ประชากรมากกว่า 2 กลุ่มที่เป็น อิสระต่อกัน สถิติที่ใช้คือ ไคสแควร์

3.2 ประชากรมากกว่า 2 กลุ่มที่ไม่อิสระต่อกัน สถิติที่ใช้คือ Cochran q test ตัวอย่างหน้า 44

การหาค่าความสัมพันธ์ระหว่างตัวแปร

1. การหาความสัมพันธ์ของข้อมูลเชิงคุณภาพ

1.1 ตารางการจร r x c สถิติที่ใช้คือ ไคสแควร์ test

1.2 ตารางการจร 2 x 2 สถิติที่ใช้คือ X^2 – test ที่ปรับแก้โดย เยทส์ (Yates’s)

ทั้งสอบแบบความถี่ในช่องต้องไม่น้อยกว่า 5 ถ้าน้อยกว่าให้ใช้แบบฟิชเชอร์

1.3 การทดสอบความแม่นยำของพีชเชอร์ สำหรับตาราง 2 x 2 ใช้วิธีหาค่า p-value

2. การหาความสัมพันธ์ของข้อมูลเชิงปริมาณ สถิติที่ใช้คือการหาสหสัมพันธ์แบบเพียร์สัน ตัวอย่างกิจกรรม 8.2.3 หน้า 47

การหาค่าตัวทำนาย

ในการหาค่าตัวทำนายนี้หมายความว่า ถ้าตัวแปรหนึ่ง ๆ มีความสัมพันธ์กับตัวแปรอีกตัวหนึ่ง หรือตัวแปรอิสระมีผลต่อตัวแปรตาม ตัวแปรอิสระ (X) จะสามารถทำนายการเกิดตัวแปรตาม (Y) ได้หรือไม่ มีความแม่นยำในการทำนายได้เพียงใด บางกรณีเรียกว่า การหาตัวพยากรณ์ (predicting, forecasting) เป็น สถิติขั้นสูง (advanced statistics) ซึ่งต้องคำนึงถึงข้อตกลงเบื้องต้นก่อนนำไปใช้

1. การวิเคราะห์ถดถอย
2. การวิเคราะห์จำแนกพหุ
3. การวิเคราะห์จำแนกประเภท

คู่มือกิจกรรมหน้า 51

การนำเสนอข้อมูล

การนำเสนอข้อมูลในรูปบทความ เป็นการนำเสนอข้อมูลที่มีลักษณะเป็นบทความสั้น ๆ ปนไป กับตัวเลขที่มีรายการน้อย นับว่าเป็นข้อมูลธรรมดาและง่ายที่สุด พบได้ทั่วไปในหน้าหนังสือพิมพ์ รายการวิทยุ โทรทัศน์ หรือสรุปรายงานต่าง ๆ

การเสนอข้อมูลในรูปบทความถึงตาราง เป็นการเสนอด้วยบทความกับตัวเลขที่จัดเป็นหมวดหมู่ซึ่งมีจำนวนไม่มากและไม่จำเป็นต้องทำเป็นรูปตาราง

การนำเสนอข้อมูลในรูปตาราง เป็นการนำข้อมูลที่มีหลายรายการหรือซ้ำ ๆ กัน มาแสดงไว้ในตารางให้เป็นแถวตามแนวตั้ง และแถวตามแนวนอนเพื่อความเป็นระเบียบเรียบร้อย และสะดวกในการเปรียบเทียบ

การนำเสนอข้อมูลในรูปกราฟและแผนภูมิ

1. การนำเสนอแบบรูปกราฟ

2. การนำเสนอแบบแผนภูมิแท่ง

3. การนำเสนอแบบแผนภูมิวงกลม

4. การนำเสนอแบบแผนภูมิรูปภาพ

5. การนำเสนอแบบแผนภูมิแผนที่

6. การนำเสนอแบบ แผนภาพกระจาย หรือ แผนภาพสหสัมพันธ์

7. **การนำเสนอแบบฮิสโตแกรม (histogram)** คือ รูปกราฟแท่งที่เรียงติดกันและฐานของแท่งเป็นตัวแปรในสเกลช่วงมาตรา หรืออัตราส่วนมาตรา มีหลักดังนี้


- ใช้กับข้อมูลที่มีการแจกแจงความถี่
- ใช้แสดงปริมาณด้วยพื้นที่ ๆ สร้างขึ้นเป็นแท่ง ตั้งอยู่บนแกนนอน
- แท่งเหล่านี้นั้นวางชิดติดกัน
- เพื่อความสะดวกในการสร้างควรทำให้ทุกอันตรภาคชั้น (class interval) มีค่าเท่า ๆ กัน ถ้าไม่เท่ากันต้องทำให้เป็นความหนาแน่นของความถี่เสียก่อน (กราฟ หน้า 71)

รูปฮิสโตแกรม คือ รูปแท่ง ส่วนภาพเส้นประที่จะสร้างต่อที่หลัง คือ รูปหลายเหลี่ยมแห่งความถี่

8. **การนำเสนอรูปหลายเหลี่ยมแห่งความถี่ (frequency polygon)** เปรียบเทียบข้อมูลแจกแจงความถี่ตั้งแต่ 2 ชุดขึ้นไป ดัดแปลงมาจากรูปฮิสโตแกรมให้ดูง่ายซึ่งมีเนื้อหาอยู่ครบตามเดิม (รูป กราฟ หน้า 72)

 - รูปสี่เหลี่ยมผืนผ้า หมายถึงว่ามีความถี่เท่ากันหมด

 - รูปประฆังสมมาตร หมายถึงว่าส่วนใหญ่มีคะแนนปานกลาง คะแนนน้อย ๆ และคะแนนมาก ๆ มีเป็นจำนวนน้อย

 - รูปประฆังเบ้ขวา หมายถึงว่าคะแนนน้อย ๆ มีเป็นจำนวนมาก คะแนนมาก ๆ มีเป็นจำนวนน้อย

 - รูปประฆังเบ้ซ้าย หมายถึงว่าคะแนนน้อย ๆ มีเป็นจำนวนน้อย คะแนนมากมีเป็นจำนวนมาก

- รูปตัว U หมายถึงว่า คะแนนน้อยและคะแนนมากมีจำนวนมากกว่ามีจำนวนมากกว่าคะแนนกลาง ๆ เช่นการแจกแจงของปริมาณรถในช่วงเช้าก่อนทำงานและ ช่วงเย็นหลังเลิกงาน

- รูปตัว J หมายถึงว่าคะแนนน้อยมีจำนวนมากกว่าคะแนนกลาง ๆ แต่น้อยกว่าคะแนนมาก เช่นการแจกแจงอายุประชากรตามอัตราตายตามกลุ่มอายุ

- รูปตัว J กลับ หมายถึงว่าคะแนนน้อยมีจำนวนมากกว่าคะแนนกลาง ๆ และคะแนนมากมีน้อยกว่าคะแนนน้อย

9. การนำเสนอแบบแผนภูมิพีระมิด (pyramidal chart) หรือพีระมิดประชากร (population pyramid) แสดงถึงโครงสร้างประชากรตามกลุ่มอายุและเพศ เป็นภาพฮิสโตแกรม 2 ภาพติดกัน แกน x แทนด้วยประชากรแยกตามเพศชาย เพศหญิง แกน Y แทนกลุ่มอายุ นิยมจัดกลุ่มอายุให้มีอันตรภาคชั้นเท่ากันเท่ากับ 5 สามารถแสดงได้ 3 รูปแบบ

- **พีระมิดประชากร** แสดงเป็นค่าจำนวนประชากร ทำให้ผู้อ่านสามารถทราบได้ว่า จำนวนประชากรแยกตามเพศและกลุ่มอายุ มีจำนวนกี่คน
- **พีระมิดประชากร** แสดงเป็นค่าร้อยละของประชากรทั้งหมด ทำให้ผู้อ่านทราบได้ว่า มีประชากรในแต่ละกลุ่มอายุและเพศ มีกี่เปอร์เซ็นต์ของประชากรทั้งหมด ประชากรวัยต่าง ๆ เช่นเด็กต่ำกว่า 5 ปี มีเป็นร้อยละเท่าไรของประชากรทั้งหมด
- **พีระมิดประชากร** แสดงเป็นค่าร้อยละของประชากรแต่ละเพศ ทำให้ผู้อ่านทราบว่าประชากรแต่ละกลุ่ม อายุ และเพศ มีเป็นร้อยละเท่าไรของจำนวนประชากรรวมแต่ละเพศนั้น ๆ เป็นการเปรียบเทียบกันเอง

เซต

หมายถึง กลุ่มของสิ่งของต่างๆ ซึ่งมีคุณสมบัติที่แน่ชัด โดยที่สามารถบอกสมาชิกของเซตนั้นได้ การดำเนินการของเซต ได้แก่ ยูเนียน อินเตอร์เซกชัน และคอมพลิเมนต์ เป็นการนำหลายๆ เซตมาดำเนินการให้เกิดเซตใหม่

$A \cup B$

คือ เซตที่มีสมาชิกที่เป็นสมาชิกของเซต A หรือ สมาชิกของ เซต B หรือทั้งสองเซต เขียนแทนด้วยสัญลักษณ์

$$A \cup B = \{X \mid X \text{ เป็นสมาชิกของ } A \text{ หรือ } X \text{ เป็นสมาชิกของ } B\}$$

$A \cap B$

คือ เซตที่มีสมาชิกที่เป็นสมาชิกของเซต A และเซต B

เขียนอธิบายลักษณะของสมาชิกได้ว่า

$$A \cap B = \{X \mid X \text{ เป็นสมาชิกของ } A \text{ และ } X \text{ เป็นสมาชิกของ } B\}$$

A^c

คอมพลิเมนต์ของเซต A ใดๆ เมื่อเทียบกับเอกภพสัมพัทธ์ U คือเซตของทุกสมาชิกที่อยู่ใน U แต่ไม่อยู่ใน A

เขียนอธิบายลักษณะของสมาชิกได้ว่า

$$A^c = \{X \mid X \text{ เป็นสมาชิกของ } U \text{ และ } X \text{ ไม่เป็นสมาชิกของ } A\}$$

การเรียงสับเปลี่ยนและการจัดหมู่

เป็นเรื่องราวของจำนวนวิธีการจัดสิ่งของที่แตกต่างกัน โดยดูอันดับสิ่งของเรียกว่าการเรียงสับเปลี่ยน และเมื่อไม่ดูอันดับสิ่งของเรียกว่าการจัดหมู่

การเรียงสับเปลี่ยนของสิ่งของ n สิ่งซึ่งแตกต่างกันทั้งหมด โดยจัดอันดับที่ละ r สิ่ง ($r < n$) มีค่าคือ

$${}_n P_r = n! / (n-r)! \quad \text{ตัวอย่างหน้า 93}$$

การจัดหมู่ของสิ่งของที่แตกต่างกัน n สิ่ง โดยที่มีหมู่ละ r สิ่ง (r มากกว่าหรือเท่ากับ n) มีค่าคือ

$${}_n C_r = n! / r!(n-r)! \quad \text{ตัวอย่างหน้า 94}$$

การทดลองเชิงสุ่ม

หมายถึง การทดลองที่ไม่สามารถพยากรณ์ผลลัพธ์ได้อย่างถูกต้องแน่นอน เนื่องจากผลลัพธ์อาจเกิดขึ้นได้หลายอย่าง ยกตัวอย่างการทอดลูกเต๋า การโยนเหรียญ

ปริภูมิตัวอย่าง (Sample space) ใช้ตัวย่อ S

หมายถึง เซตที่มีสมาชิกเป็นผลลัพธ์ที่อาจจะเป็นไปได้ทั้งหมดของการทดลองเชิงสุ่ม

เหตุการณ์ (event) ใช้ตัวย่อ E

หมายถึง เซตย่อยของปริภูมิตัวอย่าง เช่น การทอดลูกเต๋าที่ยังตรง 1 ครั้ง

$$S = \{1, 2, 3, 4, 5, 6\}$$

$$E_1 = \{2, 4, 6\} \text{ คือเซตของเหตุการณ์ที่เป็นเลขคู่}$$

$$E_2 = \{1, 2, 3\} \text{ คือเซตที่ได้แต้มต่ำกว่า 4}$$

ความน่าจะเป็น

หมายถึง ค่าความถี่สัมพัทธ์ที่บอกให้ทราบว่าเหตุการณ์นั้นๆ มีโอกาสเกิดขึ้นได้มากน้อยเพียงใด ตัวอย่างหน้า 98

การแจกแจงความน่าจะเป็น

หมายถึง การแจกแจงความถี่ของตัวแปรสุ่มชนิดไม่ต่อเนื่อง และชนิดต่อเนื่อง ด้วยความน่าจะเป็น

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง ได้แก่ การแจกแจงทวินาม และการแจกแจงปัวซอง

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่อง ได้แก่ การแจกแจงปกติ

หน้า 105-107

การแจกแจงทวินาม

คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง ที่ผลการทดลองมี 2 อย่าง คือ สำเร็จและล้มเหลว

ความน่าจะเป็นของการแจกแจงทวินาม

ตัวอย่างหน้า 111, 114

การแจกแจงปัวซอง

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง ที่ค่าเฉลี่ยของความสำเร็เกิดขึ้นในช่วงเวลาหรือขอบเขตที่

กำหนด ความน่าจะเป็นของการแจกแจงปัวซอง ตัวอย่างหน้า 118

การแจกแจงปกติ

คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่อง ที่โค้งการแจกแจงเป็นรูประฆังคว่ำสมมาตร

ความน่าจะเป็นของการแจกแจงปกติมาตรฐาน ตัวอย่างหน้า 123, 124

ความหมายของการประมาณค่า (estimation) เป็นการนำสถิติที่ได้จากตัวอย่างมาใช้ประมาณค่าพารามิเตอร์ของประชากร **ความสำคัญของการประมาณค่า** สิ่งที่น่าสนใจ ต้องการศึกษ ต้องการทราบในการวิจัยคือ ลักษณะของประชากรในเรื่องนั้น ๆ ลักษณะของประชากรคือ ค่าพารามิเตอร์ ในการวิจัยเราไม่สามารถเก็บรวบรวมข้อมูลจากทุกหน่วยของประชากรได้ เนื่องจากประชากรที่ต้องการศึกษามีเป็นจำนวนมาก และในการทำการวิจัยมีข้อจำกัดในด้าน เวลา ค่าใช้จ่าย กำลังคน และวัสดุอุปกรณ์ หรือเครื่องมือต่าง ๆ ดังนั้นจึงต้องทำการเก็บรวบรวมข้อมูลจากหน่วยตัวอย่าง เช่น โดยการสำรวจ การสังเกต หรือการทดลอง แล้วใช้วิธีทางสถิติ หาค่าสถิติจากตัวอย่าง ใช้ในการประมาณค่าพารามิเตอร์ของประชากรในเรื่องที่สนใจต้องการศึกษา หรือ ต้องการทราบเพื่อแสดงลักษณะของประชากรนั้น

ประเภทของการประมาณค่า ในการประมาณค่าพารามิเตอร์ของประชากร โดยใช้ค่าสถิติจากตัวอย่าง

แบ่งออกได้เป็น 2 ประเภท คือ

1. **การประมาณค่าเป็นจุด** (point estimation) คือ การนำเอาค่าสถิติที่ได้จากข้อมูลตัวอย่างไปใช้ประมาณค่าพารามิเตอร์โดยการประมาณค่าพารามิเตอร์หนึ่งๆ โดยกล่าวว่าค่าเท่ากับค่าเพียงค่าเดียวเท่านั้น ก็จะมีโอกาสหรือความน่าจะเป็นที่ผิดพลาดได้ง่าย
2. **การประมาณค่าเป็นช่วง** (interval estimation) คือการนำเอาค่าสถิติ ที่ได้จากข้อมูลของตัวอย่างไปใช้ประมาณค่าพารามิเตอร์ โดยการกำหนดค่าประมาณเป็นช่วงว่าค่าพารามิเตอร์ที่ต้องการทราบจะมีค่าอยู่ในช่วงนี้ ด้วยความเชื่อมั่นที่ถูกต้องเป็นร้อยละหรือเป็นเปอร์เซ็นต์เท่าใด และเรียกช่วงนั้นว่า ช่วงความเชื่อมั่น (confidence interval)

คุณสมบัติของการประมาณค่า ตัวประมาณค่าเป็นจุดที่ดีจะต้องมีคุณสมบัติดังนี้

1. **เป็นตัวประมาณค่าที่ไม่อคติ** (unbiased estimator) คือค่าที่คาดหวัง ของตัวประมาณค่าจะต้องมีค่าเท่ากับค่าพารามิเตอร์นั้น ซึ่งหมายความว่าค่าเฉลี่ยของค่าพารามิเตอร์ตัวอย่างของตัวอย่าง (sample) จะต้องเท่ากับพารามิเตอร์ที่ต้องการทราบหรือต้องการศึกษานั้น ๆ
2. **มีความพ้องและสม่ำเสมอ** (consistency) คือเมื่อตัวอย่างมีขนาดใหญ่ขึ้น หรือมีจำนวนตัวอย่างเพิ่มขึ้น โดยไม่มีขอบเขตจำกัด ตัวประมาณค่า ก็ควรมีค่าเข้าใกล้ค่าพารามิเตอร์ ที่แท้จริงมากยิ่งขึ้น มีความพ้องและสม่ำเสมออาจเรียกว่ามีความคงที่
3. **มีความแปรปรวนน้อยที่สุด** (minimum variance) คือตัวประมาณค่าที่เลือกมาใช้จะต้องมีความแปรปรวนน้อยที่สุด เลือกวิธีที่มีความแปรปรวนน้อยที่สุด
4. **มีความเพียงพอ** (sufficiency) ตัวประมาณค่าที่เลือกมาใช้จะต้องเป็นตัวประมาณค่าที่มีความเพียงพอให้สาระข้อเท็จจริงหรือให้ข่าวสารข้อมูลอย่างเพียงพอ

ความหมายของการทดสอบสมมติฐาน (hypothesis testing) การทดสอบสมมติฐานเป็นการสรุปหรือพิสูจน์คำตอบของการวิจัย ในการที่จะให้ได้ข้อสรุปหรือการตัดสินใจที่ชัดเจนตามสมมติฐานที่สนใจเฉพาะนั้นจึงจำเป็นต้องมีการยืนยันโดยใช้ข้อเท็จจริงจากชุดข้อมูลที่ได้จากตัวอย่างต่าง ๆ โดยจะต้องมีวิธีการสำหรับการปฏิเสธ หรือการยอมรับสมมติฐานอย่างถูกต้องแน่นอน วิธีทางวิทยาศาสตร์ที่สำคัญอย่างหนึ่งคือ การสรุปผลที่มีเหตุผลทางวิทยาศาสตร์ เป็นวิธีที่ยอมรับและสามารถที่จะทดสอบข้อสรุปซ้ำได้โดยนักวิจัยคนอื่น ๆ

ความสำคัญของการทดสอบสมมติฐาน คล้ายกันกับความสำคัญของการประมาณค่า การทดสอบสมมติฐานช่วยหาข้อสรุปของการทดสอบที่สามารถเชื่อถือได้ ทำให้สรุปผลหรือพิสูจน์ข้อสงสัยที่ต้องการทราบของการวิจัยได้ โดยต้องมีขั้นตอนในการพิสูจน์ ซึ่งเป็นไปตามขั้นตอนของวิธีทางวิทยาศาสตร์ และสามารถนำข้อสรุปหรือข้อค้นพบที่หาได้จากตัวอย่างนี้ไปใช้หรืออ้างในประชากรได้ ด้วยความมั่นใจว่าถูกต้องเชื่อถือได้เป็นร้อยละเท่าใด

ขั้นตอนหรือวิธีการทดสอบสถิติในการวิจัย ได้แก่

1. ตั้งหรือกำหนดสมมติฐานว่าง (H_0) และสมมติฐานทางเลือก (H_1)
2. เลือกตัวทดสอบทางสถิติ ที่เหมาะสมสำหรับการทดสอบ H_0 โดยเลือกตัวทดสอบสถิติ จากตัวทดสอบสถิติต่าง ๆ มาใช้ให้เหมาะสมกับการออกแบบการวิจัย และข้อกำหนดของการวัดข้อมูลที่กำหนดไว้
3. กำหนดระดับนัยสำคัญ (α) และขนาดตัวอย่าง (n)
4. หาลักษณะการแจกแจงของตัวอย่าง (sampling distribution) ของตัวทดสอบสถิติภายใต้สมมติฐานว่าง
5. กำหนดบริเวณของการปฏิเสธสมมติฐานหรือบริเวณวิกฤต
6. หาค่าของตัวทดสอบสถิติโดยใช้ข้อมูลที่ได้จากตัวอย่าง ถ้ามีค่าอยู่ในบริเวณวิกฤต เราก็จะปฏิเสธสมมติฐาน H_0 นั่นคือเราตัดสินใจยอมรับ H_1 เมื่อเราปฏิเสธ H_0 โดยค่าของตัวทดสอบสถิติมีค่าอยู่ในบริเวณวิกฤต (แต่ถ้าค่าอยู่นอกพื้นที่บริเวณวิกฤตเราไม่สามารถตัดสินใจปฏิเสธ H_0 ได้ที่ระดับนัยสำคัญนั้น)
7. ตีความสรุปผลการทดลองในแง่การนำไปใช้ คำนวณระดับค่าพี (p-value) ต่อท้ายเพื่อยืนยันความน่าเชื่อถือของความเป็นที่สรุปผลการทดสอบ
8. สรุปผลการทดสอบ

การเลือกสถิติที่ใช้ทดสอบ (test statistic)

ตัวทดสอบทางสถิติที่นำมาใช้ในการตัดสินใจ เพื่อสรุปเกี่ยวกับการทดสอบสมมติฐานที่สำคัญ ๆ ได้แก่ ตัวทดสอบ Z ตัวทดสอบ t ตัวทดสอบ X^2 และตัวทดสอบ F

ตัวทดสอบ Z (z-test) ใช้ทดสอบเกี่ยวกับค่าเฉลี่ยที่ทราบค่าความแปรปรวนของประชากร เมื่อข้อมูลมีการแจกแจงปกติ ข้อมูลจะถูกแปลงให้เป็นค่ามาตรฐานคือ Z ซึ่งมีค่าเฉลี่ย = 0 ละค่าความแปรปรวน = 1 หากค่าสถิติที่นำมาทดสอบเป็นค่าสัดส่วนซึ่งไม่ทราบค่าความแปรปรวนของประชากรก็สามารถใช้ตัวทดสอบ Z ได้ (หน้า 148)

ตัวทดสอบ t (t-test) เมื่อข้อมูลมีการแจกแจงแบบปกติใช้ตัวทดสอบ t ทดสอบเกี่ยวกับค่าเฉลี่ยโดยไม่ทราบค่าความแปรปรวนของประชากร เมื่อไม่ทราบค่าความแปรปรวนของประชากร สามารถใช้ค่าความแปรปรวนของข้อมูลจากสิ่งตัวอย่างแทนได้ ซึ่งตัวทดสอบ t มีลักษณะการแจกแจงหรือการกระจาย เส้นโค้งจะกว้างและแบนกว่าตัวทดสอบ Z เมื่อองศาของความเป็นอิสระ (degree of freedom = df) เพิ่มขึ้น t จะมีค่าเฉลี่ย = 0

ตัวทดสอบ X^2 (X^2 - test) มีลักษณะการแจกแจง ซึ่งจะเปลี่ยนไปตามระดับขององศาของความเป็นอิสระ จะมีค่าเป็นบวกเสมอ ตัวทดสอบ X^2 นำไปใช้ในการทดสอบเกี่ยวกับค่าความแปรปรวนของข้อมูล 1 ชุด ทดสอบเกี่ยวกับความสัมพันธ์และทดสอบว่าข้อมูลมีลักษณะการแจกแจงปกติหรือไม่

ตัวทดสอบ F (F-test) ลักษณะการแจกแจงของ F จะมีลักษณะคล้ายการแจกแจง X^2 แต่มีองศาแห่งความเป็นอิสระ 2 ค่า คือ ของเศษและส่วน

การกำหนดความผิดพลาดในการทดสอบ การตัดสินใจเกี่ยวกับการทดสอบสมมติฐานว่าง H_0 มีความผิดพลาดจากการทดสอบได้ ข้อสรุปผลการทดสอบอาจไม่ตรงกับความเป็นจริง มีความผิดพลาดที่เกิดขึ้นมีได้ 2 ประเภท คือ

1. **ความผิดพลาดประเภทที่ 1 (Type I error)** เป็นความผิดพลาดที่เกิดขึ้นจากการสรุปผลการทดสอบที่ปฏิเสธ H_0 เมื่อ H_0 เป็นจริง และความน่าจะเป็นของความผิดพลาดประเภท 1 นี้เขียนแทนด้วย α
2. **ความผิดพลาดประเภทที่ 2 (Type II error)** เป็นความผิดพลาดของการสรุปว่ายอมรับ H_0 เมื่อ H_0 ไม่เป็นจริง ความน่าจะเป็นของความผิดพลาดประเภทที่ 2 นี้เขียนแทนด้วย β

ถ้าค่าสถิติทดสอบตกอยู่บริเวณวิกฤติหรือ $p\text{-value} < \alpha$ จะสรุปว่า ปฏิเสธ H_0 ถ้าไม่ตกอยู่ในวิกฤติ หรือ $p\text{-value} > \alpha$ จะสรุปว่ายอมรับ H_0

การประมาณค่า

การประมาณค่าจะต้องระบุพารามิเตอร์ที่ต้องการประมาณ และต้องกำหนดช่วงความเชื่อมั่น โดยอาศัยทฤษฎีการแจกแจงค่าสถิติที่ได้จากตัวอย่าง ค่าสถิติที่ใช้ต้องเป็นตัวประมาณค่าที่ดี เลือกตัวประมาณค่า กำหนดระดับความเชื่อมั่น แล้วคำนวณค่าประมาณ

ค่าประมาณค่าเฉลี่ยของประชากร (μ) ได้แก่ค่าเฉลี่ยเลขคณิตของตัวอย่าง ชิดจำกัด ความเชื่อมั่นของค่าเฉลี่ย

ค่าประมาณสัดส่วนของประชากร (P) ได้แก่ ค่าสัดส่วนของตัวอย่าง (p) ชิดจำกัดความเชื่อมั่นของสัดส่วน

หลักของการประมาณค่า อาศัยทฤษฎีการแจกแจงตัวอย่างของค่าสถิติ โดยมีค่า $E = \theta$ ถ้าต้องการระดับความเชื่อมั่นสูง ความคลาดเคลื่อนมาตรฐานก็จะสูงด้วยเช่นกัน

ขั้นตอนการประมาณค่าพารามิเตอร์

1. **ระบุพารามิเตอร์ที่ต้องการประมาณค่า** พารามิเตอร์ของประชากรที่ต้องการประมาณค่าจะขึ้นอยู่กับวัตถุประสงค์หรือเป้าหมายการวิจัย
2. **เลือกตัวประมาณค่า** จะต้องเป็นตัวประมาณค่าที่มีคุณสมบัติของตัวประมาณค่าที่ดี แล้วเลือกสูตรสถิติที่สอดคล้องกับตัวประมาณค่า สำหรับประมาณค่าพารามิเตอร์ที่ต้องการนั้น
3. **กำหนดระดับความเชื่อมั่น และหาช่วงความเชื่อมั่น** หรือระดับนัยสำคัญของการประมาณค่าพารามิเตอร์ เพื่อคำนวณค่าความคลาดเคลื่อนมาตรฐาน และช่วงความเชื่อมั่น
4. **ทำการประมาณค่าพารามิเตอร์** คำนวณค่าประมาณพารามิเตอร์ อาจเป็นค่าประมาณแบบจุด หรือ ค่าประมาณแบบเป็นช่วง ซึ่งได้แก่ขีดจำกัดล่างและขีดจำกัดบนของความเชื่อมั่น

ตัวอย่าง 10.2.1 ม 10.2.3 หน้า 162,163

แบบประเมินหลังการสอนเสริมครั้งที่ 2

1. เครื่องมือที่มีรูปแบบแบ่งระดับความคิดเห็นของผู้ตอบ 4-6 ระดับ ให้ผู้ตอบเลือกตอบ เรียกรูปแบบการวัดนี้ว่าอะไร
ข. มาตราส่วนประมาณค่า (Likert)
2. สัมประสิทธิ์แอลฟาเหมาะกับการหาความเที่ยงในเครื่องมือที่มีลักษณะอย่างไร
ค. แบบวัดแบบมาตราส่วนประมาณค่า
3. การจัดการข้อมูลไม่ดีทำให้เกิดผลเสียอย่างไร
ค. งานล่าช้า ผิดพลาด ไม่ครบถ้วน เสียงบประมาณ
4. การตรวจสอบข้อมูลขั้นสุดท้ายโดยใช้โปรแกรมสำเร็จรูปที่เหมาะสมกับข้อมูลจำนวนมากคือกรณีใด
ค. ตรวจสอบรหัสข้อมูลโดยคู่มือเลขที่เป็นไปไม่ได้
5. ข้อมูลชุดหนึ่งมีดังนี้ 3, 7, 2, 6, 8 และ 10 จงหามัชฌิมเลขคณิตของข้อมูลชุดนี้
จ. 6
6. การหาความสัมพันธ์โดยใช้สถิติไคสแควร์เป็นการหาความสัมพันธ์ของข้อมูลที่มีลักษณะแบบใด
ค. ตั้งแปรอิสระเป็นข้อมูลคุณภาพ ตัวแปรตามเป็นข้อมูลคุณภาพ
7. ทอดลูกเต๋าเที่ยงตรง 2 ลูก ความน่าจะเป็นที่ได้แต้ม ที่ค่ารวมกัน 9 มีค่าเท่ากับค่าใด
ข. 1/9

8. จากสถิติการคลอดที่โรงพยาบาลแห่งหนึ่งพบว่าใน 1 ปี มีผู้มาคลอด 12,000 ราย โดยเฉลี่ยพบว่าเป็นเด็กตายคลอดร้อยละ 0.1 จงหาค่าเฉลี่ยของเด็กตายคลอดต่อเดือน
- ก. 1
9. การทดสอบสมมุติฐานแบ่งตามวัตถุประสงค์ได้ 2 ประเภท คือการทดสอบตามข้อใด
- ก. ความแตกต่างและความสัมพันธ์
- 10 ค่าประมาณค่าเฉลี่ยของประชากร (μ) ได้แก่ ตัวประมาณค่าชนิดใดของตัวอย่าง
- ก. ค่าเฉลี่ยเลขคณิต