บทที่ 1
ความรู้เบื้องต้นทางสถิติ
(Introduction of Statistical)

1.1 ความหมายของชีวสถิติ

• คือ ข้อมูลสถิติ (Statistical Data) ข้อมูลสรุปได้จากการประมวลหรือวิเคราะห์กลุ่มของข้อมูลเพื่อใช้แสดงลักษณะข้อมูลของกลุ่มนั้น

• คือ สถิติศาสตร์ (Statistics) ศาสตร์ว่าด้วยการจัดกระทำต่างๆเกี่ยวกับข้อมูลเพื่อบรรยายลักษณะของสิ่งที่ศึกษาได้อย่างเที่ยงตรงและเชื่อถือได้นำเอาไปใช้คาดคะเนและการตัดสินใจต่างๆ

1.2 ประเภทของสถิติ

• สถิติพรรณนา (Descriptive Statistics) วิธีการทางสถิติที่ใช้พรรณนาลักษณะ สิ่งต้องการศึกษาให้อยู่ในรูปของตารางข้อมูลสรุป การนำเสนอแบบต่างๆเพื่อให้เข้าใจถึงข้อมูลที่รวบรวมมาได้ แต่ไม่สามารถคาดคะเนนอกเหนือไปจากข้อมูลที่มีอยู่ได้

• สถิติอนุมาน (Inferential Statistics) วิธีการทางสถิติที่ใช้ทฤษฎีความน่าจะเป็นในการอนุมานลักษณะของประชากรจากข้อมูลของตัวอย่างเช่นการศึกษาโรคขาดสารอาหารในเด็กวัยก่อนเรียนในภาคอีสานส่วนใหญ่จะสุ่มจากเด็กวัยก่อนเรียนมาบางส่วนเพื่อประเมินหาอัตราการขาดสารอาหาร เป็นต้น

1.3 ประเภทข้อมูล

ข้อมูล (Data) หมายถึง ข้อเท็จจริงที่เก็บรวบรวมได้จากตัวอย่างหรือประชากร

1.3.1 แบ่งตามลักษณะข้อมูล ได้ 2 ประเภท ดังนี้

1. ข้อมูลเชิงปริมาณ (Quantitative Data) เป็นข้อมูลที่วัดค่าได้ว่ามากหรือน้อยในเชิงปริมาณ เช่น รายได้ อายุ ความสูงจำนวนสินค้า ฯลฯ ซึ่งแบ่งได้ 2 แบบ คือ

1.1 ข้อมูลแบบต่อเนื่อง (Continuous Data) หมายถึง ข้อมูลที่มีค่าต่อเนื่องกันในช่วงที่กำหนด สามารถแจงสมาชิกในข้อมูลได้ เช่น ความสูง อายุ ระยะทาง เป็นต้น

1.2 ข้อมูลแบบไม่ต่อเนื่อง (Discrete Data) หมายถึง ข้อมูลที่มีค่าเป็นจำนวนเต็มหรือจำนวนนับ เช่น จำนวนนักศึกษา จำนวนสมาชิกในครัวเรือน เป็นต้น

2. ข้อมูลเชิงคุณภาพ (Qualitative Data) เป็นข้อมูลที่ไม่สามารถระบุค่าได้ว่ามากหรือ น้อย อาจแทนด้วยตัวเลขก็ได้โดยตัวเลขดังกล่าวไม่มีความหมายในเชิงปริมาณ เช่น เพศ ระดับการศึกษา อาชีพ ทัศนคติ เป็นต้น

1.3.2 แบ่งตามแหล่งที่มาของข้อมูล ได้ 2 ประเภท ดังนี้

1. ข้อมูลปฐมภูมิ (Primary Data) เป็นข้อมูลที่ผู้ใช้ไปเก็บรวบรวมข้อมูลเอง ข้อ มูลที่ได้จะมีความทันสมัยมีความถูกต้องน่าเชื่อถือ แต่การรวบรวมข้อมูลต้องใช้เวลานาน ต้องใช้กำลังคนมาก เสียค่าใช้จ่ายสูง ไม่สะดวกเท่าที่ควร

2. ข้อมูลทุติยภูมิ (Secondary Data) เป็นข้อมูลที่มีการเก็บรวบรวมไว้แล้วผู้ใช้เป็นเพียงผู้ที่นำข้อมูลนั้นมาใช้ จึงเป็นการประหยัดทั้งเวลาและค่าใช้จ่าย แต่บางครั้งจะเป็นข้อมูลที่ไม่ตรงกับความต้องการหรือไม่ละเอียดพอ นอกจากนี้ผู้ใช้มักจะไม่ทราบถึงข้อผิดพลาดของข้อมูล ซึ่งมีผลทำให้การวิเคราะห์ผลอาจจะผิดพลาดได้

1.4 ระดับของการวัด (Level of Measurement)

ข้อมูลในการวิจัยจำนวนมากได้มาจากการวัด ซึ่งการวัด (Measurement) หมายถึง การกำหนดตัวเลข หรือสัญลักษณ์อื่น ๆ แทนปริมาณหรือคุณภาพหรือคุณลักษณะ ของสิ่งที่วัด มาตราการวัดมี 4 ระดับ คือ

1. มาตรานามบัญญัติ (Nominal Scale or Classification Scale)

เป็นข้อมูลที่มีลักษณะจำแนกกลุ่มหรือประเภท โดยตัวเลขหรือค่าที่กำหนดให้นำมาบวก ลบ คูณ หาร กันไม่ได้ เป็นระดับการวัดที่ต่ำ ที่สุด เป็นการกำหนดตัวเลขแทนชื่อคน แทนคุณลักษณะต่าง ๆ แทนเหตุการณ์ต่างๆ หรือแทนสิ่งต่าง ๆ เช่น เบอร์นางงามที่เข้าประกวด เบอร์นักฟุตบอล เลขทะเบียนรถต่าง ๆ การกำหนดให้เลข 0 แทน เพศหญิงเลข 1 แทนเพศชาย คุณสมบัติที่สำคัญของมาตรานี้ก็คือ ตัวเลขที่กำหนดให้จะเพียงแต่ชี้ถึง ความแตกต่างกัน คือชี้ว่าไม่ใช่สิ่งเดียวกัน ไม่ได้แทนอันดับ ขนาด ปริมาณหรือคุณภาพใด ๆ ซึ่งตัวเลขหรือค่าต่าง ๆ ที่กำหนดให้นั้นนำมาบวก ลบ คูณ หารกันไม่ได้ และจากการที่ไม่ได้ชี้ปริมาณหรือคุณภาพดังกล่าว นักจิตวิทยาบางท่านจึงไม่ยอมรับการวัดชนิดนี้ว่าเป็นการวัด (Measurement)

ข้อมูลที่ได้จากการวัดโดยใช้มาตรานามบัญญัติ (Nominal scale)

1. เพศ ☐ ชาย ☐ หญิง
2. เชื้อชาติ (ไทย, จีน, ฯลฯ)
3. ศาสนา (พุทธ, คริสต์, อิสลาม, ฯลฯ
4. อาชีพ ☐ หมอ ☐ นักเรียน ☐ ครู
5. หมายเลขโทรศัพท์ .....
6. เลขที่บ้าน .....

สรุป ถ้าข้อมูลบอกความแตกต่างได้แต่เพียงชื่อ เป็นความแตกต่างที่หยาบที่สุด เช่นชื่อนักศึกษาทั้ง 30 คน ไม่ได้ให้รายละเอียดอะไร นอกจากว่าแต่ละคนชื่อแตกต่างกัน เราเรียกข้อมูลที่สามารถจำแนกได้แต่เพียงชื่อว่า ข้อมูลระดับมาตรานามบัญญัติ(Nominal scale)

2. มาตรอันดับ (Ordinal Scale)

เป็นระดับของการวัดที่สูงกว่ามาตรานามบัญญัติ เป็นการกำหนดตัวเลขหรือสัญลักษณ์เพื่อชี้ถึงอันดับ เช่น หลังจากพิจารณาภาพที่นักเรียนวาดมาแล้วก็ได้อันดับจากภาพที่ดีที่สุดเป็นอันดับ 1 รองลงมาเป็นอันอับ 2 , 3 ,..... ตามลำดับ เป็นต้น จะเห็นได้ว่าในมาตรานี้มีคุณสมบัติของมาตรานามบัญญัติคือ ความแตกต่าง อันดับ 1 และอันดับ 2 จะเป็นคนละคนไม่เหมือนกัน สิ่งที่เราทราบเพิ่มขึ้นจากมาตรานามบัญญัติคือ ทิศทาง ของความแตกต่าง อันดับ 1 อยู่เหนือกว่าอันดับ 2 เนื่องจากมีปริมาณหรือคุณภาพมากกว่า อย่างไรก็ตาม แม้จะทราบว่าใครมากกว่า น้อยกว่า แต่ไม่อาจทราบว่ามากกว่ากันเท่าใด และช่วงระหว่างอันดับต่าง ๆ มักไม่เท่ากัน เช่นที่ 1 อาจมีคุณภาพเหนือกว่าที่ 2 มากขณะที่ 2 มีคุณภาพห่างจากที่ 3 เพียงเล็กน้อย เป็นต้น จากการที่ช่วงอันดับไม่เท่ากันดังกล่าว จึงไม่สามารถนำเอาตัวเลขในมาตรานี้มาบวก ลบ คูณ หรือหารกันได้ เช่น ผลการแข่งขันกีฬามหาวิทยาลัย ระดับความคิดเห็น ตำแหน่งทางวิชาการ

ข้อมูลที่ได้จากการวัดโดยใช้มาตรอันดับ (Ordinal scale)

เป็นข้อมูลที่นอกจากมีลักษณะจำแนกกลุ่มหรือประเภทได้แล้วยังสามารถเรียงอันดับได้ด้วย เช่น ตำแหน่ง (ที่ 1, ที่ 2, ที่ 3, ฯลฯ) ระดับความพอใจ (มากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด)

1. กิจกรรมที่นิยมทำในวันหยุด (เรียงลำดับมากที่สุดเป็นลำดับ 1)

..... ดูหนัง

..... ฟังเพลง

..... เล่นกีฬา

..... ดูโทรทัศน์

..... ช๊อปปิ้ง

สรุป ถ้าข้อมูลบอกความแตกต่างได้ เป็นการจำแนกข้อมูลที่ละเอียดขึ้น เพราะบอกความแตกต่างได้ เช่นชื่อนักศึกษาที่เรียงกันตามระดับความสูงทั้ง 30 คน เราเรียกข้อมูลที่เราสามารถจำแนกถึงความแตกต่างได้นี้ว่า ข้อมูลระดับมาตราอันดับบัญญัติ (Ordinal scale)

3. มาตรอันตรภาค (Interval Scale)

เป็นระดับของการวัดที่สูงกว่าสองมาตราที่กล่าว มาโดยมีคุณสมบัติเพิ่มขึ้นอีก 2 ประการ คือ มีศูนย์สมมุติ (Arbitrary Zero or Relative Zero) และมีหน่วยของการวัดที่เท่ากัน ตัวอย่างของมาตรานี้ ได้แก่ การวัดอุณหภูมิ เช่น ในหน่วยวัดอุณหภูมิแบบเซลเซียส จะกำหนดจุดที่น้ำกลายเป็นน้ำแข็งเป็น 0° ซ. เป็นศูนย์เทียมไม่ได้หมายความว่าถึง ณ อุณหภูมิ 0° ซ. นี้ไม่มีความร้อนอยู่เลยแต่เป็นเพียงจุดที่น้ำกลายเป็นน้ำแข็ง จากการที่มีหน่วยของการวัดที่เท่ากัน และมีศูนย์เทียมจึงสามารถเปรียบเทียบปริมาณ หรือคุณภาพได้ว่ามากกว่ากันเท่าไร เช่น 40° ซ. จะมีอุณหภูมิสูงกว่า 30° ซ. อยู่ 10° ซ. และสามารถพูดได้ว่าอุณหภูมิ 20° ซ. สูงกว่าอุณหภูมิ 15° ซ. เท่ากับอุณหภูมิ 14° ซ. สูงกว่าอุณหภูมิ 9° ซ. เพราะต่างก็สูงกว่ากัน 5° ซ. (ไม่อาจพูดได้ว่าอุณหภูมิ 60° ซ. ร้อนเป็นสองเท่าของอุณหภูมิ 30° ซ. เพราะความร้อนไม่ได้เริ่มที่จุด 0° ซ.) หรือ 60° ซ. = 2 (30° ซ.) แต่ปริมาณความร้อนของสสาร 60° ซ. ? 2 (ความร้อนของสสาร 30° ซ.) นักพฤติกรรมศาสตร์มักถือเอาว่าคะแนนการสอบเป็นการวัดในมาตรานี้ จึงตีความในลักษณะเดียวกันกับกรณีของอุณหภูมิที่กล่าวมา เช่น ในแบบทดสอบที่มีจำนวน 60 ข้อ ถ้า ก สอบได้ 50 คะแนน ข สอบได้ 30 คะแนน ค สอบได้ 25 คะแนน และ ง สอบได้ 5 คะแนน ก็กล่าวว่า ก ได้คะแนนมากกว่า ข 20 คะแนน ข ได้คะแนนมากกว่า ง 25 คะแนน ก ได้คะแนนมากกว่า ข เท่ากับ ค ได้คะแนนมากกว่า ง (ต่างกันมากกว่า 20 คะแนน) แต่ไม่สามารถตีความได้ว่า ก มีความรู้เป็น 2 เท่าของ ค เพราะจุดเริ่มต้นไม่ใช้ศูนย์แท้ ผู้สอบได้คะแนนศูนย์ไม่ได้หมายความว่าไม่มีความรู้ในวิชานั้น เป็นเพียงแต่ว่าทำข้อสอบชุดนั้นไม่ได้ ถ้าออกข้อสอบมากกว่านั้น หรือง่ายกว่านั้นเขาอาจทำได้บ้าง มาตราอันตราภาคนับว่าเป็นมาตราที่เป็นปริมาณอย่างแท้จริงไม่เหมือนมาตรานามบัญญัติและมาตราเรียงลำดับ**ข้อมูลที่ได้จากการวัดโดยใช้มาตราอันตรภาค (Interval scale)**

เป็นข้อมูลที่มีลักษณะจำแนกกลุ่ม เรียงอันดับ และแบ่งเป็นช่วง ๆ โดยแต่ละช่วงมีขนาดเท่ากัน ศูนย์ของข้อมูลประเภทนี้เป็นศูนย์สมมติ ไม่มีศูนย์แท้ เช่น คะแนนสอบ อุณหภูมิ เวลา IQ

สรุป ถ้าข้อมูลบอกความแตกต่างเป็นหน่วยที่เท่ากันได้ เป็นการจำแนกความแตกต่างที่ละเอียดขึ้น เพราะสามารถบอกความแตกต่างเป็นปริมาณหน่วยที่เท่ากัน ทำให้บอกระดับความแตกต่างที่ละเอียดมาก และบอกได้ว่าแต่ละคนแตกต่างกันเป็นปริมาณเท่าใด โดยเปรียบเทียบกับหน่วยปริมาณที่เท่ากันได้ เช่น คะแนนสอบของนักศึกษาทั้ง 30 คน เราเรียกข้อมูลที่จำแนกถึง หน่วยที่แตกต่างกันได้ว่า ข้อมูลระดับช่วงบัญญัติ (Interval scale)

4. มาตรอัตราส่วน (Ratio Scale)

เป็นระดับของการวัดที่สูงที่สุด มีความสมบูรณ์ มากกว่ามาตราวัดอันตรภาค นอกจากจะมีคุณสมบัติเหมือนมาตราวัดอันตรภาคแล้วยังมี ศูนย์แท้ (Absolute Zero) ในขณะที่มาตราอันตรภาคมีเพียงศูนย์สมมุติ ตัวอย่างการวัดในมาตรานี้ได้แก่ การวัดความยาว น้ำหนัก ส่วนสูง อายุแต่ละหน่วยของความยาวจะมีช่วงเท่ากันแต่ละหน่วยของ น้ำหนักมีขนาดเท่ากัน เช่น เอื้อมพร หนัก 40 กิโลกรัม จะหนักเป็น 2 เท่าของนิตยารัตน์ ซึ่งหนัก 20 กิโลกรัม การที่กล่าวเช่นนี้ได้เนื่องจากแต่ละหน่วยกิโลกรัมมีน้ำหนักเท่ากัน และเริ่มจากศูนย์แท้ น้ำหนักศูนย์กิโลกรัมก็คือไม่มีน้ำหนักเลยเนื่องจากการวัดระดับนี้มีความสมบูรณ์ทุกประการ จึงสามารถนำมาจัดกระทำตามหลักคณิตศาสตร์ได้ทุกประการ เช่น บวก ลบ คูณ หาร ถอดราก และยกกำลังได้

ข้อมูลที่ได้จากการวัดโดยใช้มาตราอัตราส่วน (Ratio scale) เป็นข้อมูลที่มีลักษณะจำแนกกลุ่ม เรียงอันดับ แบ่งเป็นช่วงเท่า ๆ กัน และมีศูนย์แท้ สามารถเปรียบเทียบในเชิงอัตราส่วนได้ เช่น ระยะทาง เวลา น้ำหนัก ส่วนสูง อายุ

ตัวอย่าง ของ มาตรอัตราส่วน (Ratio Scale)

1. รายได้ต่อเดือน .....
2. อายุ .....
3. ค่าใช้จ่าย .....
4. น้ำหนัก .....
5. ส่วนสูง .....

สรุป ถ้าข้อมูลบอกความแตกต่างเป็นหน่วยที่เท่ากันได้และค่าศูนย์เป็นค่าศูนย์ที่แท้จริง(Absolute Zero) เช่น ความสูงเป็นเซนติเมตรของนักศึกษาทั้ง 30 คน เราเรียกข้อมูลที่สามารถจำแนกเป็นปริมาณที่แตกต่างกันเท่ากันและมีศูนย์แท้ว่า ข้อมูลระดับมาตราอัตราส่วนบัญญัติ(Radio scale)

คุณสมบัติของแต่ละมาตรา สรุปได้ดังนี้

มาตรา	คุณสมบัติ
นามบัญญัติ (Nominal Scale)	ความแตกต่างกัน
เรียงลำดับ (Ordinal Scale)	ความแตกต่างกัน + ทิศทาง
อันตรภาค (Interval Scale)	ความแตกต่างกัน + ทิศทาง + ช่วงเท่ากัน + ศูนย์สมมุติ
อัตราส่วน (Ratio Scale)	ความแตกต่างกัน + ทิศทาง + ช่วงเท่ากัน + ศูนย์แท

สถิติสำหรับตัวแปรเดียว

ลักษณะของข้อมูล	สถิติที่ใช้	การนำเสนอข้อมูล
1. สเกลนามบัญญัติ (Nominal Scale)	ความถี่ อัตราส่วนร้อย	ตารางแจกแจงความถี่ ร้อยละ สัดส่วน
2. สเกลอันดับ (Ordinal Scale)	ความถี่ อัตราส่วนร้อย เปอร์เซ็นต์ไทส์	ตารางแจกแจงความถี่ ร้อยละ สัดส่วน
3. สเกลอันตรภาคและอัตราส่วน (Interval and Ratio Scale)	ความถี่ อัตราส่วนร้อยละ ส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวน พิสัย เปอร์เซ็นต์ไทล์ ค่าเฉลี่ย	ตารางแจกแจงความถี่ ค่า ความเบ้ ค่าความโด่ง

1.5 ประชากรและตัวอย่าง

ประชากร (Population) หมายถึง หน่วยต่าง ๆ ทุกหน่วยของสิ่งที่เราสนใจจะทำการศึกษา เป็นหน่วยที่สามารถให้ข้อมูลต่าง ๆ แก่เราได้ เช่น สนใจรายได้เฉลี่ยของข้าราชการในจังหวัดสุรินทร์ ดังนั้นประชากรคือ ข้าราชการทุกคนที่ทำงานอยู่ในจังหวัดสุรินทร์

ลักษณะของประชากรที่ศึกษาอาจมีจำนวนจำกัด (finite population) ดังตัวอย่างข้างต้นหรืออาจมีจำนวนอนันต์ (infinite population) เช่น การศึกษาเกี่ยวกับประสิทธิภาพของยาชนิดหนึ่ง ประชากรจะเป็นผลการทดสอบประสิทธิภาพของยาในผู้ป่วยที่ใช้ยานี้ ซึ่งไม่สามารถบอกถึงจำนวนทั้งหมดได้

ตัวอย่าง (Sample) หมายถึง หน่วยบางหน่วยของประชากรที่จะให้ข้อเท็จจริงต่าง ๆ ที่เราสนใจศึกษา เช่น สนใจรายได้เฉลี่ยของข้าราชการที่ทำงานอยู่ในจังหวัดสุรินทร์ ดังนั้นประชากร คือ ข้าราชการทุกคนที่ทำงานอยู่ในจังหวัดสุรินทร์และตัวอย่าง คือข้าราชการบางคนที่ทำงานอยู่ในจังหวัดสุรินทร์

1.6 ค่าพารามิเตอร์และค่าสถิติ

พารามิเตอร์ (Parameter) หมายถึง ค่าที่คำนวณมาจากทุก ๆ หน่วยของประชากรเป็นตัวที่บ่งชี้ถึงคุณลักษณะของประชากร**ค่าสถิติ (Statistic)** หมายถึง ค่าที่คำนวณได้จากข้อมูลที่เป็นตัวอย่าง เป็นตัวที่บ่งชี้ถึงคุณลักษณะของตัวอย่าง ค่าสถิติจะเป็นตัวประมาณค่าพารามิเตอร์

ตารางที่ 1.1 สัญลักษณ์ค่าพารามิเตอร์และค่าสถิติ

ความหมาย	สัญลักษณ์	
	ค่าสถิติ	ค่าพารามิเตอร์
ค่าเฉลี่ยเลขคณิต	$\bar{x}$	μ
ค่าเบี่ยงเบนมาตราฐาน	s	σ
ความแปรปรวน	s^2	σ^2

1.7 ระเบียบวิธีทางสถิติ

ระเบียบวิธีทางสถิติ แบ่งออกเป็น 4 ขั้นตอน ดังนี้

1. การรวบรวมข้อมูล
2. การนำเสนอข้อมูล
3. การวิเคราะห์ข้อมูล
4. การตีความหมาย

1. การเก็บรวบรวมข้อมูล (data collection)

เป็นการรวบรวมข้อมูล จากแหล่งข้อมูล ตามที่ได้มีการวางแผนไว้ ซึ่งอาจเป็น ได้ทั้ง ข้อมูลปฐมภูมิ หรือทุติยภูมิ

2. การนำเสนอข้อมูล (data presentation)

เป็นการจัดทำข้อมูลที่รวบรวมได้ให้อยู่ในรูปแบบที่กะทัดรัด เช่น ตาราง กราฟ แผนภูมิ ข้อความ เป็นต้น เพื่อความสะดวกในการอ่านข้อมูล ให้เข้าใจง่าย และเพื่อประโยชน์ในการวิเคราะห์ต่อไป

3. การวิเคราะห์ข้อมูล (data analysis)

เป็นขั้นตอนการประมวลผลข้อมูลซึ่งในการวิเคราะห์จำเป็นต้องใช้สูตรทางสถิติต่างๆหรือใช้การอ้าง อิงทางสถิติขึ้นกับวัตถุประสงค์ของงานนั้นๆ เช่น การวิเคราะห์แนวโน้มเข้าสู่ส่วนกลาง การวัดการกระจาย การทดสอบสมมติฐาน การประมาณค่า เป็นต้น

4. การแปลความหมาย (interpretation)

เป็นขั้นตอนของการนำผลการวิเคราะห์มาอธิบายให้บุคคลทั่ว ไปเข้า ใจ อาจจำเป็น ต้องมีการขยายความในการอธิบาย เพื่อให้งานที่ศึกษาเป็นประโยชน์ต่อคนทั่ว ไปได้จากกระบวนการทางสถิติดังกล่าว เราสามารถจำแนกเป็นสถิติศาสตร์ที่สอดคล้องกับขั้นตอนต่างๆ ได้2 ลักษณะคือ สถิติบรรยาย (หรือสถิติเชิงพรรณนา) และสถิติอ้างอิง (หรือสถิติเชิงอนุมาน)

1.8 **การวัดแนวโน้มเข้าสู่ส่วนกลาง**

ในการนำเสนอข้อมูลทั้งหมดข้างต้น ถ้ามีข้อมูลจำนวนมาก อาจทำให้ผู้อ่านหรือผู้ใช้เข้าใจได้ยาก จึงมีการนำเสนอค่าตัวแทนของข้อมูลเพื่อ ให้เห็นลักษณะของข้อมูลจากค่าตัวแทนของข้อมูลค่าใดค่าหนึ่ง ค่าตัวแทนของข้อมูล หรือการวัดแนวโน้มเข้าสู่ส่วนกลางที่นิยมใช้ได้แก่ ค่าเฉลี่ยเลขคณิต (arithmetic mean, X) ค่ามัธยฐาน (median) และค่าฐานนิยม (mode)

1.8.1 ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) ค่าเฉลี่ยเลขคณิตเป็นค่าเฉลี่ยที่นิยมใช้กันมากที่สุด จะแทนด้วย μ (มิว) เมื่อคำนวณจากข้อมูลทั้งประชากร และแทนด้วย $\bar{x}$ เมื่อคำนวณจากข้อมูลตัวอย่าง การคำนวณหาได้จากผลรวมของข้อมูลทุกค่าแล้วหารด้วยจำนวนข้อมูลทั้งหมด สำหรับการคำนวณหาค่าเฉลี่ยเลขคณิตแบ่งเป็น 2 กรณี คือ

• กรณีข้อมูลไม่มีการจัดหมวดหมู่

ข้อมูลประชากร	ข้อมูลตัวอย่าง
ถ้ามีข้อมูล $X_1, ..., X_N$ รวม N ตัว	ถ้ามีข้อมูล $X_1, ..., X_n$ รวม n ตัว

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- ถ้าข้อมูล X_1 ,...,X_K มีความถี่ หรือปรากฏรวม f_1,...,f_K ครั้งตามลำดับ

ข้อมูลประชากร

ถ้ามีข้อมูล X_1 ,..., X_N รวม N ตัว

$$\mu = \frac{\sum_{i=1}^{K} f_i X_i}{\sum_{i=1}^{K} f_i} = \frac{\sum_{i=1}^{K} f_i X_i}{N}$$

ข้อมูลตัวอย่าง

ถ้ามีข้อมูล X_1,..., X_n รวม n ตัว

$$\mu = \frac{\sum_{i=1}^{K} f_i X_i}{\sum_{i=1}^{K} f_i} = \frac{\sum_{i=1}^{K} f_i X_i}{n}$$

• กรณีข้อมูลมีการจัดหมวดหมู่

ข้อมูลประชากร

ถ้ามีข้อมูล X_1 ,..., X_N รวม N ตัว

$$\mu = \frac{\sum_{i=1}^{K} f_i X_i}{\sum_{i=1}^{K} f_i} = \frac{\sum_{i=1}^{K} f_i X_i}{N}$$

ข้อมูลตัวอย่าง

ถ้ามีข้อมูล X_1,..., X_n รวม n ตัว

$$\mu = \frac{\sum_{i=1}^{K} f_i X_i}{\sum_{i=1}^{K} f_i} = \frac{\sum_{i=1}^{K} f_i X_i}{n}$$

ถ้าข้อมูล X_1 ,...,X_K มีความถี่ หรือปรากฏรวม f_1,...,f_K ครั้งตามลำดับ

เมื่อ K คือ จำนวนชั้นของข้อมูล

X_i คือ ค่ากึ่งกลางของชั้นที่ i เมื่อ I = 1, 2, 3, ... , K

f_i คือ ความถี่ของชั้นที่ i เมื่อ I = 1, 2, 3, ... , K

ตัวอย่าง 1. โรงงานแห่งหนึ่งมีพนักงาน 36, 45, 28 และ 41 คน ที่ได้รับค่าจ้างรายวันเฉลี่ย 162.5, 170.2, 159.5 และ 167 บาทตามลำดับ จงหาค่าจ้างรายวันเฉลี่ยโดยรวมของพนักงานทั้ง 4 กลุ่มนี้

วิธีทำ ค่าจ้างรายวันเฉลี่ยโดยรวมของพนักงานทั้ง 4 กลุ่ม คำนวณจากสูตร

$$\mu = \frac{\sum_{i=1}^{K} f_i X_i}{\sum_{i=1}^{K} f_i} = \frac{\sum_{i=1}^{K} f_i X_i}{N}$$

$$= \frac{\sum(36x162.5)+(45x170.2)+(28x159.5)+(41x167)}{(36+45+28+41)}$$

$$= \frac{5850+7659+4466+6847}{150}$$

$$= \frac{24822}{150}$$

$$= 165.48$$

ดังนั้น ค่าจ้างรายวันเฉลี่ยโดยรวมของพนักงานทั้ง 4 กลุ่มเท่ากับ 165.48 บาท

คุณสมบัติของค่าเฉลี่ย

1. เป็นตัวแทนข้อมูล ที่ใช้ข้อมูลทุกค่ามาทำการคำนวณหาขนาดของค่าเฉลี่ย
2. เนื่องจากมีการนำข้อมูลทุกค่ามาคำนวณตามหลักคณิตศาสตร์จึงสามารถใช้ในการวิเคราะห์สถิติขั้นสูงได้
3. เนื่องจากมีการใช้ข้อมูลทุกค่ามาคำนวณ ดังนั้นหากมีข้อมูลบางตัวที่มีขนาดใหญ่มากๆหรือเล็กมากๆ ผิดปกติจะมีผลต่อการคำนวณขนาดของค่าเฉลี่ยด้วย
4. ข้อมูลที่มีมาตรวัดเป็นนามบัญญัติ (nominal scale) และเรียงอันดับ (ordinal scale) ไม่สามารถใช้คำนวณค่าเฉลี่ยได้

1.8.2 (Median) เป็นค่าที่บอกภาพรวมของข้อมูล โดยพิจารณาจากตำแหน่งกลางของข้อมูลที่เรียกลำดับจากน้อยไปหามากแทนด้วย Me

ก. กรณีที่ข้อมูลไม่ได้แจกแจงความถี่

- ข้อมูลเป็นเลขคี่

  มัธยฐาน = ค่าของข้อมูลลำดับที่(n+1)/2

- ข้อมูลเป็นเลขคู่

  มัธยฐาน = [ค่าของข้อมูลลำดับที่(n)/2 + ค่าของข้อมูลลำดับที่(n+1)/2] / 2

ข. กรณีที่ข้อมูลมีการแจกแจงความถี่

$$\text{Me} = L_1 + \left[\frac{\frac{N}{2} - F}{f_1} \right] l$$

$$Me = U + \left[\frac{\frac{N}{2} - F}{fu} \right] I$$

โดยที่ L แทนขอบเขตล่างของชั้นมัธยฐาน

U แทนขอบเขตบนของชั้นมัธยฐาน

l แทนความกว้างของชั้น

n แทนจำนวนข้อมูลทั้งหมด

fL แทนความถี่ของชั้นที่มีค่าสังเกตต่ำกว่าชั้นมัธยฐาน

fU แทนความถี่ของชั้นที่มีค่าสังเกตสูงกว่าชั้นมัธยฐาน

fm แทนความถี่ของชั้นที่มีค่ามธัยฐาน

***ตัวอย่าง* 2.** จากข้อมูลต่อไปนี้จงหาค่ามัธยฐาน

6.5 12.0 14.9 10.0 7.9 21.9 12.5 12.5 14.5 9.2

วิธีทำ เรียงลำดับข้อมูลจากน้อยไปมาก และเนื่องจาก N = 9 ซึ่งเป็นเลขคี่ ดังนั้น ค่าของข้อมูลตัวที่ = 5 จะเป็นค่ามัธยฐานดังนั้น ค่ามัธยฐานของข้อมูลชุดนี้เท่ากับ 12.0

ตัวอย่าง จากข้อมูลตารางแจกแจงความถี่ของคะแนนสอบวิชาสถิติเบื้องต้น จากนิสิตที่สุ่มมา 150 คน จงหาค่าคะแนนสอบเฉลี่ยของวิชานี้ จงหาค่ากลางโดยใช้มัธยฐาน

คะแนนสอบ	ความถี่ (f_i)	ค่ากลาง (X_i)
10 - 19	1	14.5
20 - 29	6	24.5
30 - 39	9	34.5
40 - 49	31	44.5
50 - 59	42	54.5
60 - 69	32	64.5
70 - 79	17	74.5
80 - 89	10	84.5
90 - 99	2	84.5
รวม	150	

วิธีทำ เรียงลำดับข้อมูลจากน้อยไปมาก และเนื่องจาก N = 9 ซึ่งเป็นเลขคี่ ดังนั้น ค่าของข้อมูลตัวที่ = 5 จะเป็นค่ามัธยฐาน

ขั้นตอนที่ 3 หาตำแหน่งของมัธยฐานสำหรับข้อมูลชุดนี้ เนื่องจาก N=150 เป็นเลขคู่ Me จะอยู่ระหว่างตำแหน่งที่ 75 และ 76ซึ่งอยู่ในช่วงคะแนน 50 – 59 หรือจะกล่าวได้ว่า ชั้นที่มีมัธยฐานอยู่ คือ ชั้นที่ 5 (สังเกตจากความถี่สะสม) นั่นคือL_5 = 49.5 N = 150 F = 47 f_5 = 42 I = 10

ขั้นตอนที่ 2 แทนค่าต่าง ๆ ลงในสูตร

$$\text{Me} = L_1 + \left[\frac{\frac{N}{2} - F}{f_1} \right] l$$

$$= 49.5 + \left[\frac{\frac{150}{2} - 47}{42} \right] 10$$

$$= 49.5 + 6.67$$

$$= 56.17$$

ดังนั้น ค่ามัธยฐานของคะแนนสอบวิชานี้เท่ากับ 56.17 คะแนน

คุณสมบัติของค่ามัธยฐาน

1. มัธยฐาน เป็นการใช้ค่าของข้อมูลที่อยู่ตำแหน่งตรงกลาง มาเป็นตัวแทน ดังนั้นข้อมูลที่มีค่ามากหรือน้อยผิดปกติจะไม่มีผลกระทบต่อค่ามัธยฐาน และถ้ามีการเปลี่ยนแปลงข้อมูลบางตัวในกลุ่มจะมีผลกระทบต่อค่ามัธยฐานน้อยมาก

2. มัธยฐาน จะเป็นค่าตัวแทนของข้อมูลได้ใกล้เคียงกับประชากรส่วนใหญ่มากกว่าค่าเฉลี่ย หากการแจกแจงข้อมูลเบ้ไปทางใดทางหนึ่ง

3. ข้อมูลที่มีมาตรวัดเป็นนามบัญญัติ (nominal scale) ไม่สามารถใช้คำนวณหาค่า มัธยฐานได้

4. กรณีที่มีข้อมูลกระจุกอยู่ที่ค่าต่ำสุด หรือสูงสุดมากเกินไปจะไม่สามารถหาค่ามัธยฐานได้ เช่น น้ำหนักของผู้ป่วย 9 คน เป็น 55 55 55 55 55 60 65 70 72 ค่ามัธยฐานเป็น 55 ซึ่งไม่ได้เป็นค่าของข้อมูลที่อยู่ครึ่งหนึ่งตามความหมายของมัธยฐาน

1.8.3 ฐานนิยม (mode) เป็นค่าที่มีความถี่สูงสุดในข้อมูลชุดหนึ่ง ฐานนิยมอาจมีค่าเดียวในชุดข้อมูลนั้น หรืออาจมีหลายค่าได้กรณีที่มีข้อมูลที่ความถี่สูงสุดเท่ากันหลายค่ากรณีที่ข้อมุลมีการแจกแจงความถี่

$$Mo = L1 + \left[\frac{d1}{d1 + d2}\right]I$$

โดยที่ L แทนขอบเขตล่างของช้นั ที่มคี วามถ่สี ูง สดุ

I แทนความกว้างของชั้น

d1 แทนผลต่างระหว่างความถี่ของชั้นที่มีความถี่สูงสุด กับชั้นติดกันที่มีข้อมูลต่ำกว่า

d2 แทนผลต่างระหว่างความถี่ของชั้นที่มีความถี่สูงสุด กับชั้นติดกันที่มีข้อมูลสูงกว่า

ตัวอย่าง 4 จากข้อมูลในตัวอย่าง 1.7 จงหาค่ากลางโดยใช้ฐานนิยม

วิธีทำ คำนวณค่าฐานนิยมของคะแนนสอบวิชาสถิติเบื้องต้น ดังนี้

ขั้นตอนที่ 1 หาชั้นที่มีฐานนิยมสำหรับข้อมูลชุดนี้ พบว่า ช่วงคะแนน 50 – 59 จะมีความถี่มากที่สุด หรือจะกล่าวได้ว่า ชั้นที่มีฐานนิยมอยู่คือ ชั้นที่ 5 นั่น คือ

L_5 = 49.5 N d1 = 42 - 31 = 31 d2 = 42 - 32 l = 10

ขั้นตอนที่ 2 แทนค่าต่าง ๆ ลงในสูตร

$$Mo = L1 + \left[\frac{d1}{d1 + d2}\right]I$$

$$= 49.5 + \left[\frac{11}{11+10}\right]10$$

$$= 49.5 + 5.24$$

$$= 54.74$$

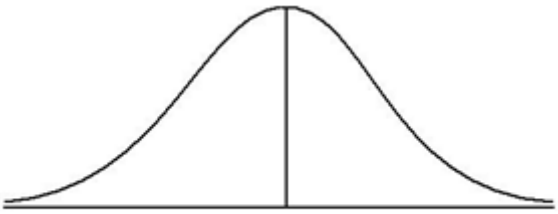
ดังนั้น ค่าฐานนิยมของคะแนนสอบวิชานี้เท่ากับ 54.74 คะแนน

คุณสมบัติของฐานนิยม

1. สามารถคำนวณได้ง่าย รวดเร็ว
2. ใช้กับข้อมูลที่มีมาตรวัดนามบัญญัติ(norminal scale)
3. ข้อมูลที่มีค่ามาก หรือน้อยผิดปกติ จะไม่มีผลกระทบต่อค่าฐานนิยม และถ้ามีการเปลี่ยนแปลงข้อมูลบางตัวในกลุ่มจะไม่มีผลกระทบต่อค่าฐานนิยม หรือมีน้อยมาก

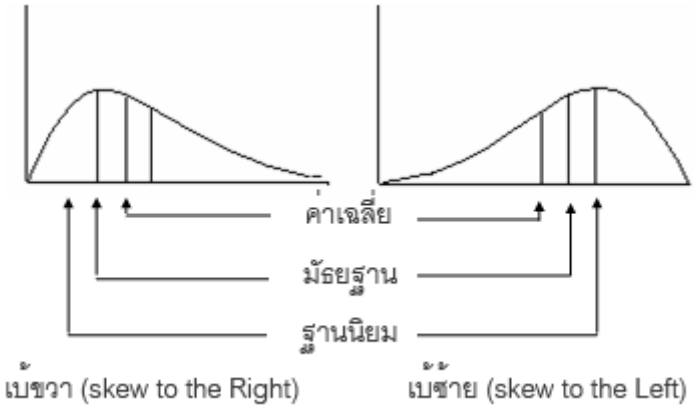
ความสัมพันธ์ระหว่าง Mean , Median และ Mode

• ข้อมูลที่มีการกระจายอย่างสมมาตรค่า Mean Median และ Mode จะเท่ากันหรืออยู่ใกล้เคียงกัน



Mean, Median, Mode

• ข้อมูลที่มีการกระจายไม่เท่ากัน ค่า Mean ,Median และ Mode จะแตกต่างกัน



แบบฝึกหัดประจำบทเรียน

1. ลักษณะของตัวอย่างที่ดีที่สามารถนำมาศึกษาได้ควรเป็นอย่างไร
2. จงบอกถึงข้อดี ข้อเสีย ของค่าเฉลี่ย ฐานนิยมและมัธยฐาน
3. ทำการสุ่มตัวอย่างนักศึกษาวิชาเอกสถิติจำนวน 10 คน ได้ข้อมูล ดังนี้

คนที่	อายุ	%fat	เพศ	grade
1	23	9.5	M	3.58
2	27	27.9	M	2.52
3	29	7.8	F	2.89
4	39	17.8	M	1.86
5	41	31.4	F	3.25
6	45	25.9	F	2.66
7	50	27.4	F	2.99
8	28	25.2	M	3.41
9	20	31.1	M	1.75
10	22	34.7	M	3.00

3.1 อยากทราบว่าตัวแปร อายุ, % fat,เพศ, gradeเป็นข้อมูลประเภทใด และอยู่ใน scale ไหน

3.2 จงสร้าง histogram ของตัวแปร age ถ้ากำหนดให้ความกว้างของแต่ละชั้นเป็น 5 และ 10 ตามลำดับ

4. จงหาค่าเฉลี่ยเลขคณิต และมัธยฐานของข้อมูลต่อไปนี้

5.1 5 4 8 3 7 2 9

5.2 18.3 20.6 19.3 22.4 20.2 18.8 16.7 20.0

5. เลขชุดหนึ่งประกอบด้วยเลข 6.2 5.1 8.7 6.2 4.1 3.3 5.4 6.2 6.7 และ 9.6

5.1 จงหา Mean Mode Median

5.2 คูณทุกตัวด้วย 10 แล้วหา Mean Mode Median

5.3 บวกทุกตัวด้วย 10 แล้วหา Mean Mode Median

6. จากการสำรวจคนงานในโรงงานหนึ่ง พบว่าคนงานชายและหญิงทั้งหมด มีรายจ่ายค่าอาหารเย็นเฉลี่ยคนละ 12 บาท โดยแยกเป็นคนงานชายทั้งหมด 20 คน มีรายจ่ายค่าอาหารเย็นเฉลี่ยคนละ 15 คน ส่วนคนงานหญิงมีรายจ่ายค่าอาหารเย็นคนละ 11 บาท จงหาว่าโรงงานแห่งนี้มีคนงานหญิงกี่คน