

หน่วยที่

7

---

## ความรู้พื้นฐานในการวิเคราะห์ข้อมูล

โดย รองศาสตราจารย์ ดร.คณิต ไข่มุกด์



## ความหมายของข้อมูล และตัวแปร

ข้อมูลเป็นข้อเท็จจริงที่เก็บรวบรวมได้ โดยใช้ตัวแปรเป็นสัญลักษณ์แทนค่าของข้อมูล ระดับของข้อมูลจะเป็นตัวกำหนดวิธีการวิเคราะห์ข้อมูล

### 1. ข้อมูลคืออะไร

ข้อมูลเป็นข้อเท็จจริงที่สังเกตได้ ในชีวิตประจำวันของเรามีข้อมูลเข้ามาให้สังเกตตลอดเวลา ถ้ามนุษย์เราเป็นผู้สังเกต เราสามารถสังเกตได้จากจากประสาทสัมผัสทั้ง 6 คือ หู ตา กาย จมูก ลิ้น และใจ เริ่มจากตื่นขึ้นมาเพราะได้ยินเสียงสุนัขเห่า(หูเป็นเครื่องมือสังเกต) พอลืมตาขึ้นมาก็เห็นทีวีเปิดอยู่(ตาเป็นเครื่องมือสังเกต) เลยใช้มือคลำข้างๆ ตัวหารีโมท(กายเป็นเครื่องมือสังเกต, มือสัมผัส) เมื่อเจอแล้วก็กดปุ่มปิดทีวี ปรากฏว่าขณะเดียวกันก็ได้กลิ่นกาแฟ(จมูกเป็นเครื่องมือสังเกต) จึงลุกจากที่นอนไปตามกลิ่นกาแฟ เมื่อเห็นแล้วก็ลองจิบดู(ลิ้นเป็นเครื่องมือสังเกต) ผลปรากฏว่าขมมากเพราะไม่ได้ใส่น้ำตาล จึงคิดไปว่าใครลุกขึ้นมาชงกาแฟแล้วลืมทิ้งไว้(ใจเป็นเครื่องมือสังเกต)

จะเห็นว่าข้อมูลที่สังเกตได้และพื้นฐานประสบการณ์จะเป็นตัวตัดสินใจและอธิบายเหตุการณ์ที่เกิดขึ้น ดังนั้นการตัดสินใจที่เกิดขึ้นจะถูกหรือผิดก็ขึ้นอยู่กับข้อมูลที่สังเกตได้และประสบการณ์ของแต่ละคน

ถ้าลองถืกลงไปจะเห็นว่าข้อมูลที่สังเกตได้มีทั้งเป็นลักษณะร่องรอย รูป รส กลิ่น เสียง สิ่งที่สังเกตได้ อาจจะเป็น เสียง ภาพ วัตถุ สี กลิ่น อุณหภูมิ ความรู้สึกและรส การบันทึกข้อมูลที่สังเกตได้ อาจจะเป็น ข้อความบรรยาย ภาพ ตัวเลข กราฟ สื่อแม่เหล็ก หรือสื่อแสงเลเซอร์ ก็ได้ ทั้งหมดนี้เป็นข้อเท็จจริงที่สังเกตได้จากการใช้มนุษย์เป็นผู้สังเกตจากเครื่องมือการรับรู้ที่มีอยู่ในตัวเอง และเครื่องมือที่มนุษย์สร้างขึ้น

ความจริงข้อมูลที่สังเกตได้ถ้าเรามีเครื่องมือที่ตีพอหรือเหมาะกับการสังเกตข้อเท็จจริงนั้นก็จะได้ข้อมูลจากการสังเกตมากกว่าเราสังเกตเอง เช่น ใช้สุนัขดมกลิ่นยาเสพติด หรือค้นหาบุคคล ใช้โทรศัพท์เพื่อรับสัญญาณเสียงจากที่ไกลๆ ใช้เครื่องรับโทรทัศน์เพื่อรับสัญญาณภาพและสัญญาณเสียงในอากาศที่คนรับไม่ได้ ใช้เครื่องรับวิทยุเพื่อรับสัญญาณเสียงที่คนรับไม่ได้ ใช้

เครื่องวัดกัมมันตภาพรังสี วัดขนาดกัมมันตภาพรังสี แม้แต่ใช้เครื่องรับฟังคลื่นแม่เหล็กไฟฟ้าจากนอกโลก และการใช้ดาวเทียมถ่ายภาพความเคลื่อนไหวของโลกทุกวัน

ข้อเท็จจริงที่สังเกตได้เหล่านี้เป็นข้อมูลทั้งสิ้น ถ้าเราจำแนกข้อมูลโดยใช้เกณฑ์ที่สามารถแสดงเป็นตัวเลขได้กับแสดงเป็นตัวเลขไม่ได้ เราจะเรียกข้อมูลที่แสดงเป็นตัวเลขได้ว่าข้อมูลเชิงปริมาณ และข้อมูลที่ไม่สามารถแสดงเป็นตัวเลขได้ เรียกว่าข้อมูลเชิงคุณลักษณะหรือเชิงคุณภาพ

## 2. ตัวแปรคืออะไร

จะเห็นว่าข้อมูลนอกจากจะมีมากมายและหลากหลายแล้ว การนำมาจัดเป็นหมวดหมู่แล้วตั้งชื่อเป็นตัวแปรเป็นเรื่อยๆ ไป ก็สามารถที่จะนำไปทำการวิเคราะห์ได้ง่ายและเป็นระเบียบ เช่น ข้อมูลปริมาณน้ำฝนแต่ละเดือนในรอบปีที่ผ่านมาที่จังหวัดนครราชสีมา ตั้งเป็นชื่อตัวแปรว่า RAINFALL ก็จะเข้าใจกันได้ง่ายขึ้น ว่าค่าที่สังเกตได้ของตัวแปรนี้คือปริมาณน้ำฝนแต่ละเดือน มีจำนวนข้อมูลเท่ากับ 12 ค่า ข้อมูลผลการสอบกลางภาคของนักศึกษาที่เรียนวิชา เทคโนโลยีสารสนเทศ 2 ตั้งเป็นชื่อตัวแปรว่า SCORE ก็จะเข้าใจว่าคะแนนที่แต่ละคนได้ของตัวแปรนี้คือคะแนนในการสอบกลางภาคที่ผ่านมา จะมีจำนวนข้อมูลเท่ากับจำนวนนักศึกษาที่เข้าสอบวิชานี้

ข้อมูลจากหัวข้อการประเมินการเรียนการสอนจากนักศึกษามีหลายประเด็นอาจจะจำแนกเป็นหลายตัวแปรตามแต่ละข้อคำถาม ข้อคำถามละ 1 ตัวแปร ได้แก่

X1 แทน การแจ้งเค้าโครงรายวิชาต้นภาคการศึกษา

X2 แทน ความถี่ในการแจ้งเนื้อหาการสอนแต่ละครั้ง

X3 แทน ความถี่ในการประเมินความรู้ของนักศึกษาก่อนเรียน

X4 แทน การแนะนำเรื่องที่จะเรียนด้วยการบอกความจำเป็น ชื่อเรื่อง หัวข้อที่จะเรียน และวัตถุประสงค์ที่จะเรียนในแต่ละครั้ง

X5 แทน การนำเข้าสู่บทเรียนด้วยการยกประเด็น เรื่องราว รายการที่เกี่ยวข้องกับเรื่องที่จะสอน

.....

ดังนั้น โดยสรุปตัวแปรก็คือชื่อของกลุ่มข้อมูลประเภทเดียวกัน ตั้งขึ้นเพื่อประโยชน์ในการอ้างถึงหรือนำไปวิเคราะห์ข้อมูล

## 3. หลักการแบ่งระดับของข้อมูล

ถ้าเราพิจารณาข้อเท็จจริงที่สังเกตได้แล้วจะเห็นว่าข้อมูลที่ได้รับจะแตกต่างกัน เช่น ชื่อของนักศึกษา 30 คน ไม่ได้ให้ข้อมูลอะไรมากไปกว่าบอกความแตกต่างของแต่ละคนด้วยชื่อเท่านั้น

แต่ถ้าเราให้นักศึกษาทั้ง 30 คนมาเรียงกันตามลำดับความสูง เราก็จะบอกลำดับที่ตามความสูงได้ว่าใน 30 คนนี้ใครอยู่ลำดับความสูงที่เท่าใด แต่ถ้าเราให้นักศึกษาทั้ง 30 คนสอบ เราก็จะได้คะแนนของแต่ละคนที่บอกความแตกต่างกันได้ว่าแต่ละคนได้คะแนนแตกต่างกันก็คะแนน และถ้าเราให้ทุกคนวัดความสูงของแต่ละคนเป็นเซ็นติเมตรแล้วบันทึกไว้ ค่าความสูงก็จะบอกความแตกต่างกันได้ว่าแต่ละคนสูงแตกต่างกันเท่าใด จะเห็นว่าข้อมูลจากตัวแปร ชื่อ ลำดับความสูง คะแนนสอบ และความสูง แสดงรายละเอียดหรือความแตกต่างไม่เท่ากัน

ดังนั้นเราจึงยึดหลักความสามารถในการบอกรายละเอียดความแตกต่างของข้อมูลในการแบ่งระดับของข้อมูล ดังนี้

ถ้าข้อมูลบอกความแตกต่างได้แต่เพียงชื่อ เป็นการจำแนกความแตกต่างที่หยาบที่สุด เช่น ชื่อของนักศึกษาทั้ง 30 คน ไม่ได้ให้รายละเอียดอะไรนอกจากว่าแต่ละคนชื่อแตกต่างกันเท่านั้น เราเรียกข้อมูลที่สามารถจำแนกได้แค่ชื่อว่า ข้อมูลระดับมาตรานามบัญญัติ(Nominal scale)

ถ้าข้อมูลบอกลำดับความแตกต่างได้ เป็นการจำแนกความแตกต่างที่ละเอียดขึ้นเพราะบอกลำดับที่ความแตกต่างได้ เช่น ชื่อของนักศึกษาที่เรียงตามลำดับความสูงทั้ง 30 คน เราเรียกข้อมูลที่สามารถจำแนกถึงลำดับความแตกต่างได้นี้ว่า ข้อมูลระดับมาตราอันดับบัญญัติ(Ordinal scale)

ถ้าข้อมูลบอกความแตกต่างเป็นหน่วยที่เท่ากันได้ เป็นการจำแนกความแตกต่างที่ละเอียดขึ้นไปอีกเพราะสามารถบอกปริมาณที่แตกต่างกันเป็นปริมาณหน่วยที่เท่ากัน ทำให้บอกระดับขึ้นความแตกต่างที่ละเอียดมากเพราะบอกได้ว่าแต่ละคนแตกต่างกันเป็นปริมาณเท่าใดโดยเปรียบเทียบกับหน่วยปริมาณที่เท่ากันได้ เช่น คะแนนสอบของนักศึกษาทั้ง 30 คน เราเรียกว่าข้อมูลที่สามารถจำแนกถึงหน่วยที่แตกต่างกันได้ว่า ข้อมูลระดับมาตราช่วงบัญญัติ(Interval scale)

ถ้าข้อมูลบอกความแตกต่างเป็นหน่วยที่เท่ากันได้และค่าศูนย์เป็นค่าศูนย์ที่แท้จริง (Absolute Zero) เช่น ความสูงเป็นเซ็นติเมตรของนักศึกษาทั้ง 30 คน เราเรียกว่าข้อมูลที่สามารถจำแนกเป็นปริมาณที่แตกต่างเท่ากันและมีศูนย์แท้ว่า ข้อมูลระดับมาตราอัตราส่วนบัญญัติ(Ratio scale)

---

## สรุป

ตัวแปรเป็นชื่อของกลุ่มข้อมูลประเภทเดียวกันที่มีค่าแตกต่างกันไป ข้อมูลเป็นชื่อเท็จจริงหรือค่าแต่ละค่าของตัวแปรซึ่งเปลี่ยนไปตามชื่อเท็จจริงที่สังเกตได้ ความสามารถในการจำแนกความแตกต่างของข้อมูลที่สังเกตได้ เป็นตัวแสดงระดับของข้อมูล เพราะความสามารถในการจำแนกข้อมูลที่แตกต่างกันนำไปสู่การใช้สถิติที่ถูกต้องในการวิเคราะห์ข้อมูล

---

## กิจกรรมต่อเนื่อง ตอนที่ 7.1

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.1 ความหมายของข้อมูล และตัวแปร
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.1 ความหมายของข้อมูล และตัวแปร
3. ทำกิจกรรมประกอบการเรียนในกลุ่มการเรียนรู้ประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.1 ความหมายของข้อมูล และตัวแปร
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.1 ความหมายของข้อมูล และตัวแปร

## ระดับมาตราของข้อมูล

การแบ่งระดับของข้อมูล จะทำให้สามารถเลือกวิธีวิเคราะห์ทางสถิติที่เหมาะสมได้ โดยแบ่งระดับการวัดข้อมูลออกเป็น 4 ระดับ ได้แก่ (1) ข้อมูลระดับมาตรานามบัญญัติ(Nominal scale) (2) ข้อมูลระดับมาตราอันดับบัญญัติ(Ordinal scale) (3) ข้อมูลระดับมาตราช่วงบัญญัติ(Interval scale) และ (4) ข้อมูลระดับมาตราอัตราส่วนบัญญัติ(Ratio scale)

### 1. ข้อมูลระดับมาตรานามบัญญัติ

ข้อมูลระดับมาตรานามบัญญัติ เป็นข้อมูลระดับที่หยาบที่สุดเพราะสามารถจำแนกความแตกต่างได้ด้วยชื่อเท่านั้น ตัวอย่างเช่น ตัวแปรการนับถือศาสนา ประกอบด้วยข้อมูล ชื่อศาสนา เช่น พุทธ คริสต์ อิสลาม และอื่นๆ ตัวแปรเพศ ประกอบด้วย ข้อมูล เพศชาย เพศหญิง ตัวแปรรหัสนักศึกษา ข้อมูล B4270911 B4270584 B4270508 และอื่นๆ ตัวแปรเบอร์เสื้อทีมักฟุตบอล ตัวเลขบนเสื้อแทนตำแหน่งในการเล่นฟุตบอล เลข 1 ถึง 90 บนเสื้อนักฟุตบอล ไม่ได้มีความหมายในการบอกความแตกต่างมากนักแสดงเพียงความแตกต่างโดยชื่อเท่านั้น ถึงแม้ เบอร์ 1 หมายถึงผู้รักษาประตู เบอร์ 7 หมายถึง เบอร์นำโชค เบอร์ 9 หมายถึง กองหน้า เบอร์ 10 หมายถึง เบอร์ทำเกมส์ เป็นต้น แต่เบอร์เสื้อไม่ได้บอกค่าความแตกต่าง เช่น เลข 1 น้อยกว่าเลข 2 หรือ เลข 1 ดีกว่า เลข 2 เป็นความหมายแต่เพียงว่านักฟุตบอลแต่ละคนมีหมายเลขเรียกแทนชื่อเท่านั้น

### 2. ข้อมูลระดับมาตราอันดับบัญญัติ

ข้อมูลระดับมาตราอันดับบัญญัติ เป็นข้อมูลระดับที่ละเอียดกว่าข้อมูลระดับมาตรานามบัญญัติ เพราะสามารถจำแนกความแตกต่างตามอันดับได้ เช่น การจัดอันดับความสวยของนางงามสามารถบอกความแตกต่างได้ว่าใครสวยกว่าใคร หรือ ให้นักศึกษามายืนเรียงกัน 30 คนตามลำดับความสูง สามารถจำแนกความแตกต่างตามอันดับความสูงได้ หรือจัดอันดับที่ให้นักศึกษาทั้ง 30 คน เรียงตามลำดับคะแนนเฉลี่ยสะสม ข้อมูลที่สามารถบอกลำดับที่ของข้อมูลแต่ไม่สามารถบอกได้ว่าแต่

ระดับที่นั่นแตกต่างกัน นั่นแตกต่างกันเท่าใด เพียงแต่บอกลำดับที่แตกต่างได้เท่านั้น เราเรียกข้อมูลนี้ว่าข้อมูลระดับมาตราอันดับบัญญัติ

### 3. ข้อมูลระดับมาตราช่วงบัญญัติ

ข้อมูลระดับมาตราช่วงบัญญัติ เป็นข้อมูลระดับที่ละเอียดกว่าข้อมูลระดับมาตราอันดับบัญญัติ เพราะสามารถบอกขนาดความแตกต่างของแต่ละข้อมูลได้ เช่น ข้อมูลคะแนนสอบของนักศึกษา จำนวน 30 คน สมมติว่านาย ก ได้ 20 คะแนน และ นาย ข ได้ 22 คะแนน ก็หมายความว่านาย ข ได้คะแนนมากกว่านาย ก 2 คะแนน

ดังนั้นใน กรณี ที่เป็นคะแนนความคิดเห็น โดยระบุว่า

เห็นด้วยมากที่สุด	ให้คะแนน 5
เห็นด้วยมาก	ให้คะแนน 4
เห็นด้วยปานกลาง	ให้คะแนน 3
เห็นด้วยน้อย	ให้คะแนน 2
เห็นด้วยน้อยที่สุด	ให้คะแนน 1

ถ้าตีความว่าตัวเลขที่กำกับ แสดงถึงหน่วยที่มีความแตกต่างเท่ากัน นั่นคือถ้าตีความว่าการให้คะแนน 5 แทน เห็นด้วยมากที่สุด จะต่างจาก คะแนน 4 ซึ่งแทน เห็นด้วยมาก อยู่ 1 คะแนน เท่ากับ คะแนน 4 แทน เห็นด้วยมาก ต่างจาก คะแนน 3 แทน เห็นด้วยปานกลาง อยู่ 1 คะแนน เท่ากัน ก็จะถือว่าข้อมูลนี้ เป็นข้อมูลระดับมาตราช่วงบัญญัติ เพราะแต่ละช่วงความคิดเห็นถือว่าห่างกัน 1 คะแนนเท่ากัน แต่ถ้าตัวเลขที่กำกับเพียงบอกอันดับความแตกต่างของความคิดเห็นเท่านั้นและไม่ถือว่าความเห็นแต่ละระดับแตกต่างเท่ากันก็จะถือว่าข้อมูลนี้ เป็นข้อมูลระดับมาตราอันดับบัญญัติ ซึ่งก็ยังเป็นข้อถกเถียงกันจนถึงปัจจุบันแล้วแต่ว่าการให้เหตุผลและความเชื่อว่าระดับความเห็นแต่ละช่วงที่กำหนดมีแตกต่างเท่ากันหรือไม่เท่ากัน

ระดับมาตรานำไปสู่ทางเลือกในการใช้สถิติเพื่อการวิเคราะห์ข้อมูล ดังนั้นเรื่องระดับมาตราเป็นเรื่องที่จำเป็นจะต้องเข้าใจอย่างดีจึงจะนำไปเลือกใช้สถิติเพื่อวิเคราะห์ข้อมูลได้ถูกต้อง

### 4. ข้อมูลระดับมาตราอัตราส่วนบัญญัติ

ข้อมูลระดับมาตราอัตราส่วนบัญญัติ เป็นข้อมูลระดับที่ละเอียดสูงสุด และแตกต่างกับข้อมูลระดับช่วงบัญญัติ ก็แต่เพียงว่า ค่าศูนย์เป็นค่าศูนย์ที่แท้จริง(absolute zero) เท่านั้น ตัวอย่างเช่น คะแนนสอบเป็นข้อมูลระดับมาตราช่วงบัญญัตินักศึกษาแต่ละคนได้คะแนนแตกต่างกันตามจำนวนหน่วยคะแนนที่ได้ แต่คะแนนศูนย์ ไม่ได้เป็นค่าศูนย์ที่แท้จริงเพราะว่า คะแนนศูนย์ คือทำข้อสอบผิดหมดทุกข้อเท่านั้นเอง ในความเป็นจริงเขาอาจจะมีความรู้อีกมากแต่ไม่ตรงกับข้อสอบก็ได้เราจึง

สมมติว่าเขาได้ ศูนย์ ดังนั้นคะแนนสอบก็เป็นข้อมูลระดับมาตราช่วงบัญญัติเท่านั้นเพราะคะแนน ศูนย์เป็นค่าศูนย์ที่สมมติขึ้น แต่ถ้าเปรียบเทียบกับ ความสูงเป็นเซ็นติเมตรของนักศึกษา เป็นข้อมูล ระดับมาตราอัตราส่วนบัญญัติ เพราะว่า นักศึกษาที่มีความสูงเป็นศูนย์ ก็คือ ผู้ที่ไม่มี ความสูงจริงๆ ดังนั้น ค่าศูนย์ของความสูง เป็น ค่าศูนย์แท้ ดังนั้นข้อมูลระดับอัตราส่วนก็คือข้อมูลระดับช่วง บัญญัติที่ศูนย์เป็นศูนย์แท้

สถิติที่ใช้ในการวิเคราะห์ข้อมูล จำแนกตามระดับของข้อมูล แบ่งออกเป็น 3 กลุ่ม คือ

กลุ่มที่ 1 สถิติที่ใช้ในการวิเคราะห์ข้อมูลระดับมาตรานามบัญญัติ

กลุ่มที่ 2 สถิติที่ใช้ในการวิเคราะห์ข้อมูลระดับมาตราอันดับบัญญัติ

กลุ่มที่ 3 สถิติที่ใช้ในการวิเคราะห์ข้อมูลระดับมาตราช่วงบัญญัติ และ อัตราส่วนบัญญัติ

ดังนั้นในโปรแกรม SPSS ก็แบ่งประเภทของข้อมูลออกเป็น 3 ระดับเช่นเดียวกัน ถ้าดูใน เมนู ก็คือ Nominal, Ordinal และ Scale

Nominal หมายถึง ข้อมูลระดับมาตรานามบัญญัติ

Ordinal หมายถึง ข้อมูลระดับมาตราอันดับบัญญัติ

Scale หมายถึง ข้อมูลระดับมาตราช่วงบัญญัติ และ อัตราส่วนบัญญัติ

การแบ่งระดับข้อมูล แบ่งตามความสามารถในการบอกความแตกต่างของข้อมูลว่าบอกความแตกต่างได้ ละเอียดมากน้อยต่างกันเพียงใด ตั้งแต่ความสามารถในการบอกความแตกต่างได้หยาบที่สุดข้อมูลคือนามบัญญัติ หรือระดับต่ำสุดจนถึงข้อมูลระดับที่สามารถบอกความแตกต่างได้ละเอียดที่สุด คืออัตราส่วนบัญญัติ หรือระดับ สูงสุด ระดับมาตราที่สูงกว่าสามารถใช้ได้กับสถิติ ที่ใช้ได้กับข้อมูลระดับมาตราที่ต่ำกว่า พอจะสรุปสถิติวิเคราะห์ที่ จำแนกตามระดับข้อมูล ได้เป็นตารางดังนี้

ข้อมูลระดับมาตรา	สถิติวิเคราะห์เชิงพรรณนา	สถิติวิเคราะห์เชิงอนุมาน
นามบัญญัติ	การแจกแจงความถี่ ร้อยละ ฐานนิยม	(1 popn) $\chi^2 - test$ Binomial test (2 popn) McNemar test (3 popn) Cochran's Q test
อันดับบัญญัติ	มัธยฐาน ส่วนเบี่ยงเบนควอไทล์ Rank correlation	(1 popn) Kolmogorov-Smirnov test One-Sample Runs test (2 popn) - Sign test - Wilcoxon Signed Ranks test - Median test - Mann-Whitney U-test - Kolmogorov-Smirnov Sample test

		- Wald-Wolfowitz Runs test - Moses test of Extreme Reactions (3 popn) - Friedman Two-way ANOVA - Kendall Coefficient of Concordance - Median test - Kruskal-Wallis One-way ANOVA
ช่วงและอัตราส่วนบัญญัติ	ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน Correlation	Z-test t-test F-test

### สรุป

ระดับความแตกต่างในการจำแนกข้อมูลว่าจำแนกความแตกต่างได้เพียงชื่อ จำแนกข้อมูลได้ถึงอันดับที่ จำแนกข้อมูลได้ถึงระดับหน่วยความแตกต่างโดยมีค่าศูนย์สมมติ และจำแนกข้อมูลได้ถึงระดับหน่วยความแตกต่างโดยมีค่าศูนย์สัมบูรณ์ กำหนดได้ด้วย ระดับมาตรานามบัญญัติ ระดับมาตราอันดับบัญญัติ ระดับมาตราช่วงบัญญัติ และระดับมาตราอัตราส่วนบัญญัติ ตามลำดับ

ระดับมาตราที่แตกต่างกันของข้อมูลมีผลต่อการเลือกสถิติในการวิเคราะห์ข้อมูล ดังการที่จะวิเคราะห์ข้อมูลได้ถูกต้องจะต้องเข้าใจในเรื่องระดับมาตราของข้อมูลเป็นอย่างดี

---

## กิจกรรมต่อเนื่อง ตอนที่ 7.2

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.2 ระดับมาตราของข้อมูล
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.2 ระดับมาตราของข้อมูล
3. ทำกิจกรรมประกอบการเรียนในกลุ่มการเรียนรู้ประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.2 ระดับมาตราของข้อมูล
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.2 ระดับมาตราของข้อมูล



## การวิเคราะห์ข้อมูลทางสถิติ

การวิเคราะห์ข้อมูลทางสถิติสามารถแบ่งออกได้เป็น 2 ประเภท ได้แก่ (1) สถิติเชิงพรรณนา เป็นสถิติที่เกี่ยวข้องกับการวิเคราะห์ และการบรรยายถึงลักษณะข้อมูลที่เก็บรวบรวมได้ และ (2) สถิติเชิงอนุมาน เป็นสถิติที่เกี่ยวข้องกับกลุ่มตัวอย่าง ที่ได้มาจากการสุ่มตัวอย่าง(random sample) แล้วนำผลที่ได้ไปอ้างอิงถึงประชากร(population)

### 1. สถิติเชิงพรรณนา

สถิติเชิงพรรณนา เป็นสถิติที่แสดงข้อสรุปของข้อมูล โดยไม่ได้นำเอาทฤษฎีความน่าจะเป็น เข้ามามีส่วนในการพยากรณ์ ข้อสรุปของประชากร จึงไม่มีการอ้างอิงถึงตัวพารามิเตอร์(parameter)สถิติเชิงพรรณนา ประกอบด้วย เนื้อหา ใหญ่ๆ 3 เรื่อง คือ

- ก. การจัดอันดับข้อมูล
- ข. การวัดแนวโน้มเข้าสู่ส่วนกลาง
- ค. การวัดการกระจาย

**การจัดอันดับข้อมูล** เริ่มจากการทำการแจกแจงข้อมูลเป็นกลุ่ม ได้แก่ ตารางแจกแจงความถี่ ร้อยละของความถี่ เพื่อจัดอันดับของตำแหน่งข้อมูล ประกอบด้วยตารางแจกแจงความถี่สะสม ร้อยละของความถี่สะสมหรือเปอร์เซ็นต์ไทล์(percentile)

**การวัดแนวโน้มเข้าสู่ส่วนกลาง** ประกอบด้วยค่าที่เป็นตัวแทนของกลุ่มข้อมูล ได้แก่ ค่าฐานนิยม(mode) ค่ามัธยฐาน(mode) และค่าเฉลี่ย(average หรือ mean)

**การวัดการกระจาย** เป็นค่าที่บอกความแตกต่างของข้อมูล ว่าห่างออกไปจากค่าที่เป็นตัวแทนของกลุ่มเพียงใด ได้แก่ พิสัย(range) ส่วนเบี่ยงเบนควอดไทล์(quate deviation) ส่วนเบี่ยงเบนเฉลี่ย(average deviation) และส่วนเบี่ยงเบนมาตรฐาน( standard deviation)

สถิติเชิงพรรณนาที่จำเป็นและมีที่ใช้กันมาก ก็คือ เปอร์เซนต์ความถี่สะสมหรือ เปอร์เซนต์ไทล์ ค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐาน สำหรับค่าฐานนิยมมีที่ใช้บ้างกรณีข้อมูลระดับมาตรานามบัญญัติ ค่ามัธยฐานมีที่ใช้ในกรณีข้อมูลระดับมาตราอันดับบัญญัติ และกรณีข้อมูลระดับมาตราช่วงบัญญัติหรืออัตราส่วนบัญญัติที่ค่าของข้อมูลสุดโต่ง ไม่ว่าจะเป็นค่าต่ำสุดหรือสูงสุดเพราะถ้าใช้ค่าเฉลี่ยจะทำให้ไม่ได้ค่ากลางที่เป็นตัวแทนข้อมูลที่แท้จริง

ตัวอย่าง จงคำนวณค่าสถิติเชิงพรรณนา ของ คะแนนสอบจากนักศึกษา จำนวน 30 คนซึ่งคะแนน ผลการสอบเป็นดังนี้ 1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 9, 9, 10

วิธีทำ ค่าเฉลี่ยเลขคณิต  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{165}{30} = 5.5$

$$\text{ค่าส่วนเบี่ยงเบนมาตรฐาน } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x)^2}{n(n - 1)}} = 2.2552$$

ค่าฐานนิยม คือค่าคะแนนที่มีความถี่สูงสุด ได้แก่ 5 และ 6

$$\begin{aligned} \text{ค่าความเบ้ (Skewness)} &= \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n - 1)s^3} \\ &= .000 \end{aligned}$$

$$\begin{aligned} \text{ค่าความโด่ง (Kurtosis)} &= \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n - 1)s^4} = 2.2605 \end{aligned}$$

ค่าเปอร์เซ็นต์ไทล์(Percentiles)คือค่าของคะแนนที่ตรงกับอันดับของตำแหน่งที่กำหนด เช่น เปอร์เซ็นต์ไทล์ที่ 10 คือ ค่าของคะแนนที่ตรงกับ ตำแหน่งที่ 10 ใน ร้อย นั่นคือเมื่อเรียงลำดับ คะแนนจากน้อยไปมาก ค่าคะแนนนี้จะมีค่ามากกว่าคนอื่นอยู่ 10 เปอร์เซนต์ ปกติตำแหน่งของ เปอร์เซ็นต์ไทล์จะเขียนเป็นตัวห้อยกับค่า P เช่น เปอร์เซ็นต์ไทล์ที่ 10 เขียนเป็น  $P_{10}$

จากสูตรคำนวณ ค่าเปอร์เซ็นต์ไทล์ จากคะแนนดิบ ให้เรียงคะแนนจากมากไปน้อยและคำนวณ ตำแหน่งเปอร์เซ็นต์ไทล์ ที่ k ว่าตรงกับคะแนนใด คะแนนนั้นก็จะเป็นค่าของ  $p_k$

$$\text{ค่าที่ตรงกับเปอร์เซ็นต์ไทล์ที่ } k \text{ คือค่าคะแนนที่ตรงกับลำดับ คะแนนที่ } (n + 1) \frac{k}{100}$$

$$\text{ค่าที่ตรงกับเปอร์เซ็นต์ไทล์ที่ } 10 \text{ คือค่าคะแนนที่ตรงกับลำดับ คะแนนที่ } (30 + 1) \frac{10}{100}$$

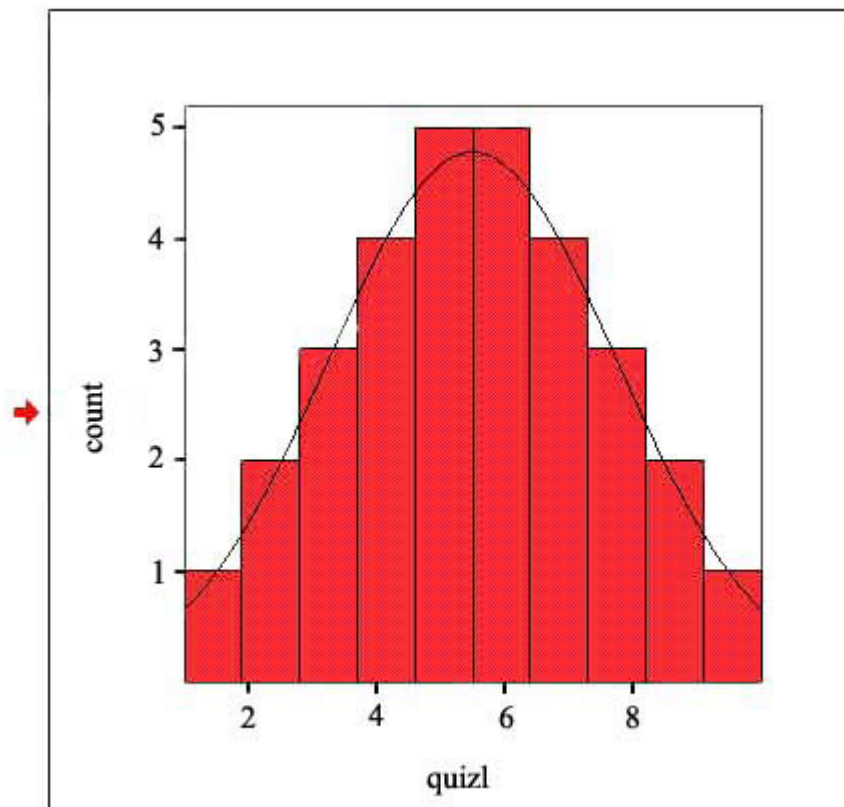
นั่นคือ คะแนนลำดับที่ 3.1 ตรงกับคะแนน 2.1 ฉะนั้น  $P_{10}=2.1$  ทำนองเดียวกัน

$$P_{20} = 3.2 \quad P_{30} = 4 \quad P_{40} = 5 \quad P_{50} = 5.5 \quad P_{60} = 6 \quad P_{70} = 7 \quad P_{80} = 7.8 \\ P_{90} = 8.9$$

เมื่อนำข้อมูลมาทำเป็นตารางแจกแจงความถี่จะได้ ดังตารางต่อไปนี้

		Frequenc y	Percent	Valid Percent	Cumulative Percent
Valid	1	1	3.3	3.3	3.3
	2	2	6.7	6.7	10.0
	3	3	10.0	10.0	20.0
	4	4	13.3	13.3	33.3
	5	5	16.7	16.7	50.0
	6	5	16.7	16.7	66.7
	7	4	13.3	13.3	80.0
	8	3	10.0	10.0	90.0
	9	2	6.7	6.7	96.7
	10	1	3.3	3.3	100.0
Total	30	100.0	100.0		

เมื่อนำมา แสดงเป็นฮิสโตแกรม เทียบกับโค้งปกติมาตรฐาน จะได้ดังรูป



## 2. สถิติเชิงอนุมาน

สถิติเชิงอนุมาน มีจุดเน้นที่จะศึกษาถึงค่าพารามิเตอร์โดยอาศัยทฤษฎีความน่าจะเป็น โดยทั่วไปเราจะแบ่งเนื้อหาและวิธีการทดสอบ ตามจำนวนตัวแปรที่นำมาทดสอบหรือนำมาวิเคราะห์

- ก. การทดสอบตัวแปรเดียว ที่สำคัญและใช้กันบ่อย ได้แก่
  - ก.1 การทดสอบค่าเฉลี่ย โดย z-test, t-test
  - ก.2 การทดสอบค่าสัดส่วน โดย t-test
  - ก.3 การทดสอบการแจกแจงแบบปกติ โดย  $\chi^2$ -test
- ข. การทดสอบสองตัวแปร ที่สำคัญและใช้กันบ่อย ได้แก่
  - ข.1 การทดสอบความแตกต่างของค่าเฉลี่ย โดย z-test, t-test
  - ข.2 การทดสอบความแตกต่างของค่าสัดส่วน โดย t-test
  - ข.3 การทดสอบความเป็นอิสระ(independent test) โดย  $\chi^2$ -test
- ค. การทดสอบตั้งแต่สามตัวแปรขึ้นไป ที่สำคัญและใช้กันบ่อย ได้แก่
  - การทดสอบความแตกต่างของค่าเฉลี่ย โดยการวิเคราะห์ความแปรปรวน หรือ F-test

**ตัวอย่าง** จากข้อมูลเดิมนำมาทดสอบว่าข้อมูลที่ได้เป็น โค้งปกติหรือไม่ ปรากฏว่าไม่สามารถปฏิเสธสมมติฐานว่างที่ว่า การแจกแจงของข้อมูลเป็น โค้งปกติไม่ว่าจะเป็นการทดสอบด้วย Kolmogorov-Smirnov หรือ Shapiro-Wilk

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
quiz1	.088	30	.200	.977	30	.756

\* This is a lower bound of the true significance.

a. Lilliefors Significance Correction

ดังนั้นการที่เรามองว่าข้อมูลสมมาตรกันอย่างเดียวไม่เพียงพอต่อการสรุปว่าเป็น โค้งปกติ เพราะจุดสำคัญอยู่ที่ ความโด่งด้วย โค้งปกติจะมี ความเบ้เท่ากับ 0 และความโด่งเท่ากับ 3

หรือสามารถทดสอบง่ายๆ ได้โดย หาอัตราส่วนของความเบ้กับค่าคลาดเคลื่อนมาตรฐาน ถ้าอัตราส่วนอยู่ระหว่าง  $\pm 2$  ก็ประมาณได้ว่า การแจกแจงนี้เป็นการแจกแจงแบบปกติ

---

## สรุป

---

### กิจกรรมต่อเนื่อง ตอนที่ 7.3

1. ศึกษาผ่านสื่อปฏิสัมพันธ์ผ่านจอภาพ การบรรยาย หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.3 การวิเคราะห์ข้อมูลทางสถิติ
2. ศึกษาด้วยตนเองผ่านสื่อบทเรียนคอมพิวเตอร์ช่วยสอน หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.3 การวิเคราะห์ข้อมูลทางสถิติ
3. ทำกิจกรรมประกอบการเรียนในกลุ่มการเรียนรู้ประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.3 การวิเคราะห์ข้อมูลทางสถิติ
4. ทำกิจกรรมเสริมประสบการณ์ในแผนกิจกรรมการเรียนประจำวิชาเทคโนโลยีสารสนเทศ 2 หน่วยที่ 7 ความรู้พื้นฐานในการวิเคราะห์ข้อมูล ตอนที่ 7.3 การวิเคราะห์ข้อมูลทางสถิติ