

# บทที่ 1

## สถิติเบื้องต้นกับคอมพิวเตอร์

### 1. ความหมายของสถิติ ข้อมูล และตัวแปร

#### 1.1 ความหมายของสถิติ (statistics)

สถิติ มีความหมายแยกออกเป็น 2 ประการ คือ

(1) ความหมายดั้งเดิมของสถิติคือ สถิติตัวเลข แผนภูมิ หรือกราฟ เกี่ยวกับการศึกษาเรื่อง ประชากร การเกษตรกรรม การเมืองและเศรษฐกิจ วิทยาศาสตร์และวิทยาศาสตร์การแพทย์ เป็นต้น ตัวอย่างเช่น สถิติการเกิด สถิติคนไม่มีงานทำ สถิติอุบัติเหตุในท้องถนน สถิติผู้ป่วยเป็น โรคมะเร็งปอด

(2) สถิติศาสตร์คือ คณิตศาสตร์ประยุกต์เป็นเครื่องมือของนักวิจัยที่ใช้ศึกษาข้อมูลที่สังเกตได้จากกลุ่มตัวอย่างหรือประชากร โดยอาศัยระเบียบวิธีทางสถิติ ซึ่งประกอบด้วย การเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล การตีความหมายข้อมูล และการนำเสนอข้อมูลในรูปตารางและกราฟ ตลอดจนการพัฒนาตัวแบบทางสถิติ (statistical model) เพื่อศึกษาสิ่งที่เกิดขึ้นในธรรมชาติ และใช้ตัวแบบทางสถิติในการทำนายค่าที่เป็นตัวเลข ซึ่งนำไปสู่การตัดสินใจและการประเมินค่าความเสี่ยงของเหตุการณ์ต่าง ๆ ได้ สถิติศาสตร์จึงเป็นวิทยาศาสตร์แขนงหนึ่งที่เกี่ยวข้องกับการใช้ข้อมูลจากกลุ่มตัวอย่างแล้วทำการสรุปอ้างอิงไปสู่ประชากร

#### 1.2 ความหมายของข้อมูล (data)

ข้อมูลคือ วัตถุวิสัยของสถิติศาสตร์ซึ่งได้จากการวัด (measurement) และการนับ (count) ตัวอย่างเช่น ข้อมูลน้ำหนัก ความสูง ความดันโลหิต อุณหภูมิ ความชื้น ได้จากการวัด ส่วนข้อมูลที่ได้จากการนับ ตัวอย่างเช่น จำนวนผู้เป็นโรคเอดส์ จำนวนมารดาที่สูบบุหรี่ระหว่างการตั้งครรภ์ จำนวนผู้สอบผ่านเกณฑ์

สเกลการวัดข้อมูลขึ้นกับลักษณะของข้อมูลหรือตัวแปร ดังนี้ คือ

(1) สเกลแบบแบ่งประเภท (nominal scale) เป็นข้อมูลที่ได้จากการนับตามคุณลักษณะหรือคุณสมบัติของสิ่งที่สนใจศึกษา ซึ่งอาจเป็น คน สัตว์ สิ่งของ หรือต้นไม้ก็ได้ เป็นคุณลักษณะที่ไม่สามารถวัดค่าได้ ทำได้เพียงแบ่งประเภทให้เท่านั้น อาจเรียกว่าข้อมูลเชิงคุณภาพ (qualitative data) หรือข้อมูลแบบจำแนกประเภท (categorical data) ตัวอย่างเช่น การนับถือศาสนา อาจแบ่งเป็น พุทธ คริสต์ อิสลาม อื่น ๆ อีกตัวอย่างคือ การเดินทางมาทำงาน อาจแบ่งเป็น รถยนต์ รถจักรยานยนต์ จักรยาน รถเมล์ เดิน เป็นต้น

(2) สเกลแบบอันดับ (ordinal scale) เป็นข้อมูลที่ได้จากการนับเช่นเดียวกับสเกลแบบแบ่งประเภท แต่แตกต่างกันที่คุณลักษณะหรือคุณสมบัติของสิ่งที่สนใจศึกษานั้น สามารถจัดอันดับความเข้มข้น ความมากน้อยของคุณลักษณะนั้น ๆ ได้โดยการจัดอันดับตามความเข้มข้นของคุณลักษณะนั้น ๆ แต่ไม่สามารถบอกได้ว่าแต่ละอันดับของคุณลักษณะนั้น ๆ ห่างกันเป็นปริมาณเท่าใด หรือระยะห่างของแต่ละอันดับเท่ากันหรือไม่ อาจเรียกว่าข้อมูลเชิงคุณภาพ หรือข้อมูลแบบจำแนกประเภท ตัวอย่างเช่น ทักษะคิดต่อการทำแกงอย่างถูกกฎหมาย อาจให้เลือกตอบว่า ไม่เห็นด้วยทุกกรณี เห็นด้วยบางกรณี เห็นด้วยทุกกรณี

(3) สเกลแบบช่วง (interval scale) เป็นข้อมูลที่ได้จากการวัดคุณลักษณะหรือคุณสมบัติของสิ่งที่สนใจศึกษาด้วยเครื่องมือวัดที่มีความห่างของหน่วยการวัดแต่ละหน่วยเท่า ๆ กัน อาจเรียกว่าข้อมูลเชิงปริมาณ (quantitative data) หรือข้อมูลแบบต่อเนื่อง (continuous data) ตัวอย่างเช่น การวัดอุณหภูมิ คะแนนสอบ ซึ่งการวัดแบบนี้ไม่มีจุดศูนย์ที่แท้จริง หมายความว่าค่าที่เป็น 0 เป็นเพียงศูนย์สมมติไม่ได้หมายถึงความไม่มีของคุณลักษณะนั้น ๆ

(4) สเกลแบบอัตราส่วน (ratio scale) เป็นข้อมูลที่ได้จากการวัดเช่นเดียวกับสเกลแบบช่วง แต่แตกต่างกันที่การวัดแบบนี้มีจุดศูนย์ที่แท้จริง ซึ่งแสดงถึงความไม่มีของคุณลักษณะนั้น ๆ อาจเรียกว่าข้อมูลเชิงปริมาณหรือข้อมูลแบบต่อเนื่อง ตัวอย่างเช่น การวัดน้ำหนัก ความสูง ปริมาณผลผลิต จำนวนบุตร

### 1.3 ความหมายของตัวแปร (variables)

ตัวแปร หมายถึง คุณลักษณะหรือคุณสมบัติของคน สัตว์ สิ่งของ หรือหน่วยตัวอย่างที่มีความแตกต่างกันระหว่างคน สัตว์ สิ่งของ หรือหน่วยตัวอย่างหนึ่ง ๆ ตัวอย่างเช่น เพศ อายุ อาชีพ น้ำหนัก ความสูง ความดันเลือด ความเชื่อ คะแนนการสอบ ระดับการศึกษา การเป็นโรค

ตัวแปรแบ่งออกได้เป็น 2 ประเภท คือ

(1) ตัวแปรเชิงปริมาณ (quantitative variables) ได้แก่ คุณลักษณะซึ่งสามารถวัดค่าได้ ตัวอย่างเช่น อายุ น้ำหนัก ความสูง ความดันเลือด คะแนนสอบ

(2) ตัวแปรเชิงคุณภาพ (qualitative variables) ได้แก่ คุณลักษณะซึ่งไม่สามารถวัดค่าได้ ทำได้เพียงแบ่งประเภท หรือจัดอันดับให้เท่านั้น ตัวอย่างเช่น เพศ อาชีพ ความเชื่อ ระดับการศึกษา การเป็นโรค ดังนั้นค่าของตัวแปรชนิดนี้จึงทำได้โดยการนับ

## 2. ความหมายของประชากรและตัวอย่าง

### 2.1 ประชากร (population)

ประชากร หมายถึง กลุ่มของคน สัตว์ สิ่งของ ต้นไม้ อาจเป็นสิ่งมีชีวิตหรือไม่มีชีวิตก็ได้ ที่มีขนาดใหญ่มากที่สุด และมีคุณลักษณะที่เราสนใจศึกษาในช่วงเวลาหนึ่ง ๆ หรือในขอบเขตหนึ่ง ๆ ประชากรอาจมีจำนวนนับได้หรือนับจำนวนไม่ได้ก็ได้ ตัวอย่างเช่น ประชากรต้นทุเรียนในสวนของเกษตรกรในจังหวัดระยอง ประชากรนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในกรุงเทพฯ ประชากรหมูในฟาร์มที่ถูกเลี้ยงด้วยอาหารผสมสารเร่งเนื้อแดงในประเทศไทย ประชากรผู้ที่อ่านหนังสือไม่ออกในประเทศไทย ประชากรชาวนาที่ยังใช้ควายไถนาในประเทศไทย ประชากรปลาตะเพียนในกลุ่มแม่น้ำเจ้าพระยา

### 2.2 ตัวอย่าง (sample)

ตัวอย่าง หมายถึง ส่วนหนึ่งของประชากรที่เราสนใจศึกษา สมมติว่าประชากรที่เราสนใจศึกษาคือ ความยาวของปลาตะเพียนในกลุ่มแม่น้ำเจ้าพระยา ในการศึกษาครั้งนี้ดำเนินการโดยจับปลาตะเพียนในกลุ่มแม่น้ำเจ้าพระยามาจำนวนหนึ่งเท่านั้นจากจุดต่าง ๆ บางจุดในกลุ่มแม่น้ำเจ้าพระยา แล้ววัดความยาวของปลาตะเพียน ปลาตะเพียนกลุ่มนี้เรียกว่าตัวอย่าง

### 3. ประเภทของสถิติ

อาจแบ่งได้เป็น 2 ประเภทใหญ่ ๆ คือ

#### 3.1 สถิติพรรณนา (descriptive statistics)

เป็นหลักการเกี่ยวกับการวิเคราะห์ข้อมูลที่เก็บรวบรวมได้ในรูปแบบของการบรรยาย โดยอาจนำเสนอข้อมูลในรูปตาราง การแจกแจงความถี่ กราฟรูปภาพต่าง ๆ หรืออาจเป็นการวัดคุณลักษณะของสิ่งที่สนใจโดยใช้การวัดแนวโน้มเข้าสู่ส่วนกลาง เช่น ค่าเฉลี่ย มัชยฐาน การวัดการกระจายของข้อมูล เช่น ส่วนเบี่ยงเบนมาตรฐาน พิสัย

#### 3.2 สถิติเชิงอนุมาน (inferential statistics)

มีวิธีการ 2 วิธีคือ

##### 3.2.1 สถิติอิงพารามิเตอร์ (parametric statistics)

ในสถิติเชิงอนุมานมีเนื้อหา 2 ส่วนใหญ่ ๆ คือ ส่วนแรกเป็นการศึกษาข้อมูลที่สังเกตได้จากกลุ่มตัวอย่าง โดยอาศัยระเบียบวิธีทางสถิติ เราเรียกค่าที่ประมวลได้จากข้อมูลตัวอย่างว่า ค่าสถิติ และใช้ค่าสถิติเป็นค่าประมาณของค่าของประชากรซึ่งมักจะไม่สามารถทราบค่า เรียกว่า ค่าพารามิเตอร์ ตัวอย่างการประมาณค่าพารามิเตอร์ของประชากรจากค่าสถิติ ได้แก่ ค่าเฉลี่ย ค่าสัดส่วน ค่าความแปรปรวน ในส่วนที่สองเป็นการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ของประชากรเพื่อสรุปอ้างอิงเกี่ยวกับพารามิเตอร์ เป็นการสรุปผลที่ได้จากกลุ่มตัวอย่างไปสู่ประชากรที่ต้องการศึกษา การใช้สถิติวิเคราะห์เพื่อสรุปผลจากตัวอย่างไปสู่ประชากรนี้เรียกว่า สถิติเชิงอนุมาน ซึ่งเป็นสถิติวิเคราะห์ที่เกี่ยวข้องกับการแจกแจงความน่าจะเป็นของข้อมูลตัวอย่าง

การแจกแจงความน่าจะเป็นที่รู้จักกันดี ได้แก่ การแจกแจงแบบปกติซึ่งมีพารามิเตอร์คือ  $\mu$  และ  $\sigma^2$  แทนค่าเฉลี่ยและความแปรปรวน เขียนเป็นสัญลักษณ์ได้คือ  $N(\mu, \sigma^2)$  แทนการแจกแจงแบบปกติที่มีพารามิเตอร์ 2 ตัวนี้ ถ้าข้อมูลของกลุ่มตัวอย่างมาจากการแจกแจงแบบปกติ ถ้าต้องการสรุปอ้างอิงเกี่ยวกับพารามิเตอร์ที่ไม่ทราบค่านี้ เราจะใช้ค่าสถิติ ได้แก่ ค่าเฉลี่ยของตัวอย่าง เป็นค่าประมาณของพารามิเตอร์  $\mu$  ซึ่งเป็นค่าเฉลี่ยของประชากร และใช้การทดสอบ t-test ในการทดสอบสมมติฐานเกี่ยวกับ  $\mu$  โดยการใช้ข้อมูลจากกลุ่มตัวอย่างเป็นหลักฐานในการสนับสนุนการสรุปอ้างอิงเกี่ยวกับพารามิเตอร์ที่ไม่ทราบค่านี้ เรียกว่า การสรุปอ้างอิงเกี่ยวกับพารามิเตอร์ (parametric inferences) ตัวอย่างสถิติ เช่น ถ้าต้องการทดสอบค่าเฉลี่ย

ของประชากรไม่เกิน 2 กลุ่ม สถิติที่ใช้ทดสอบคือ t-test ถ้าต้องการทดสอบค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่ม สถิติที่ใช้ทดสอบคือ F-test

สิ่งที่ต้องพิจารณาในการเลือกใช้สถิติสำหรับการวิจัย คือ ลักษณะของข้อมูลเป็นไปตามข้อกำหนดหรือข้อตกลงเบื้องต้น (assumption) ของการใช้สถิตินั้น ๆ หรือไม่ เช่น ข้อตกลงเบื้องต้นของการใช้สถิติ t-test

(1) ข้อมูลหรือตัวแปรที่ได้จากการวัดเป็นข้อมูลแบบต่อเนื่อง มีสเกลการวัดเป็นแบบช่วงหรือแบบอัตราส่วน

(2) ข้อมูลที่เป็นอิสระกันในแต่ละกลุ่มมีการแจกแจงแบบปกติ

ข้อตกลงเบื้องต้นของการใช้สถิติ F-test ในการวิเคราะห์ความแปรปรวน

(1) ข้อมูลหรือตัวแปรที่ได้จากการวัดเป็นข้อมูลแบบต่อเนื่องมีระดับการวัดเป็นแบบช่วงหรือแบบอัตราส่วน

(2) ความเป็นอิสระของหน่วยตัวอย่างในกลุ่มเดียวกันและระหว่างกลุ่ม

(3) ประชากรแต่ละกลุ่มมีการแจกแจงแบบปกติ

(4) ความแปรปรวนของประชากรแต่ละกลุ่มเท่ากัน

### 3.2.2 สถิติไม่อิงพารามิเตอร์ (nonparametric statistics)

ในสถิติอิงพารามิเตอร์เราจำเป็นต้องทราบเกี่ยวกับการแจกแจงความน่าจะเป็นของข้อมูล การสรุปอ้างอิงเกี่ยวกับพารามิเตอร์บางครั้งอาจไม่เหมาะสมหรือทำไม่ได้ เนื่องจากการที่จะบอกว่ากลุ่มตัวอย่างมาจากการแจกแจงแบบใดแบบหนึ่งอาจไม่มีเหตุผลเพียงพอ ในทางปฏิบัติเราสามารถบอกได้ยากกว่ากลุ่มตัวอย่างของเรามาจากการแจกแจงแบบใด ในกรณีเช่นนี้เหมาะสมที่จะใช้การสรุปอ้างอิงแบบไม่อิงพารามิเตอร์

ในกรณีที่เราไม่ทราบการแจกแจงความน่าจะเป็นของข้อมูล หรือขนาดตัวอย่างเล็ก หรือตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงไม่ใช่แบบปกติ ลักษณะข้อมูลเป็นข้อมูลเชิงคุณภาพที่มีสเกลการวัดแบบแบ่งประเภทหรือแบบอันดับ ซึ่งไม่เป็นไปตามข้อตกลงเบื้องต้นของสถิติเชิงอนุมานเกี่ยวกับพารามิเตอร์ ดังนั้นจึงอาจจำเป็นต้องใช้วิธีการของสถิติที่ไม่อิงพารามิเตอร์ ซึ่งได้แก่ Chi - Square test , Binomial test , Runs test , Kolmogorov – Smirnov test , Wilcoxon test , Sign test เป็นต้น

#### 4. บทบาทของคอมพิวเตอร์ในการวิเคราะห์ทางสถิติ

ในระเบียบวิธีทางสถิติซึ่งประกอบด้วย การเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล การตีความหมายข้อมูล และการนำเสนอข้อมูลนั้น มีวัตถุประสงค์เป็นตัวเลขจำนวนมากมาย ในขั้นตอนของการวิเคราะห์ข้อมูลก็มีการคำนวณด้วยสูตรต่าง ๆ ที่มีความยุ่งยากซับซ้อน ซึ่งการคำนวณด้วยมือหรือเครื่องคิดเลขธรรมดาทำได้ยาก และอาจมีความผิดพลาดจากการคำนวณได้สูง ทำให้อาจต้องมีการคำนวณซ้ำเพื่อตรวจสอบความถูกต้องของผลการคำนวณ แต่เมื่อใช้คอมพิวเตอร์ช่วยในการคำนวณในขั้นตอนต่าง ๆ ของระเบียบวิธีทางสถิติที่มีความยุ่งยากซับซ้อนของสูตรทางคณิตศาสตร์นั้นกลายเป็นเรื่องง่าย ใช้เวลาน้อย และมีผลการคำนวณที่ถูกต้องเชื่อถือได้ ทำให้เรามีเวลามากขึ้นในการตรวจสอบความถูกต้องของข้อมูลดิบก่อนการวิเคราะห์ทางสถิติและมีเวลามากขึ้นสำหรับการพิจารณาความหมายของผลการวิเคราะห์

โปรแกรมสำเร็จรูปทางสถิติขนาดใหญ่ที่รู้จักกันแพร่หลายได้แก่ โปรแกรม SPSS (Statistical Package for the Social Sciences) โปรแกรม SAS (Statistical Analysis System) โปรแกรม MINITAB เป็นต้น ในหนังสือเล่มนี้จะอธิบายการวิเคราะห์ทางสถิติโดยใช้โปรแกรม SPSS for Windows version 12 ที่ใช้กับเครื่องไมโครคอมพิวเตอร์ที่มีระบบปฏิบัติการ Windows สามารถรับข้อมูลที่สร้างจากโปรแกรมประเภทต่าง ๆ ได้เช่น Excel Microsoft Word เป็นต้น เนื่องจากเป็นโปรแกรมที่นิยมใช้กันมากมีแพร่หลายหาได้ง่าย ราคาไม่แพง และวิธีการใช้โปรแกรมสะดวกไม่มีความยุ่งยากซับซ้อน ใช้ง่าย

ในขั้นตอนการเก็บรวบรวมข้อมูลจะประกอบด้วย การเก็บข้อมูล (data collection) และการรวบรวมข้อมูล (data compilation) วิธีการเก็บข้อมูลก็ทำได้หลายวิธี อาจแบ่งได้เป็น 2 วิธีใหญ่ ๆ คือ

(1) การเก็บข้อมูลทุติยภูมิ (collection of secondary data) เป็นการเก็บข้อมูลจากข้อมูลที่ผู้อื่นได้ทำการเก็บรวบรวมไว้แล้ว ผู้วิจัยเพียงทำการรวบรวมข้อมูลเหล่านั้นมาตามวัตถุประสงค์ที่ต้องการศึกษา

(2) การเก็บข้อมูลปฐมภูมิ (collection of primary data) เป็นการเก็บข้อมูลจากกลุ่มตัวอย่างหรือกลุ่มประชากรที่มีคุณลักษณะที่ผู้วิจัยสนใจศึกษา โดยใช้เครื่องมือในการเก็บรวบรวมตัวอย่างเช่น แบบสอบถาม (questionnaire) แบบสัมภาษณ์ (interview schedual) แบบทดสอบ (Test) เป็นต้น หรืออาจใช้วิธีการสังเกตโดยตรง (direct observation) โดยสังเกตจาก

ปรากฏการณ์ที่เกิดขึ้น จะเห็นได้ว่าข้อมูลอาจได้มาจากการวัด (measurements) ด้วยเครื่องมือ หรือจากการสังเกต (observations) ก็ได้

เมื่อได้ข้อมูลมาแล้ว ต้องดำเนินการตรวจสอบความถูกต้อง ความสมบูรณ์ ความครบถ้วน และข้อบกพร่องต่าง ๆ ที่อาจเกิดขึ้นในขั้นตอนการเก็บรวบรวมข้อมูล

เมื่อตรวจสอบข้อมูลที่ได้จากการเก็บรวบรวมข้อมูลมาดีแล้ว ก่อนจะนำไปวิเคราะห์ประมวลผลด้วยคอมพิวเตอร์ ต้องสร้างคู่มือลงรหัส ขั้นตอนต่อไปคือ การลงรหัสข้อมูลจากแบบบันทึกข้อมูลให้เป็นตัวเลข ตามลักษณะของข้อมูลหรือตัวแปร ตัวอย่างเช่น การศึกษาความคิดเห็นของอาจารย์ที่มีต่อการให้บริการสื่อการเรียนการสอนในศูนย์เรียนรวม

ตัวอย่างการสร้างคู่มือลงรหัส

ลำดับ ที่	ชื่อตัวแปร	ความหมายของตัวแปร	ค่าของตัวแปรและความหมาย	คอลัมน์ ที่
1	ID	เลขที่ของตัวอย่าง	001 – 500	1 – 3
2	SEX	เพศ	1 = ชาย, 2 = หญิง	4
3	DEP	คณะ	1 = คณะเกษตร 2 = คณะวิศวกรรมศาสตร์ 3 = คณะศึกษาศาสตร์	5
4	EQUIP	เครื่องมือสื่อการเรียน	1 = เครื่องฉายโปรเจกเตอร์ 2 = เครื่องฉายแผ่นใส 3 = เครื่องฉายภาพทึบแสง 4 = เครื่องฉายสไลด์	6
5	ITEM1	ความคิดเห็นเกี่ยวกับคุณภาพของเครื่องฉายโปรเจกเตอร์	1 = ใช้งานไม่ได้เลย 2 = ใช้งานได้บางครั้ง 3 = ส่วนใหญ่ใช้งานได้ดี 4 = ใช้งานได้ดีมาก	7
6	ITEM2	ความคิดเห็นเกี่ยวกับคุณภาพของเครื่องฉายภาพทึบแสง	1 = ใช้งานไม่ได้เลย 2 = ใช้งานได้บางครั้ง 3 = ส่วนใหญ่ใช้งานได้ดี 4 = ใช้งานได้ดีมาก	8

การลงรหัสข้อมูลเป็นขั้นตอนการนำข้อมูลที่เก็บรวบรวมมาได้นำมาแปลงเป็นตัวเลขตามคู่มือลงรหัส ทำทีละหน่วยตัวอย่างจนครบทั้งกลุ่มตัวอย่าง ขั้นตอนต่อไปคือ การบันทึกข้อมูลลงในแผ่นดิสก์ สามารถใช้โปรแกรม Excel ช่วยในการบันทึกข้อมูลได้ ก็จะได้เพิ่มข้อมูลที่สามารถนำไปใช้วิเคราะห์ด้วยโปรแกรมสำเร็จรูปทางสถิติ เช่น โปรแกรม SPSS ได้ สำหรับรายละเอียดเกี่ยวกับการสร้างเพิ่มข้อมูล การตรวจสอบข้อมูล และการจัดการเพิ่มข้อมูลจะอธิบายในบทต่อไป