

บทที่ 1 ความรู้เบื้องต้นทางสถิติ

1.1 ความหมายของสถิติ

คำว่าสถิติมีความหมาย 2 ประการ ดังนี้ 1) สถิติหมายถึงข้อเท็จจริงที่เป็นตัวเลขหรือลายลักษณ์อักษรที่เก็บรวบรวมมาได้ เช่น สถิติการเกิด สถิติผู้ป่วยโรคต่างๆ และสถิติการเกิดอุบัติเหตุบนท้องถนน เป็นต้น 2) สถิติคือ ศาสตร์แขนงหนึ่ง หรือที่เรียกว่า สถิติศาสตร์ (Statistic) หรือระเบียบวิธีการทางสถิติ (Statistical method) ประกอบด้วย การเก็บรวบรวมข้อมูล (Data collection) การนำเสนอข้อมูล (Data representation) การวิเคราะห์ข้อมูล (Data analysis) และการตีความข้อมูล (Data interpretation)

จากความหมายดังกล่าวนี้ทำให้องค์กรต่างๆ ในปัจจุบันนำสถิติไปช่วยในการตัดสินใจ เช่น บริษัทผลิตโทรศัพท์มือถือต้องการผลิตโทรศัพท์รุ่นใหม่ที่มีฟังก์ชันการใช้งานครบครันตรงตามความต้องการของผู้บริโภค จึงต้องการเก็บรวบรวมข้อมูลจากกลุ่มผู้บริโภคแล้วนำมาวิเคราะห์ว่าฟังก์ชันใดที่ผู้บริโภคต้องการให้มีในโทรศัพท์มือถือ จากนั้นก็จะผลิตให้ตรงตามความต้องการของผู้บริโภค เป็นต้นนอกจากจะทำให้ได้มีการพัฒนาคุณภาพของสินค้าแล้วยังอาจส่งผลให้ยอดขายบริษัทดังกล่าวเพิ่มขึ้นอีกด้วย

1.2 ประชากรและตัวอย่าง (Population and Sample)

ประชากร (Population) หมายถึงหน่วยทั้งหมดในเรื่องที่ศึกษาอาจเป็น คนสัตว์ และสิ่งของ เช่น ถ้าต้องการศึกษาความคิดเห็นของนิสิตมหาวิทยาลัยบูรพาที่มีต่อการบริหารประเทศของรัฐบาล ประชากรคือนิสิตมหาวิทยาลัยบูรพาทุกคน ซึ่งถ้าเก็บข้อมูลจากประชากรทุกหน่วยจะเรียกว่าการทำสำมะโน (Census) แต่ในทางปฏิบัติผู้วิจัยไม่สามารถทำการเก็บข้อมูลความคิดเห็นของนิสิตมหาวิทยาลัยบูรพากับทุกคนได้ จึงต้องเลือกนิสิตเพียงบางคนมาสอบถาม โดยที่นิสิตที่ถูกเลือกมานั้นเรียกว่าตัวอย่างและถือว่าตัวอย่างเป็นตัวแทนของประชากร ดังนั้น **ตัวอย่าง (Sample)** จึงหมายถึงบางส่วนของประชากรหรือเซตย่อยของประชากร และตัวอย่างที่ดีจะต้องมีลักษณะเหมือนประชากรมากที่สุด ถ้ากลุ่มตัวอย่างมีลักษณะแตกต่างจากประชากรแล้วจะทำให้ข้อมูลที่ได้นั้นคลาดเคลื่อนไปจากความจริง นอกจากนี้ยังมีค่าที่ต้องรู้จัก 2 ค่าคือ ค่าพารามิเตอร์ และค่าสถิติ

ค่าพารามิเตอร์ (Parameters) หมายถึง ค่าที่บ่งบอกลักษณะที่สำคัญของประชากร

ค่าสถิติ (Statistic) หมายถึง ค่าที่บ่งบอกลักษณะของตัวอย่าง

1.3 ประเภทของสถิติ

ทางด้านสถิติศาสตร์แบ่งประเภทของสถิติ ออกได้เป็น 2 ประเภท คือ

1. สถิติเชิงพรรณนา (Descriptive Statistics)

สถิติเชิงพรรณนา หมายถึง สถิติที่ใช้ในการบรรยายถึงข้อเท็จจริงของข้อมูลที่เรารวบรวมมาได้ ได้แก่ การวิเคราะห์ข้อมูลหรือการนำเสนอข้อมูลในรูปแบบของตาราง แผนภูมิ แผนภาพ หรือคำอธิบายการจัดเก็บข้อมูลต่างๆ เพื่อแสดงความหมายในเชิงจำนวนหรือปริมาณของสิ่งต่างๆ เช่น เพศ ความสูง อายุ น้ำหนัก รายได้ เป็นต้น

หรืออาจแสดงความหมายในเชิงคุณภาพ เช่น เจตคติต่อวิชาชีพ การนับถือศาสนา เป็นต้น โดยไม่ต้องอาศัยทฤษฎีความน่าจะเป็นมาช่วยในการวิเคราะห์

2. สถิติเชิงอนุมาน (Inferential Statistics)

สถิติเชิงอนุมาน หมายถึง สถิติที่เป็นการนำค่าที่ได้จากตัวอย่างซึ่งถือเป็นค่าสถิติไปอธิบายคุณลักษณะประชากรซึ่งเป็นค่าพารามิเตอร์ โดยจุดมุ่งหมายของสถิติอนุมานคือการศึกษาคูสมบัติของประชากรโดยใช้ความรู้ที่ได้จากการศึกษากลุ่มตัวอย่างที่สุ่มมาจากประชากรกลุ่มเดียวกัน โดยอาศัยทฤษฎีความน่าจะเป็น เช่น การประมาณค่า และการทดสอบสมมติฐาน เป็นต้น

1.4 ประเภทของข้อมูล

ข้อมูลในทางสถิติแบ่งเป็น 2 ประเภท

1. ข้อมูลเชิงคุณภาพ (Qualitative data) หมายถึง ข้อมูลที่รวบรวมมาโดยพิจารณาจากคุณสมบัติของข้อมูลเป็นหลัก เช่น เพศ (ชาย และหญิง) สถานภาพของการสมรส (โสด แต่งงาน หม้าย และหย่า) ระดับการศึกษา (ปริญญาตรี ปริญญาโท และปริญญาเอก) ระดับรายได้ของประชากร (สูง ปานกลาง และต่ำ) เป็นต้น

2. ข้อมูลเชิงปริมาณ (Quantitative data) หมายถึง ข้อมูลที่รวบรวมมาได้ในรูปของตัวเลขที่สามารถนำมาคำนวณได้ เช่น น้ำหนัก อายุ ส่วนสูง อุณหภูมิ

ข้อมูลบางชนิดอาจจะเก็บรวบรวมได้หลายแบบแต่เราจะรวบรวมแบบใดขึ้นอยู่กับวัตถุประสงค์ที่จะนำไปวิเคราะห์และใช้ประโยชน์

1.5 ระดับของการวัด (Level of measurement)

ระดับของการวัดเป็นการกำหนดจำนวนตัวเลขให้กับวัตถุหรือเหตุการณ์เรื่องราวต่างๆ ดังนั้นลักษณะอย่างหนึ่งที่ชี้ให้เห็นว่าควรจะใช้วิธีการทางสถิติแบบใด ในการวิเคราะห์ข้อมูลก็คือ ระดับของการวัด ซึ่งเป็นการบอกลักษณะข้อมูลที่ใช้ในการวิจัย โดยระดับของการวัดแบ่งออกได้เป็น 4 ระดับ แต่ละระดับใช้มาตรการวัดที่แตกต่างกันดังนี้

1. ระดับนามบัญญัติ (Nominal level) คือ ข้อมูลเชิงคุณภาพเท่านั้น มักจะกำหนดตัวเลขแทนกลุ่มย่อยแต่ละกลุ่ม ตัวเลขที่กำหนดขึ้นเป็นเพียงตัวเลขที่สมมติขึ้นตามกลุ่มเท่านั้นไม่สามารถนำมาคำนวณหรือเปรียบเทียบทางคณิตศาสตร์ได้ เช่น เพศ (0 แทนกลุ่มเพศชาย และ 1 แทนกลุ่มเพศหญิง) ข้อมูลระดับนี้ไม่มีคุณลักษณะทางด้านการคำนวณทางคณิตศาสตร์จึงสามารถใช้สถิติได้เพียง จำนวน ความถี่ ร้อยละ ฐานนิยม หรือใช้สถิติแบบนอนพาราเมตริก (non-parametric) เท่านั้น

2. ระดับอันดับ (Ordinal level) คือ ข้อมูลเชิงคุณภาพเช่นเดียวกับระดับนามบัญญัติ แต่จะสามารถแสดงความแตกต่างระหว่างกลุ่มได้ กล่าวคือ เป็นการจัดเรียงอันดับสิ่งของพวกเดียวกันให้มีความลดหลั่นกันไปเป็นขั้นๆโดยไม่ทราบว่าจะแต่ละขั้นต่างกันอยู่เท่าใด เช่น ไม่สามารถบอกได้ว่าหนังสือเล่มที่ 1 ดีกว่าหนังสือเล่มที่ 2 เท่าไร ดังนั้นจะใช้กับสถิติในความหมายที่มากกว่าหรือน้อยกว่า แต่ไม่ได้แสดงถึงรายละเอียดของคุณภาพได้ ตัวอย่าง

ข้อมูลในระดับนี้ คือ ระดับการศึกษา (ให้ 1 แทนปริญญาตรี 2 แทนปริญญาโท และ 3 แทนปริญญาเอก) คนที่อยู่ในในกลุ่มจบปริญญาตรีมีการศึกษาน้อยกว่ากลุ่มคนที่จบปริญญาโทและจบปริญญาเอก สำหรับสถิติที่ใช้ในการวัดในระดับนี้ ได้แก่ มัธยมศึกษา เปรอร์เซ็นต์ไทล์ พิสัย การจัดอันดับต่างๆ และสถิติแบบนอนพารามตริก

3. ระดับอันตรภาค (Interval level) คือ ข้อมูลที่มีรายละเอียดมากกว่าระดับอันดับ สามารถวัดความแตกต่างได้แน่นอน สามารถระบุได้ว่ามากกว่าหรือน้อยกว่าเท่าใดและสามารถระบุระยะห่างด้วยช่วงที่เท่าๆกันได้ เช่น อุณหภูมิ 30 °C มากกว่าอุณหภูมิ 10 °C อยู่ 20 °C แต่ข้อมูลระดับอันตรภาคจะไม่มี 0 ที่แท้จริง ศูนย์เป็นเพียงสิ่งที่สมมติขึ้นเท่านั้น เช่น อุณหภูมิ 0 °C น้ำจะกลายเป็นน้ำแข็ง แต่ไม่ได้หมายความว่าไม่มีความร้อนอยู่เลย สถิติที่ใช้กับการวัดระดับนี้ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน การหาค่าสหสัมพันธ์ของเพียร์สัน และการทดสอบสมมติฐาน เป็นต้น

4. ระดับอัตราส่วน (Ratio level) คือ ข้อมูลเชิงปริมาณที่มีรายละเอียดมากที่สุด สามารถระบุขนาดและเปรียบเทียบบอกความแตกต่างได้ เช่น น้ำหนัก ส่วนสูง ปริมาณน้ำฝนที่ตกใน 1 วัน เป็นต้น ข้อมูลระดับนี้มี 0 ที่แท้จริง เช่น ส่วนสูง 0 เซนติเมตร หมายถึงไม่มีความสูงเลย ข้อมูลในระดับนี้สามารถนำมาวิเคราะห์ทางสถิติได้ทุกวิธี

1.6 ขั้นตอนของการดำเนินงานทางสถิติ

ขั้นตอนของการดำเนินการทางสถิติจะต้องดำเนินไปตามขั้นตอนตามระเบียบวิธีการทางสถิติซึ่งแบ่งเป็นขั้นตอนใหญ่ๆ ได้ 4 ขั้นตอนนี้

1. การเก็บรวบรวมข้อมูล (Data collection) ขั้นตอนนี้รวมถึงการวางแผนเตรียมการตลอดไปจนถึงการเก็บรวบรวมข้อมูล ซึ่งต้องอาศัยความรู้ทางวิชาการพอสมควรเพื่อศึกษาภูมิหลังของข้อมูล ความต้องการของข้อมูลในแง่ของการนำไปใช้ประโยชน์ โดยวิธีการเก็บรวบรวมข้อมูลอาจทำได้ 2 วิธี

- 1) การเก็บรวบรวมจากเอกสารต่างๆ ที่ได้มีการบันทึกไว้แล้วเรียกว่า แหล่งทุติยภูมิ (Secondary source)
- 2) การเก็บรวบรวมข้อมูลด้วยตนเองเรียกว่า แหล่งปฐมภูมิ (Primary source)

2. การนำเสนอข้อมูล (Data representation) เป็นการนำข้อมูลที่เรารวบรวมได้มาเสนอหรือแสดงให้ผู้อื่นทราบ ได้เข้าใจและเกิดภาพพจน์เกี่ยวกับข้อมูลที่ถูกต้อง การนำเสนอข้อมูลอาจทำได้โดยใช้บทความ ตาราง กราฟ และแผนภูมิต่างๆ

3. การวิเคราะห์ข้อมูล (Data analysis) โดยการนำข้อมูลที่รวบรวมได้มาทำการวิเคราะห์ทางสถิติเพื่อศึกษาหาข้อสรุปตามที่ต้องการ

4. การตีความ (Data interpretation) โดยทำการสรุปผลของการวิเคราะห์ข้อมูลว่าผลจากการวิเคราะห์ทางสถิติที่ได้ออกมาเป็นค่าต่างๆ นั้นหมายความว่าอย่างไร เพื่อนำไปใช้ในการตัดสินใจ

1.7 วิธีการนำเสนอข้อมูลสถิติ

เมื่อทำการเก็บรวบรวมข้อมูลมาแล้วขั้นต่อไปคือการนำเสนอข้อมูลให้ผู้อื่นได้ทราบการนำเสนอข้อมูลทำได้โดยวิธีการนำเสนอด้วยบทความ การนำเสนอข้อมูลด้วยตาราง การนำเสนอข้อมูลด้วยบทความกึ่งตาราง และการนำเสนอข้อมูลด้วยแผนภูมิและกราฟ ดังต่อไปนี้

1. นำเสนอข้อมูลด้วยบทความ (Text presentation) มีลักษณะเป็นข้อความสั้นๆอ่านเข้าใจง่ายพบมากในหนังสือพิมพ์วารสารและรายงานต่างๆ เช่น

จำนวนครั้งที่ผู้ป่วยนอกมารับการรักษาที่โรงพยาบาลนครเชียงใหม่ในช่วง 5 ปีเพิ่มขึ้นจาก 194,660 ครั้งใน พ.ศ. 2519 เป็น 237,810 ครั้งใน พ.ศ. 2524 หรือเพิ่มขึ้นร้อยละ 22.17

2. การนำเสนอข้อมูลด้วยตาราง (Tabular presentation) ในกรณีที่มีข้อมูลซ้ำๆ กันเป็นจำนวนมาก การนำเสนอในรูปแบบความย่อไม่เหมาะสมเนื่องจากเข้าใจยากและเปรียบเทียบตัวเลขได้ไม่ชัดเจน จึงควรใช้การนำเสนอในรูปแบบตารางแจกแจงความถี่แทนการใช้บทความโดยระบุชื่อเรื่องแหล่งที่มาของข้อมูลให้ชัดเจน ดังตารางที่ 1.1 และ 1.2

ตารางที่ 1.1 ความแข็งแรงของกิ่งกุหลาบจากธาตุเพราะชำ

ธาตุเพราะชำ	ความแข็งแรงของกิ่งกุหลาบ		รวม
	แข็งแรง	ไม่แข็งแรง	
ธาตุที่ 1	17	8	25
ธาตุที่ 2	13	12	25
รวม	30	20	50

ที่มา: บริษัทกุหลาบจำกัด

ตารางที่ 1.2 ผลผลิตป่าไม้ที่สำคัญของประเทศไทยระหว่างพ. ศ. 2523 ถึง 2525 (พันเมตริกตัน)

ผลผลิตจากป่าไม้	2523	2524	2525
ไม้ยาง	627.4	551.1	289.4
ไม้สัก	179.6	97.3	73.2
ไม้อื่นๆ	2293.7	1895.8	1435.9
ฟืน	825.2	635.7	673.0

ที่มา: กรมป่าไม้

3. การนำเสนอข้อมูลด้วยบทความกึ่งตาราง (Semi - tabular presentation) มีลักษณะคล้ายการนำเสนอด้วยบทความ แต่มีการสร้างตารางนำเสนอตัวเลขขึ้นในบทความเพื่อให้เห็นชัดเจน และสามารถ

เปรียบเทียบได้ รูปแบบการนำเสนอข้อมูลนี้ใช้ในกรณีที่ข้อมูลมีไม่มากพอที่จะทำเป็นตาราง แต่ก็มากเกินไปที่จะแสดงในรูปบทความ หรือต้องการขยายความให้ชัดเจนขึ้นด้วยการบรรยายประกอบดังนี้

จำนวนผู้ป่วยนอกของโรงพยาบาลนครเชียงใหม่มีแนวโน้มเพิ่มขึ้นทุกปี และอัตราการเพิ่มของผู้ป่วยในแต่ละช่วง 5 ปี นับตั้งแต่ พ.ศ. 2524 เป็นต้นมา อยู่ในอัตราค่อนข้างสูง ดังตารางต่อไปนี้

ตารางที่ 1.3 การเปลี่ยนแปลงจำนวนผู้ป่วยนอกในแต่ละช่วง 5 ปี

พ. ศ.	จำนวนครั้ง	ร้อยละการเปลี่ยนแปลงในช่วง 5 ปี
2514	113310	
2519	194600	+71.79
2524	237810	+22.17

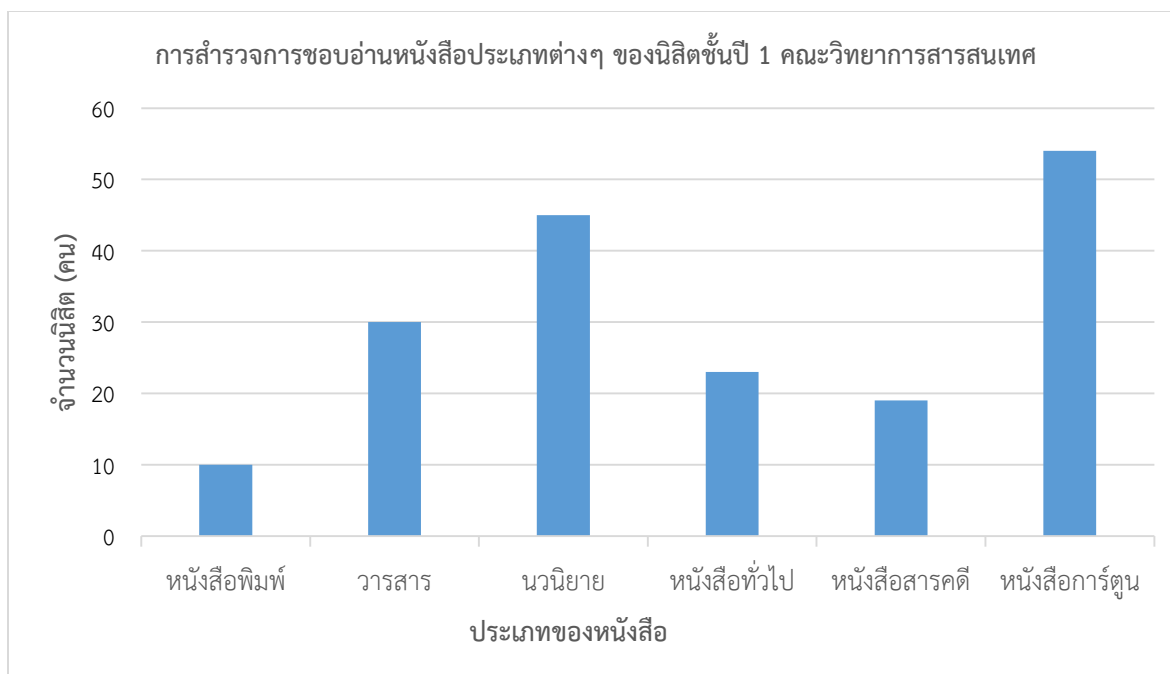
4. การนำเสนอข้อมูลด้วยแผนภูมิและกราฟ (Graphic presentation) รูปแบบของการนำเสนอข้อมูลสถิติอีกวิธีหนึ่งที่น่าสนใจ คือ การนำเสนอโดยใช้แผนภูมิและกราฟจุดประสงค์ของการนำเสนอในรูปแบบนี้นอกจากจะทำให้ทราบรูปแบบหรือแนวโน้มของข้อมูลทั้งหมดได้ชัดเจนมากขึ้นแล้วยังสามารถเปรียบเทียบกลุ่มข้อมูลได้ชัดเจนขึ้นอีกครั้งยังทำให้ผู้อ่านเข้าใจได้ง่ายรวดเร็วและสามารถดึงดูดความสนใจได้ดีกว่านำเสนอข้อมูลด้วยวิธีการอื่นๆ

แผนภูมิแท่ง

กราฟแท่งหรือแผนภูมิแท่งมีลักษณะเป็นแท่งรูปสี่เหลี่ยมมีช่องไฟระหว่างแท่งจะวางอยู่ในแนวดิ่งหรือแนวนอนก็ได้ ความยาวของแท่งขึ้นอยู่กับจำนวนข้อมูล ส่วนความกว้างของแท่งคงที่ แผนภูมิแท่งเหมาะสำหรับใช้เปรียบเทียบรายการที่แตกต่างกันหลายรายการเช่นการเปรียบเทียบข้อมูลตามเวลาการเปรียบเทียบข้อมูลตามคุณลักษณะเป็นต้นการนำเสนอข้อมูลด้วยแผนภูมิแท่งเป็นวิธีการนำเสนอที่น่าสนใจสำหรับผู้อ่านทุกระดับเพราะเป็นการที่ผู้อ่านทำความเข้าใจด้วยตนเองได้ง่าย จากตารางที่ 1.4 สามารถเขียนเป็นแผนภูมิแท่งได้ดังรูปที่ 1.1 จากตารางที่ 1.5 สามารถเขียนเป็นแผนภูมิแท่งได้ดังรูปที่ 1.2

ตารางที่ 1.4 การสำรวจการชอบอ่านหนังสือประเภทต่างๆ ของนิสิตชั้นปี 1 คณะวิทยาการสารสนเทศ

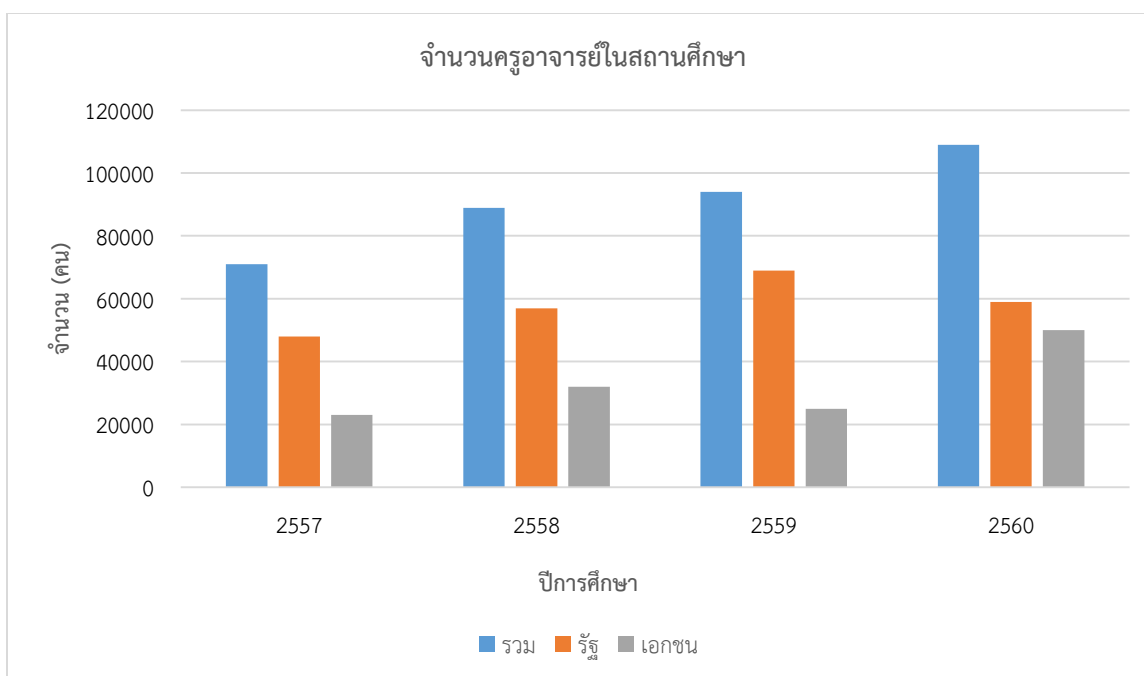
ประเภทของหนังสือ	จำนวนนิสิต (คน)
หนังสือพิมพ์	10
วารสาร	30
นวนิยาย	45
หนังสือทั่วไป	23
หนังสือสารคดี	19
หนังสือการ์ตูน	54



รูปที่ 1.1 แผนภูมิแท่งแสดงการชอบอ่านหนังสือประเภทต่างๆ ของนิสิตชั้นปี 1 คณะวิทยาการสารสนเทศ

ตารางที่ 1.5 จำนวนครุอาจารย์ในสถานศึกษา

	2557	2558	2559	2560
รวม	71,000	88,900	93,950	109,000
รัฐ	48,000	56,900	68,950	59,000
เอกชน	23,000	32,000	25,000	50,000



รูปที่ 1.1 แผนภูมิแท่งแสดงจำนวนครูอาจารย์ในสถานศึกษา

แผนภูมิวงกลม

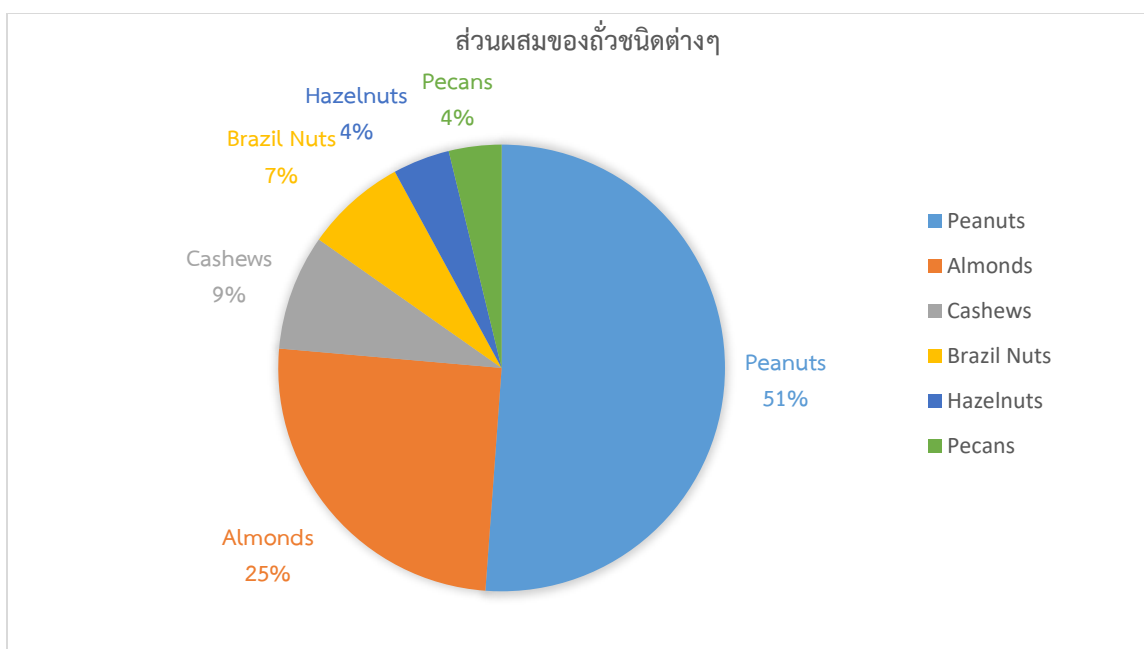
กราฟวงกลมหรือแผนภูมิวงกลมเป็นการนำเสนอข้อมูลทางสถิติเป็นรูปวงกลมด้วยการแบ่งพื้นที่ภายในวงกลมเป็นส่วนๆ เพื่อเปรียบเทียบสถิติแต่ละส่วนว่ามีมากน้อยต่างกันอย่างไร กราฟวงกลม หรือแผนภูมิวงกลม หรือแผนภาพกึ่ง จะใช้ในการแสดงสัดส่วนหรือเปอร์เซ็นต์ของข้อมูล

ในการสร้างแผนภูมิวงกลม พื้นที่ทั้งหมดของวงกลมเป็นร้อยละ มุมรอบจุดศูนย์กลางของวงกลมเท่ากับ 360 องศาหรือ 100 ส่วน เพราะฉะนั้น 1 ส่วนเท่ากับ 3.60 องศา และทำสถิติที่ต้องการนำมาเปรียบเทียบแต่ละส่วนให้เป็นร้อยละของทั้งหมด แบ่งพื้นที่ในวงกลมตามค่าร้อยละที่ได้ โดยเรียงจากมากไปหาน้อย จากนั้นแรงเงาหรือระบายสีพื้นที่ในแต่ละส่วนให้แตกต่างกัน เพื่อให้เปรียบเทียบกันได้ชัดเจนยิ่งขึ้นอาจจะเขียนตัวเลขแสดงค่าร้อยละไว้ในพื้นที่แต่ละส่วนหรือเขียนตอนล่างของกราฟด้วยก็ได้

จากตารางที่ 1.6 ส่วนผสมของถั่วชนิดต่างๆ ในหน่วยกรัม สามารถแสดงแผนภูมิวงกลมได้ดังรูปที่ 1.2

ตารางที่ 1.6 แสดงส่วนผสมของถั่วชนิดต่างๆ (กรัม)

Peanuts	Almonds	Cashews	Brazil Nuts	Hazelnuts	Pecans
167.7g	82.7g	27.4g	24.0g	13.5g	12.5g



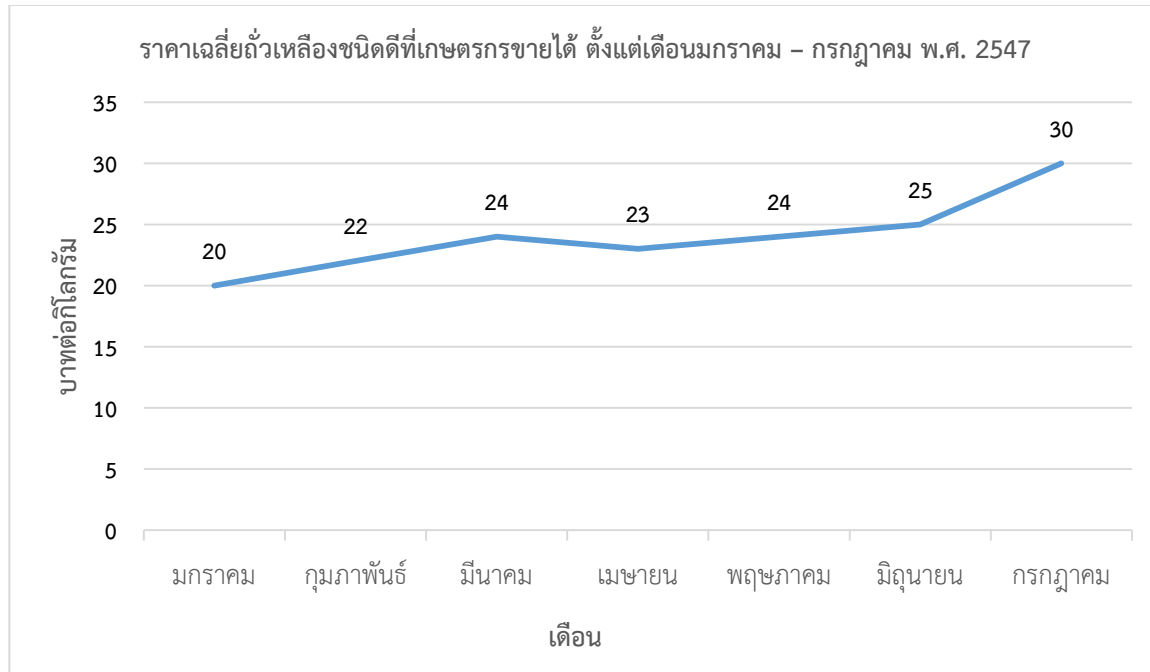
รูปที่ 1.2 แผนภูมิวงกลมแสดงส่วนผสมของถั่วชนิดต่างๆ

กราฟเส้น

กราฟเส้นหรือแผนภูมิเส้นเป็นกราฟที่สร้างได้ง่าย กราฟเส้นนี้อาจจะแสดงเป็นเส้นตรงหรือเส้นโค้งก็ได้ ทั้งนี้ขึ้นอยู่กับข้อมูล ข้อมูลอาจจะมึลักษณะเป็นอนุกรมหรือข้อมูลที่เกี่ยวข้องกับเวลา หรืออาจเป็นข้อมูลเชิงปริมาณด้วยก็ได้ เช่น จำนวนยอดขายในแต่ละเดือน แต่ละปี ปริมาณฝนตกในแต่ละอาทิตย์ ซึ่งนอกจากจะช่วยให้เห็นลักษณะความแปรปรวนของข้อมูลแล้ว ยังช่วยให้มองเห็นแนวโน้มและความสัมพันธ์ระหว่างข้อมูลในหลายช่วงเวลาได้ชัดเจนขึ้น ซึ่งมักจะใช้กราฟเส้นนำเสนอข้อมูลที่มีลักษณะต้องการแสดงถึงความเคลื่อนไหว มีการเปรียบเทียบระหว่างรายการในการเดียวกัน แสดงรายการในระยะเวลายาว หรือเมื่อต้องการใช้การพยากรณ์ข้อมูลในอนาคต การนำเสนอข้อมูลโดยใช้กราฟเส้นสามารถทำได้หลายแบบ เช่น นำเสนอด้วยกราฟเส้นเชิงเดียว ในกรณีที่มีตัวแปรเดียว ดังรูปที่ 1.4 หรือใช้กราฟเส้นเชิงซ้อนเมื่อมีจำนวนตัวแปรเพิ่มขึ้น ดังรูปที่ 1.4

ตารางที่ 1.7 ราคาเฉลี่ยถั่วเหลืองชนิดดีที่เกษตรกรขายได้ ตั้งแต่เดือนมกราคม – กรกฎาคม พ.ศ. 2547 (บาทต่อกิโลกรัม)

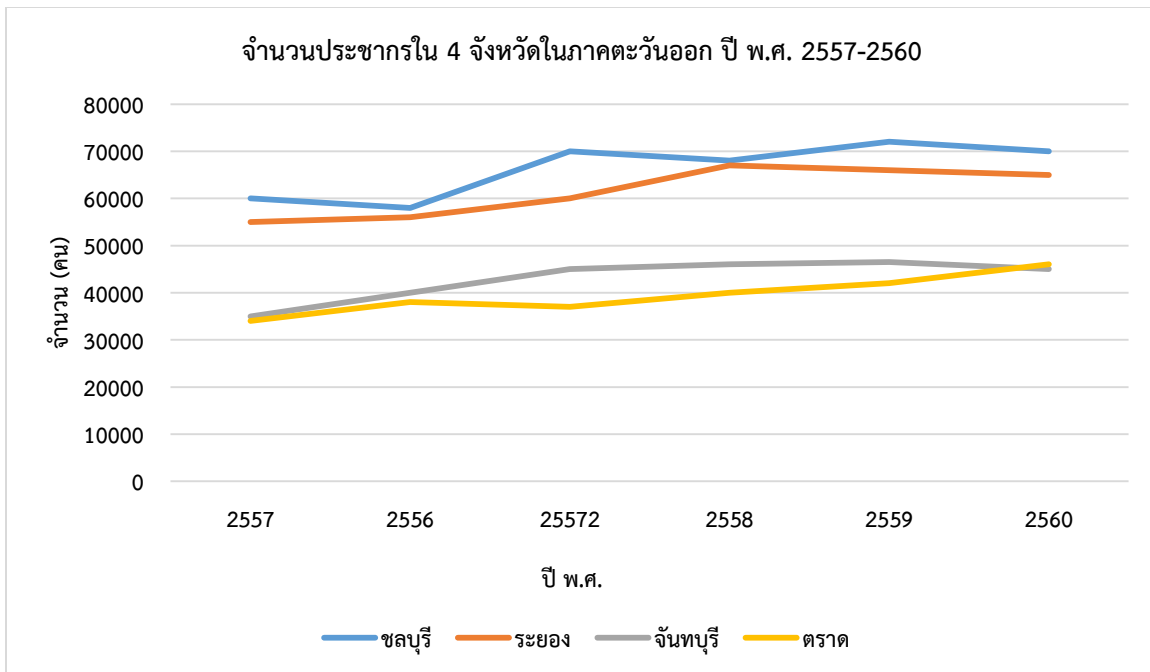
เดือน	มกราคม	กุมภาพันธ์	มีนาคม	เมษายน	พฤษภาคม	มิถุนายน	กรกฎาคม
ราคาเฉลี่ยถั่ว	20	22	24	23	24	25	30



รูปที่ 1.3 แสดงกราฟเส้นของราคาเฉลี่ยถัวเฉลี่ยของชนิดดีที่เกษตรกรขายได้ ตั้งแต่เดือนมกราคม – กรกฎาคม พ.ศ. 2547

ตารางที่ 1.8 จำนวนประชากรใน 4 จังหวัดในภาคตะวันออก ปี พ.ศ. 2557-2560

	2557	2556	2557	2558	2559	2560
ชลบุรี	60000	58000	70000	68000	72000	70000
ระยอง	55000	56000	60000	67000	66000	65000
จันทบุรี	35000	40000	45000	46000	46500	45000
ตราด	34000	38000	37000	40000	42000	46000



รูปที่ 1.4 แสดงกราฟเส้นของจำนวนประชากรใน 4 จังหวัดในภาคตะวันออก ปี พ.ศ. 2557-2560

1.8 การวิเคราะห์ข้อมูลเบื้องต้น

หากข้อมูลที่ได้รับมาเป็นข้อมูลดิบ (Raw data) ข้อมูลดิบเหล่านี้ถ้ามีจำนวนมาก เราจะไม่สามารถมองเห็นลักษณะของข้อมูลได้ ดังนั้นจึงต้องมีการจัดเตรียมข้อมูลดิบให้เป็นหมวดหมู่ เพื่อดำเนินการวิเคราะห์ต่อไป

1. การแจกแจงความถี่ (Frequency distribution) เป็นการนำข้อมูลดิบที่ถูกเก็บรวบรวม ในกรณีที่ข้อมูลดิบมีจำนวนมากเราจะนำข้อมูลเหล่านั้นมาแบ่งออกเป็นกลุ่มตามลักษณะต่างๆ หรือจัดให้ข้อมูลที่มีค่าใกล้เคียงกันหรือซ้ำกันอยู่ในกลุ่มเดียวกัน แล้วหาจำนวนข้อมูลในแต่ละหมวดหมู่ แล้วนำมาเสนอในรูปแบบตาราง เรียกว่าตารางแจกแจงความถี่

การสร้างตารางแจกแจงความถี่มีวิธีการดังนี้

- 1) กำหนดจำนวนชั้นที่ต้องการสร้างให้เหมาะสมกับข้อมูล
- 2) หาค่าพิสัย (Range) โดยการหาผลต่างระหว่างข้อมูลที่มีค่ามากที่สุดและข้อมูลที่มีค่าน้อยที่สุด
- 3) หาความกว้างของชั้นจากค่าของพิสัยหารด้วยจำนวนชั้นที่ต้องการ ดังสูตร

$$\text{ความกว้างของชั้น} = \text{พิสัย} / \text{จำนวนชั้น}$$

- 4) คำนวณหาขีดจำกัดของชั้น โดยจะกำหนดให้ขีดจำกัดล่างของชั้นต่ำสุดครอบคลุมข้อมูลที่มีค่าต่ำสุด เริ่มจากข้อมูลตัวที่น้อยที่สุด และให้ขีดจำกัดบนของชั้นที่มีค่าสูงสุดครอบคลุมข้อมูลที่มีค่าสูงสุด
- 5) นับจำนวนข้อมูลที่ตกอยู่ในแต่ละชั้นแล้วนำค่าความถี่ไปใส่ในตาราง

ตัวอย่างที่ 1.1 ข้อมูลเวลาที่นักเรียนใช้ในการอ่านหนังสือ(นาที)ต่อสัปดาห์จำนวน 50 คน

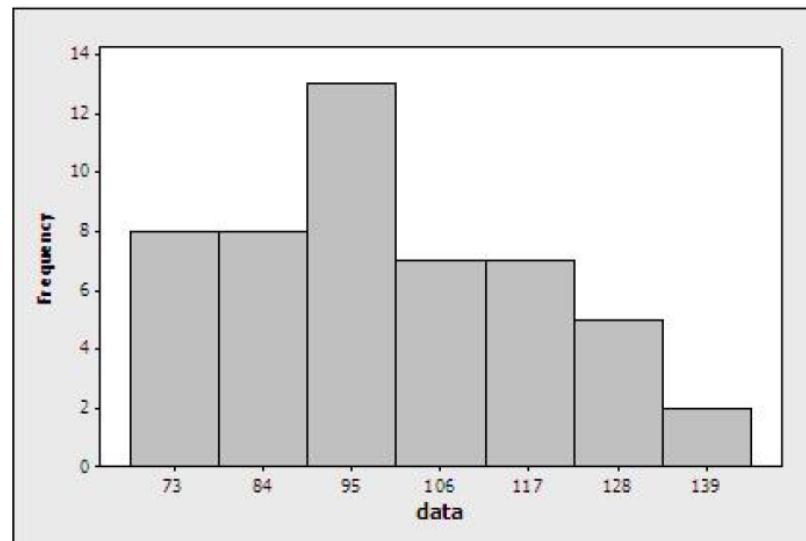
112	72	69	97	107	73	92	76	126	128
118	127	124	82	104	132	92	108	96	100
92	115	76	91	95	141	81	80	106	84
119	113	68	98	115	106	95	100	85	94
86	73	134	83	102	81	98	75	106	119

จากข้อมูลดังกล่าว หากให้สร้างตารางแจกแจงความถี่ โดยมีจำนวนชั้น 7 ชั้น จะได้ดังนี้

ตารางที่ 1.9 ตารางแจกแจงความถี่ของเวลาที่นักเรียนใช้ในการอ่านหนังสือ(นาที)ต่อสัปดาห์จำนวน 50 คน

เวลา (นาที)	จำนวนนักเรียน
68-78	8
79-89	8
90-100	13
101-111	7
112-122	7
123-133	5
134-144	2
รวม	50

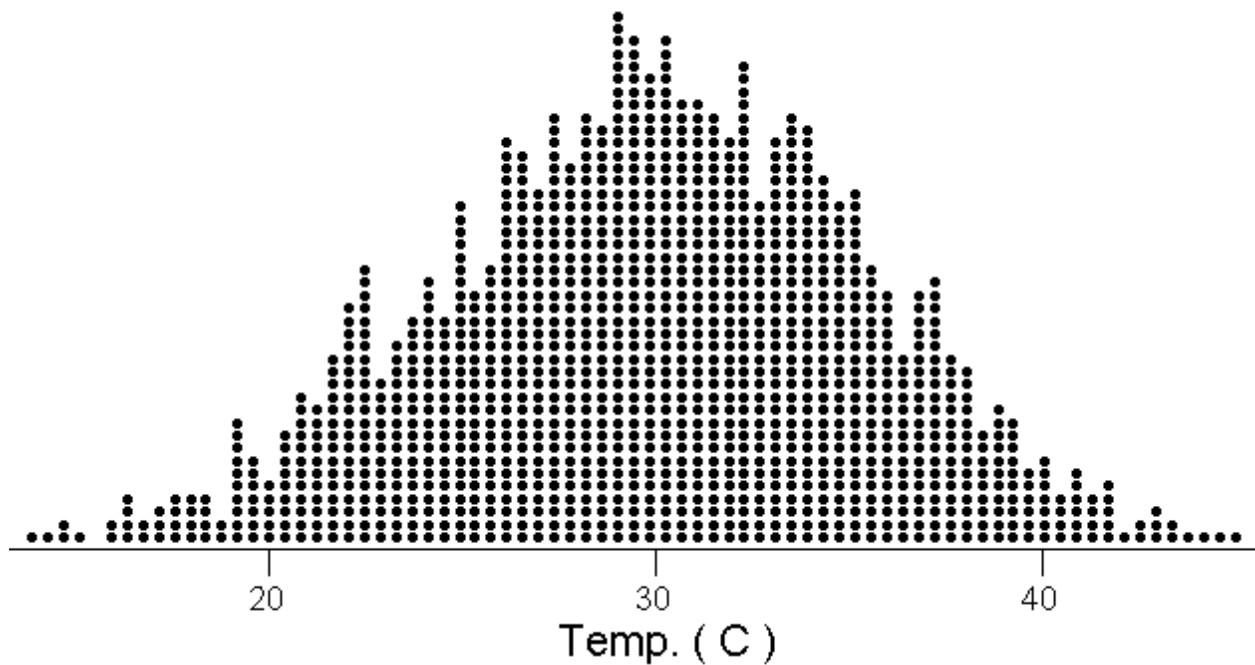
2. ฮิสโตแกรม เป็นการนำข้อมูลที่ได้แจกแจงความถี่แล้วมาแสดงเป็นภาพ ซึ่งมีลักษณะเป็นแท่งสี่เหลี่ยมผืนผ้าโดยใช้ความกว้างของชั้นเป็นความกว้างของแท่งแต่ละแท่ง และความสูงของแต่ละแท่งเท่ากับจำนวนความถี่ของแต่ละชั้น ใช้พิจารณาถึงลักษณะการแจกแจงและรูปแบบของข้อมูลที่น่ามาวิเคราะห์ได้ง่ายขึ้น เมื่อนำตารางแจกแจงความถี่ดังตารางที่ 1.9 นำมาสร้างฮิสโตแกรมจะได้ดังรูปที่ 1.5



รูปที่ 1.5 แสดงฮิสโตแกรมของเวลาที่นักเรียนใช้ในการอ่านหนังสือ(นาทิต่อสัปดาห์จำนวน 50 คน

3. แผนภาพจุด (Dot Plot)

เป็นกราฟที่มีคุณสมบัติเช่นเดียวกับ Histogram โดยใช้จุดแทนตัวข้อมูล หากค่าเป็นค่าเดียวกัน จุดก็จะวางทับกันเป็นชั้นๆ ขึ้นไป ตัวอย่างดังรูปที่ 1.6



รูปที่ 1.6 ตัวอย่างการแสดงผลข้อมูลแผนภาพจุด

ที่มา: <https://sites.google.com/site/mystatistics01/chapter2/graphical-data-presentation>

