บทที่ 1
ความรู้ทั่วไปเกี่ยวกับสถิติ

ในปัจจุบันสถิติได้เข้ามามีบทบาทในชีวิตประจำวันของคนเรามากขึ้น ส่วนใหญ่เรามักจะได้ยินหรือได้ฟังสถิติอยู่เสมอ ไม่ว่าจะเป็นจากการอภิปราย จากการสนทนา จากวิทยุ และจากโทรทัศน์ หรือในหน้าหนังสือพิมพ์รายวันก็มีตัวเลขสถิติแสดงอยู่บ่อย ๆ เช่น สถิติจำนวนคนเกิดคนตาย สถิติจำนวนประชากร สถิติผลผลิตทางการเกษตร สถิติจำนวนอุบัติเหตุบนท้องถนน ฯลฯ เมื่อได้ยินคำว่า “สถิติ” หลายคนจะนึกถึงตัวเลข ทั้งนี้เพราะสถิติที่คนทั่วไปรู้จักและเกี่ยวข้องจะอยู่ในรูปของตัวเลข ตัวเลขมักแสดงในรูปของยอดรวม ซึ่งประมวลจากข้อมูลเบื้องต้นเป็นจำนวนมาก ๆ หรืออาจอยู่ในรูปใดๆ ซึ่งได้มาด้วยการดำเนินการคำนวณอย่างหนึ่งอย่างใดตามวิธีการทางสถิติ เพื่อทราบความสัมพันธ์หรือลักษณะโดยส่วนรวมของข้อมูลเบื้องต้นเหล่านั้น สถิติในความหมายที่เป็นตัวเลขนี้ บางทีก็เรียกกันว่า ข้อมูลสถิติ

เมื่อกล่าวถึงสถิติในรูปที่เป็นตัวเลข หรือเรียกว่า ข้อมูลสถิติ ควรทำความเข้าใจถึงความแตกต่างระหว่างคำว่าข้อมูลและข้อมูลสถิติ ข้อมูล เป็นคำที่ใช้เรียกข้อเท็จจริง ซึ่งได้มาจากการเก็บรวบรวมโดยวิธีใด ๆ ก็ได้ ข้อเท็จจริงเหล่านี้ อาจเป็นตัวเลขหรือไม่เป็นตัวเลขก็ได้ ข้อมูลสถิติจะต้องเกิดจากข้อมูลจำนวนมาก ถ้าข้อมูลเพียงตัวเดียวไม่ถือว่าเป็นข้อมูลสถิติ เช่น ปริมาณน้ำฝนที่ตกใน 1 วัน เป็นข้อมูลแต่ยอดรวมหรือค่าเฉลี่ยต่อวันของปริมาณน้ำฝนที่ตกใน 1 เดือน หรือ 1 ปี จึงจะถือได้ว่าเป็นข้อมูลสถิติ

ความหมายของสถิติ

คำว่า **สถิติ (Statistics)** มีความหมาย 2 ประการ ดังนี้

ประการแรกสถิติ หมายถึง ข้อมูลหรือตัวเลขที่แทนข้อเท็จจริงเกี่ยวกับเรื่องต่าง ๆ ที่เราสนใจหรือที่อยู่รอบตัวเรา เช่น ปริมาณน้ำฝน ราคาน้ำมัน ผลผลิตทางการเกษตร ฯลฯ

ประการที่สอง หมายถึง ศาสตร์หรือวิชาที่ว่าด้วยหลักการและวิธีการทางสถิติ ซึ่งประกอบด้วย การเก็บรวบรวมข้อมูล การจัดระเบียบข้อมูล หรือการนำเสนอข้อมูล การวิเคราะห์และการแปลความหมายข้อมูล ซึ่งจะเรียกสถิติตามความหมายนี้ว่า สถิติศาสตร์

ตามความหมายสถิติในแง่ที่เป็นศาสตร์ สถิติแบ่งออกเป็น 2 ประเภทคือ

1. สถิติพรรณนา (Descriptive Statistics) หมายถึง วิชาที่ว่าด้วยการเก็บรวบรวมข้อมูล การประมวลผลหรือนำเสนอด้วยการบรรยาย ตาราง แผนภูมิ แผนภาพ การวิเคราะห์ ซึ่งประกอบด้วยการวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดการกระจาย การวัดตำแหน่ง ตลอดจนการตีความเพื่อหาข้อสรุปของข้อมูลชุดนั้น ๆ หรือระหว่างข้อมูลมากกว่าหนึ่งชุด

2. สถิติอนุมาน (Inferential Statistics) หมายถึง วิชาสถิติที่ว่าด้วยการสรุปถึงข้อเท็จจริงของข้อมูลทั้งหมด ในลักษณะการประมาณค่า (Estimation) และการทดสอบสมมติฐาน (Testing Hypothesis) โดยอาศัยเพียงบางส่วนของข้อมูลที่เก็บรวบรวมมาได้

ข้อมูลและระดับการวัด

ข้อมูล (Data) หมายถึง ข้อเท็จจริงต่าง ๆ ที่เราสนใจจะศึกษา ซึ่งอาจเป็นตัวเลขหรือข้อความที่ไม่ใช่ตัวเลขก็ได้ เช่น รายได้ของเกษตรกรในหมู่บ้านแห่งหนึ่ง จำนวนคนไข้ที่ป่วยเป็นโรคหัวใจในโรงพยาบาล สินค้าส่งออกของประเทศไทยระหว่างปีพ.ศ. 2550 - 2555 เป็นต้น

ระดับการวัดข้อมูล (Levels of Measurement) สามารถจำแนกได้หลายรูปแบบ เช่น

จำแนกตามประเภทของข้อมูล ได้แก่

1. ข้อมูลเชิงปริมาณ (Quantitative data) หมายถึง ข้อมูลที่อยู่ในรูปตัวเลข (numerical data) ที่แสดงถึงปริมาณ เช่น น้ำหนัก ส่วนสูง ความยาว อายุ คะแนน เป็นต้น
2. ข้อมูลเชิงคุณภาพ (Qualitative data) หมายถึง ข้อมูลที่แสดงถึงสถานภาพ คุณลักษณะ หรือคุณสมบัติ เช่น เชื้อชาติ ศาสนา เพศ ความนิยม เป็นต้น

จำแนกตามระดับการวัดข้อมูล แบ่งออกเป็น 4 ระดับ ได้แก่ มาตรานามบัญญัติ (Nominal scale) มาตราเรียงอันดับ (Ordinal scale) มาตราอันตรภาค (Interval scale) และมาตราอัตราส่วน (Ratio scale) ระดับการวัดที่หยาบหรือต่ำที่สุด คือ มาตรานามบัญญัติ และระดับการวัดที่สูงที่สุด คือ มาตราอัตราส่วน ซึ่งเป็นระดับการวัดที่ให้สารสนเทศของค่าสังเกตมากที่สุด รายละเอียดของระดับการวัด มีดังต่อไปนี้

1. มาตรานามบัญญัติ (Nominal scale) เป็นมาตราการวัดที่ใช้กับข้อมูลที่มีลักษณะหยาบหรือต่ำสุด อาจเป็นการกำหนดสัญลักษณ์หรือตัวเลขเพื่อจำแนกประเภทสิ่งของหรือคุณลักษณะต่างๆ เท่านั้น ไม่สามารถแสดงให้เห็นปริมาณมากน้อยหรือสูงต่ำแต่อย่างใด ดังนั้น จึงไม่สามารถนำตัวเลขเหล่านั้นมาบวก ลบ คูณ หารได้ เช่น เพศ ที่แบ่งออกได้เป็น 2 ประเภทเท่านั้น คือ ชายและหญิง โดยให้ 1 เป็นสัญลักษณ์แทนเพศชาย และ 2 เป็นสัญลักษณ์แทนเพศหญิง เท่านั้น ไม่สามารถนำสัญลักษณ์ที่เป็นตัวเลข 1 กับ 2 มาบวก ลบ คูณ หรือหารได้แต่อย่างใด หรือไม่สามารถสรุปได้ว่า 2 มากกว่า 1 เป็นต้น

2. มาตราเรียงอันดับ (Ordinal scale) เป็นมาตราการวัดที่มีความละเอียดการวัดเพิ่มขึ้นหรือสูงกว่ามาตรานามบัญญัติ เพราะสามารถบอกลำดับและความแตกต่าง แต่ไม่สามารถบอกได้ว่าคุณลักษณะหรือคุณสมบัติเหล่านี้มีปริมาณมากน้อยกว่ากันเท่าใด กล่าวอีกนัยหนึ่ง ข้อมูลในระดับนี้ไม่สามารถนำมาคำนวณทางคณิตศาสตร์ได้เช่นเดียวกับมาตรานามบัญญัติ เช่น การประกวดนางสาวไทย อันดับที่ 1 คือ นางสาวไทย อันดับที่ 2 คือ รองนางสาวไทยคนที่ 1 และอันดับที่ 3 คือ รองนางสาวไทยคนที่ 2 เช่นนี้เรารู้ถึงอันดับและความแตกต่างของความสวยของนางสาวไทยทั้ง 3 คน แต่เราไม่สามารถจัดจำแนกได้ว่า อันดับ 1 สวยมากกว่าอันดับ 2 เท่าใด เป็นต้น

3. มาตราอันตรภาคหรือมาตราช่วง (Interval scale) เป็นมาตราการวัดที่สามารถทราบได้ว่าสิ่งที่จะวัดมีช่วงมากน้อยเท่าใด โดยแต่ละช่วงมาตรานี้มีค่าเท่า ๆ กัน ทำให้เราทราบถึงความแตกต่างที่ห่างกันเป็นช่วงได้ และค่าที่ได้จากการวัดสามารถนำมาคำนวณทาง

คณิตศาสตร์ได้ เช่น เราจะบอกความแตกต่างของน้ำร้อนระหว่าง 60° C กับ 80° C เท่ากับความแตกต่างระหว่าง 100° C กับ 120° C โดยดูจากช่วงที่ห่างกันเท่ากับ20° C แต่อย่างไรก็ตามยังสามารถบอกได้ว่า ที่อุณหภูมิ 120° C มีความร้อนเป็น 2 เท่าของอุณหภูมิ 60° C ทั้ง ๆ ที่ 120 มีค่าเป็น 2 เท่าของ 60 ทั้งนี้เพราะว่าที่ 0° C ไม่ได้แปลว่าไม่มีความร้อนเลย หรือเราไม่สามารถบอกได้ว่าคนที่สอบได้คะแนน 100 คะแนนเต็มมีความรู้เป็นสองเท่ามากกว่าคนที่สอบได้คะแนน 50 เพราะคนที่สอบได้ 0 คะแนนไม่ได้แปลว่าไม่มีความรู้เลย ทั้งนี้เนื่องจากในมาตราการวัดระดับนี้**จะไม่มีศูนย์แท้**นั่นเอง

4. มาตราอัตราส่วน (Ratio scale) เป็นมาตราการวัดที่ดีที่สุดและวัดได้อย่างละเอียดที่สุด ตัวเลขที่วัดได้สามารถสื่อความหมายตรงตามค่าของสิ่งที่วัด และเป็นมาตราวัดที่ข้อมูลมี**ค่าเป็นศูนย์แท้** คือ ถ้าค่าตัวเลขที่วัดได้มีค่าเป็นศูนย์ ก็แปลว่า สิ่งที่วัดนั้นมีค่าศูนย์ด้วย ข้อมูลที่อยู่ในมาตราวัดระดับนี้ ได้แก่ อายุ น้ำหนัก ส่วนสูง ระยะทาง เป็นต้น ข้อมูลที่วัดได้สามารถนำมาทำการคำนวณทางคณิตศาสตร์ได้ ตัวอย่างเช่น คนที่มีน้ำหนัก 40 กิโลกรัม ย่อมแปลว่า มีน้ำหนักมากกว่าหรือหนักเป็นสองเท่าของคนที่มีน้ำหนักเพียง 20 กิโลกรัม ทั้งนี้เพราะว่าถ้าคนที่ไม่มีน้ำหนักเลยจะเท่ากับศูนย์ หรือระยะทางของถนนที่ยาว 50 เมตร ย่อมยาวเป็น 10 เท่าของถนนที่มีระยะทางยาวเพียง 5 เมตร เป็นต้น

แหล่งของข้อมูล

แหล่งของข้อมูล (Source of data) หมายถึง สิ่งที่ทำให้ได้มาซึ่งข้อมูลทางสถิติ แบ่งเป็น 2 ประเภท คือ

1. แหล่งปฐมภูมิ (Primary source) เป็นแหล่งของข้อมูลที่เป็นต้นกำเนิดของข้อมูลนั้น ๆ เราเรียกข้อมูลนี้ว่า ข้อมูลปฐมภูมิ เช่น ข้อมูลที่ได้จากการสัมภาษณ์ ข้อมูลที่ได้จากการทดลอง ข้อมูลที่ได้จากการกรอกแบบสอบถาม เป็นต้น

2. แหล่งทุติยภูมิ (Secondary source) เป็นแหล่งของข้อมูลที่ได้รวบรวมเอาข้อมูลเอาไว้เมื่อผู้สนใจต้องการศึกษาข้อมูลที่ได้รวบรวมเอาไว้ เราเรียกข้อมูลนี้ว่า ข้อมูลทุติยภูมิ เช่น ข้อมูลที่คัดเลือกมาจากทะเบียนราษฎร์ ข้อมูลที่คัดมาจากบัญชีรายชื่อผู้ป่วย ข้อมูลจากหนังสือพิมพ์ เป็นต้น

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูล แบ่งเป็น 3 วิธี ดังนี้

1. การเก็บรวบรวมข้อมูลจากงานทะเบียน เช่น มหาวิทยาลัยจะมีการบันทึกหมายเหตุรายวันในเรื่องจำนวนอาจารย์ เจ้าหน้าที่ที่มาปฏิบัติราชการ จำนวนอาจารย์ เจ้าหน้าที่ที่ลา จำนวนอาจารย์ที่ไปราชการ เป็นต้น

2. การเก็บรวบรวมข้อมูลโดยการสำรวจ เป็นการเก็บรวบรวมข้อมูลจากหน่วยที่ศึกษาโดยตรง เช่น การสำรวจความคิดเห็นของประชาชนในการร่างรัฐธรรมนูญ ซึ่งหน่วยที่ศึกษา คือ ประชาชนคนไทย การเก็บรวบรวมข้อมูลโดยการสำรวจจะแบ่งออกเป็น 2 ประเภท

สำมะโน เก็บรวบรวมข้อมูลจากทุกๆหน่วยศึกษาของประชากร ซึ่งอาจเป็นการแจงนับโดยการนับ วัดหรือชั่ง การสัมภาษณ์ที่มีการเผชิญหน้ากันตลอดจนโดยการอาศัยสื่อกลางต่าง ๆ

การสำรวจตัวอย่าง เก็บรวบรวมข้อมูลจากบางหน่วยศึกษาของประชากร โดยแต่ละหน่วยของประชากรที่ถูกแจงนับจะเป็นไปโดยสุ่ม

3. การเก็บรวบรวมข้อมูลจากการทดลอง ข้อมูลบางประเภทไม่สามารถหาได้จากการสำรวจ แต่จัดทำได้จากการทดลอง เช่น การศึกษาวิธีการปลูกพืชที่แตกต่างกัน 3 วิธี การเปรียบเทียบผลผลิตของข้าว 4 พันธุ์ และการเปรียบเทียบคุณภาพของยารักษาโรคผิวหนัง 4 ชนิด เป็นต้น

การนำเสนอข้อมูลสถิติ

ในการรวบรวมข้อมูลสถิตินั้นเราเก็บข้อมูลได้แล้วก็นำไปสู่การนำเสนอข้อมูลที่เรารวบรวมได้นำเสนอหรือแสดงให้ผู้อื่นทราบและเข้าใจ ซึ่งเราจะนำเสนอข้อมูลได้ในหลายลักษณะด้วยกัน ซึ่งขึ้นอยู่กับลักษณะของข้อมูลนั้น ข้อมูลบางอย่างอาจจะไม่จำเป็นต้องนำไปวิเคราะห์ทางสถิติ อาจจะนำเสนอข้อมูลอย่างง่าย ๆ ซึ่งจะทำให้น่าสนใจมากขึ้น

1. การนำเสนอข้อมูลโดยบทความ (Text Presentation)

การนำเสนอข้อมูลโดยบทความจะมีลักษณะการเสนอเป็นบทความสั้น ๆ และมีข้อมูลตัวเลขอยู่ด้วย ซึ่งทำให้อ่านเข้าใจง่าย อาจเป็นการนำเสนอบทความทางวิทยุ โทรทัศน์ หรืออาจจะเป็นบทความในหนังสือพิมพ์ วารสาร และรายงานต่าง ๆ

ตัวอย่าง 1.1

ในปีพ.ศ. 2556 มีจำนวนผู้จบการศึกษาระดับปริญญาตรีจากมหาวิทยาลัยของรัฐประมาณ 550,000 คน ซึ่งคาดว่าในปีพ.ศ. 2557 เพิ่มขึ้นเป็น 970,000 คน

ในปีพ.ศ. 2556 ชาวสวนเงาะ จังหวัดเชียงใหม่ สามารถส่งเงาะออกสู่ตลาดเป็นจำนวนเงิน 100 ล้านบาท ซึ่งมากกว่าปีพ.ศ. 2555 จำนวน 40 ล้านบาท

2. การนำเสนอโดยบทความกึ่งตาราง (Semi-Tabular Presentation)

การนำเสนอโดยบทความกึ่งตาราง เป็นการนำเสนอข้อมูลโดยแยกตัวเลขออกจากข้อความ หรือการนำเสนอบทความแต่มีการตั้งแนวตัวเลขขึ้นในบทความ เพื่อให้เห็นตัวเลขชัดเจน และเปรียบเทียบสะดวกเมื่อต้องการ

ตัวอย่าง 1.2

ภูมิลำเนาของนักศึกษาวิทยาลัยแห่งหนึ่งระหว่างปี 2559 – 2562

	จำนวน (คน)			
ภูมิลำเนา	**2559**	**2560**	**2561**	**2562**
1. กรุงเทพฯ	2,540	2,590	2,556	2,618
2. ภาคเหนือ	350	244	310	287
3. ภาคกลาง	825	1,300	1,310	1,544
4. ภาคใต้	408	325	368	387
5. ภาคตะวันออกเฉียงเหนือ	520	458	488	481
รวม	**4,643**	**4,917**	**5,032**	**5,317**

3. การนำเสนอโดยตาราง (Tabular Presentation)

การนำเสนอข้อมูลโดยใช้ตารางกรอกข้อมูลที่เป็นตัวเลขโดยแบ่งเป็นแถวตั้ง (Columns) และแถวนอน (Row) เพื่อจัดข้อมูลให้เป็นระเบียบ ซึ่งลักษณะของตารางขึ้นอยู่กับจุดมุ่งหมายของการนำเสนอข้อมูล

ตารางที่ 1.1 จำนวนของลูกจ้างภาคเอกชนและภาครัฐบาลที่ได้รับผลประโยชน์เพิ่มเติมไตรมาส 3 พ.ศ. 2562 (หน่วย : คน)

ผลประโยชน์เพิ่มเติม	**จำนวน (คน)**
โบนัส	144,400
ค่าล่วงเวลา	1,879,900
เงินสดอื่น ๆ	2,566,000
อาหาร	1,472,100
เสื้อผ้า	139,600
ที่อยู่อาศัย	737,600
อื่น ๆ	652,800
รวม	**7,592,400**

4. การนำเสนอด้วยกราฟหรือแผนภูมิ (Graphical Presentation)

เมื่อได้จัดข้อมูลที่จะนำเสนอแล้ว เราอาจจะพิจารณาในการนำเสนอข้อมูลด้วยกราฟหรือแผนภูมิ ซึ่งเป็นวิธีที่ใช้ได้ดี เพราะรูปภาพที่แสดงข้อมูลจะทำให้เกิดความน่าสนใจ ทำให้อ่านเข้าใจได้ง่าย และรวดเร็วกว่าวิธีอื่น ๆ การนำเสนอด้วยกราฟหรือแผนภูมิมีหลายลักษณะดังนี้

4.1 แผนภูมิวงกลม (Pie Chart)

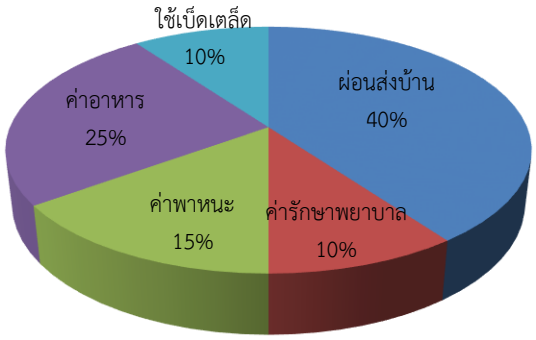
4.2 แผนภูมิแท่งหรือกราฟแท่ง (Bar Chart)

4.3 กราฟเส้น (Line Graphs)

4.4 แผนภูมิภาพ (Pictogram)

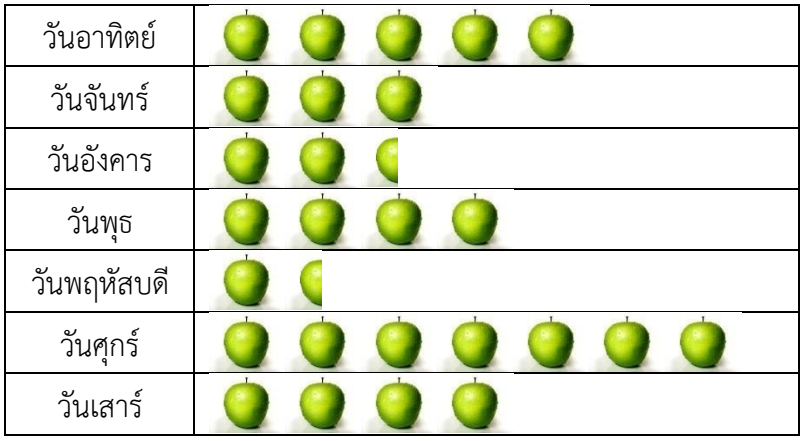
4.1 กราฟวงกลมหรือแผนภูมิวงกลม (Pie Chart) เป็นการนำเสนอข้อมูลที่รวบรวมได้ในรูปวงกลม โดยมีการแบ่งพื้นที่ภายในวงกลมออกเป็นส่วน ๆ ในการเปรียบเทียบ แต่มีหลายลักษณะของกลุ่มประชากร ดังตัวอย่างต่อไปนี้

รูปที่ 1.1 รายจ่ายต่าง ๆ ในแต่ละเดือน



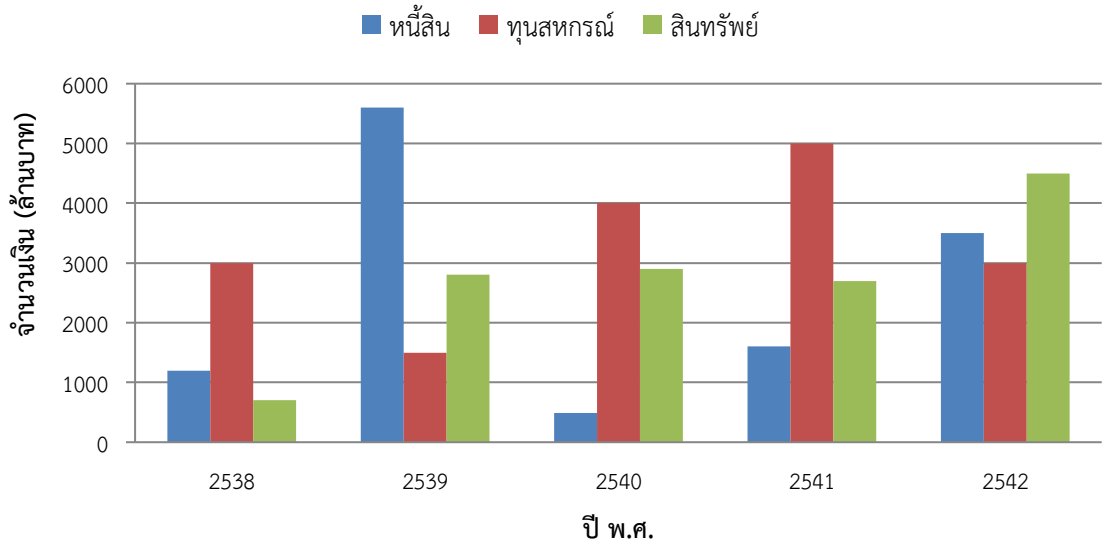
4.2 แผนภูมิรูปภาพ (Pictogram) เป็นการใช้รูปภาพแทนค่าตัวเลขจำนวนหนึ่งของข้อมูลที่นำเสนอ เช่น ภาพของรถยนต์ 1 คัน แทนจำนวนรถที่นำเสนอ 1,000 คน หรือภาพของคน 1 ภาพ แทนประชากรที่นำเสนอ 100 คน ซึ่งรูปภาพนั้นจะแทนของจริงจำนวนเท่าไรก็ได้ แล้วแต่ปริมาณมากน้อยของข้อมูลที่นำเสนอ จะทำให้ผู้อ่านเข้าใจได้ง่าย แปลความหมายได้ทันที่และน่าสนใจมากขึ้น

รูปที่ 1.2 แผนภูมิรูปภาพแสดงจำนวนแอปเปิ้ลเขียวที่ขายได้ตลอดทั้งสัปดาห์



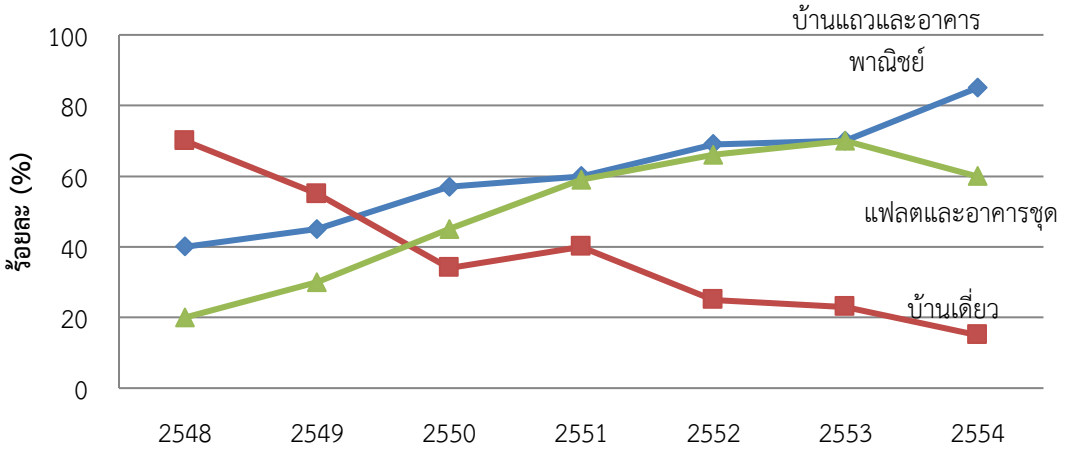
4.3 แผนภูมิแท่งหรือกราฟแท่ง (Bar Chart) คือ แผนภูมิที่ประกอบด้วยรูปสี่เหลี่ยมผืนผ้าที่มีความยาวของแต่ละรูปเป็นขนาดของข้อมูล มีช่องไฟแต่ละช่องความกว้างจะคงที่ จะใช้กับกาเปรียบเทียบรายการข้อมูลที่แตกต่างกันหลายรายการ หรือข้อมูลที่จำแนกตามลักษณะคุณภาพ เวลา หรือความถี่ ซึ่งทำให้ผู้คนเข้าใจง่ายด้วยตนเอง

รูปที่ 1.3 จำนวนเงินที่แสดงในปีต่าง ๆ ตั้งแต่ พ.ศ. 2538 - 2542



4.4 กราฟเส้น (Line Graphs) การนำเสนอโดยกราฟเส้นจะเป็นที่นิยมใช้มากซึ่งใช้กับข้อมูลอนุกรมเวลา (Time Series Data) ซึ่งแสดงการเปลี่ยนแปลงลำดับก่อนหลังของเวลาที่ข้อมูลนั้นเกิดขึ้นและมีจำนวนมาก เป็นการสร้างที่ง่าย อาจจะเป็นเส้นตรงหรือเส้นโค้งก็ได้ ขึ้นอยู่กับลักษณะข้อมูลที่มีอยู่ใช้เปรียบเทียบระหว่างหลายรายการในระยะยาว

รูปที่ 1.4 แผนภูมิเส้นแสดงการเปรียบเทียบสัดส่วนประเภทที่อยู่อาศัยสร้างเสร็จ ปี พ.ศ. 2548 - 2554



การแจกแจงความถี่ด้วยตาราง

การแจกแจงความถี่ด้วยวิธีนี้ เป็นการนำเสนอข้อมูลเชิงปริมาณที่เก็บรวบรวมได้ให้อยู่ในรูปของตารางที่เรียกว่า **ตารางแจกแจงความถี่** (Frequency table) การแจกแจงวิธีนี้สามารถทำได้ 3 วิธี คือ

1. การแจกแจงจัดเรียง (Listed distribution) เป็นวิธีการนำเสนอข้อมูลเชิงปริมาณแบบง่ายที่สุด โดยมีข้อมูลแตกต่างกันไม่มากนักและไม่มีข้อมูลค่าใดมีค่าซ้ำกัน โดยนำข้อมูลทุกค่าที่มีความถี่เท่ากับ 1 มาเรียงตามลำดับ โดยทั่วไปเรียงจากน้อยไปมาก ทำให้เราทราบว่าค่าใดเป็นค่าสูงสุดและต่ำสุด

2. การแจกแจงความถี่ชนิดไม่จัดข้อมูลเป็นหมวดหมู่ (Ungrouped frequency distribution) ในกรณีที่ข้อมูลแตกต่างกันไม่มากนักและมีบางค่าซ้ำกัน ในการนำเสนอข้อมูลดังกล่าวจะสร้างสดมภ์ที่แสดงรอยขีด (Tally) เพื่อแสดงความถี่ของข้อมูลแต่ละค่าและการคำนวณหาความถี่ของข้อมูลแต่ละค่าจะทำได้ง่ายและไม่สับสน

ตัวอย่าง 1.3 จงสร้างตารางแจกแจงความถี่จากข้อมูลจำนวนหนังสือที่เรียบเรียงโดยอาจารย์ 100 คน ในมหาวิทยาลัยแห่งหนึ่ง ต่อไปนี้

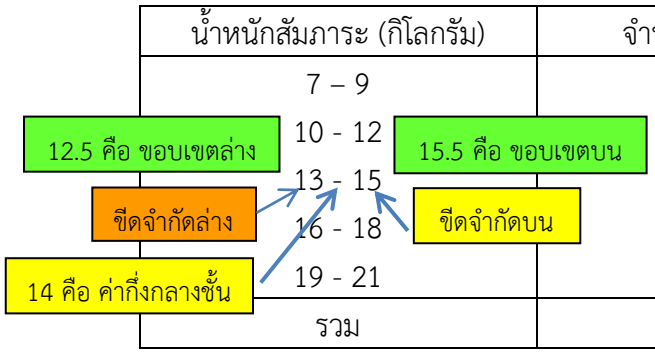
0 2 1 2 0 3 0 2 2 1 0 5 4 1 1 0 1 2 0 4 3 0 0 1 1 3 1 2 2 1 1 0 0
0 0 1 2 2 3 4 5 0 0 0 1 1 0 0 0 1 0 1 1 2 1 1 2 6 2 3 3 4 4 1 1 3
5 2 0 0 1 2 0 1 2 0 1 1 0 2 6 0 0 1 0 0 1 2 9 1 7 0 1 1 0 2 0 0 1
0

<u>วิธีทำ</u> ข้อมูลมีค่าสูงสุดเป็น 9 และค่าต่ำสุดเป็น 0 ซึ่งแตกต่างกันไม่มากนักและมีข้อมูลบางค่าซ้ำกัน จึงสร้างตารางแจกแจงความถี่ได้เป็นดังนี้

จำนวนหนังสือ	รอยขีด	ความถี่ (f)
0	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ///	33
1	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	30
2	𝍸 𝍸 𝍸 ///	18
3	𝍸 //	7
4	𝍸	5
5	///	3
6	//	2
7	/	1
8		0
9	/	1
รวม		100

3. การแจกแจงความถี่ชนิดจัดข้อมูลเป็นหมวดหมู่ (Grouped frequency distribution) ในกรณีที่มีข้อมูลแตกต่างกันมาก แล้วสดมภ์แรกของการแจกแจงความถี่จะมีความยาวมาก ดังนั้นเราสามารถสร้างตารางแจกแจงความถี่ของข้อมูลดังกล่าวได้กะทัดรัดยิ่งขึ้น โดยการจัดหมวดหมู่ให้แก่ข้อมูล ดังเช่น

น้ำหนักสัมภาระ (กิโลกรัม)	จำนวนสัมภาระ (f)
7 – 9	2
10 - 12	8
13 - 15	14
16 - 18	19
19 - 21	7
รวม	50



วิธีการสร้างตารางแจกแจงความถี่

ตัวอย่าง 1.4 คะแนนของนักศึกษา 150 คน เป็นดังนี้

27 79 69 40 51 88 55 48 36 61 53 44 12 51 65 42 58 55 69 63 70 48
61 55 60 25 47 78 61 54 57 76 73 62 36 67 40 51 59 68 27 46 62 43
54 83 59 94 72 57 82 45 54 52 71 53 82 69 60 35 41 65 62 75 60 42 55
34 49 45 49 64 40 61 73 44 59 46 71 86 43 69 54 31 56 51 75 44 66 53
80 71 53 56 91 60 41 29 56 57 35 51 43 39 56 27 62 44 85 61 59 89
60 51 71 53 58 26 77 68 62 57 48 69 76 52 49 45 54 41 33 61 80 57
42 45 59 44 68 73 55 70 39 58 69 51 85 46 55 67

<u>วิธีทำ</u> 1. กำหนดจำนวนชั้น ในที่นี้ต้องการตารางแจกแจงความถี่ที่มี 9 ชั้น

2. คำนวณค่าพิสัย (Range) ของข้อมูล เมื่อพิสัยคือผลต่างระหว่างค่าสูงสุดและค่าต่ำสุด

จากข้อมูล มีค่าสูงสุดและค่าต่ำสุดเป็น **94** และ **12** ดังนั้น พิสัย = 94 – 12 = 82

3. คำนวณความกว้างของแต่ละชั้น โดยความกว้างของแต่ละชั้นคือผลหารระหว่างพิสัยและ

จำนวนชั้น คือ $\frac{82}{9}=9.11$ เพื่อความสะดวกจะใช้ความกว้างของแต่ละชั้นเป็น 10

4. กำหนดขีดจำกัดล่างของชั้นแรก แล้วคำนวณขีดจำกัดของแต่ละชั้น ชั้นแรกเป็น 10 ขอบเขตล่างของชั้นแรกจึงเท่ากับ 9.5 บวกค่าดังกล่าวด้วยความกว้างของชั้นเท่ากับ 10 จะได้ขอบเขตบนของชั้นแรกเป็น 19.5 และขีดจำกัดบนของชั้นแรกเป็น 19 ดังนั้นขีดจำกัดของชั้นแรกคือ 10 – 19

5. เขียนรอยขีด

6. คำนวณความถี่ โดยการนับรอยขีดในชั้นนั้นๆ ซึ่งผลรวมของความถี่ทั้งหมดทุกชั้นต้องจำนวนข้อมูล จากวิธีการข้างต้นได้ตารางแจกแจงความถี่ดังนี้

คะแนน	รอยขีด	ความถี่ (f)
10 – 19	/	1
20 - 29	𝍸 /	6
30 - 39	𝍸 ////	9
40 - 49	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 /	31
50 - 59	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //	42
60 - 69	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //	32
70 - 79	𝍸 𝍸 𝍸 //	17
80 - 89	𝍸 𝍸	10
90 - 99	//	2
รวม		150

การวัดแนวโน้มเข้าสู่ส่วนกลาง

การวัดแนวโน้มเข้าสู่ส่วนกลาง เป็นการหาค่าใดค่าหนึ่งเพื่อใช้เป็นตัวแทนของข้อมูล โดยพิจารณาค่ากลางเป็นค่าตัวแทนของข้อมูล การวัดแนวโน้มเข้าสู่ส่วนกลางที่นิยมใช้มีอยู่ 3 วิธี ซึ่งขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ของการนำไปใช้

1. ค่าเฉลี่ยเลขคณิต (Arithmetic mean) เป็นค่าเฉลี่ยที่นิยมใช้กันมากที่สุด แทนด้วย (มิว) เมื่อคำนวณจากข้อมูลประชากร และแทนด้วย $\bar{X}$ (เอกซ์บาร์) เมื่อคำนวณจากข้อมูลตัวอย่าง การคำนวณหาได้จากผลรวมของข้อมูลทุกค่าแล้วหารด้วยจำนวนข้อมูลทั้งหมด การหาค่าเฉลี่ยเลขคณิตแบ่งออกเป็น 2 กรณี คือ

1) การหาค่าเฉลี่ยเลขคณิตในกรณีที่ข้อมูลไม่ได้แจกแจงความถี่

ค่าเฉลี่ยเลขคณิตของข้อมูลประชากร

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

ค่าเฉลี่ยเลขคณิตของข้อมูลตัวอย่าง

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

ตัวอย่าง 1.5 จงหาค่าเฉลี่ยของกลุ่มตัวอย่างที่กำหนดให้ต่อไปนี้

22 23 24 24 25 25 21 23 15 20

<u>วิธีทำ</u>

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{22+23+24+24+25+25+21+23+15+20}{10}$$

$$= \frac{222}{10} = 22.2$$

ดังนั้น ค่าเฉลี่ยของข้อมูลนี้ เท่ากับ 22.2

2) การหาค่าเฉลี่ยในกรณีที่ข้อมูลแจกแจงความถี่

ในการหาค่าเฉลี่ยของข้อมูลที่แจกแจงความถี่ ซึ่งไม่ทราบค่าที่แท้จริงของข้อมูลในแต่ละชั้น ดังนั้น ถ้าให้ค่ากึ่งกลางของแต่ละชั้นเป็นตัวแทนของข้อมูลในชั้นนั้น ๆ แล้ว จะได้ว่า

ค่าเฉลี่ยเลขคณิตของข้อมูลประชากร

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

ค่าเฉลี่ยเลขคณิตของข้อมูลตัวอย่าง

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i X_i}{n}$$

เมื่อกำหนดให้ x_i คือ ค่ากึ่งกลางของชั้นคะแนนที่ i ในชุดข้อมูล

f_i คือ ความถี่ของชั้นที่ i

N คือ จำนวนข้อมูลของประชากร

n คือ จำนวนข้อมูลของกลุ่มตัวอย่าง

ตัวอย่าง 1.6 จงหาค่าเฉลี่ยของข้อมูลตัวอย่างที่กำหนดให้ต่อไปนี้

คะแนน	ค่ากึ่งกลาง (x_i)	ความถี่ (f)	ผลรวมคะแนนในชั้น (fx_i)
20 - 24	22	4	88
25 - 29	27	6	162
30 - 34	32	12	384
35 - 39	37	16	592
40 - 44	42	12	504
45 - 49	47	7	329
50 - 54	52	3	156
รวม		60	2,215

<u>วิธีทำ</u> 1. หาผลรวมของคะแนนได้เท่ากับ 2,215

2. คำนวณค่าเฉลี่ยได้จะได้ว่า

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i X_i}{n} = \frac{2,215}{60} = 36.92$$

ดังนั้น ค่าเฉลี่ยของข้อมูลชุดนี้ เท่ากับ 36.92

2. มัธยฐาน (Median : Me) คือ ค่าที่บอกตำแหน่งกึ่งกลางของข้อมูลทั้งหมดเมื่อได้จัดเรียงข้อมูลจากค่าน้อยไปหาค่ามาก หรือ จากค่ามากไปหาค่าน้อย โดยวิธีการหาได้ 2 วิธี ดังนี้

1) การหาค่ามัธยฐานกรณีข้อมูลที่ไม่ได้แจกแจงความถี่

1. ถ้าข้อมูลเป็นจำนวนคี่ มัธยฐาน คือ ค่าของข้อมูลที่อยู่ตำแหน่ง $\frac{N+1}{2}$

2. ถ้าข้อมูลเป็นจำนวนคู่ คำนวณมัธยฐานโดยการเปรียบเทียบระหว่างข้อมูลที่อยู่ในตำแหน่ง $\frac{N}{2}$ และ $\frac{N}{2}+1$ การหามัธยฐานให้นำเอาคะแนนที่อยู่ตรงตำแหน่งกึ่งกลางทั้งสองค่านั้นมาบวกกันหารสอง

ตัวอย่าง 1.7 จงหามัธยฐานของข้อมูลที่กำหนดให้ 1, 3, 2, 2, 5, 3, 4, 4, 3

<u>วิธีทำ</u> จัดเรียงข้อมูลใหม่ได้ดังนี้ 1, 2, 2, 3, 3, 3, 4, 4, 5

เนื่องจากข้อมูลเป็นจำนวนคี่ ดังนั้น ตำแหน่งกึ่งกลาง คือ $\frac{N+1}{2} = \frac{9+1}{2} = 5$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ เท่ากับ 3

ตัวอย่าง 1.8 จงหามัธยฐานของข้อมูลที่กำหนดให้ 6.5, 12, 14.9, 10, 7.9, 21.9, 9.2, 14.5

<u>วิธีทำ</u> จัดเรียงลำดับข้อมูลใหม่ 6.5, 7.9, 9.2, 10, 12, 14.5, 14.9, 21.9

เนื่องจากข้อมูลเป็นจำนวนคู่ ดังนั้น ตำแหน่งกึ่งกลาง คือ $\frac{N}{2} = \frac{8}{2} = 4$ ซึ่งมีค่าเท่ากับตำแหน่งที่ 4 ตรงกับค่าคะแนน **10** และตำแหน่งกึ่งกลาง คือ $\frac{N}{2}+1 = \frac{8}{2}+1 = 5$ ซึ่งมีค่าเท่ากับตำแหน่งที่ 5 ตรงกับค่าคะแนน **12**

ดังนั้น มัธยฐานของข้อมูลชุดนี้ เท่ากับ $\frac{10+12}{2} = 11$

2) การหาค่ามัธยฐานกรณีข้อมูลที่แจกแจงความถี่

คำนวณหามัธยฐานได้จากสูตรต่อไปนี้

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

เมื่อ L_1 คือ ขอบเขตล่างของชั้นที่มัธยฐานอยู่

N คือ จำนวนข้อมูลทั้งหมด หรือความถี่ทั้งหมด

F คือ ผลรวมของความถี่ของทุกชั้นที่มีข้อมูลต่ำกว่าชั้นที่มีมัธยฐานอยู่

f_1 คือ ความถี่ของชั้นที่มีมัธยฐานอยู่

I คือ ความกว้างของชั้น

ตัวอย่าง 1.9 จงหามัธยฐานของข้อมูลที่กำหนดให้

คะแนน	ความถี่ (f)	ความถี่สะสม
20 - 24	4	4
25 - 29	6	10
30 - 34	12	22
35 - 39	16	38
40 - 44	12	50
45 - 49	7	57
50 - 54	3	60

<u>วิธีทำ</u> เนื่องจาก $\frac{N}{2} = \frac{60}{2} = 30$ ดังนั้น มัธยฐานอยู่ในชั้น 35 – 39

และได้ว่า $L_1 = 34.5,\ F = 22,\ f_1 = 16,\ I = 5$

จาก

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

$$= 34.5 + \left[\frac{\frac{60}{2} - 22}{16}\right](5)$$

$$= 37$$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ เท่ากับ 37

3. ฐานนิยม (Mode : Mo) คือ ค่าของข้อมูลที่ปรากฏบ่อยครั้งที่สุด หรือมีความถี่สูงสุด ทั้งนี้ข้อมูลชุดหนึ่งอาจไม่มีฐานนิยมเลยก็ได้ แต่ถ้ามีก็อาจมีได้มากกว่าหนึ่งค่า โดยมีวิธีการหาได้ 2 วิธี ดังนี้

1) การหาค่าฐานนิยมกรณีข้อมูลที่ไม่ได้แจกแจงความถี่

ฐานนิยม คือ ค่าของข้อมูลที่เกิดขึ้นบ่อยที่สุดหรือมีความถี่สูงสุด

ตัวอย่าง 1.10 จงหาฐานนิยมของข้อมูลต่อไปนี้

1. 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12 และ 18

   Mo = 9

2. 3, 5, 8, 10, 12, 15, 16

   ข้อมูลชุดนี้ไม่มี Mo เนื่องจากไม่มีข้อมูลค่าใดมีความถี่สูงสุด

3. 2, 3, 4, 4, 4, 5, 5, 7, 7, 7 และ 9

Mo = 4 และ 7

2) การหาค่าฐานนิยมกรณีข้อมูลที่แจกแจงความถี่

คำนวณหาฐานนิยมได้จากสูตรต่อไปนี้

$$Mo = L_1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] I$$

เมื่อ L_1 คือ ขอบเขตล่างของชั้นที่ฐานนิยมอยู่

Δ_1 คือ ผลต่างระหว่างความถี่ของชั้นที่มีฐานนิยมอยู่ กับชั้นติดกันที่มีข้อมูลต่ำกว่า

Δ_2 คือ ผลต่างระหว่างความถี่ของชั้นที่มีฐานนิยมอยู่ กับชั้นติดกันที่มีข้อมูลสูงกว่า

I คือ ความกว้างของชั้น

ตัวอย่าง 1.11 จงหาฐานนิยมของคะแนนนักศึกษา 150 คน

คะแนน	**ความถี่** (f)
20 - 24	4
25 - 29	6
30 - 34	12
35 - 39	16
40 - 44	12
45 - 49	7
50 - 54	3

<u>วิธีทำ</u> เนื่องจากความถี่สูงสุด = 16 ดังนั้น ฐานนิยมอยู่ในชั้น 35 – 39

และได้ว่า $L_1 = 34.5,\ \Delta_1 = 16 - 12 = 4,\ \Delta_2 = 16 - 12 = 4,\ I = 5$

จาก $Mo = L_1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] I$

จะได้

$$\begin{aligned} Mo &= 34.5 + \left[\frac{4}{4+4}\right](5) \\ &= 34.5 + \left[\frac{1}{2}\right](5) \\ &= 34.5 + 2.5 \\ &= 37 \end{aligned}$$

ดังนั้น ฐานนิยมของข้อมูลชุดนี้ เท่ากับ 37

การวัดการกระจายของข้อมูล

พิจารณาข้อมูล 3 ชุด ต่อไปนี้

ชุดที่ 1 : 10, 8, 9, 10, 11, 12

ชุดที่ 2 : 5, 6, 8, 10, 12, 14, 15

ชุดที่ 3 : 1, 2, 5, 10, 15, 18, 19

จะพบว่า ข้อมูลทั้ง 3 ชุด มี $\mu = 10$ เท่ากัน และอาจจะเข้าใจว่าข้อมูลทั้ง 3 ชุดมีลักษณะคล้ายคลึงกัน แต่จะเห็นว่าข้อมูลชุดที่ 1 มีค่าของข้อมูลแต่ละค่าใกล้เคียงกับ $\mu = 10$ มากที่สุด นั่นคือ จะได้ว่า ในข้อมูลชุดที่ 1 ค่าของข้อมูลแต่ละตัว **มีการกระจายห่างจากค่าเฉลี่ยน้อยที่สุด**

การวัดการกระจายมีอยู่หลายวิธี ดังต่อไปนี้

1. พิสัย (Range) คือ ผลต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูล

นั่นคือ ถ้ามีข้อมูล $X_1, X_2, \ldots, X_N$ รวม N ตัว แล้ว พิสัย $= X_{\max} - X_{\min}$

ตัวอย่าง 1.12 จงหาพิสัยของข้อมูลที่กำหนดให้ ดังต่อไปนี้ 2, 8, 6, 7, 10, 5, 6, 3

<u>วิธีทำ</u> จาก พิสัย = ค่าสูงสุด – ค่าต่ำสุด

= 10 – 2 = 8

นั่นคือ พิสัยของข้อมูลชุดนี้ คือ 8

2. พิสัยกึ่งควอไทล์ คือ ผลต่างระหว่างควอไทล์ที่ 3 และควอไทล์ที่ 1 นั่นคือ

$$Q.R. = Q_3 - Q_1$$

การหาพิสัยกึ่งควอไทล์ในกรณีที่ข้อมูลไม่ได้แจกแจงความถี่ โดยตำแหน่งของควอไทล์ที่ r หรือ Q_r คือ $\frac{r(N+1)}{4}$; $r = 1, 2, 3$

ตัวอย่าง 1.13 จงหาพิสัยกึ่งควอไทล์ของข้อมูลที่กำหนดให้ ต่อไปนี้

10, 16, 2, 19, 30, 15, 23, 13

<u>วิธีทำ</u> ก่อนอื่นเรียงลำดับข้อมูลจากน้อยไปหามาก ได้ดังนี้ 2, 10, 13, 15, 16, 19, 23, 30

ตำแหน่งของ Q_1 คือ $\frac{1(8+1)}{4} = 2.25$

ดังนั้น $Q_1 = 10 + (0.25)(13-10) = 10.75$

และตำแหน่งของ Q_3 คือ $\frac{3(8+1)}{4} = 6.75$

ได้ว่า $Q_3 = 19 + (0.75)(23-19) = 22$

ดังนั้น $Q.R. = Q_3 - Q_1 = 22 - 10.75 = 11.25$

3. ความแปรปรวน (Variance) คือ ค่าเฉลี่ยของกำลังสองของผลต่างระหว่างค่าของข้อมูลแต่ละค่ากับค่าเฉลี่ย หรือความแปรปรวน ก็คือ กำลังสองของส่วนเบี่ยงเบนมาตรฐานนั่นเอง โดยมีวิธีการหาได้ 2 วิธี ดังนี้

1) การหาค่าความแปรปรวนกรณีข้อมูลที่ไม่ได้แจกแจงความถี่

ค่าความแปรปรวนของประชากร (σ^2)

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N} \quad \text{หรือ} \quad \sigma^2 = \frac{\sum_{i=1}^{N}X_i^2 - \frac{\left[\sum_{i=1}^{N}X_i\right]^2}{N}}{N}$$

ค่าความแปรปรวนของตัวอย่าง (S^2)

$$S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} \quad \text{หรือ} \quad S^2 = \frac{\sum_{i=1}^{n}X_i^2 - \frac{\left[\sum_{i=1}^{n}X_i\right]^2}{n}}{n-1}$$

2) การหาค่าความแปรปรวนกรณีข้อมูลที่แจกแจงความถี่

ค่าความแปรปรวนของประชากร (σ^2)

$$\sigma^2 = \frac{\sum_{i=1}^{N}f_i(X_i-\mu)^2}{N} \quad \text{หรือ} \quad \sigma^2 = \frac{\sum_{i=1}^{N}f_iX_i^2 - \frac{\left[\sum_{i=1}^{N}f_iX_i\right]^2}{N}}{N}$$

ค่าความแปรปรวนของตัวอย่าง (S^2)

$$S^2 = \frac{\sum_{i=1}^{n}f_i(X_i-\bar{X})^2}{n-1} \quad \text{หรือ} \quad S^2 = \frac{\sum_{i=1}^{n}X_i^2 - \frac{\left[\sum_{i=1}^{n}X_i\right]^2}{n}}{n-1}$$

ตัวอย่าง 1.14 จงหาความแปรปรวนของข้อมูลตัวอย่างที่กำหนดให้ ต่อไปนี้

คะแนน	**ความถี่** (f)
20 - 24	4
25 - 29	6
30 - 34	12
35 - 39	16
40 - 44	12
45 - 49	7
50 - 54	3

<u>วิธีทำ</u> สร้างตารางวิเคราะห์ข้อมูลก่อน ดังนี้

คะแนน	**ค่ากึ่งกลาง** (x_i)	**ความถี่** (f)	**ผลรวมคะแนนในชั้น** (fx_i)	fx^2
20 - 24	22	4	88	1,936
25 - 29	27	6	162	4,374
30 - 34	32	12	384	12,288
35 - 39	37	16	592	21,904
40 - 44	42	12	504	21,168
45 - 49	47	7	329	15,463
50 - 54	52	3	156	8,112
รวม		**60**	**2,215**	**85,245**

แทนค่าในสูตร

$$S^2 = \frac{\sum_{i=1}^{n} f_i X_i{}^2 - \frac{\left[\sum_{i=1}^{n} f_i X_i\right]^2}{n}}{n-1}$$

จะได้

$$S^2 = \frac{85,245 - \frac{[2,215]^2}{60}}{60-1}$$
$$= \frac{3474.583}{59}$$
$$= 58.89$$

ดังนั้น ความแปรปรวนของข้อมูลชุดนี้ เท่ากับ 58.89

4. ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) คือ ค่ารากที่สองที่เป็นบวกของความแปรปรวน จะได้ว่า

ค่าเบี่ยงเบนมาตรฐานของประชากร (σ)

$$\sigma = \sqrt{\sigma^2}$$

ค่าเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง (S)

$$S = \sqrt{S^2}$$

ตัวอย่าง 1.15 จงหาค่าเบี่ยงเบนมาตรฐานของข้อมูลตัวอย่างที่กำหนดให้ต่อไปนี้

2, 3, 4, 4, 5, 5, 1, 3

<u>วิธีทำ</u> จากสูตร

$$S = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - \frac{\left[\sum_{i=1}^{n} X_i\right]^2}{n}}{n-1}}$$

คำนวณหาค่าต่าง ๆ ดังนี้

$$\sum_{i=1}^{n} X_i^2 = 2^2+3^2+4^2+4^2+5^2+5^2+1^2+3^2 = 105$$

$$\sum_{i=1}^{n} X_i = 2+3+4+4+5+5+1+3 = 27$$

แทนค่าในสูตรจะได้

$$S = \sqrt{\frac{\sum_{i=1}^{n} X_i^{\,2} - \frac{\left[\sum_{i=1}^{n} X_i\right]^2}{n}}{n-1}}$$

$$= \sqrt{\frac{105 - \frac{(27)^2}{8}}{8-1}}$$

$$= \sqrt{\frac{105 - 91.125}{7}}$$

$$= \sqrt{\frac{13.875}{7}}$$

$$= \sqrt{1.982}$$

$$= 1.407$$

ดังนั้น ค่าเบี่ยงเบนมาตรฐานของข้อมูล เท่ากับ 1.407

ตัวอย่าง 1.16 จงหาค่าเบี่ยงเบนมาตรฐานของข้อมูลตัวอย่างที่กำหนดให้ต่อไปนี้

คะแนน	**ความถี่** (f)
20 - 24	4
25 - 29	6
30 - 34	12
35 - 39	16
40 - 44	12
45 - 49	7
50 - 54	3

<u>วิธีทำ</u> สร้างตารางวิเคราะห์ข้อมูลก่อน ดังนี้

คะแนน	**ค่ากึ่งกลาง** (x_i)	**ความถี่** (f)	**ผลรวมคะแนนในชั้น** (fx_i)	fx^2
20 - 24	22	4	88	1,936
25 - 29	27	6	162	4,374
30 - 34	32	12	384	12,288
35 - 39	37	16	592	21,904
40 - 44	42	12	504	21,168
45 - 49	47	7	329	15,463
50 - 54	52	3	156	8,112
รวม		**60**	**2,215**	**85,245**

แทนค่าในสูตร

$$S = \sqrt{\frac{\sum_{i=1}^{n} f_i X_i^2 - \frac{\left[\sum_{i=1}^{n} f_i X_i\right]^2}{n}}{n-1}}$$

จะได้

$$S = \sqrt{\frac{85,245 - \frac{[2,215]^2}{60}}{60-1}}$$

$$= \sqrt{\frac{3474.583}{59}}$$

$$= \sqrt{58.89}$$

$$= 7.674$$

ดังนั้น ค่าเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้ เท่ากับ 7.674

แบบฝึกหัดท้ายบท

1. ให้นักศึกษาหาข้อมูลจากข่าวสารต่าง ๆ หรือสมมติข้อมูลขึ้น แล้วเลือกวิธีในการนำเสนอข้อมูลมา 2 วิธี

2. ผลการสอบวิชาหลักสถิติของนักศึกษาชั้นปีที่ 1 ของมหาวิทยาลัยแห่งหนึ่ง จำนวน 30 คน ได้คะแนนสอบ ดังนี้

50	49	27	46	48	26	30	47	24	17	31
42		45	33	44	29	35	43	20	37	22
19	38	30	18	39	30	40	21	25		

จงสร้างตารางแจกแจงความถี่ (หาขีดจำกัดชั้น ขอบเขต ความถี่ ความถี่สะสม ความถี่สัมพัทธ์ และจุดกึ่งกลางชั้น)

3. กำหนดข้อมูลต่อไปนี้ 6, 25, 9, 18, 29, 7, 31, 9, 23, 10, 30, 28, 8, 15 จงหาพิสัยกึ่งควอร์ไทล์ของข้อมูล

4. จงหาค่าเฉลี่ยและมัธยฐานของตัวอย่าง ต่อไปนี้

5.5	6.7	7.3	7.7	6.8	6.9	7.2	6.4	8.1	7.5

5. จงหามัธยฐาน พิสัย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐาน ของข้อมูลประชากร ดังต่อไปนี้

2	4	3	1	5	4	7	6	5	6	3
5	5	4	7	3	4	3	8	6	5	5
9	2									

6. จากการสำรวจอายุการเข้าทำงานของพนักงานในบริษัทแห่งหนึ่ง ดังนี้

อายุของพนักงาน	จำนวน
15 - 17	4
18 - 20	7
21 - 23	9
24 - 26	12
27 - 29	10
30 - 32	5
33 - 35	3

6.1 จงหาความถี่สะสมและความถี่สัมพัทธ์

6.2 จงหาค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน และมัธยฐาน

7. จงหาค่าเฉลี่ยและมัธยฐานของตัวอย่าง ต่อไปนี้

5.5	6.7	7.3	7.7	6.8	6.9	7.2	6.4	8.1	7.5