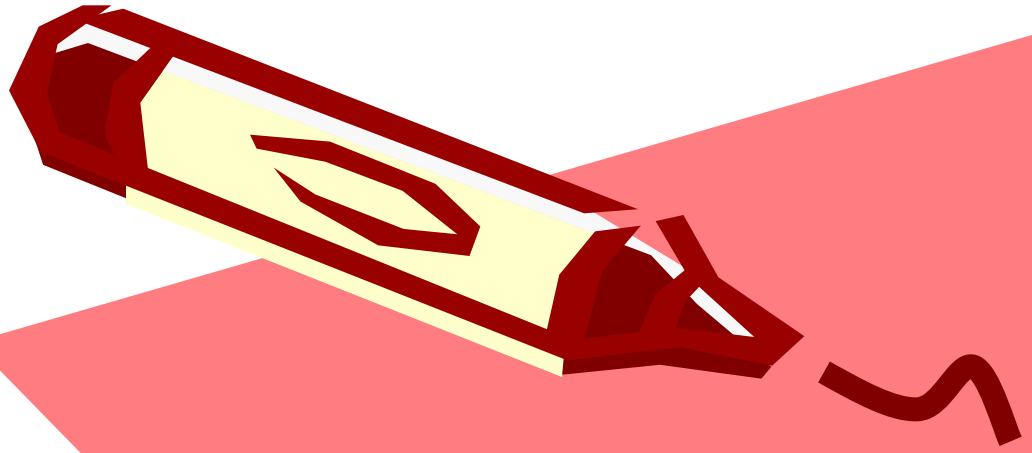


บทที่ 1
ความรู้เบื้องต้นเกี่ยวกับสถิติ

4113107 สถิติธุรกิจ
อาจารย์รัชนีกร ทบประดิษฐ์

ความหมายสถิติ

คำว่า สถิติ (Statistics) จำแนกได้ 2 ความหมาย คือ

สถิติ หมายถึง ข้อมูลหรือตัวเลขที่แทนข้อเท็จจริงเกี่ยวกับเรื่องต่างๆที่เราสนใจหรือที่อยู่รอบตัวเรา

สถิติ หมายถึง ศาสตร์หรือวิชาที่ว่าด้วยหลักการและวิธีการทางสถิติ ซึ่งประกอบด้วย การเก็บรวบรวมข้อมูล การจัดระเบียบข้อมูล หรือการนำเสนอข้อมูล การวิเคราะห์และการแปลความหมายข้อมูล

ตามสถิติในแง่ที่เป็นศาสตร์ สถิติแบ่งออกเป็น 2 ประเภทคือ

1) สถิติเชิงพรรณนา (descriptive statistics)
2) สถิติเชิงอนุมาน (inferential statistics)

ข้อมูลและระดับการวัดข้อมูล

ข้อมูล (data) หมายถึง ข้อเท็จจริงต่างๆ ที่เราสนใจจะศึกษา ซึ่งอาจเป็นตัวเลขหรือมิใช่ตัวเลขก็ได้ โดยแบ่งได้เป็น 2 ประเภท คือ

- ข้อมูลเชิงปริมาณ คือ ค่าที่วัดได้ออกมาเป็นตัวเลข
- ข้อมูลเชิงคุณภาพ คือ ค่าที่วัดได้ไม่สามารถวัดออกมาเป็นตัวเลขได้

ระดับการวัดข้อมูล (levels of measurement)

ระดับการวัดข้อมูลแบ่งออกเป็น 4 ระดับหรือมาตรา ดังนี้

- มาตรานามบัญญัติ
- มาตราเรียงลำดับ
- มาตราอันตรภาค
- มาตราอัตราส่วน

ระดับการวัดข้อมูล

มาตรานามบัญญัติ (nominal scale)

เป็นมาตราการวัดที่ใช้กับข้อมูลที่มีลักษณะหยาบหรือต่ำสุด อาจเป็นการกำหนดสัญลักษณ์หรือตัวเลขเพื่อจำแนกประเภทสิ่งของหรือคุณลักษณะต่างๆ เท่านั้น ไม่สามารถแสดงให้เห็นปริมาณมากน้อยหรือสูงต่ำแต่อย่างใด ดังนั้น จึงไม่สามารถนำตัวเลขเหล่านั้นมาบวก ลบ คูณ หารได้ เช่น เพศ คือ เพศชายและหญิง

มาตราเรียงอันดับ (ordinal scale)

เป็นมาตราการวัดที่มีความละเอียดการวัดเพิ่มขึ้นหรือสูงกว่ามาตรานามบัญญัติ เพราะสามารถบอกลำดับและความแตกต่าง แต่ไม่สามารถบอกได้ว่าคุณลักษณะหรือคุณสมบัติเหล่านี้มีปริมาณมากน้อยกว่ากันเท่าใด กล่าวอีกนัยหนึ่ง ข้อมูลในระดับนี้ไม่สามารถนำมาคำนวณทางคณิตศาสตร์ได้เช่นเดียวกับมาตรานามบัญญัติ เช่น การประกวดนางสาวไทย

อันดับที่ 1 คือ นางสาวไทย

อันดับที่ 2 คือ รองนางสาวไทยคนที่ 1

และอันดับที่ 3 คือ รองนางสาวไทยคนที่ 2

ระดับการวัดข้อมูล

มาตราอันตรภาค (interval scale)

เป็นมาตราการวัดที่สามารถทราบได้ว่าสิ่งที่จะวัดมีช่วงมากน้อยเท่าใด โดยแต่ละช่วงมาตรานี้มีค่าเท่าๆ กัน ทำให้เราทราบถึงความแตกต่างที่ห่างกันเป็นช่วงได้ และค่าที่ได้จากการวัดสามารถนำมาคำนวณทางคณิตศาสตร์ได้ เช่น เราจะบอกความแตกต่างของน้ำร้อนระหว่าง 60° C กับ 80° C เท่ากับความแตกต่างระหว่าง 100° C กับ 120° C โดยดูจากช่วงที่ห่างกันเท่ากับ20° C

มาตราอัตราส่วน (ratio scale)

เป็นมาตราการวัดที่ดีที่สุดและวัดได้อย่างละเอียดที่สุด ตัวเลขที่วัดได้สามารถสื่อความหมายตรงตามค่าของสิ่งที่วัด และเป็นมาตราวัดที่ข้อมูลมีค่าเป็นศูนย์แท้ คือ ถ้าค่าตัวเลขที่วัดได้มีค่าเป็นศูนย์ ก็แปลว่า สิ่งที่วัดนั้นมีค่าศูนย์ด้วย ข้อมูลที่อยู่ในมาตราวัดระดับนี้ ได้แก่ เวลา อายุ น้ำหนัก ส่วนสูง ระยะทาง เป็นต้น ข้อมูลที่วัดได้สามารถนำมาทำการคำนวณทางคณิตศาสตร์ได้

แหล่งของข้อมูล

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

แหล่งปฐมภูมิ (Primary source) เป็นแหล่งของข้อมูลที่เป็นต้นกำเนิดของข้อมูลนั้นๆ เราเรียกข้อมูลนี้ว่า ข้อมูลปฐมภูมิ เช่น ข้อมูลที่ได้จากการสัมภาษณ์ ข้อมูลที่ได้จากการทดลอง ข้อมูลที่ได้จากการกรอกแบบสอบถาม เป็นต้น

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

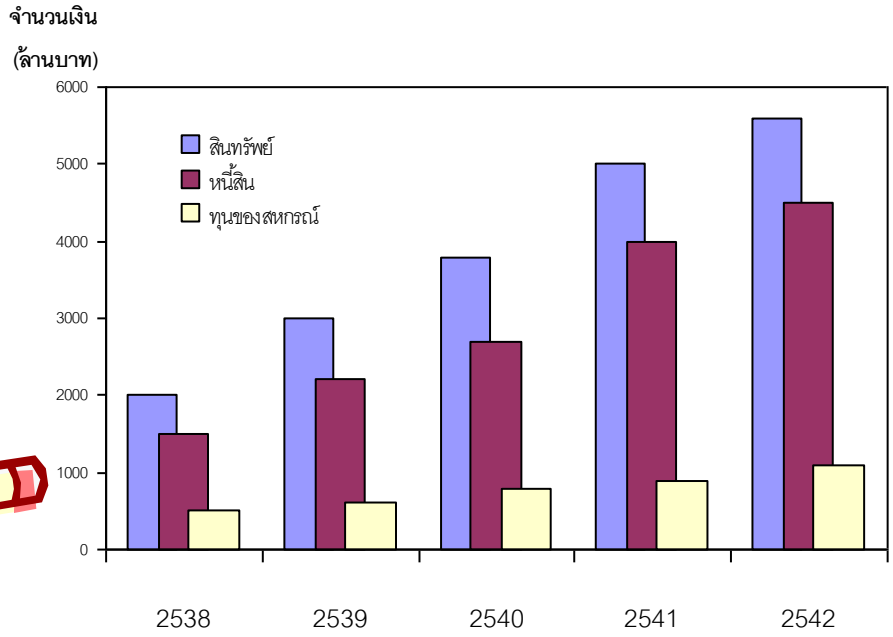
2. การเก็บรวบรวมข้อมูลโดยการสำรวจ เป็นการเก็บรวบรวมข้อมูลจากหน่วยที่ศึกษาโดยตรง เช่น การสำรวจความคิดเห็นของประชาชนในการร่างรัฐธรรมนูญ ซึ่งหน่วยที่ศึกษา คือ ประชาชนคนไทย การเก็บรวบรวมข้อมูลโดยการสำรวจจะแบ่งออกเป็น 2 ประเภท

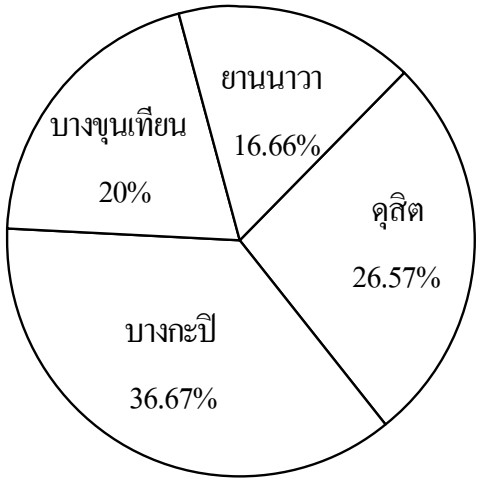
- **การสำมะโน** เก็บรวบรวมข้อมูลจากทุกๆหน่วยศึกษาของประชากร
- **การสำรวจตัวอย่าง** เก็บรวบรวมข้อมูลจากบางหน่วยศึกษาของประชากร

3. การเก็บรวบรวมข้อมูลจากการทดลอง ข้อมูลบางประเภทไม่สามารถหาได้จากการสำรวจ แต่จัดทำได้จากการทดลอง เช่น การศึกษาวิธีการปลูกพืชที่แตกต่างกัน 3 วิธี

การนำเสนอข้อมูล

- การนำเสนอด้วยบทความ
- การนำเสนอด้วยตาราง
- การนำเสนอด้วยกราฟ
- การนำเสนอด้วยแผนภูมิ อาจเป็นแผนภูมิแท่ง, วงกลม หรือรูปภาพ





การแจกแจงความถี่ด้วยตาราง

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

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

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

การแจกแจงความถี่ด้วยตาราง

ตัวอย่าง 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

วิธีทำ 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

การแจกแจงความถี่ด้วยตาราง

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

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

ขีดจำกัดล่าง

ขีดจำกัดบน

ค่ากึ่งกลางชั้น คือ 14

ขอบเขตบน คือ 15.5

ขอบเขตล่าง คือ 12.5

ผลต่างระหว่างขอบเขตล่างและบนของแต่ละชั้น คือ ความกว้างของชั้น

การแจกแจงความถี่ด้วยตาราง

วิธีการสร้างตารางแจกแจงความถี่

ตัวอย่าง 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

วิธีทำ 1. กำหนดจำนวนชั้น ในที่นี้ต้องการตารางแจกแจงความถี่ที่มี 9 ชั้น

2. คำนวณค่าพิสัย (Range) ของข้อมูล เมื่อพิสัย คือ ผลต่างระหว่างค่าสูงสุดและค่าต่ำสุด

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

3. คำนวณความกว้างของแต่ละชั้น โดยความกว้างของแต่ละชั้น คือ ผลหารระหว่างพิสัยและ

จำนวนชั้น คือ $82/9 = 9.11$ เพื่อความสะดวกจะใช้ความกว้างของแต่ละชั้นเป็น 10

การแจกแจงความถี่ด้วยตาราง

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

5. เขียนรอยขีด

6. คำนวณความถี่ โดยการนับรอยขีดในชั้นนั้นๆ ซึ่งผลรวมของความถี่ทั้งหมดทุกชั้นต้องเท่ากับจำนวนข้อมูล 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

ค่าเฉลี่ย (Mean : $\bar{X}$) เป็นค่ากลางที่คำนวณได้จากข้อมูลทุกค่าที่มีในชุดข้อมูล โดยวิธีการหาได้ 2 วิธี ดังนี้

การหาค่าเฉลี่ยในกรณีที่ข้อมูลไม่ได้แจกแจงความถี่

ค่าเฉลี่ย = ผลรวมทั้งหมดหารด้วยจำนวนข้อมูล

หรือ $$\bar{x} = \frac{1}{N}\sum x_i = \frac{\sum x}{N}$$

ตัวอย่าง 1 จงหาค่าเฉลี่ยของข้อมูลที่กำหนดให้ต่อไปนี้ 22 23 24 24 25 25 21 23 15 20

วิธีทำ $$\bar{x} = \frac{1}{N}\sum x_i = \frac{22+23+24+24+25+25+21+23+15+20}{10}$$

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

นั่นคือ ค่าเฉลี่ยของข้อมูลนี้เท่ากับ 22.2 คะแนน

การวัดแนวโน้มเข้าสู่ส่วนกลาง

การหาค่าเฉลี่ยในกรณีที่ข้อมูลแจกแจงความถี่แล้ว

ในการหาค่าเฉลี่ยของข้อมูลที่แจกแจงความถี่แล้ว ซึ่งไม่ทราบค่าที่แท้จริงของข้อมูลในแต่ละชั้น ดังนั้นถ้าให้ค่ากึ่งกลางของแต่ละชั้นเป็นตัวแทนของข้อมูลในชั้นนั้นๆ แล้ว

$$\bar{x} = \frac{1}{\sum f_i} \sum f_i x_i = \frac{\sum fx}{\sum f}$$

เมื่อกำหนดให้ x_i คือ ค่ากึ่งกลางของชั้นคะแนนที่ i ในชุดข้อมูล

f_i คือ ความถี่ของชั้นที่ i

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ตัวอย่าง 2 จงหาค่าเฉลี่ยของข้อมูลที่กำหนดให้ต่อไปนี้

ขีดจำกัดคะแนน	ค่ากึ่งกลาง x_i	ความถี่ (f)	ผลรวมคะแนนในชั้น fx_i
19.5 – 24.5	22	4	88
24.5 – 29.5	27	6	162
29.5 – 34.5	32	12	384
34.5 – 39.5	37	16	592
39.5 – 44.5	42	12	504
44.5 – 49.5	47	7	329
49.5 – 54.5	52	3	156
รวม		60	2,215

วิธีทำ คำนวณค่าเฉลี่ยได้

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{2,215}{60} = 36.92$$

นั่นคือ ค่าเฉลี่ยของข้อมูลชุดนี้ คือ 36.92

การวัดแนวโน้มเข้าสู่ส่วนกลาง

มัธยฐาน (Median : Me) คือ ค่าที่บอกตำแหน่งกึ่งกลางของข้อมูลทั้งหมดเมื่อได้จัดเรียงข้อมูลจากค่าน้อยไปหาค่ามาก หรือจากค่ามากไปหาค่าน้อย โดยวิธีการหาได้ 2 วิธี ดังนี้

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่

- ถ้าข้อมูลเป็นจำนวนคี่ มัธยฐาน คือ ค่าของข้อมูลตัวที่ $\frac{N+1}{2}$

- ถ้าข้อมูลเป็นจำนวนคู่ คำนวณมัธยฐานโดยการเปรียบเทียบระหว่างข้อมูลตัวที่ $\frac{N}{2}$ และ $\frac{N}{2}+1$ การหามัธยฐานให้นำเอาคะแนนที่อยู่ตรงตำแหน่งกึ่งกลางทั้งสองค่านั้นมาบวกกันหารสอง

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

วิธีทำ 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 คือ 3

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ตัวอย่าง 4 จงหามัธยฐานของข้อมูลที่กำหนดให้ต่อไปนี้

6.5, 12, 14.9, 10, 7.9, 21.9, 9.2, 14.5

วิธีทำ 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

ดังนั้นมัธยฐานของข้อมูลชุดนี้ คือ $\frac{10+12}{2} = 11$

การวัดแนวโน้มเข้าสู่ส่วนกลาง

กรณีข้อมูลที่แจกแจงความถี่แล้ว

คำนวณหามัธยฐานได้จากสูตร ต่อไปนี้

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

เมื่อ L_1 คือ ขอบเขตล่างของชั้นที่มัธยฐานอยู่

N คือ จำนวนข้อมูลทั้งหมด หรือความถี่ทั้งหมด

F คือ ผลรวมของความถี่ของทุกชั้นที่มีข้อมูลต่ำกว่าชั้นที่มีมัธยฐานอยู่

f_1 คือ ความถี่ของชั้นที่มีมัธยฐานอยู่

I คือ ความกว้างของชั้น

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ตัวอย่าง 5 จงหามัธยฐานของข้อมูลที่กำหนดให้ต่อไปนี้

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 – 29.5	6	10
29.5 – 34.5	12	22
34.5 – 39.5	16	38
39.5 – 44.5	12	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

การวัดแนวโน้มเข้าสู่ส่วนกลาง

วิธีทำ เนื่องจาก $\frac{N}{2} = \frac{60}{2} = 30$ ดังนั้นมัธยฐานอยู่ในชั้น 34.5 – 39.5

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

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

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

$$= 37$$

นั่นคือ มัธยฐานของข้อมูลชุดนี้ คือ 37

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ฐานนิยม (Mode : Mo) คือ ค่าของข้อมูลที่ปรากฏย่อยครั้งที่สุดหรือมีความถี่สูงสุด ทั้งนี้ข้อมูลชุดหนึ่งอาจไม่มีฐานนิยมเลยก็ได้ แต่ถ้ามีก็อาจมีได้มากกว่าหนึ่งค่า โดยวิธีการหาได้ 2 วิธี ดังนี้

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่

ฐานนิยม คือ ค่าของข้อมูลที่มีความถี่สูงสุด

ตัวอย่าง 6 จงหาฐานนิยมของข้อมูลต่อไปนี้

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

Mo = 9

ข. 3, 5, 8, 10, 12, 15, 16

ข้อมูลชุดนี้ไม่มี Mo เนื่องจากไม่มีข้อมูลค่าใดมีความถี่สูงสุด

ค. 2, 3, 4, 4, 4, 5, 5, 7, 7, 7 และ 9

Mo = 4 และ 7

การวัดแนวโน้มเข้าสู่ส่วนกลาง

กรณีข้อมูลที่แจกแจงความถี่แล้ว

คำนวณหาฐานนิยมได้จากสูตร ต่อไปนี้

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

เมื่อ L_1 คือ ขอบเขตล่างของชั้นที่ฐานนิยมอยู่

Δ_1 คือ ผลต่างระหว่างความถี่ของชั้นที่มีฐานนิยมอยู่ กับชั้นติดกันที่มีข้อมูลต่ำกว่า

Δ_2 คือ ผลต่างระหว่างความถี่ของชั้นที่มีฐานนิยมอยู่ กับชั้นติดกันที่มีข้อมูลสูงกว่า

I คือ ความกว้างของชั้น

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ตัวอย่าง 7 จงหาฐานนิยมของคะแนนนักศึกษา 150 คน

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 – 29.5	6	10
29.5 – 34.5	12	22
34.5 – 39.5	16	38
39.5 – 44.5	12	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

วิธีทำ เนื่องจากความถี่สูงสุด = 16 ดังนั้นฐานนิยมอยู่ในชั้น 34.5 – 39.5

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

จะได้ว่า $$Mo = L_1 + \left[\frac{\Delta_1}{\Delta_1+\Delta_2}\right]I = 34.5 + \left[\frac{4}{4+4}\right](5) = 37$$

นั่นคือ ฐานนิยมของข้อมูลชุดนี้ คือ 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 ชุด มีลักษณะคล้ายคลึงกัน แต่จะเห็นว่าข้อมูลชุดที่ 1 มีค่าของข้อมูลแต่ละค่าใกล้เคียงกับ $\mu = 10$ มากที่สุด นั่นคือ จะได้ว่า ในข้อมูลชุดที่ 1 ค่าของข้อมูลแต่ละตัว **มีการกระจายห่างจากค่าเฉลี่ยน้อยที่สุด**

การวัดการกระจายของข้อมูล

การวัดการกระจายมีอยู่หลายวิธี ดังต่อไปนี้

พิสัย (Range) คือ ผลต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูล

นั่นคือ ถ้ามีข้อมูล $X_1, X_2, \ldots, X_N$ รวม N ตัว แล้ว **พิสัย = ค่าสูงสุด - ค่าต่ำสุด**

ตัวอย่าง 8 จงหาพิสัยของข้อมูลที่กำหนดให้ดังต่อไปนี้ 2, 8, 6, 7, 10, 5, 6, 3

วิธีทำ จากพิสัย = ค่าสูงสุด – ค่าต่ำสุด

= 10 – 2 = 8

นั่นคือ พิสัยของข้อมูลชุดนี้ คือ 8

การวัดการกระจายของข้อมูล

พิสัยกึ่งควอไทล์ คือ ผลต่างระหว่างควอไทล์ที่ 3 และควอไทล์ที่ 1 นั่นคือ

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

การหาพิสัยกึ่งควอไทล์ในกรณีที่ข้อมูลไม่ได้แจกแจงความถี่

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

ตัวอย่าง 9 จงหาพิสัยกึ่งควอไทล์ของข้อมูลที่กำหนดให้ ต่อไปนี้

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

วิธีทำ ก่อนอื่นเรียงลำดับข้อมูลจากน้อยไปหามาก ได้ดังนี้

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

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

ดังนั้น

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

การวัดการกระจายของข้อมูล

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

จะได้ว่า

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

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

นั่นคือ พิสัยกึ่งควอไทล์ของข้อมูลเท่ากับ 11.25

การวัดการกระจายของข้อมูล

ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation ; S.D.) คือ ค่ารากที่สองที่เป็นบวกของค่าเฉลี่ยของกำลังสองของผลต่างของคะแนนใดๆ กับค่าเฉลี่ยของข้อมูลนั้น นั่นคือ

$$S.D. = \sqrt{\frac{\sum(x-\bar{x})^2}{N}}$$

ในกรณีที่ข้อมูลแจกแจงความถี่แล้ว

ส่วนเบี่ยงเบนมาตรฐาน หาได้จากสูตรดังนี้

$$S.D. = \sqrt{\left(\frac{\sum fx^2}{\sum f}\right)-\left(\frac{\sum fx}{\sum f}\right)^2}$$

การวัดการกระจายของข้อมูล

ตัวอย่าง 10 จงหา S.D. ของข้อมูลที่กำหนดให้ ต่อไปนี้ 2, 3, 4, 4, 5, 5, 1, 3

วิธีทำ 1) คำนวณหา $\overline{X}$ ของข้อมูลได้ดังนี้

$$\overline{x} = \frac{2+3+4+4+5+5+1+3}{8} = 3.38$$

2) หากำลังสองของผลต่างของ x และ $\overline{X}$ หรือหา $(x-\overline{x})^2$ ได้ดังนี้

$(2-3.38)^2 = (-1.38)^2 = 1.90$ $(3-3.38)^2 = (-0.38)^2 = 0.14$

$(4-3.38)^2 = (0.62)^2 = 0.38$ $(4-3.38)^2 = (0.62)^2 = 0.38$

$(5-3.38)^2 = (1.62)^2 = 2.62$ $(5-3.38)^2 = (1.62)^2 = 2.62$

$(1-3.38)^2 = (-2.38)^2 = 5.66$ $(3-3.38)^2 = (-0.38)^2 = 0.14$

3) คำนวณค่า S.D. ดังนี้

$$S.D. = \sqrt{\frac{1.90+0.14+0.38+0.38+2.62+2.62+5.66+0.14}{8}}$$

$$= \sqrt{\frac{13.84}{8}} = \sqrt{1.73} = \pm 1.315$$

การวัดการกระจายของข้อมูล

ตัวอย่าง 11 จงหา S.D. ของข้อมูลที่กำหนดให้ ต่อไปนี้

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 – 29.5	6	10
29.5 – 34.5	12	22
34.5 – 39.5	16	38
39.5 – 44.5	12	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

การวัดการกระจายของข้อมูล

วิธีทำ สร้างตารางวิเคราะห์ข้อมูลก่อน ดังนี้

ขีดจำกัดคะแนน	ค่ากึ่งกลาง x_i	ความถี่ (f)	ผลรวมคะแนนในชั้น fx_i	fx^2
19.5 – 24.5	22	4	88	1936
24.5 – 29.5	27	6	162	4374
29.5 – 34.5	32	12	384	12,288
34.5 – 39.5	37	16	592	21,904
39.5 – 44.5	42	12	504	21,168
44.5 – 49.5	47	7	329	15,463
49.5 – 54.5	52	3	156	8,112
รวม		60	2,215	85,245

แทนค่าในสูตร จะได้

$$S.D. = \sqrt{\left(\frac{\sum fx^2}{\sum f}\right) - \left(\frac{\sum fx}{\sum f}\right)^2} = \sqrt{\left(\frac{85,245}{60}\right) - \left(\frac{2,215}{60}\right)^2} = \sqrt{57.91}$$

ดังนั้น ค่าเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้ คือ ± 7.61

การวัดการกระจายของข้อมูล

ความแปรปรวน (Variance : s^2) ความแปรปรวนเป็นกำลังสองของส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดเดียวกัน นั่นคือ

$$s^2 = \frac{\sum(x-\bar{x})^2}{N}$$

ในกรณีที่ข้อมูลแจกแจงความถี่แล้ว

ความแปรปรวน หาได้จากสูตร

$$s^2 = \left(\frac{\sum fx^2}{\sum f}\right) - \left(\frac{\sum fx}{\sum f}\right)^2$$

การวัดการกระจายของข้อมูล

ตัวอย่าง 12 จงหาความแปรปรวนของข้อมูลที่กำหนดให้ ต่อไปนี้

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 – 29.5	6	10
29.5 – 34.5	12	22
34.5 – 39.5	16	38
39.5 – 44.5	12	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

การวัดการกระจายของข้อมูล

วิธีทำ สร้างตารางวิเคราะห์ข้อมูลก่อน ดังนี้

ขีดจำกัดคะแนน	ค่ากึ่งกลาง x_i	ความถี่ (f)	ผลรวมคะแนนในชั้น fx_i	fx^2
19.5 – 24.5	22	4	88	1936
24.5 – 29.5	27	6	162	4374
29.5 – 34.5	32	12	384	12,288
34.5 – 39.5	37	16	592	21,904
39.5 – 44.5	42	12	504	21,168
44.5 – 49.5	47	7	329	15,463
49.5 – 54.5	52	3	156	8,112
รวม		60	2,215	85,245

แทนค่าในสูตร จะได้

$$S^2 = \left(\frac{\sum fx^2}{\sum f}\right) - \left(\frac{\sum fx}{\sum f}\right)^2 = \left(\frac{85,245}{60}\right) - \left(\frac{2,215}{60}\right)^2 = 57.91$$

ดังนั้น ความแปรปรวนของข้อมูลชุดนี้ คือ 57.91

แบบฝึกหัดบทที่ 1

1. ผลการสอบวิชาวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนแห่งหนึ่งจำนวน 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 (หาขีดจำกัด, ขอบเขต, จำนวนข้อมูล, ความถี่, ความถี่สะสม, ความถี่สัมพัทธ์, จุดกึ่งกลางชั้น)

2. จงหาพิสัย มัธยฐาน ความแปรปรวน และค่าเบี่ยงเบนมาตรฐานประชากรของข้อมูล ต่อไปนี้

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

แบบฝึกหัดบทที่ 1

3. 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

3.1 จงหาความถี่สะสมและความถี่สัมพัทธ์

3.2 จงหาค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน และมัธยฐาน

4. จงหาค่าเฉลี่ยและมัธยฐานของตัวอย่าง ต่อไปนี้

5.5 6.7 7.3 7.7 6.8 6.9 7.2

6.4 8.1 7.5