"จากตัวเลขสู่ความหมาย: สถิติพรรณนาช่วยให้เข้าใจ ข้อมูลเชิงปริมาณได้ลึกซึ้งยิ่งขึ้น"

การวิเคราะห์ด้วยการใช้วิธีการคำนวณเชิงปริมาณจะมีคำสำคัญอยู่ 2 คำ คือ คำว่า “พารามิเตอร์” และคำว่า “สถิติ” โดยความหมายของคำว่า “พารามิเตอร์” (Parameter) เป็นการวิเคราะห์ด้วยการคำนวณจากข้อมูลหรือกลุ่มประชากรทั้งหมดที่สนใจ ในขณะที่คำว่า “สถิติ” (Statistics) เป็นการวิเคราะห์ด้วยการคำนวณจากกลุ่มตัวอย่าง (Bluman, 2008: 104) ซึ่งการคำนวณเชิงปริมาณทั้ง2 ลักษณะมีข้อดีและข้อเสียแตกต่างกันดังได้อธิบายไปแล้วในบทที่ 1 ดังนั้นจะมิได้ขยายความอย่างไรก็ตามในโลกปัจจุบันการวิเคราะห์ข้อมูลด้วยประชากรทั้งหมดที่สนใจไม่สามารถทำได้เนื่องจากใช้เวลาและงบประมาณสูง ดังนั้นในการพิจารณาถึงความน่าจะเป็นจึงนิยมใช้ “สถิติพรรณนาที่คำนวณเชิงปริมาณ” (Descriptive statistics for quantitative variables) โดยการนำกลุ่มตัวอย่างข้อมูลที่ดีมาพิจารณาแล้วนำข้อมูลดังกล่าวมาอธิบายต่อกลุ่มประชากรทั้งหมดว่ามีความน่าจะเป็นอย่างไรฉะนั้นการศึกษาสถิติพรรณนาเพื่อคำนวณเชิงปริมาณจึงมีความสำคัญมากและเพื่อให้เห็นลักษณะสำคัญของการคำนวณค่าสถิติจึงกำหนดการศึกษา คือ 1.การวัดแนวโน้มเข้าสู่ส่วนกลาง 2.การวัดการกระจายของข้อมูล 3.การวิเคราะห์ด้วยการใช้ Box plot

1.การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measures of central tendency or location)

การพิจารณาถึงค่ากลางของข้อมูลเชิงปริมาณสามารถคำนวณเพื่อหาค่ากลางต่อการวัดแนวโน้มเข้าสู่ส่วนกลางของการแจกแจงข้อมูลสามารถคำนวณได้หลายวิธี รวมทั้งข้อมูลแบบไม่จัดกลุ่มและข้อมูลแบบจัดกลุ่ม คือ (Welkowitz, Joan, Cohen, Barry H., and Lea, R. Brooke, 2011: 58-60; Panik, Michael J, 2012: 23-26)

1. ค่าเฉลี่ย (Mean) เมื่อมีการวัดค่ากลางของการแจกแจงด้วยการหาค่าเฉลี่ย ส่วนใหญ่จะนึกถึงและอ้างอิงไปยัง “ค่าเฉลี่ยเลขคณิต” (Arithmetic mean) แต่ในการการวัดค่ากลางยังมีอีกหลายวิธีในการหาค่าเฉลี่ย ดังนั้นเพื่อให้เห็นวิธีอื่นๆ ด้วยจึงจะอธิบายค่าเฉลี่ยในการวัดค่ากลางของการแจกแจงที่นี้ 3 วิธีคือ 1.ค่าเฉลี่ยเลขคณิต 2.ค่าเฉลี่ยเรขาคณิตและ 3.ค่าเฉลี่ยฮาร์โมนิค ดังนี้

1.1 ค่าเฉลี่ยเลขคณิต (Arithmetic mean)การคำนวณลักษณะนี้นิยมใช้เป็นส่วนใหญ่ (Foster, C., 2014: 32-33) ซึ่งสามารถวิเคราะห์ได้ 4 วิธี คือ

วิธีที่ 1 การหาค่าเฉลี่ยเลขคณิตสำหรับข้อมูลที่ไม่ได้จัดกลุ่ม

ก. ค่าเฉลี่ยเลขคณิตประชากร

สูตร $$\mu = \frac{\sum x_i}{N}$$

μ = ค่าเฉลี่ยเลขคณิตประชากร

$\sum x_i$ = ผลรวมของค่า $X_1, X_2, X_3, X_4 \ldots X_N$

N = จำนวนประชากรทั้งหมด

ข. ค่าเฉลี่ยเลขคณิตตัวอย่าง

สูตร $$\bar{x} = \frac{\sum x_i}{N} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

$\bar{x}$ = ค่าเฉลี่ยเลขคณิตตัวอย่าง

$\sum x_i$ = ผลรวมของค่า $X_1, X_2, X_3, X_4 \ldots X_n$

n = จำนวนตัวอย่าง

ตัวอย่าง 3.1 สุ่มช่วงปีของข้อมูลการซื้อขายหลักทรัพย์ของประเทศไทยในรอบ 10 ปี ตั้งแต่ปี พ.ศ.2542-2551 มีอัตราการซื้อขายดังนี้ (หน่วยเป็นพันล้านบาท)

ตารางที่ 3.1 ข้อมูลการซื้อขายหลักทรัพย์ของประเทศไทยในรอบ 10 ปี

ปี	2542	2543	2544	2545	2546	2547	2548	2549	2550	2551
มูลค่าการซื้อขายหลักทรัพย์	1,609.8	923.7	1,577.8	2,047.4	4,670.3	5,024.4	4,031.2	3,956.3	4,188.8	3,919.9

ที่มา: สำนักงานสถิติแห่งชาติ, 2552: http://service.nso.go.th/nso/nsopublish/download/files/socioSocIndicator.pdf. Web.19 สิงหาคม 2560.

$$\bar{x} = \frac{\sum x_i}{N} = \frac{1{,}609.8 + 923.7 + 1{,}577.8 + 2{,}047.4 + 4{,}670.3 + 5{,}024.4 + 4{,}031.2 + 3{,}956.3 + 4{,}188.8 + 3{,}919.9}{10}$$

= 3,194.9 พันล้านบาท

ดังนั้น ข้อมูลการซื้อขายหลักทรัพย์ของประเทศไทยในรอบ 10 ปี ตั้งแต่ปี พ.ศ.2542-2551 มีค่าเฉลี่ยเลขคณิต 3,194.9 พันล้านบาท

ตัวอย่าง 3.2 สุ่มช่วงปีของอัตราการแลกเปลี่ยนเงินตราต่างประเทศ สกุลเงินประเทศอังกฤษ (ปอนด์) เทียบกับเงินไทย ตั้งแต่ปี พ.ศ.2543-2552 (1 ปอนด์ต่อบาท)

ตารางที่ 3.2 อัตราการแลกเปลี่ยนเงินตราต่างประเทศระหว่างประเทศไทยกับประเทศอังกฤษ

ปี	2543	2544	2545	2546	2547	2548	2549	2550	2551	2552
อัตราการแลกเปลี่ยนเงินปอนด์ต่อจำนวนเงินไทย	60.7	64.0	64.5	67.8	73.7	73.1	69.7	69.1	61.6	53.6

ที่มา: สำนักงานสถิติแห่งชาติ, 2553: http://service.nso.go.th/nso/nsopublish/download /files/socioSoc53.pdf. Web.19สิงหาคม 2560.

$$\bar{x} = \frac{\sum x_i}{N} = \frac{60.7 + 64.0 + 64.5 + 67.8 + 73.7 + 73.1 + 69.7 + 69.1 + 61.6 + 53.6}{10}$$

= 65.8บาทต่อปอนด์

ดังนั้น อัตราการแลกเปลี่ยนเงินไทยกับเงินปอนด์ของประเทศอังกฤษมีอัตราการแลกเปลี่ยนค่าเฉลี่ยเลขคณิต 65.8 บาทต่อปอนด์ในรอบ 10 ปี (พ.ศ.2543-2553)

วิธีที่ 2 การหาค่าเฉลี่ยเลขคณิตสำหรับข้อมูลที่จัดกลุ่ม

วิธีการการกำหนดข้อมูลที่จัดกลุ่มนี้นิยมใช้ในกรณีมีข้อมูลเป็นจำนวนมาก ดังนั้นจึงมีการจัดกลุ่มเป็นช่วงของกลุ่มคล้ายกับตารางแจกแจงความถี่

ก. ค่าเฉลี่ยเลขคณิตประชากรสำหรับข้อมูลที่จัดกลุ่ม

สูตร $$\mu = \frac{\sum x_i f_i}{\sum f_i}$$

X_i =ค่ากึ่งกลางของแต่ละชั้น i

f_i = ค่าความถี่ของแต่ละชั้น i

N = $\sum f_i$ = จำนวนข้อมูลประชากร

ข. ค่าเฉลี่ยเลขคณิตตัวอย่างสำหรับข้อมูลที่จัดกลุ่ม

สูตร $$\bar{x} = \frac{\sum x_i f_i}{n}$$

x_i =ค่ากึ่งกลางของแต่ละชั้น i

f_i = ค่าความถี่ของแต่ละชั้น i

n = $\sum f_i$ = จำนวนข้อมูลประชากร

ตัวอย่าง 3.3 สมมตินักศึกษามีการใช้จ่ายทางการเงินในการบริโภคสินค้าและบริการภายในมหาวิทยาลัยต่อวัน โดยการสุ่มตัวอย่างมาจำนวน 50 คนและมีการแจกแจงความถี่ ดังตาราง

ตารางที่ 3.3 การใช้จ่ายทางการเงินในการบริโภคสินค้าและบริการของนักศึกษา

ขอบเขตจำนวนชั้น (ค่าใช้จ่ายเงินหน่วยเป็นบาท)	จำนวนนักศึกษา (ความถี่) = f_i
151-200	14
201-250	15
251-300	8
301-350	9
351-400	4
รวม	50

วิธีทำ x_i=ค่ากึ่งกลางของแต่ละชั้นi=(ขอบเขตล่าง+ขอบเขตบน)/2

ตารางที่ 2.4คำนวณการใช้จ่ายทางการเงินในการบริโภคสินค้าและบริการของนักศึกษา

ขอบเขตจำนวนชั้น (ค่าใช้จ่ายเงินหน่วยเป็นบาท)	จำนวนนักศึกษา (ความถี่) = f_i	ค่ากึ่งกลาง ของแต่ละชั้น i	f_iX_i
150.5-200.5	14	176.0	2,465.0
200.5-250.5	15	225.5	3,382.5
250.5-300.5	8	275.5	2,204.0
300.5-350.5	9	325.5	2,929.5
350.5-400.5	4	375.5	1,502.0
รวม	50		12,483.0

สูตร $\bar{x} = \frac{\sum x_i f_i}{n} = \frac{12,483}{50} = 249.7$ บาทต่อวัน

ดังนั้น นักศึกษามีการใช้จ่ายเงินในการบริโภคสินค้าและบริการภายในมหาวิทยาลัยต่อวันมีค่าเฉลี่ยเลขคณิต 249.7 บาทต่อวัน

วิธีที่ 3 การหาค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนัก (Bluman, 2008: 112-113)

การหาค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนัก เกิดจากข้อมูลในแต่กลุ่มมีน้ำหนักในการวัดค่าแตกต่างกันหรือความสำคัญไม่เหมือนกัน ดังนั้นการใช้วิธีนี้สามารถลดความคลาดเคลื่อนในการหาค่าเฉลี่ยเลขคณิต

สูตร $$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + .. + w_n x_n}{w_1 + w_2 + .. + w_n} = \frac{\sum w_i x_i}{\sum w_i}$$

w_i = ถ่วงน้ำหนักของข้อมูลนั้นๆ

x_i =ค่าคะแนน

ตัวอย่างที่ 3.4เกษตรกรได้วิเคราะห์ถึงค่าเฉลี่ยถ่วงน้ำหนักของสินค้าทางการเกษตรที่นำมาจำหน่ายในตลาดกลางซึ่งประกอบด้วยสินค้าการเกษตรทั้งสิ้น 5 ชนิดต่อราคาต่อกิโลกรัม ดังนี้

ตารางที่ 3.5 สินค้าทางการเกษตรที่นำมาจำหน่าย

สินค้าทางการเกษตร	ราคาต่อกิโลกรัม	จำนวนที่ขายได้ (กิโลกรัม)
พริก	20	15
ผักบุ้ง	12	20
คะน้า	22	17
ผักกาดขาว	14	23
กะหล่ำปลี	13	19

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{(20x15)+(12x20)+(22x17)+(14x23)+(13x19)}{81} = \frac{1{,}483}{81} = 18.31$$

ดังนั้น ค่าเฉลี่ยถ่วงน้ำหนักของการขายสินค้าทางการเกษตรเฉลี่ย 18.31 บาทต่อกิโลกรัม

2. ค่าเฉลี่ยเรขาคณิต (Geometric mean)เป็นค่าเกิดจากการนำข้อมูลมาถอดรากที่*N*เพื่อหาค่ากลาง โดยค่า x_i ไม่มีค่าบางตัวเป็น 0

ในการหาค่าเฉลี่ยเรขาคณิตจะมีข้อแตกต่างจากค่าเฉลี่ยเลขคณิตเนื่องจากการหาค่าเฉลี่ยเรขาคณิตจะพิจารณาค่าเฉลี่ยเข้าสู่ส่วนกลางที่เน้นสัดส่วนการเปลี่ยนแปลงเป็นเปอร์เซ็นต์ที่อาจมีการเพิ่มค่าในแต่ละช่วงเวลาๆ ซึ่งจะทำให้เกิดการคลาดเคลื่อนน้อยกว่าค่าเฉลี่ยเลขคณิต

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่ค่าเฉลี่ยเรขาคณิต = $\sqrt[N]{x_1 x_2 x_3 \dots x_n}$

$x_1 x_2 x_3 \dots x_n$ = ข้อมูลที่พิจารณา   *N* = จำนวนข้อมูลทั้งหมดที่พิจารณา

กรณีข้อมูลค่าที่พิจารณา $x_1 x_2 x_3 \ldots x_n$ เกิดมีความถี่ f_i คือ $\sum_{i=1}^{k} f_i$และทุกค่าต้องเป็นบวก ดังนั้นใช้สูตรดังนี้

กรณีข้อมูลแจกแจงความถี่ค่าเฉลี่ยเรขาคณิต = $\sqrt[N]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \ldots x_k^{f_k}}$

เนื่องจากการคำนวณในการถอดรากที่ *N* ของจำนวนข้อมูลที่พิจารณานั้น มีความยุ่งยากซึ่งก่อให้เกิดความไม่สะดวกโดยเฉพาะเมื่อมีการคำนวณหาข้อมูลเป็นจำนวนมาก ดังนั้น เพื่อให้เกิดความสะดวกและรวดเร็วจึงมีการนำวิธีการใช้ลอการิทึมในการคำนวณ คือ (Shanmugam, Ramalingam and Chattamvelli, Rajan., 2015: 59-60)

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่ $\log \text{G. M.} = \frac{1}{N} \sum_{i=1}^{N} \log x_i$

กรณีข้อมูลที่แจกแจงความถี่ $\log \text{G. M.} = \frac{1}{N} \sum_{i=1}^{N} f_i \log x_i$

x_i = จุดกึ่งกลางของอัตราภาคชั้นที่ i, f_i= ความถี่ของอันตรภาคของชั้นที่ i, k = จำนวนอันตรภาคชั้น

ตัวอย่างที่ 3.5 กำหนดค่าให้ ดังนี้ 13.82, 36.54, 33.46, 32.16, 30.82, 35.82 จงหาค่าเฉลี่ยเรขาคณิต

ให้ค่า	log 13.82 = 1.1405	log 36.54 = 1.5628	log 33.46 = 1.5245
	log 32.16 = 1.5073	log 30.82 = 1.4895	log 35.82 = 1.5541

วิธีทำ $\log \text{G. M.} = \frac{1}{N} \sum_{i=1}^{N} \log x_i = \frac{1}{6} \sum_{i=1}^{6} \log x_i = \frac{8.7787}{6} \approx 1.46312$

ดังนั้น ค่าเฉลี่ยเรขาคณิตของข้อมูล $\log x$ = 1.46312(เปิดตาราง 0.4639 $\approx \log 2.91$)

$= \log 10 + \log 2.91 = \log(10^1 x 2.91) \approx 29.10$

3. ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean) บางครั้งอาจเรียกว่า “ค่าเฉลี่ยแบบไม่ต่อเนื่อง” (sub-contrary mean) ซึ่งเป็นหนึ่งในรูปแบบค่าเฉลี่ยในรูปแบบพีทาโกรัส โดยปกติค่าเฉลี่ยฮาร์โมนิคนิยมใช้ในสถานการณ์หาอัตราเฉลี่ยที่สามารถบ่งบอกถึงค่าสัมประสิทธ์เลขคณิตของชุดข้อมูลที่เราสนใจศึกษา

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่ค่าเฉลี่ยฮาร์โมนิค= $\frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}} = \frac{1}{(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_N})}$

โดยค่า $x_1 x_2 x_3 \ldots x_n$ เป็นข้อมูล *N* จำนวนที่มีค่าเป็นจำนวนบวกทุกจำนวน

กรณีข้อมูลที่ได้แจกแจงความถี่ค่าเฉลี่ยฮาร์โมนิค $= \frac{N}{\sum_{i=1}^{N} \frac{f_i}{x_i}} = \frac{1}{(f_1\frac{1}{x_1}+f_2\frac{1}{x_2}+\cdots+f_n\frac{1}{x_N})}$

x_i มีความถี่ f_i ค่า $\sum_{\mathrm{i}=1}^{k} f_i$, ค่า k คือ จำนวนอันตรภาคชั้น, ค่า x_i คือ ค่ากึ่งกลางของอันตรภาคชั้นที่ i

ตัวอย่างที่ 3.6 การสอบพิมพ์คอมพิวเตอร์ของนักศึกษาเศรษฐศาสตร์โดยมีการจับเวลาในการพิมพ์ทั้งสิ้น จำนวน 1 ชั่วโมง ภายใต้จำนวน 2 หน้า ซึ่งสุ่มนักศึกษามาวิเคราะห์จำนวน 7 คน พบว่า นักศึกษาใช้เวลาในการพิมพ์คอมพิวเตอร์ดังนี้ 13 8 10 11 7 9 และ 12 นาทีตามลำดับ จงวิเคราะห์ค่าเฉลี่ยฮาร์โมนิค

วิธีทำ นักศึกษาจำนวน 7 คนมีสัดส่วนในการพิมพ์คอมพิวเตอร์ คือ 13 8 10 11 7 9 12

ดังนั้น นักศึกษาคนที่ 1 ทำงาน 1 นาที ได้งาน $\frac{1}{13}$

นักศึกษาคนที่ 2 ทำงาน 1 นาที ได้งาน $\frac{1}{8}$

นักศึกษาคนที่ 3 ทำงาน 1 นาที ได้งาน $\frac{1}{10}$

นักศึกษาคนที่ 4 ทำงาน 1 นาที ได้งาน $\frac{1}{11}$

นักศึกษาคนที่ 5 ทำงาน 1 นาที ได้งาน $\frac{1}{7}$

นักศึกษาคนที่ 6 ทำงาน 1 นาที ได้งาน $\frac{1}{9}$

นักศึกษาคนที่ 7 ทำงาน 1 นาที ได้งาน $\frac{1}{12}$

ค่าเฉลี่ยฮาร์โมนิค$= \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}} = \frac{1}{(\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_N})} = \frac{1}{(\frac{1}{13}+\frac{1}{8}+\frac{1}{10}+\frac{1}{11}+\frac{1}{7}+\frac{1}{9}+\frac{1}{12})} \approx 1.37$

ดังนั้น ค่าเฉลี่ยของอัตราการพิมพ์คอมพิวเตอร์ทั้ง 7 คน คือ 1.37 นาทีต่องานหนึ่งหน่วยซึ่งในเวลา 60 นาที จะได้งานเท่ากับ 7 x $\frac{60}{1.37} \approx$ 306.57 หน่วย

เปรียบเทียบการใช้ค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยเรขาคณิตและค่าเฉลี่ยฮาร์โมนิค

ในการคำนวณหาค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยเรขาคณิตและค่าเฉลี่ยฮาร์โมนิคมีการนำไปใช้ในแต่ละลักษณะแตกต่างกันที่สอดคล้องกับข้อมูลที่พิจารณา กล่าวคือ ค่าเฉลี่ยเลขคณิตจะใช้ได้เพียงนำเมื่อข้อมูลทั้งหมดมารวมกันแล้วหาค่าเฉลี่ย ขณะที่ค่าเฉลี่ยเรขาคณิตจะนิยมใช้วิเคราะห์ถึงข้อมูลที่มูลค่า ค่าตอบแทนในรูปแบบหน่วยของอัตราดอกเบี้ย อัตราการลงทุนต่างๆ สุดท้ายค่าเฉลี่ยฮาร์โมนิคจะใช้ในการหาสัดส่วนของความไวต่อค่าเฉลี่ยอย่างไรก็ตามการใช้ค่าเฉลี่ยทั้ง 3 ประเภทมีข้อดีและข้อเสียแตกต่างกัน โดยสามารถเปรียบเทียบให้เห็นข้อดีและข้อเสีย ดังตาราง

ตารางที่ 3.6 เปรียบเทียบข้อดี-ข้อเสียการใช้ค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยเรขาคณิตและค่าเฉลี่ยฮาร์โมนิค

เปรียบเทียบค่าเฉลี่ย	**ค่าเฉลี่ยเลขคณิต**	**ค่าเฉลี่ยเรขาคณิต**	**ค่าเฉลี่ยฮาร์โมนิค**
ข้อดี	1.คำนวณไม่ยุ่งยาก 2.สามารถใช้ได้กับข้อมูลทุกค่าที่นำมารวมกันแล้วหาค่าเฉลี่ย	ข้อมูลใกล้เคียงความเป็นจริงเพราะพิจารณาถึงสัดส่วนค่าร้อยละอัตราส่วนต่างๆ ส่วนใหญ่หาค่าเพื่อวิเคราะห์ทางธุรกิจ เศรษฐศาสตร์	1.แม้ข้อมูลมีไม่ครบสามารถคำนวณได้ 2.เปรียบเทียบกับข้อมูลได้หลายชุดข้อมูล
ข้อเสีย	1.ใช้ได้เฉพาะข้อมูลเชิงปริมาณเท่านั้น 2.ถ้าข้อมูลมีความห่างกันมากหรือผิดปกติสูงเมื่อนำข้อมูลมารวมกันหาค่าเฉลี่ยจะทำให้ค่าอาจสูงหรือต่ำเกินความเป็นจริง	1. ใช้ได้เฉพาะข้อมูลเชิงปริมาณเท่านั้น 2. คำนวณยุ่งยากในการถอดรากหรือการเปลี่ยนค่าเป็นค่าลอการิทึม	1. ใช้ได้เฉพาะข้อมูลเชิงปริมาณเท่านั้น 2. ข้อมูลที่ได้มาแตกต่างกันมากจะทำให้เกิดค่าเฉลี่ยที่ผิดปกติเกิดความคลาดเคลื่อนสูง

ที่มา: กัลยา วานิชย์บัญชา, 2557: 46-56

2. ค่ามัธยฐาน (Median) คือ ข้อมูลที่พิจารณาด้วยการแยกออกครึ่งหนึ่งของข้อมูลเป็นข้อมูลค่า "กลาง" โดยมีข้อมูลสังเกตครึ่งบนและครึ่งล่างเป็นองค์ประกอบ (Goos, Peter, and Meintrup, David., 2015: 56-57) โดยการคำนวณค่ามัธยฐานแบ่งเป็น 2 วิธี คือ วิธีที่ 1 ค่ามัธยฐานแบบไม่จัดกลุ่ม วิธีที่ 2 ค่ามัธยฐานแบบจัดกลุ่ม (กัลยา, 2557: 57-59)

วิธีที่ 1 ค่ามัธยฐานแบบไม่จัดกลุ่ม

การหาค่ามัธยฐานแบบไม่จัดกลุ่มมี 2 ขั้นตอน คือ

1.นำข้อมูลมาเรียงลำดับจากมากไปหาน้อยหรือจากน้อยไปหามากก็ได้

2.หาค่าข้อมูลมัธยฐานที่เป็นข้อมูลกลาง โดยการหาค่ากลางของข้อมูลนั้นสามารถพิจารณาข้อมูล 2 แบบ คือ

1.ข้อมูลที่มีจำนวนเป็นเลขคี่ จะใช้สูตร $\frac{(n+1)}{2}$(Wilcox, Rand R., 2009: 15)

2.ข้อมูลที่มีจำนวนเป็นเลขคู่จะใช้สูตร เช่นเดียวกับจำนวนเป็นเลขคี่ แต่ต้องคำนวณหาระหว่างค่ากลางอีกครั้ง ใช้สูตร $\frac{(\text{ตัวเลขก่อนค่ากลาง} + \text{ตัวเลขหลังค่ากลาง})}{2}$

ตัวอย่างที่ 3.7 กรณีข้อมูลเป็นเลขคี่สมมติว่าคนงานก่อสร้างจำนวน 7 คนได้ทำการขนอิฐจากที่หนึ่งมายังที่หนึ่ง คิดเป็นจำนวนก้อนอิฐที่ขนได้ดังนี้ 1125, 2054, 1384, 1440, 1702, 1593 และ 1668 ก้อน จงหาค่ามัธยฐานของการขนอิฐของคนงานก่อสร้าง

วิธีทำ จำนวนคนงานก่อสร้างที่ขนอิฐจำนวน 7 คน คือ $n = 7$

เรียงลำดับข้อมูลจากน้อยไปหามากได้ ดังนี้ 1125, 1384, 1440,1593, 1668, 1702, 2054

ค่ามัธยฐาน = $\frac{(n+1)}{2} = \frac{(7+1)}{2} = \frac{(8)}{2} = 4$

ดังนั้น ค่ามัธยฐาน คือ 1125, 1384, 1440, 1593, 1668, 1702, 2054

สรุปได้ว่าค่ามัธยฐานการขนอิฐของคนงานก่อสร้าง= 1593ก้อน

ตัวอย่างที่ 3.8 กรณีข้อมูลเป็นเลขคู่สมมติว่าคะแนนสอบของนักศึกษาในการสอบกลางภาครายวิชาเศรษฐศาสตร์มหภาค จำนวน 8 คน มีคะแนนดังนี้

21	25	18	20	19	23	27	26

จงหาค่ามัธยฐานของคะแนนสอบนักศึกษากลุ่มนี้

วิธีทำ จำนวนนักศึกษาในการสอบกลางภาคทั้งสิ้น 8 คน ดังนั้น $n = 8$

เรียงลำดับข้อมูลจากน้อยไปหามากได้ ดังนี้ 18, 19, 20, 21, 23, 25, 26 และ 27

ค่ามัธยฐาน = $\frac{(n+1)}{2} = \frac{(8+1)}{2} = \frac{(9)}{2} = 4.5$

ดังนั้น ค่ามัธยฐานจะอยู่ระหว่างค่ากลาง 21 และ 23 คือ 18, 19, 20, 21, 23, 25, 26, 27

หาค่ากลาง = $\frac{(\text{ตัวเลขก่อนค่ากลาง} + \text{ตัวเลขหลังค่ากลาง})}{2} = \frac{(21+23)}{2} = \frac{(44)}{2} = 22$

สรุปได้ว่า ค่ามัธยฐานของคะแนนนักศึกษาในการสอบกลางภาค = 22 คะแนน

วิธีที่ 2 ค่ามัธยฐานแบบจัดกลุ่ม

การหาค่ามัธยฐานแบบจัดกลุ่มมี 2 ขั้นตอน คือ

1.นำข้อมูลมาสร้าง โดยการหาจำนวนชั้นของอัตราภาค สูตร $2^c = n$

2.หาความกว้างของอันตรภาคชั้น ใช้สูตร $\frac{\text{ค่าสูงสุดของข้อมูล} - \text{ค่าต่ำสุดของข้อมูล}}{\text{จำนวนชั้นของอัตราภาค}}$

3.นำข้อมูลมาเรียงลำดับจากน้อยไปหามาก และหาค่ากลางของข้อมูลนั้นด้วยสูตร $\frac{(n)}{2}$ และหลังจากนั้นนำข้อมูลมาหาความถี่ รวมทั้งความถี่สะสมด้วยการสร้างเป็นตาราง

4.สูตรในการคำนวณหาค่ามัธยฐานแบบจัดกลุ่ม คือ

ค่ามัธยฐาน (Median)= L + $\frac{(\frac{n}{2} - CF)}{f_m}$. I

L = ค่าขอบเขตล่างของชั้นที่มีค่ามัธยฐานอยู่

CF = ความถี่สะสมของชั้นก่อนจะถึงชั้นที่มีค่ามัธยฐานอยู่

f_m = ความถี่ที่มีค่ามัธยฐานอยู่

I = ความกว้างของอันตรภาคชั้น

ตัวอย่างที่ 3.9 จากตารางตัวอย่าง 3.4 ให้คำนวณหาค่ามัธยฐานของการใช้จ่ายเงินในการบริโภคสินค้าและบริการของนักศึกษา

ตารางที่ 3.7 การใช้จ่ายเงินในการบริโภคสินค้าและบริการของนักศึกษา

ขอบเขตจำนวนชั้น (ค่าใช้จ่ายเงินหน่วยเป็นบาท)	จำนวนนักศึกษา (ความถี่) = f_i	ความถี่สะสม
150.5-200.5	14	14
200.5-250.5	15	29
250.5-300.5	8	37
300.5-350.5	9	46
350.5-400.5	4	50
รวม	50	

วิธีทำ ค่ากลางของข้อมูล =$\frac{n}{2}$ = $\frac{50}{2}$= 25 ซึ่งอยู่ชั้นที่ 2 ของขอบเขตความกว้างชั้น 200.5-250.5

ค่ามัธยฐาน = L +$\frac{(\frac{n}{2} - cf)}{f_m}$. I = 200.5 + $\frac{(\frac{50}{2} - 14)}{15}$ x 50 = 200.5 + $\frac{(11)}{15}$x 50

= 200.5 + 36.67 ≈237.17

ดังนั้น ค่ามัธยฐานของการใช้จ่ายเงินของนักศึกษาเท่ากับ 237.17บาท

ข้อดีของมัธยฐาน คือ สามารถใช้ได้ง่ายทั้งข้อมูลที่เป็นเชิงปริมาณและเชิงคุณภาพลักษณะแบบสเกลอันดับ

ข้อเสียของมัธยฐาน คือ ไม่ได้ใช้ข้อมูลทั้งหมดในการคำนวณ

3. ค่าฐานนิยม (Mode) คือชุดข้อมูลที่ศึกษามีข้อมูลปรากฏบ่อยที่สุดหรือมีความถี่ของข้อมูลสูงสุดที่ซ้ำกัน ซึ่งเรียกว่า "ค่าฐานนิยม" (model value) (Shanmugam, Ramalingam and Chattamvelli, Rajan., 2015: 58-59) โดยการคำนวณหาค่าฐานนิยมทำได้ 2 วิธี คือ

วิธีที่ 1 ค่าฐานนิยมที่ไม่ได้จัดกลุ่ม

ใช้วิธีพิจารณาชุดข้อมูลที่ศึกษาว่าข้อมูลใดมีข้อมูลที่ซ้ำกันมากที่สุด

ตัวอย่างที่ 3.10 ให้หาฐานนิยมต่อไปนี้

8 9 6 9 2 6 7 9 3 5 1 0

วิธีทำ ค่าที่มีความถี่ซ้ำกันมากที่สุด คือ 9 ดังนั้น ค่าฐานนิยม = 9

ตัวอย่างที่ 3.11 ให้หาฐานนิยมต่อไปนี้

6 5 2 9 6 3 2 4 9 1 0 8 5 3 7 9 6 4 3 1

วิธีทำ ค่าที่มีความถี่ซ้ำกันมากที่สุดและมีค่าเท่ากัน คือ 6, 9 และ 3 ดังนั้น ค่าฐานนิยม = 6, 9, 3

ตัวอย่างที่ 3.22 ให้หาฐานนิยมต่อไปนี้

1 2 3 4 5 6 7 8 9 0

วิธีทำ ค่าที่มีความถี่ซ้ำกันไม่มี ดังนั้น ค่าฐานนิยม = ไม่มี

วิธีที่ 2 ค่าฐานนิยมที่จัดกลุ่ม

การหาค่าฐานนิยมแบบจัดกลุ่มมี 2 ขั้นตอน คือ

1. นำข้อมูลมาสร้าง โดยการหาจำนวนชั้นของอัตราภาค สูตร $2^c = n$
2. หาความกว้างของอันตรภาคชั้น ใช้สูตร $\frac{\text{ค่าสูงสุดของข้อมูล} - \text{ค่าต่ำสุดของข้อมูล}}{\text{จำนวนชั้นของอัตราภาค}}$
3. พิจารณาชั้นที่มีจำนวนความถี่มากที่สุดให้ใช้ชั้นนั้น
4. สูตรในการคำนวณหาค่าฐานนิยมแบบจัดกลุ่ม คือ

ค่าฐานนิยม (Mode) = $L_0 + [\frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \cdot I]$ หรือใช้ $= L_0 + [\frac{\delta_u}{\delta_u + \delta_1} \cdot I]$

L_0 = ค่าขอบเขตล่างของชั้นที่มีความถี่มากที่สุด

f_m = ความถี่ของชั้นที่มีความถี่มากที่สุด

f_1 = ความถี่ของชั้นก่อนที่มีความถี่มากที่สุด

f_2 = ความถี่ของชั้นหลังที่มีความถี่มากที่สุด

I = ความกว้างของอันตรภาคชั้น

δ_u = ความแตกต่างระหว่างความถี่ของชั้นกับความถี่ของชั้นก่อนฐานนิยม

δ_1 = ความแตกต่างระหว่างความถี่ของชั้นกับความถี่ของชั้นหลังฐานนิยม

ตัวอย่างที่ 3.23 จากตารางตัวอย่าง 2.4 ให้คำนวณหาค่าฐานนิยมของการใช้จ่ายเงินในการบริโภคสินค้าและบริการของนักศึกษา(โดยสมมติให้ค่ากลางของข้อมูลเท่ากับ 278 และn = 50)

ตารางที่ 3.8 การใช้จ่ายเงินในการบริโภคสินค้าและบริการของนักศึกษาเพื่อหาค่าฐานนิยม

ขอบเขตจำนวนชั้น (ค่าใช้จ่ายเงินหน่วยเป็นบาท)	จำนวนนักศึกษา (ความถี่) = f_i
150.5-200.5	14
200.5-250.5	15
250.5-300.5	8
300.5-350.5	9
350.5-400.5	4
รวม	50

วิธีทำ ค่าความถี่มีจำนวนมากที่สุด คือ 15 ซึ่งอยู่ในชั้นที่ 2 ระหว่างค่าใช้จ่าย 200.5-250.5

สูตร ค่าฐานนิยม (Mode) = $L_0 + [\frac{f_m - f_1}{(f_m - f_1)+(f_m - f_2)} \cdot I]$

$$= 200.5 + \frac{15-14}{(15-14)+(15-8)} \times 50$$

$$= 200.5 + \left(\frac{1}{1+7}\right) \times 50$$

$$= 200.5 + 6.25 = 206.30$$

ดังนั้น ค่าฐานนิยมของการใช้จ่ายเงินของนักศึกษาเท่ากับ 206.30 บาท

ข้อดีของฐานนิยม คือ ไม่ต้องกังวลกับข้อมูลที่มีค่าผิดปกติและสามารถวัดค่ากลางได้ทั้งข้อมูลเชิงปริมาณและเชิงคุณภาพที่เป็นมาตรวัดนามบัญญัติ

ข้อเสียของฐานนิยม คือ กลุ่มข้อมูลที่พิจารณาไม่มีข้อมูลที่มีความถี่มากที่สุดหรือข้อมูลที่ซ้ำกันก็ไม่สามารถหาค่าฐานนิยมได้ และถ้ากลุ่มข้อมูลในลักษณะจัดกลุ่มแล้วมีค่าฐานนิยมที่มีความถี่มากที่สุดเท่ากันหลายๆ ชั้น เมื่อคำนวณหาค่าฐานนิยมจะได้ค่าแตกต่างกัน

สรุปได้ว่าการหาค่าแนวโน้มเข้าสู่ส่วนกลางทั้งการหาค่าเฉลี่ยเลขคณิต ค่ามัธยฐานและค่าฐานนิยมนั้น มีความแตกต่างกันในการวัดค่ากลาง นอกจากนั้นต้องเลือกใช้ให้สัมพันธ์กับมาตรวัดทั้ง 4 ประเภท

ด้วยกัน คือ มาตราวัดแบบแบ่งกลุ่ม (nominal scales) มาตราวัดแบบเรียงอันดับ (ordinal scales) มาตราวัดแบบช่วง (interval scales) มาตราวัดแบบอัตราส่วน (radio scales) โดยสามารถสรุปการใช้ค่ากลางได้ ดังตารางที่ 3.9

ตารางที่ 3.9 การวัดแนวโน้มเข้าสู่ส่วนกลางสัมพันธ์กับมาตรวัดต่างๆ

ระดับมาตรวัด	การวัดแนวโน้มเข้าสู่ส่วนกลาง
มาตราวัดแบบแบ่งกลุ่ม (nominal scales)	ฐานนิยม
มาตราวัดแบบเรียงอันดับ (ordinal scales)	ฐานนิยม, มัธยฐาน
มาตราวัดแบบช่วง (interval scales)	ฐานนิยม, มัธยฐาน, ค่าเฉลี่ยเลขคณิต
มาตราวัดแบบอัตราส่วน (radio scales)	ฐานนิยม, มัธยฐาน, ค่าเฉลี่ยเลขคณิต

ที่มา: K. R. B. Jankowski and K.J.Flannelly, 2015: 40.

อย่างไรก็ตามการวัดค่าแนวโน้มเข้าสู่ส่วนกลาง นอกจากการหาค่าเฉลี่ยเลขคณิต ค่ามัธยฐานและค่าฐานนิยมยังสามารถใช้ค่า Midrange ในการคำนวณหาค่ากลางได้เช่นกัน (กัลยา, 2557: 62)

$$\text{Midrange} = \frac{\text{ค่าสูงสุด} + \text{ค่าต่ำสุด}}{2}$$

ตัวอย่างที่ 3.24 จงหาค่า Midrange ด้วยข้อมูลต่อไปนี้

45 62 89 32 51 49 34 78 66

วิธีทำ ค่าสูงสุด คือ 89 และค่าต่ำสุด คือ 32

$$\text{สูตรMidrange} = \frac{\text{ค่าสูงสุด} + \text{ค่าต่ำสุด}}{2} = \frac{89+32}{2} = 60.50$$

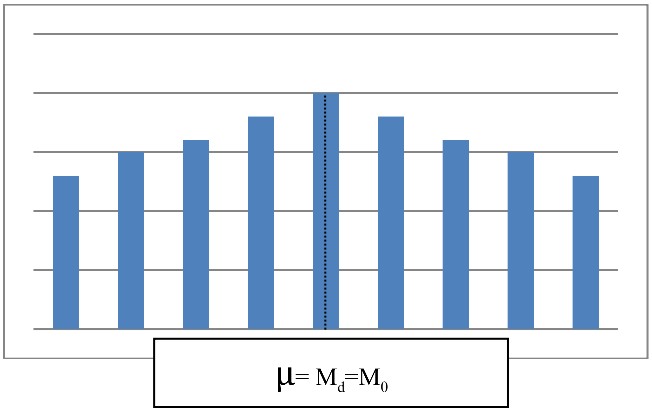
ข้อดีของการหาค่า Midrange คือ การคำนวณค่อนข้างง่ายไม่มีความซับซ้อนและนิยมใช้ในการหาค่าทางการรายได้ การเงินหรือการพยากรณ์อากาศ

ข้อเสียของการหาค่า Midrange คือ เมื่อข้อมูลค่าสูงสุดหรือค่าต่ำสุดมีความผิดปกติสูงจะทำให้การได้มาของค่ากลางมีความคลาดเคลื่อนมากและไม่มีความน่าเชื่อถือ

เปรียบเทียบความสัมพันธ์ของค่าเฉลี่ยเลขคณิตค่ามัธยฐานและค่าฐานนิยมของการแจกแจงลักษณะต่างๆ

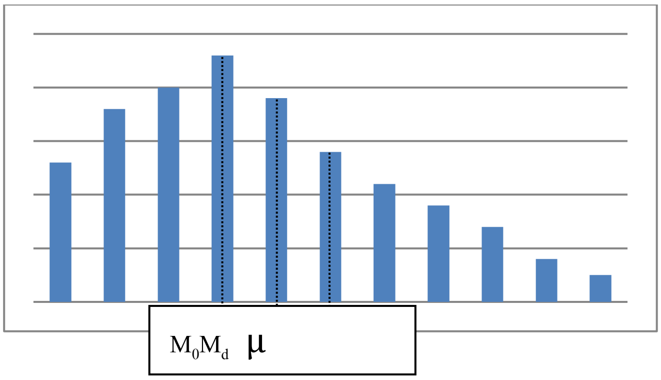
ความสัมพันธ์ของค่าเฉลี่ยเลขคณิต ค่ามัธยฐานและค่าฐานนิยมมีความสัมพันธ์ได้ทั้งสิ้น 3 แบบด้วยกัน คือ 1.ลักษณะการหาค่ากลางทั้งสามมีความสัมพันธ์สมมาตรกัน (symmetric) 2.ลักษณะการหาค่ากลางทั้งสามมีความสัมพันธ์เบ้ไปทางบวก (positively skewed distribution) 3.ลักษณะการหาค่ากลางทั้งสามมีความสัมพันธ์เบ้ไปทางลบ (negatively skewed distribution)ดังนี้ (Lind, Douglas A., Marchal, William G. &Wathen, Samuel A., 2005: 66-68., Schindler, Thomas M., 2005: 31-33)

1.ลักษณะการหาค่ากลางทั้งสามมีความสัมพันธ์สมมาตรกัน (symmetric) คือ การกระจายข้อมูลค่ากลางทั้งสามมีการกระจายเท่าๆ กันไปทางซ้ายและทางขวาด้วยการเบี่ยงเบนอย่างสม่ำเสมอ ซึ่งค่าเฉลี่ยเลขคณิตเท่ากับค่ามัธยฐานเท่ากับค่าฐานนิยม คือ ค่าเฉลี่ยเลขคณิต (μ) = ค่ามัธยฐาน(M_d) = ค่าฐานนิยม(M_o)



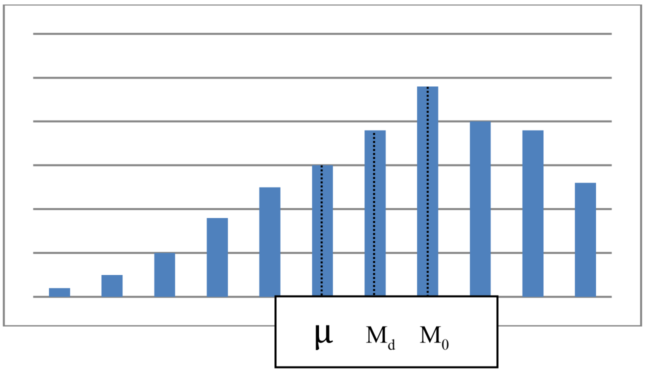
รูปที่ 3.1 ค่าเฉลี่ยเลขคณิต (μ) = ค่ามัธยฐาน (M_d) = ค่าฐานนิยม (M_o)

2.ลักษณะการหาค่ากลางทั้งสามมีความสัมพันธ์เบ้ไปทางบวก (positively skewed distribution) คือการกระจายข้อมูลค่ากลางที่ไม่สมมาตรกันแต่มีการเบ้ไปทางบวก โดยค่าเฉลี่ยเลขคณิตจะมีค่ามากกว่าค่ามัธยฐานและค่าฐานนิยม ตามลำดับ การที่ค่าเฉลี่ยเลขคณิตมีค่ามากกว่าค่ากลางทั้ง 2 นั้นเพราะค่าเฉลี่ยเลขคณิตมีอิทธิพลสูงกว่า โดยสามารถพิจารณาความสัมพันธ์ได้ คือ ค่าเฉลี่ยเลขคณิต (μ) >ค่ามัธยฐาน(M_d) > ค่าฐานนิยม(M_o)



รูปที่ 3.2 ค่าเฉลี่ยเลขคณิต (μ) >ค่ามัธยฐาน(M_d) >ค่าฐานนิยม(M_o)

3.ลักษณะการหาค่ากลางทั้งสามมีความสัมพันธ์เบ้ไปทางลบ (negatively skewed distribution)คือการกระจายข้อมูลค่ากลางมีการเบี่ยงเบนไปทางด้านลบหรือมีการเบ้ โดยค่าฐานนิยมจะมีอิทธิพลสูงกว่าค่ามัธยฐานและค่าเฉลี่ยเลขคณิต ตามลำดับ โดยสามารถพิจารณาความสัมพันธ์ได้ คือ ค่าฐานนิยม(M_o)> ค่ามัธยฐาน(M_d) >ค่าเฉลี่ยเลขคณิต (μ)



รูปที่ 3.3 ค่าฐานนิยม(M_o)>ค่ามัธยฐาน(M_d) >ค่าเฉลี่ยเลขคณิต (μ)

การวัดข้อมูลเชิงปริมาณในตำแหน่งต่างๆ นอกเหนือค่าแนวโน้มเข้าสู่ส่วนกลาง

การพิจารณากลุ่มค่าตัวเลขที่สนใจมิใช่จะศึกษาค่าแนวโน้มเข้าสู่ส่วนกลางเพียงอย่างเดียวเท่านั้น ยังมีวิธีการวัดข้อมูลเชิงปริมาณในรูปแบบอื่นๆ ของตำแหน่งที่นอกเหนือค่ากลาง ดังนั้นในที่นี้จะอธิบายถึงค่าที่นิยมใช้กันมากที่สุด คือ การหาค่าเปอร์เซ็นต์ไทล์, ค่าควอไทล์และค่าเดไซน์ ดังต่อไปนี้ (กัลยา, 2557: 63-67)

1. ค่าเปอร์เซ็นต์ไทล์ (Percentile) จะเป็นการแบ่งข้อมูลเป็นส่วนๆ ทั้งสิ้น 100 ส่วนด้วยการเรียงลำดับจากส่วนละ 1% จนไปถึง 100% และเมื่อเทียบกับค่ามัธยฐานจะแบ่งข้อมูลเป็น 2 ส่วน ส่วนละ 50% ดังนั้นการหาค่าเปอร์เซ็นต์ไทล์จะมีข้อดีที่แตกต่างจากการวัดค่ากลาง คือ สามารถวัดได้ทั้งการหาค่ากลางและหาค่าที่มิใช่ค่ากลางที่เป็นข้อมูลเชิงปริมาณ

วิธีการหาค่าเปอร์เซ็นต์ไทล์การหาค่าจะต้องแบ่งข้อมูลเป็น 100 ส่วน ส่วนละเท่าๆ กันอย่างละ 1% โดยนำข้อมูลที่ได้มาเรียงลำดับจากน้อยไปมาก อาทิเช่น ค่าเปอร์เซ็นต์ไทล์ที่ 75 หมายถึง จำนวนที่มีการเรียงลำดับจากน้อยไปมากอยู่ลำดับที่ 75 เป็นต้น ในการหาค่าเปอร์เซ็นต์ไทล์สามารถแบ่งได้ 2 วิธี คือ

วิธีที่ 1 ข้อมูลที่ไม่ได้จัดกลุ่ม มีขั้นตอนดังนี้

1. เรียงลำดับข้อมูลค่าจากน้อยไปมาก
2. สูตรในการคำนวณ $\frac{p(n+1)}{100}$
3. คำตอบที่ได้เป็นจุดทศนิยมให้ปัดขึ้นเมื่อหลังจุดทศนิยมมีตัวเลขตั้งแต่ 5 ขึ้นไป แต่ถ้าหลังจุดทศนิยมมีตัวเลขน้อยกว่า 5 ให้ปัดลง

ตัวอย่างที่ 3.25 ให้หาค่าเปอร์เซ็นไทล์ที่ 70 จากข้อมูลข้างล่างนี้

7.2 4.9 8.5 6.5 6.2 7.7 5.8

วิธีทำ เรียงลำดับจากน้อยไปมากได้ดังนี้ 4.9 5.8 6.2 6.5 7.2 7.7 8.5

สูตร ค่าเปอร์เซ็นไทล์ที่ 70 = $P_{70} = \frac{p(n+1)}{100} = \frac{70(7+1)}{100} = \frac{560}{100} = 5.6 \approx 6$

ดังนั้น ค่า P_{70}คือ ค่าลำดับที่ 6 ของข้อมูลเรียงลำดับจึงได้ค่า = 7.7

วิธีที่ 2 ข้อมูลที่จัดกลุ่ม มีขั้นตอนดังนี้

สูตรที่ใช้คำนวณ คือ $P_r = L + \frac{i}{f_r}(\frac{nr}{100} - \sum f_i)$

L = ขอบเขตจำกัดล่างของชั้นที่มีค่าเปอร์เซ็นต์ไทล์r อยู่

I = ความกว้างของชั้น

n = จำนวนข้อมูลที่พิจารณา

$\sum f_i$ = ผลรวมของความถี่สะสมของชั้นก่อนถึงค่าเปอร์เซ็นต์ไทล์

f_r = ความถี่ของชั้นที่ค่าของเปอร์เซ็นต์ไทล์

r = ตำแหน่งค่าเปอร์เซ็นต์ไทล์ที่พิจารณา

ตัวอย่างที่ 3.26 จากตารางตัวอย่าง 3.4 ให้คำนวณหาค่าเปอร์เซ็นต์ไทล์ที่ 42

ตารางที่ 3.9 การใช้จ่ายเงินในการบริโภคสินค้าและบริการของนักศึกษาเพื่อหาค่าเปอร์เซ็นต์ไทล์

ขอบเขตจำนวนชั้น (ค่าใช้จ่ายเงินหน่วยเป็นบาท)	จำนวนนักศึกษา (ความถี่) = f_i	ความถี่สะสม
150.5-200.5	14	14
200.5-250.5	15	29
250.5-300.5	8	37
300.5-350.5	9	46
350.5-400.5	4	50
รวม	50	

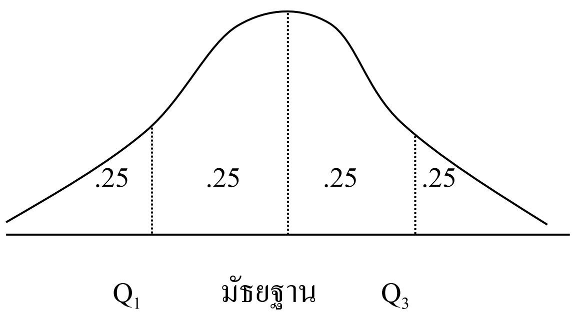
วิธีทำให้หาค่าเปอร์เซ็นต์ไทล์ที่ 42 ดังนั้น P_{42}จะอยู่ที่ระดับความถี่สะสมระหว่างค่า 37 กับค่า 46 ซึ่งหมายความว่า ค่า P_{42} จะอยู่ที่ชั้น 4

จากตารางที่ 2.9 จะได้ค่า ดังนี้ $L = 300.5, I = 50, \sum f_i = 46, f_{42} = 9$

แทนค่าในสูตร $P_{42} = L + \frac{I}{f_r}(\frac{nr}{100} - \sum f_i) = 300.5 + \frac{50}{9}(\frac{50*42}{100} - 46)$

$= 300.5 + (-138.89) = 161.61$ บาท

2. ค่าควอไทล์ (Quartiles) เป็นการแบ่งข้อมูลออกเป็น 4 ส่วนเท่าๆ กัน ส่วนละ 25% หรือคิดเป็น 1/4 ของข้อมูลทั้งหมดโดยมีการเรียงลำดับจากน้อยไปมากและเมื่อเทียบกับค่ามัธยฐานจะแบ่งเป็น 2 ส่วนเท่าๆ กัน ส่วนละ 50% ดังรูป



รูปที่ 3.4 การแบ่งส่วนเป็น 4 ส่วนเท่าๆ กันของค่าควอไทล์

จากรูปที่ 3.4 จะพบว่าค่า Q_3มีค่ามากกว่าค่า Q_1 เพราะเมื่อเรียงลำดับจากน้อยไปหามากจะได้ค่า $Q_1 = 25\%$ ขณะที่ค่า $Q_3 = 75\%$ จึงสามารถแบ่งเป็นค่าควอไทล์ล่างและควอไทล์บนได้ว่า ควอไทล์ล่าง Q_L

(Lower Quartile) จะมีค่าข้อมูลที่น้อยกว่า Q_1 ส่วนควอไทล์บน Q_U (Upper Quartile) จะมีค่าข้อมูลมากกว่า Q_3 ขึ้นไป ส่วนค่ามัธยฐานเทียบค่าควอไทล์ คือ Q_2 ที่อยู่ระหว่างกลางของค่า Q_1 และค่า Q_3 ที่มีสัดส่วนการแบ่งเป็นอย่างละ 50% ของข้อมูลทั้งหมด

วิธีการหาค่าควอไทล์ การหาค่าสามารถหาได้ 2 วิธี ดังนี้

วิธีที่ 1 ข้อมูลที่ไม่ได้จัดกลุ่ม มีขั้นตอนดังนี้

1. นำข้อมูลที่พิจารณามาเรียงลำดับจากน้อยไปมาก
2. สูตรคำนวณ กรณีคำนวณหาค่า Q_1 ใช้สูตร $\frac{(n+1)}{4}$
3. สูตรคำนวณ กรณีคำนวณหาค่า Q_3 ใช้สูตร $(3 \times \frac{(n+1)}{4})$
4. เมื่อคำนวณได้ตัวเลขที่ไม่ลงตัวให้หาค่าใกล้เคียงที่สุดด้วยการการปัดจุดทศนิยมโดยถ้าจุดทศนิยมต่ำกว่า 0.5 ให้ปัดตัวเลขลงและถ้าจุดทศนิยมตั้งแต่ 0.5 ขึ้นไปให้ปัดตัวเลขขึ้นไป

วิธีที่ 2 ข้อมูลที่จัดกลุ่ม มีขั้นตอนดังนี้

1. สร้างตารางความถี่สะสม
2. หาตำแหน่งควอไทล์ที่เราพิจารณาโดยใช้สูตร $Q_r = \frac{r}{4}(N)$
3. หาค่าอันตรภาคชั้นของควอไทล์ที่ต้องการพิจารณา
4. ใช้สูตรคำนวณ คือ $Q_r = L + \frac{I}{f_r}(\frac{rN}{4} - \sum f_i)$

L = ขอบเขตจำกัดล่างของชั้นที่มีค่าควอไทล์ r อยู่

I = ความกว้างของชั้น

n = จำนวนข้อมูลที่พิจารณา

$\sum f_i$ = ผลรวมของความถี่สะสมของชั้นก่อนถึงค่าควอไทล์

f_r = ความถี่ของชั้นที่ค่าของควอไทล์

r = ตำแหน่งค่าควอไทล์ที่พิจารณา

3. ค่าเดไซล์ (Deciles) การหาค่าเดไซล์จะแบ่งข้อมูลเป็น 10 ส่วน ส่วนละเท่าๆ กัน โดยจะมีค่า D_1, D_2, D_3,…,D_9 ซึ่งจะแทนค่าเดไซล์ คือ D_k

วิธีการหาค่าเดไซล์ สามารถหาได้ 2 วิธี ดังนี้

วิธีที่ 1 ข้อมูลที่ไม่จัดกลุ่ม มีขั้นตอนดังนี้

1. เรียงลำดับข้อมูลจากน้อยไปมาก
2. สูตรในการคำนวณเพื่อหา $\frac{k}{10} \times (n+1)$

วิธีที่ 2 ข้อมูลที่จัดกลุ่ม มีขั้นตอนดังนี้

1. สร้างตารางความถี่สะสม
2. หาตำแหน่งของค่าเดไซล์ด้วยการใช้สูตร $D_r = \frac{r}{10}(N)$

3. หาค่าอันตรภาคชั้นของเดไซล์อยู่

4. คำนวณค่าเดไซล์จากสูตร $D_r = L + \frac{I}{f}(\frac{rN}{10} - \sum f_i)$

L =ขอบเขตจำกัดล่างของชั้นที่มีค่าเดไซล์ r อยู่

I = ความกว้างของชั้น

n = จำนวนข้อมูลที่พิจารณา

$\sum f_i$ = ผลรวมของความถี่สะสมของชั้นก่อนถึงค่าเดไซล์

f_r = ความถี่ของชั้นที่ค่าของเดไซล์

r = ตำแหน่งค่าเดไซล์ที่พิจารณา

2.การวัดการกระจายของข้อมูล (Measures of variation or spread)

การวัดการกระจายข้อมูลเป็นการวัดเพื่อวิเคราะห์สืบเนื่องจากการวัดค่าแนวโน้มเข้าสู่ส่วนกลาง เนื่องจากการวัดค่าแนวโน้มเข้าสู่ส่วนกลางเป็นการวัดขั้นแรกของการวิเคราะห์เท่านั้น ซึ่งยังมีข้อบกพร่องในหลายๆ ประเด็นด้วยกัน คือ (Chikkodi, C.M., and Satyaprasad, B.G., 2009: 10.1-10.2)

1. ในการวัดค่าแนวโน้มเข้าสู่ส่วนกลางเป็นเพียงการวัดจากชุดข้อมูลเท่านั้นแต่ไม่ให้ความสำคัญของค่าของข้อมูลในแต่ละอันว่ามีค่าต่างกันมากหรือน้อยเพียงใดซึ่งไม่ได้สังเกตในรายละเอียดของข้อมูลในชุดนั้น

2. การวัดค่าแนวโน้มเข้าสู่ส่วนกลางไม่ให้ความสำคัญของความสัมพันธ์ในชุดข้อมูล

3. ไม่สามารถอธิบายถึงการกระจายข้อมูลว่าเป็นข้อมูลปกติหรือไม่ให้เพียงแนวคิดของการวัดค่าเท่านั้น

ดังนั้นการวัดการกระจายของข้อมูลจึงมีความสำคัญต่อเนื่องจากการวัดค่าแนวโน้มเข้าสู่ส่วนกลางที่จะมาแก้ไขข้อบกพร่องดังกล่าว ดังนี้

1. ทำให้การวิเคราะห์ข้อมูลมีความถูกต้องมากขึ้น จากข้อมูลที่ได้มาว่ามีความน่าเชื่อหรือไม่

2. สามารถที่จะสังเกตลักษณะรายละเอียดของข้อมูลในชุดข้อมูลนั้นที่ทำการศึกษา

3. สามารถนำข้อมูลมาแสดงค่าของการกระจายด้วยรูปภาพที่วิเคราะห์ได้ชัดเจนขึ้น

ตัวอย่างที่ 3.27 วิเคราะห์อัตราการออมของเกษตรกรในรอบ 3 ปี ดังนี้(หน่วย: เปอร์เซ็นต์ต่อรายได้)

ปี 2557	8	8	8	8	8	8	8	8	8	8	8	8
ปี 25589		7	6	5	8	14	6	5	12	6	10	8
ปี 2559	3	4	5	6	7	8	9	10	10	11	11	12

จากตัวอย่างที่ 2.27 เมื่อหาค่าเฉลี่ยเลขคณิตทั้ง 3 ปี พบว่า ค่าเฉลี่ยเลขคณิตจะเท่ากับ 8 เปอร์เซ็นต์ต่อรายได้ ซึ่งเท่ากันทั้งปี 2557, ปี 2558 และปี 2559 ตามลำดับ แต่ถ้าลองสังเกตของการกระจายข้อมูลในแต่ละปีของอัตราการออมของเกษตรกรในรอบ 3 ปี พบว่า ปี 2557 มีอัตราการออมของเกษตรกรมีการกระจายเท่ากันทุกๆ เดือน ในขณะที่ ปี 2558 และปี 2559 มีการกระจายในแต่ละเดือนมากบ้างน้อยบ้างของค่าสูงสุดและต่ำสุดแตกต่างกันอย่างเห็นได้ชัด ดังนั้น การพิจารณาตัวเลขเพียงการใช้การวิเคราะห์แนวโน้มเข้าสู่ส่วนกลางไม่ว่าจะเป็นการหาค่าเฉลี่ยเลขคณิต ค่ามัธยฐานหรือค่าฐานนิยมเป็นส่วนเดียวมิได้ ซึ่งจะทำให้การวิเคราะห์เกิดความคลาดเคลื่อน ฉะนั้นการวิเคราะห์การวัดการกระจายข้อมูลจึงเป็นพิจารณาใช้ร่วมด้วยกันกับการวิเคราะห์แนวโน้มเข้าสู่ส่วนกลางเพื่อให้ได้ข้อมูลที่มีความใกล้เคียงกับความเป็นจริงมากที่สุด

ดังนั้นการวิเคราะห์การวัดการกระจายข้อมูลต้องทำควบคู่ไปกับการวิเคราะห์แนวโน้มเข้าสู่ส่วนกลางเพื่อใช้ในการตรวจสอบข้อมูลและประกอบการตัดสินใจ โดยการวิเคราะห์ด้วยการวัดการกระจายข้อมูลมีหลายวิธี ดังนี้ 1.ค่าพิสัยและค่าสัมประสิทธิ์ค่าพิสัย (Range and Coefficient of Range) 2.ค่าเบี่ยงเบนเฉลี่ย (Mean Deviation) 3.ค่าความแปรปรวนและค่าความเบี่ยงเบนมาตรฐาน (Variance and Standard Deviation) 4.พิสัยควอไทล์ (Quartile Deviation) 5.ค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variance) (กัลยา, 2557: 68-73., Chikkodi, C.M., and Satyaprasad, B.G., 2009: 10.3-10.44., Lind, Douglas A., Marchal, William G. &Wathen, Samuel A., 2005: 66-68., Schindler, Thomas M., 2005: 72-84., Goos, Peter, and Meintrup, David.2015: 64-70)

1.ค่าพิสัยและค่าสัมประสิทธิ์ค่าพิสัย (Range and Coefficient of Range)

ค่าพิสัย (Range)เป็นเกณฑ์หาค่าการกระจายอย่างง่ายพิจารณาจากข้อมูลที่แตกต่างกันของข้อมูลที่มีค่ามากที่สุดกับข้อมูลที่มีค่าน้อยที่สุด (Goos, Peter, and Meintrup, David.2015: 65)

สูตร ค่าพิสัย = ค่าสูงสุด – ค่าต่ำสุด

การหาค่าพิสัยนั้นเป็นการพิจารณาเพียง 2 ค่าเท่านั้น คือ ค่าสูงสุดกับค่าต่ำสุด ขณะที่ข้อมูลอื่นๆ นอกเหนือ 2 ค่านี้จะไม่พิจารณาหรือไม่มีอิทธิพลใดๆ ทั้งสิ้น โดยการหาค่าพิสัยนิยมใช้ในโรงงานอุตสาหกรรมเป็นส่วนใหญ่เพื่อใช้ในการควบคุมกระบวนการผลิตสินค้าและบริการ

ค่าสัมประสิทธิ์พิสัย (Coefficient of Range) ค่าเกิดจากความแตกต่างระหว่างค่าสูงสุดกับค่าต่ำสุด หารด้วยค่าสูงสุดบวกค่าต่ำสุด ใช้ร่วมกับการหาค่าพิสัยเพื่อพิจารณาถึงความสัมพันธ์ระหว่างค่าสูงสุดกับค่าต่ำสุดของข้อมูล (Chikkodi, C.M., and Satyaprasad, B.G., 2009: 10.3)

สูตร ค่าสัมประสิทธิ์พิสัย $= \dfrac{\text{ค่าสูงสุด} - \text{ค่าต่ำสุด}}{\text{ค่าสูงสุด} + \text{ค่าต่ำสุด}}$

ตัวอย่างที่ 3.28 จากตัวอย่างที่ 3.27 จงหาค่าพิสัยและค่าสัมประสิทธิ์พิสัย

ปี 2557	8	8	8	8	8	8	8	8	8	8	8	8
ปี 2558	9	7	6	5	8	14	6	5	12	6	10	8
ปี 2559	3	4	5	6	7	8	9	10	10	11	11	12

วิธีทำ สูตรค่าพิสัย = ค่าสูงสุด – ค่าต่ำสุด

ปี 2557 ค่าพิสัย = 8 – 8 = 0

ปี 2558 ค่าพิสัย = 14 – 5 = 9

ปี 2559 ค่าพิสัย = 12 – 3 = 9

สูตรค่าสัมประสิทธิ์พิสัย $= \dfrac{\text{ค่าสูงสุด} - \text{ค่าต่ำสุด}}{\text{ค่าสูงสุด} + \text{ค่าต่ำสุด}}$

ปี 2557 ค่าสัมประสิทธิ์พิสัย $= \dfrac{8-8}{8+8} = 0$

ปี 2558 ค่าสัมประสิทธิ์พิสัย $= \dfrac{14-5}{14+5} = 0.47$

ปี 2559 ค่าสัมประสิทธิ์พิสัย $= \dfrac{12-3}{12+3} = 0.60$

จากการคำนวณทั้ง 3 ชุด พบว่า ปี 2557 ไม่มีค่าพิสัยและค่าสัมประสิทธิ์พิสัยมีค่าเท่ากับศูนย์
ปี 2558 มีค่าพิสัยเท่ากับ 9 และค่าสัมประสิทธิ์พิสัยเท่ากับ 0.47
ปี 2559 มีค่าพิสัยเท่ากับ 9 และค่าสัมประสิทธิ์พิสัยเท่ากับ 0.60

จากตัวอย่างที่ 3.27 ได้มีการคำนวณค่าเฉลี่ยเลขคณิตทั้ง 3 ชุด พบว่า ค่าเฉลี่ยเลขคณิตเท่ากับ 8 เปอร์เซ็นต์ต่อรายได้เท่ากันทั้ง 3 ชุด แต่เมื่อนำมาคำนวณหาค่าพิสัยและค่าสัมประสิทธิ์พิสัยจากตัวอย่างที่ 2.28 พบว่า ชุดที่ 1 ของปี 2557 ไม่มีค่าพิสัยและไม่มีการกระจายข้อมูลแต่อย่างใด ในขณะที่ชุดที่ 2 และชุดที่ 3 ของปี 2558 และปี 2559 มีค่าพิสัยเท่ากัน คือ 9 แต่เมื่อมาพิจารณาถึงค่าสัมประสิทธิ์พิสัยของการกระจายข้อมูล พบว่า ชุดที่ 2 ของปี 2558 มีค่าสัมประสิทธิ์พิสัย 0.47 บ่งบอกถึงการกระจายข้อมูลน้อยกว่า ชุดที่ 3 ของปี 2559 ที่มีค่าสัมประสิทธิ์พิสัย 0.60

ดังนั้นการหาค่าในแต่ละชุดข้อมูลมิสามารถหาเพียงค่าแนวโน้มเข้าสู่ส่วนกลางเพียงอย่างเดียว ต้องมีการพิจารณาถึงค่าการกระจายข้อมูลประกอบกับการวิเคราะห์อื่นๆ ด้วย

ข้อดีและข้อเสียของค่าพิสัย

ข้อดี 1. การคำนวณง่ายไม่ยุ่งยาก รวดเร็ว

2.ใช้ในการควบคุมคุณภาพเทคนิคต่างๆ ของโรงงานอุตสาหกรรมหรือใช้ในการเปรียบเทียบสภาพอากาศอย่างง่ายราคาส่วนแบ่งตลาดหรือค่าสูงต่ำของตลาดหุ้น

ข้อเสีย 1. สนใจเพียงข้อมูลที่มีค่าสูงสุดและค่าต่ำสุดทำให้ค่าระหว่างกลางไม่มีประโยชน์ซึ่งก่อให้เกิดความคลาดเคลื่อนในการวิเคราะห์ข้อมูลได้

2. การคำนวณเพื่อวิเคราะห์ข้อมูลด้วยการใช้สูตรค่าพิสัยไม่เพียงพอต่อการตอบคำถามในทางคณิตศาสตร์

3. ถึงแม้กลุ่มข้อมูลมีขนาดใหญ่ก็ตามแต่ก็ไม่เกิดผลต่อการวิเคราะห์ถึงการกระจายข้อมูลได้อย่างชัดเจน

2.ค่าเบี่ยงเบนเฉลี่ย (Mean Absolute Deviation: M.A.Dor M.D or A.D)

ค่าเบี่ยงเบนเฉลี่ยเป็นการคำนวณจากค่าเฉลี่ยเลขคณิตที่มีค่าสัมบูรณ์เท่านั้น โดยมิได้นำเครื่องหมายบวกหรือลบนำมาพิจารณาของค่าเฉลี่ยเลขคณิตแต่อย่างใด (Goos, Peter, and Meintrup, David.,2015: 65)

การหาค่าเบี่ยงเบนเฉลี่ย สามารถหาได้ 2 วิธีคือ (Chikkodi, C.M., and Satyaprasad, B.G., 2009: 10.11)

วิธีที่ 1 ข้อมูลที่ไม่จัดกลุ่มดังนี้

สูตร ค่าเบี่ยงเบนเฉลี่ยประชากร (M.D หรือδ) $= \frac{\sum_{i=1}^{N}|x_i - u_i|}{N} = \frac{\sum_{i=1}^{N}|x_i - Me|}{N} = \frac{\sum_{i=1}^{N}|x_i - z|}{N}$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนประชากร $= \frac{\delta}{u_i} = \frac{\delta}{Me} = \frac{\delta}{z}$

สูตร ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือδ) $= \frac{\sum_{i=1}^{N}|x_i - u_i|}{n} = \frac{\sum_{i=1}^{N}|x_i - Me|}{n} = \frac{\sum_{i=1}^{N}|x_i - z|}{n}$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนตัวอย่าง $= \frac{\delta}{\bar{x}} = \frac{\delta}{Me} = \frac{\delta}{z}$

ตัวอย่างที่ 3.29 จงหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างของข้อมูลตัวอย่างต่อไปนี้

45 52 68 92 87 73 64 59 81

วิธีทำ โจทย์ให้มา n = 9,

สูตร ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือδ) $= \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$

หาค่า $\bar{x} = \frac{\sum x_i}{n} = \frac{45+52+68+92+87+73+64+59+81}{9} = 69$

ตารางที่ 3.10 คำนวณหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์เบี่ยงเบนเฉลี่ยตัวอย่าง

ข้อมูลตัวเลข	$\boldsymbol{X_i - \bar{x}}$	$\boldsymbol{\lvert X_i - \bar{x}\rvert}$
45	45-69 = -24	24
52	52-69 = -17	17
68	68-69 = -1	1
92	92-69 = 23	23
87	87-69 = 18	18
73	73-69 = 4	4
64	64-69 = -5	5
59	59-69 = -10	10
81	81-69 = 12	12
		$\sum_{i=1}^{n}\lvert X_i - \bar{x}\rvert$ = 114

แทนค่าในสูตร ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือδ) $= \frac{\sum_{i=1}^{n}|X_i - \bar{x}|}{n} = \frac{114}{9} \approx 12.67$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง $= \frac{\delta}{\bar{x}} = \frac{12.67}{69} \approx 0.184$

สรุปได้ว่า ค่าเบี่ยงเบนเฉลี่ยตัวอย่างมีค่า 12.67 และมีค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างมีค่า 0.184 ซึ่งบ่งบอกถึงการกระจายของข้อมูลชุดนี้อยู่ที่ 12.67 ขณะที่สัมประสิทธิ์การกระจายของค่าเบี่ยงเบนเฉลี่ยระหว่างตัวเลขข้อมูลอยู่ที่ 0.184

ตัวอย่างที่ 3.30 จงหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่ามัธยฐานของข้อมูลตัวอย่างต่อไปนี้

14 25 61 39 36 54 47 22 66 58 43

วิธีทำ โจทย์ให้มาn = 11,เรียงลำดับจากน้อยไปมาก คือ 14, 22, 25, 36, 39, 43, 47, 54, 58, 61,66

หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยจากค่ามัธยฐาน ดังนั้น

Median = $\frac{n+1}{2} = \frac{11+1}{2} = \frac{12}{2} = 6$ ดังนั้นค่าอยู่ที่ลำดับที่ 6 คือ 43

ตารางที่ 3.11คำนวณหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์เบี่ยงเบนเฉลี่ยตัวอย่างจากค่ามัธยฐาน

ข้อมูลตัวเลข	$X_i - Me$	$\lvert X_i - Me \rvert$
14	14-43 = -29	29
22	22-43 = -21	21
25	25-43 = -18	18
36	36-43 = -7	7
39	39-43 = -4	4
43	43-43 = 0	0
47	47-43 = 4	4
54	54-43 = 11	11
58	58-43 = 15	15
61	61-43 = 18	18
66	66-43 = 23	23
		$\sum_{i=1}^{n} \lvert X_i - Me \rvert$ = 150

แทนค่าในสูตร ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือδ) $= \frac{\sum_{i=1}^{n}|X_i - Me|}{n} = \frac{150}{11} \approx 13.64$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง $= \frac{\delta}{Me} = \frac{13.64}{150} \approx 0.091$

สรุปได้ว่า ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่ามัธยฐานมีค่า 13.64 และมีค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่ามัธยฐานมีค่า 0.091 ซึ่งบ่งบอกถึงการกระจายของข้อมูลชุดนี้อยู่ที่ 13.64 ขณะที่สัมประสิทธิ์การกระจายของค่าเบี่ยงเบนเฉลี่ยระหว่างตัวเลขข้อมูลอยู่ที่ 0.091

ตัวอย่างที่ 3.31 จงหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่าฐานนิยมของข้อมูลตัวอย่างที่เกษตรกรมีการจับปลาได้ในแต่ละวันต่อไปนี้

จับปลาได้	56	89	34	43	62	74	23	ตัว
ความถี่การจับได้	12	36	41	28	29	16	25	ครั้ง

วิธีทำ โจทย์ให้มาn = 7 และในแต่ละช่วงมีความถี่ในการจับปลาซ้ำได้ต่างกัน

หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยจากค่าฐานนิยมที่มีความถี่มากที่สุด พบว่า จำนวนที่จับปลาได้ 34 ตัว มีความถี่ในการจับถึง 41 ครั้ง ดังนั้น จึงมีค่าฐานนิยมมากที่สุด คือ 34

ตารางที่ 3.12คำนวณหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์เบี่ยงเบนเฉลี่ยตัวอย่างจากค่าฐานนิยม

ข้อมูลตัวเลข	**ความถี่= f**	$X_i - z$	$\lvert X_i - z \rvert$	f * $\lvert X_i - z \rvert$
56	12	56-34= 22	22	264
89	36	89-34 = 55	55	1980
34	41	34-34 = 0	0	0
43	28	43-34 = 9	9	252
62	29	62-34 = 28	28	812
74	16	74-34 = 40	40	640
23	25	23-34 = -11	11	275
	n = 187			$\sum_{i=1}^{n} f * \lvert X_i - z \rvert = 4223$

แทนค่าในสูตร ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือδ) $= \frac{\sum_{i=1}^{n} f*|X_i - z|}{n} = \frac{4223}{187} \approx 22.58$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง $= \frac{\delta}{z} = \frac{22.58}{34} \approx 0.664$

สรุปได้ว่า ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่าฐานนิยมมีค่า 22.58 และมีค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่าฐานนิยมมีค่า 0.664 ซึ่งบ่งบอกถึงการกระจายของข้อมูลชุดนี้อยู่ที่ 22.58 ขณะที่สัมประสิทธิ์การกระจายของค่าเบี่ยงเบนเฉลี่ยระหว่างตัวเลขข้อมูลอยู่ที่ 0.664

วิธีที่ 2ข้อมูลที่จัดกลุ่ม ดังนี้

สูตร ค่าเบี่ยงเบนเฉลี่ยประชากร (M.D หรือδ)$= \frac{\sum_{i=1}^{N} f_i|X_i - u_i|}{\sum f_i} = \frac{\sum_{i=1}^{N} f_i|X_i - Me|}{\sum f_i} = \frac{\sum_{i=1}^{N} f_i|X_i - z|}{\sum f_i}$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนประชากร $= \frac{\delta}{u_i} = \frac{\delta}{Me} = \frac{\delta}{z}$

สูตร ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือδ) $= \frac{\sum_{i=1}^{N} f_i|X_i - u_i|}{\sum f_i} = \frac{\sum_{i=1}^{N} f_i|X_i - Me|}{\sum f_i} = \frac{\sum_{i=1}^{N} f_i|X_i - z|}{\sum f_i}$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนตัวอย่าง $= \frac{\delta}{\bar{x}} = \frac{\delta}{Me} = \frac{\delta}{z}$

ตัวอย่างที่ 3.32ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากข้อมูลที่ให้ต่อไปนี้

ช่วงเวลาทำงาน	5-8	9-12	13-16	17-20	21-24	ชั่วโมง
จำนวนคนงาน	15	11	8	23	5	คน

วิธีทำ โจทย์กำหนดให้5 ช่วง และให้ค่าความถี่แต่ละช่วงมา

ตารางที่ 3.12คำนวณหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์เบี่ยงเบนเฉลี่ยตัวอย่าง

ช่วงเวลาทำงาน	**ความถี่= f_i**	**ค่ากึ่งกลางของแต่ละชั้น i (X_i)**	**f_i* X_i**	**X_i - $\bar{x}$**	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert X_i - \bar{x}\rvert$
5-8	15	6.5	97.5	6.5-14 = -7.5	7.5	112.5
9-12	11	10.5	115.5	10.5-14 = -3.5	3.5	38.5
13-16	8	14.5	116	14.5-14 = 0.5	0.5	4.0
17-20	23	18.5	425.5	18.5-14 = 4.5	4.5	103.5
21-24	5	22.5	112.5	22.5-14 = 8.5	8.5	42.5
	$\sum f_i$ = n = 62		$\sum f_i X_i$= 867			$\sum f_i\lvert X_i - \bar{x}\rvert$ = 301

สูตร $\bar{x} = \frac{\sum f_i X_i}{n} = \frac{867}{62} \approx 13.98$ ปัดให้เป็นจำนวนเต็มเพื่อให้ง่ายในการคำนวณ จึงได้ตัวเลขเท่ากับ 14

แทนค่าในสูตรค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือ δ) $= \frac{\sum_{i=1}^{n} f_i\lvert X_i - \bar{x}\rvert}{\sum f_i} = \frac{301}{62} \approx 4.85$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนตัวอย่าง $= \frac{\delta}{\bar{x}} = \frac{4.85}{14} \approx 0.346$

สรุปได้ว่า ค่าเบี่ยงเบนเฉลี่ยตัวอย่างมีค่า 4.85 และมีค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างมีค่า 0.346 ซึ่งบ่งบอกถึงการกระจายของข้อมูลชุดนี้อยู่ที่ 4.85 ขณะที่สัมประสิทธิ์การกระจายของค่าเบี่ยงเบนเฉลี่ยระหว่างช่วงข้อมูล 0.346

ตัวอย่างที่ 3.33 จากข้อมูลตัวอย่างที่ 2.32ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่ามัธยฐานของข้อมูลที่ให้ต่อไปนี้

วิธีทำ ให้หาค่าด้วยการใช้วิธีมัธยฐาน โดยหาค่ากลางของข้อมูลด้วยสูตร $\frac{(n)}{2} = \frac{(62)}{2} = 31$ แสดงว่าข้อมูลมัธยฐานตกอยู่ที่ชั้นที่ 3 คือ ช่วงเวลาทำงาน 13-16 ชั่วโมง

ตารางที่ 3.12 คำนวณหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์เบี่ยงเบนเฉลี่ยตัวอย่างจากค่ามัธยฐาน

ช่วงเวลาทำงาน	**ความถี่ = f_i**	**ความถี่สะสม *CF***	**ค่ากึ่งกลางของแต่ละชั้น *i* (X_i)**	**X_i - Me**	$\lvert X_i - \boldsymbol{Me}\rvert$	$f_i\lvert x_i - \boldsymbol{Me}\rvert$
5-8	15	15	6.5	6.5-25 = -18.5	18.5	277.5
9-12	11	26	10.5	10.5-25 = -14.5	14.5	159.5
13-16	8	34	14.5	14.5-25 = -10.5	10.5	84.0
17-20	23	57	18.5	18.5-25 = -6.5	6.5	149.5
21-24	5	62	22.5	22.5-25 = -2.5	2.5	12.5
	$\sum f_i$ = n = 62					$\sum f_i\lvert x_i - \boldsymbol{Me}\rvert$ = 683

จากตารางที่ 3.12 จะได้ค่า L = 13, CF = 26, f_m = 8, I = 3 แทนค่าในสูตรได้

ค่ามัธยฐาน (Median) = $L + \frac{(\frac{n}{2} - CF)}{f_m}.\ I = 13 + \frac{(\frac{62}{2} - 26)}{8} \times 3 = 13 + \frac{31}{8} \times 3 = 13 + 11.63 = 24.63 \approx 25$

สูตร ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือ δ) = $\frac{\sum_{i=1}^{n} f_i\lvert X_i - Me\rvert}{\sum f_i} = \frac{683}{62} \approx 11.02$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนตัวอย่าง = $\frac{\delta}{Me} = \frac{11.02}{25} \approx 0.441$

สรุปได้ว่า ค่าเบี่ยงเบนเฉลี่ยตัวอย่างมีค่า 11.02 จากค่ามัธยฐานและมีค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างมีค่า 0.441จากค่ามัธยฐาน ซึ่งบ่งบอกถึงการกระจายของข้อมูลชุดนี้อยู่ที่ 11.02 ขณะที่สัมประสิทธิ์การกระจายของค่าเบี่ยงเบนเฉลี่ยระหว่างช่วงข้อมูล 0.441

ตัวอย่างที่ 3.34 จากข้อมูลตัวอย่างที่ 2.32ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่าฐานนิยมของข้อมูลที่ให้ต่อไปนี้

วิธีทำ ให้หาค่าด้วยการใช้วิธีฐานนิยม โดยค่าฐานนิยมอยู่ที่ชั้น 4 ระหว่างช่วงเวลาในการทำงาน 17-20 ชั่วโมงและมีค่าความถี่ของคนทำงานจำนวน 23 คน

ตารางที่ 3.13 คำนวณหาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์เบี่ยงเบนเฉลี่ยตัวอย่างจากค่าฐานนิยม

ช่วงเวลาทำงาน	ความถี่ = f_i	ค่ากึ่งกลางของแต่ละชั้น i (X_i)	$X_i - M_\circ$	$\lvert \boldsymbol{Xi} - M_\circ \rvert$	$f_i \lvert \boldsymbol{Xi} - M_\circ \rvert$
5-8	15	6.5	6.5-18 = -11.5	11.5	172.5
9-12	11	10.5	10.5-18 = -7.5	7.5	82.5
13-16	8	14.5	14.5-18 = -3.5	3.5	28.0
17-20	23	18.5	18.5-18 = 0.5	0.5	11.5
21-24	5	22.5	22.5-18 = 4.5	4.5	22.5
	$\sum f_i$ = n = 62				$\sum f_i \lvert \boldsymbol{Xi} - M_\circ \rvert$= 317

จากตารางที่ 3.13 จะได้ค่า $L_0 = 17$, $f_m = 23$, $f_1 = 8$, $f_2 = 5$, $I = 3$

แทนค่าฐานนิยม (Mode) $= L_0+[\frac{f_m - f_1}{f_m - f_1)+(f_m - f_2)} . I] = 17+[\frac{23-8}{(23-8)+(23-5)} \text{ x } 3] = 17+[\frac{15}{15+18}] \text{ x } 3$

$= 17 + 1.36 = 18.36 \approx 18$

สูตร ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง (M.D หรือ δ) $= \frac{\sum_{i=1}^{n} f_i \lvert x_i - z \rvert}{\sum f_i} = \frac{317}{62} \approx 5.11$

ค่าสัมประสิทธิ์ค่าเบี่ยงเบนตัวอย่าง $= \frac{\delta}{z} = \frac{\delta}{M_\circ} = \frac{5.11}{18} \approx 0.284$

สรุปได้ว่า ค่าเบี่ยงเบนเฉลี่ยตัวอย่างมีค่า 5.11 จากค่าฐานนิยมและมีค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างมีค่า 0.284 จากค่าฐานนิยม ซึ่งบ่งบอกถึงการกระจายของข้อมูลชุดนี้อยู่ที่ 5.11 ขณะที่สัมประสิทธิ์การกระจายของค่าเบี่ยงเบนเฉลี่ยระหว่างช่วงข้อมูล 0.284

3. ค่าความแปรปรวนและค่าความเบี่ยงเบนมาตรฐาน (Variance and Standard Deviation)

ก. ความแปรปรวน (Variance) เป็นค่าที่เกิดจากผลรวมของข้อมูลตัวเลขที่ศึกษาลบด้วยค่าเฉลี่ยเลขคณิต โดยทั้งหมดยกกำลังสองแล้วหารด้วยจำนวนข้อมูลทั้งหมดหรืออีกนัยหนึ่งคือ ค่าแปรปรวนเป็นส่วนหนึ่งเกิดจากค่าเฉลี่ยเลขคณิตภายใต้ค่าเฉลี่ยทั้งหมด (Lind, Douglas A., Marchal, William G. &Wathen, Samuel A., 2005: 74., Everitt, B. S., 2002: 388)

ค่าความแปรปรวนสามารถหาได้ 2 วิธี คือ (กัลยา, 2557: 70)

วิธีที่ 1 ข้อมูลที่ไม่จัดกลุ่ม

สูตร ค่าความแปรปรวนประชากร (σ^2) $= \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N} = \frac{\sum_{i=1}^{N}X_i{}^2-(\sum_{i=1}^{N}X_i{}^2\ )/N}{N}$

สูตรค่าความแปรปรวนตัวอย่าง (s^2) $= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}x_i{}^2\ -(\sum_{i=1}^{n}x_i{}^2\ )/n}{n-1}$

วิธีที่ 2 ข้อมูลจัดกลุ่ม

สูตร ค่าความแปรปรวนประชากร (σ^2) $= \frac{\sum_{i=1}^{N}f_i(X_i-\mu)^2}{N} = \frac{\sum_{i=1}^{N}f_iX_i{}^2\ -(\sum_{i=1}^{N}f_iX_i{}^2\ )/N}{N}$

สูตร ค่าความแปรปรวนตัวอย่าง (s^2) $= \frac{\sum_{i=1}^{n}f_i(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}f_ix_i{}^2\ -(\sum_{i=1}^{n}f_ix_i{}^2\ )/n}{n-1}$

โดยความหมาย

σ^2 = ความแปรปรวนประชากร

X_i = ข้อมูลที่เราศึกษาของประชากร

μ = ค่าเฉลี่ยเลขคณิตประชากร

f_i = ความถี่ของข้อมูล

N = จำนวนข้อมูลทั้งหมดที่ศึกษาของประชากร

s^2 = ความแปรปรวนตัวอย่าง

x_i = ข้อมูลที่เราศึกษาของตัวอย่าง

$\bar{x}$ = ค่าเฉลี่ยเลขคณิตตัวอย่าง

n = จำนวนข้อมูลทั้งหมดที่ศึกษาของตัวอย่าง

ตัวอย่างที่ 3.35 ข้อมูลต่อไปนี้จงหาความแปรปรวนของข้อมูลตัวอย่าง

32	45	56	89	75	42	18	34	22

วิธีทำ โจทย์กำหนดให้ข้อมูล n = 9

สูตรค่าความแปรปรวนตัวอย่าง (s^2) $= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}x_i{}^2\ -(\sum_{i=1}^{n}x_i{}^2\ )/n}{n-1}$

หาค่า $\bar{x} = \frac{\sum_{i=1}^{n}x}{n} = \frac{32+45+56+89+75+42+18+34+22}{9} = \frac{413}{9} \approx 46$

ตารางที่ 3.14 คำนวณหาค่าความแปรปรวนตัวอย่าง

จำนวน	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
32	32-46 = -14	$(-14)^2$ = 196
45	45-46 = -1	$(-1)^2$ = 1
56	56-46 = 10	$(10)^2$ = 100
89	89-46 = 43	$(43)^2$ = 1,849
75	75-46 = 29	$(29)^2$ = 841
42	42-46 = -4	$(-4)^2$ = 16
18	18-46 = -28	$(-28)^2$ = 784
34	34-46 = -12	$(-12)^2$ = 144
22	22-46 = -24	$(-24)^2$ = 576
		$\sum_{i=1}^{n}(x_i - \bar{x})^2$ = 4,507

แทนค่าในสูตร ค่าความแปรปรวนตัวอย่าง $(s^2) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{4,507}{9-1} = \frac{4,507}{8} \approx 563.38$

สรุปได้ว่า ค่าความแปรปรวนของกลุ่มตัวอย่างชุดนี้มีค่าเท่ากับ 563.38

ตัวอย่างที่ 3.36 จากตัวอย่างที่ 2.32 จงหาความแปรปรวนของข้อมูลตัวอย่าง

วิธีทำ สูตร ค่าความแปรปรวนตัวอย่าง $(s^2) = \frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{n-1}$

หาค่า $\bar{x} = \frac{\sum_{i=1}^{n} f_i M}{n} = \frac{867}{62} \approx 14$

ตารางที่ 3.15 คำนวณหาความแปรปรวนของข้อมูลตัวอย่าง

ช่วงเวลาทำงาน	**ความถี่ = f_i**	**ค่ากึ่งกลาง ($M = x_i$)**	$f_i M$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
5-8	15	$\frac{5+8}{2}$ = 6.5	15x6.5 = 97.5	6.5-14 = -7.5	$(-7.5)^2$ = 56.25	843.75
9-12	11	$\frac{9+12}{2}$ = 10.5	11x10.5 = 115.5	10.5-14 = -3.5	$(-3.5)^2$ = 12.25	134.75
13-16	8	$\frac{13+16}{2}$ = 14.5	8x14.5 = 116.0	14.5-14 = 0.5	$(-0.5)^2$ = 0.25	2.00
17-20	23	$\frac{17+20}{2}$ = 18.5	23x18.5 = 425.5	18.5-14 = 4.5	$(-4.5)^2$ = 20.25	465.75
21-24	5	$\frac{21+24}{2}$ = 22.5	5x22.5 = 112.5	22.5-14 = 8.5	$(-8.5)^2$ = 72.25	361.25
	$\sum f_i$ = n = 62		$\sum_{i=1}^{n} f_i M$ = 867			$\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$ = 1,807.50

แทนค่าในสูตร ค่าความแปรปรวนตัวอย่าง (s^2) $= \frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{n-1} = \frac{1,807.50}{62-1} \approx 29.63$

สรุปได้ว่า ค่าความแปรปรวนของข้อมูลตัวอย่างช่วงเวลาทำงาน 29.63 ชั่วโมง^2

ข. ความเบี่ยงเบนมาตรฐาน (Standard Deviation) เป็นค่าที่เกิดจากการนำค่าความแปรปรวนมาถอดรากที่สองเพื่อให้ได้ค่าความเบี่ยงเบนมาตรฐาน (กัลยา, 2557: 71., Urdan, Timothy C., 2005 : 16)

ความเบี่ยงเบนมาตรฐานมี 2 วิธี คือ

วิธีที่ 1 ข้อมูลที่ไม่จัดกลุ่ม

สูตร ค่าความเบี่ยงเบนมาตรฐานประชากร (σ) $= \sqrt{\frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N} X_i^2-(\sum_{i=1}^{N} X_i^2)/N}{N}}$

สูตรค่าความเบี่ยงเบนมาตรฐานตัวอย่าง (S) $= \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2-(\sum_{i=1}^{n} x_i^2)/n}{n-1}}$

วิธีที่ 2 ข้อมูลจัดกลุ่ม

สูตร ค่าความเบี่ยงเบนมาตรฐานประชากร (σ) $= \sqrt{\frac{\sum_{i=1}^{N} f_i(X_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N} f_i X_i^2-(\sum_{i=1}^{N} f_i X_i^2)/N}{N}}$

สูตร ค่าความเบี่ยงเบนมาตรฐานตัวอย่าง (S) $= \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} f_i x_i^2-(\sum_{i=1}^{n} f_i x_i^2)/n}{n-1}}$

4. พิสัยควอไทล์ (Inter-Quartile Deviation) เป็นค่าที่เกิดจากความต่างระหว่าง Q_3-Q_1 โดยค่า Q_3 แทนค่าควอไทล์บนและ Q_1 แทนค่าควอไทล์ล่าง ซึ่งคำนิยามของค่าพิสัยควอไทล์ Q_1 เป็นค่าควอไทล์ล่างของข้อมูลหนึ่งในสี่ที่ศึกษาจากข้อมูลทั้งหมด ขณะที่ค่าพิสัยควอไทล์ Q_3 เป็นค่าควอไทล์บนของข้อมูลสามในสี่ที่ศึกษาจากข้อมูลทั้งหมด (Shanmugam, Ramalingam and Chattamvelli, Rajan, 2015: 73)

สูตร พิสัยควอไทล์ = IQR = Q_3-Q_1

ตัวอย่างที่ 3.37 จงหาค่าควอไทล์บน ควอไทล์ล่างและค่าพิสัยควอไทล์ของข้อมูลต่อไปนี้

52 49 63 12 15 64 87 92 11 56 78 66

วิธีทำ เรียงลำดับจากน้อยไปหามากที่สุด คือ 11, 12, 15, 49, 52, 56, 63, 64, 66, 78, 87, 92

จำนวนข้อมูลทั้งหมด n = 12

หาค่าควอไทล์บน Q_3 ใช้สูตร 3(n+1)/4 = 3(12+1)/4 = 9.75 ซึ่งข้อมูลจะอยู่ระหว่างตัวเลขลำดับที่ 9 กับลำดับที่ 10 ดังนั้น เลือกลำดับต่ำสุด คือ ลำดับที่ 9 = 66

หาค่าควอไทล์ล่าง Q_1 ใช้สูตร (n+1)/4 = (12+1)/4 = 3.25 ซึ่งข้อมูลจะอยู่ระหว่างตัวเลขลำดับที่ 3 กับลำดับที่ 4 ดังนั้น เลือกลำดับที่สูงสุด คือ ลำดับที่ 4 = 49

สูตร พิสัยควอไทล์ = IQR = Q_3-Q_1 = 66-49 = 17

ดังนั้นสรุปได้ว่าค่า ควอไทล์บน Q_3 = 66 ค่าควอไทล์ล่าง Q_1 = 49 โดยค่าพิสัยควอไทล์เท่ากับ 17

5. ค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variance) เป็นการวัดค่าการกระจายที่ไม่มีหน่วยวัดข้อมูล ด้วยการนำค่าเบี่ยงเบนมาตรฐานหารด้วยค่าเฉลี่ยเลขคณิต (กัลยา, 2557: 72) โดยข้อดีของการหาค่าสัมประสิทธิ์ความแปรผัน กล่าวคือ ในการวิเคราะห์เปรียบเทียบกับข้อมูล 2 ประเภทที่มีหน่วยวัดต่างกันทำได้ยาก ดังนั้นการวิเคราะห์ด้วยสัมประสิทธิ์ความแปรผันสามารถก้าวข้ามหน่วยวัดต่างๆ ที่ต่างกันได้ ด้วยการวัดค่าการกระจายข้อมูลซึ่งทำให้เห็นถึงการกระจายของข้อมูลจากการเปรียบเทียบได้ชัดเจนขึ้น

สูตร ค่าสัมประสิทธิ์ความแปรผันของประชากร (C.V) = $\frac{\sigma}{\mu}$

ค่าสัมประสิทธิ์ความแปรผันของตัวอย่าง (C.V) = $\frac{SD}{\bar{X}}$

และทำให้อยู่ในรูปเปอร์เซ็นต์จะได้สูตร ดังนี้ ค่าสัมประสิทธิ์ความแปรผันของประชากร (C.V) = $\frac{\sigma}{\mu}*100$

ค่าสัมประสิทธิ์ความแปรผันของตัวอย่าง (C.V) = $\frac{SD}{\bar{X}}*100$

ตัวอย่างที่ 3.38 เปรียบเทียบการเกิดอุบัติเหตุของรถจักรยานยนต์กับรถยนต์ โดยอุบัติเหตุของรถจักรยานยนต์ในรอบเดือนที่ผ่านมาเป็นจำนวน 750 ครั้งต่อเดือน มีส่วนเบี่ยงเบนมาตรฐานเป็น 86 ขณะที่การเกิดอุบัติเหตุรถยนต์ในรอบเดือนเดียวกันเป็นจำนวน 245 ครั้งต่อเดือนและมีส่วนเบี่ยงเบนมาตรฐานเป็น 120 ตามลำดับ อยากทราบถึงการกระจายค่าสัมประสิทธิ์ความแปรผันของการเกิดอุบัติเหตุว่ามีลักษณะเป็นอย่างไร

วิธีทำ การเกิดอุบัติเหตุของรถจักรยานยนต์ μ = 750 ครั้งต่อเดือน, σ = 86

สูตร ค่าสัมประสิทธิ์ความแปรผันของประชากร (C.V) = $\frac{\sigma}{\mu}*100 = \frac{86}{750}*100 \approx 11.47$

การเกิดอุบัติเหตุของรถยนต์ μ = 245 ครั้งต่อเดือน, σ = 120

สูตร ค่าสัมประสิทธิ์ความแปรผันของประชากร (C.V) = $\frac{\sigma}{\mu}*100 = \frac{120}{245}*100 \approx 48.98$

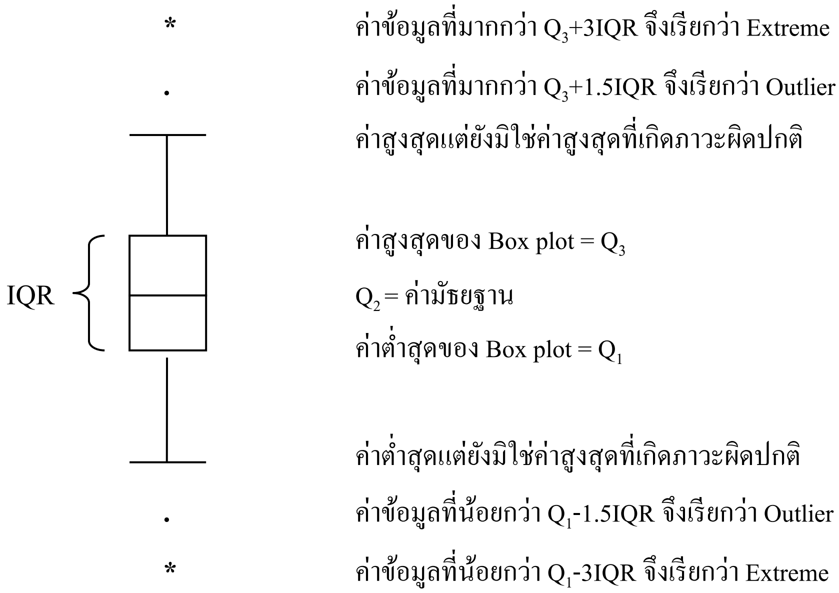
สรุปได้ว่าการเกิดอุบัติเหตุของรถยนต์มีการกระจายของค่าสัมประสิทธิ์ความแปรผันมากกว่าการเกิดอุบัติเหตุของรถจักรยานยนต์เกือบ 5 เท่าของการกระจาย

3. การวิเคราะห์ด้วยการใช้ Box plot (Box plot)

การวิเคราะห์ด้วยการใช้ Box plot นั้น เป็นการกระจายข้อมูลด้วยการสร้างเป็น Box plot เพื่อหาค่าที่ผิดปกติของข้อมูลการกระจายที่มีค่าสูงเกินหรือต่ำเกิน โดยข้อมูลนั้นจะมีการสร้างเป็นค่ามัธยฐาน เปอร์เซ็นต์ไทล์ที่ 25, 75 กระจายข้อมูลจากกึ่งกลางไปด้านบนและด้านล่าง (กัลยา, 2557: 73-74)

การวิเคราะห์ด้วย Box plot ต้องใช้สถิติ 5 ค่าด้วยกัน ดังนี้

1. ค่าต่ำสุดแต่ยังมิใช่ค่าต่ำสุดที่เกิดภาวะผิดปกติ จะมีค่าไม่ต่ำกว่า Q_1 + 1.5IQR
2. ควอไทล์ที่ 1 (Q_1)
3. ค่ามัธยฐานหรือค่าควอไทล์ที่ 2 (Q_2)
4. ควอไทล์ที่ 3 (Q_3)
5. ค่าสูงสุดแต่ยังมิใช่ค่าสูงสุดที่เกิดภาวะผิดปกติ จะมีค่าไม่เกิน Q_1 - 1.5IQR



รูปที่ 3.5 การกระจายข้อมูลด้วยการสร้างเป็น Box plot
ที่มา: กัลยา, 2557: 73

การวิเคราะห์ด้วย Box plot ต้องพิจารณากล่องสี่เหลี่ยมของ Box plot กล่าวคือ ถ้ากล่อง Box plot ทั้ง 2 เท่ากันแสดงว่ามีค่าเป็นสมมาตร ถ้ากล่อง Box plot ด้านบนมากกว่ากล่อง Box plot ด้านล่าง แสดงว่า การกระจายมีการเบ้ขวา ขณะที่ถ้ากล่อง Box plot ด้านบนน้อยกว่ากล่อง Box plot ด้านล่าง แสดงว่า การกระจายมีการเบ้ซ้าย ดังนั้น กล่อง Box plot มีนัยสำคัญที่บ่งบอกถึงการกระจายข้อมูลมากน้อยเพียงใดและมีแนวโน้มเบ้ไปในทิศทางใดของข้อมูลที่เราพิจารณา

สรุป

การคำนวณด้วยการใช้สถิติพรรณนาในเชิงปริมาณเป็นพื้นฐานของการแสดงข้อมูลเชิงประจักษ์ในการนำเสนอ โดยเฉพาะการนำเสนอข้อมูลอยู่ในรูปการวัดแนวโน้มเข้าสู่ส่วนกลางซึ่งมีหลายวิธี เช่น ค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยเรขาคณิต ค่าเฉลี่ยฮาร์โมนิค เป็นต้น โดยในการหาค่าวัดแนวโน้มเข้าสู่ส่วนกลางในแบบต่างๆ ก็มีข้อดี ข้อเสียแตกต่างกันในการนำเสนอข้อมูล นอกจากนั้นก็มีการหาค่ามัธยฐาน ค่าฐานนิยม เป็นต้น นอกเหนือในการวัดแนวโน้มเข้าสู่ส่วนกลางก็จะมีหลายวิธี เช่น การหาค่าเปอร์เซนต์ไทล์ การหาค่าควอไทล์และการหาค่าเดไซล์

ขณะที่การวัดการกระจายข้อมูลจะมีการนำเสนอไปพร้อมกับข้อมูลในการวัดแนวโน้มเข้าสู่ส่วนกลางด้วยเพื่อให้เกิดการนำเสนอข้อมูลให้สมบูรณ์ยิ่งขึ้น โดยการคำนวณหาข้อมูลในการวัดการกระจายนั้น ประกอบด้วย การหาค่าพิสัยและค่าสัมประสิทธิ์ค่าพิสัย ค่าเบี่ยงเบนเฉลี่ย ค่าความแปรปรวนและค่าความเบี่ยงเบนมาตรฐาน พิสัยควอไทล์และค่าสัมประสิทธิ์ความแปรผัน และอีกสิ่งหนึ่งที่สำคัญในการพิจารณาต่อการนำเสนอข้อมูล คือ การวิเคราะห์ด้วย Box Plot ที่สามารถทำให้เห็นถึงค่ากลางและการกระจายข้อมูลได้ชัดเจนในเชิงรูปภาพ

คำถามทบทวน

1. จงหาค่าเฉลี่ยเลขคณิตสำหรับข้อมูลที่ไม่ได้จัดกลุ่มด้วยข้อมูลต่อไปนี้

ก.	4	5	9	8	3	4	6
ข.	9	11	61	54	23	13	22
ค.	12	36	45	11	8	51	42

2. จงหาค่าเฉลี่ยเลขคณิตสำหรับข้อมูลที่จัดกลุ่มด้วยข้อมูลต่อไปนี้ เพื่อวิเคราะห์พฤติกรรมของนักศึกษาในการใช้ความถี่ของการใช้โทรศัพท์เคลื่อนที่

ขอบเขตจำนวนชั้น (จำนวนเวลาในการใช้โทรศัพท์เคลื่อนที่ต่อครั้งที่มีการใช้)	จำนวนครั้งในการใช้ (ความถี่ต่อวัน)
น้อยกว่า 1 นาที	30
1-5	45
6-10	62
11-15	56
16-20	72
21-25	26
26-30	10
มากกว่า 30 นาทีขึ้นไป	9
รวม	310

3. จงหาค่ามัธยฐานสำหรับข้อมูลที่ไม่ได้จัดกลุ่มด้วยข้อมูลต่อไปนี้

ก.	12	15	14	9	11	8
ข.	51	64	21	17	81	12
ค.	32	11	54	74	9	25

4. จงหาค่ามัธยฐานสำหรับข้อมูลที่จัดกลุ่มต่อไปนี้

ขอบเขตจำนวนชั้น (จำนวนเวลาในออกกำลังกายต่อวัน)	จำนวนประชากร (ความถี่)
น้อยกว่า 10 นาที	5
11-15	12
16-20	16
21-25	25
26-30	8
31-35	28
36-40	16
มากกว่า 40 นาทีขึ้นไป	10
รวม	120

5. จงหาค่าฐานนิยมสำหรับข้อมูลที่ไม่จัดกลุ่มต่อไปนี้

ก.	5	6	2	8	9	11	8	7
ข.	16	11	25	31	8	13	12	20
ค.	31	42	13	15	29	26	24	19

6. จงหาค่าฐานนิยมสำหรับข้อมูลที่จัดกลุ่มต่อไปนี้

ขอบเขตจำนวนชั้น (จำนวนคะแนนของนักศึกษาในการสอบ)	จำนวนนักศึกษา (ความถี่)
60-64	23
65-69	18
70-74	34
75-79	9
80-84	7
85-89	5
รวม	96

7. ให้หาค่าเปอร์เซนไทล์ที่ 80, ค่าควอไทล์ที่ 80 และค่าเดไซล์ที่ 80 จากข้อมูลต่อไปนี้

ก.	3.6	4.8	2.8	1.5	6.3	4.5
ข.	4.9	5.9	6.7	2.5	4.4	5.0
ค.	7.2	6.3	5.1	4.4	3.9	3.3

8. จงหาค่าพิสัยและค่าสัมประสิทธิ์พิสัย จากข้อมูลต่อไปนี้

ก.	2	6	5	4	9	11	12	8	7
ข.	13	11	12	8	19	17	21	7	5
ค.	56	45	48	55	71	39	42	64	33

9. ข้อมูลต่อไปนี้

45 62 58 43 68 35 51

ก. ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง

ข. ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่ามัธยฐาน

ค. ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่าฐานนิยม

10. จากข้อมูลต่อไปนี้

ช่วงเวลาการพักผ่อน	3-5	6-8	9-11	12-14	15-17	ชั่วโมง
จำนวนคน	14	25	8	5	3	คน

ก. ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่าง

ข. ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่ามัธยฐาน

ค. ให้หาค่าเบี่ยงเบนเฉลี่ยตัวอย่างและค่าสัมประสิทธิ์ค่าเบี่ยงเบนเฉลี่ยตัวอย่างจากค่าฐานนิยม

11. ข้อมูลต่อไปนี้

12 23 35 45 9 18 22 14 19 34 42

ก. ให้หาความแปรปรวนและค่าความเบี่ยงเบนมาตรฐาน

ข. ให้หาค่าควอไทล์บน ควอไทล์ล่างและค่าพิสัยควอไทล์

12. สถานีตำรวจบุรีรัมย์ได้ดำเนินการจับกุมผู้กระทำความผิดทางกฎหมายด้วยการฝ่าฝืนในการสวมหมวกนิรภัยในการที่จะก่อให้อุบัติเหตุ โดยในรอบเดือนที่ผ่านมาได้จับกุมและออกใบสั่งเป็นจำนวน 560 ราย มีส่วนเบี่ยงเบนมาตรฐานเป็น 62 โดยการเกิดอุบัติเหตุจากการไม่สวมหมวกนิรภัยในรอบเดือนเดียวกันเป็นจำนวน 310 รายและค่าเบี่ยงเบนมาตรฐานเป็น 89 อยากทราบถึงการกระจายค่าสัมประสิทธิ์ความแปรผันของการเกิดอุบัติเหตุจากการไม่สวมหมวกนิรภัยว่ามีลักษณะเป็นอย่างไร

เอกสารอ้างอิง

กัลยา วานิชย์บัญชา. (2557). **หลักสถิติ.** พิมพ์ครั้งที่ 14. กรุงเทพฯ : โรงพิมพ์สามลดา.

สำนักงานสถิติแห่งชาติ. (2552). **เครื่องชี้ภาวะเศรษฐกิจและสังคมไทยที่สำคัญ พ.ศ.2552 ด้านการเงิน-การธนาคาร: ดัชนีตลาดหลักทรัพย์และมูลค่าการซื้อขาย ณ สิ้นปี พ.ศ. 2542-2551.**[ออนไลน์]. แหล่งที่มา:http://service.nso.go.th/nso/nsopublish/download/files/socioSocIndicator.pdf. (วันที่ค้นข้อมูล. 18 สิงหาคม 2560).

สำนักงานสถิติแห่งชาติ. (2553). **เครื่องชี้ภาวะเศรษฐกิจและสังคมไทยที่สำคัญ พ.ศ.2553ด้านการเงิน-การธนาคาร: อัตราการแลกเปลี่ยนเงินตราต่างประเทศ สกุลที่สำคัญเทียบเท่าเงินไทย พ.ศ.2543-2552.** [ออนไลน์]. แหล่งที่มา: http://service.nso.go.th/nso/nsopublish/download/files/socioSoc53.pdf. (วันที่ค้นข้อมูล. 18 สิงหาคม 2560).

Bluman, Allan G., A brief version elementary statistics: A step by step approach. 4th ed., New York: McGraw-Hill, 2008.

Chikkodi, C.M., and Satyaprasad, B.G. (2009). **Business Statistics.** Mumbai, IN: Himalaya Publishing House. ProQuest ebrary. Web. 3 August 2017.

Everitt, B. S. (2002). **The Cambridge Dictionary of Statistics (2).** Cambridge University Press. ProQuest ebrary. Web 3 August 2017.

Foster, C. (2014). **Being mean about the mean.** Mathematics in School*,* 43(1), 32–33.

Goos, Peter, and Meintrup, David. (2015). **Statistics with JMP: Graphs, Descriptive Statistics and Probability (1).** New York, GB: John Wiley & Sons, Incorporated. ProQuestebrary. Web. 9 August 2017.

K. R. B. Jankowski and K.J.Flannelly. (2015). **Measures of Central Tendency in Chaplaincy, Health Care, and Related Research.** Journal of Health Care Chaplaincy, 21: 39-49.doi: 10.1080/08854726.2014.989799

Lind, Douglas A., Marchal, William G. &Wathen, Samuel A. (2005). **Statistical Techniques in Business & Economics. (12th ed.)**.New York, US: The McGraw-Hill Companies, Inc.

Panik, Michael J**. Statistical Inference: A Short Course (1).** Hoboken, US: Wiley, 2012. ProQuest ebrary. Web. 9 August 2017.

Schindler, Thomas M. (2015). **Meaning it! A Refresher on Mean, Median, and Mode.** AMWA Journal, 30(1), 31-33.

Shanmugam, Ramalingam and Chattamvelli, Rajan. (2015). **Statistics for Scientists and Engineers (1).** Hoboken, US: John Wiley & Sons, Incorporated. ProQuestebrary. Web 8 August 2017.

Urdan, Timothy C. (2005). Statistics in Plain English. Mahwah, US: Lawrence Eribaum. ProQuest ebrary. Web 3 August 2017.

Welkowitz, Joan, Cohen, Barry H., and Lea, R. Brooke**. Introductory Statistics for the Behavioral Sciences (7).** Hoboken, US: John Wiley & Sons, Incorporated, 2011. ProQuestebrary. Web. 8 August 2017.

Wilcox, Rand R. (2009). **Basic Statistics: Understanding Conventional Methods and Modern Insights.** New York, US: Oxford University Press. ProQuestebrary. Web 8 August 2017.