

การวัดการกระจายของข้อมูล

BUA3127 สถิติธุรกิจ

Dr.Khwanchat Wongjantip

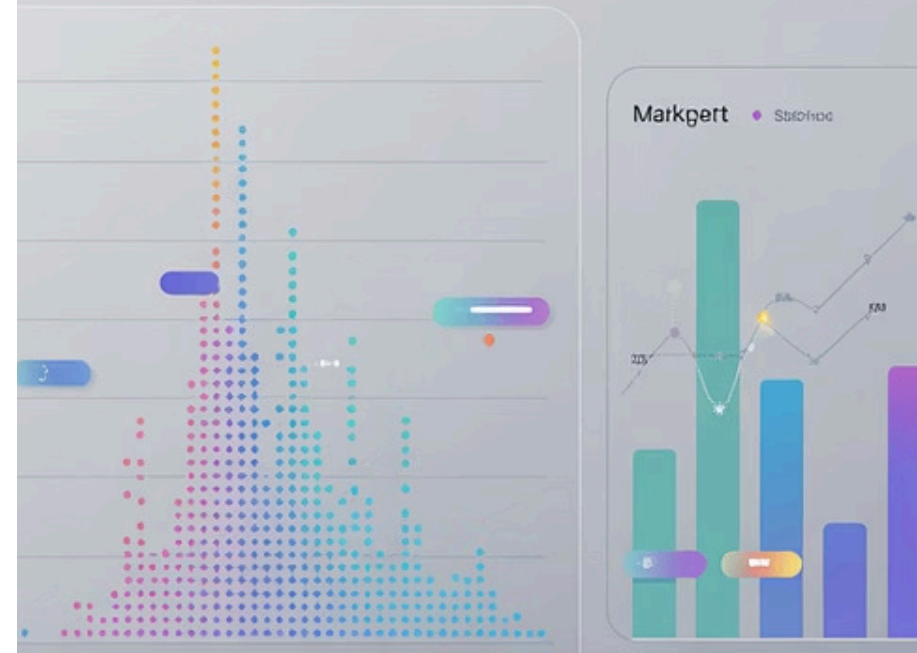
Reports

Integrations

Support

Dashboard

Data analytics platform



การวัดการกระจายของข้อมูล

การวัดการกระจายของข้อมูลเป็นการวัดความแตกต่างของข้อมูลแต่ละตัวในชุดข้อมูล ซึ่งแบ่งออกเป็น 2 ประเภทหลัก:

1

การวัดการกระจายแบบสัมบูรณ์

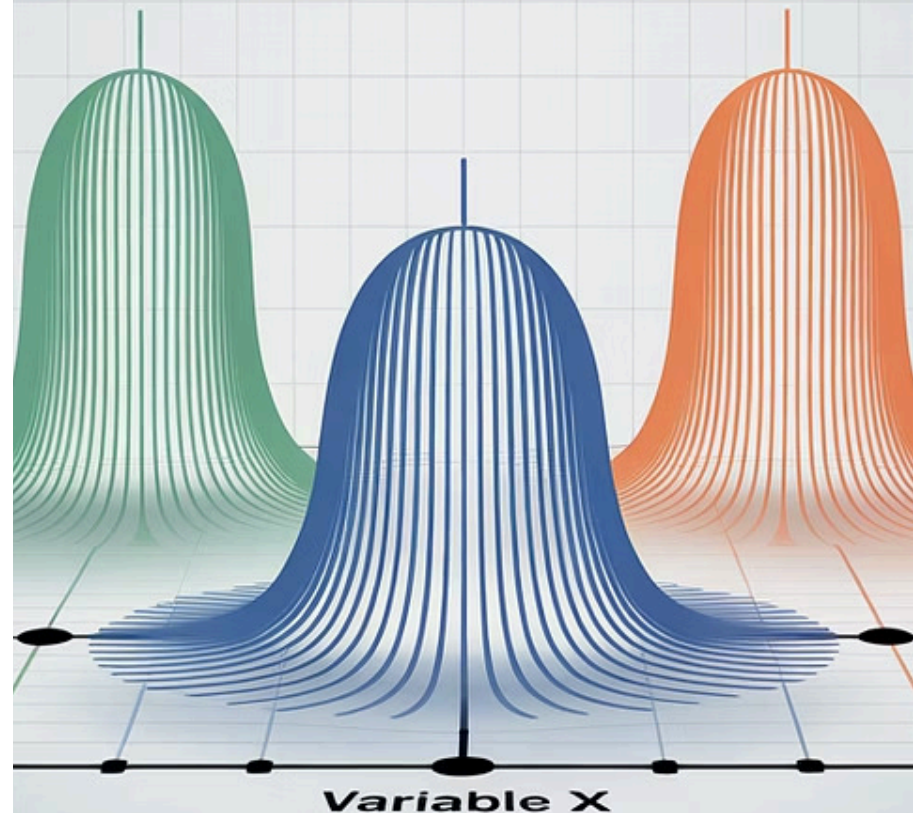
การวัดที่มีหน่วยเดียวกับข้อมูล เช่น พิสัย ส่วนเบี่ยงเบนควอไทล์ ส่วนเบี่ยงเบนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

2

การวัดการกระจายแบบสัมพัทธ์

การวัดที่ไม่มีหน่วย ใช้สำหรับเปรียบเทียบการกระจายระหว่างข้อมูลต่างชุดกัน เช่น สัมประสิทธิ์ของพิสัย สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์

Statistical Dispersion



พิสัย (Range)

พิสัยเป็นการวัดการกระจายที่ง่ายที่สุด โดยวัดจากความแตกต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูล

กรณีข้อมูลไม่ได้แจกแจงความถี่

$$R = x_{max} - x_{min}$$

เมื่อ x_{max} คือข้อมูลตัวที่มีค่ามากที่สุด และ x_{min} คือข้อมูลตัวที่มีค่าน้อยที่สุด

กรณีข้อมูลแจกแจงความถี่

$$R = x_{max} - x_{min}$$

เมื่อ x_{max} คือขอบบนของชั้นที่ข้อมูลตัวที่มีค่ามากที่สุด และ x_{min} คือขอบล่างของชั้นที่ข้อมูลตัวที่มีค่าน้อยที่สุด

ตัวอย่างการหาพิสัย

ตัวอย่าง 4.1

จงหาพิสัยของข้อมูลแต่ละชุดต่อไปนี้

ชุดที่ 1) 11, 18, 14, 15, 16, 19, 16

ชุดที่ 2) 10, 14, 17, 17, 11, 17, 11, 19

วิธีทำ

- 1. $R = 19 - 11 = 8$ ดังนั้นพิสัยของข้อมูลชุดที่ 1 เท่ากับ 8
- 2. $R = 19 - 10 = 9$ ดังนั้นพิสัยของข้อมูลชุดที่ 2 เท่ากับ 9

ตัวอย่าง 4.2

จงหาพิสัยของข้อมูลแจกแจงความถี่ต่อไปนี้

คะแนน	ความถี่
5 - 9	13
10 - 14	14
15 - 19	13
20 - 24	17

ขอบล่างของชั้นที่มีค่าน้อยที่สุด = 4.5

ขอบบนของชั้นที่มีค่ามากที่สุด = 24.5

$R = 24.5 - 4.5 = 20$

ส่วนเบี่ยงเบนควอไทล์ (Quartile Deviation)

ส่วนเบี่ยงเบนควอไทล์เป็นการวัดการกระจายโดยใช้ค่าควอไทล์ที่ 3 และควอไทล์ที่ 1 มาคำนวณ บางครั้งเรียกว่า พิสัยระหว่างควอไทล์ (Interquartile Range)

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

ขั้นตอนที่ 1

เรียงข้อมูลจากน้อยไปหามาก

ขั้นตอนที่ 3

หาตำแหน่งของควอไทล์ที่ 3 จากสูตร $3(N+1)/4$

ขั้นตอนที่ 2

หาตำแหน่งของควอไทล์ที่ 1 จากสูตร $(N+1)/4$

ขั้นตอนที่ 4

แทนค่าเพื่อหา Q.D. จากสูตร $(Q_3 - Q_1)/2$

ตัวอย่างการหาส่วนเบี่ยงเบนควอไทล์

ตัวอย่าง 4.3

จงหาพิสัยระหว่างควอไทล์ของข้อมูลต่อไปนี้

19, 18, 16, 18, 19, 19, 13, 18, 19, 19, 16, 18, 19, 14, 19, 16, 11, 19, 11, 18

วิธีทำ

จัดเรียงข้อมูลตามลำดับค่าจากน้อยไปหามาก:

11, 11, 13, 14, 16, 16, 16, 18, 18, 18, 18, 18, 19, 19,
19, 19, 19, 19, 19, 19

ในที่นี้ $N = 20$ ดังนั้น:

ตำแหน่งควอไทล์ที่ 1 คือ $(20+1)/4 = 5.25$

แสดงว่าควอไทล์ที่ 1 คือข้อมูลที่อยู่ระหว่างตัวที่ 5
และตัวที่ 6

ดังนั้น $Q_1 = 16$

ตำแหน่งของควอไทล์ที่ 3 คือ $3(20+1)/4 = 15.75$

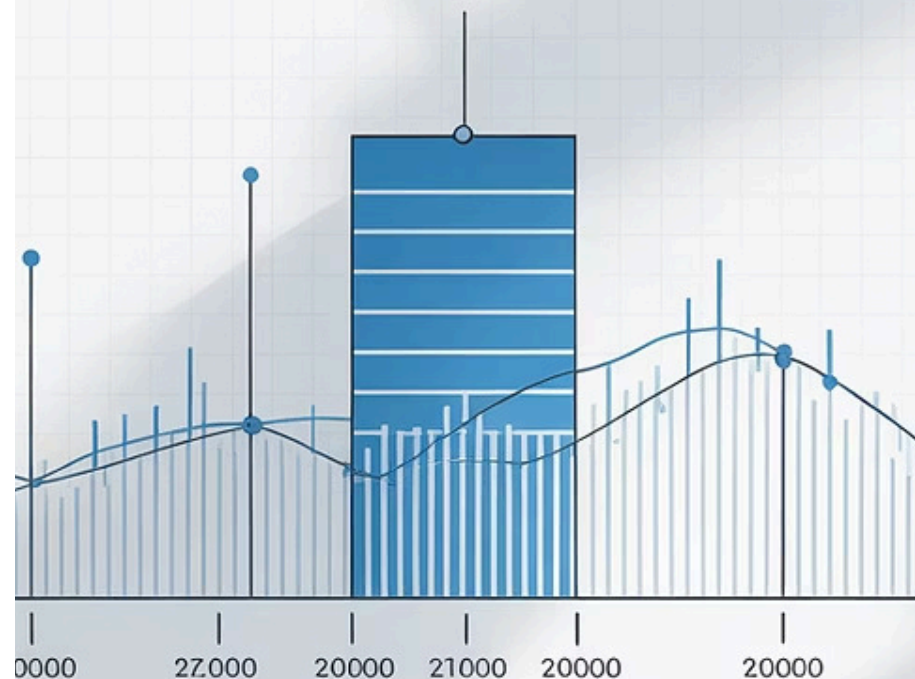
แสดงว่าควอไทล์ที่ 3 คือข้อมูลที่อยู่ระหว่างตัวที่ 15
และตัวที่ 16

ดังนั้น $Q_3 = 19$

ส่วนเบี่ยงเบนควอไทล์:

$$Q.D. = \frac{19 - 16}{2} = 1.5$$

BOX PLOT ANALYSIS



ส่วนเบี่ยงเบนควอไทล์ (กรณีข้อมูลแจกแจงความถี่)



เรียงข้อมูลจากน้อยไปหามาก

และหาความถี่สะสม



หาตำแหน่งของควอไทล์

ตำแหน่งของควอไทล์ที่ r คือ $Nr/4$ เมื่อ $r = 1, 3$

$f(x)$

คำนวณหาค่า Q_r

$$Q_r = L + I \left[\frac{\frac{Nr}{4} - \sum F_l}{f_m} \right]$$

$Q.D.$

คำนวณ $Q.D.$

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

เมื่อ L = ขีดจำกัดล่างที่แท้จริงของชั้นที่ Q_r อยู่, i = ความกว้างของชั้น, $\sum F_l$ = ความถี่สะสมของชั้นก่อนหน้าชั้นที่ Q_r อยู่, f_m = ความถี่ของชั้นที่ Q_r อยู่

ตัวอย่างการหาส่วนเบี่ยงเบนควอไทล์ (ข้อมูลแจกแจงความถี่)

ตัวอย่าง 4.4

คะแนน	ความถี่	ความถี่สะสม
5 - 9	13	13
10 - 14	14	27
15 - 19	13	40
20 - 24	17	57
25 - 29	16	73
30 - 34	14	87
35 - 39	12	99
40 - 44	13	112

วิธีทำ

ตำแหน่งควอไทล์ที่ 1 คือ $112(1)/4 = 28$

ตำแหน่งควอไทล์ที่ 3 คือ $112(3)/4 = 84$

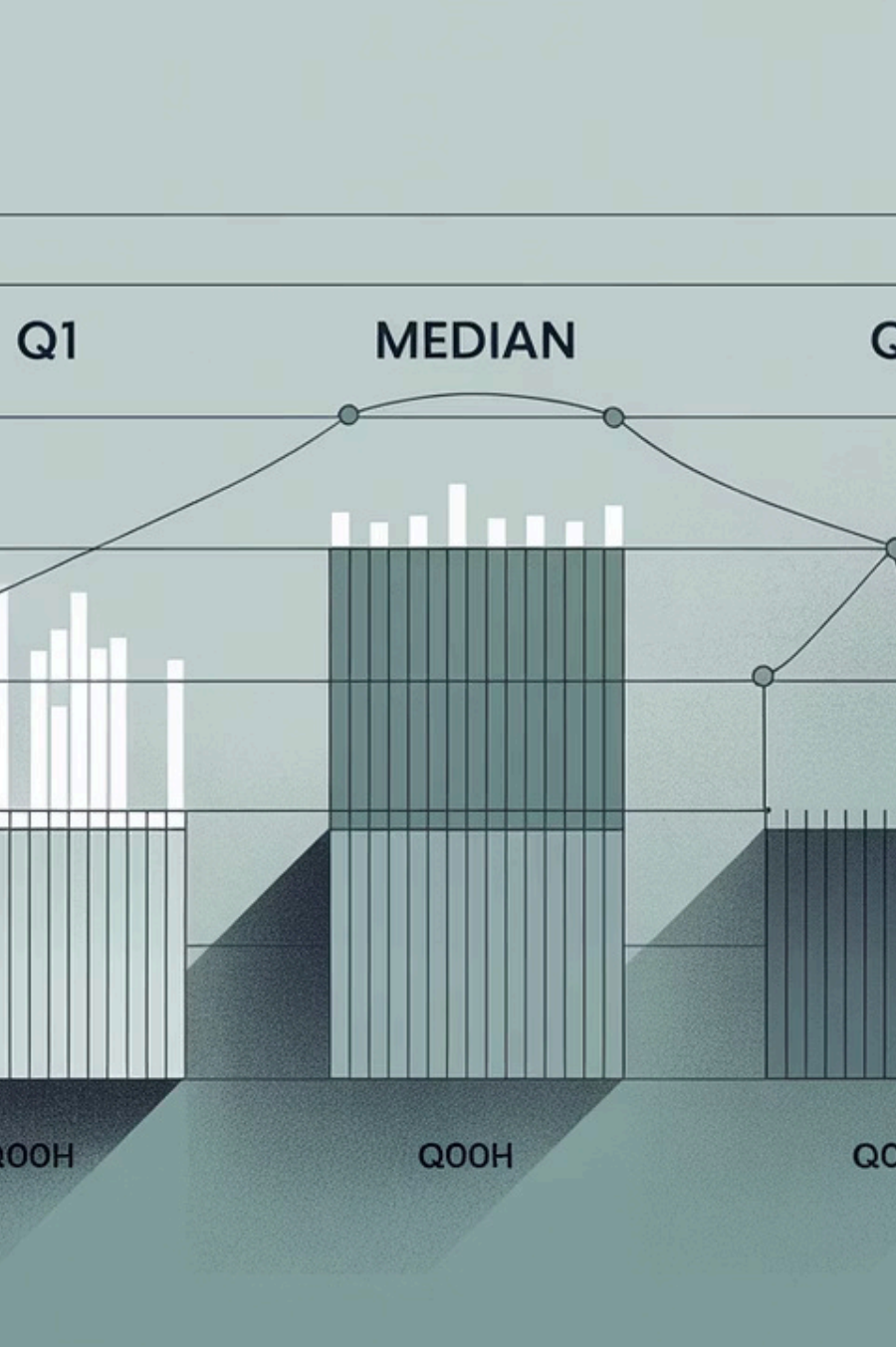
Q_1 อยู่ในชั้น 10-14 เนื่องจาก 28 อยู่ระหว่างความถี่สะสม 27 และ 40

$$Q_1 = 9.5 + 5 \left[\frac{28 - 27}{13} \right] = 9.5 + 0.38 = 9.88$$

Q_3 อยู่ในชั้น 30-34 เนื่องจาก 84 อยู่ระหว่างความถี่สะสม 73 และ 87

$$Q_3 = 29.5 + 5 \left[\frac{84 - 73}{14} \right] = 29.5 + 3.93 = 33.43$$

ดังนั้น $Q.D. = (33.43 - 9.88)/2 = 11.78$



ข้อสังเกตที่น่าสนใจเกี่ยวกับส่วนเบี่ยงเบนควอไทล์

การใช้ร่วมกับมัธยฐาน

หากนำเสนอค่ากลางของข้อมูลด้วยสถิติมัธยฐานแล้ว ควรใช้การวัดการกระจายของข้อมูลโดยวิธีพิสัยควอไทล์หรือส่วนเบี่ยงเบนควอไทล์

ข้อมูลอันตรภาคชั้นแบบเปิด

กรณีข้อมูลเป็นอันตรภาคชั้นแบบเปิด ควรใช้การวัดการกระจายของข้อมูลโดยวิธีพิสัยควอไทล์หรือส่วนเบี่ยงเบนควอไทล์

ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation)

ส่วนเบี่ยงเบนเฉลี่ยเป็นการวัดการกระจายที่ได้จากค่าเฉลี่ยของค่าสัมบูรณ์ (absolute value) ของความต่างระหว่างข้อมูลแต่ละตัวกับค่าเฉลี่ยของข้อมูลชุดนั้น

ส่วนเบี่ยงเบนเฉลี่ยของประชากร

กรณีข้อมูลไม่แจกแจงความถี่:

$$M.D. = \frac{\sum |x_i - \mu|}{N}$$

กรณีข้อมูลแจกแจงความถี่:

$$M.D. = \frac{\sum f_i |x_i - \mu|}{N}$$

ส่วนเบี่ยงเบนเฉลี่ยของตัวอย่าง

กรณีข้อมูลไม่แจกแจงความถี่:

$$M.D. = \frac{\sum |x_i - \bar{x}|}{n}$$

กรณีข้อมูลแจกแจงความถี่:

$$M.D. = \frac{\sum f_i |x_i - \bar{x}|}{n}$$

ตัวอย่างการหาส่วนเบี่ยงเบนเฉลี่ย

ตัวอย่าง 4.5

ข้อมูลชุดหนึ่งเป็น 12, 14, 16, 18, 20 จงหาค่าเบี่ยงเบนเฉลี่ย

วิธีทำ

- 1. หาค่าเฉลี่ย: $\bar{x} = (12 + 14 + 16 + 18 + 20)/5 = 16$
- 2. หาค่า M.D. โดยแทนสูตร:

$$M.D. = \frac{|12 - 16| + |14 - 16| + |16 - 16| + |18 - 16| + |20 - 16|}{5}$$
$$= \frac{4 + 2 + 0 + 2 + 4}{5} = \frac{12}{5} = 2.4$$

ตัวอย่าง 4.6

จงหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูลต่อไปนี้

คะแนน	ความถี่
5 - 9	13
10 - 14	14
15 - 19	13
20 - 24	17

หาค่าเฉลี่ย: $\bar{x} = 854/57 = 14.98$

หาค่า M.D. = $291.06/57 = 5.11$

สรุปได้ว่าข้อมูลกลุ่มตัวอย่างมีค่าการกระจายเท่ากับ 5.11

ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

ส่วนเบี่ยงเบนมาตรฐานเป็นการวัดการกระจายที่นิยมใช้กันมากที่สุด เพราะเป็นการคำนวณมาจากข้อมูลทุกตัวที่มีอยู่

ความแปรปรวน (Variance) หมายถึง ส่วนเบี่ยงเบนมาตรฐานยกกำลังสอง

1

กรณีข้อมูลไม่ได้แจกแจงความถี่

ความแปรปรวนของประชากร:

$$\sigma^2 = \frac{\sum X_i^2}{N} - \mu^2$$

ส่วนเบี่ยงเบนมาตรฐานของประชากร:

$$\sigma = \sqrt{\frac{\sum X_i^2}{N} - \mu^2}$$

2

กรณีข้อมูลแจกแจงความถี่

ความแปรปรวนของประชากร:

$$\sigma^2 = \frac{\sum f_i X_i^2}{N} - \mu^2$$

ส่วนเบี่ยงเบนมาตรฐานของประชากร:

$$\sigma = \sqrt{\frac{\sum f_i X_i^2}{N} - \mu^2}$$



ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

1

กรณีข้อมูลไม่ได้แจกแจงความถี่

ความแปรปรวนของตัวอย่าง:

$$S^2 = \frac{\sum X_i^2}{n} - \bar{x}^2$$

ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง:

$$s = \sqrt{\frac{\sum X_i^2}{n} - \bar{x}^2}$$

2

กรณีข้อมูลแจกแจงความถี่

ความแปรปรวนของตัวอย่าง:

$$S^2 = \frac{\sum f_i X_i^2}{n} - \bar{x}^2$$

ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง:

$$s = \sqrt{\frac{\sum f_i X_i^2}{n} - \bar{x}^2}$$

ตัวอย่างการหาส่วนเบี่ยงเบนมาตรฐาน

ตัวอย่าง 4.7

ข้อมูลชุดหนึ่งเป็น 12, 14, 16, 18, 20 จงหาค่าเบี่ยงเบนมาตรฐาน

วิธีทำ

- หาค่าเฉลี่ย: $\bar{x} = (12 + 14 + 16 + 18 + 20)/5 = 16$
- หาค่า $\sum X^2_i = 12^2 + 14^2 + 16^2 + 18^2 + 20^2 = 1320$
- หาค่า S^2 จาก $S^2 = (\sum X^2_i - n(\bar{x})^2)/(n-1) = (1320 - 5(16)^2)/(5-1) = 10$
- ค่าส่วนเบี่ยงเบนมาตรฐาน $S = \sqrt{10} = 3.16$

ตัวอย่าง 4.8

จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลต่อไปนี้

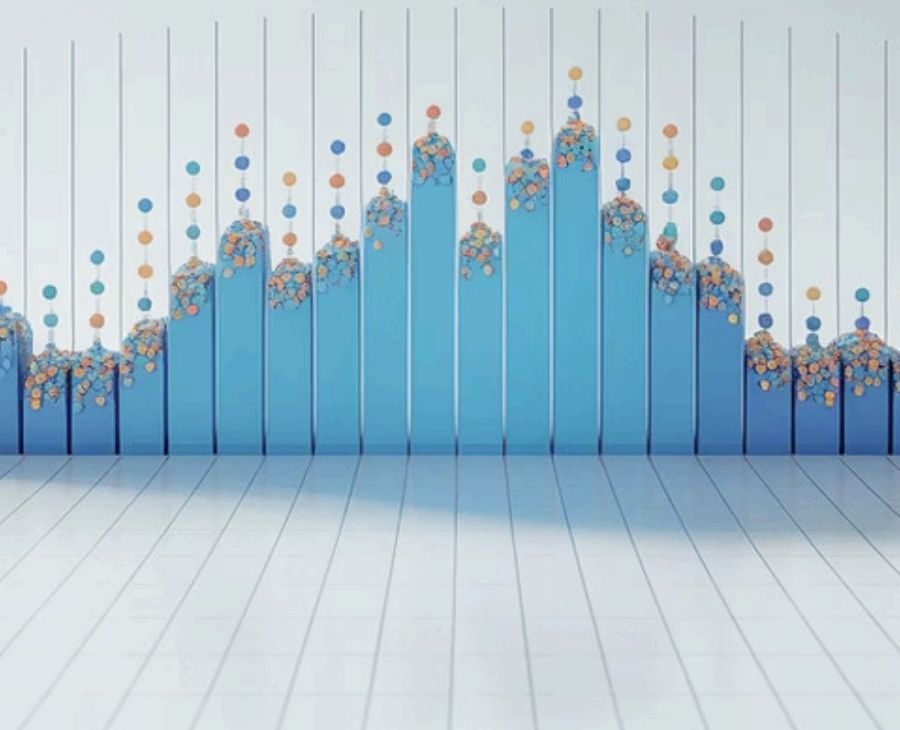
คะแนน	ความถี่
5 - 9	13
10 - 14	14
15 - 19	13
20 - 24	17

หาค่าเฉลี่ย: $\bar{x} = 854/57 = 14.98$

หาค่า $S^2 = (14638 - 57(14.98)^2)/(57-1) = 32.99$

ค่าส่วนเบี่ยงเบนมาตรฐาน $S = \sqrt{32.99} = 5.74$

STANDARD DEVIATION



ข้อสังเกตเกี่ยวกับส่วนเบี่ยงเบนมาตรฐาน

❗ ผลบวกของส่วนเบี่ยงเบนของข้อมูลแต่ละตัวกับค่าเฉลี่ยจะมีค่าเท่ากับ 0 เสมอ

กรณีข้อมูลไม่แจกแจงความถี่

$$\sum (x - \mu) = 0$$

$$\sum (x - \bar{x}) = 0$$

กรณีข้อมูลแจกแจงความถี่

$$\sum f(x - \mu) = 0$$

$$\sum f(x - \bar{x}) = 0$$

การวัดการกระจายแบบสัมพัทธ์

การวัดการกระจายสัมพัทธ์เป็นการวัดการกระจายที่ใช้เปรียบเทียบการกระจายซึ่งกันและกันระหว่างข้อมูลแต่ละชุด อันเป็นการจัดหน่วยของการวัดและอยู่ในภาวะที่สามารถนำมาเปรียบเทียบกันได้

1

สัมประสิทธิ์ของพิสัย

$$C.R. = \frac{X_{max} - X_{min}}{X_{max} + X_{min}}$$

2

สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์

$$C.Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

3

สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย

$$C.A. = \frac{M.D.}{\bar{X}}$$

4

สัมประสิทธิ์การแปรผัน

$$C.V. = \frac{S}{\bar{X}}$$



การนำค่าการกระจายข้อมูลไปใช้ในการนำเสนอผลการวิจัย

ในการวิจัย เมื่อเราวิเคราะห์ข้อมูลโดยใช้การวัดแนวโน้มสู่ส่วนกลางด้วย $\bar{x}$ และการวัดการกระจายด้วย S นั้น ในขั้นตอนแปลความหมายจะแปลความหมายเฉพาะ $\bar{x}$ เท่านั้น จะไม่นิยามแปลความหมายของ S เว้นเสียแต่ถ้า $S = 0$

ในกรณีที่ค่าของ S สูงมากๆ จะเป็นการบ่งบอกว่าค่า $\bar{x}$ ของข้อมูลชุดนี้ไม่เหมาะที่จะเป็นตัวแทนของข้อมูลชุดนี้ได้ ถ้าเป็นเช่นนี้ผู้วิจัยควรนำเสนอเป็นข้อมูลดิบตามที่ผู้ตอบแบบสอบถามให้มา

ค่าของ S ที่เขียนไว้ในตารางนำเสนอเป็นการบ่งบอกว่ากลุ่มผู้ตอบแบบสอบถามมีความเห็นใกล้เคียงกัน หรือมีความเห็นแตกต่างกันเท่านั้น

หากค่า $S = 0$ ควรหมายเหตุไว้ใต้ตารางการนำเสนอด้วย เช่น กรณีศึกษาความคิดเห็นแสดงว่าเกิดจากทุกคนมีความเห็นตรงกัน หรืออาจเกิดจากมีผู้แสดงความคิดเห็นต่อเรื่องนั้นเพียงคนเดียว

ตัวอย่างการนำเสนอข้อมูลในงานวิจัย

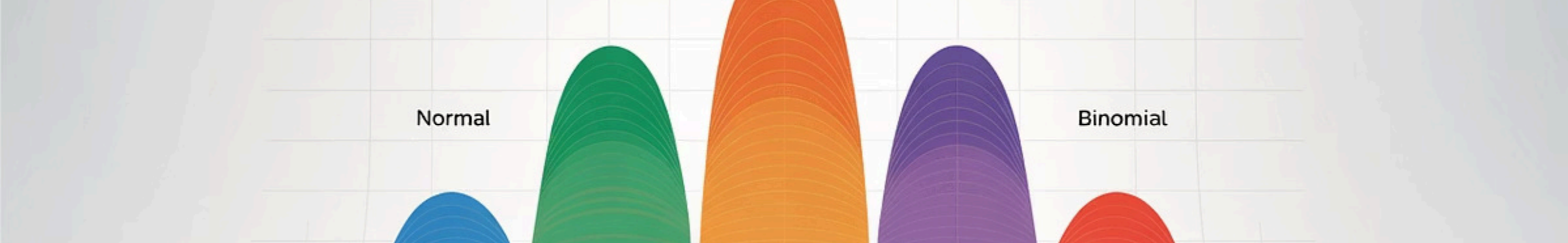
การแปลความหมายค่าเฉลี่ยในงานวิจัย (ตัวอย่างเกณฑ์)

ระดับความพึงพอใจ

- 4.51-5.00 = มากที่สุด
- 3.51-4.50 = มาก
- 2.51-3.50 = ปานกลาง
- 1.51-2.50 = น้อย
- 1.00-1.50 = น้อยที่สุด

ตัวอย่างตารางนำเสนอข้อมูล

รายการ ประเมิน	Mean	S.D.	ระดับ
ด้านที่ 1	4.78	0.44	มากที่สุด
ด้านที่ 2	4.33	0.58	มาก
ด้านที่ 3	4.67	0.58	มากที่สุด



สรุปความสำคัญของการวัดการกระจายข้อมูล

ค่า S ต่ำ

แสดงว่าข้อมูลมีการกระจายน้อย ค่าเฉลี่ยสามารถเป็นตัวแทนของข้อมูลได้ดี ผู้ตอบแบบสอบถามมีความคิดเห็นสอดคล้องกัน

ค่า S สูง

แสดงว่าข้อมูลมีการกระจายมาก ค่าเฉลี่ยอาจไม่เหมาะสมที่จะเป็นตัวแทนของข้อมูล ผู้ตอบแบบสอบถามมีความคิดเห็นแตกต่างกันมาก

ค่า S = 0

แสดงว่าข้อมูลไม่มีการกระจายเลย ทุกค่าเท่ากันหมด อาจเกิดจากผู้ตอบแบบสอบถามทุกคนมีความเห็นตรงกัน หรือมีผู้ตอบเพียงคนเดียว

การวัดการกระจายของข้อมูลมีความสำคัญในการวิเคราะห์ข้อมูลทางสถิติ ช่วยให้เราเข้าใจลักษณะของข้อมูลได้ดียิ่งขึ้น และสามารถตัดสินใจเลือกใช้สถิติที่เหมาะสมในการวิเคราะห์ข้อมูลต่อไป