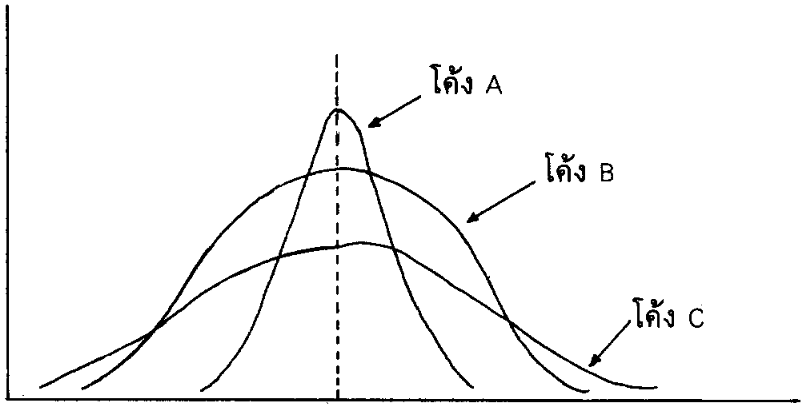
4. มาตราวัดการกระจาย

1. มาตราวัดการกระจาย
2. มาตราวัดการกระจายโดยใช้ระยะทาง
3. การหาส่วนเบี่ยงเบนเฉลี่ย และความแปรปรวน
4. สัมประสิทธิ์ความแปรปรวน

1. มาตราวัดการกระจาย



ลองพิจารณาโค้ง A, B, C ในภาพ จะเห็นว่า A มีการแผ่ขยายหรือความผันแปรน้อยกว่า B, และโค้ง B มีความผันแปรน้อยกว่า C แต่ทั้ง 3 โค้ง มีจุดกึ่งกลางอยู่ที่เดียวกัน ดังนั้น หากเราหาเฉพาะค่าวัดแนวโน้มเข้าสู่ส่วนกลาง อาทิเช่น ค่าเฉลี่ย ค่ามัธยฐาน หรือค่าฐานนิยม เมื่อเห็นว่าอยู่ที่เดียวกัน อาจทำให้เราเข้าใจผิดในสาระสำคัญบางประการว่าประชากรทั้ง 3 นี้ไม่ต่างกัน ทั้งนี้เป็นเพราะเราลืมวัดลักษณะที่สำคัญอีกอย่างหนึ่ง นั่นคือการกระจายของข้อมูล ประโยชน์ที่จะได้รับคือ

1. จะทำให้เราสามารถพิจารณาได้ว่าจะให้ความเชื่อถือกับค่าแนวโน้มเข้าสู่ส่วนกลางได้มากน้อยเพียงไร เช่น ค่าเฉลี่ยของโค้ง A จะทำหน้าที่เป็นตัวแทนได้ดีกว่าค่าเฉลี่ยของโค้ง C
2. เป็นมาตราซึ่งใช้วัดเปรียบเทียบการกระจายระหว่างกลุ่ม

2. มาตราวัดการกระจายโดยใช้ระยะทาง

เราอาจวัดการกระจายโดยหาผลต่างระหว่างค่า 2 ค่า ซึ่งเป็นการวัดระยะทางได้แก่ พิสัย interfractile range และ ส่วนเบี่ยงเบนควอไตล์

พิสัย คือ ผลต่างของข้อมูลที่มีค่าสูงสุดกับค่าต่ำสุด

Interfractile range หมายถึงส่วนหนึ่งของข้อมูลที่อยู่ใต้ fractile เช่น ค่ามัธยฐาน เป็น .5 fractile เพราะมีข้อมูลอยู่ครึ่งหนึ่งที่มีค่าน้อยกว่ามัธยฐาน นั่นคือ fractile ก็คือ เปอร์เซนต์นั่นเอง เช่นจะมี 25% ของข้อมูลที่มีค่าต่ำกว่า .25 fractile ส่วน Interfractile จะวัดการแผ่กระจายระหว่าง fractile 2 ค่า fractile ที่แบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน เรียกว่า decile ส่วน quartile จะแบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน ในขณะที่ percentile จะแบ่งข้อมูลออกเป็น 100 ส่วนเท่า ๆ กัน

พิสัยของ Interquartile $= Q_3 - Q_1$

ส่วนเบี่ยงเบนควอไตล์ $= \dfrac{(Q_3 - Q_1)}{2}$

ตัวอย่าง สมมุติข้อมูลกลุ่มหนึ่งมีเลขที่ จัดเรียงลำดับแล้ว 12 ตัว ดังนี้.

863	903	957	$\boxed{1041}$	← $\frac{1}{3}$	fractile
1,138	1,204	1,354	$\boxed{1,624}$	← $\frac{2}{3}$	fractile
1,698	1,745	1,802	1,883		

การหา $\frac{1}{3}$ และ $\frac{2}{3}$ fractile ต้องแบ่งข้อมูลเป็น 3 ส่วนเท่า ๆ กันก่อน และจะหาค่า **quartile-** โดยแบ่งข้อมูลเป็น 4 ส่วน ดังนี้

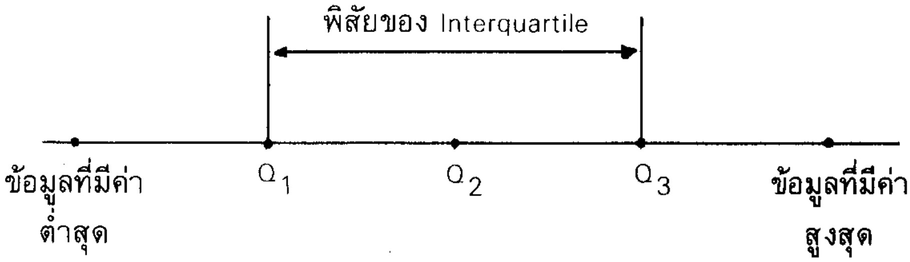
863	903	$\boxed{957}$	← quartile ที่ 1 หรือ Q_1
1,041	1,138	$\boxed{1,204}$	← quartile ที่ 2 หรือ Q_2 = Median
1,354	1,624	$\boxed{1,698}$	← quartile ที่ 3 หรือ Q_3
1,745	1,802	$\boxed{1,883}$	← quartile ที่ 4 หรือ Q_4

พิสัย ของ Interquartile $= Q_3 - Q_1$

$= 1698 - 957 = 741$

ส่วนเบี่ยงเบนควอไตล์ $= (Q_3 - Q_1)/2$

$= 741/2 = 370.5$



รูปแสดงพิสัยของ Interquartile

3. การวัดการกระจายโดยหาส่วนเบี่ยงเบนเฉลี่ย และความแปรปรวน

ส่วนเบี่ยงเบนเฉลี่ยของประชากร $= \dfrac{\Sigma|X - \mu|}{N}$

ส่วนเบี่ยงเบนเฉลี่ยของตัวอย่าง $= \dfrac{\Sigma|x - \bar{x}|}{n}$

ความแปรปรวนของประชากร

$$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$$

ส่วนเบี่ยงเบนมาตรฐานของประชากร หรือ populatim standard deviation

$$\sigma^2 = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$

ความแปรปรวนของตัวอย่าง

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{\Sigma x^2 - n\bar{x}^2}{n - 1} = \frac{\Sigma x^2 - (\Sigma x)^2|n}{n - 1}$$

ค่าเบี่ยงเบนมาตรฐานของตัวอย่าง หรือ standard error

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{\Sigma x^2 - n\bar{x}^2}{n - 1}} = \sqrt{\frac{\Sigma x^2 - (\Sigma x)^2|n}{n - 1}}$$

ประโยชน์ของส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐานจะทำให้ทราบว่ามีข้อมูลจำนวนเท่าใดที่อยู่ห่างจากค่ากึ่งกลาง ดังเช่น คุณสมบัติของโค้งปกติ เมื่อทราบค่าส่วนเบี่ยงเบนมาตรฐาน (σ) จะทราบได้ทันทีว่า มี 68% ของข้อมูลอยู่ระหว่าง 1σ ของค่าเฉลี่ย มีข้อมูลอยู่ 95% ที่อยู่ระหว่าง 2σ ของค่าเฉลี่ย และอีก 99% อยู่ระหว่าง 3σ ของค่าเฉลี่ย สำหรับข้อมูลที่ยังไม่ทราบว่ามีการแจกแจงแบบใด ได้มีนักคณิตศาสตร์ชาวรัสเซีย ชื่อ P.L. Chebyshev (1821-1894) ได้คิดทฤษฎีความน่าจะเป็น เรียกว่า Chebyshev's Theorem ดังนี้

"ไม่ว่าข้อมูลจะมีรูปร่างการแจกแจงเป็นแบบใดก็ตาม จะมีอย่างน้อยที่สุด 75% ของข้อมูลอยู่ในช่วง 2σ จากค่าเฉลี่ย และจะมีอย่างน้อยที่สุด 89% ของข้อมูลอยู่ในช่วง 3σ จากค่าเฉลี่ย" นอกจากนี้ เรายังใช้ส่วนเบี่ยงเบนมาตรฐานสำหรับเปลี่ยนค่าตัวแปร X ให้เป็นตัวแปรมาตรฐาน โดยใช้สูตร $Z = (X - \mu)/\sigma$

4. สัมประสิทธิ์ความแปรปรวน (Coefficient of Variation)

ถ้า $\sigma_1 = 100$ และ $\sigma_2 = 10$ ไม่ได้หมายความว่า σ_1 ใหญ่เป็น 10 เท่าของ σ_2 ในการนำส่วนเบี่ยงเบนมาตรฐานของ 2 กลุ่มมาเปรียบเทียบกัน จะต้องพิจารณาค่าเฉลี่ยและหน่วยที่ใช้วัดข้อมูลด้วย เช่น อาจใช้หน่วยต่างกัน เช่น σ_1 เป็นส่วนเบี่ยงเบนของค่าใช้จ่ายจึงมีหน่วยเป็นบาท ส่วน σ_2 คือส่วนเบี่ยงเบนของเวลาทำงานจึงมีหน่วยเป็นชั่วโมง และสมมุติว่า $\mu_1 = 1500$ บาท และ $\mu_2 = 50$ ชั่วโมง จะหาค่าสัมประสิทธิ์ความแปรปรวน ดังนี้

$$\text{Coefficient of Variation} = CV = \frac{\sigma}{\mu}(100)$$

ดังนั้น

$$CV_1 = \frac{100}{1500}(100) = 6.67\%$$

$$CV_2 = \frac{10}{50}(100) = 20\%$$

จะเห็นว่า CV_1 กลับเล็กกว่า CV_2 และพึงสังเกตว่า CV เป็นอิสระจากหน่วย จึงใช้เปรียบเทียบระหว่างกลุ่มได้