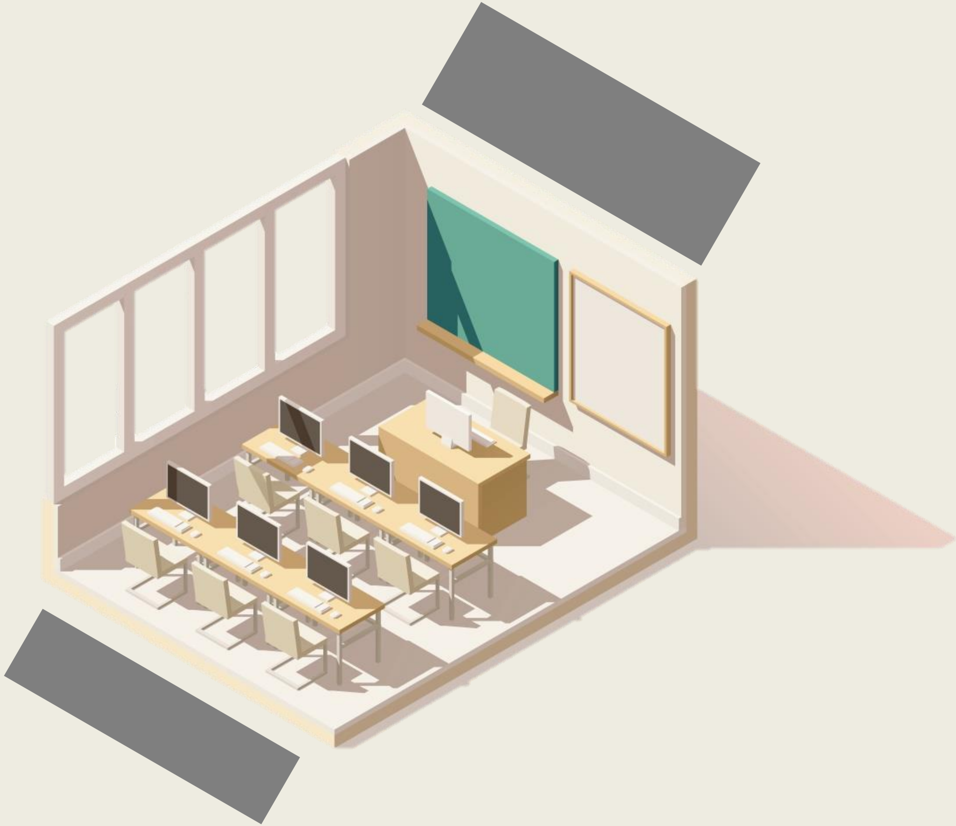
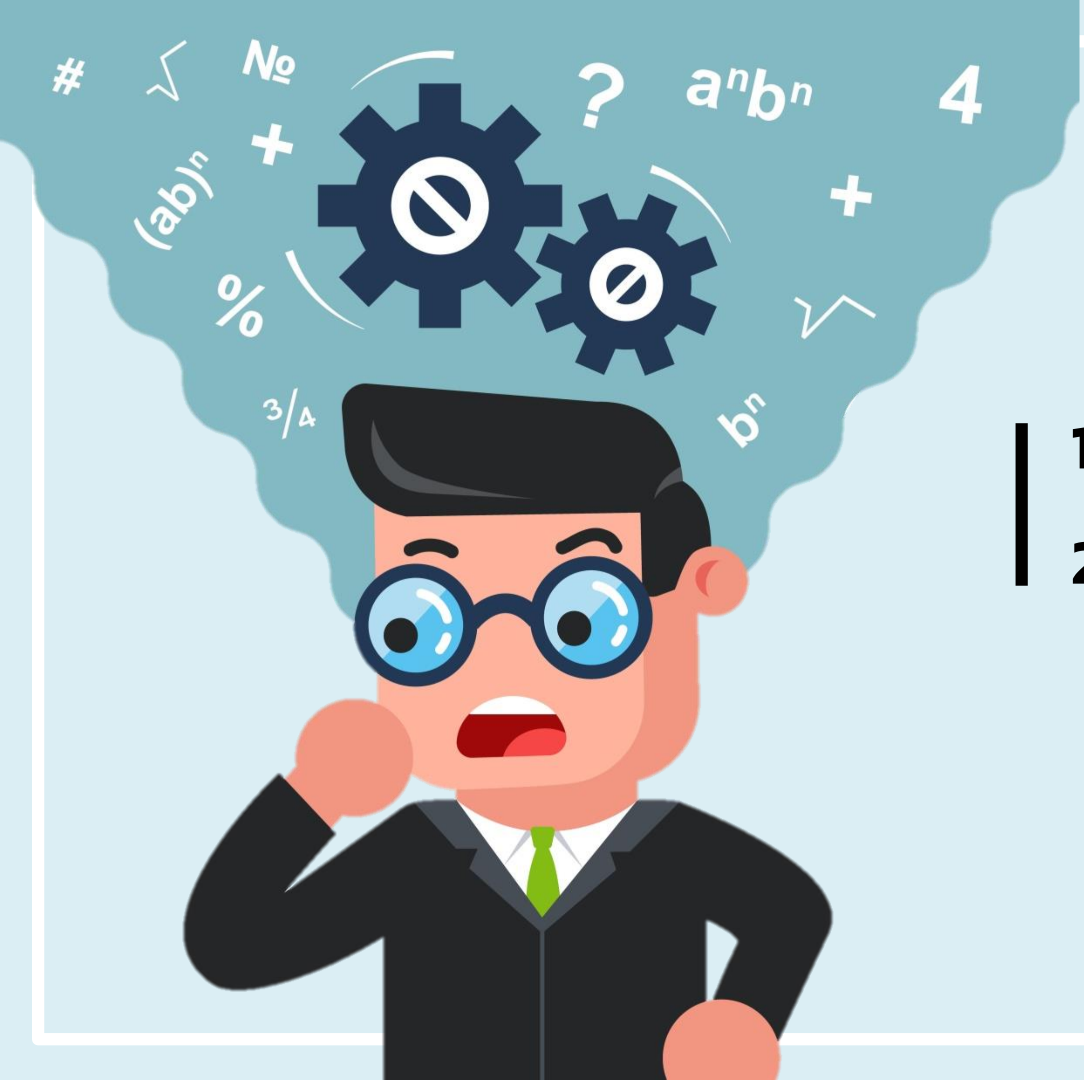


บทที่ 3
การวัดการกระจาย



สาระการเรียนรู้

1. การวัดการกระจายสัมบูรณ์
2. การวัดการกระจายสัมพัทธ์

การวัดแนวโน้มเข้าสู่ส่วนกลางของข้อมูลเพียงอย่างเดียว ทำให้ไม่ทราบลักษณะของข้อมูล จึงมีการวัดการกระจายของข้อมูลจากค่ากลางควบคู่ไปด้วย จะทำให้เห็นลักษณะของข้อมูลชัดเจนยิ่งขึ้น การวัดการกระจายของข้อมูลเป็นการพิจารณาลักษณะของข้อมูลว่ามีการกระจายหรือแตกต่างจากค่ากลางของข้อมูลมากน้อยเพียงใด

โดยทั่วไปการวัดการกระจายของข้อมูล มี 2 แบบ ดังนี้

1) การวัดการกระจายสัมบูรณ์ (Absolute Variation)

2) การวัดการกระจายสัมพัทธ์ (Relative Variation)

1 การวัดการกระจายสัมบูรณ์

การวัดการกระจายสัมบูรณ์

เป็นการวัดการกระจายของข้อมูลเพียงชุดเดียว เมื่อต้องการทราบว่าข้อมูลชุดนั้น มีการกระจายมากน้อยเพียงใด สามารถทำได้ 4 วิธี คือ

1) พิสัย (Range)
2) ส่วนเบี่ยงเบนควอร์ไทล์ (Quartile Deviation)
3) ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation)
4) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

1. พิสัย

คือ ค่าที่ใช้วัดการกระจายของข้อมูลที่ได้จากผลต่างระหว่างข้อมูลที่มีค่าสูงสุดกับข้อมูลที่มีค่าต่ำสุด

- 1) การหาพิสัยของข้อมูลที่ไม่ได้แจกแจงความถี่

ถ้าให้ X_{max} เป็นค่าสูงสุดของข้อมูล และ X_{min} เป็นค่าต่ำสุดของข้อมูล

พิสัย $= X_{max} - X_{min}$

- 2) การหาพิสัยของข้อมูลที่แจกแจงความถี่

พิสัย = ขอบบนของอันตรภาคชั้นคะแนนสูงสุด - ขอบล่างของอันตรภาคชั้นคะแนนต่ำสุด

2. ส่วนเบี่ยงเบนควอไทล์

คือ ค่าที่ใช้วัดการกระจายข้อมูลที่ได้จากครึ่งหนึ่งของผลต่างของควอร์ไทล์ที่ 3 กับควอร์ไทล์ที่ 1 ใช้สัญลักษณ์ Q.D แทนส่วนเบี่ยงเบนควอร์ไทล์

ถ้าส่วนเบี่ยงเบนควอร์ไทล์มีค่ามากแสดงว่ามีการกระจายมาก ถ้าส่วนเบี่ยงเบนควอร์ไทล์ มีค่าน้อยแสดงว่ามีการกระจายน้อย มีสูตรในการคำนวณดังนี้

2. ส่วนเบี่ยงเบนควอร์ไทล์

1) การหาส่วนเบี่ยงเบนควอร์ไทล์ของข้อมูลที่ไม่ได้แจกแจงความถี่

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

โดยที่ Q_3 แทนควอร์ไทล์ที่ 3 และ Q_1 แทนควอร์ไทล์ที่ 1

1) การหาส่วนเบี่ยงเบนควอร์ไทล์ของข้อมูลที่ไม่ได้แจกแจงความถี่
ขั้นตอนการหาค่าส่วนเบี่ยงเบนควอร์ไทล์ ดังนี้

(1) เรียงลำดับข้อมูลจากน้อยไปมาก

(2) หาตำแหน่งของ Q_3 และ Q_1 โดยใช้สูตรตำแหน่งของ $Q_r = \frac{r(N+1)}{4}$
เมื่อ r = 1, 2, 3
นั่นคือ ตำแหน่งของ $Q_3 = \frac{3(N+1)}{4}$ และตำแหน่งของ $Q_1 = \frac{N+1}{4}$

(3) หาค่า Q_3 และ Q_1 ของข้อมูล โดยดูว่าตำแหน่งของ Q_3 และ Q_1 ตรงกับค่าของข้อมูลตัวใดในข้อมูลชุดนั้นๆ

(4) คำนวณหาค่าส่วนเบี่ยงเบนควอร์ไทล์

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

2. ส่วนเบี่ยงเบนควอร์ไทล์

2) การหาส่วนเบี่ยงเบนควอร์ไทล์ของข้อมูลที่แจกแจงความถี่

2) การหาส่วนเบี่ยงเบนควอร์ไทล์ของข้อมูลที่แจกแจงความถี่ มีขั้นตอนการหาค่าส่วนเบี่ยงเบนควอร์ไทล์ดังนี้

(1) หาความถี่สะสม

(2) หาตำแหน่งของ Q_3 และ Q_1 โดยใช้สูตร ตำแหน่งของ $Q_r = \frac{rN}{4}$ เมื่อ r = 1, 2, 3

นั่นคือ ตำแหน่งของ $Q_3 = \frac{3N}{4}$ และตำแหน่งของ $Q_1 = \frac{N}{4}$

(3) หาชั้นที่คาดว่าจะมี ค่า Q_3 และ Q_1 อยู่ จากตำแหน่งของ Q_3 และ Q_1 โดยเทียบกับความถี่สะสม

(4) หาค่า Q_3 และ Q_1 ของข้อมูล จากสูตร $Q_r = L_0 + I\left(\frac{\frac{rN}{4} - F}{f}\right)$ เมื่อ r = 1, 2, 3

นั่นคือ $Q_3 = L_0 + I\left(\frac{\frac{3N}{4} - F}{f}\right)$ และ $Q_1 = L_0 + I\left(\frac{\frac{N}{4} - F}{f}\right)$

(5) คำนวณหาค่าส่วนเบี่ยงเบนควอร์ไทล์

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

3. ส่วนเบี่ยงเบนเฉลี่ย

คือ ค่าที่ใช้วัดการกระจายของข้อมูลรอบ ๆ ค่าเฉลี่ย (Mean) โดยการหาค่าเฉลี่ยของผลรวมของผลต่างระหว่างคะแนนแต่ละตัวกับค่าเฉลี่ย **ใช้สัญลักษณ์ M.D แทน ส่วนเบี่ยงเบนเฉลี่ย**

ถ้าส่วนเบี่ยงเบนเฉลี่ยมีค่ามาก แสดงว่ามีการกระจายมาก ถ้าส่วนเบี่ยงเบนเฉลี่ยมีค่าน้อย แสดงว่ามีการกระจายน้อย

3. ส่วนเบี่ยงเบนเฉลี่ย

1) การหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูลที่ไม่ได้แจกแจงความถี่

ส่วนเบี่ยงเบนเฉลี่ยของประชากร	ส่วนเบี่ยงเบนเฉลี่ยของตัวอย่าง
$M.D = \dfrac{\sum_{i=1}^{N} \lvert X_i - \mu \rvert}{N}$	$M.D = \dfrac{\sum_{i=1}^{n} \lvert X_i - \bar{X} \rvert}{n}$
เมื่อ X_i คือ ค่าของข้อมูลตัวที่ i	เมื่อ X_i คือ ค่าของข้อมูลตัวที่ i
N คือ จำนวนข้อมูลประชากร	n คือ จำนวนข้อมูลตัวอย่าง

4. ส่วนเบี่ยงเบนมาตรฐาน

เป็นค่าที่ใช้วัดการกระจายของข้อมูล เพื่อศึกษาดูว่าข้อมูลมีการกระจายห่างออกจากค่าเฉลี่ยเลขคณิตมากน้อยเพียงใด เป็นวิธีการทางสถิติที่ดีและนิยมใช้มากที่สุด

ส่วนเบี่ยงเบนมาตรฐานที่มีค่ามาก แสดงว่าข้อมูลมีค่าแตกต่างจากค่าเฉลี่ยมาก นั่นคือข้อมูลแต่ละตัวมีการกระจายห่างกันมาก ส่วนเบี่ยงเบนมาตรฐานที่มีค่าน้อย แสดงว่าข้อมูลมีค่าแตกต่างจากค่าเฉลี่ยน้อย

4. ส่วนเบี่ยงเบนมาตรฐาน

เป็นค่าที่ใช้วัดการกระจายของข้อมูล เพื่อศึกษาดูว่าข้อมูลมีการกระจายห่างออกจากค่าเฉลี่ยเลขคณิตมากน้อยเพียงใด เป็นวิธีการทางสถิติที่ดีและนิยมใช้มากที่สุด

ส่วนเบี่ยงเบนมาตรฐานที่มีค่ามาก แสดงว่าข้อมูลมีค่าแตกต่างจากค่าเฉลี่ยมาก นั่นคือข้อมูลแต่ละตัวมีการกระจายห่างกันมาก ส่วนเบี่ยงเบนมาตรฐานที่มีค่าน้อย แสดงว่าข้อมูลมีค่า แตกต่างจากค่าเฉลี่ยน้อย

4. ส่วนเบี่ยงเบนมาตรฐาน

1) ข้อมูลที่ไม่ได้แจกแจงความถี่

ความแปรปรวนของประชากร	ความแปรปรวนของตัวอย่าง
$\sigma^2 = \dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{N}$	$S^2 = \dfrac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$
$\sigma^2 = \dfrac{\sum_{i=1}^{N}X_i^2}{N} - (\mu)^2$	$S^2 = \dfrac{1}{n-1}\left(\sum_{i=1}^{n}X_i - n(\bar{X})^2\right)$
ส่วนเบี่ยงเบนมาตรฐานของประชากร	ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
$\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{N}}$	$S = \sqrt{\dfrac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}$
$\sigma = \sqrt{\dfrac{\sum_{i=1}^{N}X_i^2}{N} - (\mu)^2}$	$S = \sqrt{\dfrac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n(\bar{X})^2\right)}$
โดยที่ X_i คือ ค่าของข้อมูลตัวที่ i N คือ จำนวนข้อมูลประชากร	โดยที่ X_i คือ ค่าของข้อมูลตัวที่ i n คือ จำนวนข้อมูลตัวอย่าง

4. ส่วนเบี่ยงเบนมาตรฐาน

2) ข้อมูลที่แจกแจงความถี่

ความแปรปรวนของประชากร	ความแปรปรวนของตัวอย่าง
$\sigma^2 = \dfrac{\sum_{i=1}^{k} f_i X_i^2}{N} - (\mu)^2$	$S^2 = \dfrac{1}{n-1}\left(\sum_{i=1}^{k} f_i X_i^2 - n(\overline{X})^2\right)$
ส่วนเบี่ยงเบนมาตรฐานของประชากร	ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
$\sigma = \sqrt{\dfrac{\sum_{i=1}^{k} f_i X_i^2}{N} - (\mu)^2}$	$S = \sqrt{\dfrac{1}{n-1}\left(\sum_{i=1}^{k} f_i X_i^2 - n(\overline{X})^2\right)}$
โดยที่ X_i คือ ค่ากึ่งกลางของข้อมูลชั้นที่ i N คือ จำนวนข้อมูลประชากร f_i คือ ค่าความถี่ของข้อมูลชั้นที่ i k คือ จำนวนชั้น	โดยที่ X_i คือ ค่ากึ่งกลางของข้อมูลชั้นที่ i n คือ จำนวนข้อมูลตัวอย่าง f_i คือ ค่าความถี่ของข้อมูลชั้นที่ i k คือ จำนวนชั้น

ข้อสังเกตเกี่ยวกับการใช้ค่าวัดการกระจาย

1) พิสัย

1) การวัดการกระจายสัมบูรณ์โดยใช้พิสัยเหมาะสำหรับวัดการกระจายอย่างคร่าว ๆ เมื่อต้องการทราบค่าการกระจายอย่างรวดเร็ว เพราะใช้เวลาน้อยในการคำนวณและใช้คะแนนเพียง 2 ค่าเท่านั้น

2) ถ้าค่าของข้อมูลบางตัวมีค่ามากหรือน้อยผิดปกติจากค่าของข้อมูลอื่น ๆ จะทำให้ค่าพิสัยที่คำนวณได้มีค่าสูงกว่าที่ควรจะเป็นมาก

3) ถ้าข้อมูลที่แจกแจงความถี่เป็นอันตรภาคชั้นเปิดจะไม่สามารถหาค่าพิสัยได้

2) ส่วนเบี่ยงเบนควอร์ไทล์

1) ส่วนเบี่ยงเบนควอร์ไทล์เป็นการวัดการกระจายที่ดีกว่าการวัดด้วยค่าพิสัย แต่ก็ยังใช้เพียงบางค่า ไม่ได้ใช้ข้อมูลทุกค่าในการคำนวณ

2) ส่วนเบี่ยงเบนควอร์ไทล์เหมาะสำหรับใช้วัดการกระจาย กรณีมีคะแนนบางค่าสูงหรือต่ำกว่าคะแนนตัวอื่น ๆ ในชุดมาก

3) ส่วนเบี่ยงเบนควอร์ไทล์เหมาะสำหรับใช้ควบคู่กับการวัดแนวโน้มเข้าสู่ส่วนกลางด้วยค่ามัธยฐาน

ข้อสังเกตเกี่ยวกับการใช้ค่าวัดการกระจาย

3) ส่วนเบี่ยงเบนเฉลี่ย

1) ส่วนเบี่ยงเบนเฉลี่ยเป็นการวัดการกระจายที่ละเอียดกว่าการวัดด้วยค่าพิสัยและส่วนเบี่ยงเบนควอร์ไทล์ เพราะได้ใช้คะแนนทุก ๆ ตัวในการคำนวณ

2) การคำนวณค่าส่วนเบี่ยงเบนเฉลี่ยไม่ได้คำนึงถึงเครื่องหมายของผลต่างระหว่างคะแนนแต่ละตัวกับค่าเฉลี่ย จึงไม่เป็นที่นิยมใช้

4) ส่วนเบี่ยงเบนมาตรฐาน

1) การวัดการกระจายสัมบูรณ์โดยใช้ส่วนเบี่ยงเบนมาตรฐานเป็นวิธีที่ดีที่สุด เพราะต้องใช้ข้อมูลทุกค่ามาคำนวณ และนิยมใช้มากที่สุดในการนำไปใช้ในสถิติขั้นสูง

2) ค่าส่วนเบี่ยงเบนมาตรฐาน เป็นค่าการกระจายที่มีหน่วยเช่นเดียวกับหน่วยของข้อมูลที่เก็บรวบรวมมา ส่วนค่าความแปรปรวน ซึ่งมีค่าเป็นกำลังสองของส่วนเบี่ยงเบนมาตรฐานเป็นค่าการกระจายที่มีหน่วยเป็นกำลังสองของหน่วยของข้อมูลที่เก็บรวบรวมมา

2 การวัดการกระจายสัมพัทธ์

เป็นการวัดการกระจายของข้อมูลตั้งแต่สองชุดขึ้นไป โดยใช้อัตราส่วนของค่าที่ได้จากการวัดการกระจายสัมบูรณ์กับค่ากลางของข้อมูลชุดนั้น สามารถนำไปใช้เปรียบเทียบการกระจายของข้อมูลแต่ละกลุ่มว่ากลุ่มใดมีการกระจายมากน้อยกว่ากัน มักจะคำนวณออกมาในรูปร้อยละ เรียกว่า “สัมประสิทธิ์ของการกระจาย” แบ่งออกเป็น 4 วิธี คือ

1. สัมประสิทธิ์ของพิสัย

$$\text{สัมประสิทธิ์ของพิสัย} = \frac{\text{ค่าสูงสุดของข้อมูล} - \text{ค่าต่ำสุดของข้อมูล}}{\text{ค่าสูงสุดของข้อมูล} + \text{ค่าต่ำสุดของข้อมูล}} = \frac{X_{max} - X_{min}}{X_{max} + X_{min}}$$

2. สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์

$$\text{สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

3. สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย

$$\text{สัมประสิทธิ์ส่วนเบี่ยงเบนเฉลี่ยของประชากร} = \frac{M.D}{\mu}$$

$$\text{สัมประสิทธิ์ส่วนเบี่ยงเบนเฉลี่ยของตัวอย่าง} = \frac{M.D}{\overline{X}}$$

4. สัมประสิทธิ์การแปรผัน

$$\text{สัมประสิทธิ์ของความแปรผันของประชากร} = C.V. = \frac{\sigma}{\mu}$$

$$\text{สัมประสิทธิ์ของความแปรผันของตัวอย่าง} = C.V. = \frac{S}{\overline{X}}$$

สรุป

การวัดการกระจายของข้อมูล เป็นการศึกษาลักษณะของข้อมูลว่ามีการกระจายหรือแตกต่างจากค่ากลางของข้อมูลมากน้อยเพียงใด มี 2 แบบ คือ การวัดการกระจายสัมบูรณ์ และการวัดการกระจายสัมพัทธ์

การวัดการกระจายสัมบูรณ์ เป็นการวัดการกระจายของข้อมูลเพียงชุดเดียว สามารถทำได้ 4 วิธี คือ 1) พิสัย 2) ส่วนเบี่ยงเบนควอร์ไทล์ 3) ส่วนเบี่ยงเบนเฉลี่ย 4) ส่วนเบี่ยงเบนมาตรฐาน

การวัดการกระจายสัมพัทธ์ เป็นการวัดการกระจายของข้อมูล โดยใช้อัตราส่วนของค่าที่ได้จากการวัดการกระจายสัมบูรณ์ กับค่ากลางของข้อมูล สามารถทำได้ 4 วิธี คือ

1) สัมประสิทธิ์ของพิสัย
2) สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์
3) สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย
4) สัมประสิทธิ์การแปรผัน