

QQ Plot

1. เรียงข้อมูลของแต่ละ sample จากมากไปน้อย
2. กำหนดจำนวน quantile จากนั้นคำนวณ quantile ของข้อมูลในแต่ละ sample
3. จับคู่ข้อมูลของทั้ง 2 sample ในแต่ละ quantile
4. plot แต่ละคู่ข้อมูลจากข้อ 3
5. plot เส้นของ $y = x$ ถ้าแต่ละจุดอยู่ใกล้เส้น $y = x$ จะเป็นไปได้ว่าทั้ง 2 sample มาจาก distribution เดียวกัน

ตัวอย่างการคำนวณ

Sample 1: $X = [9, 6, 7, 3, 8]$

Sample 2: $Y = [4, 8, 6, 7, 10, 5, 12]$

1. เรียงข้อมูลของแต่ละ sample จากมากไปน้อย

$$X = [3, 6, 7, 8, 9]$$

$$Y = [4, 5, 6, 7, 8, 10, 12]$$

2. กำหนดจำนวน quantile จากนั้นคำนวณ quantile ของข้อมูลในแต่ละ sample
กำหนดให้จำนวน quantile คือ 5

สำหรับ Sample 1 (smaller sample): the quantiles are at positions:

$$\frac{i - 0.5}{n} ; i = 1, 2, \dots, n$$

โดยที่ n คือจำนวนข้อมูลใน sample

$$\text{Quantile of sample 1} = \frac{1-0.5}{5}, \frac{2-0.5}{5}, \frac{3-0.5}{5}, \frac{4-0.5}{5}, \frac{5-0.5}{5} = 0.1, 0.3, 0.5, 0.7, 0.9$$

สำหรับ Sample 2 (larger sample): หาค่าที่สอดคล้องกับ Quantile of sample 1 โดยใช้ interpolation:

0.1 \rightarrow Interpolate between 1st and 2nd quantile of X_2

0.3 \rightarrow Interpolate between 2nd and 3rd quantile of X_2

0.5 \rightarrow 4th quantile of (X_2) (middle value)

0.7 \rightarrow Interpolate between 4th and 5th quantile of X_2

0.9 \rightarrow Interpolate between 6th and 7th quantile of X_2

ใช้ linear interpolation, จะได้ว่า

$$Q_{0.1}(X_2) = 4 + 0.1 \times (5 - 4) = 4.1$$

$$Q_{0.3}(X_2) = 5 + 0.3 \times (6 - 5) = 5.3$$

$$Q_{0.5}(X_2) = 6$$

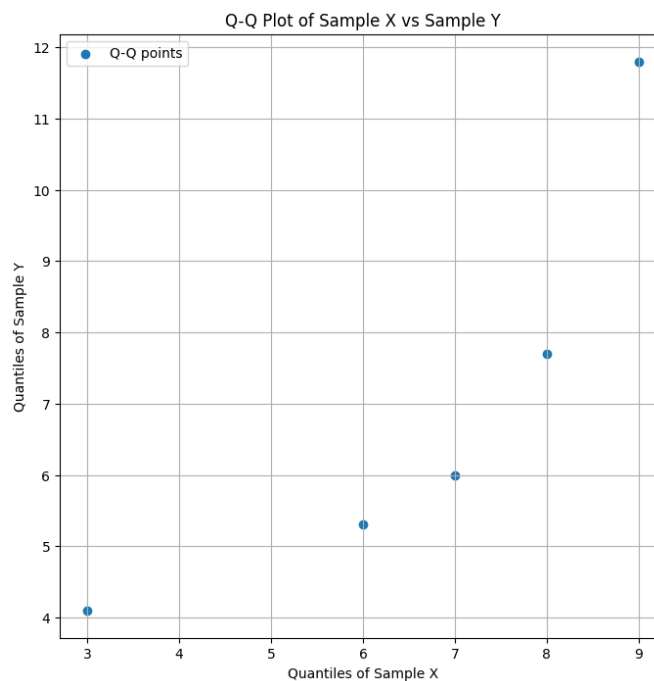
$$Q_{0.7}(X_2) = 7 + 0.7 \times (8 - 7) = 7.7$$

$$Q_{0.9}(X_2) = 10 + 0.9 \times (12 - 10) = 10 + 1.8 = 11.8$$

3. จับคู่ข้อมูลของทั้ง 2 sample ในแต่ละ quantile

(3,4.1), (6,5.3), (7,6), (8,7.7), (9,11.8)

4. plot แต่ละคู่ข้อมูลจากข้อ 3



5. plot เส้นของ $y = x$

เนื่องจากว่าแต่ละจุดไม่ได้อยู่ใกล้เส้น $y = x$ เป็นไปได้ว่าทั้ง 2 sample มาจากคนละ distribution

