STO2011 ชีวสถิติ

# Biostatistics

# What is your understanding of the word statistics / biostatistics?

# STATISTICS

ตัวเลขที่เก็บรวบรวมแทนข้อเท็จจริงต่างๆ
เป็นศาสตร์ที่ประกอบด้วย การเก็บรวบรวมข้อมูล
การนำเสนอข้อมูล การวิเคราะห์ข้อมูล การแปล
ความหมายข้อมูล

# BIOSTATISTICS

วิธีการทางสถิติ เน้นหนักทางด้านสิ่งมีชีวิต
ชีววิทยา ชีวการแพทย์ รวมทั้งทางด้านการแพทย์
และการสาธารณสุข

# วิธีการทางสถิติ

- การเก็บรวบรวมข้อมูล
- การการนำเสนอข้อมูล
- การวิเคราะห์ข้อมูล
- การแปลความหมายข้อมูล

# STATISTICAL DATA

ข้อเท็จจริงที่เป็นตัวเลข (เช่น ความสูง น้ำหนัก รายได้) หรือไม่ใช่ตัวเลข (มักกล่าวในรูปข่าวสารเป็นข้อมูลเชิงวิเคราะห์) เกี่ยวกับเรื่องหนึ่งเรื่องใดที่สนใจ

# ชนิดของข้อมูล (จำแนกตามแหล่งที่มา)

- ข้อมูลปฐมภูมิ ข้อมูลที่เก็บรวบรวมจากแหล่งกำเนิดข้อมูลโดยตรง
- ข้อมูลทุติยภูมิ ข้อมูลที่ไม่ได้เก็บรวบรวมจากแหล่งกำเนิดข้อมูลโดยตรง

# ชนิดของข้อมูล (จำแนกตามลักษณะข้อมูล)

- ข้อมูลคุณภาพ ข้อมูลที่แสดงถึงลักษณะที่แตกต่างของตัวแปร เช่น เพศ (ชาย,หญิง) การสอบ (ผ่าน, ไม่ผ่าน) ทัศนคติ (ไม่ชอบ, ชอบ)
- ข้อมูลปริมาณ ข้อมูลที่แสดงถึงปริมาณของตัวแปร เช่น อายุ น้ำหนัก ส่วนสูง จำนวนเม็ดเลือด ค่าความดันโลหิต

# ชนิดของข้อมูล (จำแนกตามลักษณะการกระจายข้อมูล)

- ข้อมูลจำแนกตามคุณภาพ ข้อมูลที่แสดงถึงคุณสมบัติ สภาพฐานะ ความคิดเห็น
- ข้อมูลจำแนกตามปริมาณ ข้อมูลที่แสดงข้อเท็จจริงตามค่าที่วัดได้
- ข้อมูลจำแนกตามกาลเวลา ข้อมูลที่แสดงข้อเท็จจริงตามกาลเวลา
- ข้อมูลจำแนกตามภูมิศาสตร์ ข้อมูลที่แสดงข้อเท็จจริงระหว่างท้องถิ่น

# Measurement data

**นามบัญญัติ**

- Nominal measurement

**จัดอันดับ**

- Ordinal measurement

**ภาคชั้น**

- Interval measurement

**อัตราส่วน**

- Ratio measurement

# การเก็บรวบรวมข้อมูล

- จากทะเบียนประวัติ
- จากการสำรวจ
- จากการทดลอง
- จากการสังเกต

# การแจกแจงความถี่

- ใช้ในการจัดข้อมูลที่มีอยู่ให้เป็นกลุ่มๆ
- ใช้เมื่อมีข้อมูลจำนวนมากหรือมีตัวเลขซ้ำกัน
- มีขั้นตอน ดังนี้

- พิสัย Range = Xmax - Xmin
- จำนวนชั้น k = 1+3.322 log n

  n = จำนวนข้อมูล
- ความกว้างอันตรภาคชั้น I = R/k

# การหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ของข้อมูลที่ไม่ได้แจกแจงความถี่

# การหาควอไทล์

ควอไทล์ เป็นค่าที่แบ่งจำนวนข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน เมื่อข้อมูลดังกล่าวถูกเรียงจากค่าน้อยไปหาค่ามาก

เนื่องจากค่าที่แบ่งจำนวนข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน มีอยู่ 3 ค่า ดังนั้นเราจึงตั้งชื่อเรียกแต่ละค่าว่า

- ควอไทล์ที่ 1 ใช้สัญลักษณ์ $Q_1$
- ควอไทล์ที่ 2 ใช้สัญลักษณ์ $Q_2$
- ควอไทล์ที่ 3 ใช้สัญลักษณ์ $Q_3$

แต่ละค่ามีความหมายดังนี้

- Q1 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 1 ใน 4 ของข้อมูลทั้งหมด
- Q2 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 2 ใน 4 ของข้อมูลทั้งหมด
- Q3 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 3 ใน 4 ของข้อมูลทั้งหมด

จากความหมายดังกล่าว ถ้าเราสังเกตให้ดี จะพบว่า Q2 คือ มัธยฐาน นั่นเอง

ตำแหน่งของ $Q_r$ คือ ตำแหน่งที่ $(N+1)\frac{r}{4}$

EX. ให้ 27 , 29 , 25 , 30 , 37 , 24 , 22 เป็นข้อมูลชุดหนึ่ง จงหาควอไทล์ที่หนึ่ง

EX. ผลการชั่งน้ำหนักของเด็กนักเรียน 9 คน ผลปรากฏดังนี้ (หน่วยเป็นกิโลกรัม)

**35 , 37 , 32 , 34 , 38 , 40 , 46 , 41 , 49** จงหาน้ำหนักของนักเรียนที่อยู่ ณ ตำแหน่งควอไทล์ที่ 3

# การหาเดไซล์

เดไซล์ เป็นค่าที่แบ่งจำนวนข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน เมื่อข้อมูลดังกล่าวถูกเรียงจากค่าน้อยไปหาค่ามาก

เนื่องจากค่าที่แบ่งจำนวนข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน มีอยู่ 9 ค่า ดังนั้นเราจึงตั้งชื่อเรียกแต่ละค่าว่า

- เดไซล์ที่ 1 ใช้สัญลักษณ์ D1
- เดไซล์ที่ 2 ใช้สัญลักษณ์ D2
- เดไซล์ที่ 9 ใช้สัญลักษณ์ D9

แต่ละค่ามีความหมายดังนี้

- D1 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 1 ใน 10 ของข้อมูลทั้งหมด
- D2 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 2 ใน 10 ของข้อมูลทั้งหมด

จะมีลักษณะดังนี้ไปเรื่อย ๆ จนถึง

- D9 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 9 ใน 10 ของข้อมูลทั้งหมด

ตำแหน่งของ $D_r$ คือ ตำแหน่งที่ $(N+1)\frac{r}{10}$

EX. ให้ 10 , 13 , 8 , 11 , 15 , 17 เป็นข้อมูลชุดหนึ่ง จงหาเดไซล์ที่ 5

EX. ให้ 10 , 17 , 13 , 25 , 26 , 27 เป็นข้อมูลชุดหนึ่ง จงหาเดไซล์ที่ 7

# การหาเปอร์เซ็นไทล์

เปอร์เซ็นไทล์ เป็นค่าที่แบ่งจำนวนข้อมูลออกเป็น 100 ส่วนเท่า ๆ กัน เมื่อข้อมูลดังกล่าวถูกเรียงจากค่าน้อยไปหาค่ามาก

เนื่องจากค่าที่แบ่งจำนวนข้อมูลออกเป็น 100 ส่วนเท่า ๆ กัน มีอยู่ 99 ค่า ดังนั้นเราจึงตั้งชื่อเรียกแต่ละค่าว่า

- เปอร์เซ็นไทล์ที่ 1 ใช้สัญลักษณ์ P1
- เปอร์เซ็นไทล์ที่ 2 ใช้สัญลักษณ์ P2
- เปอร์เซ็นไทล์ที่ 99 ใช้สัญลักษณ์ P99

แต่ละค่ามีความหมายดังนี้

- P1 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 1 ใน 100 ของข้อมูลทั้งหมด
- P2 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 2 ใน 100 ของข้อมูลทั้งหมด

จะมีลักษณะดังนี้ไปเรื่อย ๆ จนถึง

- P99 คือ ค่าที่มีจำนวนข้อมูลน้อยกว่าค่านี้อยู่ประมาณ 99 ใน 100 ของข้อมูลทั้งหมด

ตำแหน่งของ $P_r$ คือ ตำแหน่งที่ $(N+1)\frac{r}{100}$

EX. ให้ 10 , 13 , 8 , 11 , 15 , 17 , 20 , 14 , 22 เป็นข้อมูลชุดหนึ่ง จงหาเปอร์เซ็นไทล์ที่ 55

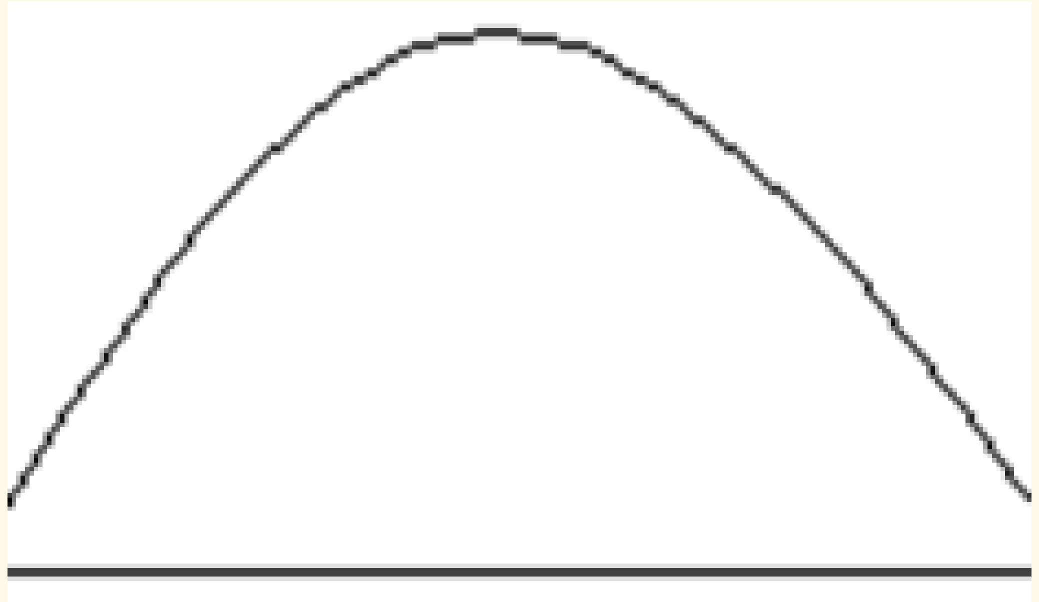
# ความสัมพันธ์ระหว่าง การแจกแจงความถี่ ค่ากลาง และการกระจายของข้อมูล

โดยทั่วๆ ไป เส้นโค้งของความถี่อาจแบ่งออกได้ 3 แบบ คือ

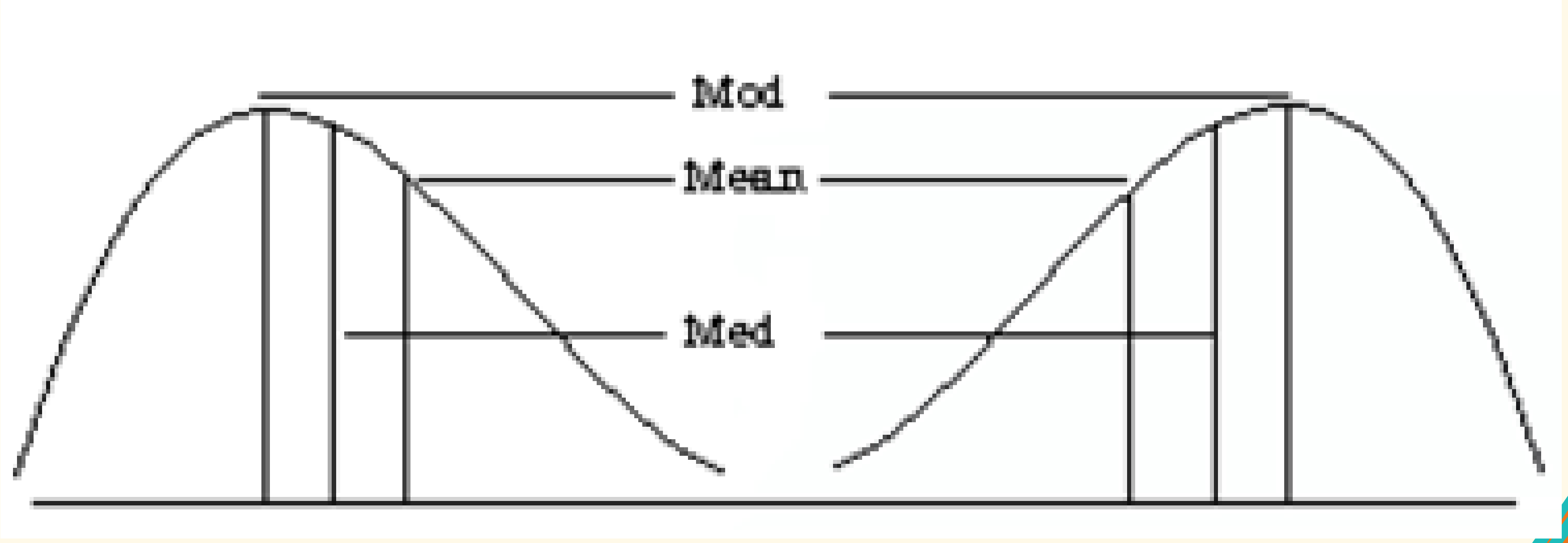
- เส้นโค้งปกติ หรือ รูประฆัง
  (normal or dell-shaped curve)
- เส้นโค้งเบ้ลาดทางขวาหรือทางบวก
  (positively skewed curve)
- เส้นโค้งเบ้ลาดทางซ้ายหรือทางลบ
  (negatively skewed curve)

## ลักษณะของเส้นโค้งปรากฏในรูปดังนี้

เส้นโค้งปกติ



เส้นโค้งเบ้ทางขวา     เส้นโค้งเบ้ทางซ้าย



รูปของการแจกแจงความถี่เป็นเส้นโค้งปกติ จะเห็นว่า ค่าเฉลี่ยเลขคณิต มัธยฐาน และ ฐานนิยม อยู่ที่จุดเดียวกัน คือ จุดที่มีความสูงสูงสุด

รูปของการแจกแจงความถี่เป็นเส้นโค้งเบ้ลาดทางขวา หรือส่วนของเส้นโค้งที่มีความชันน้อยอยู่ทางด้านขวาของรูป ค่าเฉลี่ยเลขคณิตจะมีค่ามากที่สุด รองลงมาเป็นมัธยฐาน และฐานนิยม ตามลำดับ

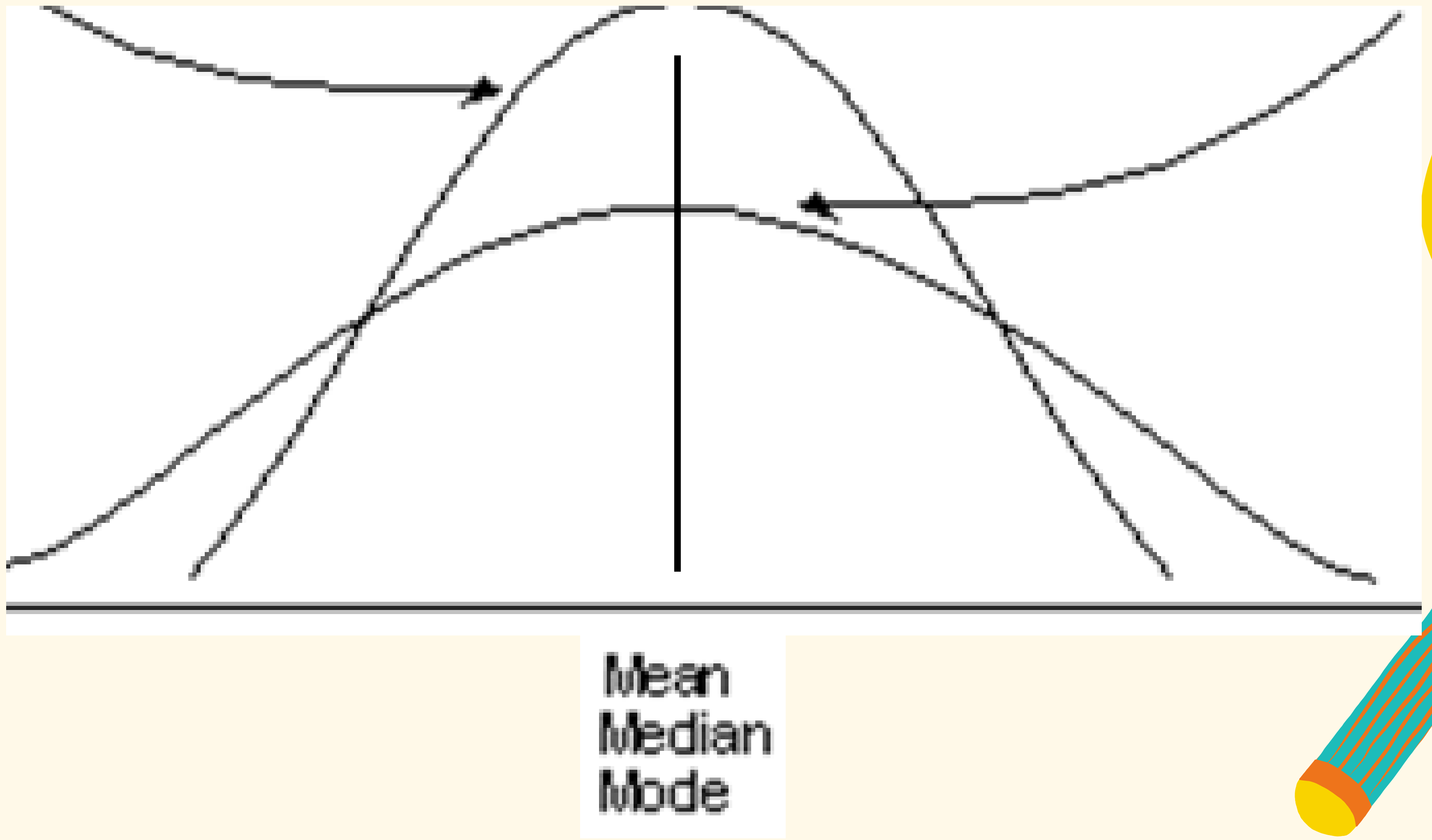
รูปของการแจกแจงความถี่เป็นเส้นโค้งเบ้ลาดทางซ้าย หรือส่วนของเส้นโค้งมีความชันน้อยอยู่ทางด้านซ้ายของรูป ฐานนิยมจะมีค่ามากที่สุด รองลงมาเป็นมัธยฐาน และค่าเฉลี่ยเลขคณิตมีค่าน้อยที่สุด

# เส้นโค้งปกติ

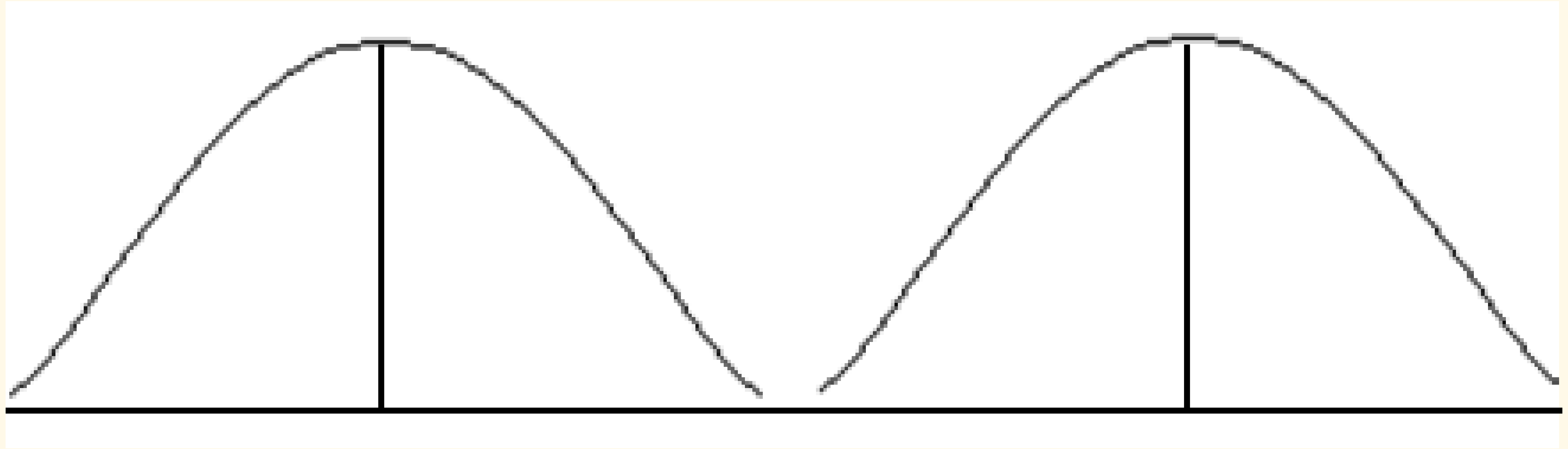
เส้นโค้งปกติจะมีความโด่งมากหรือน้อยขึ้นอยู่กับการกระจายของข้อมูล

- ถ้าข้อมูลมีการกระจายมาก เส้นโค้งปกติจะมีความโด่งน้อยหรือค่อนข้างแบน
- ถ้าข้อมูลมีการกระจายน้อย เส้นโค้งปกติจะมีความโด่งมาก

ข้อมูลที่มีการกระจายน้อย ข้อมูลที่มีการกระจายมาก



เส้นโค้งปกติ 2 รูป ซึ่งมีส่วนเบี่ยงเบนมาตรฐานเท่ากันแต่ค่าเฉลี่ยเลขคณิตไม่เท่ากัน จะมีลักษณะเหมือนกันแต่ตั้งอยู่บนตำแหน่งต่างกัน

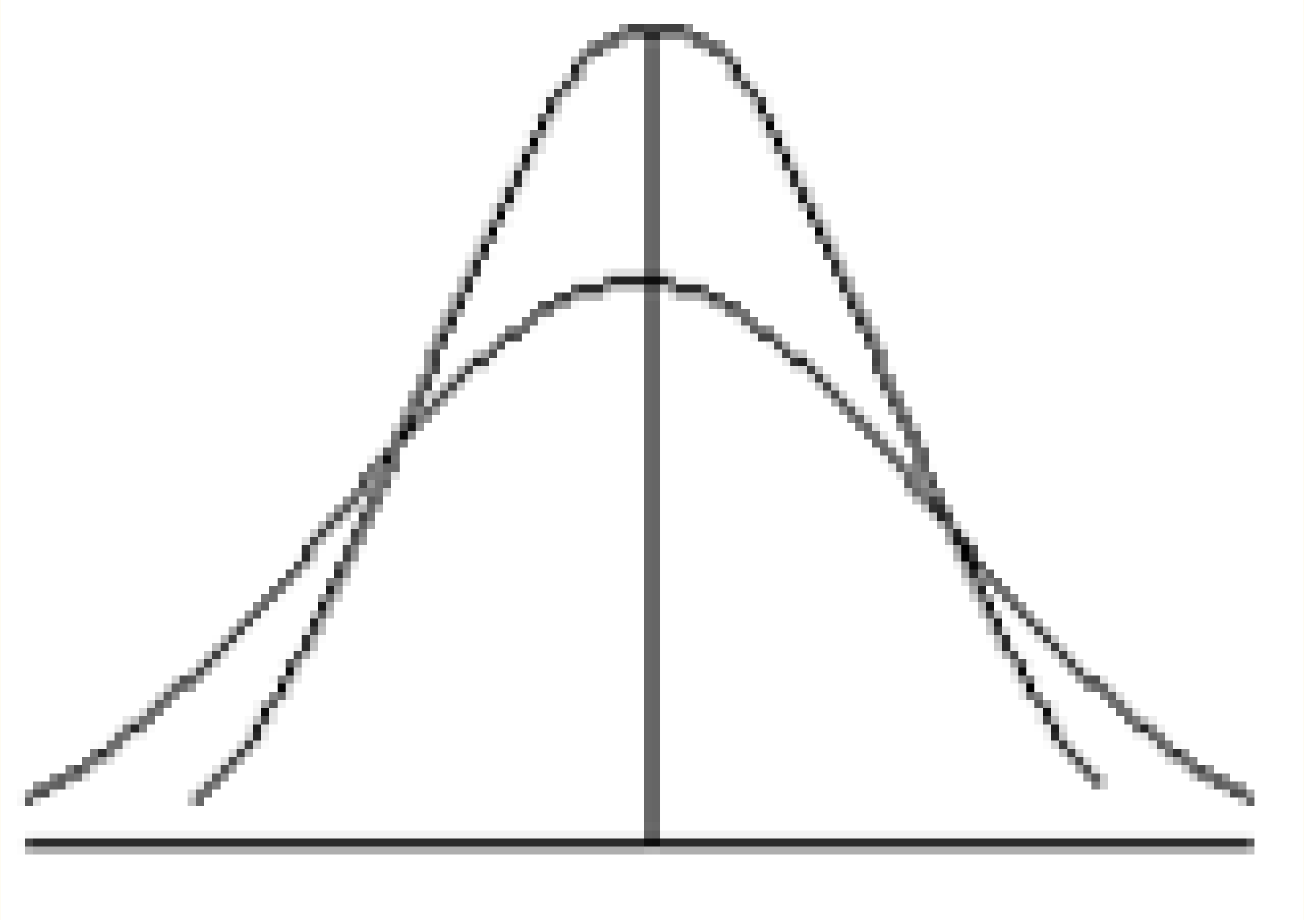


$\mu_1$ $\mu_2$

$$\sigma_1 = \sigma_2$$

$$\mu_1 \neq \mu_2$$

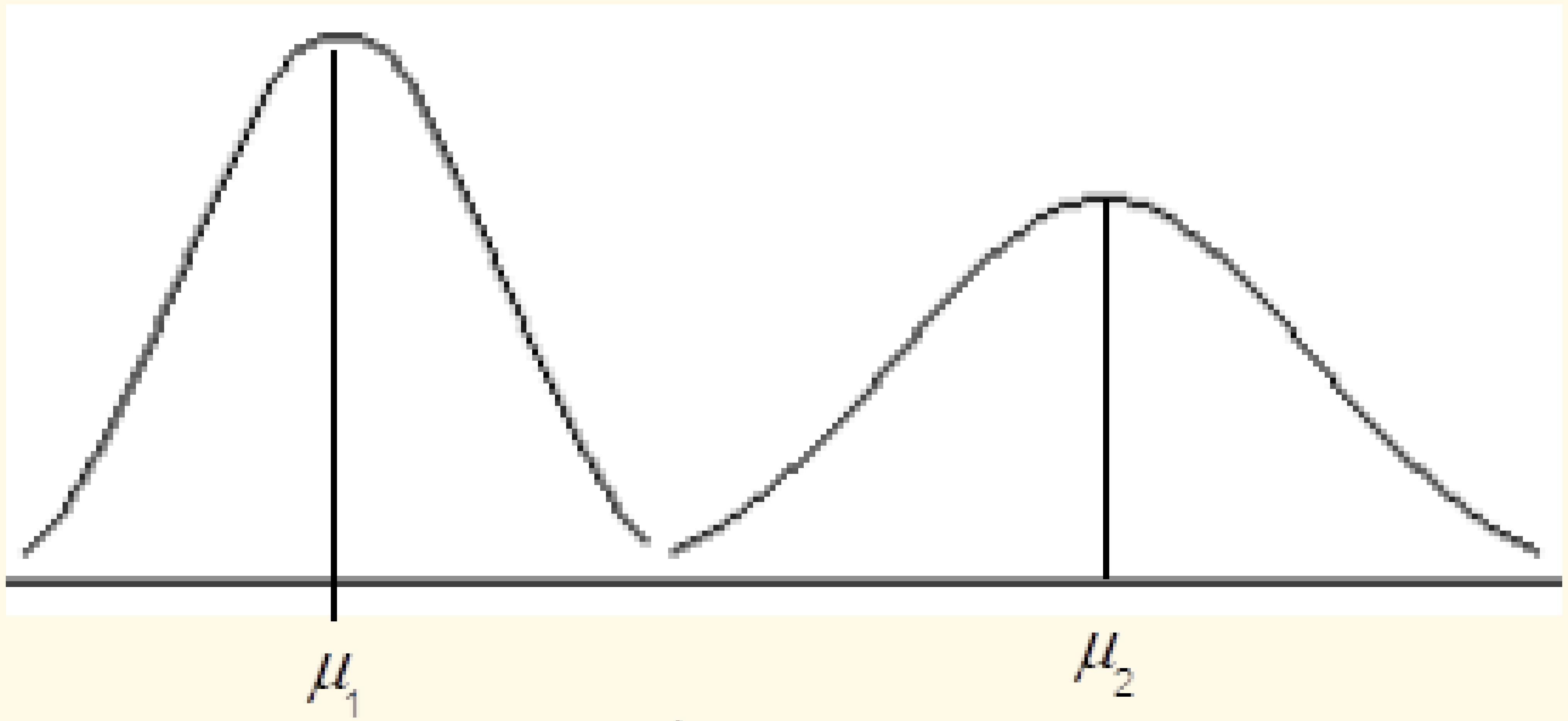
เส้นโค้งปกติ 2 รูป ซึ่งมีค่าเฉลี่ยเลขคณิต
เท่ากันแต่ส่วนเบี่ยงเบนมาตรฐานไม่เท่ากัน
จะมีจุดที่แสดงค่าเฉลี่ยเลขคณิตอยู่ที่
ตำแหน่งเดียวกันบนแกนนอน แต่เส้นโค้ง
ที่มีส่วนเบี่ยงเบนมาตรฐานมากกว่าจะเตี้ย
กว่า ดังนั้นถ้าข้อมูลมีการกระจายมากเส้น
โค้งจะเตี้ยลง และขยายออกไปไกลขึ้น



$$\mu_1 = \mu_2$$

$$\sigma_1 \neq \sigma_2$$

เส้นโค้งปกติ 2 รูป ที่มีค่าเฉลี่ยเลขคณิตต่างกันและส่วนเบี่ยงเบนมาตรฐานต่างกัน

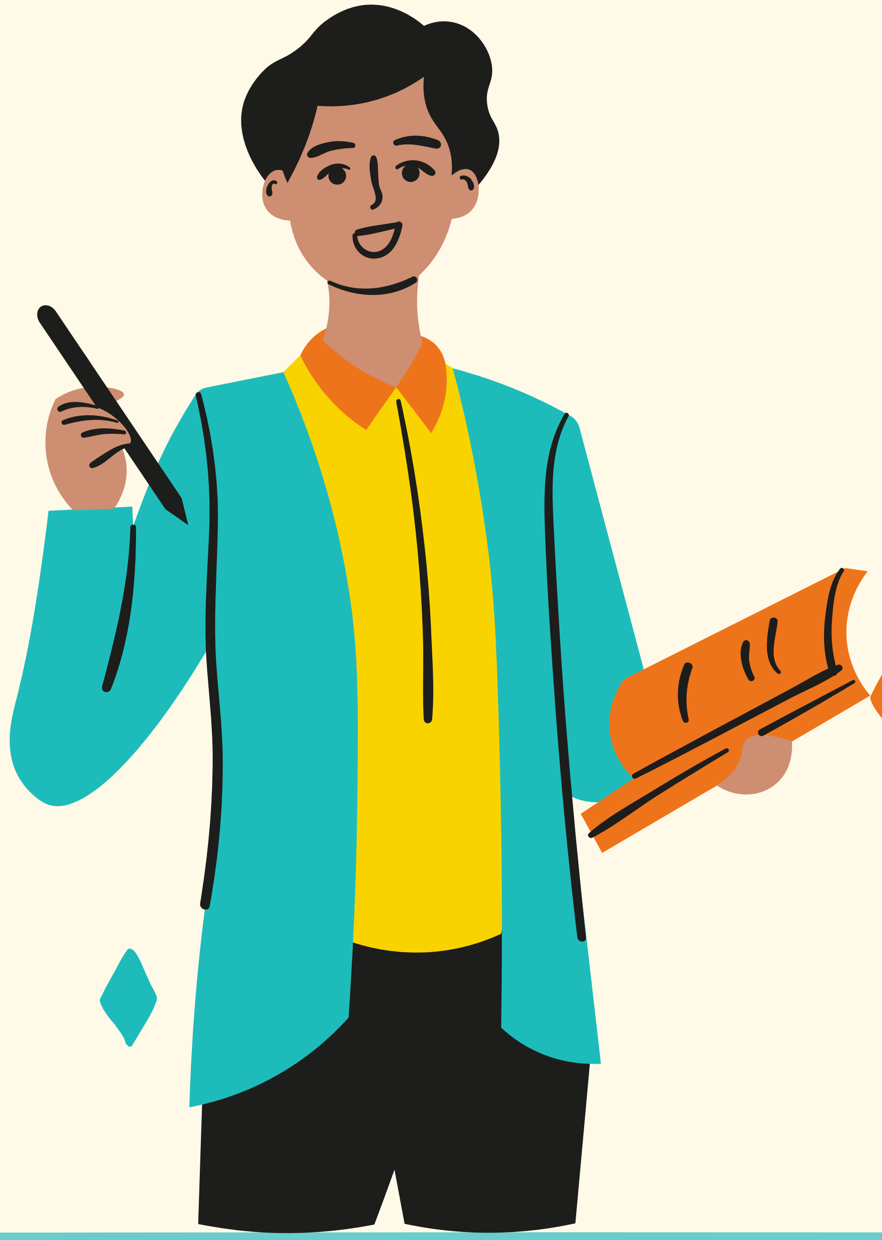



$$\sigma_1 \neq \sigma_2$$

$$\mu_1 \neq \mu_2$$

# สมบัติของเส้นโค้งปกติ

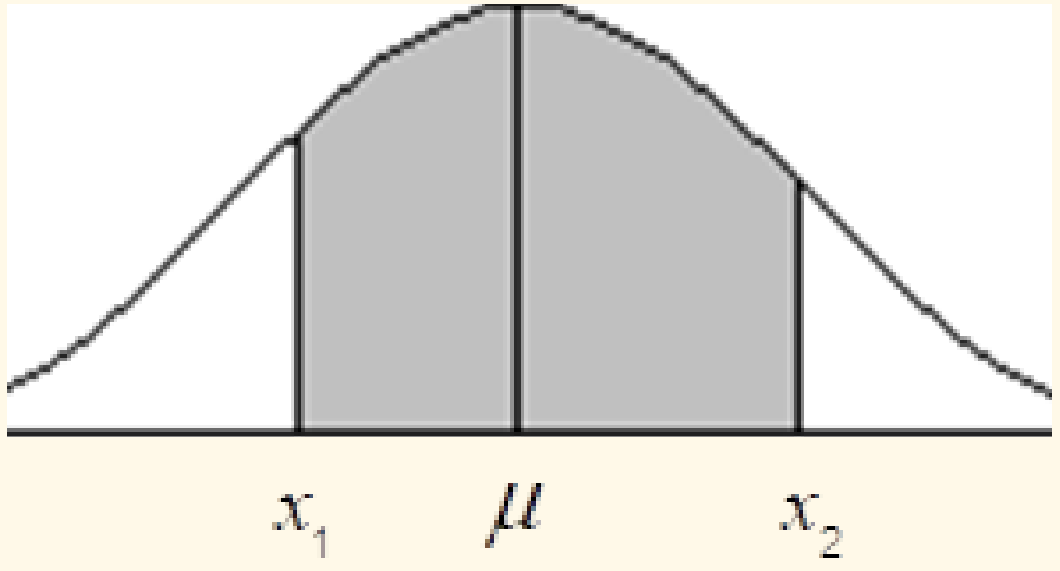
- ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม จะมีค่าเท่ากันและอยู่ ณ จุดที่เส้นตรงที่ลากผ่านจุดโด่งของเส้นโค้งนั้นตั้งฉากกับแกนนอน

- เส้นโค้งปกติจะเข้าใกล้แกนนอนเมื่อเราต่อปลายทั้งสองข้างให้ห่างจากค่าเฉลี่ยเลขคณิตออกไป แต่จะไม่ตัดแกนนอน

- พื้นที่ใต้โค้งปกติที่ค่าเท่ากับ 1 เสมอ หรือ 100% ค่าเฉลี่ยเลขคณิตจะอยู่ตรงกลางแกนนอน และเป็นจุดที่แบ่งพื้นที่ใต้โค้งออกเป็น 2 ส่วนเท่าๆกัน แต่ละส่วนมีค่า 0.5 หรือ 50%

- เส้นโค้งปกติจะมีเส้นตั้งฉากที่ลากผ่านค่าเฉลี่ยเลขคณิตเป็นแกนสมมาตร

# การหาพื้นที่ภายใต้เส้นโค้งปกติ

พื้นที่ภายใต้เส้นโค้งปกติเหนือแกน x ทั้งหมด มีค่า = 1 พื้นที่ภายใต้เส้นโค้งของตัวแปรใดๆ เช่น พื้นที่ภายใต้เส้นโค้งระหว่าง $x_1$ กับ $x_2$ คือ พื้นที่ที่แรเงาดังรูป




เนื่องจากเส้นโค้งปกติจะมีได้หลายเส้นขึ้นอยู่กับค่าของ $\mu$ และ $\sigma$ ถ้ามีการสร้างตารางแสดงพื้นที่ภายใต้เส้นโค้งปกติ ณ ค่า x ต่างๆ จะมีหลายตาราง ดังนั้น จึงได้มีการแปลงค่า x ต่างๆ ให้เป็นค่ามาตรฐาน

ตารางเส้นโค้งปกติมาตรฐาน คือ ตารางเส้นโค้งปกติที่มีค่าเฉลี่ยเท่ากับ 0 และ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 นั่นคือ ถ้าต้องการคำนวณหาพื้นที่ระหว่าง $x_1$ กับ $x_2$ ก็แปลงค่า $x_1$ และ $x_2$ ให้เป็น $z_1$ และ $z_2$ แล้วนำค่า $z_1$ และ $z_2$ เปิดตารางก็จะได้พื้นที่ที่ต้องการ

เส้นโค้งปกติซึ่งมีค่าเฉลี่ยเลขคณิตเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 เรียกว่า เส้นโค้งปกติมาตรฐาน ในการหาพื้นที่ใต้เส้นโค้งปกติมาตรฐาน ระหว่างค่า Z = 0 ถึง Z ใดๆ เราใช้ตารางแสดงพื้นที่ใต้เส้นโค้งปกติมาตรฐาน ซึ่งแสดงพื้นที่ใต้เส้นโค้งปกติระหว่างค่า Z = 0 และ Z = 0.00 , 0.01, 0.02,...,3.89 เช่น พื้นที่ใต้เส้นโค้งระหว่าง Z = 0 และ Z = 1.45 ที่อ่านได้จากตาราง = 0.4265

ในตารางไม่มีค่า Z ที่เป็นจำนวนลบ แต่สามารถหาพื้นที่ใต้เส้นโค้งปกติมาตรฐาน ระหว่าง Z ที่เป็นจำนวนลบ และ Z = 0 ได้ เนื่องจากเส้นโค้งปกติตั้งฉากกับแกนนอน ซึ่งเป็นแกนสมมาตร เช่น พื้นที่ใต้เส้นโค้งปกติ ระหว่าง Z = -1.45 และ Z = 0 หาได้จากการหาพื้นที่ใต้เส้นโค้งปกติ ระหว่าง Z = 0 ถึง Z = 1.45 ซึ่งจะได้เท่ากับ 0.4265

เนื่องจากพื้นที่ใต้เส้นโค้งปกติ เท่ากับ 1 ดังนั้นพื้นที่ทางขวามือของ Z = 0 กับพื้นที่ทางซ้ายมือของ Z = 0 มีค่าเท่ากัน คือ 0.5 ดังนั้นสามารถหาพื้นที่ใต้เส้นโค้งปกติ ทางขวามือหรือทางซ้ายมือของ Z ใดๆ และพื้นที่ใต้เส้นโค้งปกติ ระหว่างค่า Z สองค่าใดๆ ได้

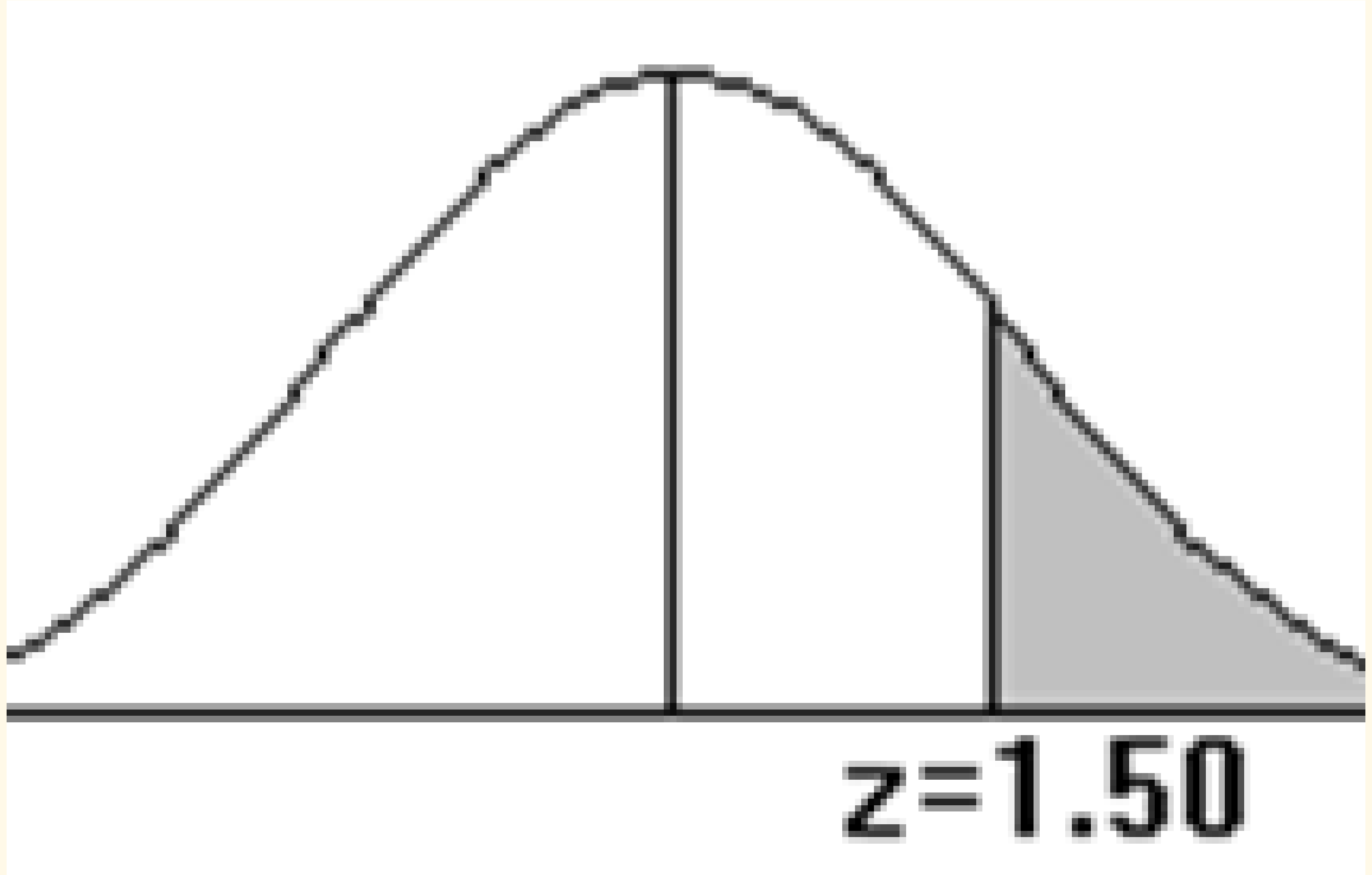
# สรุปการหาพื้นที่ภายใต้เส้นโค้งปกติ

มีขั้นตอนดังต่อไปนี้

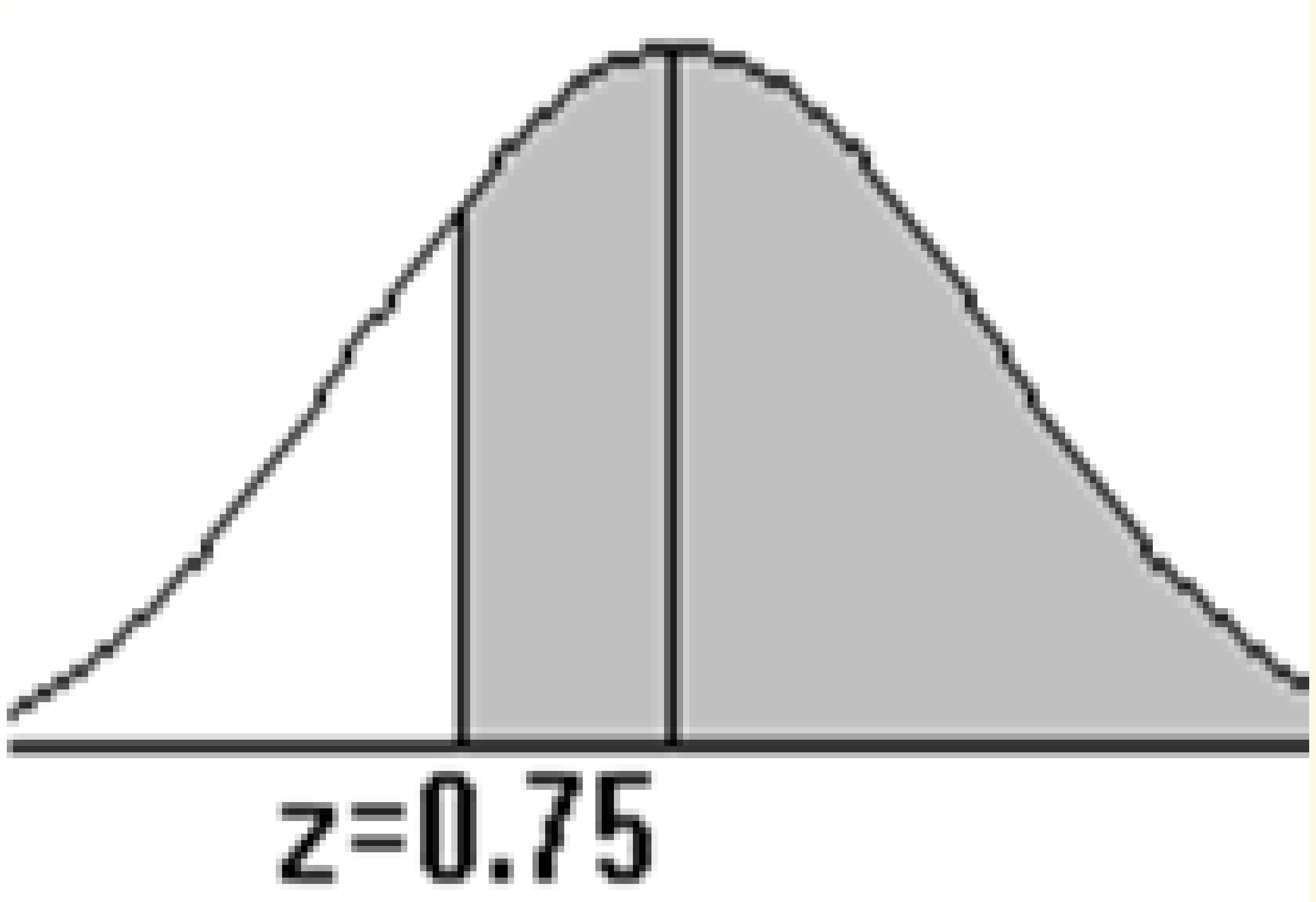
- เปลี่ยนโจทย์หรือข้อมูล( X )ให้อยู่ในรูปคะแนนมาตรฐาน
- นำคะแนนมาตรฐานที่ได้ไปเขียนบริเวณพื้นที่ที่จะหาภายใต้เส้นโค้งปกติ
- เปิดตารางเส้นโค้งปกติมาตรฐานเพื่อหาพื้นที่โดยอาศัยรูปประกอบ

ให้หาพื้นที่ส่วนที่แรเงา **(**เปิดตารางค่า **Z)**

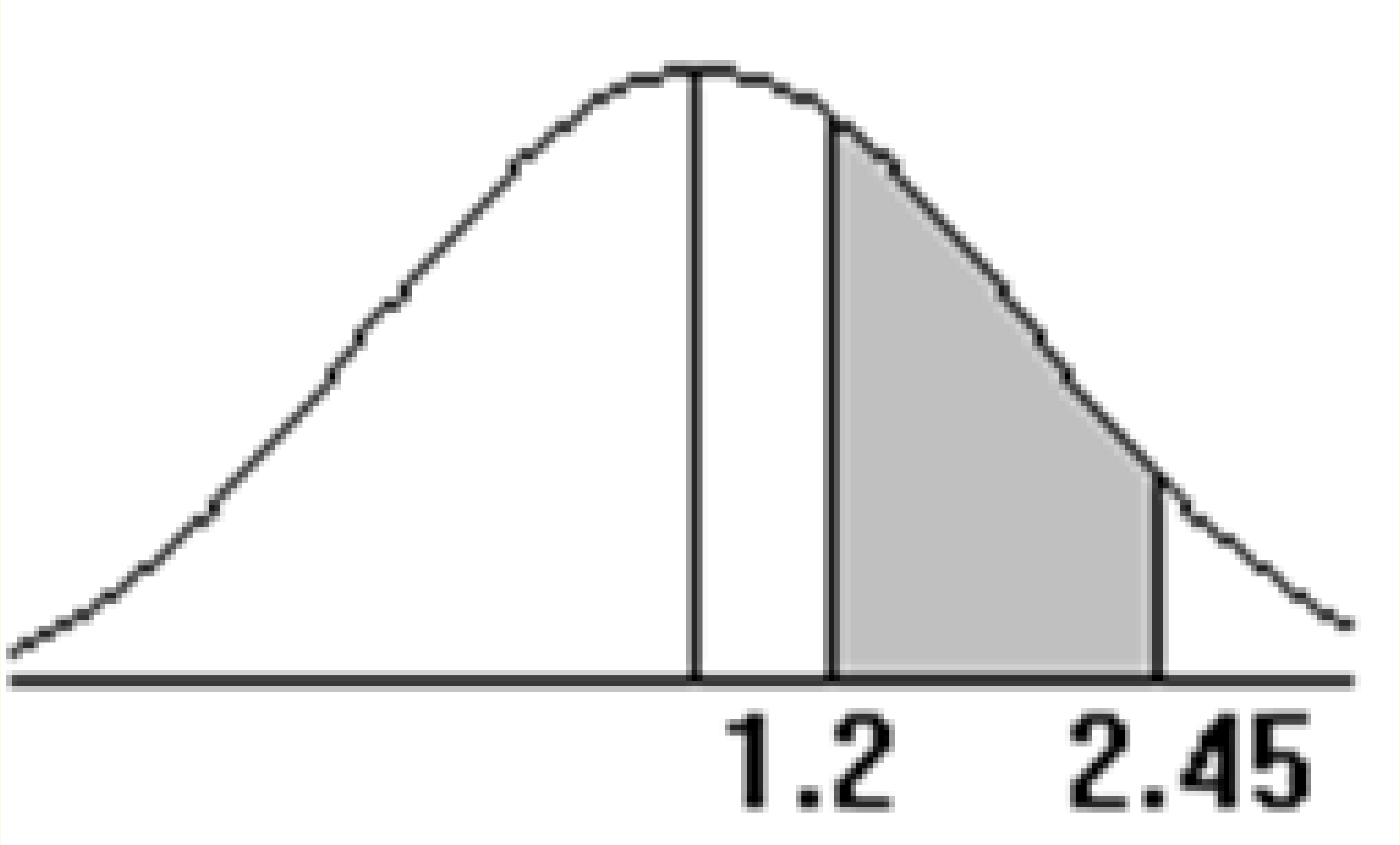
ให้หาพื้นที่ส่วนที่แรเงา (เปิดตารางค่า **Z**)



ให้หาพื้นที่ส่วนที่แรเงา (เปิดตารางค่า **Z**)



ให้หาพื้นที่ส่วนที่แรเงา (เปิดตารางค่า Z)



ให้หาพื้นที่ส่วนที่แรเงา (เปิดตารางค่า Z)

## จงหาพื้นที่ภายใต้เส้นโค้งปกติมาตรฐาน ตามเงื่อนไขต่อไปนี้

1. ระหว่าง $z = 0$ และ $z = 0.75$
2. $-1.3 < z < 0$
3. $-2.5 < z < 2.5$
4. $-1.5 < z < 2.93$ $\otimes$
5. $0.73 < z < 2.15$ $\oplus$
6. $-1.6 < z < -0.4$ $\varnothing$
7. $z < -1.8$
8. $z < 1.32$ $\cup$
9. $z > 2.76$ $\supset$
10. $z > -0.59$