# สถิติและการวิเคราะห์ข้อมูลเบื้องต้น

*นายณพพงษ์ บำรุงพงษ์*
*นักวิชาการสาธารณสุขชำนาญการ*
*สำนักงานป้องกันควบคุมโรคที่ 4 จังหวัดสระบุรี*

# วัตถุประสงค์

1. เพื่อเลือกใช้สถิติพรรณนาในการแสดงการเกิดและการกระจายทางระบาดวิทยาของโรคได้อย่างถูกต้อง
2. เพื่อวิเคราะห์สถานการณ์ แนวโน้มการเปลี่ยนแปลง และตรวจจับการระบาดได้
3. จัดทำรายงานสถานการณ์โรคที่เฝ้าระวังทางระบาดได้

# หัวข้อสำคัญ

1. อัตรา อัตราส่วน และสัดส่วน
2. Mean Median SD Range
3. การนำเสนอข้อมูลด้วย กราฟมาตรส่วนเรขาคณิต, Histogram, Bar chart, Pie chart, Spot map
4. การวิเคราะห์สถานการณ์โรค
5. การเขียนรายงานสถานการณ์โรค

# ระบาดวิทยา (Epidemiology)

การศึกษาเกี่ยวกับการเกิดและการกระจายของปัญหาสุขภาพในประชากรที่จำเพาะ ตลอดจนปัจจัยที่มีอิทธิพลต่อการเกิดปัญหา และการนำความรู้ไปประยุกต์ใช้ควบคุมปัญหาสุขภาพดังกล่าว

**"The study of the occurrence and distribution of health-related states, events, and processes in specified populations, including the study of the determinants influencing such processes, and the application of this knowledge to control relevant health problems"**

**Source: Dictionary of Epidemiology, 6th Ed. , 2014**

# การวัดทางระบาดวิทยา

- DESCRIBE → Measure of Frequency
  - การวัดขนาดของโรคหรือปัญหาทางด้านสาธารณสุข
  - ผู้ป่วยด้วยโรคไข้เลือดออกมีจำนวนเท่าใดในจังหวัด ก ในปี 2558
- EXPLAIN → Measure of Association
  - การวัดความสัมพันธ์ระหว่างปัจจัยและการเกิดโรค
  - การสูบบุหรี่เพิ่มความเสี่ยงในการเป็นโรคหัวใจหลอดเลือดหรือไม่
- PREDICT → Measure of Impact
  - การวัดผลกระทบของการมีปัจจัยหรือไม่มีปัจจัยต่อการเกิดโรค
  - วัคซีนป้องกันโรคหัดมีประสิทธิภาพในการป้องกันการป่วยได้ดีเพียงใด
- CONTROL
  - มาตรการที่เหมาะสมสำหรับชุมชน (ภายใต้ข้อจำกัดต่างๆ) คืออะไร

# Measure of Frequency

# Frequency of Health Problem

แสดงได้ 2 ลักษณะ

## 1. จำนวนผู้ป่วย (หรือจำนวนผู้เสียชีวิต)

- เข้าใจง่าย บอกถึงปริมาณปัญหาที่แท้จริง (ที่ต้องจัดเตรียมทรัพยากรรองรับ)
- แต่ไม่สะท้อนความเสี่ยงของปัญหาโดยตรง หากเปรียบเทียบกับประชากรอื่นที่ขนาดแตกต่างกัน

## 2. อัตราป่วย (หรืออัตราตาย)

- ใช้เปรียบเทียบระหว่างประชากรแต่ละกลุ่ม
- ตัวตั้ง คือ จำนวนผู้ป่วย (หรือตาย)
- ตัวหาร คือ จำนวนประชากรผู้มีโอกาสเกิดโรค (Population at risk)
- นิยมแสดงเป็นจำนวน ต่อประชากร 100,000 คน

# Ratio / Proportion / Rate

- **Ratio** (อัตราส่วน)
  - เท่ากับ A / B โดยที่ A จะเป็นหรือไม่เป็นส่วนหนึ่งของ B ก็ได้ และไม่จำเป็นต้องมีหน่วยเดียวกัน
  - เช่น การตายมารดาต่อแสนการเกิดมีชีพ หรือ อัตราส่วนจำนวนเตียงผู้ป่วยต่อจำนวนประชากร
- **Proportion** (สัดส่วน)
  - เท่ากับ A / B โดยที่ A เป็นส่วนหนึ่งของ B และมีหน่วยเดียวกัน
  - มีค่าอยู่ในช่วงตั้งแต่ 0 ถึง 1 (แต่อาจจะแสดงเป็น ต่อร้อย ต่อหมื่น ต่อแสน ฯลฯ)
  - เช่น สัดส่วนของเพศชายในประชากรทั้งหมด
- **Rate** (อัตรา)
  - ในวิชาคณิตศาสตร์ คือ  การเปลี่ยนแปลงของปริมาณหนึ่งเทียบกับอีกปริมาณหนึ่ง
  - ในระบาดวิทยา คือ จำนวนผู้ป่วยที่เกิดขึ้น (A) / เวลาที่ทำการสังเกตุ (Time)

# การวัดอัตราการเกิดโรค

การนับจำนวนการเกิดโรค:

- ไม่สามารถนำมาเปรียบเทียบกันได้
- ไม่สามารถระบุความเสี่ยงของการเกิดโรค

**ใช้อัตราในการเปรียบเทียบ**

# อัตรา

X = จำนวนคนที่เกิดเหตุการณ์ภายใต้เงื่อนเวลา
Y = จำนวนคนที่เกี่ยวข้องกับเหตุการณ์ภายใต้เงื่อนเวลาเดียวกัน
K = ค่าคงที่ กำหนดให้มีค่า 100, 1,000, 10,000, 100,000
โดยทั่วไปเลือกใช้ค่าที่ให้ผลลัพธ์เป็นจำนวนเต็มอย่างน้อย 1 หลัก เช่น 5.4/1,000 (ไม่ใช้ .54/1,000)

$$\frac{x}{y} * k$$

x = คนที่ป่วยเป็นโรคเบาหวาน
y = คนที่เป็นกลุ่มเสี่ยงต่อการป่วยเป็นโรคเบาหวาน

ตัวตั้งเป็นส่วนหนึ่งของตัวหาร

# ข้อพึงระวังในการใช้คำว่า อัตรา

โดยทั่วไป มักมีการใช้คำว่า อัตรา ในการแสดงถึงขนาดของโรคในประชากร เช่น

อัตราป่วย (Morbidity rate) ของโรคไข้เลือดออกของประเทศไทย พ.ศ. 2554 เท่ากับ 107.06 ต่อประชากรแสนคน

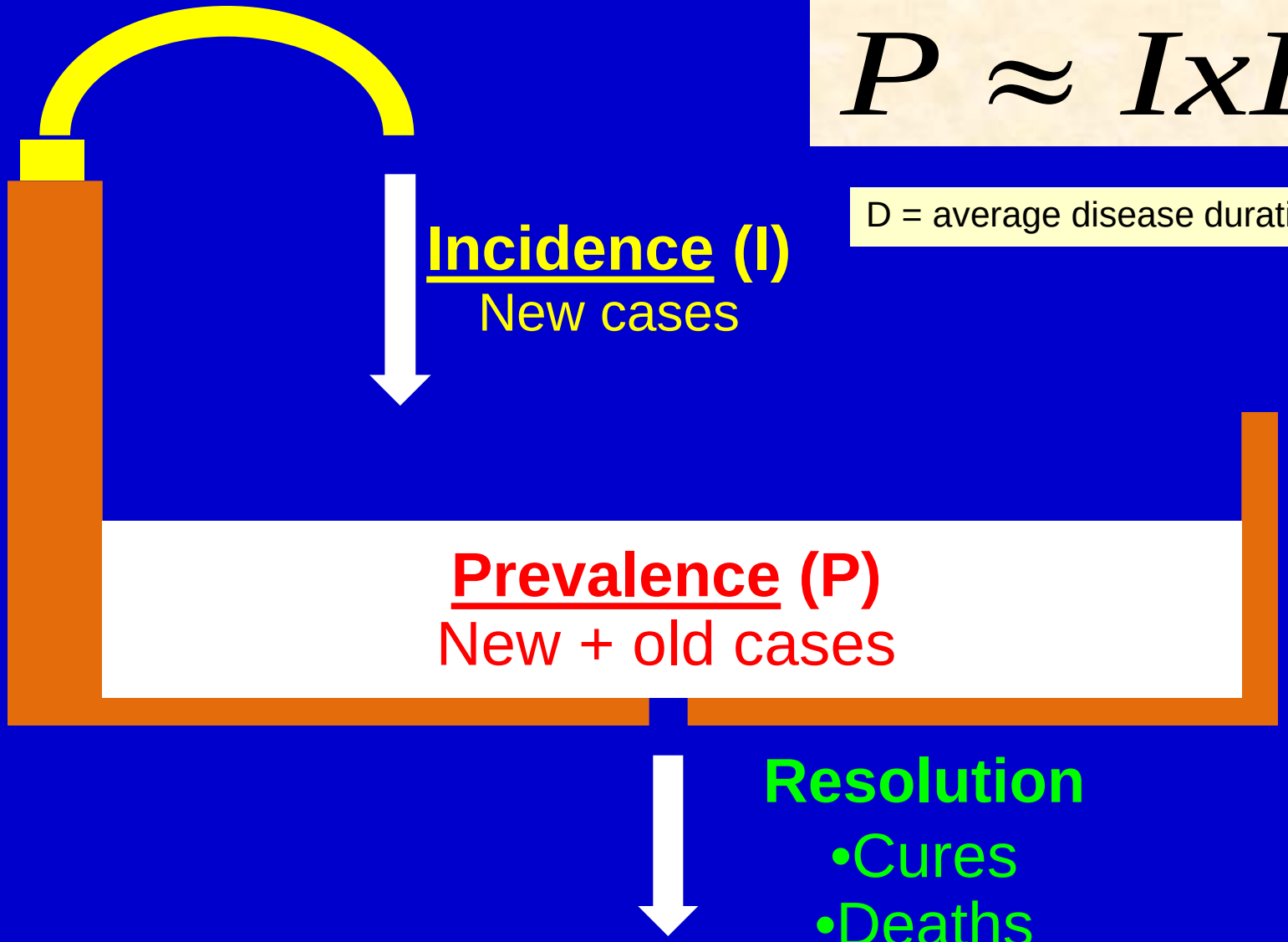
อัตราตาย (Mortality rate) ของโรคไข้เลือดออกของประเทศไทย พ.ศ. 2554 เท่ากับ 0.11 ต่อประชากรแสนคน

*ในที่นี้ ไม่ใช่ อัตรา (Rate) ในความหมายที่ถูกต้องทางระบาดวิทยา เป็นแต่เพียง สัดส่วน เท่านั้น เนื่องจากไม่ได้พิจารณาร่วมกับระยะเวลาในการติดตามประชากร*

# Prevalence and Incidence

$$P \approx IxD$$

D = average disease duration



# อุบัติการณ์ (Incidence)

จำนวนผู้ป่วยรายใหม่ในประชากรกลุ่มเสี่ยง ณ ช่วงเวลาใดเวลาหนึ่ง

*(สนใจเฉพาะรายใหม่)*

# ความสำคัญของอุบัติการณ์ (Incidence)

1. ทำให้ทราบถึงความเสี่ยง (risk) หรือความเร็ว (rate) ของการเกิดโรคของคนในชุมชนที่จะเกิดโรคในช่วงเวลาหนึ่ง
2. นิยมใช้กับโรคเฉียบพลัน หรือโรคเรื้อรังที่สามารถระบุเวลาที่เริ่มเป็นโรคได้
3. เหมาะสมที่จะใช้ในการศึกษาเพื่อหาสาเหตุของการเกิดโรค (Etiologic study) หรือประเมินผลของมาตรการควบคุมป้องกันโรค

# ชนิดของอุบัติการณ์ (Incidence)

**Incidence Proportion*** (Risk, Cumulative Incidence, Attack Rate)

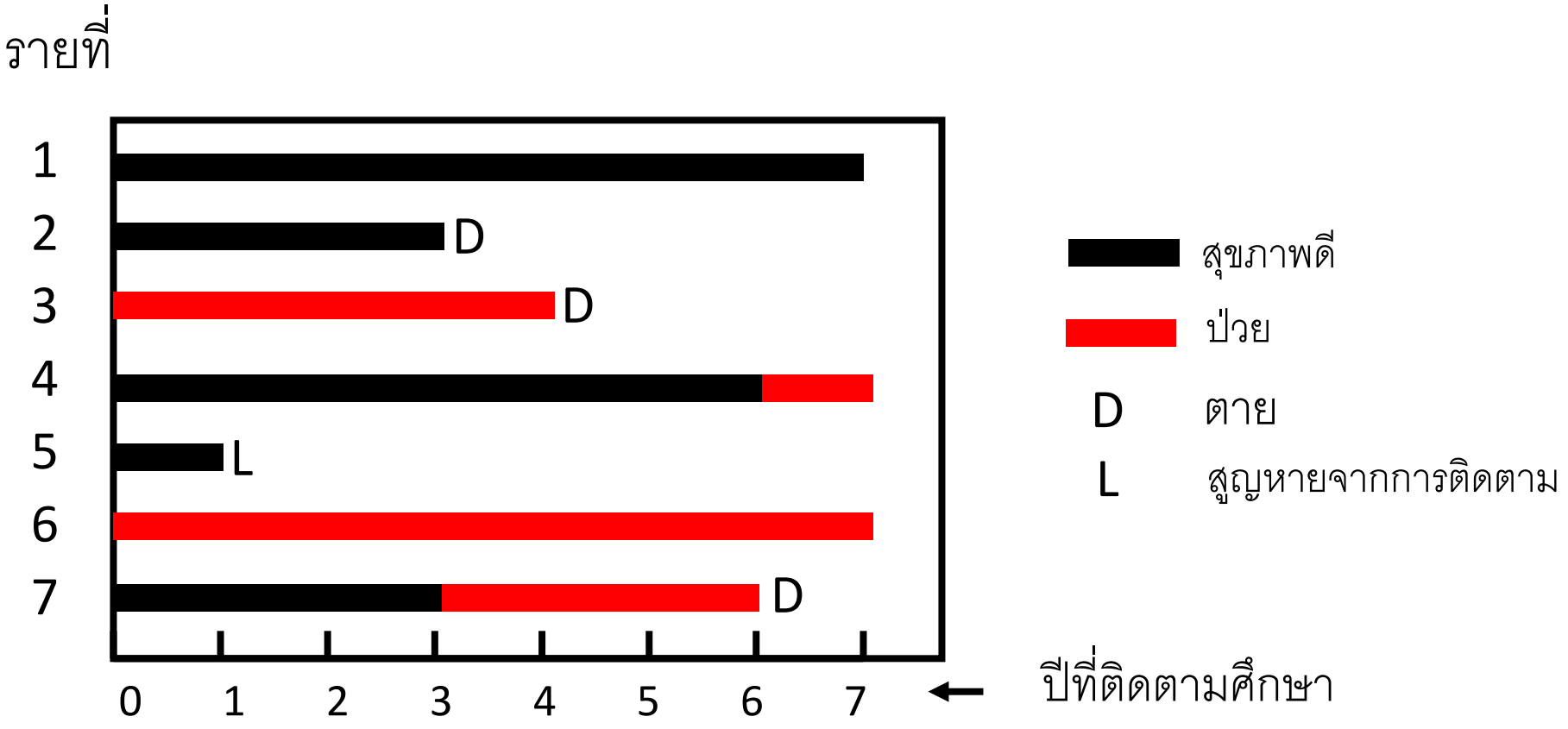
$$= \frac{\text{จำนวนของผู้ป่วยด้วยโรคที่สนใจรายใหม่ที่ค้นพบในช่วงระยะเวลาหนึ่ง}}{\text{จำนวนประชากรที่มีโอกาสเกิดโรค (at risk) ณ ตอนเริ่มต้นศึกษา}}$$

**Incidence Rate** (Rate, Incidence Density, Person-Time Incidence)

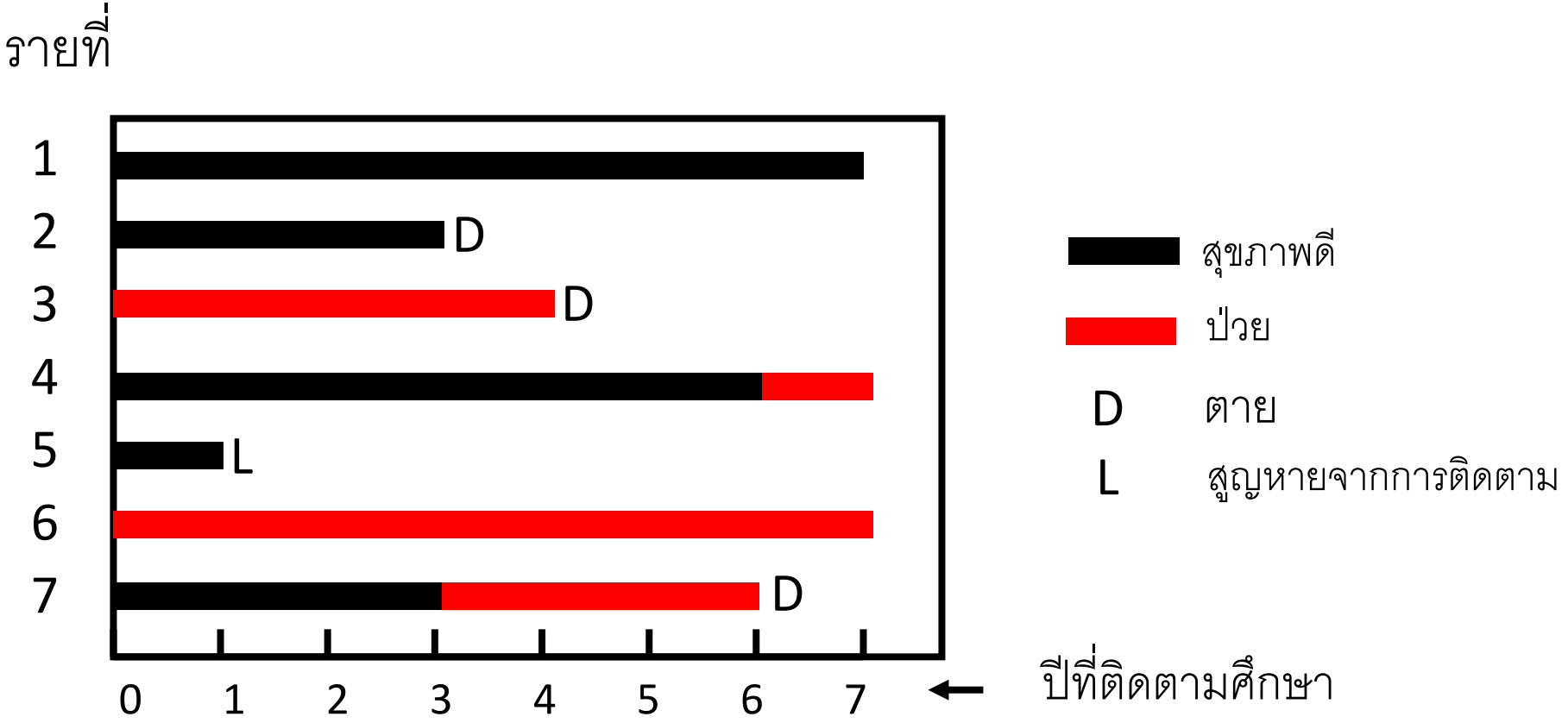
$$= \frac{\text{จำนวนของผู้ป่วยด้วยโรคที่สนใจรายใหม่ระหว่างระยะเวลาที่กำหนด (D)}}{\text{จำนวน person-time (PT) ที่มีโอกาสเกิดโรค (at risk) ทั้งหมดที่ติดตามศึกษา}}$$

* ใช้ในกรณีที่ 1. สามารถติดตามประชากรที่อยู่ในการศึกษาได้หมดหรือเกือบทั้งหมดในช่วงเวลาที่ศึกษา และ 2. ต้องระบุเวลาที่ทำการติดตามเสมอ

# การติดตามประชากร 7 คนเป็นเวลา 7 ปี



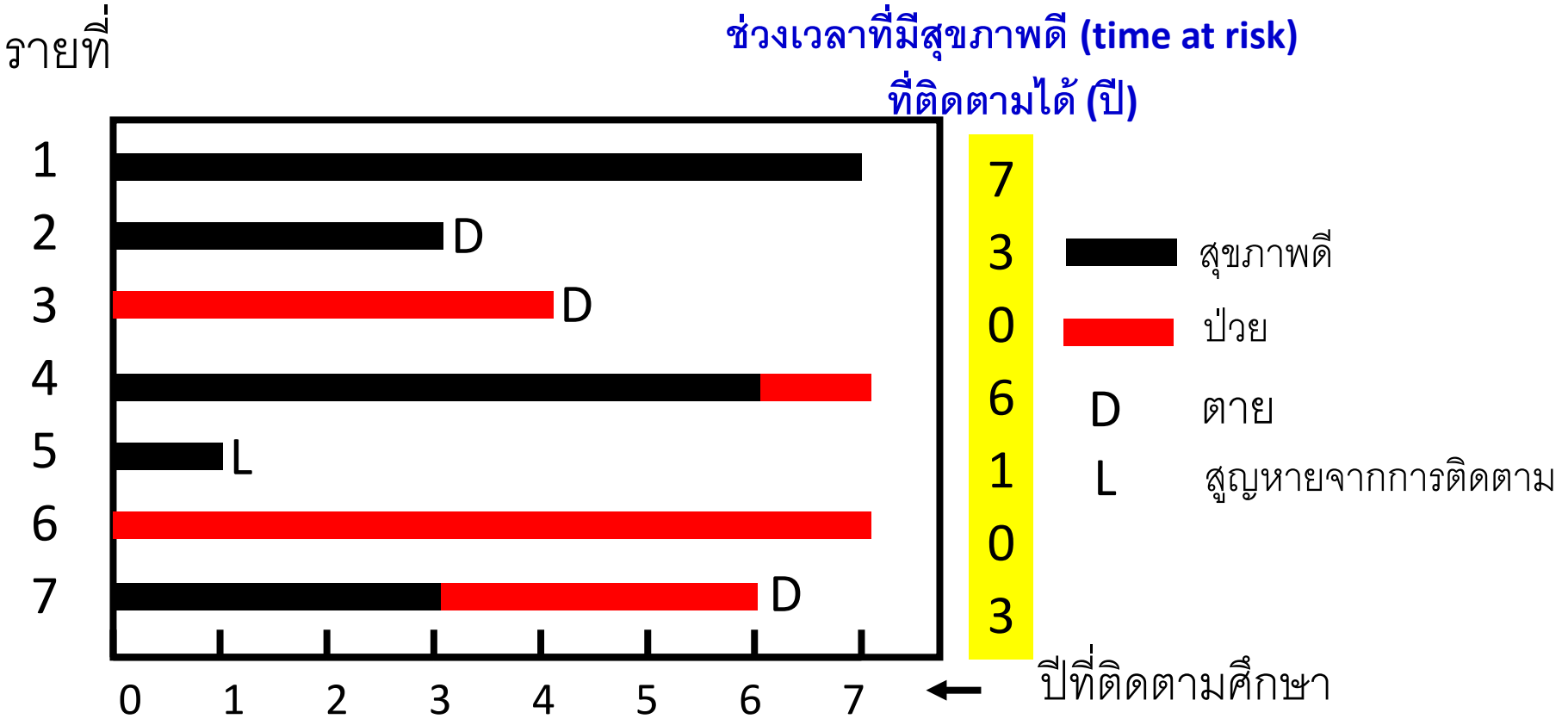
# ตัวอย่างของ Incidence Proportion




**ความเสี่ยงของการเกิดโรคใน 7 ปี = 2/5 = 0.4 = 40%**

- ตัวตั้ง คือ คนที่ 4 และ 7 เนื่องจากเป็นผู้ป่วยรายใหม่
- ตัวหาร คือ นำผู้ป่วย คนที่ 3 และ 6 ไปลบกับประชากรทั้งหมด 7-2 เพราะเป็นผู้ป่วยเก่าตั้งแต่เริ่ม

# ตัวอย่างของ Incidence Rate




**เวลาที่ติดตาม = 7+3+0+6+1+0+3 = 20 คน-ปี**

**incidence rate ในช่วงเวลาที่ติดตาม 7 ปี**

**= incident cases / PT = 2 / 20 = 0.10 / คน-ปี**

<u>แปลผล</u>

เมื่อติดตามกลุ่มเสี่ยงไป 20 Person/Year

จะพบผู้ป่วย 2 ราย (คนที่ 4 และ 7)

หรือต้องรอเวลาเท่าไหร่ถึงจะเจอผู้ป่วย 1 ราย

= 1/Incidence rate คือ 1*20/2=10 ปี

# ความชุก (Prevalence)

จำนวนผู้ป่วยด้วยโรค ๆหนึ่งในประชากรกลุ่มเสี่ยง ณ ช่วงเวลาใดเวลาหนึ่ง *(นับรวมทั้งรายใหม่และรายเก่า)*

# ความสำคัญของความชุก (Prevalence)

1. เป็นการบอกสภาวะสุขภาพของชุมชนในขณะนั้น เพื่อการวางแผนการจัดสรรทรัพยากรหรือการจัดบริการทางการแพทย์แก่ชุมชนได้ถูกต้อง
2. นิยมใช้กับโรคเรื้อรัง หรือโรคที่ไม่สามารถระบุเวลาที่เริ่มเป็นโรคได้อย่างชัดเจน
3. ไม่เหมาะสมที่จะใช้ในการศึกษาเพื่อหาสาเหตุของการเกิดโรค (Etiologic study)

# Prevalence

## 1. Point prevalence

- ความชุก ณ จุดเวลา (ในทางปฏิบัติ คือ ช่วงเวลาที่แคบมากๆ)
- เช่น อัตราความชุกของผู้ป่วยเบาหวาน ณ วันที่ 1 ตุลาคม 2552

$$= \frac{\text{จำนวนผู้ป่วยเบาหวาน ณ วันที่ 1 ตุลาคม 2552}}{\text{จำนวนประชากร ณ วันที่ 1 ตุลาคม 2552}} \times 100{,}000$$

Population at risk

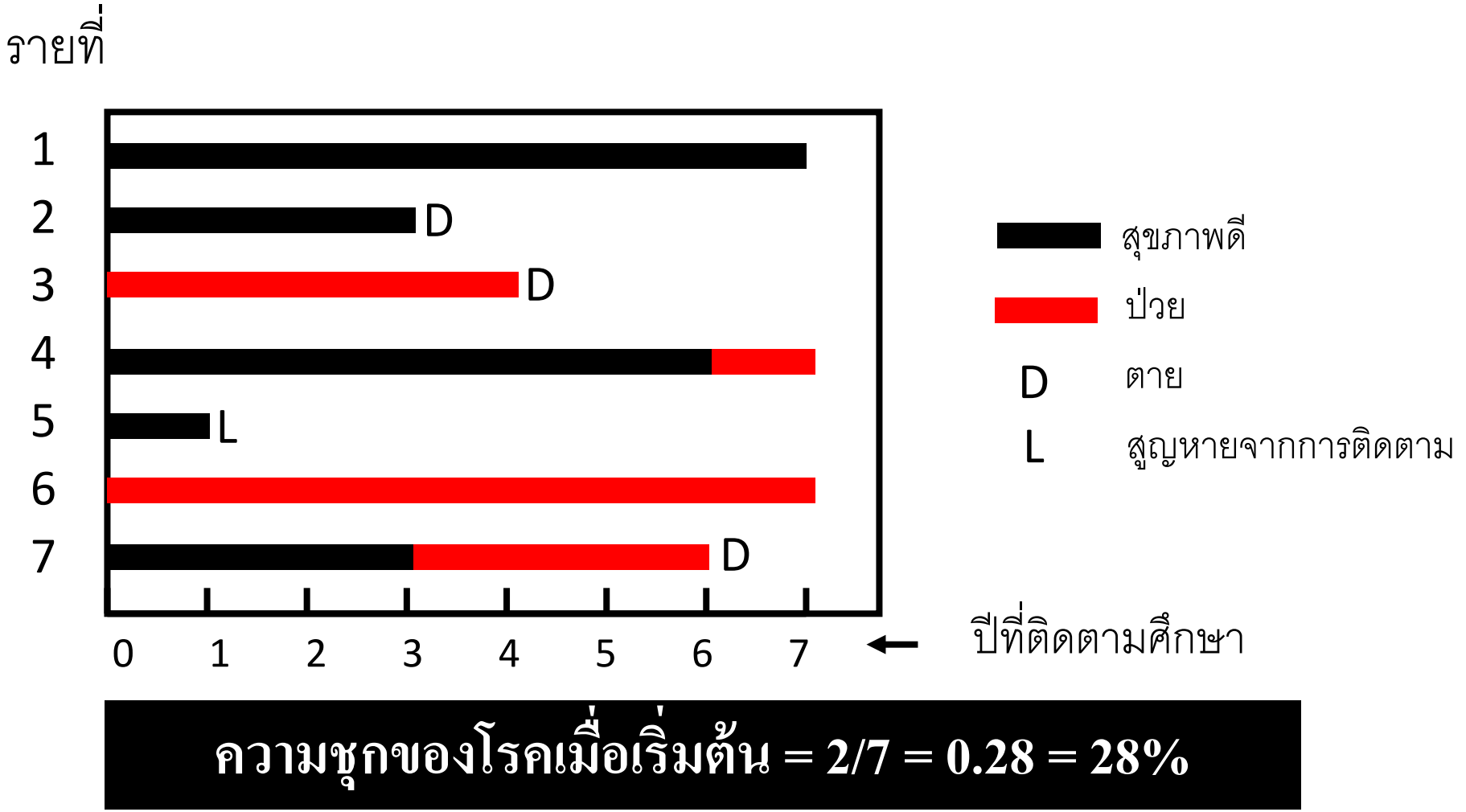
## 2. Period prevalence

- ความชุก ณ ช่วงเวลา
- เช่น อัตราความชุกของผู้ป่วยโรคความดันโลหิตสูงในปี 2552

$$= \frac{\text{จำนวนผู้ป่วยเมื่อ 1 ม.ค. 2552} + \text{ผู้ป่วยใหม่ในปี 2552}}{\text{จำนวนประชากรกลางปี 2552}} \times 100{,}000$$

Population at risk

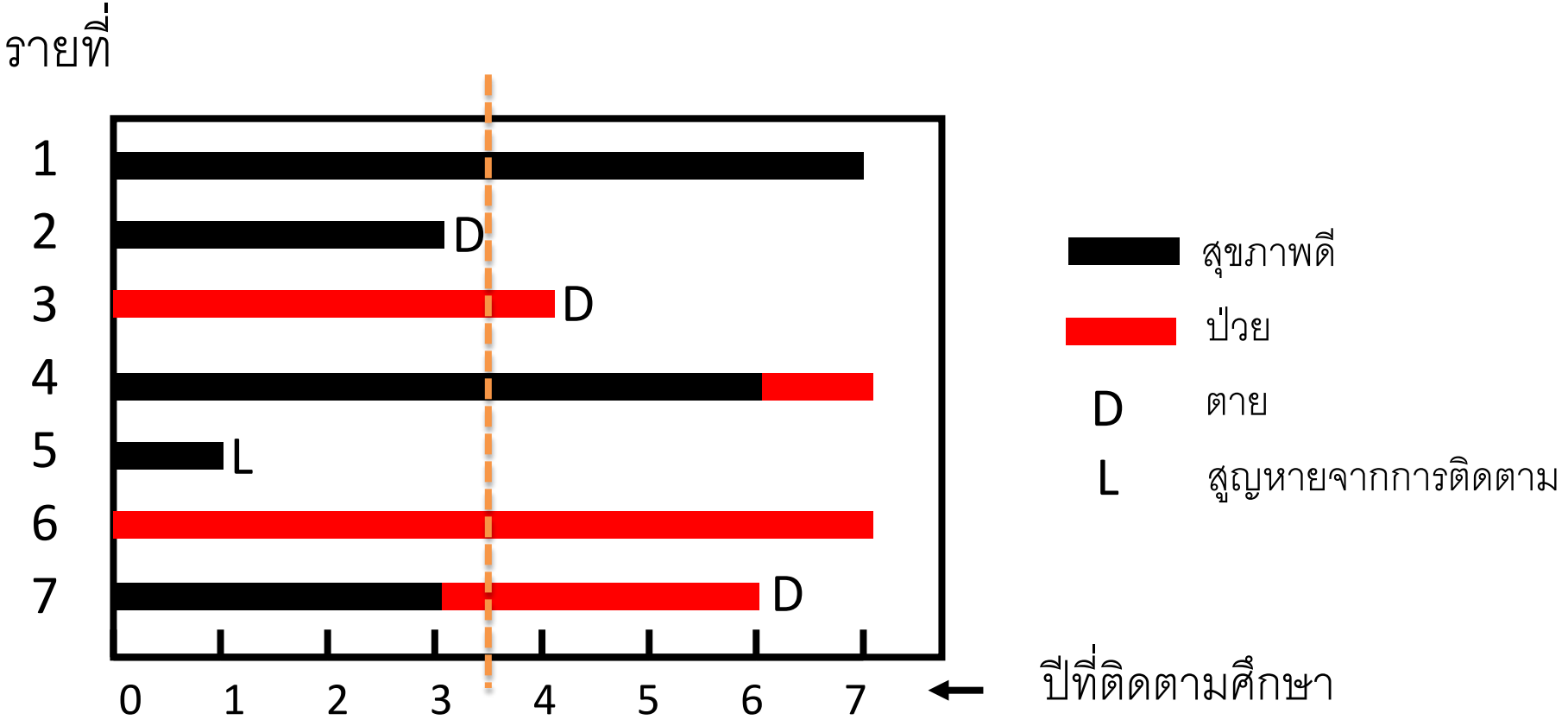
# ตัวอย่างของ Point Prevalence




**ความชุกของโรคเมื่อเริ่มต้น = 2/7 = 0.28 = 28%**

- ตัวตั้ง คือ คนที่ 3 และ 6 เนื่องจากเป็นผู้ป่วยตั้งแต่เริ่มต้น
- ตัวหาร คือ จำนวนประชากรทั้งหมด 7 คน

# ตัวอย่างของ Period Prevalence




**ความชุกของโรคใน 7 ปี = 4 / 5 = 0.8 = 80%**

- ตัวตั้ง คือ คนที่ 3 4 6 และ 7
- ตัวหาร คือ ประชากรกึ่งกลางช่วงเวลาที่ศึกษา 5 คน

# Attack Rate

คือ สัดส่วนของประชากรที่มีการป่วยในประชากรกลุ่มเสี่ยง ณ ช่วงเวลาที่กำหนด นั่นก็คือ Incidence proportion ซึ่งนิยมใช้ในการสอบสวนโรค

A.R. = X/Y * 100

X = จำนวนผู้ป่วย
Y = ประชากรกลุ่มเสี่ยง

# ตัวอย่าง

การระบาดของโรคไข้หวัดใหญ่ พบผู้ป่วย 26 คน จากห้องเรียนที่มีนักเรียนทั้งหมด 96 คน ให้คำนวณหาอัตราป่วย (attack rate)

วิธีคำนวณ Attack rate = X / Y x 100

เมื่อ X = 26 ราย , Y = 96 คน

Attack rate = 26 / 96 x 100

= 27.1 %

# อัตราการตาย (Mortality Rates)

- อัตราตายอย่างหยาบ
- อัตราตายเฉพาะ
- อัตราผู้ป่วยตาย

X = ตัวตั้งเหมือนกัน
Y = ตัวหารต่างกัน

# อัตราตายอย่างหยาบ (Crude Mortality Rate)

จำนวนผู้เสียชีวิตในประชากรทั้งหมดในช่วงระยะเวลาหนึ่ง

$$C.D.R. = X/Y * k$$

# ตัวอย่าง

Crude death rate (C.D.R.) ประเทศไทย พ.ศ. 2554 มี ประชากรกลางปีจำนวน 49,459,000 คน เสียชีวิตในปี 2554 จำนวน 252,592 ราย หาอัตราการตายอย่างหยาบ

เมื่อ X = 252,592 ราย Y = 49,459,000 คน

k = 100,000

ตายรวมไม่ทราบสาเหตุ

C.D.R. = 252,592 / 49,459,000 x 100,000

= 510.7 / ประชากร 100,000 คน

# อัตราตายจำเพาะ (Specific Mortality Rate)

จำนวนผู้เสียชีวิตด้วยโรคหรือภาวะใดภาวะหนึ่งในประชากรกลุ่มเสี่ยงในช่วงระยะเวลาที่หนึ่ง

เช่น อายุ เพศ และเชื้อชาติ สาเหตุ ในกลุ่มประชากรที่กำหนดในช่วงเวลาหนึ่ง

ตัวหารเฉพาะ

$$S.D.R. = X/Y * k$$

X = ตายในกลุ่มอายุ.....ปี

Y = ปชก.กลุ่มอายุ.....ปีด้วย

# ตัวอย่าง

Specific death rate (S.D.R.) ประเทศไทย พ.ศ. 2554

เมื่อ X = จำนวนคนตายในกลุ่มอายุ 0-4 ปีในปี 2554

= 25,583 ราย

Y = ประชากรกลางปีกลุ่มอายุ 0-4 ปี

= 6,387,000 คน

k = 100,000

**หาอัตราตายกลุ่มอายุ 0-4 ปี ประเทศไทย พ.ศ.2554**

# อัตราตายเฉพาะ (Specific Mortality rate)

**ประโยชน์**

- ใช้บ่งชี้สภาวะอนามัย บริการทางการแพทย์ สิ่งอำนวยความสะดวกด้านสุขภาพอนามัย สภาวะเศรษฐกิจและสังคม และอนามัยสิ่งแวดล้อมของชุมชน

- ใช้เปรียบเทียบได้ดีกว่า C.D.R.

- บอกลักษณะเฉพาะของโรคหรือการเสี่ยงต่อการตายด้วยโรคได้ดีกว่า

# อัตราป่วยตาย (Case Fatality Rate)

$$CFR = \frac{X}{Y} \times k$$

X = จำนวนผู้ป่วยที่ตายด้วยโรคนี้

Y = จำนวนผู้ป่วยด้วยโรคนี้

k = 100

**ความสำคัญ 1. ใช้บ่งชี้ถึงความรุนแรงของโรค**

**2. ใช้บ่งชี้ถึงคุณภาพของบริการทางการแพทย์**

# ตัวอย่าง

## CFR โรคบาดทะยักในเด็กแรกเกิด ประเทศไทย พ.ศ. 2553

เมื่อ X = จำนวนเด็กตายด้วยโรคบาดทะยักในเด็กแรกเกิด

= 122 ราย

Y = จำนวนเด็กป่วยด้วยโรคบาดทะยักในเด็กแรกเกิด

= 566 ราย

CFR = 122 / 566 x 100

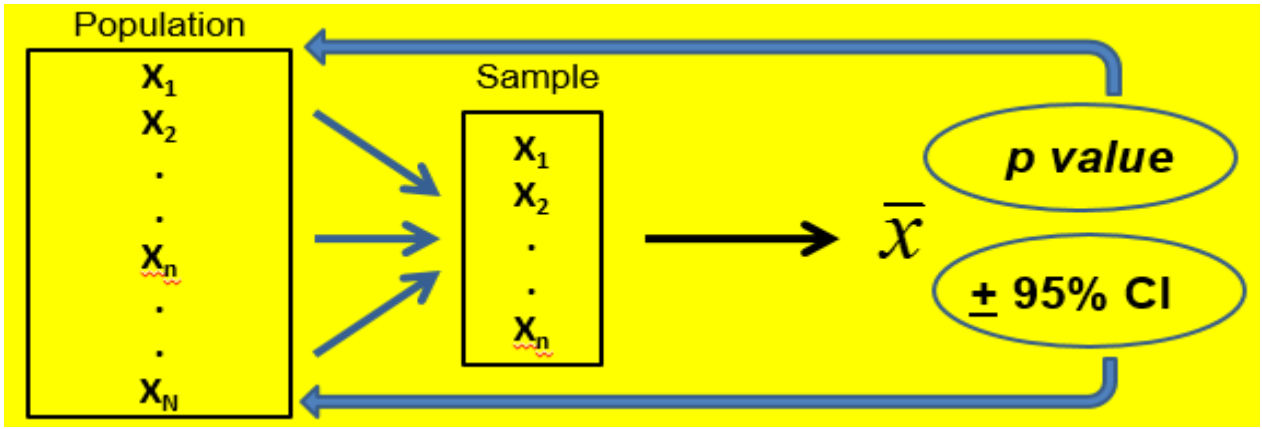
= 21.6 %

**ป่วย 100 คน ตาย 21.6 คน**
**หรือประมาณ 1 ใน 5**
**ป่วย 5 คน ตาย 1 คน**

# สถิติเชิงพรรณนา
# (Descriptive Statistics)

**ประชากร (population)** หมายถึง ***<u>ขอบเขตของข้อมูลทั้งหมดที่เรากำลังทำการศึกษา</u>*** หรืออาจหมายถึง กลุ่มของสิ่งของทั้งหมดที่ให้ข้อมูลตามที่เราต้องการศึกษา เช่น ศึกษาเกี่ยวกับคนไข้สูติ-นรีเวชของโรงพยาบาลมหาราชนครเชียงใหม่ ในปี 2548 ทั้งหมด ซึ่งอาจดูได้จากประวัติ ผู้ป่วย เป็นต้น

**ตัวอย่าง (sample)** หมายถึง ***<u>ส่วนหนึ่งของประชากรซึ่งถูกเลือกมาศึกษา</u>*** เนื่องจากในบางครั้ง พบว่า การศึกษาบางอย่างไม่อาจทำทั้งหมดของประชากรได้ เพราะต้องเสียค่าใช้จ่ายมาก เสียเวลา และอาจหาประชากรทั้งหมดไม่ได้ หรือไม่สามารถทำกับประชากรทั้งหมดได้ จึงจำเป็นต้องเลือกตัวอย่างมาศึกษา



# ตัวแปร

คุณลักษณะหรือคุณสมบัติของสิ่งต่าง ๆ อาจเป็นสิ่งมีชีวิต หรือ สิ่งไม่มีชีวิตซึ่งให้ค่าที่แตกต่างกัน เช่น เพศ อายุ ความเครียด ระดับไขมันในเลือด เป็นต้น

**ส่วนที่ 2 ข้อมูลทั่วไป**

1. สถานที่ศึกษา ☐ โรงเรียน ☐ วิทยาลัย.....

2. ขณะนี้ท่านมีอายุ ..... ปี ..... เดือน

3. ภูมิลำเนาเดิม ☐ กรุงเทพฯ และปริมณฑล ☐ ภาคกลาง ☐ ภาคเหนือ

☐ ภาคอีสาน ☐ ภาคใต้ ☐ ภาคตะวันออก ☐ ภาคตะวันตก

4. กำลังศึกษาในระดับใด ☐ มัธยมศึกษาปีที่..... ☐ วิทยาลัยปีที่ .....

5. รายได้ของท่านโดยเฉลี่ยต่อเดือน ..... บาท และรายได้เพียงพอใช้จ่ายหรือไม่

☐ เพียงพอ ☐ ไม่เพียงพอ

6. แหล่งของรายได้ ☐ จากบิดา มารดา ☐ จากการทำงานด้วยตัวเอง ☐ อื่นๆ โปรดระบุ.....

7. ปัจจุบันสถานะการสูบบุหรี่ของท่าน ☐ 1. ไม่เคยสูบบุหรี่ ☐ 2. กำลังสูบบุหรี่ ☐ 3.เลิกสูบแล้ว

# ลักษณะของตัวแปรทางสถิติ

- แบ่งตามลักษณะของตัวแปร
  - เชิงปริมาณ
  - เชิงคุณภาพ
- แบ่งตามมาตรการวัดตัวแปร
  - นามมาตรา (Nominal scale)
  - อันดับมาตรา (Ordinal scale)
  - อันตรภาคมาตรา (Interval scale)
  - อัตราส่วนมาตรา (Ratio scale)

# ลักษณะของตัวแปร

## ตัวแปรเชิงปริมาณ (Quantitative Variable)

- มีความแตกต่างกันตามความถี่ จำนวน หรือปริมาณมากน้อย
- สามารถเรียงลำดับเปรียบเทียบได้ว่านามใดดีกว่าหรือด้อยกว่าอีกนามหนึ่ง
- เช่น อายุ น้ำหนัก ส่วนสูง ระดับไขมันในเส้นเลือด คะแนนสอบ

## ตัวแปรเชิงคุณภาพ (Qualitative Variable)

- มีคุณสมบัติแตกต่างกันในแง่ของชนิดหรือประเภทหรือคุณลักษณะ
- เช่น เพศ อาชีพ ภูมิลำเนา ภาวะทางเศรษฐกิจ ความรุนแรงของโรค

# มาตรการวัดตัวแปร

## นามมาตรา (Nominal Scale)

- เป็นการจำแนกข้อมูลโดยใช้นามเป็นเกณฑ์
- ไม่สามารถเปรียบเทียบได้ว่านามหนึ่งดีกว่านามหนึ่ง
- เช่น เพศ อาชีพ

## อันดับมาตรา (Ordinal Scale)

- เป็นการจำแนกข้อมูลโดยใช้นามเป็นเกณฑ์
- สามารถเรียงลำดับเปรียบเทียบได้ว่านามใดดีกว่าหรือด้อยกว่าอีกนามหนึ่ง
- เช่น ระดับการศึกษา ระดับความนิยม ระดับรายได้ ความรุนแรงของโรค ผลการเรียนวิชาสถิติ (A, B, C, D)

## อันตรภาคมาตรา (Interval Scale)

- เป็นการจำแนกข้อมูลโดยใช้นามและลำดับ
- สามารถวัดปริมาณความแตกต่างของนามได้
- ไม่มี “0” ที่แท้จริง
- เช่น อุณหภูมิ ไอคิว ผลการเรียนวิชาสถิติ (0-100)

## อัตราส่วนมาตรา (Ratio Scale)

- สามารถเปรียบเทียบอัตราส่วนระหว่างปริมาณได้
- มี “0” ที่แท้จริง
- เช่น รายได้ อายุ น้ำหนัก

# สถิติเชิงพรรณนา (Descriptive Statistics)

- การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measure of Central Location)
  - เพื่อหาค่าที่เป็นตัวแทนที่ดีที่สุดของกลุ่มประชากร
- การวัดการกระจาย (Measure of dispersion)
  - เพื่อดูว่าประชากรแต่ละคนมีความแตกต่างกันมากน้อยเพียงใด
  - เพื่อดูความแตกต่างของประชากรแต่ละคนจากค่ากลาง

# การวัดแนวโน้มเข้าสู่ส่วนกลาง
# Measure of Central Location

- Mean (Arithmetic Mean) ค่าเฉลี่ยเลขคณิต
- Median มัธยฐาน
- Mode ฐานนิยม

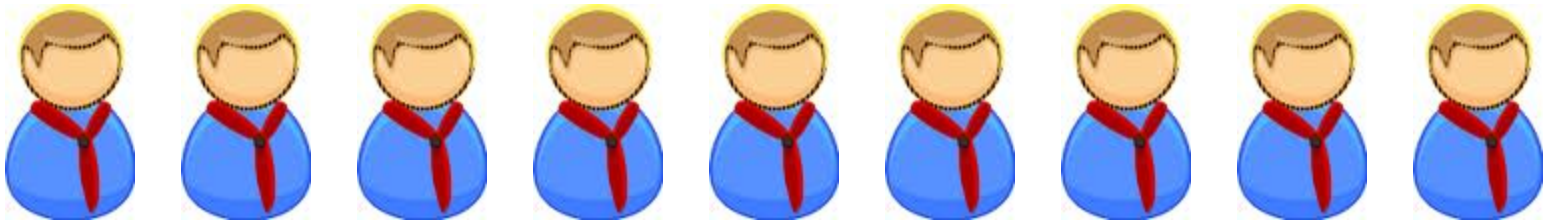
# ค่าเฉลี่ยเลขคณิต (Arithmetic Mean)

- ค่าเฉลี่ย (Mean หรือ Average)
- ผลรวมของข้อมูลหารด้วยจำนวนข้อมูล
- สัญลักษณ์ $\overline{x}$

$$\overline{x} = \frac{\sum x_i}{n}$$

# ตัวอย่าง

จงคำนวณค่าเฉลี่ยน้ำหนักตัวของนักเรียน 9 คน



24 28 17 32 27 25 22 19 31

$$\bar{x} = \frac{\sum x_i}{n} = \frac{24+28+17+32+27+25+22+19+31}{9}$$

$$= \frac{225}{9} = 25 \text{ kg}$$

ค่าเฉลี่ย คือ 25 kg

# ตัวอย่าง

จงคำนวณค่าเฉลี่ยน้ำหนักตัวของนักเรียน 9 คน

24 28 17 32 27 25 22 19 90

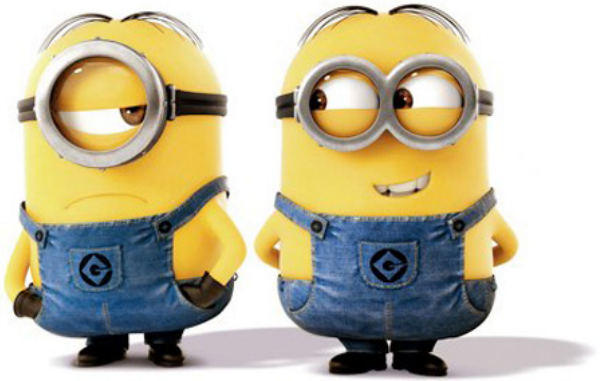
$$\bar{x} = \frac{\sum x_i}{n} = \frac{24+28+17+32+27+25+22+19+90}{9}$$

$$= \frac{284}{9} = 31.6 \text{ kg}$$

ค่าเฉลี่ย คือ 31.6 kg

# คุณสมบัติของค่าเฉลี่ย

1. เป็นตัวแทนข้อมูล ที่ใช้ข้อมูลทุกค่ามาทำการคำนวณหาขนาดของค่าเฉลี่ย
2. เนื่องจากมีการนำข้อมูลทุกค่ามาคำนวณตามหลักคณิตศาสตร์จึงสามารถใช้ในการวิเคราะห์สถิติขั้นสูงได้
3. เนื่องจากมีการใช้ข้อมูลทุกค่ามาคำนวณ ดังนั้นหากมีข้อมูลบางตัวที่มีขนาดใหญ่มากๆหรือเล็กมากๆ ผิดปกติจะมีผลต่อการคำนวณขนาดของค่าเฉลี่ยด้วย
4. ข้อมูลที่มีมาตรวัดเป็นนามบัญญัติ(norminal scale) และเรียงอันดับ (ordinal scale) ไม่สามารถใช้คำนวณค่าเฉลี่ยได้

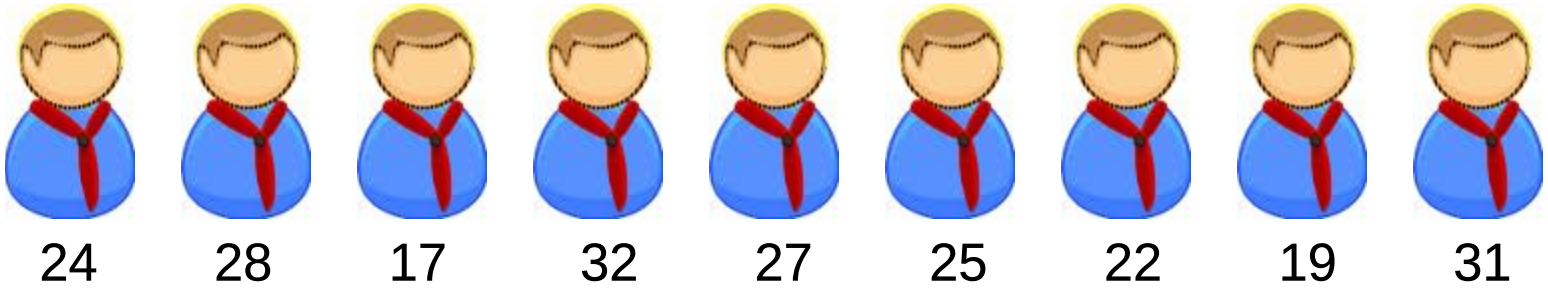


# มัธยฐาน (Median)

- ค่าข้อมูลที่อยู่ตำแหน่งตรงกลางของชุดข้อมูลที่เรียงลำดับ
- จำนวนข้อมูลที่มีค่ามากกว่ามัธยฐาน เท่ากับจำนวนข้อมูลที่มีค่าน้อยกว่ามัธยฐาน
- วิธีการหาค่ามัธยฐาน
  - จัดเรียงลำดับข้อมูลจากน้อยไปหามาก
  - หาข้อมูลที่อยู่ในตำแหน่งตรงกลางจากสูตร $Middle\ rank = \frac{(n+1)}{2}$
  - กรณีจำนวนข้อมูลเป็นเลขคี่ ค่ามัธยฐานคือข้อมูลตำแหน่งกลาง
  - กรณีจำนวนข้อมูลเป็นเลขคู่ ค่ามัธยฐานจะอยู่ระหว่างข้อมูล 2 ตำแหน่งกลาง และเท่ากับค่าเฉลี่ยของข้อมูล 2 ตำแหน่งนั้น

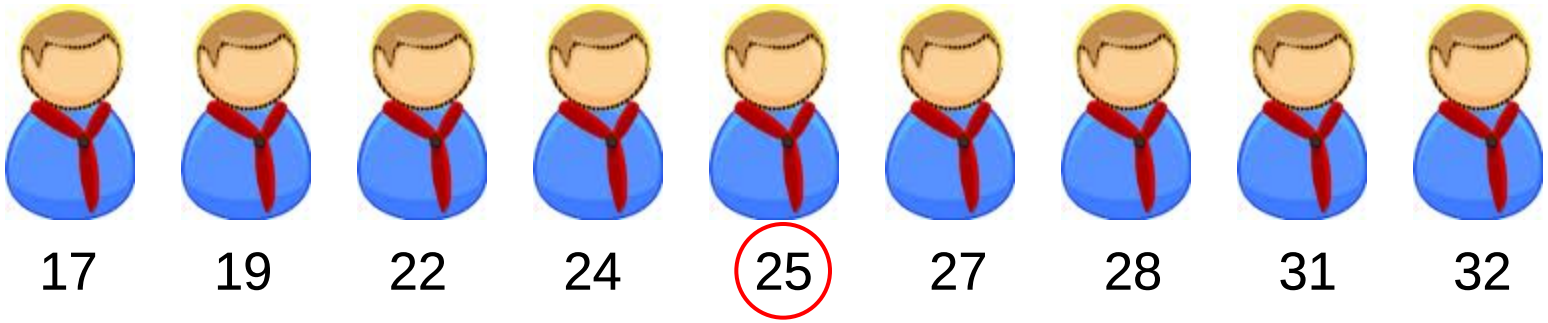
# ตัวอย่าง

## จงคำนวณค่ามัธยฐานน้ำหนักตัวของนักเรียน 9 คน



24 28 17 32 27 25 22 19 31

จัดลำดับข้อมูลจากน้อยไปหามาก



17 19 22 24 25 27 28 31 32

$$Middle\ rank = \frac{(n+1)}{2} = \frac{(9+1)}{2} = 5$$

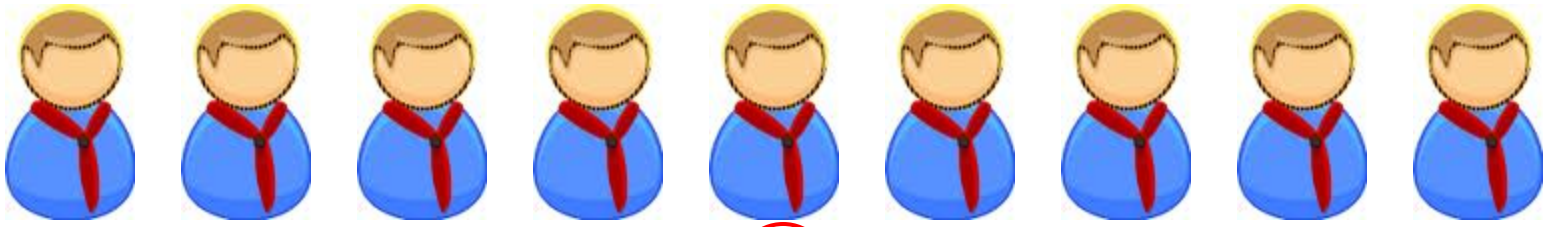
ค่ามัธยฐาน คือ 25 kg

# ตัวอย่าง

## จงคำนวณค่ามัธยฐานน้ำหนักตัวของนักเรียน 9 คน

24 28 17 32 27 25 22 19 90

จัดลำดับข้อมูลจากน้อยไปหามาก



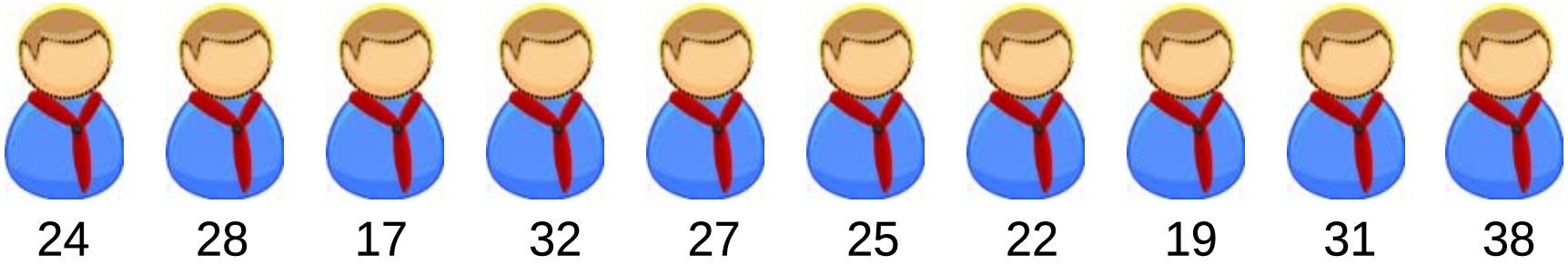
17 19 22 24 25 27 28 32 90

$$Middle\ rank = \frac{(n+1)}{2} = \frac{(9+1)}{2} = 5$$

ค่ามัธยฐาน คือ 25 kg

# ตัวอย่าง

## จงคำนวณค่ามัธยฐานน้ำหนักตัวของนักเรียน 10 คน



24 28 17 32 27 25 22 19 31 38

จัดลำดับข้อมูลจากน้อยไปหามาก

17 19 22 24 25 27 28 31 32 38

$$Middle\ rank = \frac{(n+1)}{2} = \frac{(10+1)}{2} = 5.5$$

ค่ามัธยฐาน คือ $\frac{25+27}{2}$ = 26 kg

# คุณสมบัติของค่ามัธยฐาน

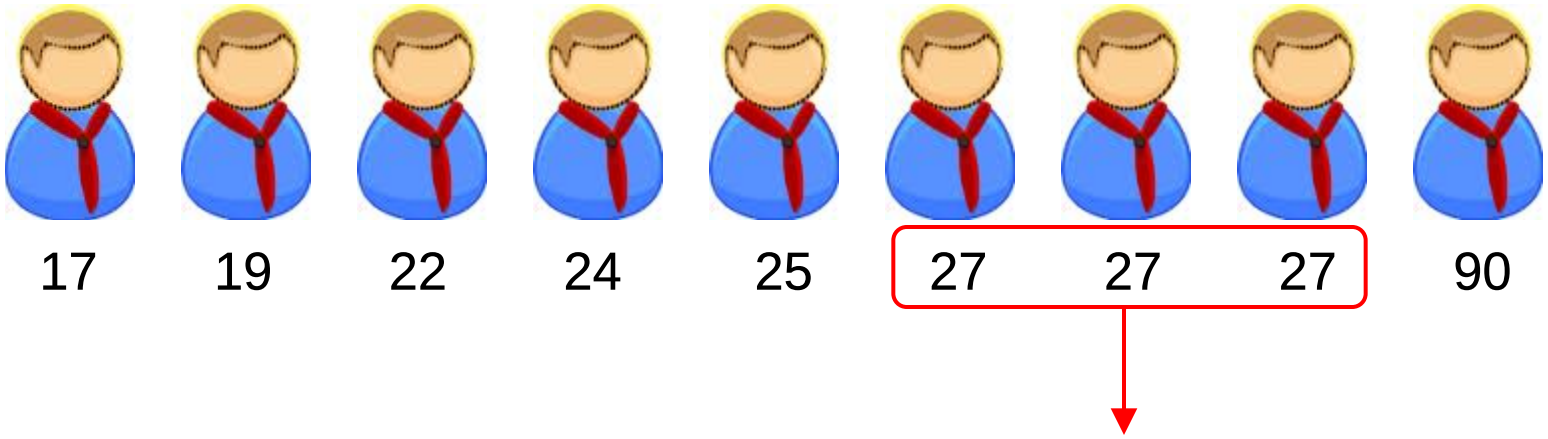
- ข้อมูลที่มีค่าต่างไปจากกลุ่มมากๆ ไม่มีผลต่อค่ามัธยฐาน
- นิยมใช้ในกรณีที่ข้อมูลมีลักษณะเบ้

# ฐานนิยม (Mode)

- ค่าที่มีความถี่สูงที่สุด หรือค่าที่ซ้ำกันมากที่สุด
- มักใช้กับตัวแปรเชิงคุณภาพ

# ตัวอย่าง

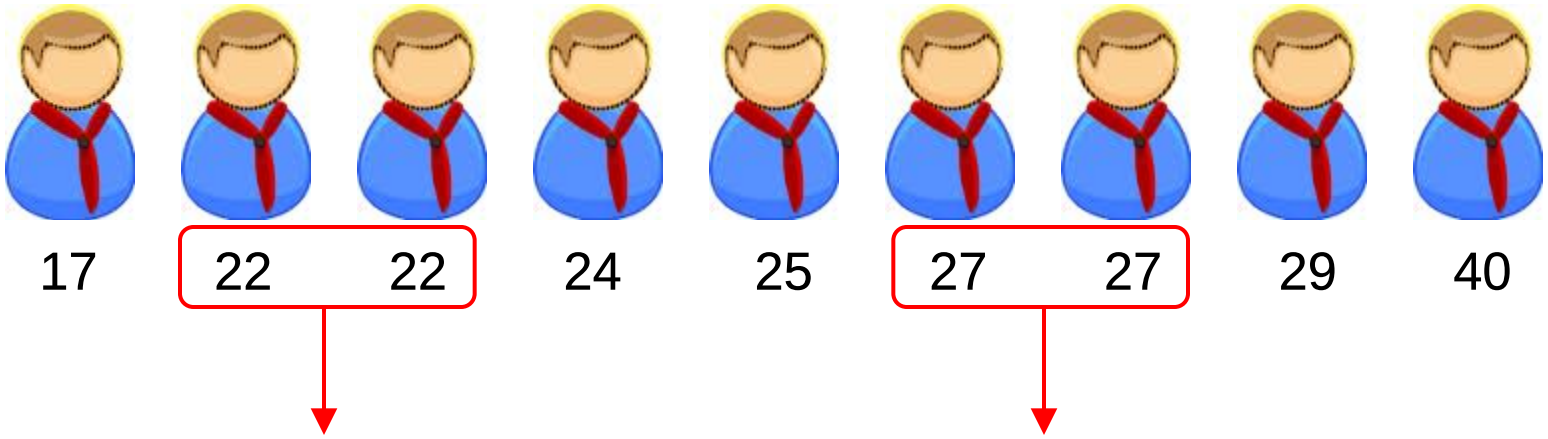
จงคำนวณค่าฐานนิยมของน้ำหนักตัวของนักเรียน 9 คน




ค่าฐานนิยม คือ 27 kg

# ตัวอย่าง

จงคำนวณค่าฐานนิยมของน้ำหนักตัวของนักเรียน 9 คน



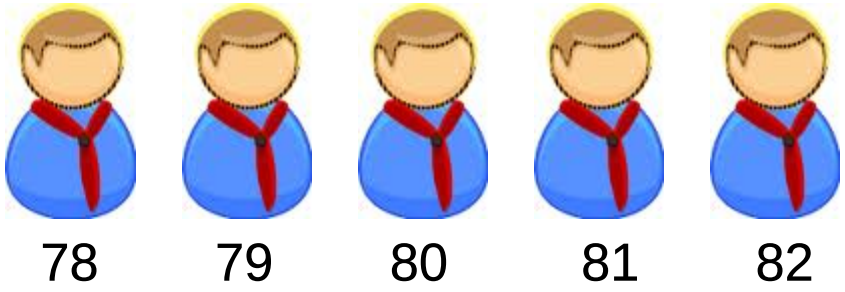

ค่าฐานนิยม คือ 22 และ 27 kg

# การวัดการกระจาย
# Measure of Dispersion

- Variance (ความแปรปรวน)
- Standard Deviation: SD (ค่าเบี่ยงเบนมาตรฐาน)
- Range (พิสัย)
- Percentile (เปอร์เซ็นไทล์)
- Quartile (ควอไทล์)
- Interquartile (อินเตอร์ควอไทล์)

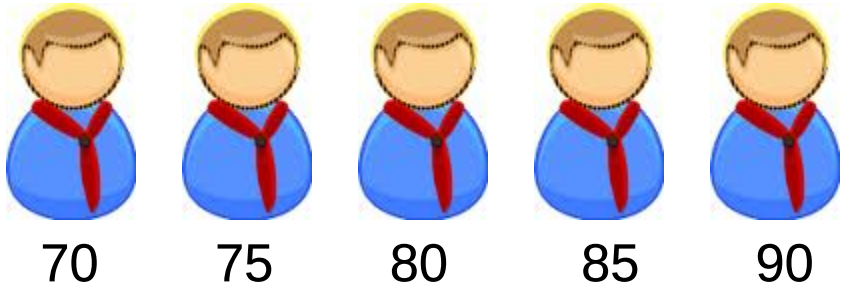
# ตัวอย่าง

จงคำนวณค่าเฉลี่ยคะแนนสอบของนักเรียน 5 คน ในแต่ละห้องเรียน



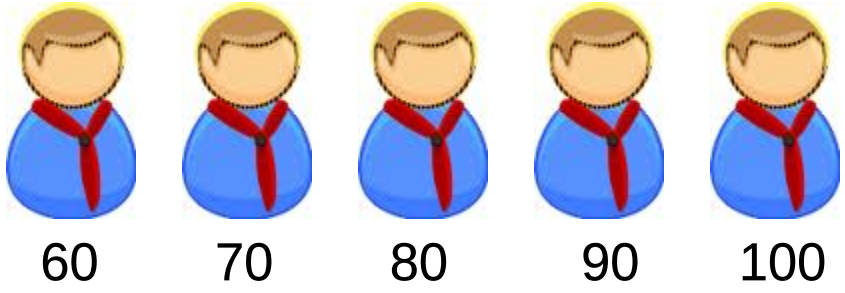
78 79 80 81 82

Mean = 80 คะแนน



70 75 80 85 90

Mean = 80 คะแนน

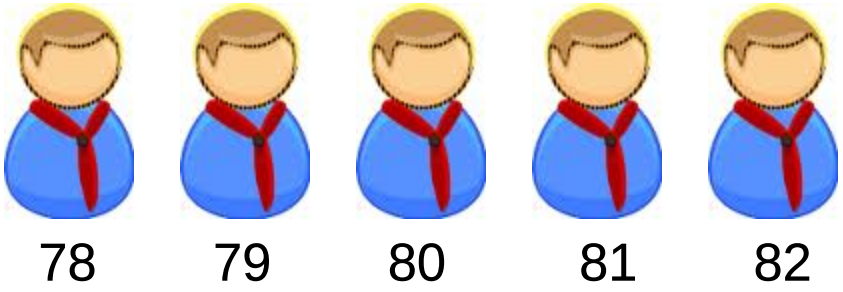


60 70 80 90 100

Mean = 80 คะแนน

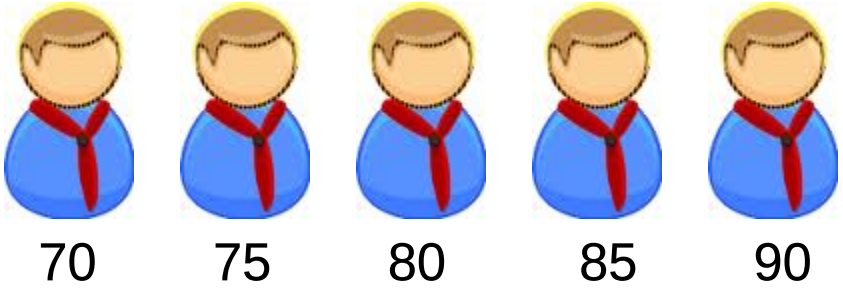
# ตัวอย่าง

## จงคำนวณค่ามัธยฐานคะแนนสอบของนักเรียน 5 คน ในแต่ละห้องเรียน



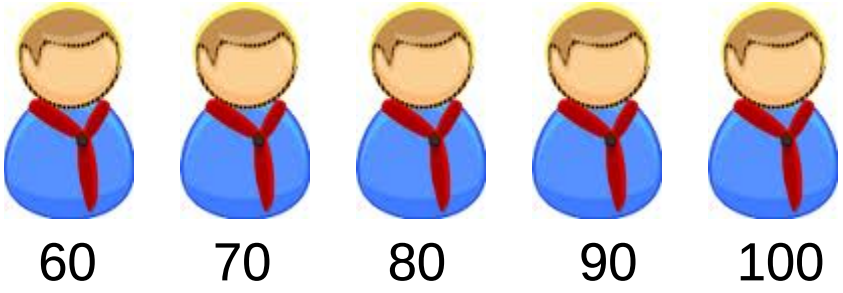
78 79 80 81 82

Median = 80 คะแนน



70 75 80 85 90

Median = 80 คะแนน



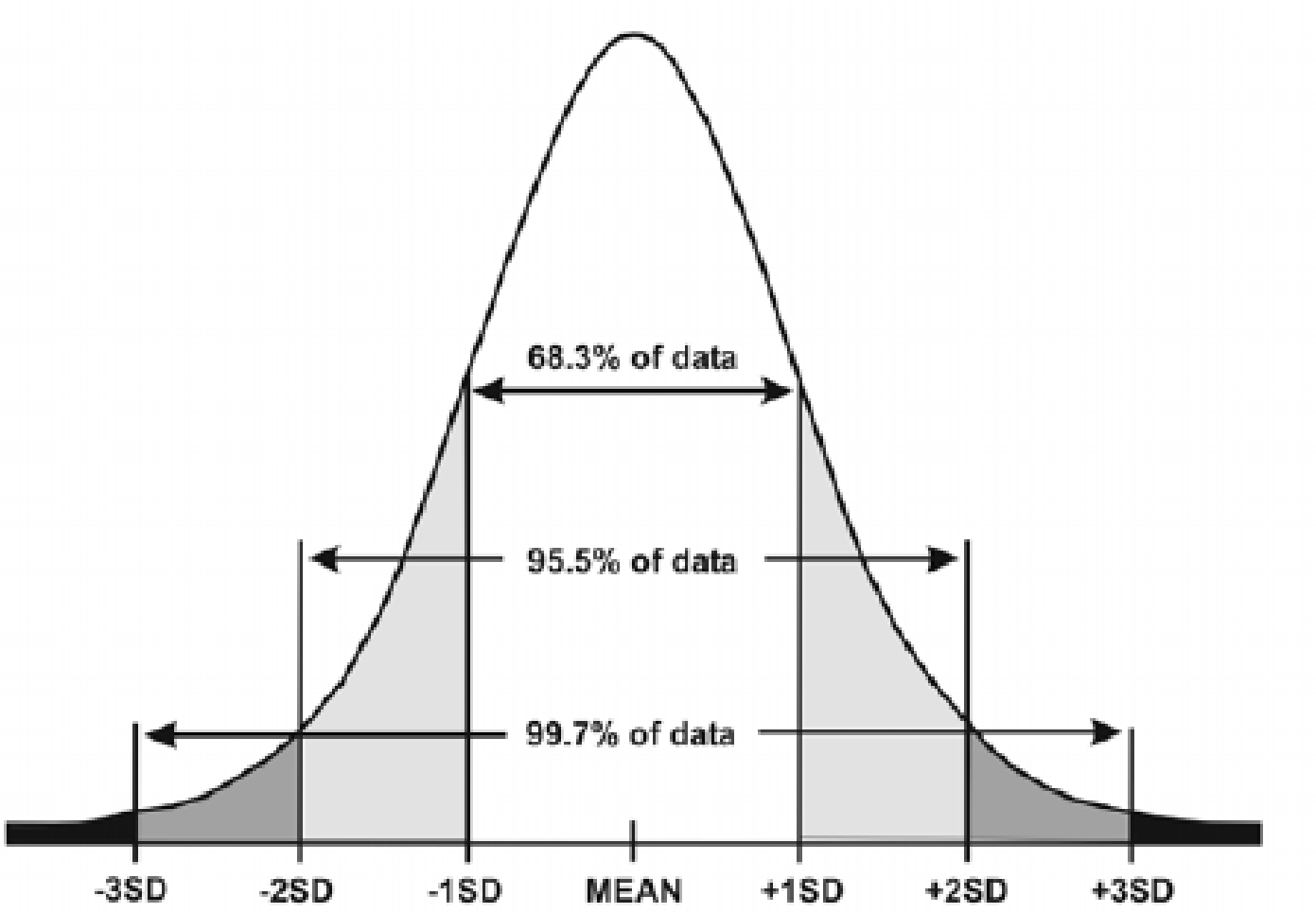
60 70 80 90 100

Median = 80 คะแนน

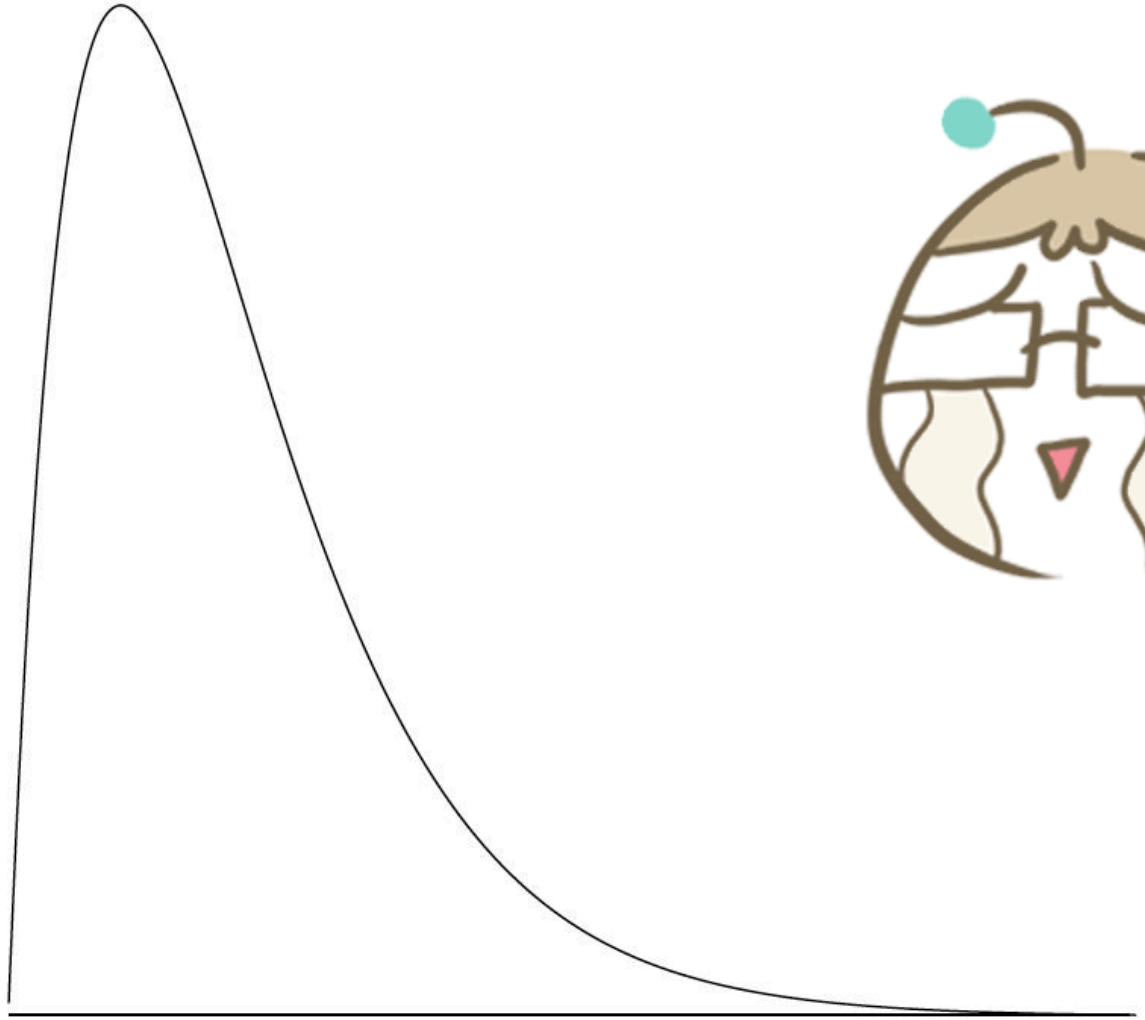
# ความแปรปรวน (Variance) และค่าเบี่ยงเบนมาตรฐาน (SD)

- Variance = $s^2 = \frac{sum\ of\ \ square\ differences}{n-1} = \frac{\sum (x_i - \bar{x})^2}{n-1}$
- SD = s = $\sqrt{Variance} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$
- ใช้ประกอบกับค่าเฉลี่ย

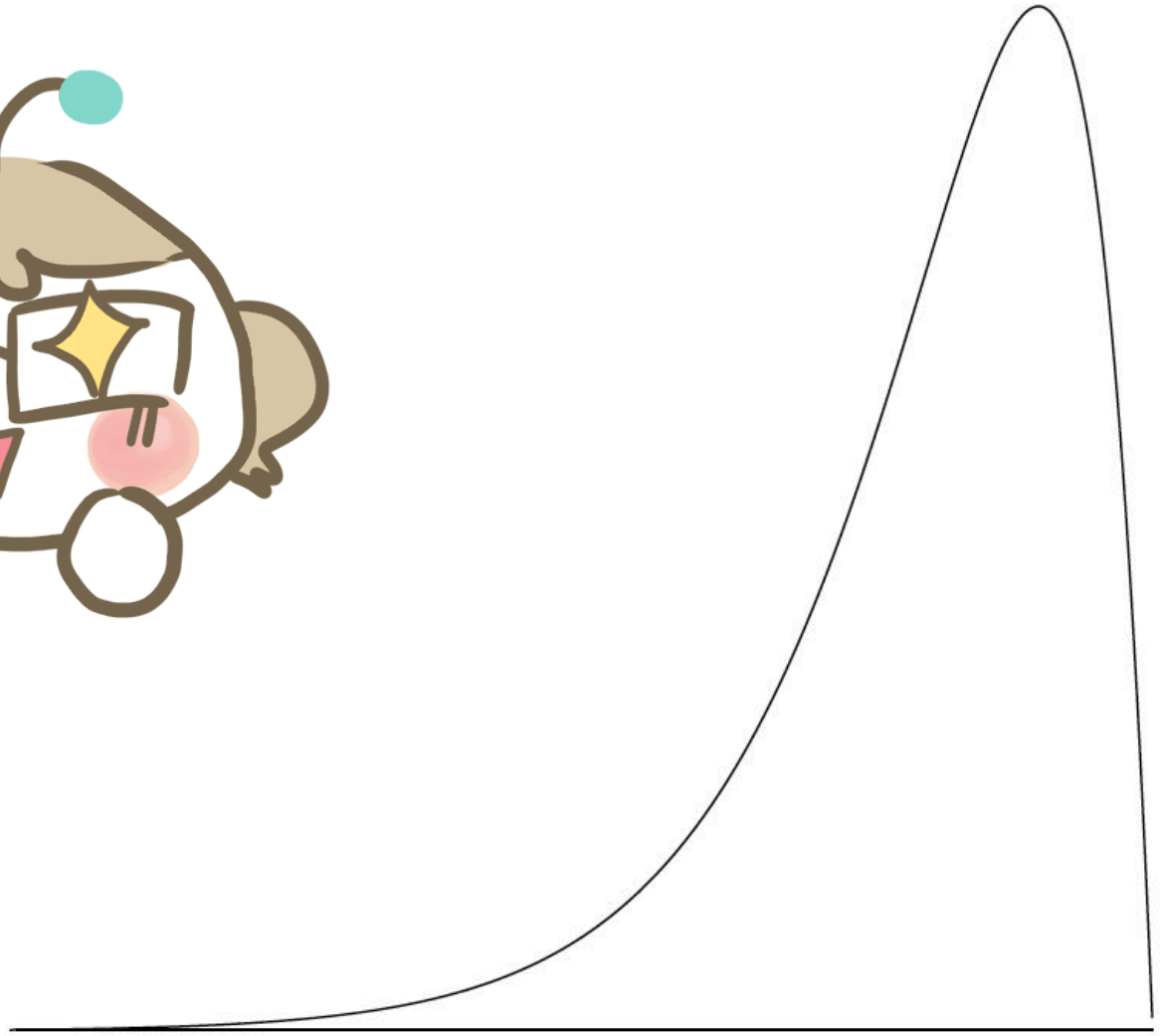
# การวัดแนวโน้มเข้าสู่ส่วนกลางและการวัดการกระจายโดยใช้ Mean และ Standard Deviation



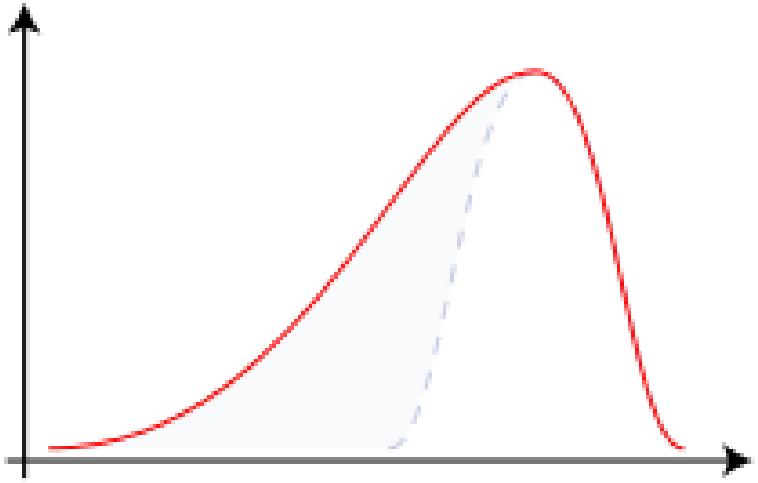
# การกระจายแบบเบ้ขวา
# (Skew to the right หรือ Positive skew)



# การกระจายแบบเบ้ซ้าย
# (Skew to the left หรือ Negative skew)

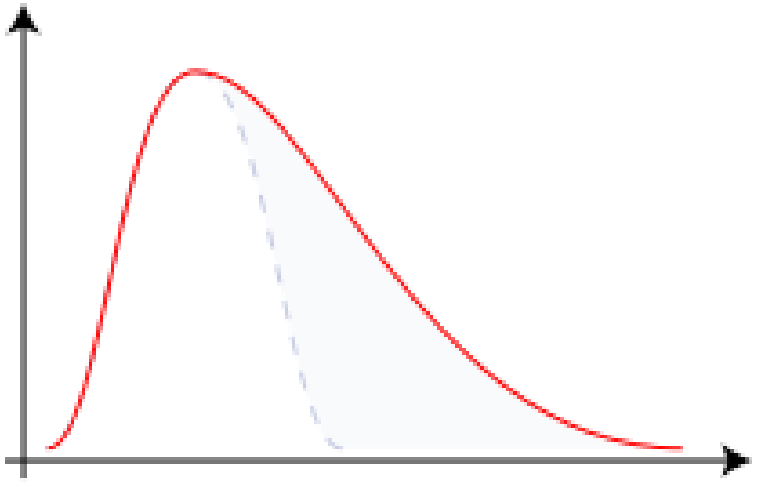


การกระจายแบบเบ้ซ้าย



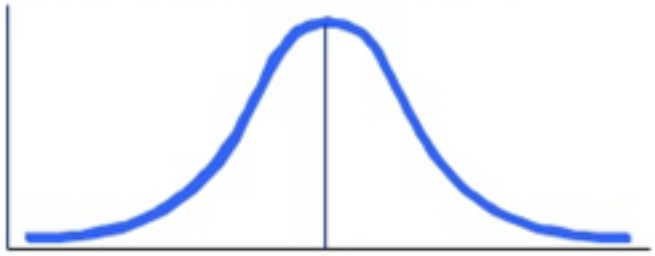
Negative Skew

การกระจายแบบเบ้ขวา



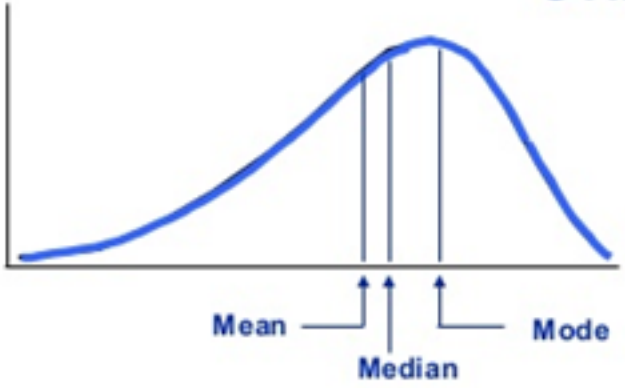
Positive Skew

# Skewness

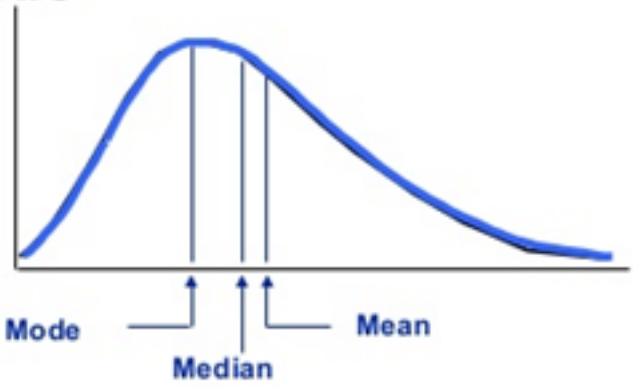


Mode = Mean = Median

**SYMMETRIC**




**SKEWED LEFT**
(negatively)



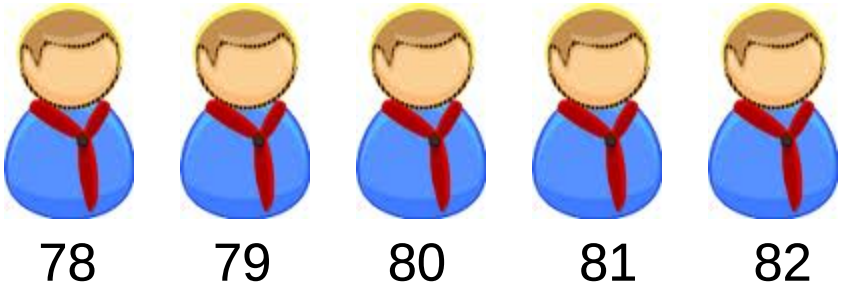

**SKEWED RIGHT**
(positively)

# พิสัย (Range)

- ความแตกต่างระหว่างค่าน้อยที่สุดกับค่ามากที่สุดในชุดข้อมูล
- ค่ามากที่สุด - ค่าน้อยที่สุด
  - ทางสถิติ พิสัย...
  - ทางระบาดวิทยา ตั้งแต่... ถึง...
- ใช้ประกอบกับค่ามัธยฐาน

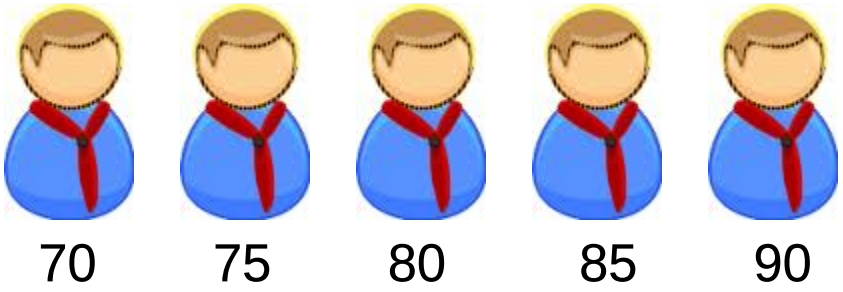
# ตัวอย่าง

## ค่ามัธยฐานและพิสัยของคะแนนสอบของนักเรียน 5 คน ในแต่ละห้องเรียน



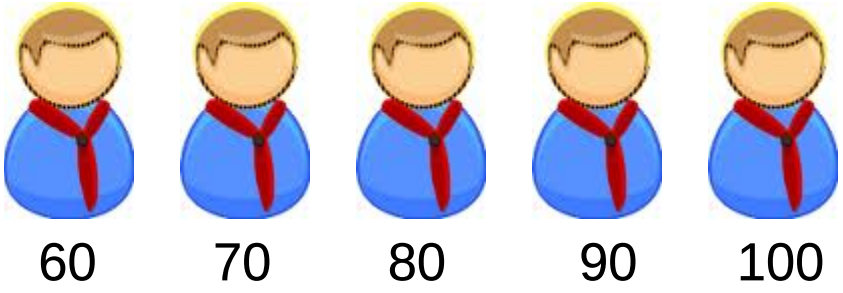
78 79 80 81 82

Median = 80 คะแนน
Range 78-82 คะแนน



70 75 80 85 90

Median = 80 คะแนน
Range 70-90 คะแนน

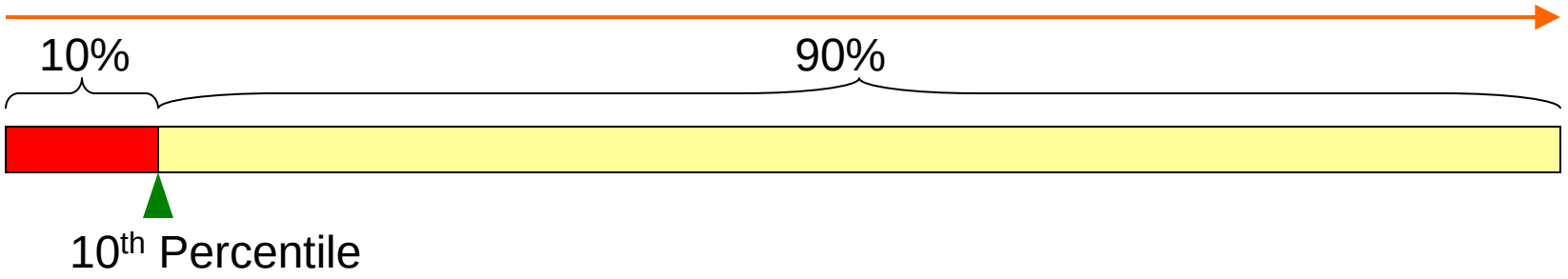


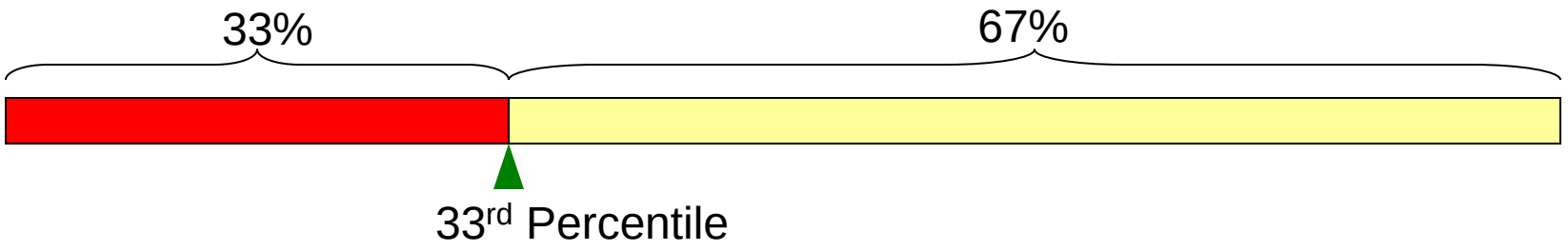
60 70 80 90 100

Median = 80 คะแนน
Range 60-100 คะแนน

# Percentiles

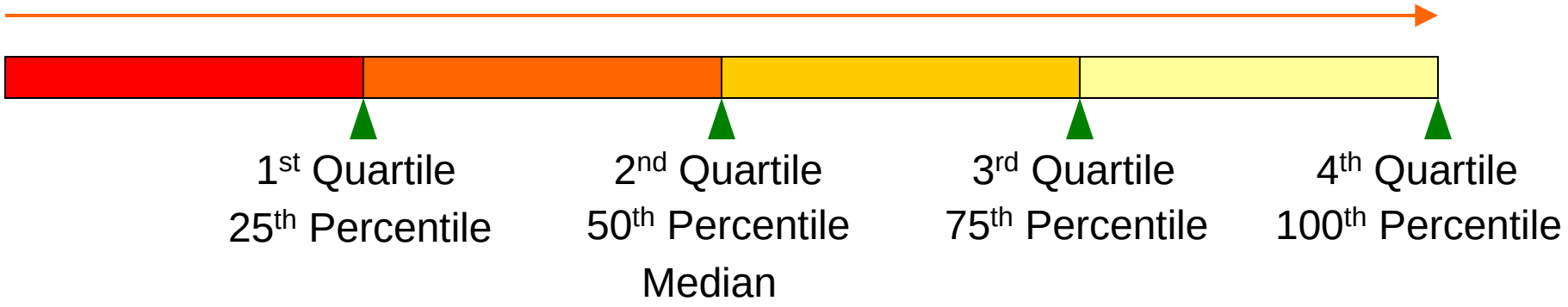
- เกิดจากการแบ่งข้อมูลเป็น 100 ส่วนเท่าๆ กัน เมื่อข้อมูลถูกเรียงจากน้อยไปหามาก
- มีค่าอยู่ระหว่าง 0-100
- Percentiles ที่ P หมายถึง ค่าที่มีจำนวนข้อมูลน้อยกว่าร้อยละ P ของชุดข้อมูลทั้งหมด



# Quartiles

- เกิดจากการแบ่งข้อมูลเป็น 4 ส่วนเท่าๆ กัน เมื่อข้อมูลถูกเรียงจากน้อยไปหามาก
- แบ่งเป็น Quartile ที่ 1, 2, 3 และ 4

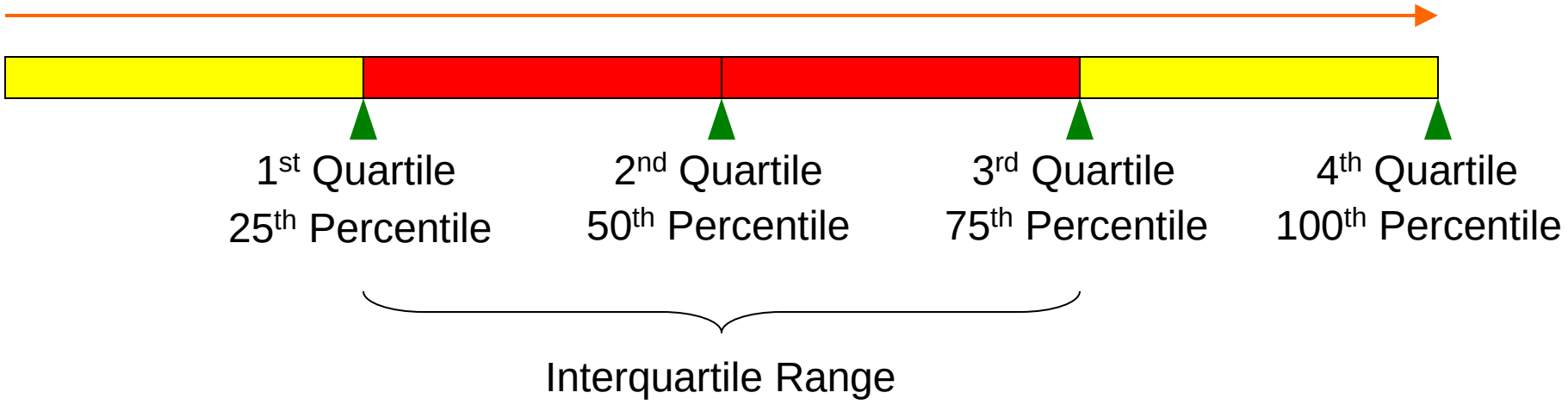



Q1 คือ ค่า ณ ตำแหน่งที่ $\frac{n+1}{4}$

Q3 คือ ค่า ณ ตำแหน่งที่ $\frac{3(n+1)}{4}$

# Interquartile Range

- เป็นการวัดการกระจายที่นิยมใช้ประกอบกับค่ามัธยฐาน
- เป็นตัวแทนของข้อมูลครึ่งหนึ่งที่อยู่ในช่วงกลางของชุดข้อมูล ระหว่าง P25-P75



# ตัวอย่าง

- Age (in years) of a group of 10 villagers :

5 11 21 24 27 28 30 42 50 52

- IQR = Q3 – Q1 = 42 – 21 = 21

- Interquartile Deviation (ส่วนเบี่ยงเบนควอไทล์)

  = IQR/2 = (Q3 – Q1) / 2

  = 21/2 = 10.5

# สรุปแนวทางการเลือกใช้วิธีการวัดแนวโน้มเข้าสู่ส่วนกลางและวิธีการวัดการกระจาย

| รูปแบบการกระจาย | การวัดแนวโน้มเข้าสู่ส่วนกลาง | การวัดการกระจาย |
|---|---|---|
| การกระจายแบบปกติ (Normal Distribution) | ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) | ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) |
| การกระจายแบบเบ้ (Skew Distribution) | ค่ามัธยฐาน (Median) | พิสัย (Range) หรือ Interquartile Range |

- ขอขอบคุณ – สไลด์ พญ.ดารินทร์ อารีย์โชคชัย นพ.ปณิธี ธัมมวิจยะ และนพ.ยงเจือ เหล่าศิริถาวร
- สำนักระบาดวิทยา กรมควบคุมโรค