

# การเลือก ใช้สถิติ ในงานวิจัย

รศ.ดร.นพ. วิฑูรย์ โล่ห์สุนทร

ภาควิชาเวชศาสตร์ป้องกันและสังคม  
คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

# การเลือกใช้สถิติในงานวิจัย

## ★ สถิติเชิงพรรณนา (Descriptive statistics)

Summarization and presentation of data

- Table
- Graphs
- Numerical summaries

## ★ สถิติเชิงอนุมาน (Inferential statistics)

Generalization of conclusions: sample --> population

- Hypothesis testing
- Estimation and confidence intervals
- Modeling

# Types of Data

```
graph TD; A[Types of Data] --> B[ข้อมูลเชิงคุณภาพ (Qualitative Data)]; A --> C[ข้อมูลเชิงปริมาณ (Quantitative Data)]; B --- B_text[บอกคุณลักษณะ เช่น เพศ หมู่เลือด เชื้อชาติ อาชีพ]; C --- C_text[บอกปริมาณ เช่น อายุ น้ำหนัก ส่วนสูง เส้นรอบเอว];
```

ข้อมูลเชิงคุณภาพ  
(Qualitative Data)

บอกคุณลักษณะ  
เช่น เพศ หมู่เลือด  
เชื้อชาติ อาชีพ

ข้อมูลเชิงปริมาณ  
(Quantitative Data)

บอกปริมาณ  
เช่น อายุ น้ำหนัก  
ส่วนสูง เส้นรอบเอว

วิเคราะห์ข้อมูลและนำเสนอข้อมูลที่เหมาะสมกับชนิดของข้อมูล

# Types of Data

ข้อมูลเชิงคุณภาพ  
(Qualitative Data)

ข้อมูลเชิงปริมาณ  
(Quantitative Data)

เพศ

อาชีพ

อายุ (ปี)

สถานภาพสมรส

ดัชนีมวลกาย ( $\text{kg}/\text{m}^2$ )

ระดับไขมันในเลือด ( $\text{mg}/\text{dl}$ )

# Categorical Data (Univariate Analysis)

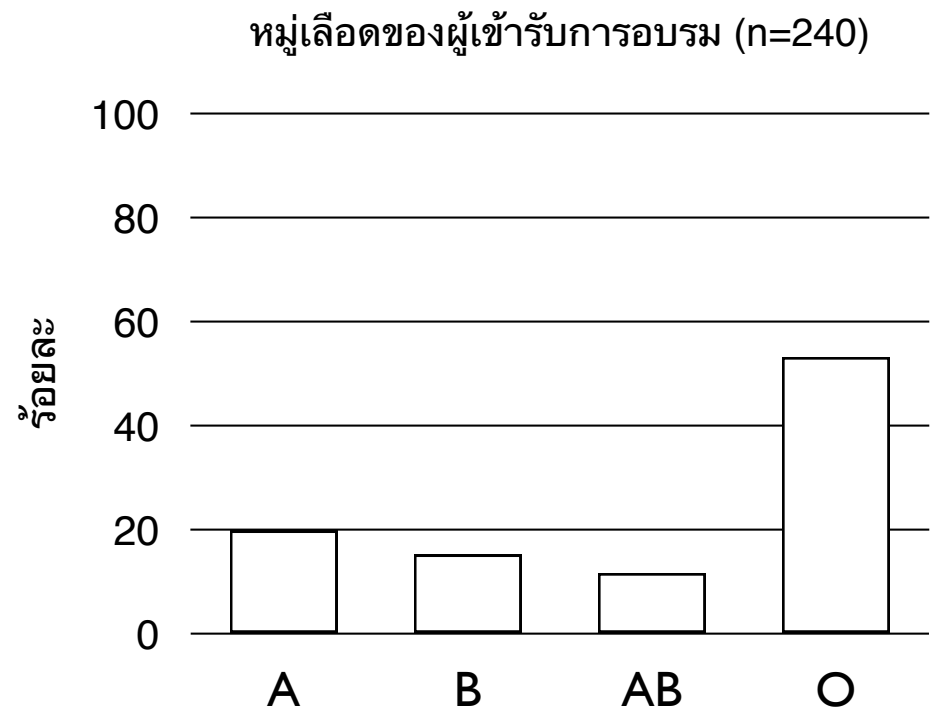
พิจารณาข้อมูลหมู่เลือดของผู้เข้ารับการอบรมจำนวน 240 คน

O, O, O, A, O, B, AB, A, O, O, A, O, O, O, A, AB, A, A, O, B, O, AB, O, AB  
A, O, O, AB, O, B, A, O, O, B, B, A, A, O, O, O, AB, B, O, O, A, A,.....O

For categorical data we usually summarize with counts.

A simple visual summary is the bar graph.

หมู่เลือด	จำนวน	ร้อยละ
A	48	20
B	36	15
AB	28	11.7
O	128	53.3
รวม	240	100



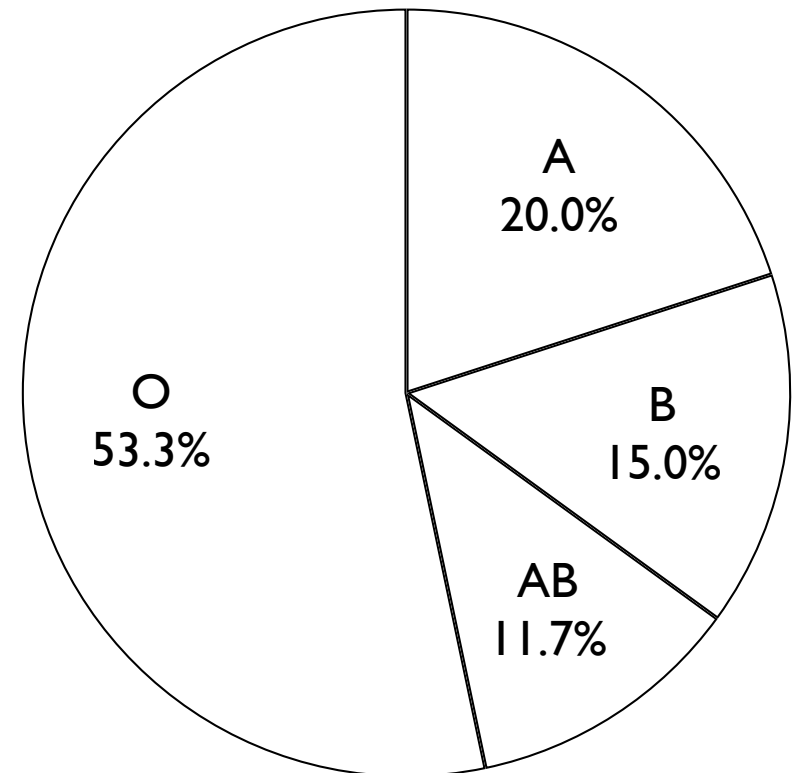
# Categorical Data (Univariate Analysis)

พิจารณาข้อมูลหมู่เลือดของผู้เข้ารับการอบรมจำนวน 240 คน

O, O, O, A, O, B, AB, A, O, O, A, O, O, O, A, AB, A, A, O, B, O, AB, O, AB  
A, O, O, AB, O, B, A, O, O, B, B, A, A, O, O, O, AB, B, O, O, A, A,.....O

หมู่เลือดของผู้เข้ารับการอบรม (n=240)

หมู่เลือด	จำนวน	ร้อยละ
A	48	20
B	36	15
AB	28	11.7
O	128	53.3
รวม	240	100



# Categorical Data (STATA Output)

-> tabulation of race

Race	Freq.	Percent	Cum.
White	96	50.79	50.79
Black	26	13.76	64.55
Other	67	35.45	100.00
Total	189	100.00	

-> tabulation of smoke

Smoked during pregnancy	Freq.	Percent	Cum.
No	115	60.85	60.85
Yes	74	39.15	100.00
Total	189	100.00	

# Categorical Data (SPSS Output)

(เลือกเมนู: Analyze --> Descriptive Statistics --> Frequencies...)

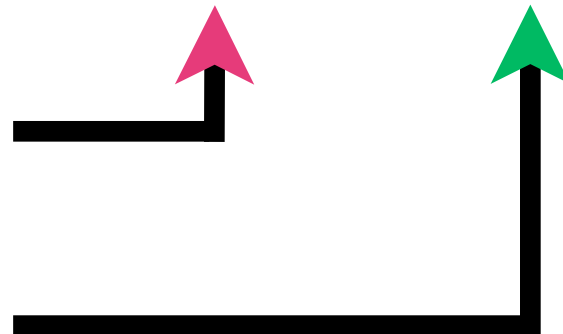
## ระดับการศึกษาสูงสุด

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	น้อยกว่าปริญญาตรี	153	21.9	22.0	22.0
	ปริญญาตรี	233	33.4	33.6	55.6
	ปริญญาโท	168	24.1	24.2	79.8
	ปริญญาเอก	140	20.1	20.2	100.0
	Total	694	99.4	100.0	
Missing	System	4	.6		
Total		698	100.0		

ช่อง Percent จะคิดคำนวณ % จากจำนวนทั้งหมด โดยรวม Missing เข้าไปด้วย



ช่อง Valid Percent จะคิดคำนวณ % จากจำนวนข้อมูลที่มี โดยไม่นับรวม Missing





# Quantitative Data (Univariate Analysis)

พิจารณาข้อมูลน้ำหนักของผู้เข้ารับการอบรมจำนวน 240 คน

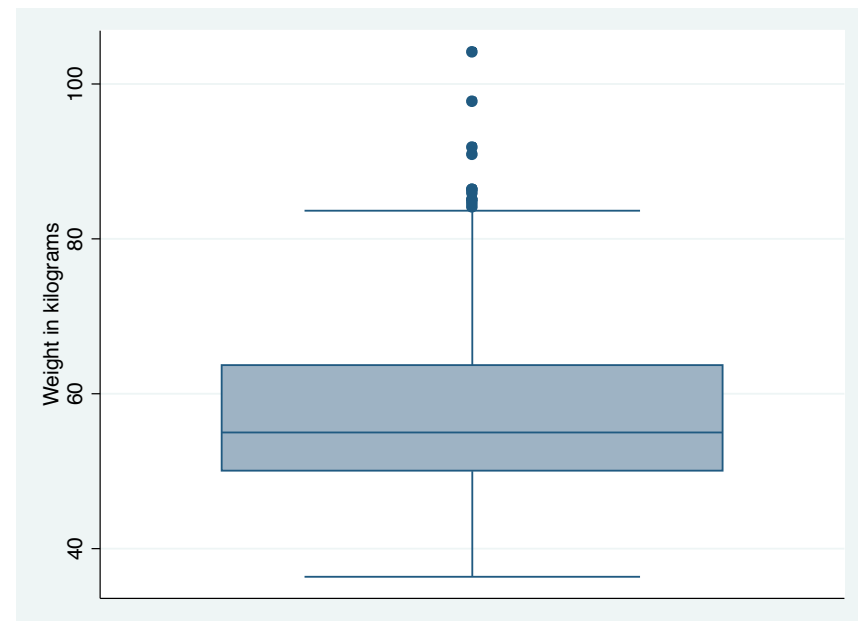
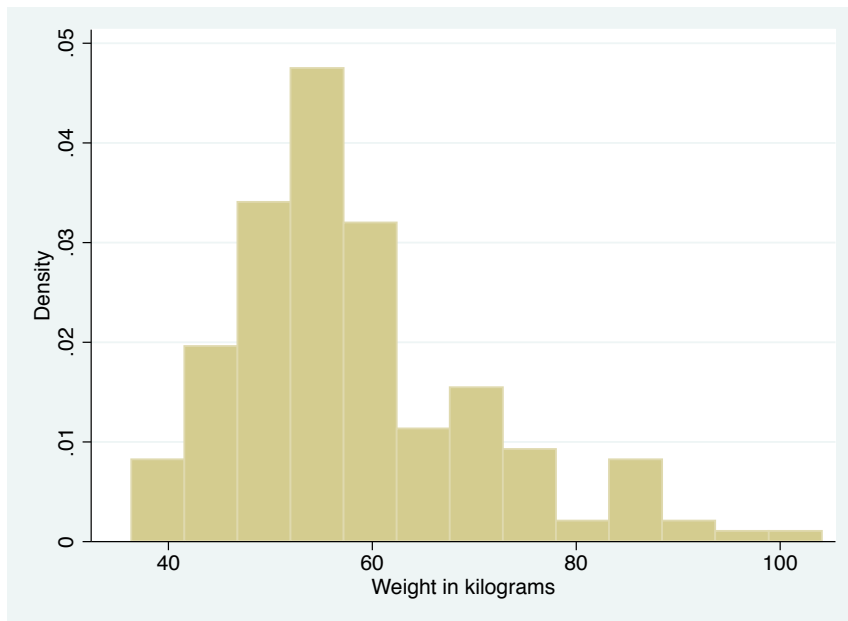
49.1, 60.0, 52.8, 49.7, 90.0, 80.5, 44.0, 67.8, 72.2, 49.0, 103.0, 55.0, 90.3  
77.1, 45.7, 66.0, 72.0, 72.8, 80.5, 90.2, 55.1, 44.9, 50.2, 44.8,.....70.0

Weight is a quantitative variable.

Here we are more interested in characteristics of the distribution of weight:

- 🎯 where is the center of the weight distribution (e.g. the average)?
- 🎯 how much does weight vary?
- 🎯 are there some values far from the bulk of the data?

We would like some visual tools to help us answer these questions.



# Quantitative Data (Univariate Analysis)

## ค่ากลางที่เหมาะสม Mean VS. Median

พิจารณาน้ำหนักตัวของผู้ป่วยจำนวน 10 คนต่อไปนี้

50 52 54 54 57 57 58 59 59 60

Mean (SD) = 56.0 (3.33)

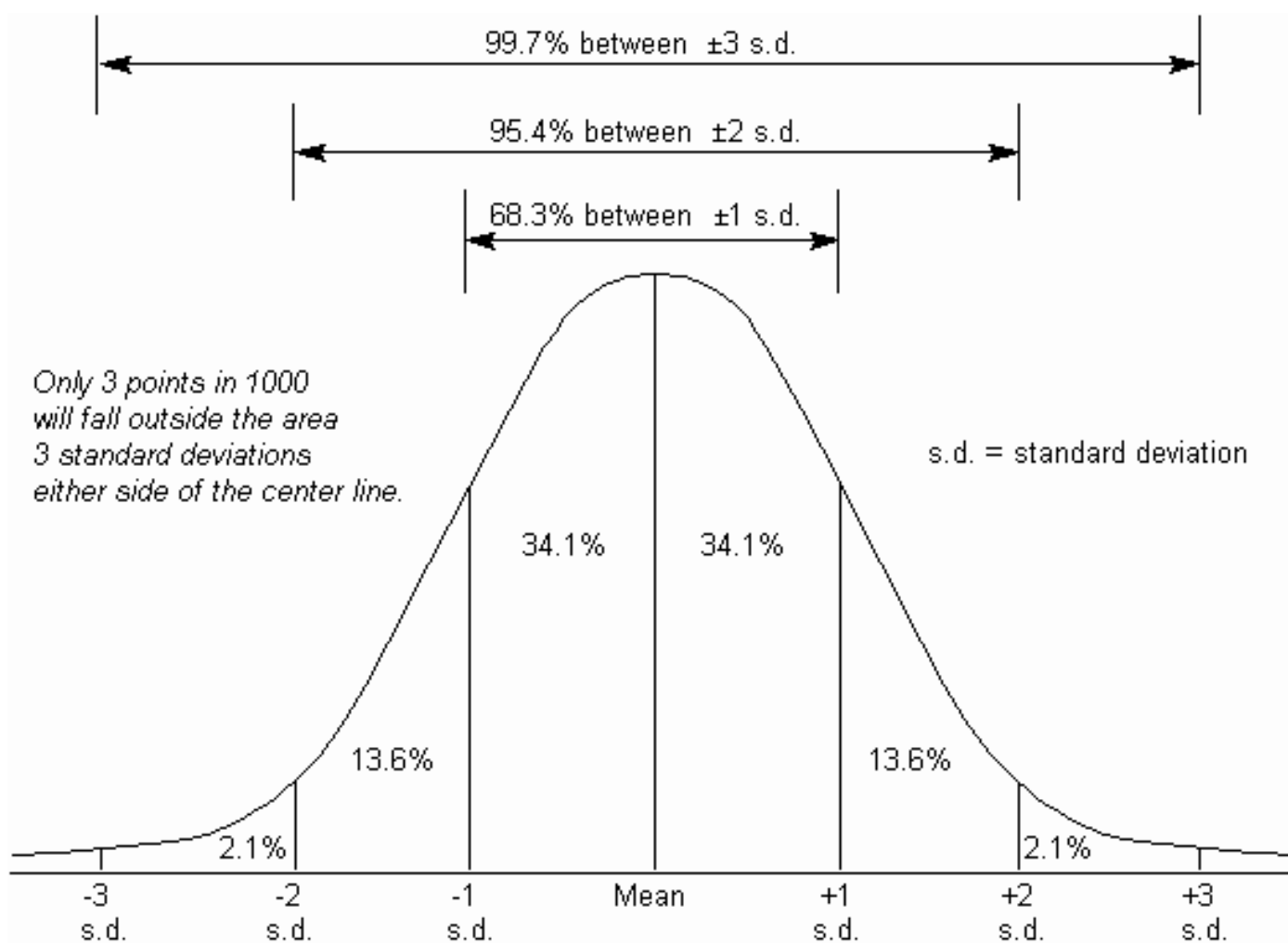
Median (IQR) = 57.0 (5.5)

พิจารณาน้ำหนักตัวของผู้ป่วยจำนวน 10 คนต่อไปนี้

50 52 54 54 57 57 58 59 59 130

Mean (SD) = 63.0 (23.73)

Median (IQR) = 57.0 (5.5)



Source: [http://www.syque.com/quality\\_tools/toolbook/Variation/measuring\\_spread.htm](http://www.syque.com/quality_tools/toolbook/Variation/measuring_spread.htm)

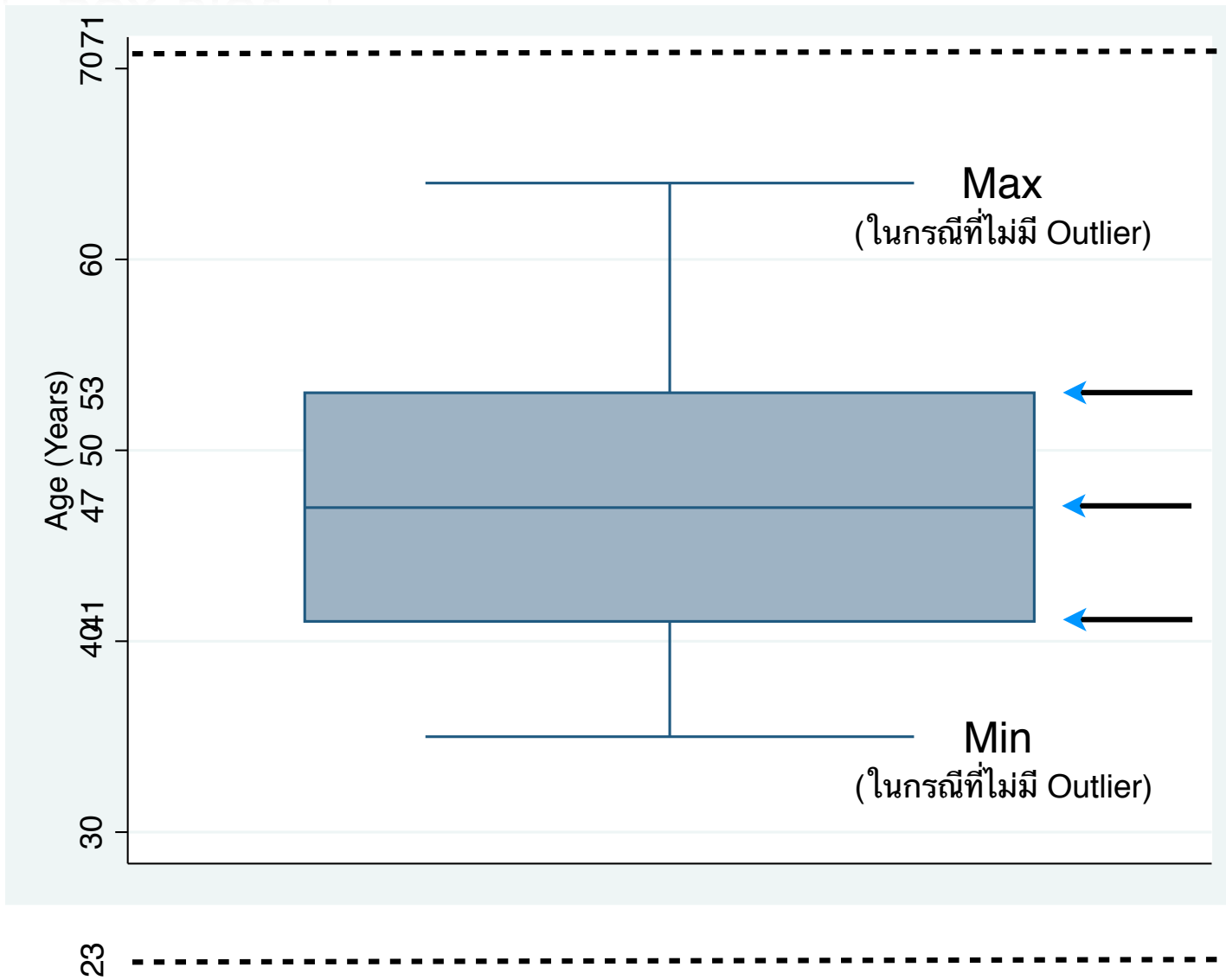
## ถ้าอายุของผู้เข้าร่วมการศึกษามีการกระจายแบบปกติ โดยมีค่า Mean (SD) = 45 (5.0)

ประมาณ 68.3% ของผู้เข้าร่วมการศึกษามีอายุอยู่ระหว่าง  $45 \pm 5$  --> 40 ถึง 50 ปี

ประมาณ 95.4% ของผู้เข้าร่วมการศึกษามีอายุอยู่ระหว่าง  $45 \pm 10$  --> 35 ถึง 55 ปี

ประมาณ 99.7% ของผู้เข้าร่วมการศึกษามีอายุอยู่ระหว่าง  $45 \pm 15$  --> 30 ถึง 60 ปี

# Box plot



ข้อมูลที่มีค่ามากกว่า  
 $Q3 + 1.5(IQR)$   
 $53 + 1.5 \times 12 = 71$   
ถือว่าเป็น Outlier

Q3

Median

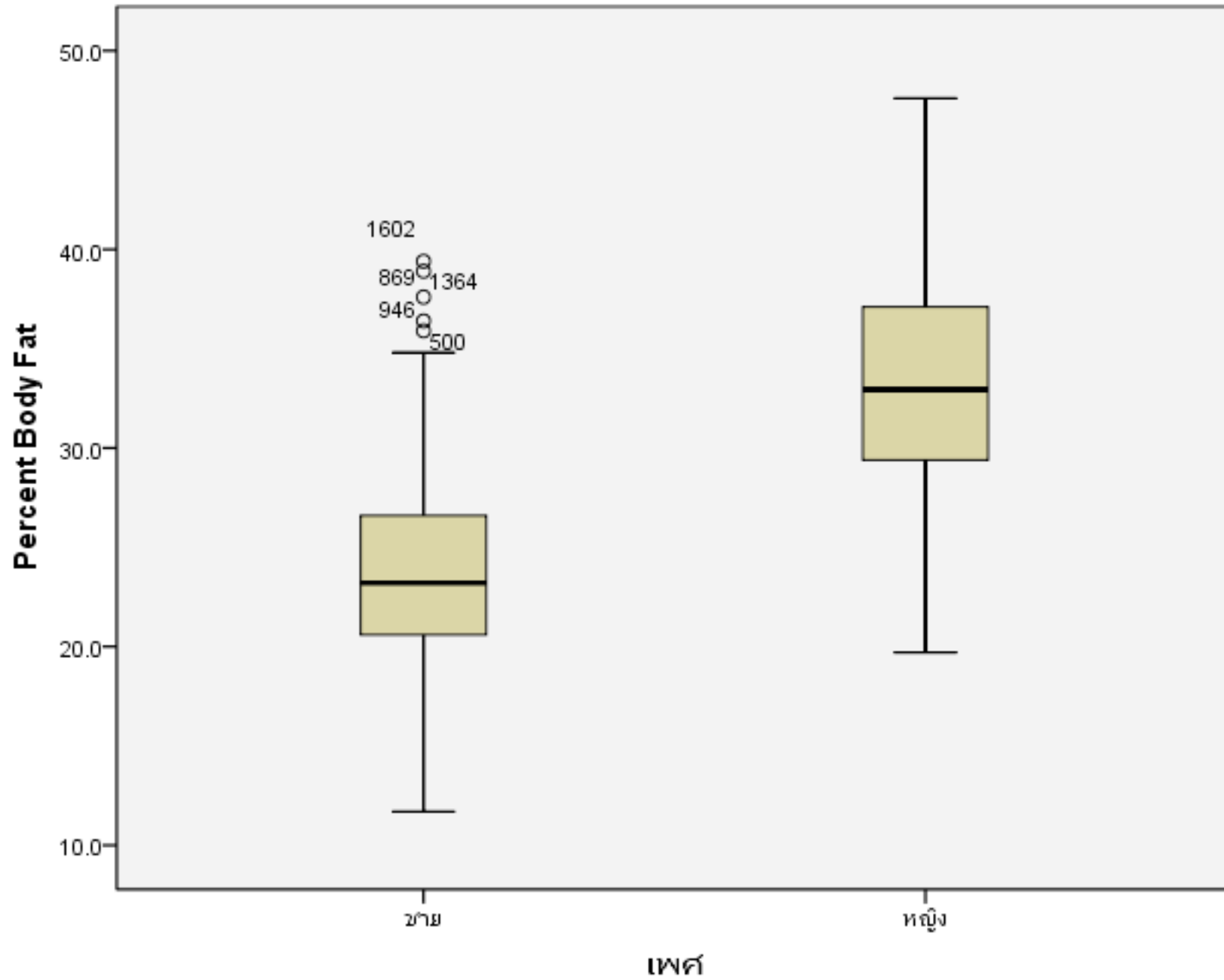
Q1

$$\begin{aligned} IQR &= Q3 - Q1 \\ &= 53 - 41 \\ &= 12 \end{aligned}$$

ข้อมูลที่มีค่าน้อยกว่า  
 $Q1 - 1.5(IQR)$   
 $41 - 1.5 \times 12 = 23$   
ถือว่าเป็น Outlier

Median (IQR) = 47 (12) หรือ 47 (41, 53)

# Box plot



# Types of Data

ข้อมูลเชิงคุณภาพ  
(Qualitative Data)

บอกคุณลักษณะ  
เช่น เพศ กลุ่มเลือด  
เชื้อชาติ อาชีพ

วิเคราะห์ข้อมูลและนำเสนอข้อมูลที่เหมาะสมกับชนิดของข้อมูล

จำนวน (ร้อยละ)  
ของแต่ละกลุ่ม/คุณลักษณะ

ข้อมูลเชิงปริมาณ  
(Quantitative Data)

บอกปริมาณ  
เช่น อายุ น้ำหนัก  
ส่วนสูง เส้นรอบเอว

Mean (SD) หรือ  
Median (IQR)

# การนำข้อมูลเชิงปริมาณมาจัดกลุ่ม

พิจารณาข้อมูลน้ำตาลในเลือด (FBS) ของผู้มารับการตรวจสุขภาพจำนวน 1,000 คน

70, 80, 89, 99, 90, 80, 84, 77, 105, 109, 103, 95, 90, 200, 140, 90, 110,  
77, 105, 96, 92, 102, 80, 90, 105, 144, 90, 83,.....126.0

Mean (SD) = 96 (23.3)

Median (IQR) = 95 (87, 105)

---

ระดับน้ำตาลในเลือด (mg/dl)	จำนวน	(ร้อยละ)
<100	831	(83.1)
100-125	100	(10.0)
$\geq 126$	69	(6.9)

---

การพิจารณาจัดกลุ่มต้องคำนึงถึงเกณฑ์มาตรฐาน ความรู้ทางคลินิก/เชิงวิชาการ  
เพื่อประโยชน์ในการรายงานผลหรือเปรียบเทียบระหว่างกลุ่ม

# จงพิจารณาจัดกลุ่มข้อมูลด้านล่าง

Number of physician visits during the first trimester

Number	Freq.	Percent	Cum.
0	100	52.91	52.91
1	47	24.87	77.78
2	30	15.87	93.65
3	7	3.70	97.35
4	4	2.12	99.47
6	1	0.53	100.00
Total	189	100.00	

จากตารางจะเห็นได้ว่าบางกลุ่มมีจำนวนค่อนข้างน้อยจึงควรพิจารณาขบรวมกลุ่ม โดยอาศัยความรู้ทางคลินิก/เชิงวิชาการ มาประกอบการพิจารณาจัดกลุ่มใหม่



## Quiz 1

ถ้าผู้วิจัยต้องการทดสอบว่าน้ำหนักทารกแรกเกิดระหว่างกลุ่มที่มารดาสูบบุหรี่ และกลุ่มที่มารดาไม่สูบบุหรี่ขณะตั้งครรภ์มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติหรือไม่ ผู้วิจัยควรเลือกใช้การทดสอบทางสถิติในข้อใด

- A. Paired t-test
- B. Unpaired t-test
- C. One-way ANOVA
- D. Pearson's correlation
- E. Pearson's chi-square test

คำอธิบาย:

ตัวแปรต้นคือ

ตัวแปรตามคือ

## Quiz 2

A randomized trial of thiazide (anti-hypertensive drug) on bone density was performed in males and females at risk for osteoporosis. Hip bone density measured at baseline (pre-randomization) and at 6 months. What is the appropriate statistical test to determine whether there are significant differences in bone density across treatment group?

Data to be used in this analysis

Variable	<u>Variable</u> label	Time measured	Codes/description
<u>trt</u>	Treatment group	Baseline	1=Placebo 2=Low dose thiazide 3=High dose thiazide
<u>gender</u>	Gender	Baseline	0=Male 1=Female
<u>age</u>	<u>Age</u> in years	Baseline	Age at time of randomization
<u>basehip</u>	Hip bone density baseline	Baseline	Hip bone density at baseline
<u>fuhip</u>	Hip bone density 6 months	6 months	Hip bone density at 6 months

- A. Paired t-test
- B. Unpaired t-test
- C. One-way ANOVA
- D. Pearson's correlation
- E. Pearson's chi-square test

## Quiz 3

It is commonly believed that blood pressure readings taken by a physician tend to be different from those taken by a nurse. This is referred to as the 'white coat' effect. The data below are two readings (one by a physician and the other by a nurse) taken during the physical exam from 12 male participants in the Cardiovascular Health Study. What test statistic will you use to test the hypotheses?

- A. Paired t-test
- B. Unpaired t-test
- C. One-way ANOVA
- D. Pearson's correlation
- E. Pearson's chi-square test

Subject #	RN BP	MD BP
1	120	125
2	124	126
3	130	138
4	118	117
5	140	143
6	128	128
7	140	146
8	135	133
9	126	127
10	130	135
11	126	126
12	127	131

## Quiz 4

Patients with (cases) and without (controls) oral cancer were surveyed regarding their smoking.

	<b>Case</b>	<b>Control</b>
Smoker	160	80
Non-smoker	40	120
Total	200	200

What is the appropriate statistical test to determine whether oral cancer is associated with smoking?

- A. Paired t-test
- B. Unpaired t-test
- C. One-way ANOVA
- D. Pearson's correlation
- E. Pearson's chi-square test