Summarization of data
(Measures of central location)

Professor Narjis A-H Ajeel
Dept. of Community Medicine, College of Medicine,
University of Basrah

Learning objectives

At the end of the lecture you should be able to:

1. Define the three measures of central location
2. Calculate measures of central location for both grouped and ungrouped data
3. Choose the most appropriate measure of central location for each data set.

Measures of Central Location

- Measures of central location are numbers that tend to cluster around the "middle" of a set of values.
- They are also known as **"measures of central tendency"**

There are 3 measures of central location

1. The **mean**
2. The **median**
3. The **mode**

1-The mean (arithmetic mean)

The mean is the average of all the data values (observations) in a distribution.

$$\text{Mean} = \frac{\text{Sum of the observations}}{\text{Number of observations}}$$

- **Sample mean is denoted as $\bar{x}$**
- **Population mean is denoted as μ (mu)**

$$\bar{x} = \frac{\sum x}{n}$$

Example (1)

The reported time on the Internet of 10 students are 0, 7, 12, 5, 33, 14, 8, 0, 9, 22 hours/week. Find the mean time on the Internet.

$$\bar{x} = \frac{\sum x}{n} = \frac{\overset{\boxed{0}}{x_1} + \overset{\boxed{7}}{x_2} + \ldots + \overset{\boxed{22}}{x_{10}}}{10} =$$

$$\bar{x} = \frac{\sum x}{n} = \frac{0+7+...+22}{10} = 11$$

The Mean for Grouped Data

- The mean of a sample of data organized in a frequency distribution is computed by the following formula:

$$\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{\Sigma fX}{n}$$

- Where:
- $\sum fx$ is the sum of the product of X times the frequency
- Σf is the sum of the frequencies = n (the sample size)

Example 2: Calculate the mean number of previous pregnancies

Number of Previous pregnancies (x)	Frequency (f)	fX
0	18	18x0=0
1	35	35x1=35
2	24	24x2=48
3	18	18x3=54
4	6	6x4=24
Total	∑f=101	∑ fX=161

$$\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{\Sigma fX}{n}$$

=161/101=1.59 ~2 previous pregnancy

Example (3):

Calculate of the mean age of 100 children attending the outpatient clinic

[redacted]

Age (yrs)	f	Midpoint (X)	fX
1 –	18	2	18x2=36
3 –	20	4	20x4=80
5 –	39	6	39x6=234
7 –	17	8	17x8=136
9 – 11	6	10	6x10=60
Total	∑f=100		∑fx=546

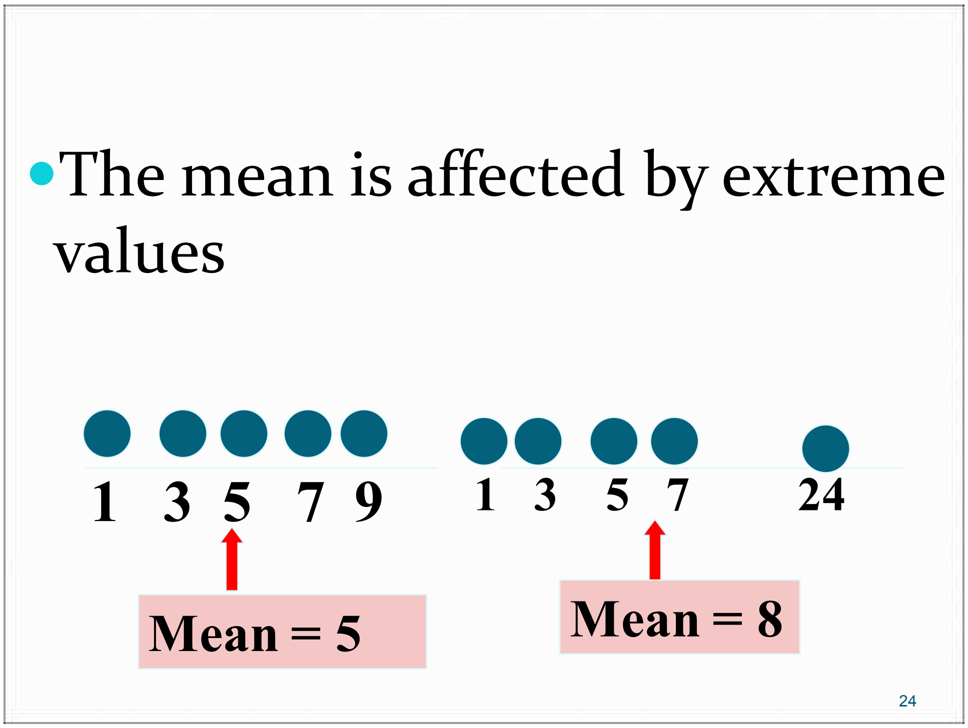
$$\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{\Sigma fX}{n} = 546/100 = 5.46\text{yrs}$$

19

$$\bar{x} = \frac{\sum fx}{n} = \frac{\overset{\boxed{18x2}}{fx_1} + \overset{\boxed{20x4}}{fx_2} + \ldots + \overset{\boxed{6x10}}{fx_5}}{100} = \boxed{5.46 \text{ years}}$$

Properties of the mean

1. The most commonly used measure of central location
2. Uses every value (uses all observations)
3. For each set of data there is only one mean.
4. Influenced (affected) by extreme values (high and low)



2. The Median is the "middle" value when the observations are arranged in ascending or descending order.

Steps for finding the median

- **For ungrouped data**

1. Arrange observations in ascending or descending order
2. Find the position of the median:

$$\text{Median position} = \frac{n + 1}{2}$$

- If n is odd, the median is the middle observation
- If n is even, the median is the average of the two middle observations

Example

Find the median of the time on the internet for the 10 adults of Example (1).

Median position $= \frac{10+1}{2}$

$= 5.5$

Even number of observations

8.5

0, 0, 5, 7, [8, 9], 12, 14, 22, 33

- Suppose only 9 adults were sampled

0, 0, 5, 7, 8, 9,12,14,22

Median position $= \frac{9+1}{2}$

$= 5$

Odd number of observations

0, 0, 5, 7, [8], 9, 12, 14, 22

- <u>**For tabulated (grouped data)**</u>

1. Calculate the cumulative frequency
2. Find the position of the median:

$$\text{Median position} = \frac{n+1}{2}$$

3. **Find the median or the median class (from the table)**
4. **To find the exact value of the median for continuous quantitative variable, apply the following formula:**

$$Median = L + \frac{n/2 - CF}{f}.w$$

- **Where:**
  - ***L*** **is the lower limit of the median class**
  - ***n*** **is the sample size**
  - ***CF*** **is the cumulative frequency preceding the median class**
  - ***f*** **is the frequency of the median class**
  - **W is the width of the class interval.**

Example 2: Calculate the median number of previous pregnancies

Number of Previous pregnancies (x)	Frequency (f)	Cf
0	18	18
1	35	53
2	24	77
3	18	95
4	6	101
Total	∑f=101	

($n+1/2=51$) , The median is 1 pregnancy

Example

X	f	C. Frequency
1 –	18	18
3 –	20	38
5 –	39	77
7 –	17	94
9 – 10	6	100
Total	100	

($n+1/2=50.5$) , The median class is 5 –

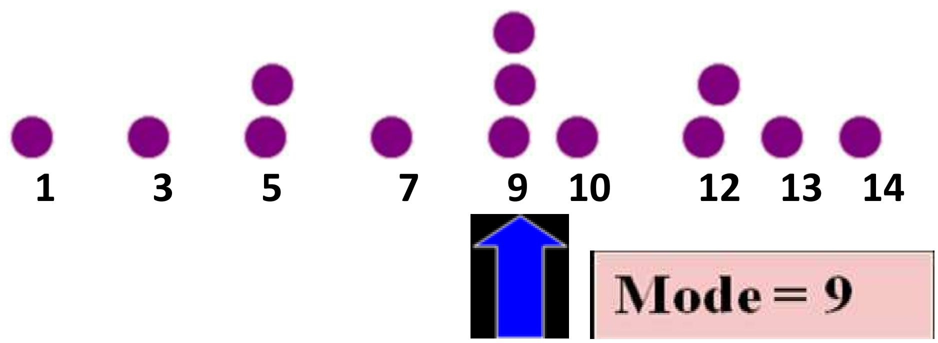
The median $= 5 + \left(\frac{100/2 - 38}{39}\right)(2) = 5.62\,yr.$

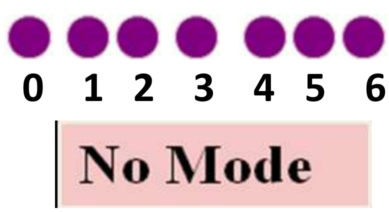
36

Properties of the median

1. It divides the observations into two equal halves (50% of the observations above and 50% below the median)
2. Uses only one or two values
3. For each set of data there is only one median.
4. It is not affected by extreme values

3. **The Mode** is the most frequently occurring observation (value) .





Example 2: Calculate the mode number of previous pregnancies

Number of Previous pregnancies (x)	Frequency (f)
0	18
1	35
2	24
3	18
4	6
Total	101

The mode is 1 previous pregnancy

Example 3.

X	f
1 –	18
3 –	20
5 –	<u>39</u>
7 –	17
9 – 10	6
Total	100

The modal class is (5-)years

36

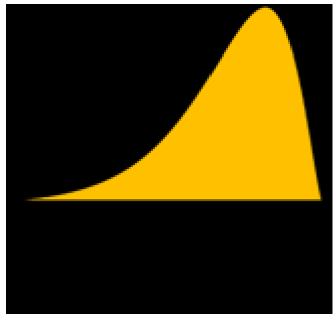
Properties of the mode

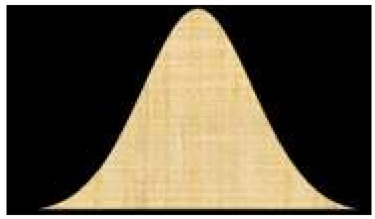
1. It is not affected by extreme values
2. A set of data may have no mode, one mode, two modes or more.

Relationship between Mean, Median, and Mode

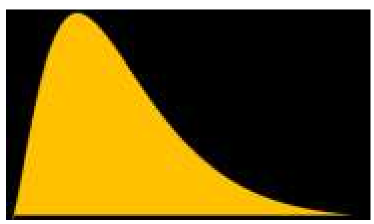
- This depends on the type of distribution

Types of distributions

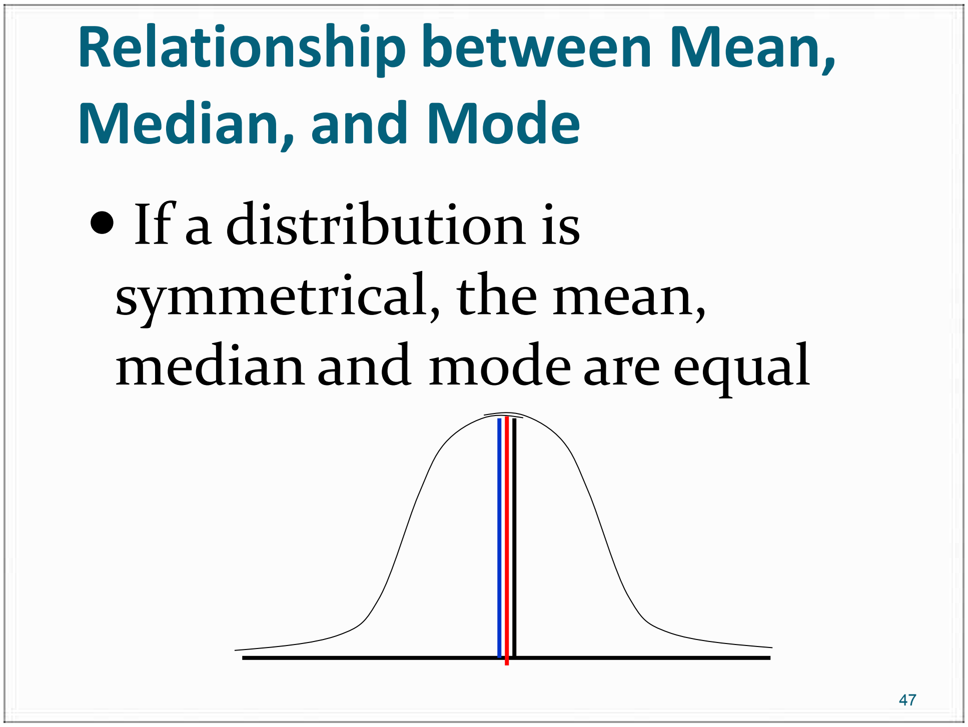




Symmetric
(Not Skewed)



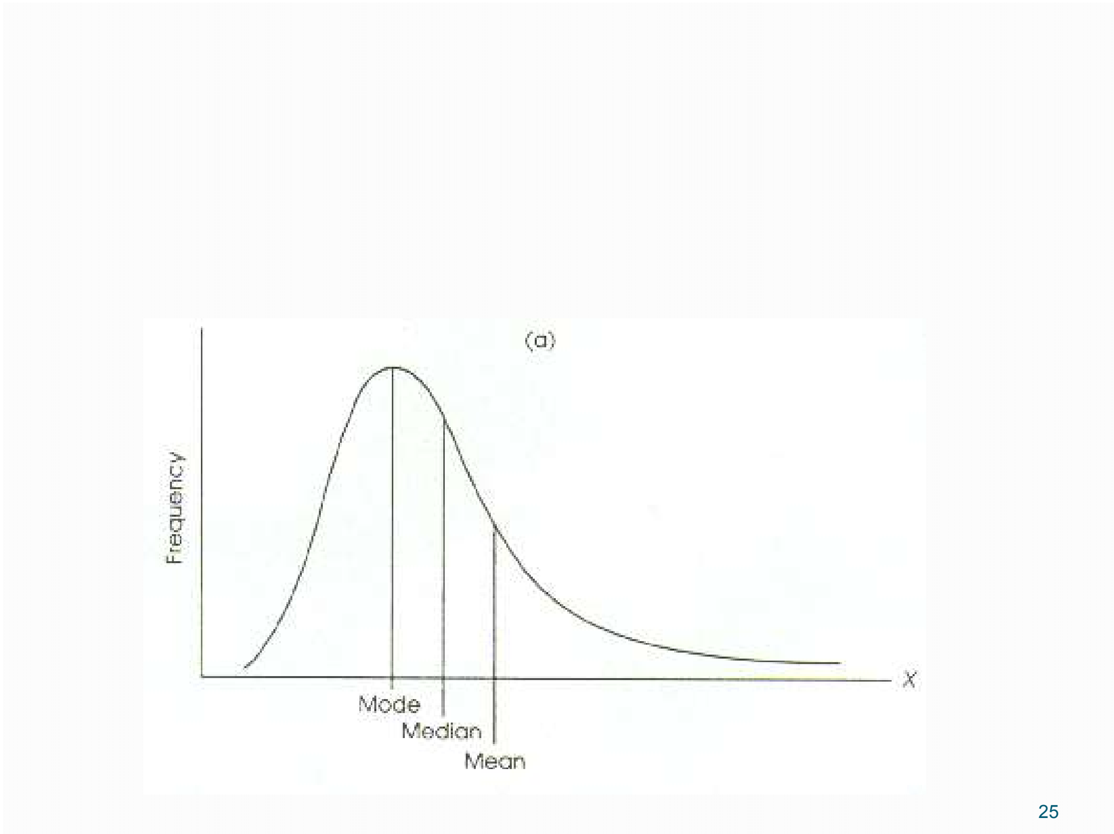
Positively
Skewed



Positively (Right) skewed distribution

Positively skewed: The Mean is to the right of the Median and the Mode, the mean is the largest (highest) value

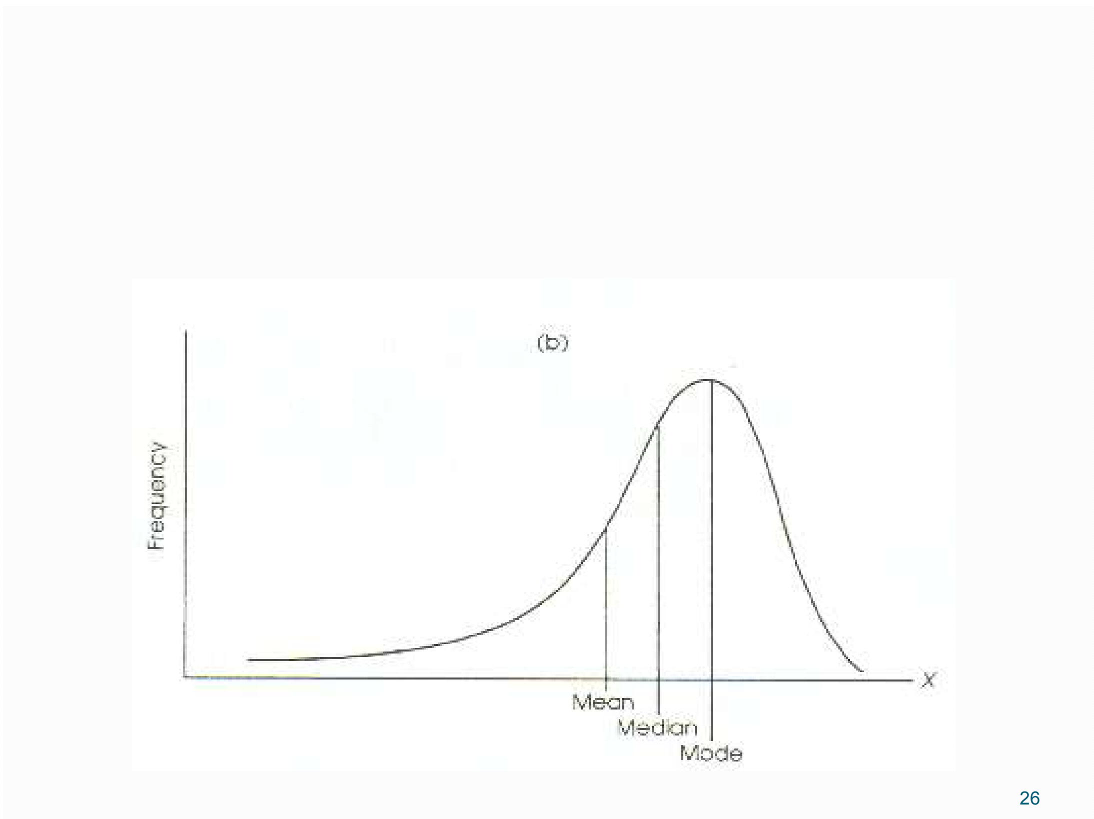
Mode < Median < Mean


(a)
Frequency
X
Mode
Median
Mean
25

Negatively (Left) skewed distribution

Negatively skewed: The Mean is to the left of the Median and the Mode, t**he mean is the smallest (lowest) value.**

Mean < Median < Mode



Questions

1. **Which of the following statements is true about the median?**

a. It is a measure of spread of the data.

b. It is a useful summary measure when the data are skewed to the right

c. It is greater than the arithmetic mean when the data are skewed to the right.

d. Can be distorted by outliers.

2. **Which of the following statements about the mean is TRUE?**

a. The mean is better for summarizing small samples because it is less affected by extreme value

b. When sample values are arranged in ascending way, the mean is the middle value

c. It coincides with the median when the distribution is symmetrical.

d. The mean is always larger than the median

3. **Which descriptive statistic gives the value that occurs most often within a sample?**

   a. mean

   b. median

   c. mode

   d. None of the above

Thank you