



Descriptive Statistics

The Third lecture



Measures of Central Tendency

We will examine in this lecture:

- Mean
- Weighted Mean
- Median
- Mode
- Fractiles (Quartiles-Deciles-Percentiles)

Measure of Central Tendency

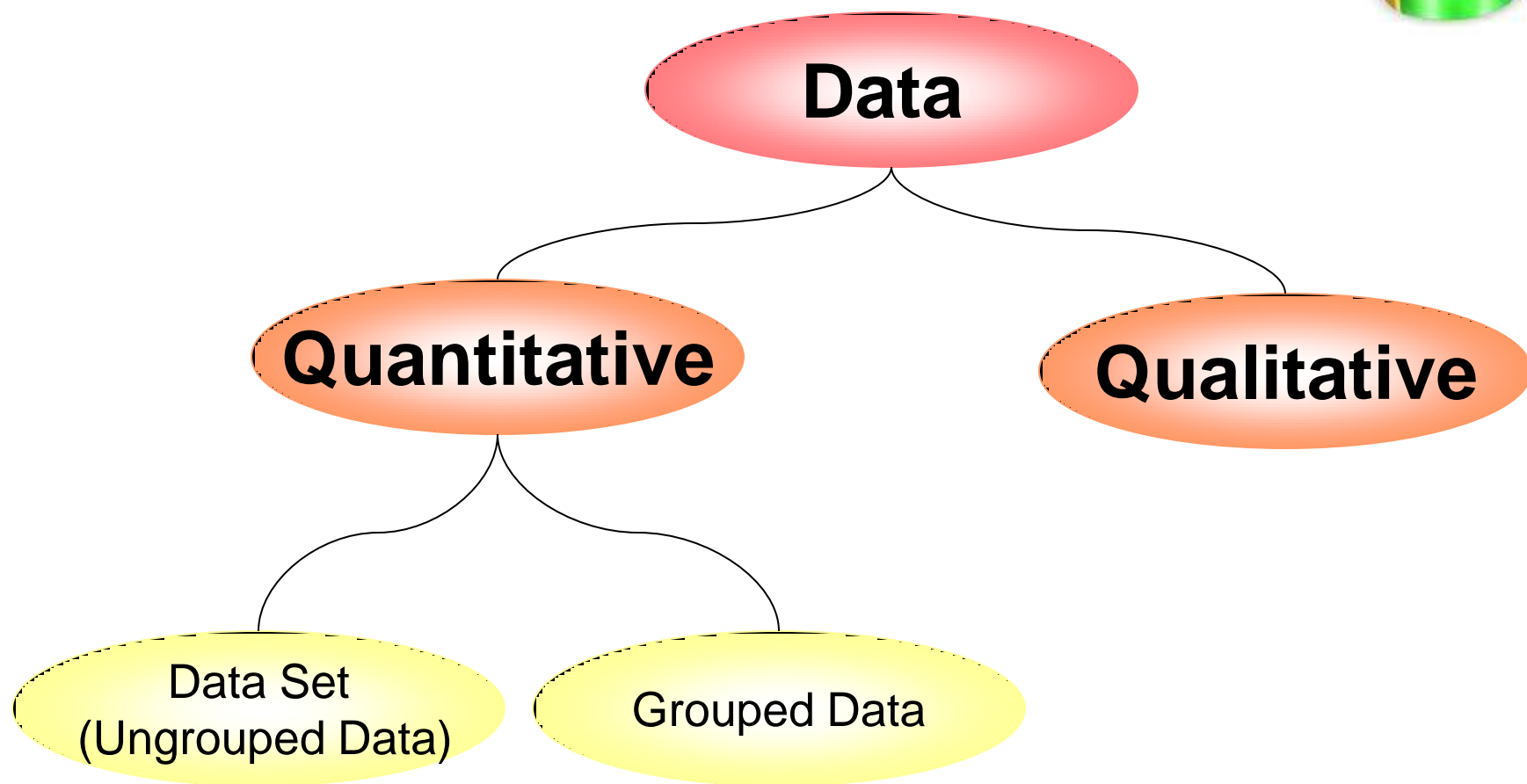


A **Measure of Central Tendency** is a value that represents a typical, or central, entry of data set.

The three most commonly used of central tendency



- **Mean**
- **Median**
- **mode**





First:

The Mean

The Mean of Ungrouped Data



The **Mean** of a data set x_1, x_2, \dots, x_n is the sum of the data entries divided by the numbers of entries.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Finding a sample mean of finite population



Example (1):

the following data represent the marks of 5 students in a course:

60,72,40,80,63

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\bar{x} = \frac{60 + 72 + 40 + 80 + 63}{5} = \frac{315}{5} = 63$$

The mean of grouped data



The **mean** of grouped data is :

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{f_1 + f_2 + \dots + f_k}$$

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i}$$

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

Where x_1, x_2, \dots, x_k are the midpoints and f_1, f_2, \dots, f_k are the frequencies of a class.

Finding the mean of grouped Data

Example (2)



Find the mean of student's age of the given data

Class intervals	frequency f_i	True classes	midpoints x_i	$x_i f_i$
5-6	2	4.5-6.5	5.5	11
7-8	5	6.5-8.5	7.5	37.5
9-10	8	8.5-10.5	9.5	76.0
11-12	4	10.5-12.5	11.5	46.0
13-14	1	12.5-14.5	13.5	13.5
Total	20			184

$$2 \times 5.5$$

$$5 \times 7.5$$



The mean of student's age is :

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

$$\bar{x} = \frac{184}{20} = 9.2$$

Mean Property:



the sum of the deviation of a set of values from their mean is 0.

If we have the observation x_1, x_2, \dots, x_n and the deviation from their mean d_1, d_2, \dots, d_n

so $d_i = x_i - \bar{x} \quad , \quad i = 1, 2, \dots, n$

then
$$\sum_{i=1}^n d_i = \sum_{i=1}^n (x_i - \bar{x}) = 0$$

Example (3):

the following data represent the marks of 5 students in a course:

60, 72, 40, 80, 63, and the mean $\bar{x} = 63$



x_i	$d_i = x_i - \bar{x}$
60	-3
72	9
40	-23
80	17
63	0
Total	0

Some advantage of using the mean:



1. For a given set of data there is one and only one mean (uniqueness) .
2. It takes every entry into account.
3. It is easy to understand and to compute.

Some disadvantage of using the mean:



1. Affected by extreme values. Since all values enter into the computation.
2. It can't be calculated with the open table.
3. It can't be used with qualitative data.

Example (4):

The mean of the data 1, 2, 3, 3, 2, 2, 3, 100 is 14.5



Second:

The Weighted Mean

The Weighted Mean



Is the mean of a data set x_1, x_2, \dots, x_n whose entries have varying weights w_1, w_2, \dots, w_n .

A weighted mean is given by:

$$\bar{x}_w = \frac{x_1 w_1 + x_2 w_2 + \dots + x_n w_n}{w_1 + w_2 + \dots + w_n}$$

$$\bar{x}_w = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$



Example (5)

Find the weighted mean \bar{x}_w of student's marks in three courses if we have the marks 40,70,65 and the study hours for these courses are 2,3,4 respectively.

$$\bar{x}_w = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

$$\bar{x}_w = \frac{40 \times 2 + 70 \times 3 + 65 \times 4}{2 + 3 + 4} = \frac{550}{9} = 61.11$$



Third:

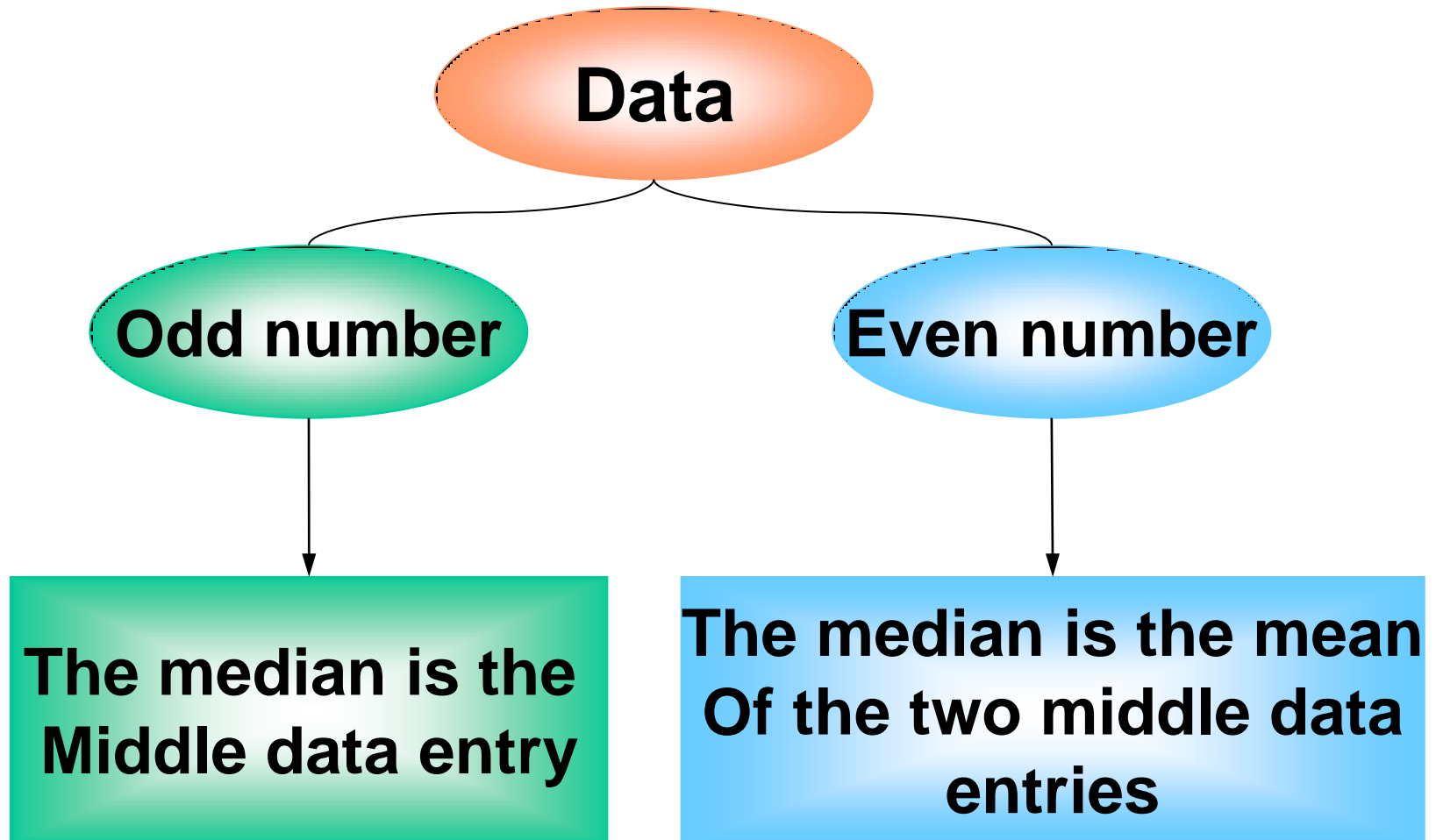
The median



The median:

The **median** of a data set is the value that lies in the middle of the data when the data set is ordered.

The median of a data set:





Example (6)

Find the median of the student's marks
60,72,40,80,63 :

First order the data

40,60,63,72,80



The median



Example (7)

Find the median of the student's marks
72, 60, 72, 40, 80, 63 :

First order the data

40, 60, 63, 72, 72, 80

$$Med = \frac{63 + 72}{2} = 67.5$$

Some advantage of using the median:



1. Don't affected by the extreme values.
2. It can be calculated with the open table.
3. It can be used with qualitative data.

Example (8):

Find the median of A, A, B, C, D

Med

Some disadvantage of using the median:



1. It don't takes every entry into account.
2. It is not easy to use in statistical analyses.



Fourth:

The mode

The mode



The **mode** of a data set is the data entry that occurs with the greatest frequency.

Finding the mode of a data set



Example (9)

Find the mode of the given data:

2, 6, 9, 4, 6, 10, 6

$$\text{Mod} = 6$$



Example (10)

Find the mode of the given data:

4, 2, 7, 4, 7, 10, 7

Mod = 7



Example (11)

Find the mode of the given data:

4, 7, 4, 7, 8, 9, 7, 4, 10

Mod = 7, 4



Example (12)

Find the mode of the given data:

4,9,8,12,11,7,15

There is no mode



Example (13)

Find the mode of the given data:

4,4,5,5,6,6,7,7

There is no mode

Some advantage of using the mode:



1. Don't affected by the extreme values.
2. It can be calculated with the open frequencies table.
3. It can be used with qualitative data.
4. It is easy measurement.

Example (14):

find the mode **A, A, B, C, D**

$$Mod = A$$

Some disadvantage of using the mode

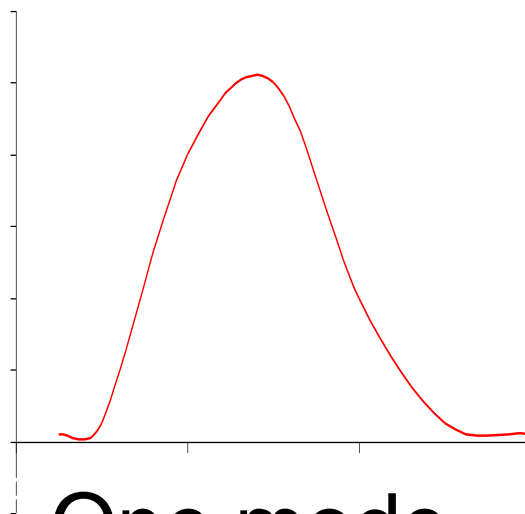


1. It doesn't take every entry into account.
2. In such cases, the mode may not exist or may not be very meaningful.
3. Some data have no mode.

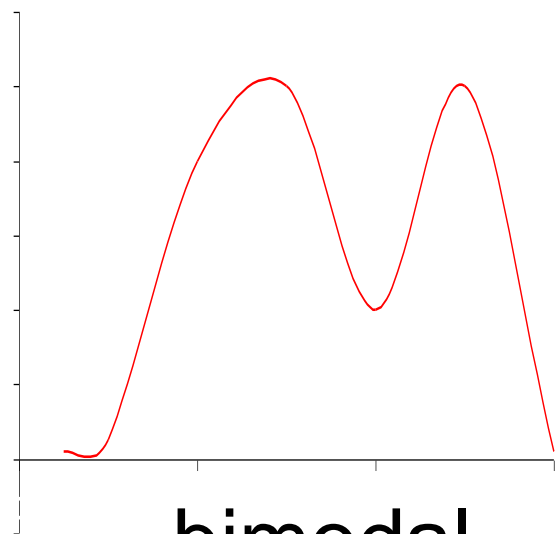


Set of data may have:

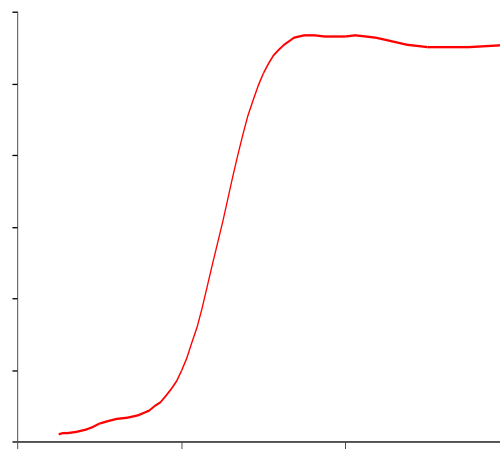
- one mode
- more than one mode (bimodal)
- no mode



One mode



bimodal



No mode

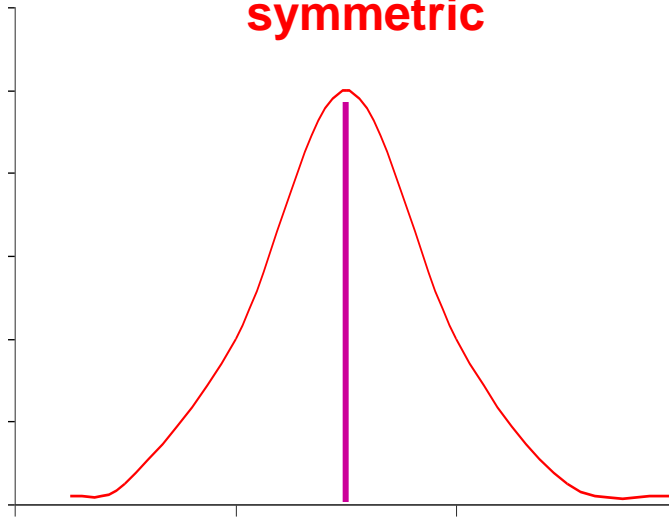


The relation between the mean, median, and mode:

The frequency distribution with one mode :

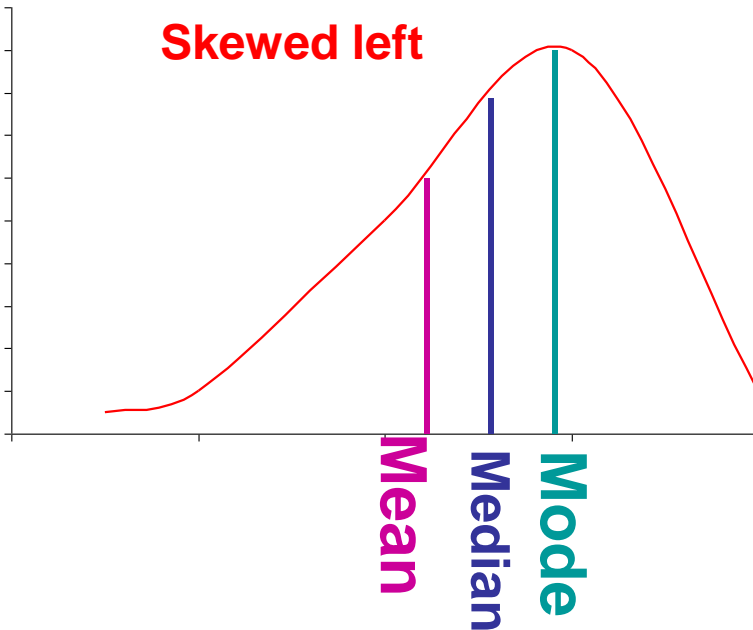
$$\frac{(\text{mean} - \text{mode})}{3} = (\text{mean} - \text{median})$$

symmetric

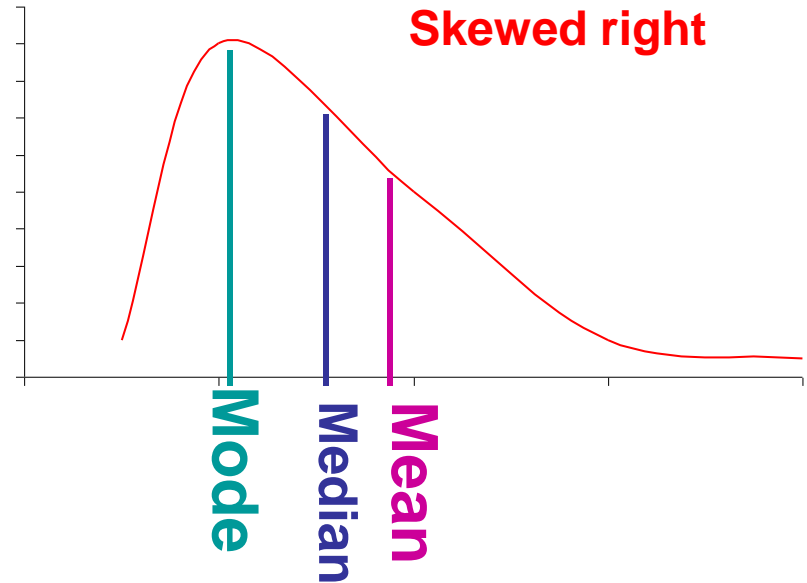


Mode = Mean = Median

Skewed left



Skewed right



Example (15):

Find the mean, median, and the mode of these data , Determine which measure of central tendency is the best to represent the data?

20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

The mean:

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{n} = \frac{475}{20} = 23.8$$

Sort the data from lowest to highest values



20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Median

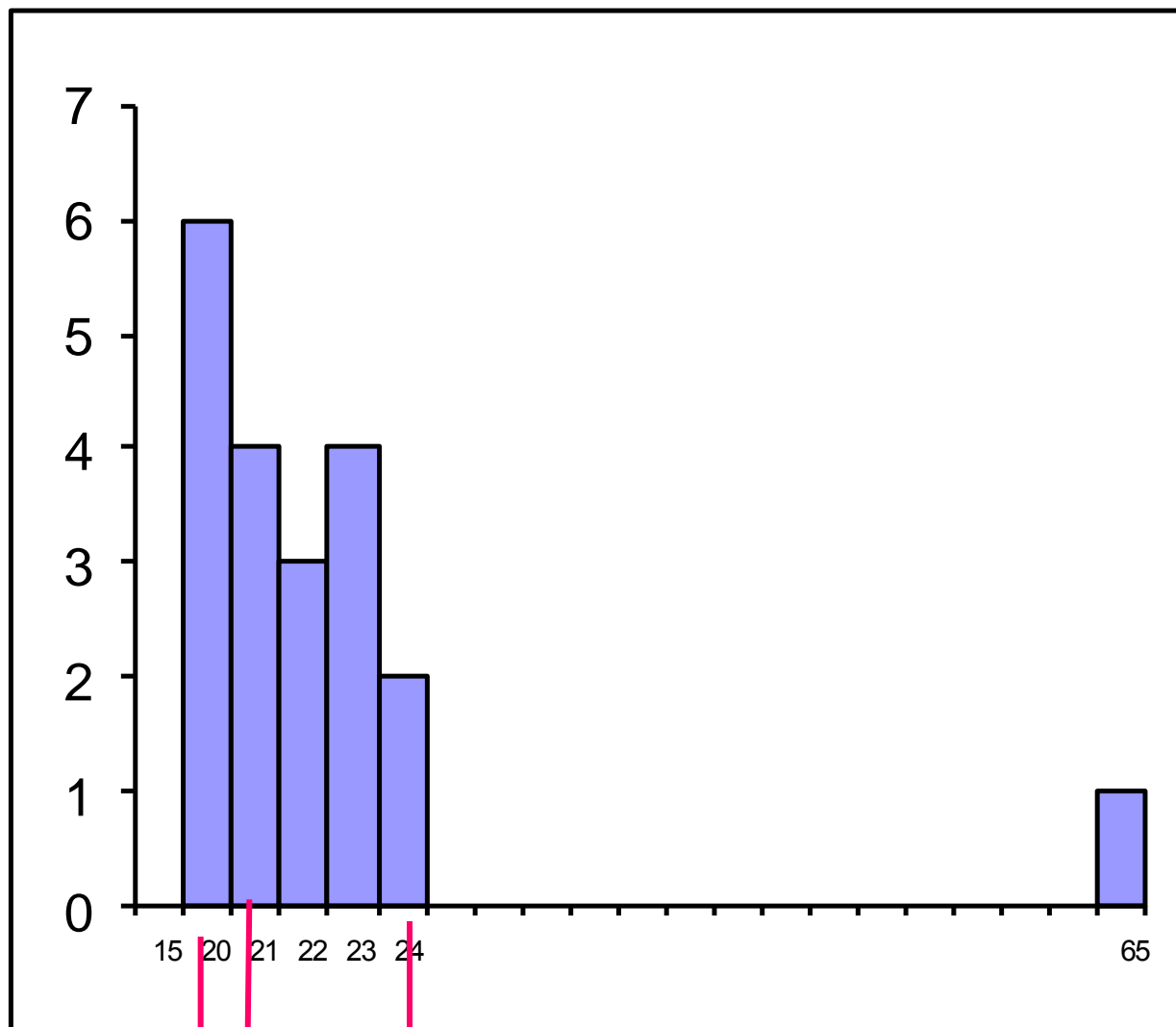
$$Med = \frac{21 + 22}{2} = 21.5$$

Mode

$$Mod = 20$$

Mean

$$\bar{x} = 23.8$$



$Mod = 20$

$Med = 21.5$

$\bar{x} = 23.8$



Examples

Example (16):

1. Find the mean, median, and mode of these data, if possible. If not explain why?
2. Determine which measure of central tendency is the best to represent the data

6, 6, 9, 9, 6, 5, 5, 5, 7, 5, 5, 5, 8

The mean

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{n} = \frac{81}{13} = 6.2308$$

Sort the data from lowest to highest values



5, 5, 5, 5, 5, 5, 6, 6, 6, 7, 8, 9, 9

Median

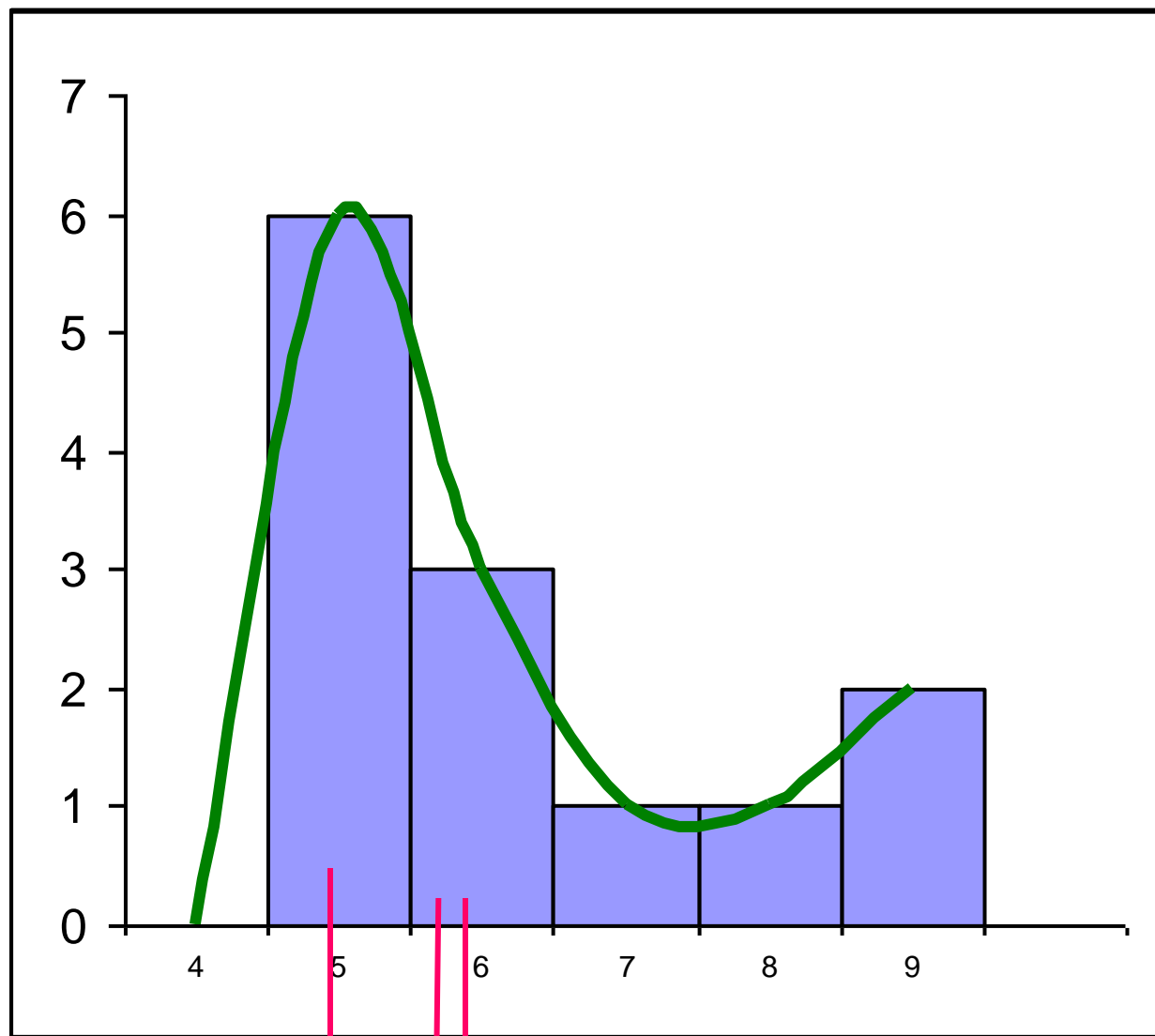
$$Med = 6$$

Mode

$$Mod = 5$$

Mean

$$\bar{x} = 6.2308$$



$Mod = 5$

$Med = 6$

$\bar{x} = 6.2308$

Example (17):

- 1. Find the mean, median, and mode of this data, if possible. If not explain why?**
- 2. Determine which measure of central tendency is the best to represent the data**

The responses by a sample of 1040 people who were asked if their next vehicle purchase will be foreign or domestic

Domestic : 346

foreign : 450

Don't know : 244

Domestic : 346

foreign : 450

Don't know : 244



Mean It can't be find because the data are qualitative.

Median It can't be find because the data are not ordered

Mode $Mod = \text{foreign}$

Example (18):

1. Find the mean, median, and mode of this data, if possible. If not explain why?
2. Determine which measure of central tendency is the best to represent the data

0		8
1		5 6 8
2		1 3 4 5
3		0 9
4		0 0

Key: $0|8 = 0.8$

Mode $Mod = 4$

0		8
1		5 6 8
2		1 3 4 5
3		0 9
4		0 0

Key: $0|8 = 0.8$

Mean

$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{n}$$

$$= \frac{0.8 + 1.5 + 1.6 + 1.8 + 2.1 + 2.3 + 2.4 + 2.5 + 3 + 3.9 + 4 + 4}{12}$$

$$= \frac{29.9}{12} = 2.4917$$

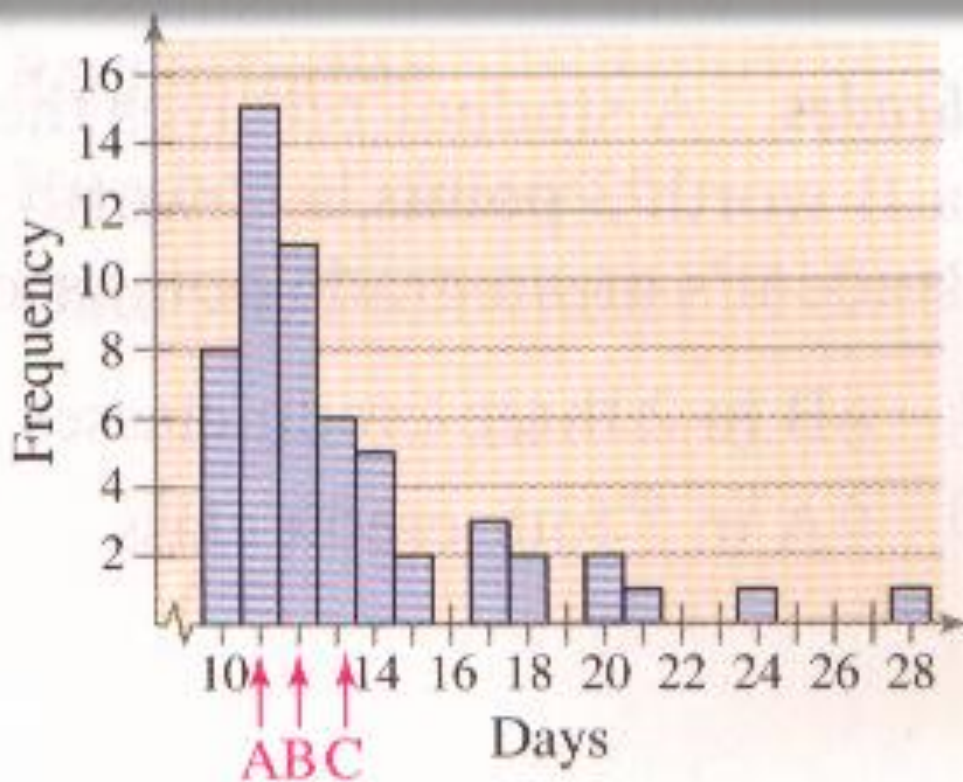
Median

$$Med = \frac{2.3 + 2.4}{2} = 2.35$$



Example (19):

the letters A,B, and C are marked on the horizontal axis. Determine which is the mean, median , and the mode. Justify your answer.



Mode $Mod = A$

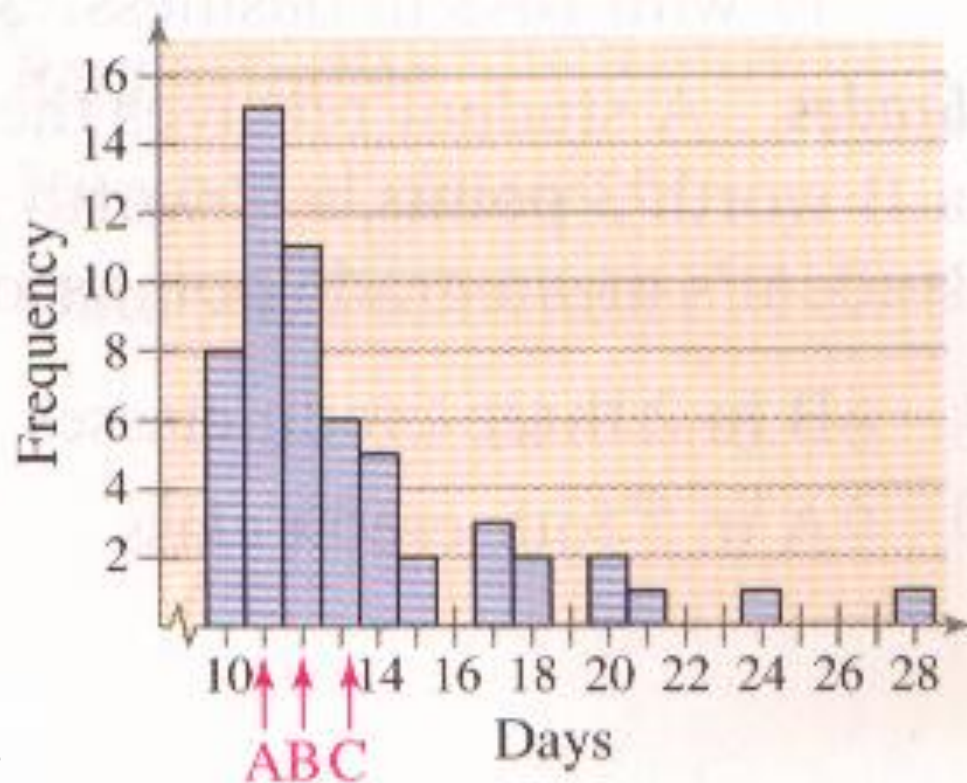
Because: it is the greatest frequency.

Mean $\bar{x} = C$

Because: there is an extreme value ,or the graph is skewed-right.

Median $Med = B$

Because: the median is between the mean and mode

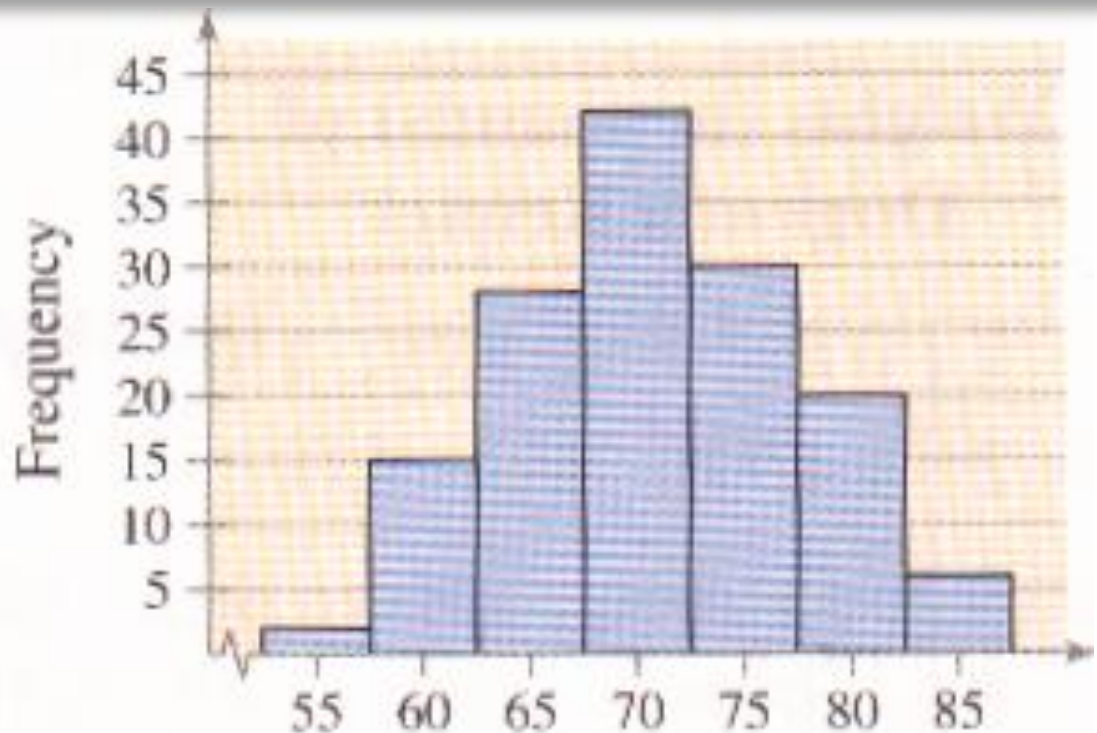


Example (20):



Determine which measure of central tendency is the best to represent the graphed data without performing any calculation

**The mean ,
because the
data are
quantitative
and there is
no outliers**





Fractiles:

Fractiles are numbers that partition, or divide, an ordered data set into equal parts like Quartiles, Deciles and Percentiles.

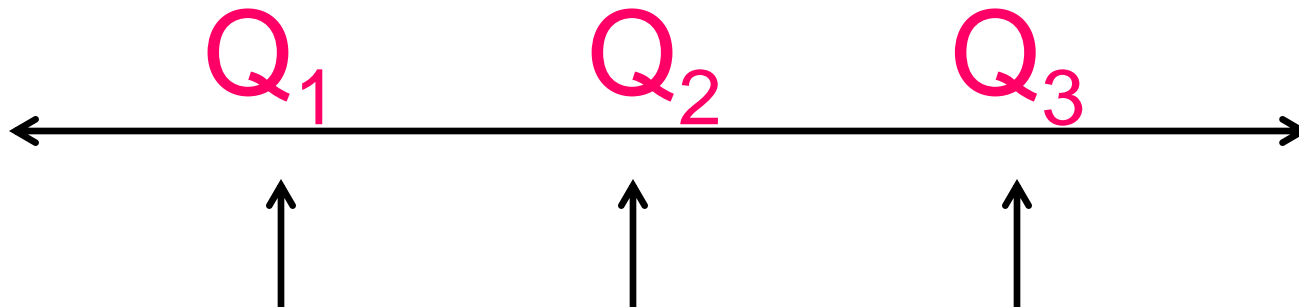
Example:

The median is a fractile because it divides an ordered data set into two equal parts.



Quartiles:

Quartiles are numbers that divide a data set into 4 equal parts. Quartiles symbolized by Q_1 , Q_2 and Q_3





Example (21):

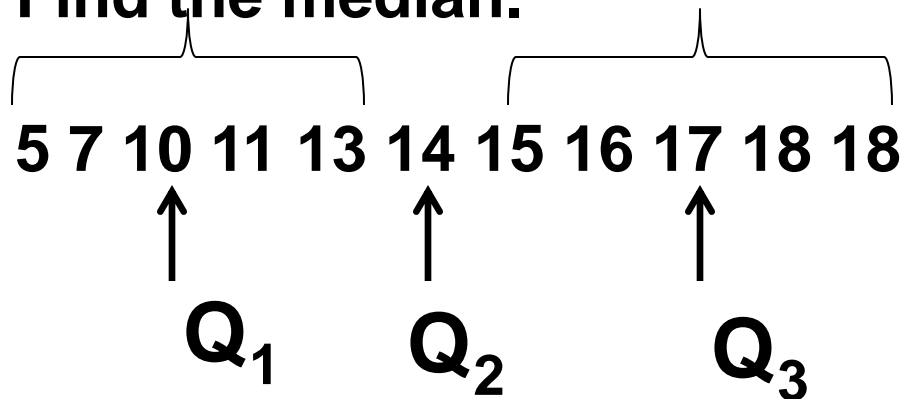
Find Q_1 , Q_2 and Q_3 for the data set
13, 18, 17, 15, 14, 7, 10, 11, 5, 18, 16

Solution:

Arrange the data in order.

5 7 10 11 13 14 15 16 17 18 18

Find the median.





Example (22):

Find Q_1 , Q_2 and Q_3 for the data set
15, 13, 6, 5, 12, 50, 22, 18.

Solution:

Arrange the data in order.

5 6 12 13 15 18 22 50

Find the median.

$$Med = \frac{13 + 15}{2} = 14$$

$$Q_2 = 14$$



Find the median of the data values less than 14.

5 6 12 13

$$Q_1 = \frac{6 + 12}{2} = 9$$

$$Q_1 = 9$$

Find the median of the data values greater than 14.

15 18 22 50

$$Q_3 = \frac{18 + 22}{2} = 20$$

$$Q_3 = 20$$

Hence, $Q_1=9$, $Q_2=14$, and $Q_3=20$



Definition:

The interquartile range (**IQR**) of a data set is the difference between the third and first quartiles.

$$\text{Interquartile range (IQR)} = Q_3 - Q_1$$

Example (23):

From Example (22), $Q_1=9$, and $Q_3=20$. Find **IQR**.

$$\begin{aligned}\text{IQR} &= Q_3 - Q_1 \\ &= 20 - 9 \\ &= 11\end{aligned}$$



Deciles:

Deciles are numbers that divide a data set into 10 equal parts. Quartiles symbolized by $D_1, D_2 \dots D_9$

Percentiles:

Percentiles are numbers that divide a data set into 100 equal parts. Percentiles symbolized by $P_1, P_2 \dots P_{99}$



Remarks:

There are relationships among percentiles ,deciles and quartiles.

Deciles are denoted by D_1, D_2, \dots, D_9 , and they correspond to $P_{10}, P_{20}, \dots, P_{90}$.

Quartiles are denoted by Q_1, Q_2, Q_3 , and they correspond to $P_{25}, P_{50}, \dots, P_{75}$.

The median is the same as P_{50} or Q_2 or D_5 .