Methods of Mathematical Presentation

(Summary Statistics)

(1)Measures of Central Tendency

Mid-range

Mode

Median

Arithmetic mean

Objectives

After this session participants will be able to do the following

- Compute and interpret the following measures of central tendency:
  - Mode
  - Median
  - Arithmetic mean
- Choose and apply the suitable measure of central tendency

1- Mode

The observation or observations of highest frequency

A) Ungrouped data Examples: **Weight (kg)**

69,	67,	70,	73,	69,	71		

Mode = 69 Kg

12,	14,	16,	18,	16,	14		

Mode = 14, 16 Kg

18	11	16	14	19	15	13	12

No Mode

16	12	16	14	18	16	14	12

Mode = 16 Kg

The mode is the most frequently occurring value in a set of discrete data. There can be more than one mode if two or more values are equally common.

Example

Suppose the results of an end of term Statistics exam were distributed as follows:

Student	Score
1	94
2	81
3	56
4	90
5	70
6	65
7	90
8	90
9	30

B) Grouped data

Examples:

Weight (kg)	Frequency
25-	14
30-	2
35-	14
40-	9
60-75	4
Total	43

Interval of 1st mode = 25-30 Kg

Interval of 2nd mode = 35-40 Kg

$$1^{st}\ \text{mode} = \frac{25+30}{2} = 27.5\ \text{Kg}$$

$$2^{nd}\ \text{mode} = \frac{35+40}{2} = 37.5\ \text{Kg}$$

Example

Bld. Gr.	Frequency
A	10
B	14
AB	25
0	9
Total	58

Mode is AB


診症區域 只可一人陪同
CONSULTATION AREA
ONE ACCOMPANIAN ONLY
Triage Station

Advantages

- Easy
- Used with all types of variables
- Not affected with extreme observations
- Obtained from closed-ended or open-ended tables

Disadvantages

- Neglects the less frequent observations
- Sometimes there is no mode
- The distribution may be bi-modal or multi-modal

2- Median

Observation which lies in the middle of the ordered observations

A) Ungrouped data

Odd number of observations:

- **Arrange observations → Ascending order**
- **Rank of median = (n + 1)/2**

Example:

- **Row data → 24 – 18 – 22 – 20 - 16 kg**
- **Arranged data → 16 – 18 – 20 – 22 – 24 kg**
- **Rank = (5+1)/2**
- **Median = value of 3rd observation = 20 kg**

Even number of observations:

- **Arrange observations → Ascending order**
- **Rank of two middle observations = (n/2), (n/2)+1**

Example:

- **Row data → 26 - 24 – 18 – 22 – 20 - 16 kg**
- **Arranged data → 16 – 18 – 20 – 22 – 24 - 26 kg**
- **Rank = (6/2), (6/2)+1 = 3, 4**
- **Median = Average of 2 middle observations =**

(20+22) / 2 = 21 kg

Advantages

- Used with quantitative variables and qualitative ordinal
- Not affected by extreme (outlying) observations
- Suitable for biological values (not normal)

Disadvantage

- Does not take all observations into consideration

The median is the value halfway through the ordered data set, below and above which there lies an equal number of data values.

The median is the 0.5 quantile

Example

With an odd number of data values, for example 21, we have:

Data	96 48 27 72 39 70 7 68 99 36 95 4 6 13 34 74 65 42 28 54 69
Ordered Data	4 6 7 13 27 28 34 36 39 42 *48* 54 65 68 69 70 72 74 95 96 99
Median	48, leaving ten values below and ten values above

With an even number of data values, for example 20, we have:

Data	57 55 85 24 33 49 94 2 8 51 71 30 91 6 47 50 65 43 41 7
Ordered Data	2 6 7 8 24 30 33 41 43 *47 49* 50 51 55 57 65 71 85 91 94
Median	Halfway between the two 'middle' data points - in this case halfway between 47 and 49, and so the median is 48

Calculate the median age, weight and height of the group

3- The arithmetic mean

A) Ungrouped data

$$\text{Mean} = \frac{\text{Sum of all observations } (\sum X)}{\text{Number of observations } (n)}$$

Example:

- 24 – 20 – 22 – 16 – 18 kg
- X1 – X2 – X3 – X4 – X5

$$\text{Mean} = \frac{24+20+22+16+18}{5} = \frac{100}{5} = 20 \text{ Kg}$$

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

Exercise : 14 subjects

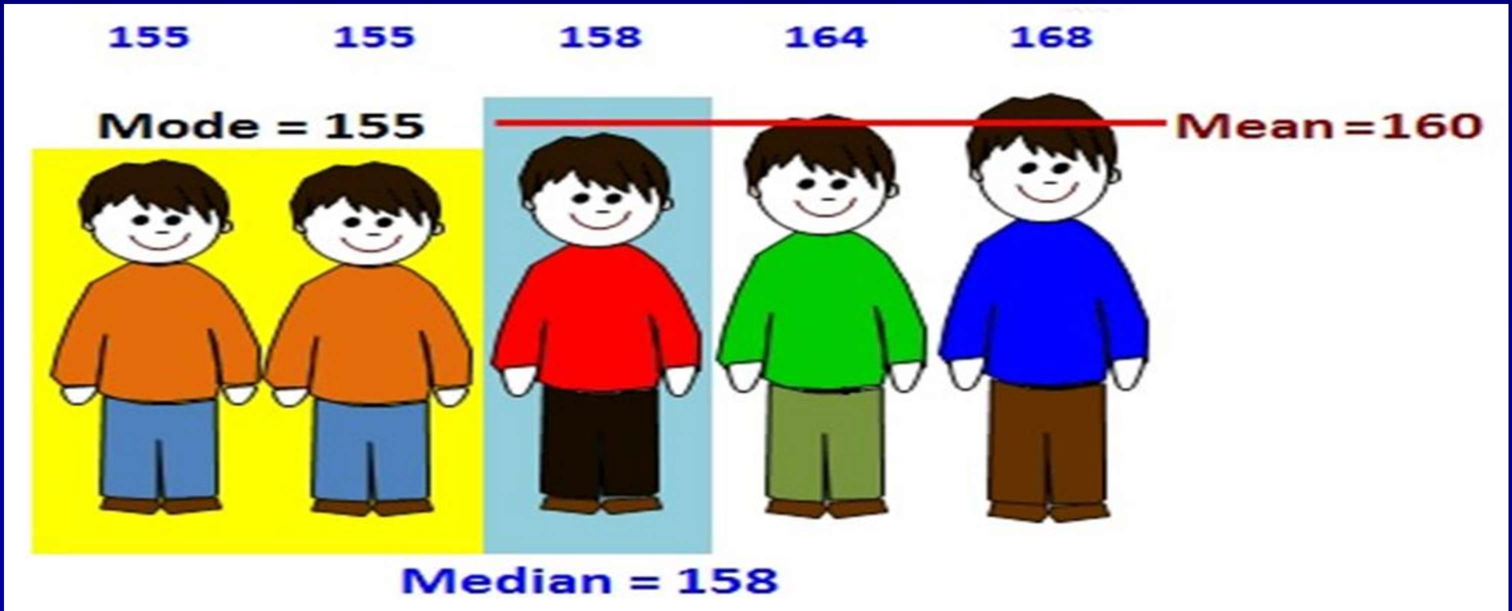
Body Mass Index: 24.4 30.4 21.4 25.1 21.3 23.8 20.8 22.9 20.9 23.2 21.1 23.0 20.6 26.0

Compute the mean, median, mid range, and mode

Mean = 23.2

Median = 22.9

Mode = 20.6



Find Mean, Median and Mode of Ungroup Data

The weekly pocket money for 9 first year pupils was found to be:

3 , 12 , 4 , 6 , 1 , 4 , 2 , 5 , 8

Mean	Median	Mode
5	4	4

B) Grouped data

Example:

Weight (kg)	Frequency f_j	Mid-point of interval X_j	f_jX_j
15-	3	20	60
25-	6	30	180
35-	8	40	320
45-	2	50	100
55-65	1	60	60
Total	20 Σf_j		720 $\Sigma f_j X_j$

$$\bar{X} = \frac{720}{20} = 36 \text{ Kg}$$

Example

Number of attacks of diarrhea	Frequency f_j	Mid-point of interval X_j	$f_j X_j$
1-	5	1.5	7.5
3-	4	3.5	14
5-	3	5.5	16.5
7-9	5	8	40
Total	17 Σf_j		78 $\Sigma f_j X_j$

$$\bar{X} = \frac{78}{17} = 4.6 \cong 5 \text{ attacks}$$

Exercise: 14 subjects

Weight: 83.9 99.0 63.8 71.3 65.3 79.6 70.3 69.2 56.4 66.2 88.7 59.7 64.6 78.8

Height: 185 180 173 168 175 183 184 174 164 169 205 161 177 174

For each variable compute the mean and median

Wt Mean = 72.63 , Median = 69.75

Ht Mean = 176.57 , Median = 174.5

Geometric Mean

Geometric mean is defined as the positive root of the product of observations. Symbolically,

$$\boxed{G = (x_1 x_2 x_3 \cdots\cdots\cdots\cdots x_n)^{1/n}}$$

It is also often used for a set of numbers whose values are meant to be **multiplied together or are exponential in nature,** such as data on the growth of the human population or interest rates of a financial investment.

Find geometric mean of rate of growth: 34, 27, 45, 55, 22, 34

Harmonic Mean

Typically, it is appropriate for situations when the average of **rates** is desired.

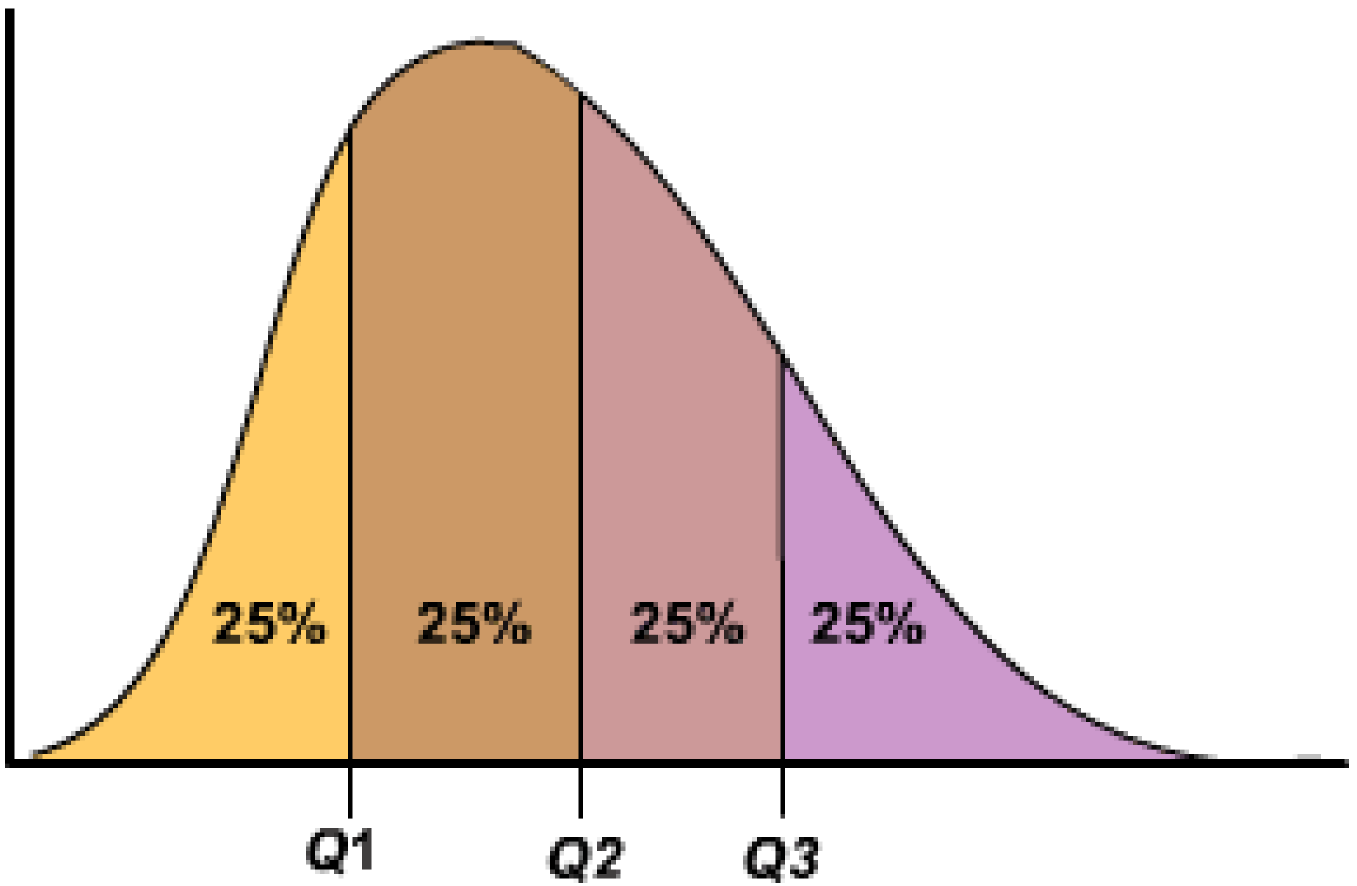
The harmonic mean is the number of variables divided by the sum of the reciprocals of the variables.

Useful for ratios such as speed (=distance/time) etc.

Quantile

Quantiles are a set of 'cut points' that divide a sample of data into groups **containing (as far as possible) equal numbers of observations**.

Examples of quantiles include quartile, quintile, percentile.


25%
25%
25%
25%
Q1
Q2
Q3

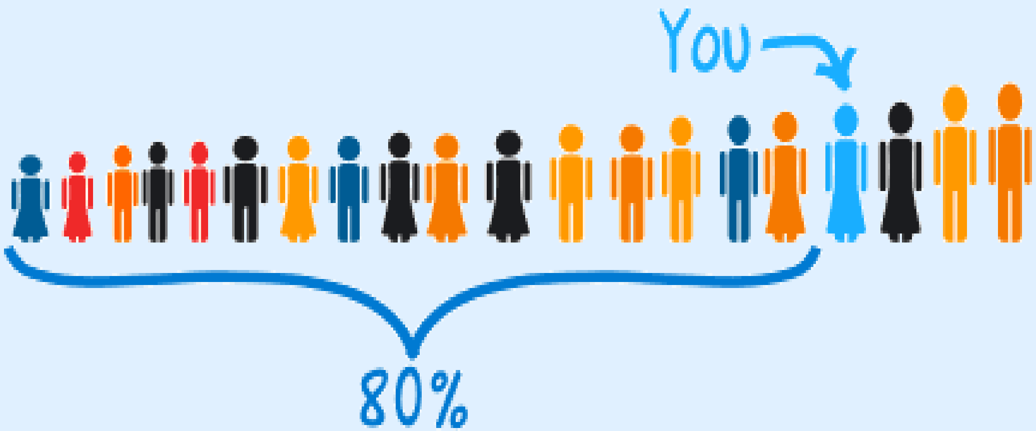
Percentile

Percentiles are values that divide a sample of data into one hundred groups containing (as far as possible) equal numbers of observations.

For example, 30% of the data values lie below the 30th percentile

Example: You are the fourth tallest person in a group of 20

80% of people are shorter than you:



That means you are at the **80th percentile**.

If your height is 1.85m then "1.85m" is the 80th percentile height in that group.

Quartile

Quartiles are values that **divide a sample of data into four groups** containing (as far as possible) equal numbers of observations.

A data set has three quartiles.

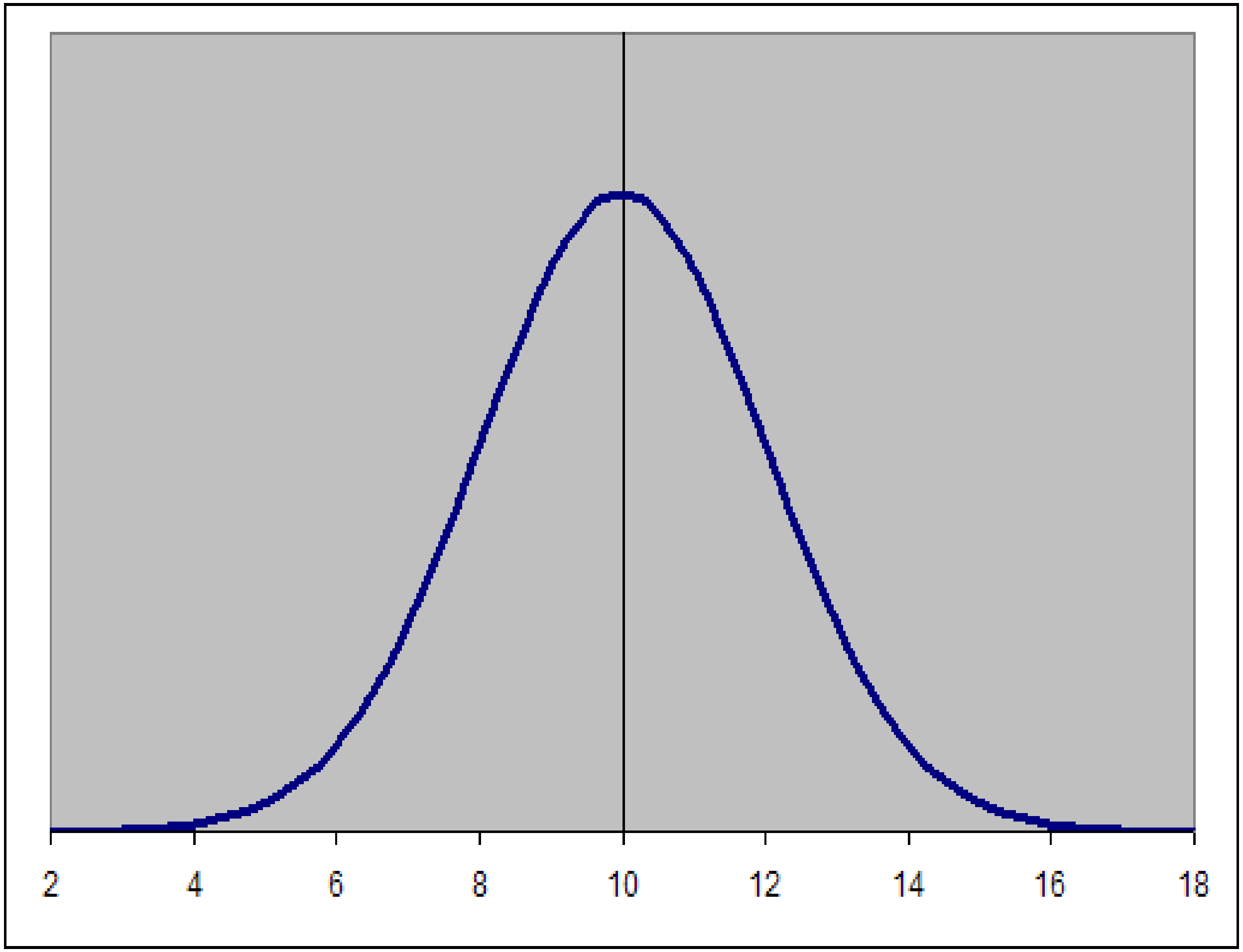
References to quartiles often relate to just the outer two, the upper and the lower quartiles; the second quartile being equal to the median. The lower quartile is the data value a quarter way up through the ordered data set; the upper quartile is the data value a quarter way down through the ordered data set.

Data	6 47 49 15 43 41 7 39 43 41 36
Ordered Data	6 7 15 36 39 41 41 43 43 47 49
Median	41
Upper quartile	43
Lower quartile	15

The Shape of the Distribution

The normal distribution is pattern for the distribution of a set of data which follows a bell shaped curve

This distribution is sometimes called the Gaussian distribution


2
4
6
8
10
12
14
16
18

The bell shaped curve has several properties:

1- The curve concentrated in the center and decreases on either side. This means that the data has less of a tendency to produce unusually extreme values, compared to some other distributions.

2- The bell shaped curve is symmetric. This tells you that the probability of deviations from the mean are comparable in either direction.

Measures of Skewness and Kurtosis

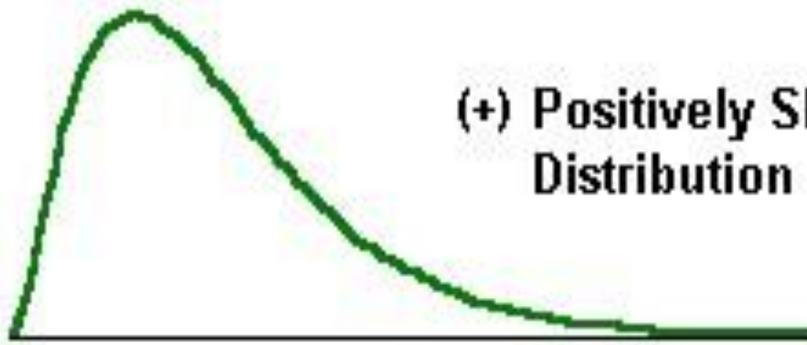
A fundamental task in many statistical analyses is to characterize the *location* and *variability* of a data set.

A further characterization of the data includes skewness and kurtosis.

Skewness is a measure of symmetry, or more precisely, the lack of symmetry.

A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.

If Mean = Median → Symmetry or zero skewness distribution



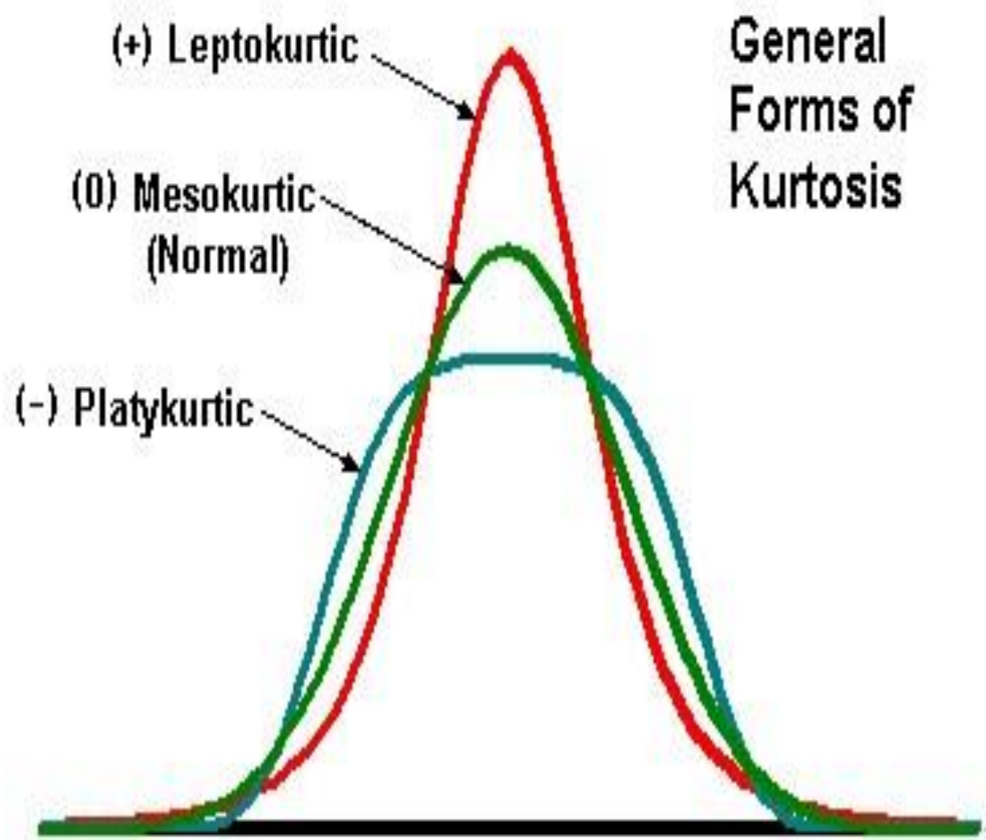
(+) Positively Skewed Distribution

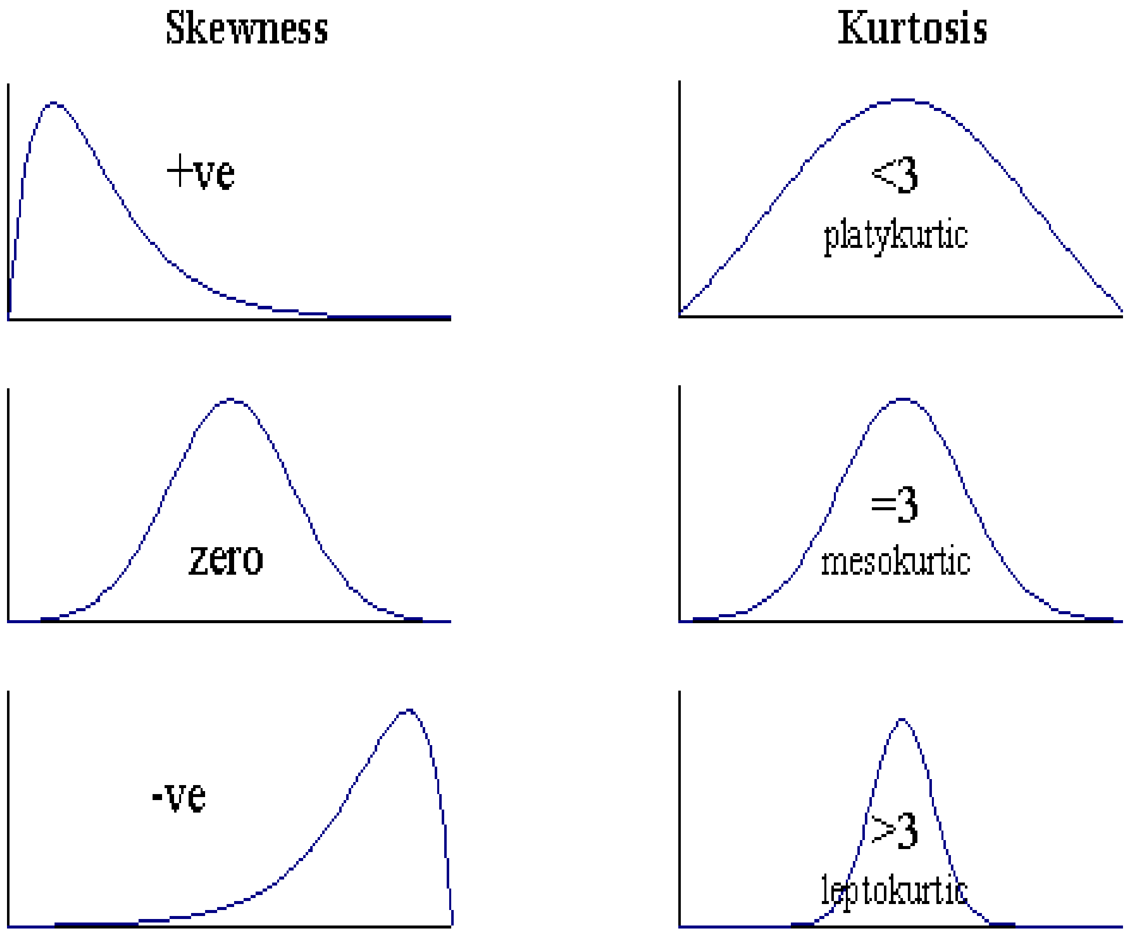
(-) Negatively Skewed Distribution

Kurtosis

Kurtosis is a **measure of whether the data are peaked or flat relative to a normal distribution.** That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails.

Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case.


(+) Leptokurtic
(0) Mesokurtic
(Normal)
(-) Platykurtic
General
Forms of
Kurtosis


Skewness
Kurtosis
+ve
<3
platykurtic
zero
=3
mesokurtic
-ve
>3
leptokurtic

Find the appropriate measure of central tendency for each variable

Respondent	Sex	Social Class	No. of Years in Party	Education	Marital Status	Number of Children
A	M	High	32	High school	Married	5
B	M	Medium	17	High school	Married	0
C	M	Low	32	High school	Single	0
D	M	Low	50	Eighth grade	Widowed	7
E	M	Low	25	Fourth grade	Married	4
F	M	Medium	25	High school	Divorced	3
G	F	High	12	College	Divorced	3
H	F	High	10	College	Separated	2
I	F	Medium	21	College	Married	1
J	F	Medium	33	College	Married	5
K	M	Low	37	High school	Single	0
L	F	Low	15	High school	Divorced	0
M	F	Low	31	Eighth grade	Widowed	1

The following table presents the annual person-hours of time lost due to traffic congestion for a group of cities for 2007. This statistic is a measure of traffic congestion

(*continued*)

City	Annual Person-Hours of Time Lost to Traffic Congestion per Year per Person
Baltimore	25
Boston	22
Buffalo	5
Chicago	22
Cleveland	7
Dallas	32
Detroit	29
Houston	32
Kansas City	8
Los Angeles	38
Miami	27
Minneapolis	22
New Orleans	10
New York	21
Philadelphia	21
Pittsburgh	8
Phoenix	23
San Antonio	21
San Diego	29
San Francisco	29
Seattle	24
Washington, DC	31

a. Calculate the mean and median of this distribution.

b. Compare the mean and median. Which is the higher value? Why?

c. If you removed Los Angeles from this distribution and recalculated, what would happen to the mean? To the median? Why?

d. Report the mean and median as you would in a formal research report.

Data below represent the weight of obese patients visiting the Dietetics clinic.

Describe the sample of patients using the appropriate statistical computations And graphical presentations

192	110	195	180	170	215
152	120	170	130	130	125
135	185	120	155	101	194
110	165	185	220	180	
128	212	175	140	187	
180	119	203	157	148	
260	165	185	150	106	
170	210	123	172	180	
165	186	139	175	127	
150	100	106	133	124	