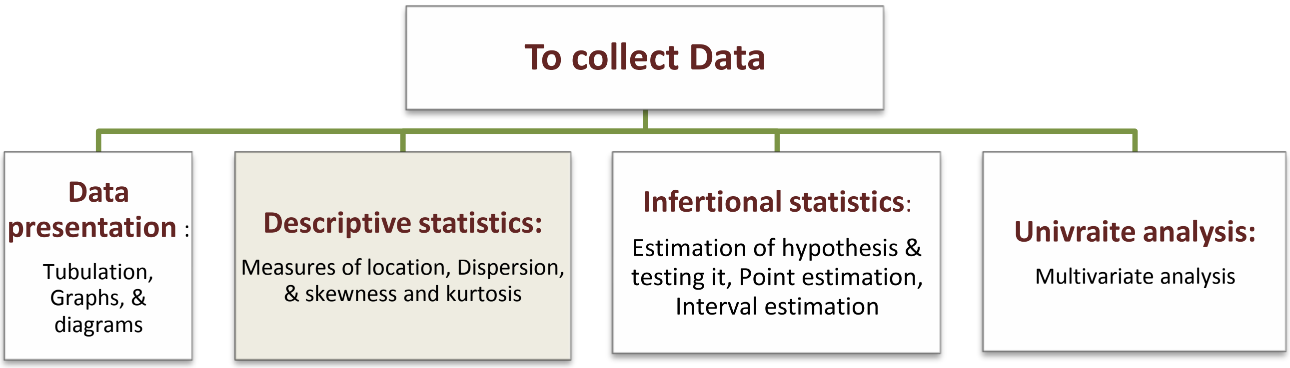
Measurement of Central Tendency

Descriptive Statistics:

The goal of descriptive statistics is to summarize a collection of data in a clear and understandable way



Central tendency:

Central tendency is a single summary score that best describes the central location of an entire distribution of scores; it represents the center of the distribution.

One distribution can have multiple locations where scores cluster, therefore for every given situation the best measurement must be decided

Measures of Central Tendency = Measures of Location = Measures of Averages

- ✓ Mean (The sum of all scores divided by the number of scores),
- ✓ Median (The value that divides the distribution in half when observations are ordered)
- ✓ Mode (The most frequent score)

FIRST
see the link

http://www.nutshellmath.com/textbooks_glossary_demos/glossary_content/mean,_median,_and_mode.html

1- Mean (Arithmetic Mean):

Sum of all the observations divided by the number of the observations

The arithmetic mean is the most common measure of the central location of a sample.

Mean is the balance point of a distribution.

In population: $\mu = \dfrac{\sum_{i=1}^{N} X_i}{N}$ In sample: $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

- Population

"Sigma" the sum of X, add up all scores

"mu" $\longrightarrow \mu = \dfrac{\Sigma X}{N}$

"N", the total number of scores in a **Population**

- Sample

"sigma", the sum of X, add up all scores

"X bar" $\longrightarrow \overline{X} = \dfrac{\Sigma X}{n}$

"n", the total number of scores in a **Sample**

Example of mean :

- Calculate the mean for the following data:

Ex.1 {1,3,6,7,2,3,5}

Number of observations: 7 = "n" = How Many Numbers between the Brackets

Sum of observations: 27 = ΣX = How Much is the sum of these numbers

Mean = 27 / 7 = 3.9 = By Applying the $\overline{X} = \dfrac{\Sigma X}{n}$ ➔ Mean= $\overline{X}$

Ex.2 Raw-score distribution

Name	X
Student1	20
Student2	23
Student3	15
Student4	21
Student5	15
Student6	21
Student7	15
Student8	20

Frequency distribution

f	X
3	15
2	20
2	21
1	23

In other words:

Just to make it simple

$$\overline{X} = \frac{\sum fX}{n}$$

$$\overline{X} = \frac{15+15+15+20+20+21+21+23}{8}$$

$$\overline{\chi} = \frac{\sum f\chi}{N}$$

$$\text{Mean} = \frac{[(3\times15)+(2\times20)+(2\times21)+(1\times23)]}{8}$$

Mean = 18.75

Ex.3

52, 76, 100, 136, 186, 196, 205, 150, 257, 264, 264, 280, 282, 283, 303, 313, 317, 317, 325, 373, 384, 384, 400, 402, 417, 422, 472, 480, 643, 693, 732, 749, 750, 791, 891

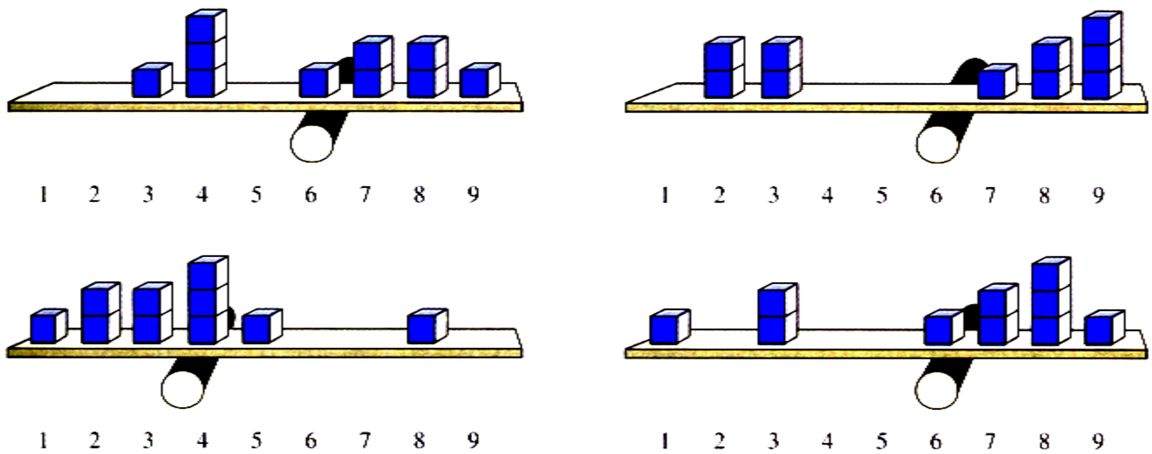
$$\overline{X} = \frac{\Sigma X}{n}$$

$$\overline{X} = \frac{13005}{35} = 371.60$$

Advantages (pros) and Disadvantages (cons) of the Mean:

- Pros
  - ✓ Mathematical center of a distribution.
  - ✓ Good for interval and ratio data.
  - ✓ Does not ignore any information.
  - ✓ Inferential statistics is based on mathematical properties of the mean.
- Cons
  - ✓ Influenced by extreme scores and skewed distributions.
  - ✓ May not exist in the data (look at example 3)

Mean : **Is the balance point of a distribution.**



2- Median :

Definition:

The value that is larger than half the population and smaller than half the population

Median is the middle number (odd # of data values)

and the average of the two middle numbers (even # of data values)

when the numbers are written from least to greatest.

If ***n is odd*** → 5 , 8 , $\boxed{9}$, 10 , 28 Median = 9

If ***n is even*** → 6 , 17 , $\boxed{19 ,\ 20}$, 21 , 27 Median = $\frac{19+20}{2}$ = 19.5

Pros and Cons of Median:

- Pros
  - ✓ Not influenced by extreme scores or skewed distributions.
  - ✓ Good with ordinal data.
  - ✓ Easier to compute than the mean.

- Cons
  - ✓ May not exist in the data.
  - ✓ Doesn't take actual values into account.

3- Mode:

The most frequently occurring value

Data	{1,3,7,3,2,3,6,7}	Mode : 3
Data	{1,3,7,3,2,3,6,7,1,1}	Mode : 1,3
Data	{1,3,7,0,2,-3, 6,5,-1}	Mode : none

Example of Mode

52, 76, 100, 136, 186, 196, 205, 150, 257, 264, 264, 280, 282, 283, 303, 313, 317,317, 325, 373, 384, 384, 400, 402, 417, 422, 472, 480, 643, 693, 732, 749, 750, 791, 891

Mode: most frequent observation

Mode(s) for hotel rates: 264, 317, 384

Pros and Cons of the Mode:

- Pros
  - ✓ Good for nominal & ordinal data.
  - ✓ Easiest to compute and understand.
  - ✓ The score comes from the data set.
- Cons
  - ✓ Ignores most of the information in a distribution.
  - ✓ Small samples may not have a mode.

Example of central tendency:

Suppose the age in years of the first 10 subjects enrolled in your study are:

34, 24, 56, 52, 21, 44, 64, 34, 42, 46

- Then the mean age of this group is 41.7 years

- To find the median, first order the data:

21, 24, 34, 34, 42, 44, 46, 52, 56, 64

The median is $\frac{42+44}{2}$ = 43 years

- The mode is 34 years.

Comparison of Mean and Median

- Mean is sensitive to a few very large (or small) values "outliers" so sometime mean does not reflect the quantity desired.
- Median is "resistant" to outliers
- Mean is attractive mathematically

Example: Regarding the Previous example

Suppose the next patient enrolls and their age is 97 years. How does the mean and median change?

Mean= 46.7

To get the median, order the data:

21, 24, 34, 34, 42, 44, 46, 52, 56, 64, 97

Median= 44

If the age were recorded incorrectly as 977 instead of 97,

What would the new median be? Median= 44

What would the new mean be? Mean= 126.7

Example.2: Calculating the Mean from a Frequency Distribution

# of Children(Y)	Frequency(f)	Frequency*Y (fY)
0	12	0
1	25	25
2	733	1466
3	333	999
4	183	732
5	26	130
6	15	90
7	12	84
Total	1339	3526

$$\overline{Y} = \frac{\sum fY}{N} = \frac{3526}{1339} = 2.6$$

Example.3

Table 6.1 Calculation of Arithmetic Mean for a Series of Serum Albumin Levels (g%) of 24 Pre-School Children

2.90	3.75	3.66
3.57	3.45	3.76
3.73	3.71	3.43
3.55	3.84	3.69
3.72	3.30	3.77
3.88	3.62	3.43
2.98	3.76	3.68
3.61	3.38	3.76

The total of all these values, i.e. $\Sigma x = 85.93$.

Total number of observations $(n) = 24$

Therefore the arithmetic mean, $\bar{x} = \frac{\Sigma x}{n} = \frac{85.93}{24} = 3.58$ g%

Table 6.2 Calculation of Arithmetic Mean of Protein Intake of 400 Families

Protein intake/consumption unit/day (g)	No. of families	Midpoint of class interval	Multiply f & x
Class interval	f	x	fx
15–25	30	20	600
25–35	40	30	1200
35–45	100	40	4000
45–55	110	50	5500
55–65	80	60	4800
65–75	30	70	2100
75–85	10	80	800
Total	400		19000

Arithmetic mean:

$$\frac{\Sigma fx}{n} = \frac{30 \times 20 + 40 \times 30 + \ldots + 10 \times 80}{400}$$

$$= \frac{19000}{400} = 47.50 \text{ g}$$

Examples From A Book

Illustration. Calculation of median for the data given in Table 6.1.

Arranging all the 24 values in ascending order of magnitude, we get the following data:

2.90	3.57	3.73
2.98	3.61	3.75
3.30	3.62	3.76
3.38	3.66	3.76
3.43	3.68	3.76
3.43	3.69	3.77
3.45	3.71	3.84
3.55	3.72	3.88

The 12th value is 3.66 and 13th is 3.68; median is the average of these two.

$$\text{Median} = \frac{3.66 + 3.68}{2} = 3.67 \text{ g\%}$$

Table 6.3 Calculation of Median for the Data of Table 6.2

Protein intake/con-sumption unit/ day (g)	No. of families (frequency)	Cumulative frequency
15–25	30	30
25–35	40	70
35–45	100	170
45–55	110	280
55–65	80	360
65–75	30	390
75–85	10	400
Total	400	

Median class is 45–55 $\quad n = 400$

$$\text{Median} = L + \frac{(n/2 - F) \times C}{f}$$

$$= 45 + \frac{(200 - 170) \times 10}{110} = 45 + 2.73 = 47.73 \text{ g}$$

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

or using the formula

$$\text{Mode} = L_M + \frac{d_1 C}{d_1 + d_2}$$

where L_M = Lower limit of modal class

d_1 = frequency in modal class *minus* frequency in the preceding class.

d_2 = frequency in modal class *minus* frequency in the succeeding class

C = class interval of modal class

For the data of Table 6.2,

$$\text{Mode} = 45 + \frac{10 \times 10}{10 + 30} = 47.5$$

Geometric Mean & Harmonic Mean:

GEOMETRIC MEAN (GM)

The arithmatic mean (AM) is not the appropriate measure of central tendency if the distribution of the observations is not symmetrical (e.g., when the values change exponentially). In such situations, another measure called Geometric Mean (GM) is useful to summarise the data. By definition, the GM of N observations is the N^{th} root of the product of the observations. In symbols,

$$GM = \sqrt[N]{X_1\ X_2\ X_3\ \ldots\ldots\ldots\ X_N}$$

The calculation of GM is made simple using logarithms. Thus,

$$\log GM = \frac{1}{N}[\log X_1 + \log X_2 + \ldots + \log X_N] = \frac{\Sigma \log Xi}{N}$$

As an example, let us calculate the GM of the following five antibody titres: 4,8,16,32,64. The logarithms of these values are 0.60, 0.90, 1.20, 1.51 and 1.81, respectively.

$$\therefore \log GM = \frac{0.60 + 0.90 + 1.20 + 1.51 + 1.81}{5} = \frac{6.02}{5}$$

$$= 1.20$$

$$\therefore \quad GM = \text{Antilog of } 1.20$$

$$= 15.85$$

(Note: The AM of the five values is 24.8)

The GM has some limitations. Even if one observation is negative, the GM cannot be calculated. If any of the observations is zero, the GM will be zero.

The Shape of Distributions

Distributions can be either <u>symmetrical</u> or <u>skewed</u>, depending on whether there are more frequencies at one end of the distribution than the other.

<u>Symmetrical Distributions:</u>

A distribution is symmetrical if the frequencies at the right and left tails of the distribution are identical, so that if it is divided into two halves, each will be the mirror image of the other.

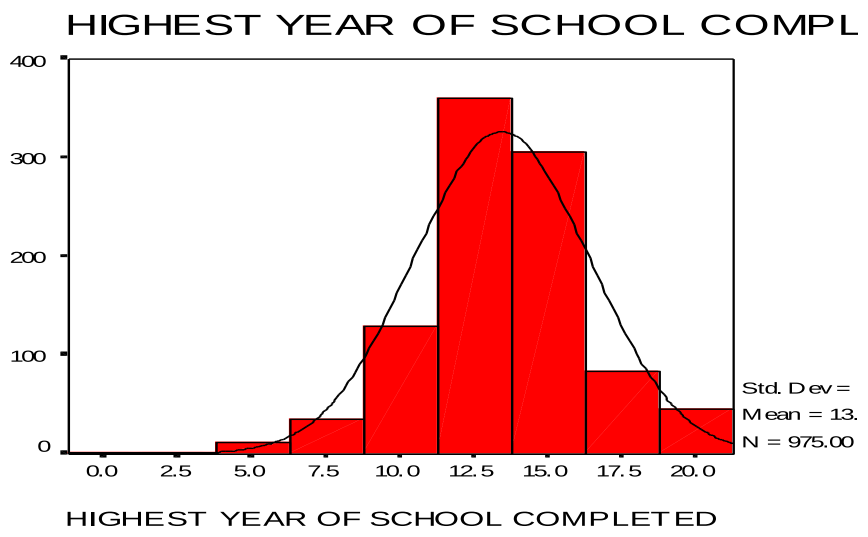
In a symmetrical distribution the mean, median, and mode are identical.

Almost Symmetrical distribution

(bell shaped)

Mean = 13.4

Mode = 13.0



Skewed Distribution:

a distribution is skewed if Few extreme values on one side of the distribution or on the other.

- **Positively skewed distributions:**

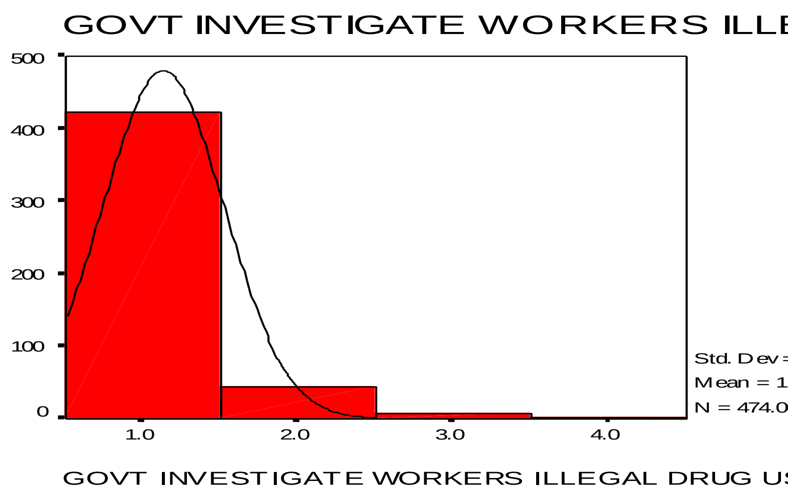
  Distributions which have few extremely high values

  (Mean > Median)

  Tails to the left

  Mean = 1.13

  Median = 1.0



- **Negatively skewed distributions:**

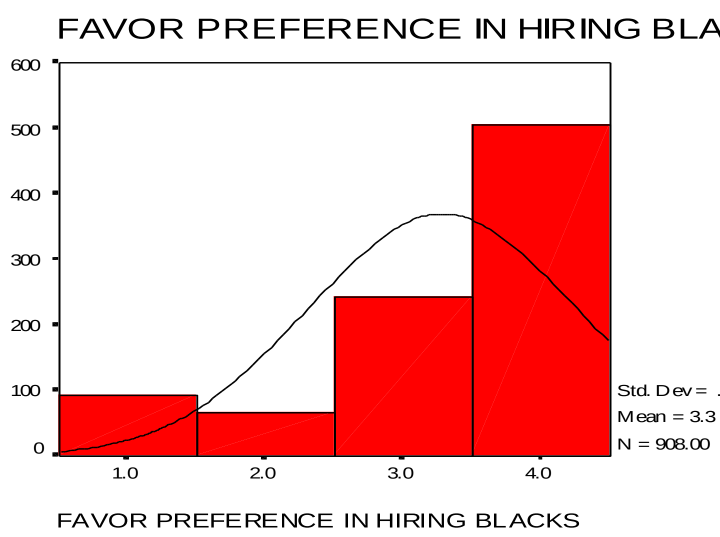
  Distributions which have few extremely low values

  (Mean < Median)

  Tails to the right

  Mean = 3.3

  Median = 4.0



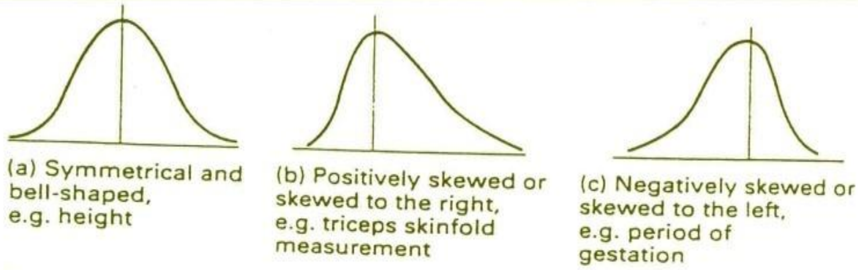
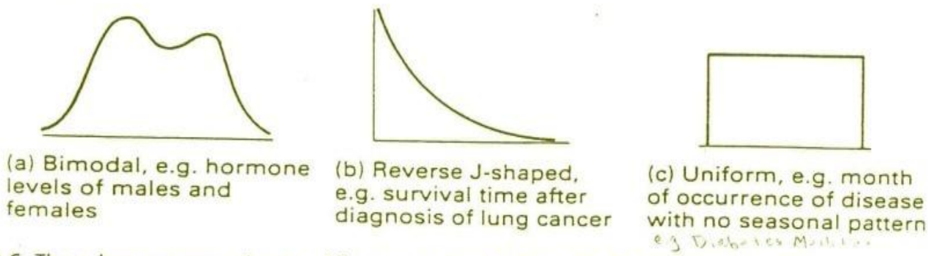


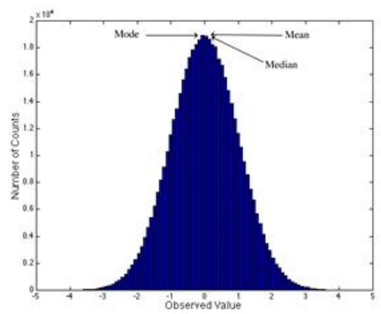
Fig. 3.5 Three common shapes of frequency distributions with an example of each.



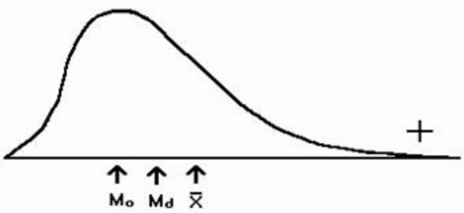
3.6 Three less-common shapes of frequency distributions with an example of each.

Distributions

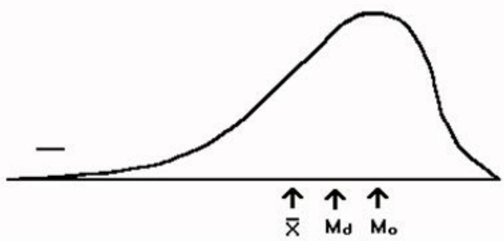
- Bell-Shaped (also known as symmetric" or "normal")



- Skewed:
  - positively (skewed to the right) – it tails off toward larger values



  - negatively (skewed to the left) – it tails off toward smaller values



Choosing a Measure of Central Tendency according to the variable :

- IF variable is Nominal → Mode

- IF variable is Ordinal → Mode or Median (Or both)

- IF variable is Interval-Ratio and distribution is Symmetrical
  → Mode, Median, or Mean

- IF variable is Interval-Ratio and distribution is Skewed
  → Mode or median