بسم الله الرحمن الرحيم

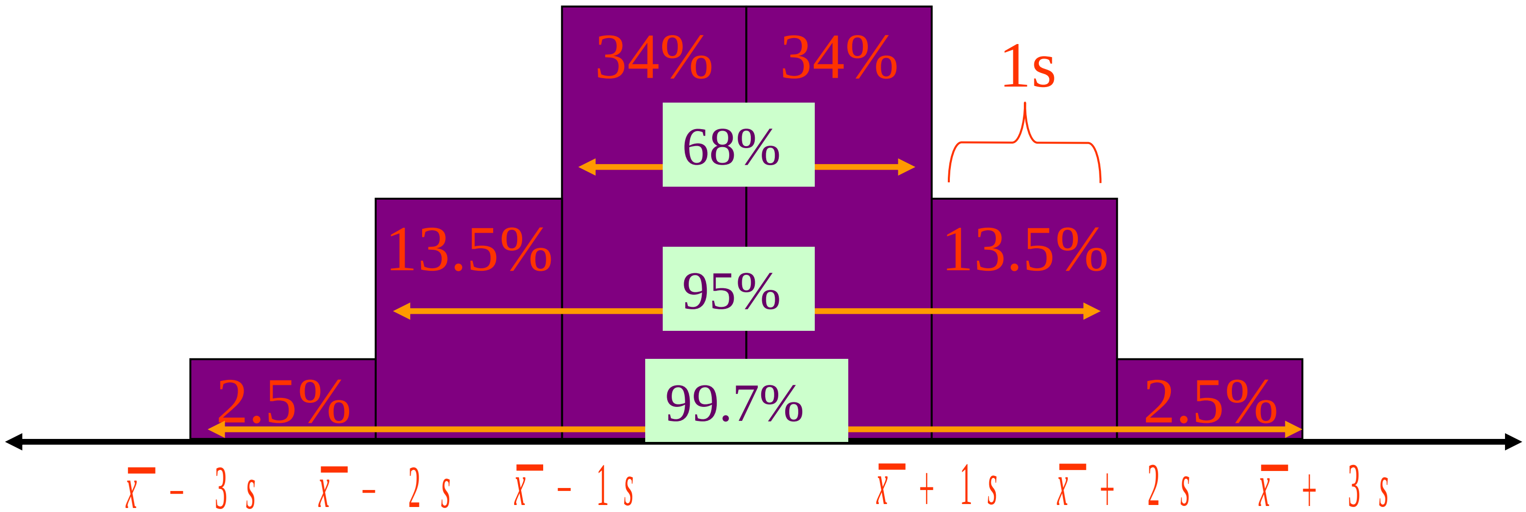
Biostatistics

L VI

PROF. DR. WAQAR AL-KUBAISY

17-7-2023

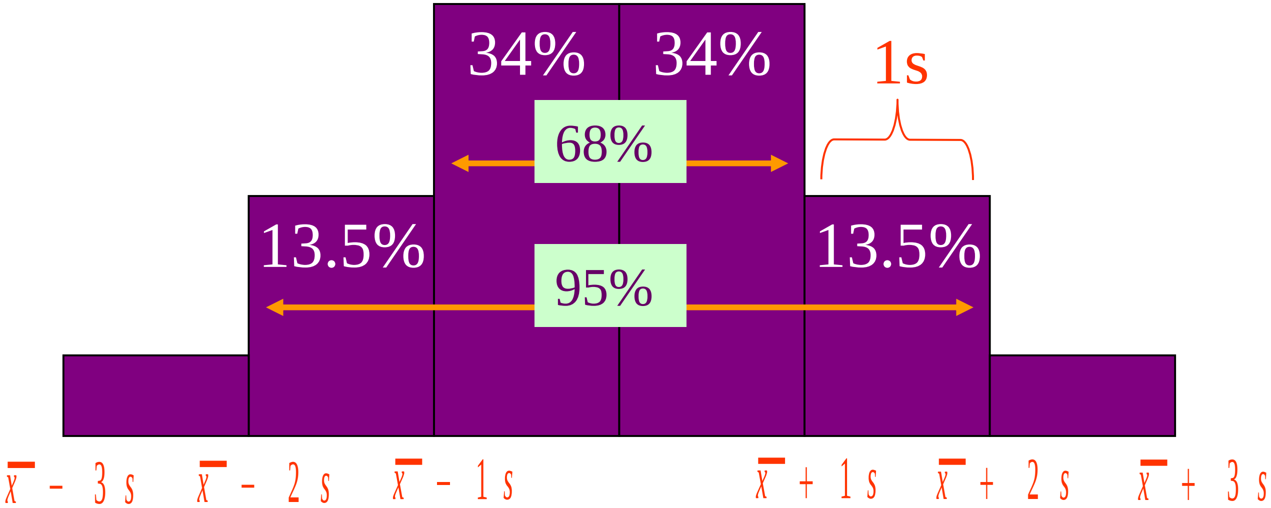
Interpreting Standard Deviation



For bell-shaped shaped distributions, the following statements hold:

- Approximately 68% of the data fall between $\bar{x} - 1s$ and $\bar{x} + 1s$
- Approximately 95% of the data fall between $\bar{x} - 2s$ and $\bar{x} + 2s$
- Approximately 99.7% of the data fall between $\bar{x} - 3s$ and $\bar{x} + 3s$

For NORMAL distributions, the word 'approximately' may be removed from The above statements.



Example: Suppose the Hb levels of 150 women has a roughly bell-shaped distribution with a mean of 12 mg/dl. and standard deviation of 0.10 g/dl.

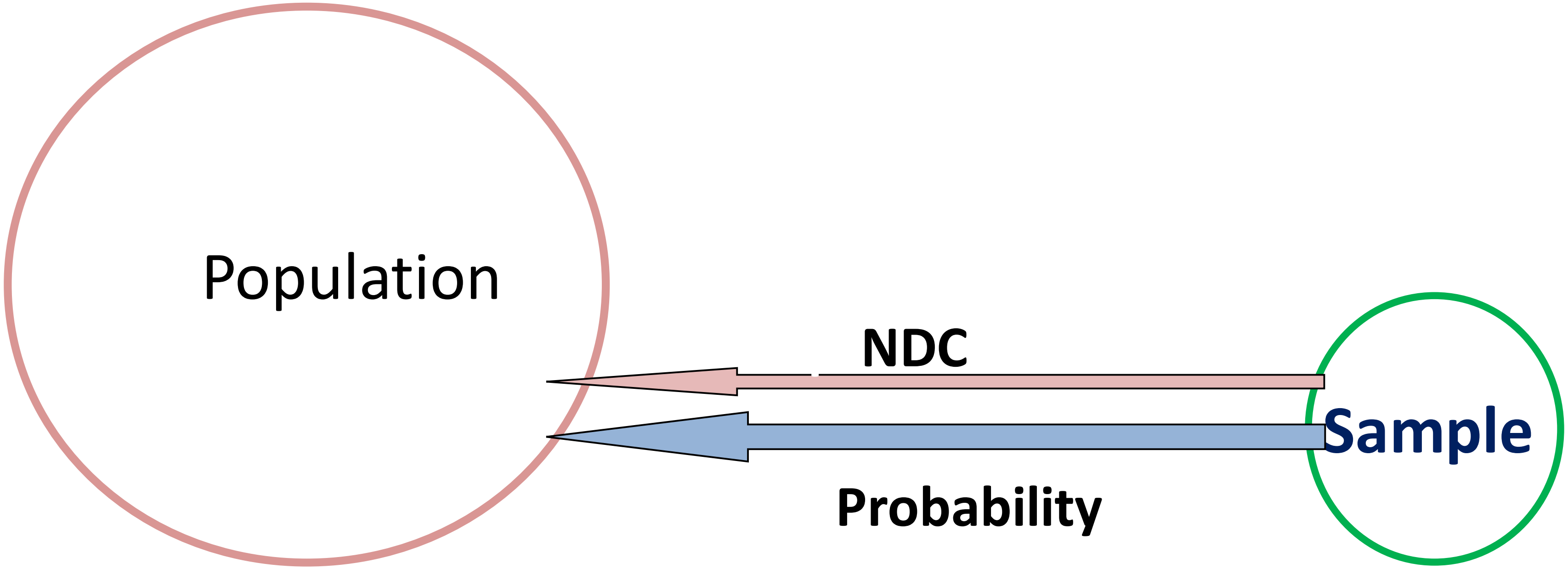
a) Give the interval of the amount of Hb level that approximately 68% of the women will have

12-0.1 to 12+0.1 = 11.9 to 12.1g/dl.

b) Give the interval of the amount of Hb level that approximately 95% of the women will have

12-2(0.1) to 12+2(0.1) = 11.8 to 12.2g/dl.

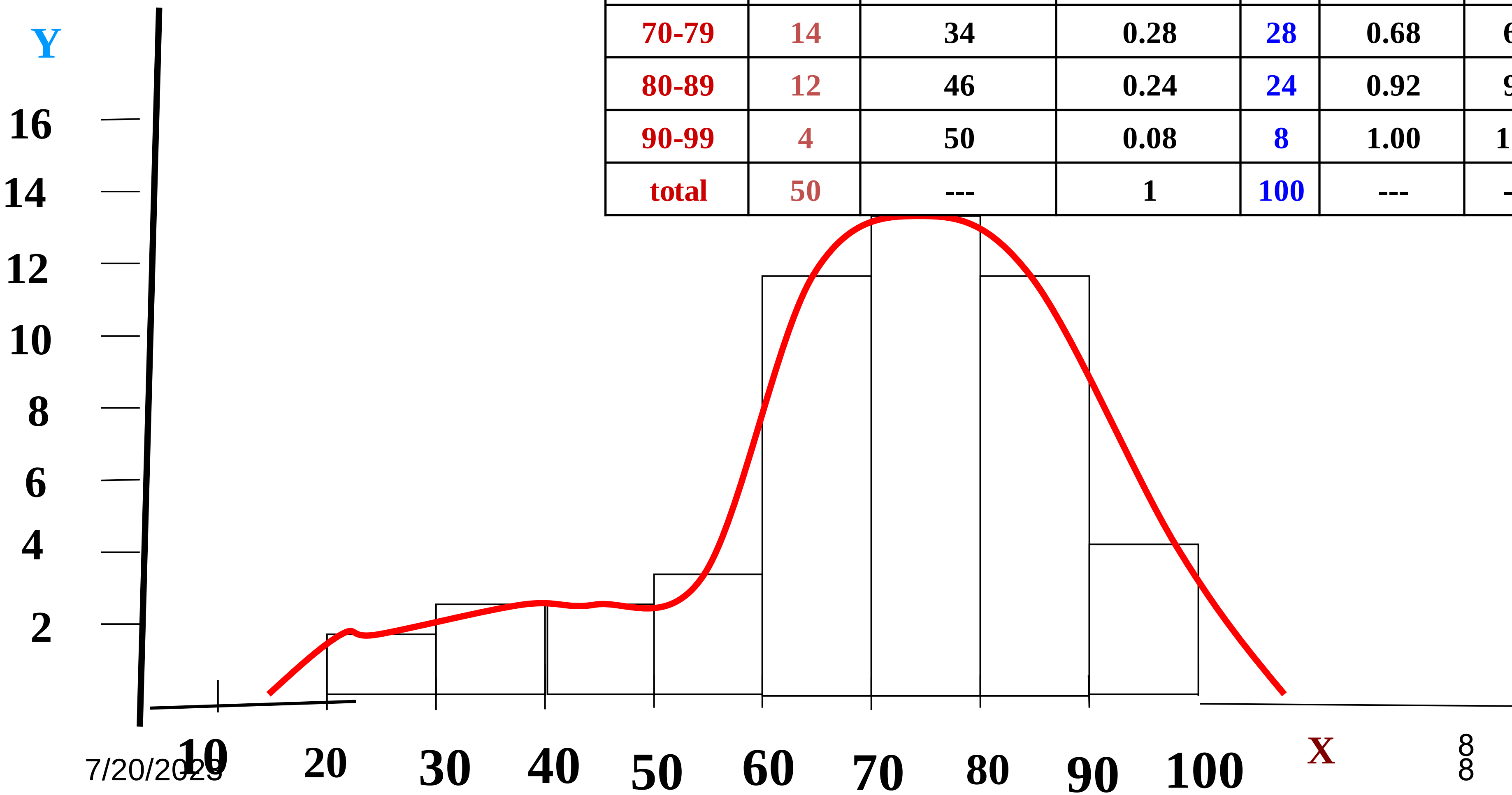
Important or Uses of SD



a sound generalized information about the population from which the sample has been drown, depending on evidence of this sample

Normal Distribution Curve

Age (year)	Freq.	Commutative frequency	Relative frequency	% R.F.	Cumulative R.F.	% cum. Freq.
20-29	1	1	0.02	2	0.02	2
30-39	2	3	0.04	4	0.06	6
40-49	2	5	0.04	4	0.1	10
50-59	3	8	0.06	6	0.16	16
60-69	12	20	0.24	24	0.4	40
70-79	14	34	0.28	28	0.68	68
80-89	12	46	0.24	24	0.92	92
90-99	4	50	0.08	8	1.00	100
total	50	---	1	100	---	---

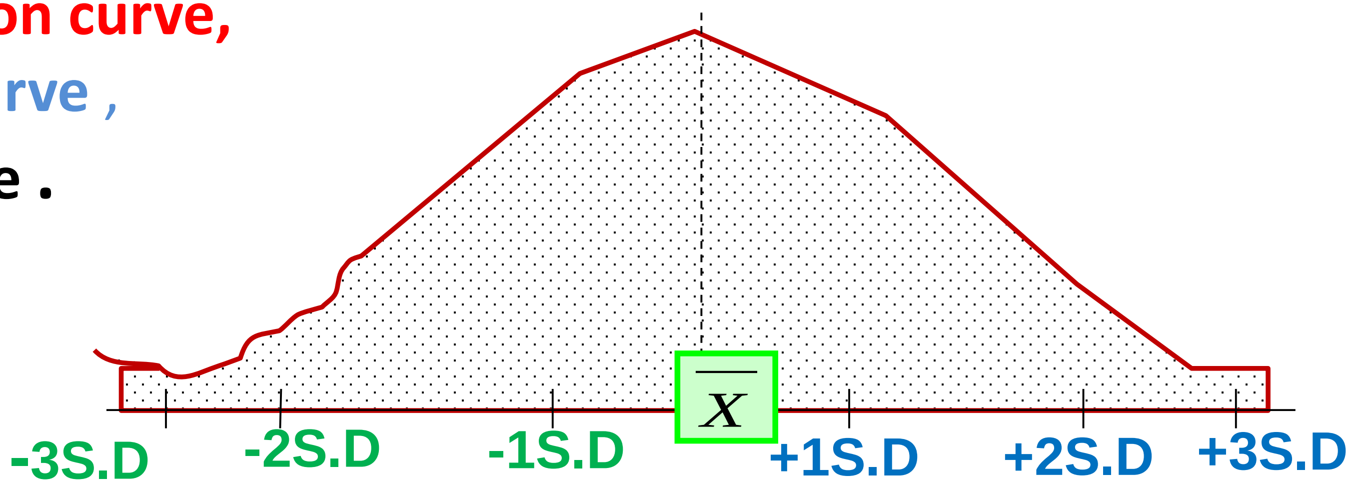


In large population → Graphically → Form of **Curve**

Normal Distribution curve,

Gaussian Curve ,

Bell Curve .



In NDC

- **All the observation are lying in area under the curve**
- **Average measures (mean Md , Mo) in the center of in the center of observation .**
- **Rest of observations distributed around the average measures .**
- **in a homogenous form**
- ***Half of them higher than the mean*** $\overline{X}$
- **Half of them lesser than the mean** $\overline{X}$

So

- **the distribution of observation in NDC is symmetrical .**

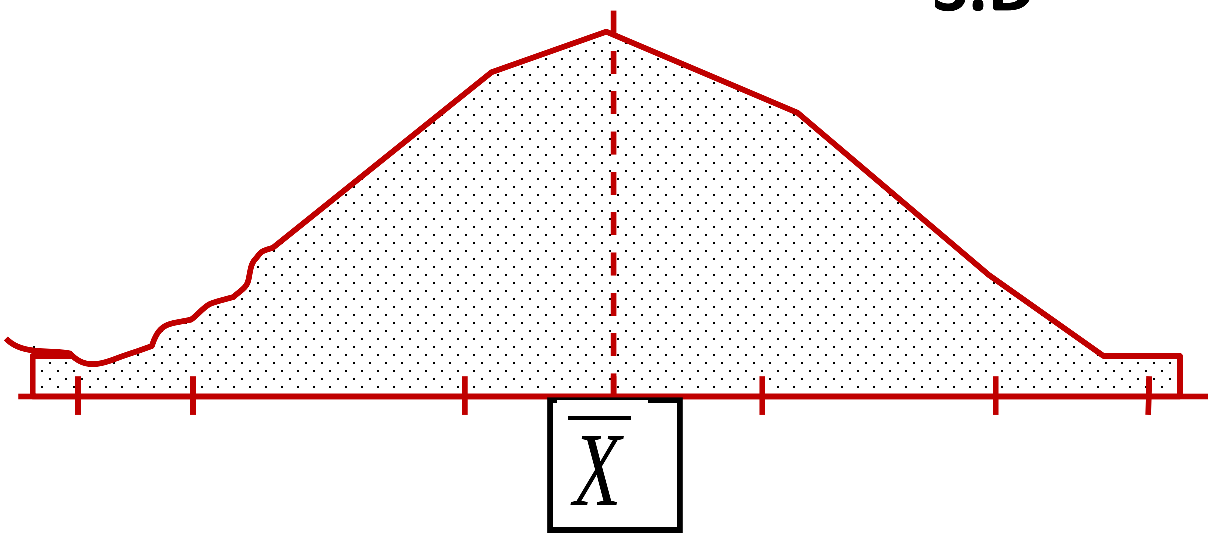
under the NDC divided by

1- **measures of C.T**

$\bar{X}$

2-**measures of variability**

S.D



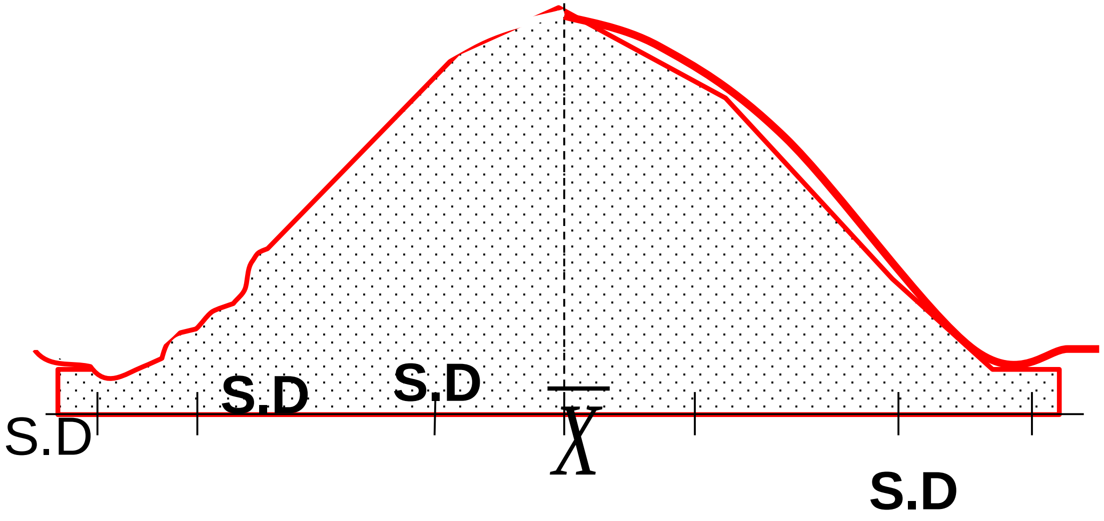
10

By Measures of C.T $\overline{X}$

Divided the area under the curve into two equal halves of observation,

50 % **of observation their values less than** $\overline{X}$ **value**

and **50 %** of observation **their values higher than** $\overline{X}$



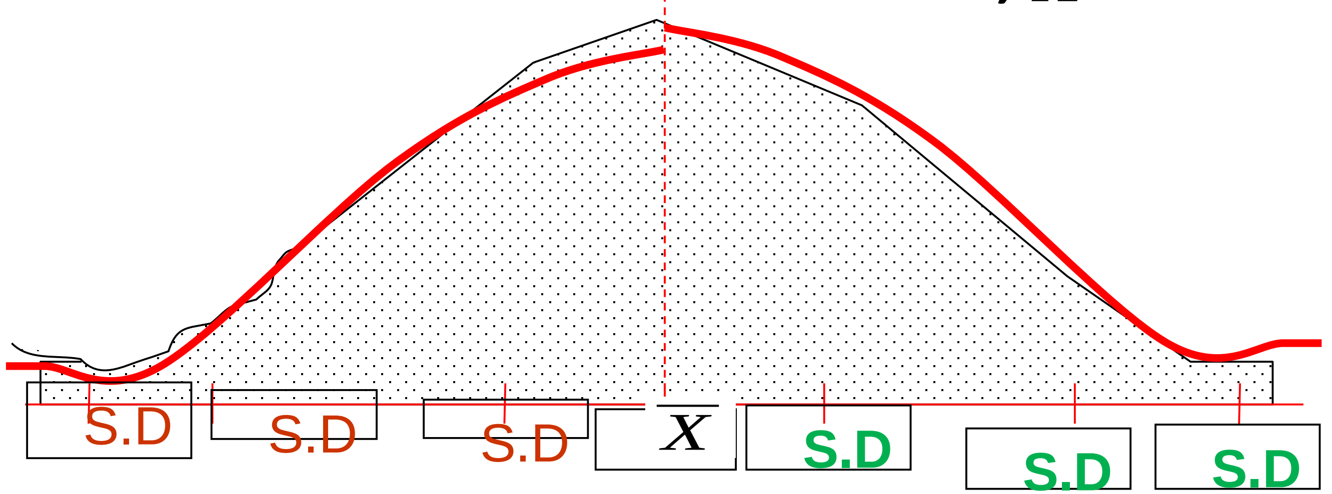
By Measures of variability (S.D)

S.D and its multiplicity (one S.D, two S.D, three S.D divided the area under the NDC into small areas, each area containing certain and fixed proportion of observation

Within $\pm$1 S.D from the $\overline{X}$

Within $\pm$2 S.D from the $\overline{X}$

Within $\pm$3 S.D from the , $\overline{X}$



Within ±1 S.D from the X

68% of observations,(34%o each side)

68% of observation deviated from the $\overline{X}$ **by not more than ±1 S.D** **???????**

Within ±2 S.D from the $\overline{X}$

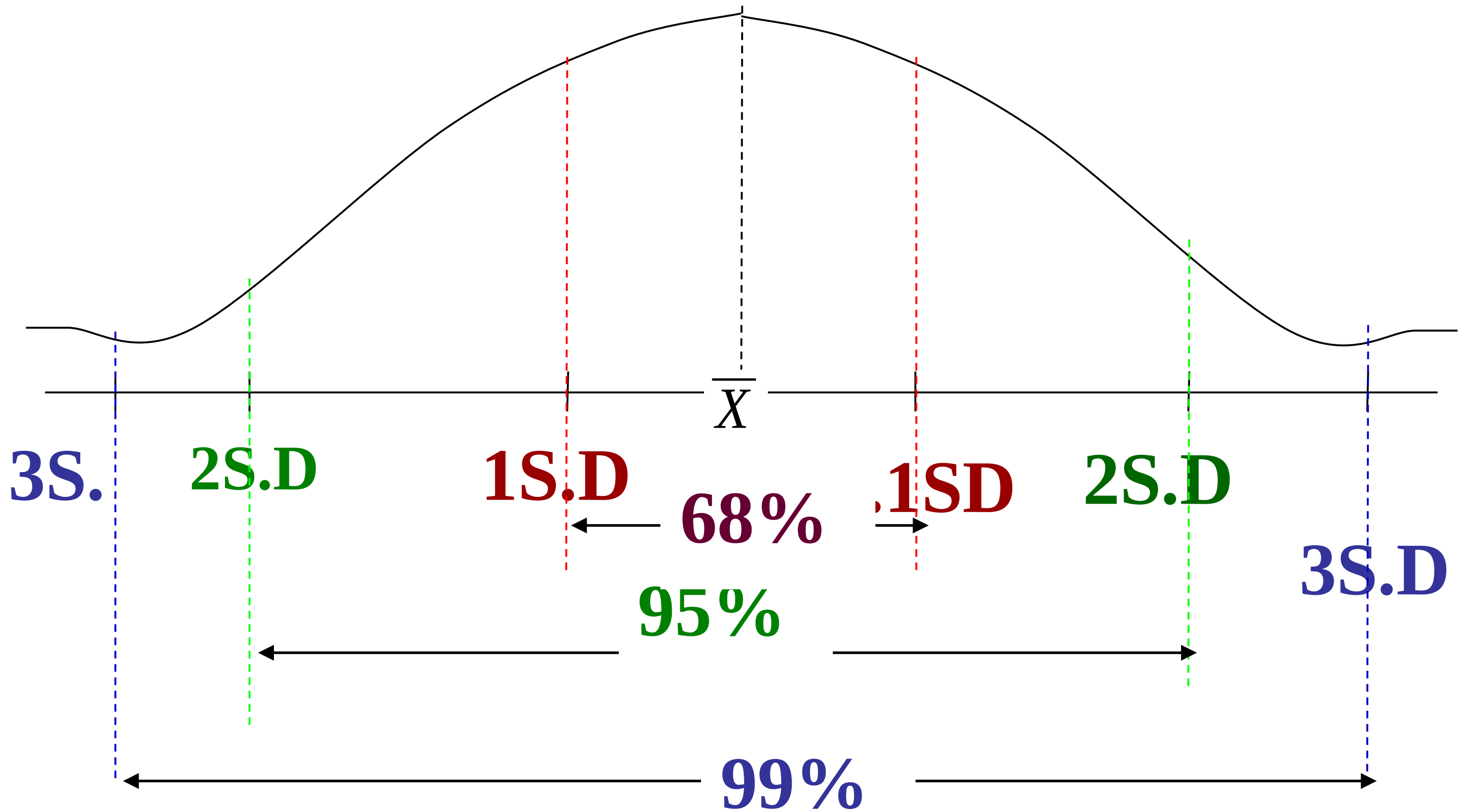
95% of observations lie,

95% of observations deviated from the $\overline{X}$ by not more than ±2 S.D . **???????**

Within ±3 S.D from the $\overline{X}$

99% of observations are located,

99% of observations deviated from the $\overline{X}$ by **not more than ±3 S.D .** **???????**



$\bar{X}$
3S.
2S.D
1S.D
.1SD
2S.D
3S.D
68%
95%
99%

??????????

Characteristics of the NDC

1.Bell shape .

2.Symmetrical distribution of observations on both sides

3.Unimodal ??????????.

4.Curving downward on both sides from the mean toward the horizontal, but never touch it .

5.Mean, Median and Mode of distribution are identical or coincide .

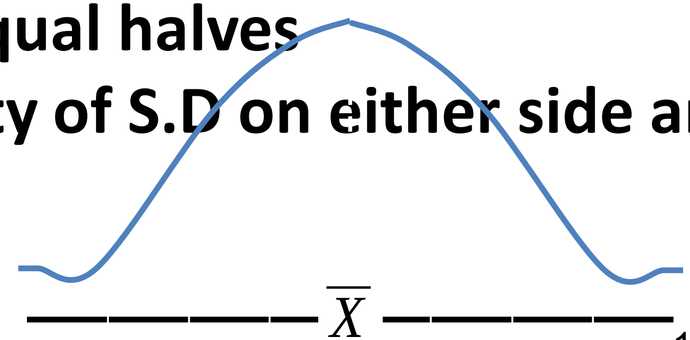
6.All the Medical, Biological phenomenon following its distribution .

7-Area under curve divided by

Mean into two equal halves

8. Between $\overline{X}$ and certain multiplicity of S.D on either side an area containing fixed proportion of observation

68% 99% 95% .



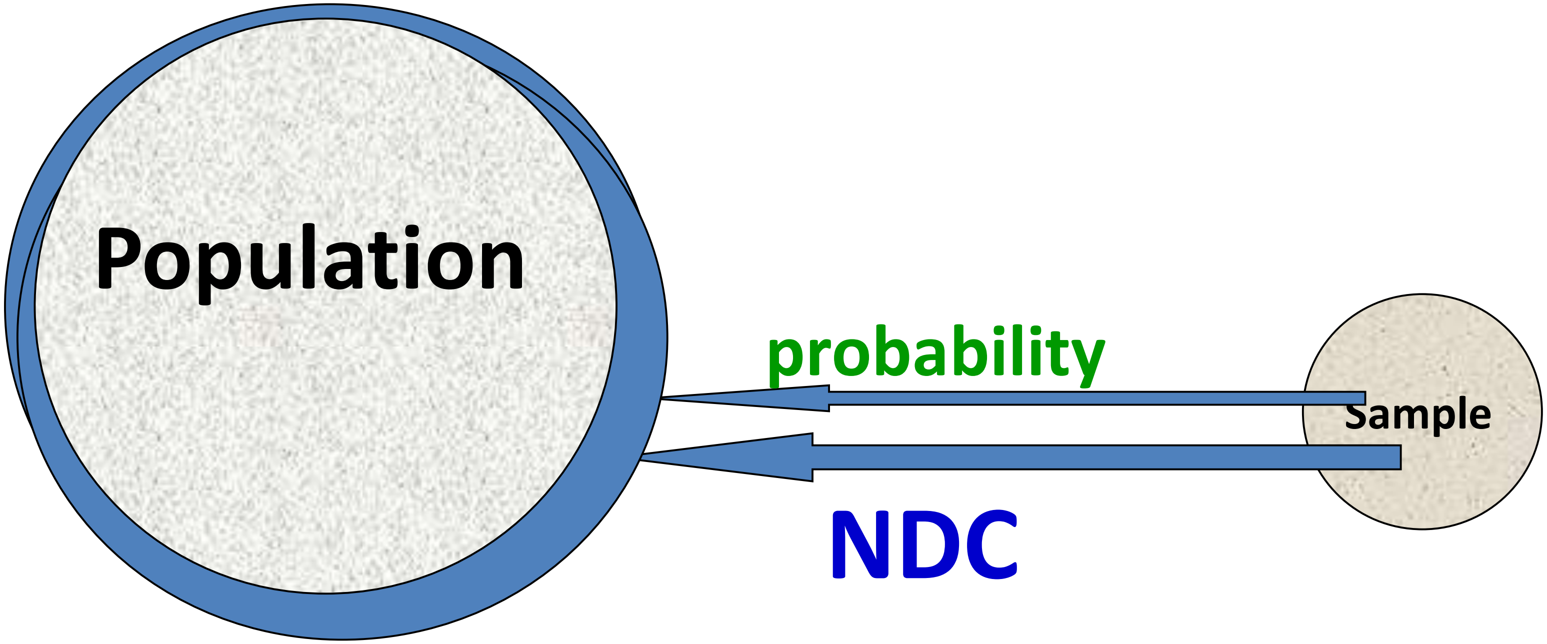
Importance

1-Most of the phenomenon in Medical field follow this distribution .

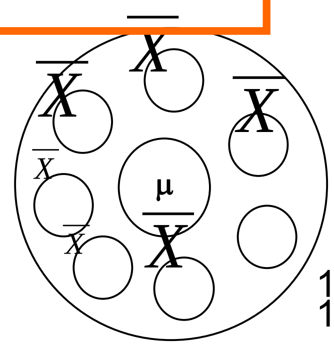
2-It is for justification and calculation of confidence interval .

3-It is form the basis of most of significance testing hypothesis . That is most test of significance depend on the theory of ND

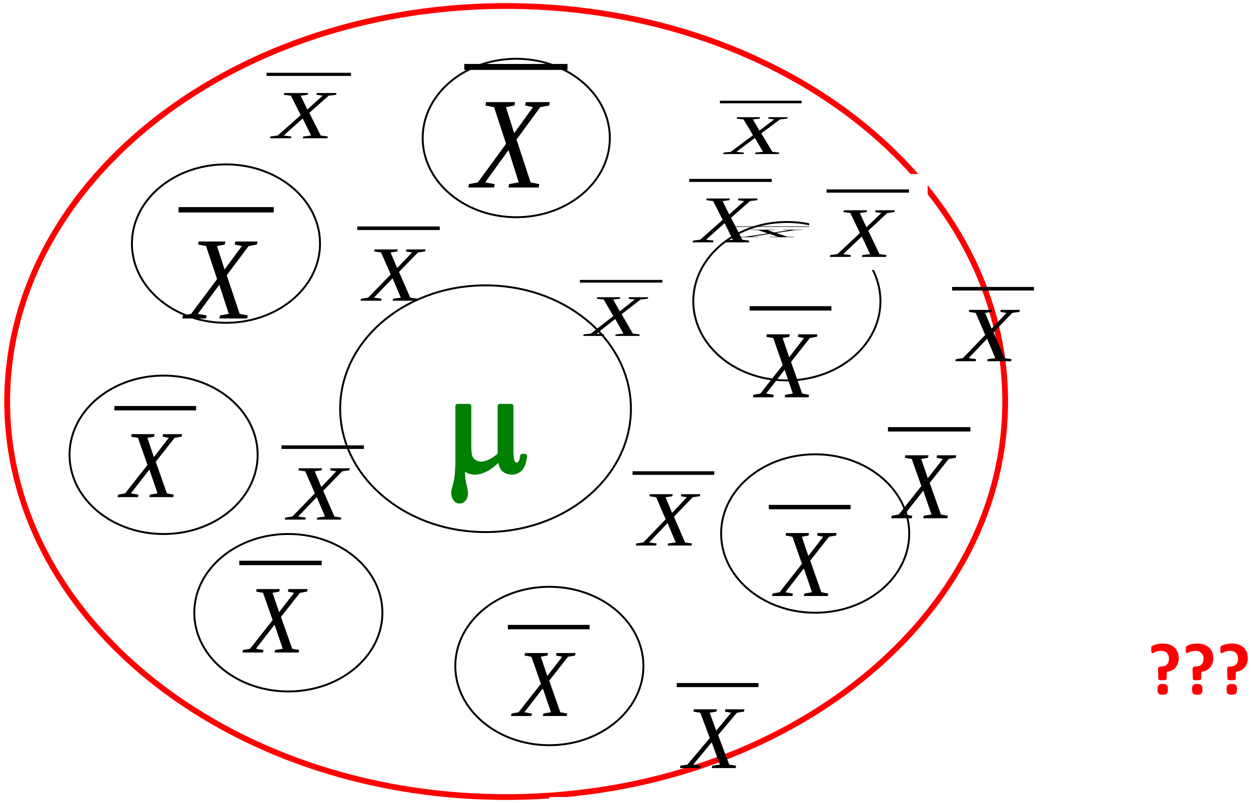
16



a sound generalized information about the population from which the sample has been drown, depending on evidence of this sample



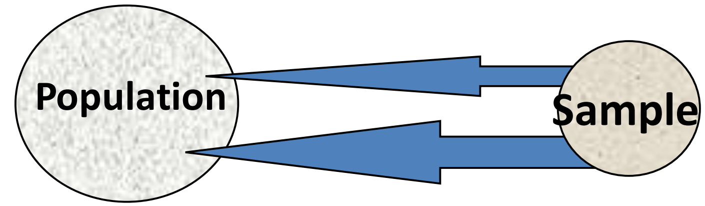
17



Different samples $\rightarrow$ different $\overline{X}_S$ even if the samples size are equal

There is a variation in the $\overline{X}s$ of different samples
This variation is due to sampling variation.

Sampling Variability



Mean $\pm$ S.D of sample . $\overline{X}$

The interest of sample not in its own right but what it tell us about the population which this sample represent

The aim of Biostatistics is to have

a sound generalized information about the population from which the sample has been drown, depending on evidence of this sample

sample

mean $\pm$ S.D

statistics values

Sample Statistics

true mean ,

Population Parameters

population mean

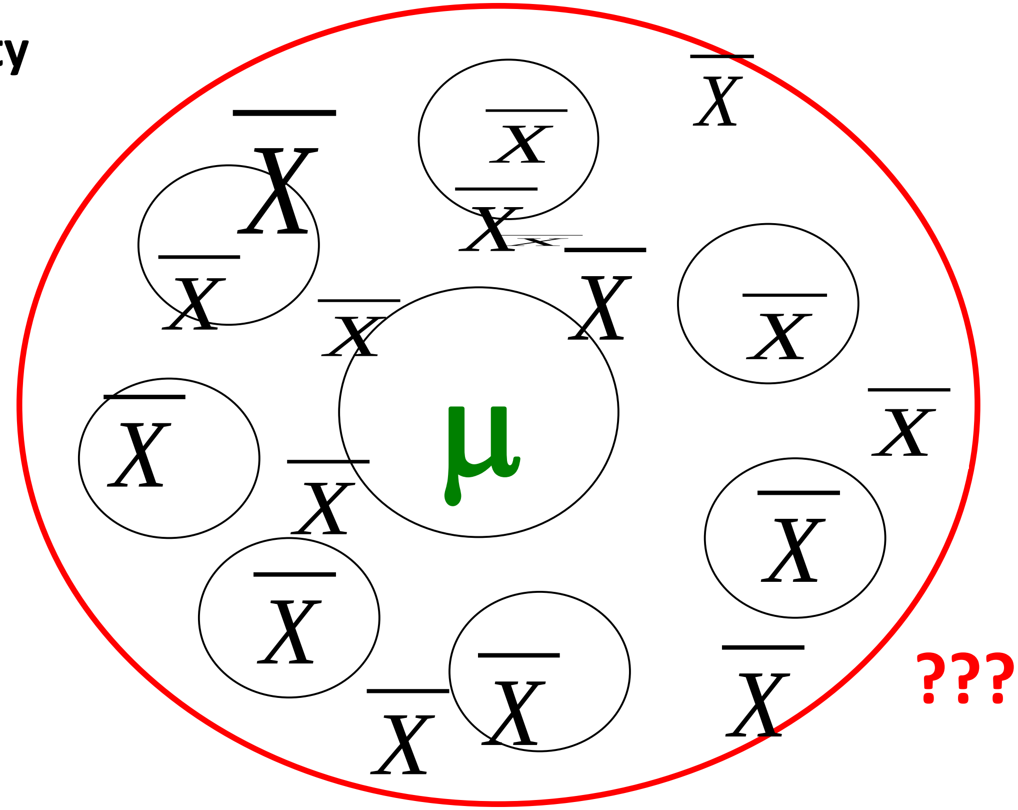
μ (mue)

S.D of population

σ (sigma

mean of universe

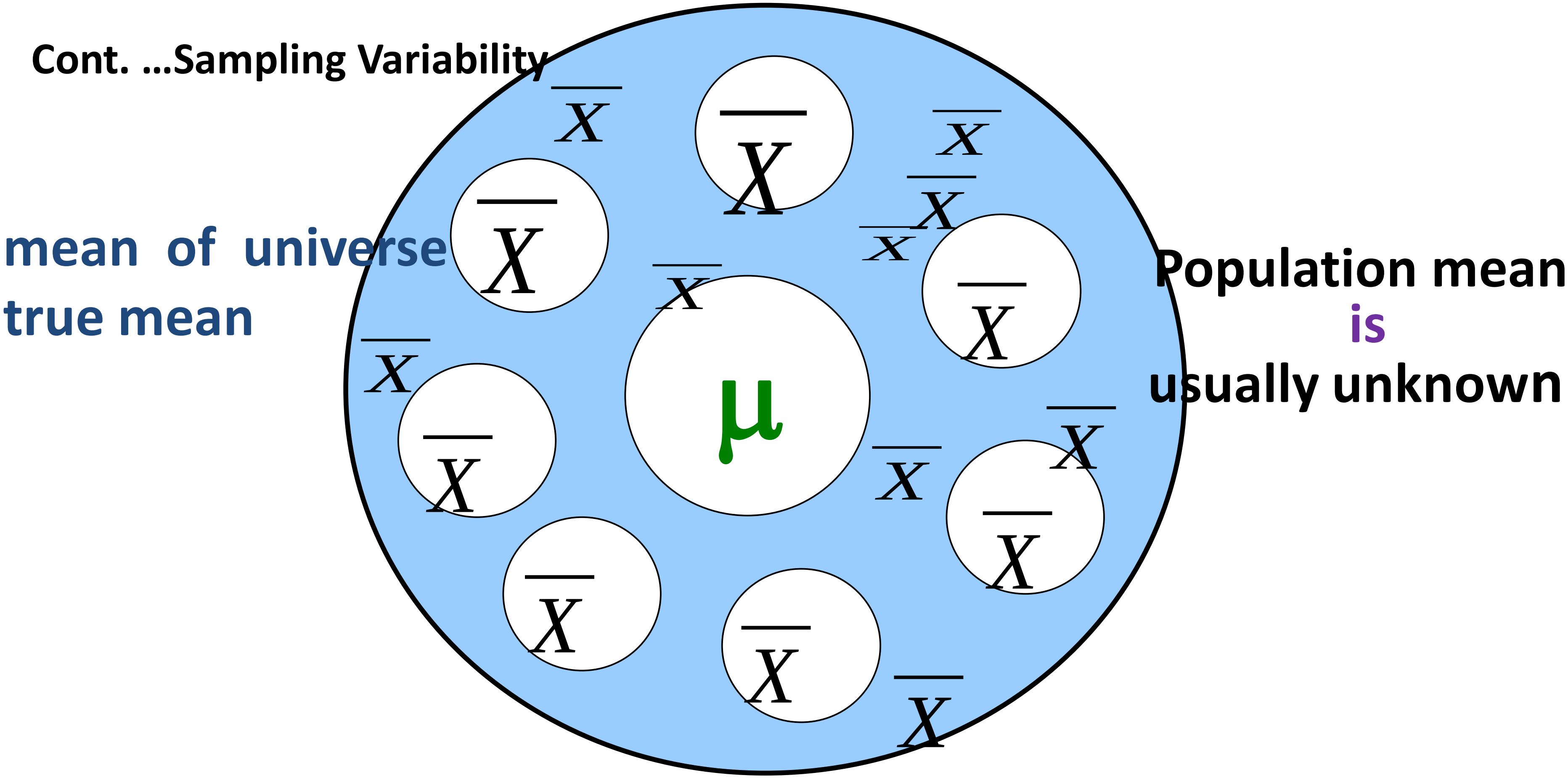
Cont. …Sampling Variability



Different samples $\rightarrow$ different $\overline{X}_S$ even if the samples size are equal

There is a variation in the $\overline{X}s$ of different samples
This variation is due to sampling variation.

Cont. …Sampling Variability

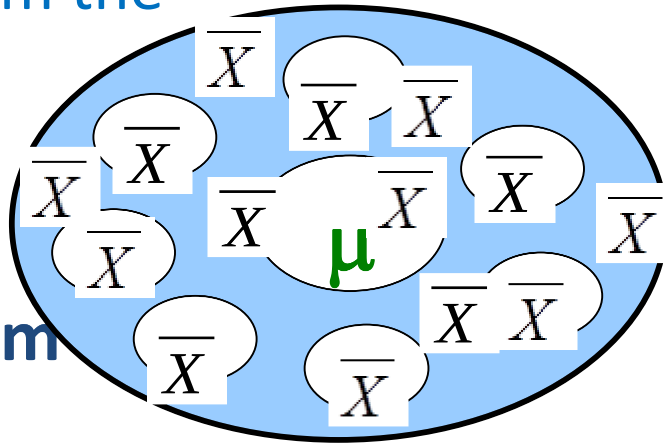


the sample measurement (mean± S.D) is not exactly reflect its population .

There is a difference between sample mean $\overline{X}$ and population mean μ

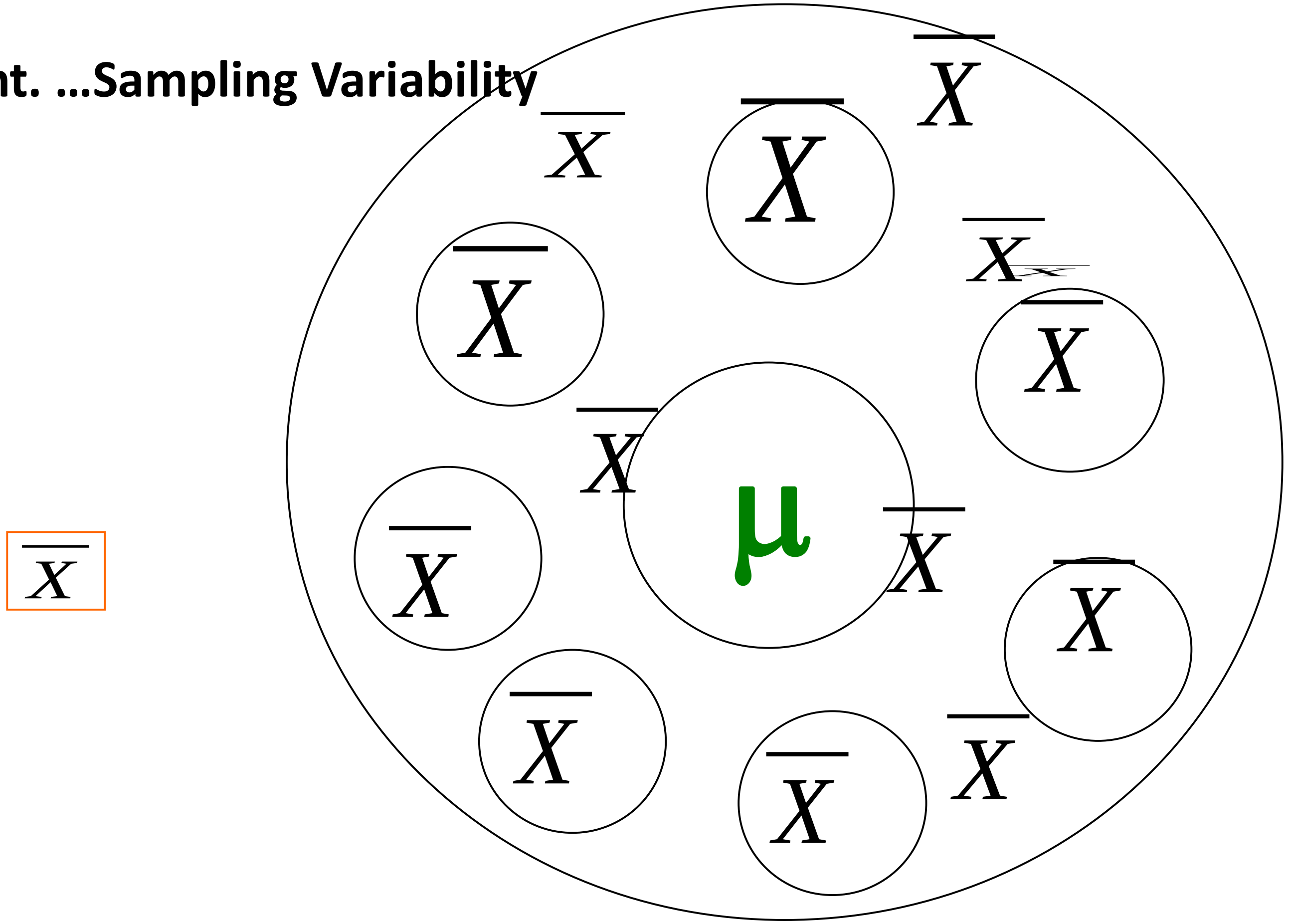
There is a difference between
sample statistics and population parameters,
this variation is called Sampling Error

There is a difference between sample means and population mean.????????



- **Deviation of the samples mean $\overline{X}$) from the population mean (μ)**
- ✓ **this will be the S.D of sample mean**
- ✓ **($\overline{X}$, from the population mean (μ)**
- ➢ **Average of S.D of sample means from population mean which is**
- **known as Standard Error**

Cont. …Sampling Variability



This mean that samples $\bar{X}_s$ distributed **around population mean,**

or Samples $\bar{X}$ s scatter around the μ .

The measurement of this scattering equal to

S.D of the sample

24
24

Standard Error S.E

- It is the average deviation of the sample mean ($\overline{X}$) from the true (population) mean (μ) of the population . *So*
- *it is equal to the S.D of sample mean* $\overline{X}$ *divided by the square root of* the *sample size (N)*

$$S.E = \frac{S.D}{\sqrt{N}}$$

depend on
- sample size
- S.D of sample

???????

The larger the sample size (N) $\rightarrow$ smaller the S.E

The smaller the S.D of sample $\rightarrow$ smaller the S.E

Standard Error S.E

Example

8 plasma values of uric acid

the mean ($\overline{X}$) **of uric acid is** **3±0.31**

$$S.E = \frac{0.31}{\sqrt{8}} = 0.11$$

$$\boxed{S.E = \frac{S.D}{\sqrt{N}}}$$

16 plasma values of uric acid the mean ($\overline{X}$)**of uric acid is 3±0.31**

$$= \frac{0.31}{\sqrt{16}} = 0.0775$$

$$\frac{0.21}{\sqrt{16}} = 0.0525$$

$$\frac{0.41}{\sqrt{16}} = 0.1025$$

- **Distribution of samples means ($\overline{X}_S$) around the population mean (μ) in NDC area**
  - is similar to that
  - of the distribution of X (values) around sample mean $\overline{X}$

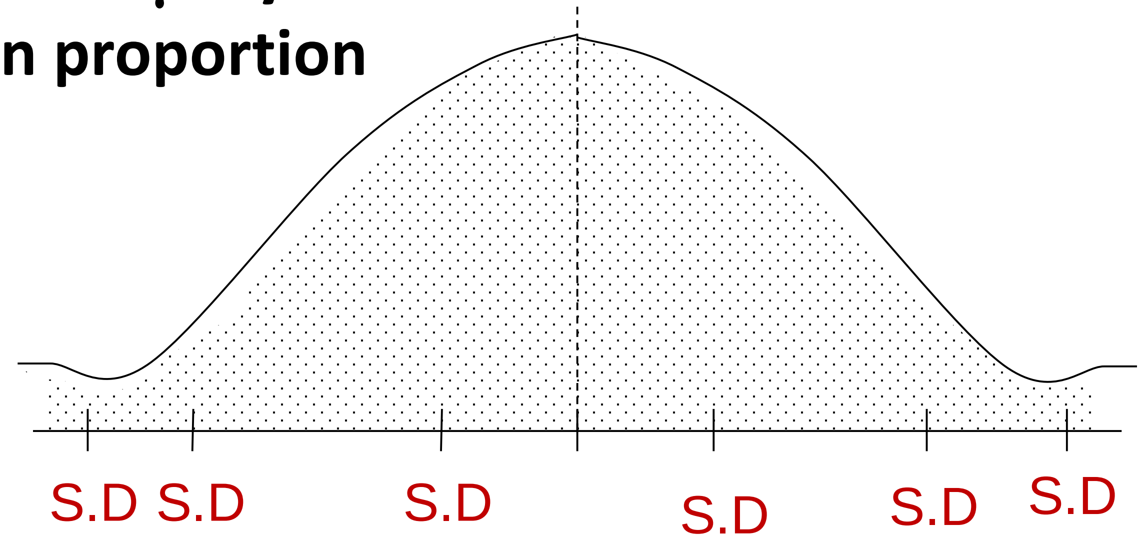
Sample means $\overline{X}_S$ deviated from μ by

✓ **S.E and its Multiplicity, so**

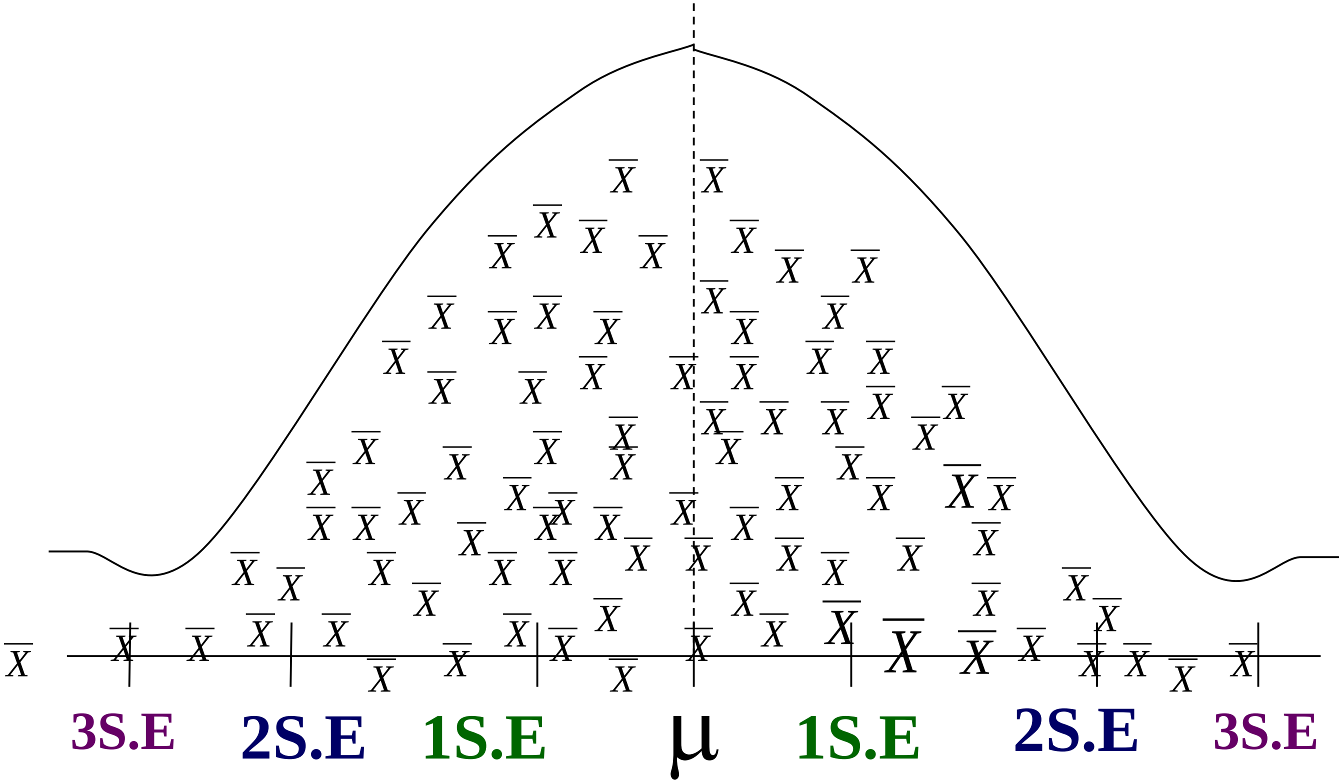
($\overline{X}$) deviated from μ by

1 S.E, 2 S.E and 3 S.E in proportion

68% 95% 99% .



27



1 S.E, 2 S.E and 3 S.E in proportion

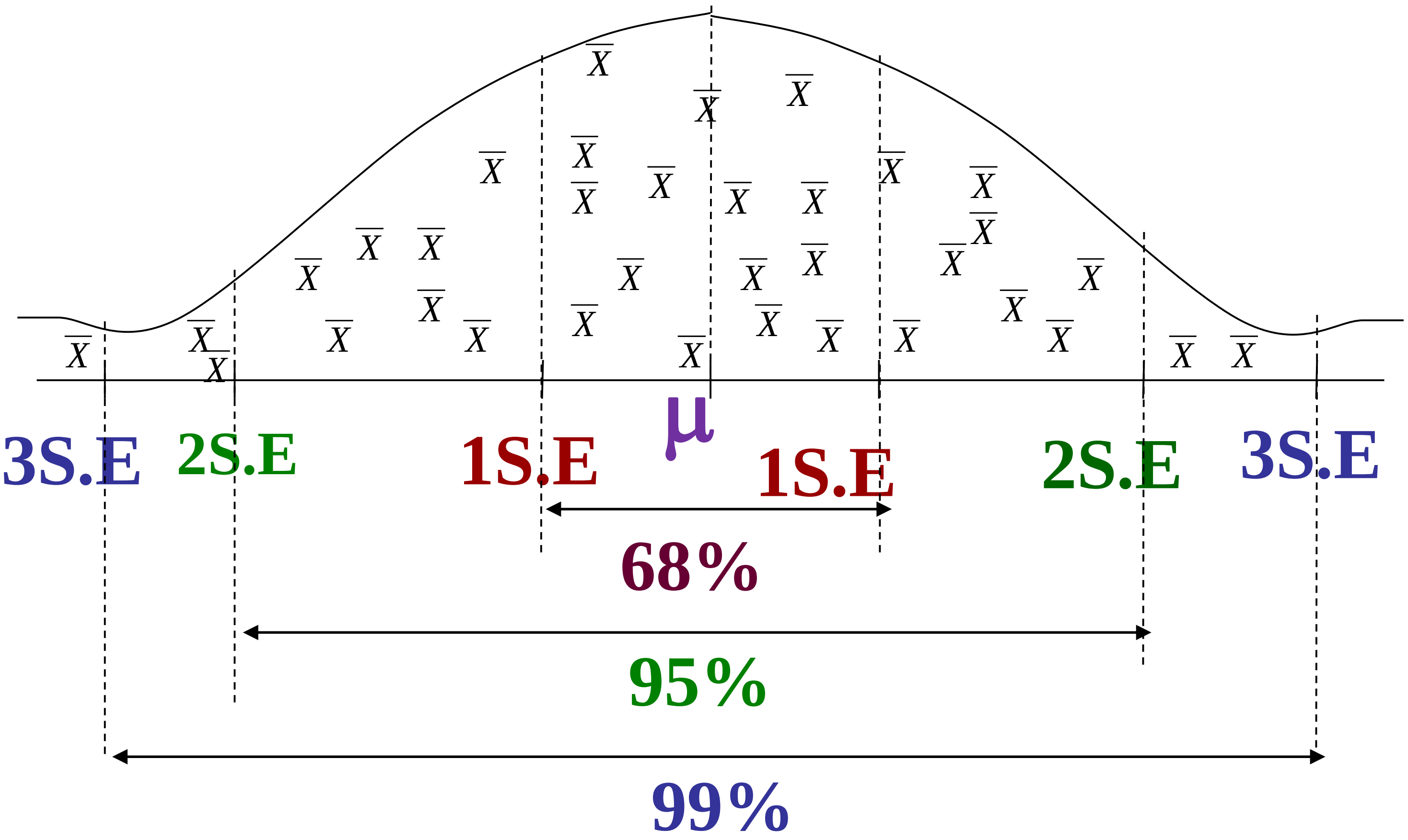
68% 95% 99% .

28

Remember that the

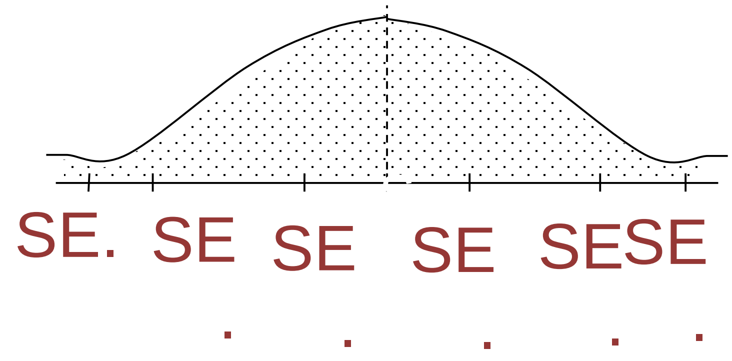
- **SD is**
  - **measure of spread of the data in a single sample .**
- **The S.E. is**
  - **a measure of spread in ALL sample means from a population.**
- **We notice that the as sample size n increases the S.E decrease**

29



???????

30

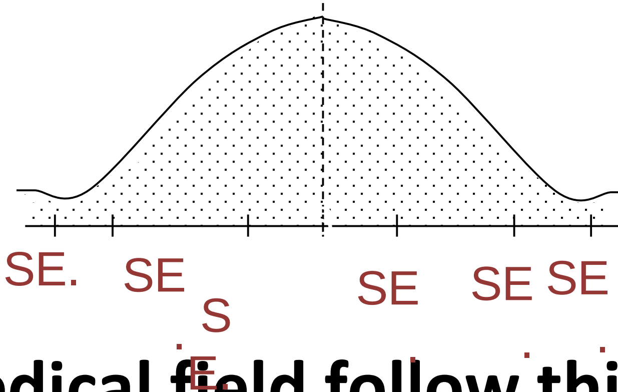


Importance

1-Most of the phenomenon in Medical field follow this distribution .

2-It is for justification and calculation of confidence interval .

3-It is form the basis of most of significance testing hypothesis . That is most test of significance depend on the theory of NDC .



Importance

1-Most of the phenomenon in Medical field follow this distribution .

2-It is for justification and calculation of confidence interval .

3-It is form the basis of most of significance testing hypothesis .
That is most test of significance depend on the theory of NDC .

Confidence Interval

The properties of NDC can be applied in distribution of ($\overline{X}_S$)

- **Distribution of samples** mean **(**$\overline{X}_S$ **)** around **the population** mean or universe mean (μ) in NDC area **is similar to that of the distribution of X (values) around** sample mean ($\overline{X}$)

- deviated from μ by **S.E and its multiplicity**, so deviated from μ by **1S.E, 2S.E and 3S.E** in **proportion**

- **Deviation of $\overline{X}$ from μ by more than 2 S.E is a rare event or uncommon,**

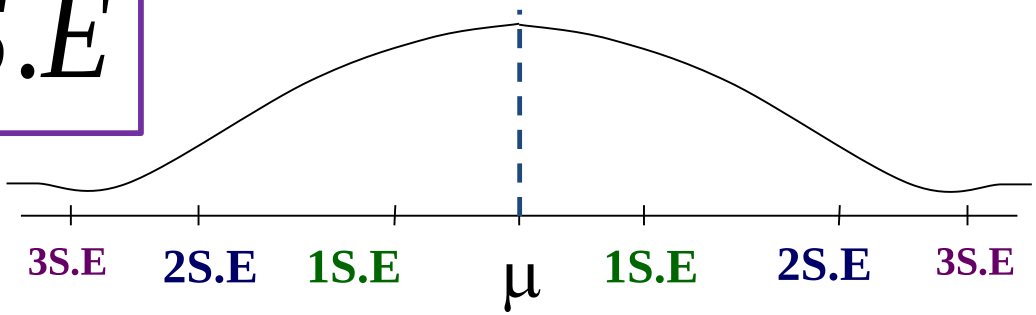
It is not more the 0.05 (5%) .

Deviation of from μ by more than 3 S.E **is very very rare event** **it is not more the 0.01 (1%)**

So by follow the NDC, we could find that the **rang at which** population mean μ is **located depending on relation to the sample mean** $\overline{X}$

5% or **0.05** of the **sample means(** $\overline{X}$**s) deported from the μ by more than ±2S.E** (**out side the limit of** $\overline{X}$ **±2S.E**) **. So approximately 95%** of the **samples mean** $\overline{X}$ **s will lie within 2S.E above or below μ.**

$$\boxed{\mu = \overline{X} \pm 2S.E}$$



Standard Error S.E

Example

8 plasma values of uric acid

the mean ($\overline{X}$) **of uric acid is 3±0.31**

$$S.E = \frac{0.31}{\sqrt{8}} = 0.11$$

$$\boxed{S.E = \frac{S.D}{\sqrt{N}}}$$

So **by this fact we could construct or conduct the population mean (μ) based on sample mean ($\overline{x}$**

Population mean (μ) within 95%

$$= \overline{X} \pm 2S.E$$

$$= \overline{X} \pm 1.96S.E$$

$$= 3 \pm 1.96 \times 0.11$$

$$= 3 \pm 0.215$$

95% of population mean μ rang (2.785) 2.8 – 3.215, such rang we call it Confidence Interval .

So 95% confidence interval of μ

$$= \overline{X} \pm 1.96S.E$$

Cont. Confidence Interval

Similarly

99% confidence interval of μ

$$= \overline{X} \pm 3S.E$$

$$= \overline{X} \pm 2.58S.E$$

= ????????

=????????

Cont. Confidence Interval

Similarly

99% confidence interval of μ $= \overline{X} \pm 3S.E$

$$= \overline{X} \pm 2.58S.E$$

= 3±2.58X0.11

=3±0.2838

=2. 7162- 3.2838

Standard Error S.E

Example

8 plasma values of uric acid

the mean ($\overline{X}$)**of uric acid is 3±0.31**

$$S.E = \frac{S.D}{\sqrt{N}}$$

$$S.E = \frac{0.31}{\sqrt{8}} = 0.11$$

16 plasma values of uric acid the mean ($\overline{X}$))**of uric acid is 3±0.31**

$$= \frac{0.31}{\sqrt{16}} = 0.0775$$

$$\frac{0.21}{\sqrt{16}} = 0.0525$$

Confidence Interval

It is the rang of the variability of population mean (μ) around the sample mean $\overline{X}$

99% C.I population mean μ $= \overline{X} \pm 2.58 S.E$

95% C.I population mean μ $= \overline{X} \pm 1.96 S.E$

Confidence Interval

- 95% chance that the error in as our estimate of $\overline{X}$ is not numerically grater than 1.96 S.E .

- In other word, if variable is normally distributed, then we may say within **certainty that 95% of all observation**
- **will fall with a rang $\pm$1.96 S.E from the μ, , or**

- **95% certainty** we have, that our sample mean
- ✓ does not differ from population mean (μ,) by **not more than $\pm$1.96 S.E** .
- ✓ **Only 5%** of the sample mean $\overline{X}$ deport from μ by more than 1.96 S.E .

$\overline{X}$

Confidence Interval

- 99% chance that the error in as our estimate of $\overline{X}$ is not numerically grater than 2.58 S.E .

- In other word, if variable is normally distributed, then we may say within **certainty that 99% of all observation**
- **will fall with a rang ±2.58 S.E from the μ, , or**

- **99% certainty** we have, that our sample mean
- ✓ does not differ from population mean (μ,) by **not more than ±2.58 S.E** .
- ✓ **Only 1%** of the sample mean $\overline{X}$ deport from μ by more than 2.58 S.E .

$\overline{X}$


Thank you!

Calculate measures of CT
Calculate measures of Dispersion
Within which range the 95% of the population mean
Within which range the 99% of the population mean

age of 50 patients

68, 62, 62, 66, 68, 65, 64, 71,77, 74, 20, 33, 38. 42, 47. 50,55, 56, 60 72, 80 74, 75, 74, 77,80, 81, 89, 86, 85, 83,72, 70, 71, 79, 76, 77, 80, 90, 97, 94, 90,65, .60, 67, 63 88, 84, 84, 87

????????