COURSE CODE: ASC201

COUSE TITLE: PROBABILITY & STATISTICS

Course Instructor: Engr. Asif Hussain

Descriptive vs. Inferential Statistics

- **Descriptive**
  - Methods for summarizing data
  - Summaries usually consist of graphs and numerical summaries of the data
- **Inferential**
  - Methods of making decisions or predictions about a populations based on sample information.

Descriptive Statistics		Inferential Statistics	
Measures of Central Tendency	Measures of Dispersion	Hypothesis Testing	Regression Analysis
Mean	Range	Z test	
Median	Standard Deviation	F test	Linear Regression
Mode	Variance	T test	
	Absolute Deviation		

Mean

Mean – the average of a group of numbers.

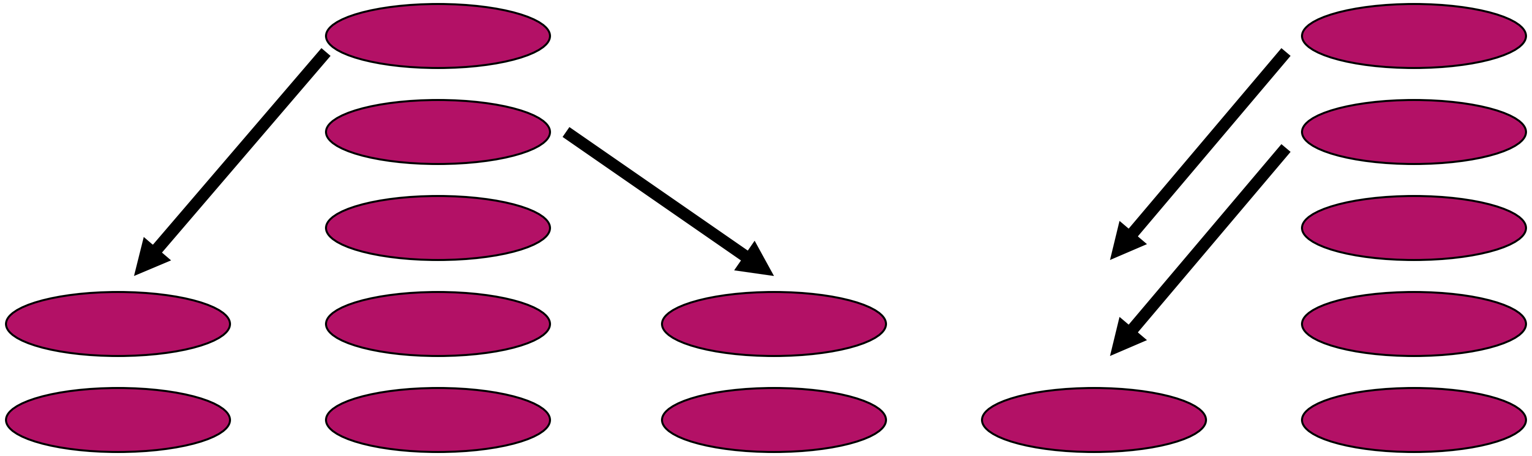
2, 5, 2, 1, 5

Mean = 3

Population Mean is denoted by **μ**

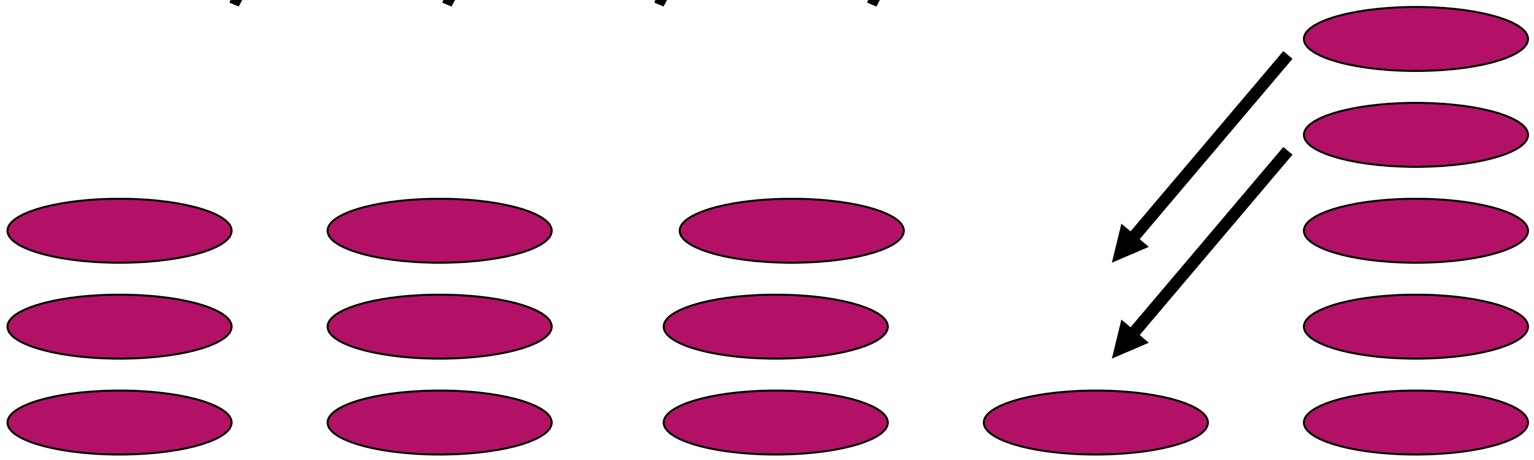
Mean is found by evening out the numbers

2, 5, 2, 1, 5



Mean is found by evening out the numbers

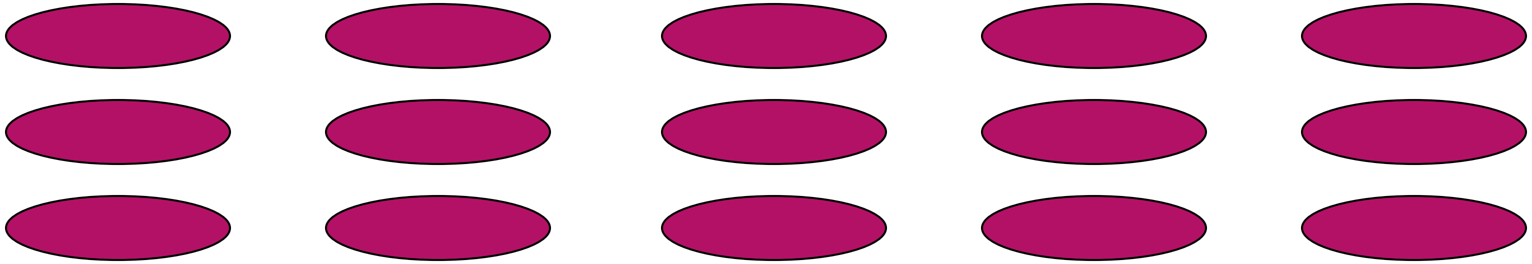
2, 5, 2, 1, 5



Mean is found by evening out the numbers

2, 5, 2, 1, 5

mean = 3



The mean (or arithmetic mean) of n observations (variates) x_1, x_2, x_3, x_4, ….., x_n is given by

Mean = $\frac{x_1+x_2+x_3+x_4+....+x_n}{n}$

In words, mean = $\frac{\textbf{Sum of the Variables}}{\textbf{Total Number of Variates}}$

Symbolically, A = $\frac{\sum x_i}{n}$; i = 1, 2, 3, 4, …., n.

Note: $\sum x_i$ = nA, i,e., sum of variates = mean × number of variates.

How to Find the Mean of a Group of Numbers

- Step 1 – Add all the numbers.

8, 10, 12, 18, 22, 26

$$8+10+12+18+22+26 = 96$$

How to Find the Mean of a Group of Numbers

- Step 2 – Divide the sum by the number of addends.

# of addends

$$\begin{array}{c} 16 \\ \hline 96 \\ \hline 6 \end{array} \longleftarrow \text{sum}$$

How to Find the Mean of a Group of Numbers

8, 10, 12, 18, 22, 26

The mean or average of these numbers is 16.

What is the mean of these numbers?

7, 10, 16

11

What is the mean of these numbers?

2, 9, 14, 27

13

What is the mean of these numbers?

26, 33, 41, 52

38

Median

Median is in the Middle

Definition

Median – the middle number in a set of ordered numbers.

1, 3, 7, 10, 13

Median = 7

How to Find the Median in a Group of Numbers

- Step 1 – Arrange the numbers in order from least to greatest.

21, 18, 24, 19, 27

18, 19, 21, 24, 27

How to Find the Median in a Group of Numbers

- Step 2 – Find the middle number.

18, 19, 21, 24, 27

This is your median number.

How to Find the Median in a Group of Numbers

- Step 3 – If there are two middle numbers, find the mean of these two numbers.

18, 19, 21, 25, 27, 28

How to Find the Median in a Group of Numbers

- Step 3 – If there are two middle numbers, find the mean of these two numbers.

$$21 + 25 = 46$$

$$2\overline{)46} \quad 23 \leftarrow \text{median}$$

What is the median of these numbers?

16, 10, 7

7, 10, 16

10

What is the median of these numbers?

29, 8, 4, 11, 19

4, 8, 11, 19, 29

11

What is the median of these numbers?

31, 7, 2, 12, 14, 19

2, 7, 12, 14, 19, 31

$12 + 14 = 26$

$$2\overline{)26}^{\;13}$$

What is the median of these numbers?

53, 5, 81, 67, 25, 78

5, 25, 53, 67, 78, 81

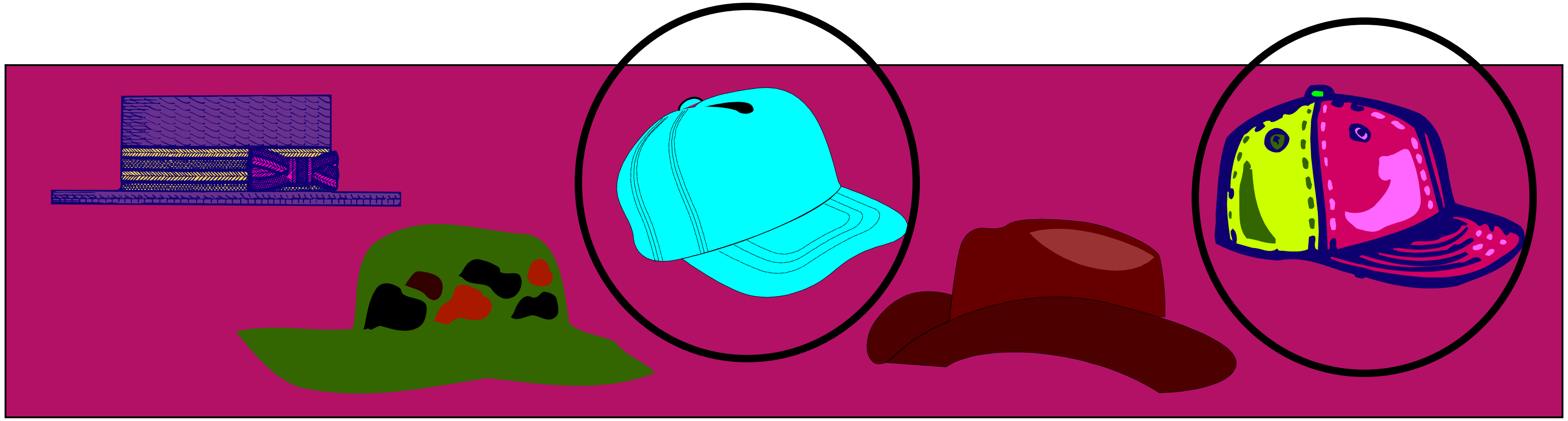
$$53 + 67 = 120$$

$$2\overline{)120}\quad 60$$

Mode

Mode is the most Popular

A mode – the most popular or that which is in fashion.



Baseball caps are a mode today.

Mode – the number that appears most frequently in a set of numbers.

1, 1, 3, 7, 10, 13

Mode = 1

How to Find the Mode in a Group of Numbers

- Step 1 – Arrange the numbers in order from least to greatest.

21, 18, 24, 19, 18

18, 18, 19, 21, 24

How to Find the Mode in a Group of Numbers

Step 2 – Find the number that is repeated the most.

21, 18, 24, 19, 18

18, 18, 19, 21, 24

Which number is the mode?

29, 8, 4, 8, 19

4, 8, 8, 19, 29

8

Which number is the mode?

1, 2, 2, 9, 9, 4, 9, 10

1, 2, 2, 4, 9, 9, 9, 10

9

Range

Range is the distance

Between greatest and the least value

Definition

Range – the difference between the greatest and the least value in a set of numbers.

1, 1, 3, 7, 10, 13

Range = 12

How to Find the Range in a Group of Numbers

Step 1 – Arrange the numbers in order from least to greatest.

21, 18, 24, 19, 27

18, 19, 21, 24, 27

How to Find the Range in a Group of Numbers

- Step 2 – Find the lowest and highest numbers.

21, 18, 24, 19, 27

18, 19, 21, 24, 27

How to Find the Range in a Group of Numbers

Step 3 – Find the difference between these 2 numbers.

18, 19, 21, 24, 27

27 – 18 = 9

The range is 9

What is the range?

29, 8, 4, 8, 19

4, 8, 8, 19, 29

29 – 4 = 25

What is the range?

23, 7, 9, 41, 19

7, 9, 23, 19, 41

$41 - 7 = 34$

Exercise 1

The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	2.8
4.4	4.0	5.2	3.0	4.8

Assume that the measurements are a simple random sample.
(a) What is the sample size for the above sample?
(b) Calculate the sample mean for this data.
(c) Calculate the sample median..
(d) Compute the 20% trimmed mean for the above data set.

- (a) 15.
- (b) X = (3.4 + 2.5 + 4.8 + · · · + 4.8) /15 = 3.787
- (c) Sample median is the 8th value, after the data is sorted from smallest to largest: **3.6**
- (d) After trimming total 40% of the data (20% highest and 20% lowest), the data becomes:

  2.9 3.0 3.3 3.4 3.6

  3.7 4.0 4.4 4.8

  So. the trimmed mean is $$\bar{x}_{\mathrm{tr}20} = \frac{1}{9}(2.9 + 3.0 + \cdots + 4.8) = 3.678.$$

Exercise 2

A manufacturer of electronic components is interested in determining the lifetime of a certain type of battery. A sample, in hour's of life, is as follows:

123,116,122,110,175, 120,125, 111, 118, 117

Find the sample mean and median

Solution:

Mean = $\bar{X}$ = 123.7 and median = $\tilde{X}$ = 119

Exercise 3

Consider a small unit of a factory where there are 5 employees : a supervisor and four labourers. The workers earn a salary of Rs. 5,000 per month each while the supervisor gets Rs. 15,000 per month. Calculate the mean, median , mode and range of the salaries.

Solution:

Mean = (5000 + 5000 + 5000 + 5000 + 15000)/5 = 35000/5 = 7000
So, the mean salary is Rs. 7000 per month

To obtain the median, let us arrange the salaries in ascending order:
5000, 5000, 5000, 5000, 15000
Median = Rs. 5000/-

Mode = Number of times an observation is repeated = Rs.5000/-

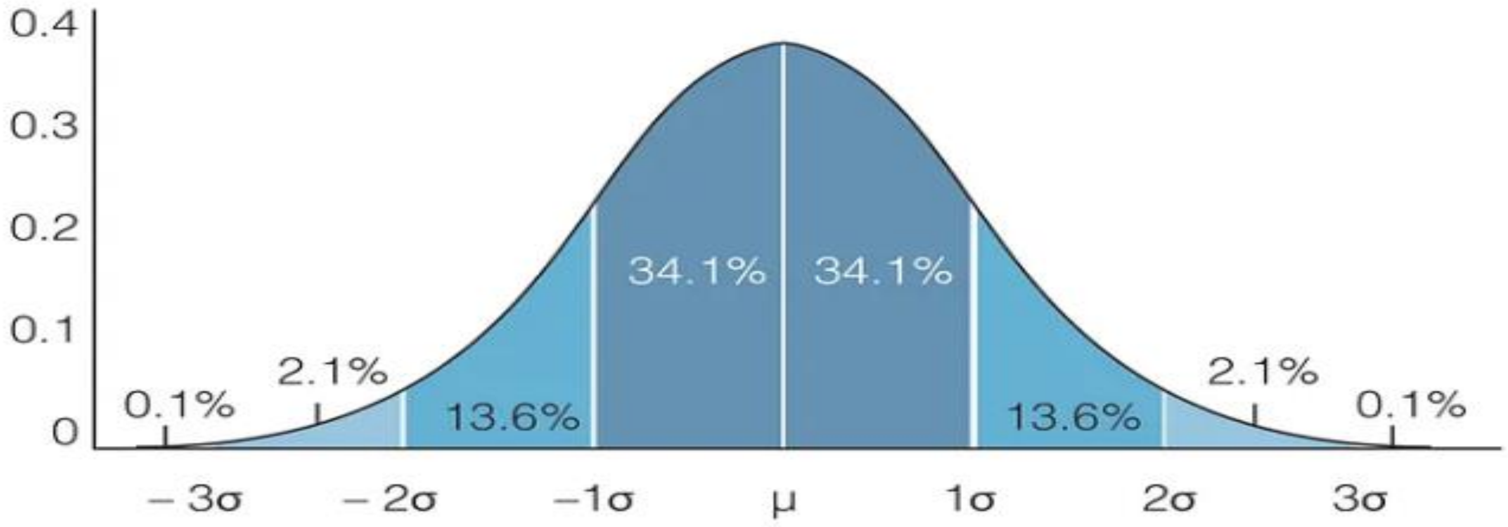
Range=15000-5000=10000

VARIANCE

- **Variance** is the average squared deviation from the mean of a set of data.
- It is used to find the **standard deviation**.

Variance is denoted by $\boldsymbol{\sigma}^2$

Distribution of Variance



μ = Expected Value
-1σ to 1σ = 1 Standard Deviation (ie: ~2/3 of the time, your results/variance will fall within this range)
-2σ to 2σ = 2 Standard Deviations (ie: 95% of the time, your results/variance will fall within this range)
-3σ to 3σ = 3 Standard Deviations (ie: 99.7 of the time, your results/variance will fall within this range)

VARIANCE

- Find the Mean of the data.
  - Mean is the average so add up the values and divide by the number of items.
- Subtract the mean from each value – the result is called the deviation from the mean.
- Square each deviation of the mean.
- Find the sum of the squares.
- Divide the total by the number of items.

Variance – Review

- (1) **Calculate** the **mean**

  → $\bar{x}$

- (2) **Calculate** the **deviation** for each value

  → $x_i - \bar{x}$

- (3) **Square** each of the deviations

  → $(x_i - \bar{x})^2$

- (4) **Sum** the **squared** deviations

  → $\sum (x_i - \bar{x})^2$

- (5) **Divide** the **sum of squares** by (n-1) for a sample

  → $\sum (x_i - \bar{x})^2 / (n-1)$

VARIANCE FORMULA

The variance formula includes the Summation Notation, $\sum$ which represents the sum of all the items to the right of Sigma.

$$o^2 = \frac{\sum\left(x - \bar{X}\right)^2}{N}$$

For population variance

$$s^2 = \frac{\sum\left(x - \bar{X}\right)^2}{n-1}$$

For sample variance

Mean is represented by μ & $\bar{x}$ and ***n* & *N*** is the number of items.

Degree Of Freedom

What is Degrees of Freedom?
Degrees of freedom are the maximum number of logically independent values, which may vary in a data sample. Degrees of freedom are calculated by subtracting one from the number of items within the data sample.

The values of the five(5) integers must have an average of six(6). If four items within the data set are {3, 8, 5, and 4}, the fifth number must be 10. Because the first four numbers can be chosen at random, the degree of freedom is four(4).

STANDARD DEVIATION

- **Standard Deviation** shows the variation in data.
- If the data is close together, the standard deviation will be small.
- If the data is spread out, the standard deviation will be large.
- **Standard Deviation** is often denoted by the lowercase Greek letter sigma, σ

STANDARD DEVIATION

Find the **variance**.

a) Find the **Mean** of the data.
b) Subtract the mean from each value.
c) Square each deviation of the mean.
d) Find the sum of the squares.
e) Divide the total by the number of items.

- Take the square root of the variance.

Standard Deviation Formula

The standard deviation formula can be represented using Sigma Notation:

$$s = \sqrt{\frac{\sum\left(x - \bar{X}\right)^2}{n-1}}$$

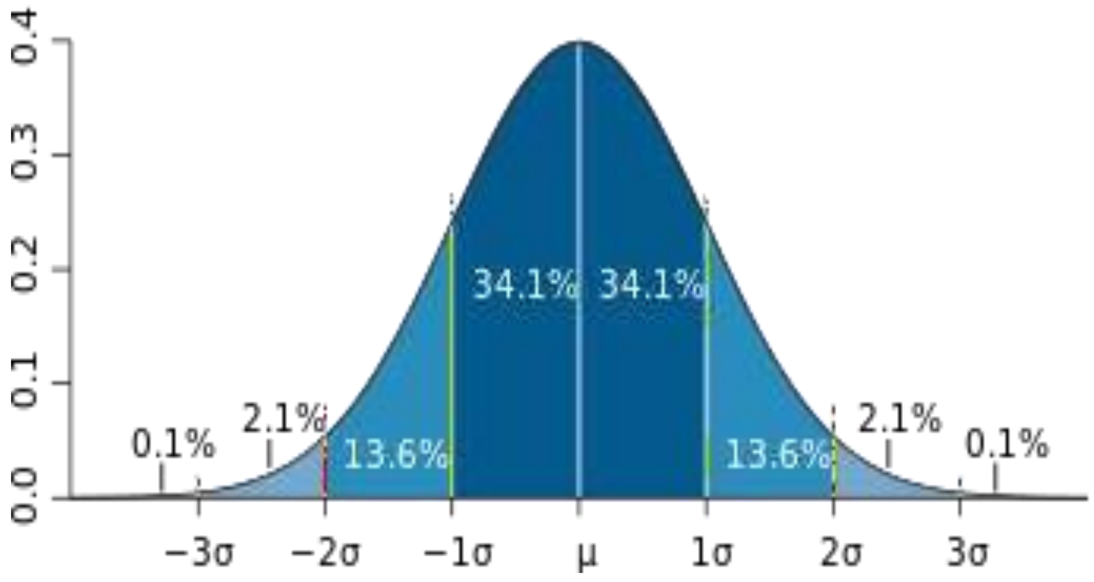
$$o = \sqrt{\frac{\sum\left(x - \mu\right)^2}{N}}$$

sample standard deviation **population standard deviation**

The standard deviation formula is the square root of the variance.

GRAPH

- The bell curve is commonly seen in statistics as a tool to understand **standard deviation**.



- The **graph** of a normal distribution represents a great deal of data in real life. The mean, or average, is represented by the Greek letter μ, in the center.

Find The Variance And Standard Deviation

Example : 1

The math test scores of five students are: 92,88,80,68 and 52.

“Consider test scores values are (x)”

1) Find the **Mean** ($\bar{x}$):

$(92+88+80+68+52)/5 = 76.$

2) Find the **deviation from the mean**:

$(x - \bar{x})$

92-76=16

88-76=12

80-76=4

68-76= -8

52-76= -24

3) Square the deviation from the mean:

$(x - \bar{x})^2$

$(16)^2 = 256$

$(12)^2 = 144$

$(4)^2 = 16$

$(-8)^2 = 64$

$(-24)^2 = 576$

4) Find the sum of the squares of the deviation from the mean $(x - \bar{x})^2$:

256+144+16+64+576
= 1056

5) Divide by the number of data items to find the **variance**:

$$\frac{\sum (x - \bar{X})^2}{N}$$

1056/5 = 211.2

6) Find the square root of the variance:

$$\sqrt{211.2} = 14.53$$

Thus the standard deviation of the test scores is 14.53.

Standard Deviation Example

Example : 2

Measure the height of 4 out of 60 Students in your room.

165
145
153
150

1) Now you have to find the mean of the height's measured :

$$\textbf{Mean } (\bar{x}) = \frac{165+145+153+150}{4} = 153.25$$

2) Next step is to find the variance.

- It is the average of the squares of the differences from mean.
- Subtract individual height's from the mean and square each value.

3) Find the **deviation from the mean:**

$(x - \bar{x})$

165-153.25= 11.75

145-153.25= -8.25

153-153.25= -0.25

150-153.25= -3.25

4) Square the **deviation from the mean:**

$(x - \bar{x})^2$

$(11.75)^2 = 138.0625$

$(-8.25)^2 = 68.0625$

$(-0.25)^2 = 0.0625$

$(-3.25)^2 = 10.5625$

5) Find the sum of the squares of the deviation from the mean $(x - \bar{x})^2$

138.0625+68.0625+0.0625+10.5625=216.75

- Sum of the square of deviation is: 216.75

- For population standard deviation, we would calculate variance without subtracting “1” from the denominator.
- But here we subtracting “1” from the denominator.
- This Process is called degree of freedom.

6) Divide by the number of data items and subtracting 1 from the denominator to find the **variance**:

$$s^2 = \frac{\sum\left(x-\bar{X}\right)^2}{n-1} \qquad \frac{216.75}{4-1} = 72.25$$

7) Find the square root of the variance:

$$s = \sqrt{\frac{\sum\left(x-\bar{X}\right)^2}{n-1}} = \sqrt{72.25} = 8.5$$

Thus the standard deviation of the heights of the members is 8.5.

Exercise 3

The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.

3.4 2.5 4.8 2.9 3.6

2.8 3.3 5.6 3.7 2.8

4.4 4.0 5.2 3.0 4.8

Compute the sample-variance and sample standard deviation

$$s^2 = \frac{\mathbf{1}}{\mathbf{15-1}}[(\mathbf{3.4-3.787})\mathbf{2} + (\mathbf{2.5-3.787})\mathbf{2} + (\mathbf{4.8-3.787})\mathbf{2} + \cdots + (\mathbf{4.8-3.787})\mathbf{2}]$$

= 0.94284

s = $\sqrt{s^2}$ = $\sqrt{0.9428}$ = 0.971.

Exercise 4

The nicotine contents, in milligrams, for 40 cigarettes of a certain brand were recorded as follows:

1.09, 1.92 ,2.31, 1.79, 2.28 ,1.74, 1.47 ,1.97, 0.85 ,1.24

1.58 ,2.03 ,1.70 ,2.17, 2.55 ,2.11, 1.86 ,1.90, 1.68, 1.51

1.64, 0.72, 1.69, 1.85 ,1.82, 1.79, 2.46, 1.88, 2.08, 1.67

1.37 ,1.93, 1.40 ,1.64, 2.09, 1.75 ,1.63, 2.37, 1.75, 1.69

a) **Find the sample mean**

b) **Find the sample standard deviation**

(a) $\bar{X}$ = 1.7743 and

(b) s = 0.3905

Task: 1

An engineer is interested in testing the "bias" in a pH meter. Data are collected on the meter by measuring the pH of a neutral substance (pH — 7.0). A sample of size 10 is taken with results given by

7.07, 7.00 ,7.10, 6.97, 7.00, 7.03, 7.01 ,7.01 ,6.98, 7.08

Calculate:

a) **Mean**

b) **Sample Variance**

c) **Sample Standard Deviation**

d) **Degree of freedom**

- (a) Mean = 7.0250
- (b) Variance = S^2 = 0.001939
- (c) Standard deviation = S = 0.044
- (d) degrees of freedom = n-1 = 9

Task: 2

A tire manufacturer wants to determine the inner diameter of a certain grade of tire. Ideally, the diameter would be 570 mm. The data are as follows

572,572,573,568,569,575,565,570

(a) **Find the sample mean and median**

(b) **Find the sample variance, standard deviation, and range.**

- (a) Mean = $\bar{X}$ = 570.5 and median = $\tilde{X}$ = 571
- (b) Variance = s^2 = 10; standard deviation= s = 3.162; range=10;

Conclusion

- As we have seen, standard deviation measures the dispersion of data.
- The greater the value of the standard deviation, the further the data tend to be dispersed from the mean.

Sampling Procedure / Method of Sampling

- Sampling techniques or methods of sampling means *"How we select a sample from population for study/research?"*
- *To draw valid conclusion from research sample should be carefully selected which represents the whole population. Two types of sampling techniques:*

Probability Sampling	**Non-Probability Sampling**
• Every member of the population has <u>equal chance</u> of being selected for research.	• Every member of the population has <u>not equal chance</u> of being selected for research.

Sampling Procedure / Method of Sampling

Probability Sampling	Non-Probability Sampling
• Simple Random Sampling	• Convenience Sampling
• Stratified Sampling	• Judgement Sampling
• Systematic Sampling	• Quota Sampling
• Cluster Sampling	• Snowball Sampling

Probability Sampling

Simple Random Sampling

- In this technique every member of the sample is selected purely random basis with equal chance.
- Picking chits from bowl, lottery system, random number generator etc. are the methods of simple random sampling.

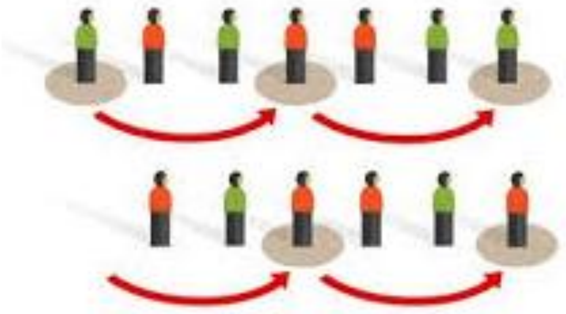
Stratified Sampling

- In this technique population is divided into mutually exclusive groups, and then every member of the group has equal chance of being selected for research.

(e.g. gender, age range, income bracket, job role).

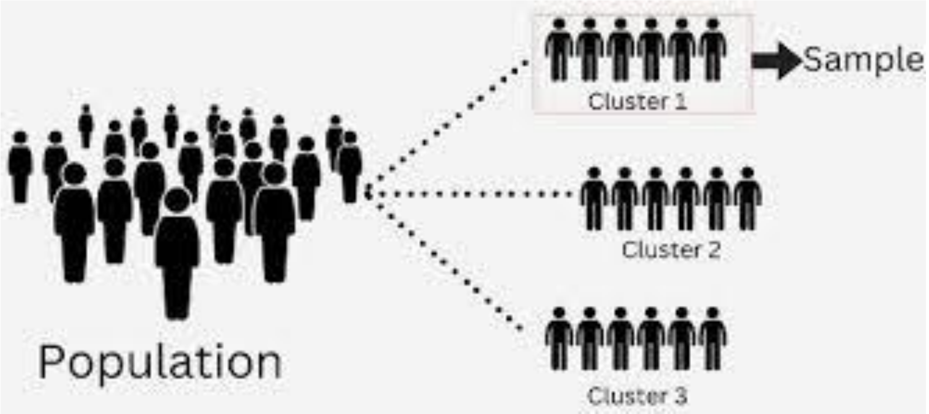
Systematic Sampling

- In this technique population is arranged in ascending or descending order, then researcher randomly picks first items from population.
- Sampling interval=total population/sample size=1000/100=10
- Suppose first number picked by researcher is 7, the next will be 7+10=17, next will be 17+10=27, 27+10=37 and so on.



Cluster/Area Sampling

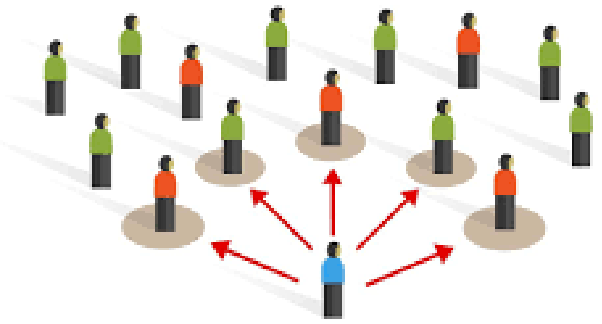
- This technique is used when large population (geographically dispersed) is under study. Whole population is divided into small groups which is called clusters.
- For example a company want to study the performance of a particular product in the country. The country is divided into cluster (cities, towns, metropolitan areas etc.)



Non-Probability Sampling

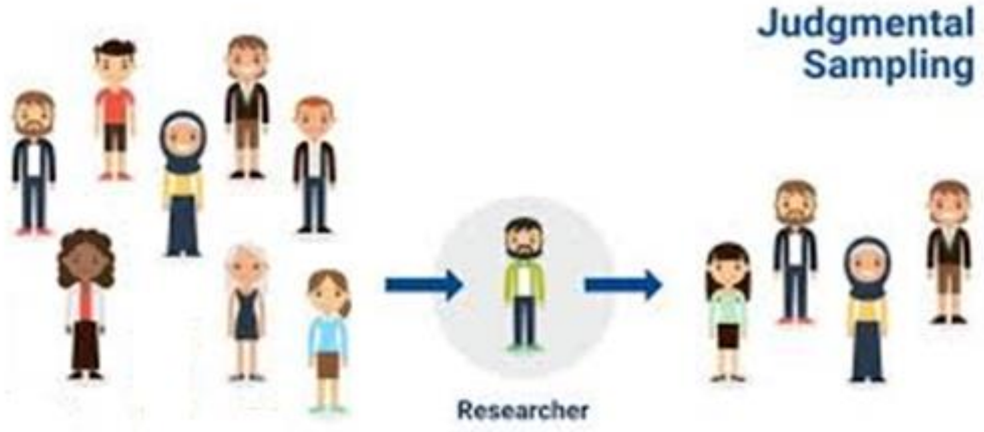
Convenience Sampling

- Convenience sampling is also called grab sampling, availability sampling, accidental sampling etc.
- It is type of sampling in which data is collected from the *"conveniently available respondents".*
- It is low cost and fast sampling technique.



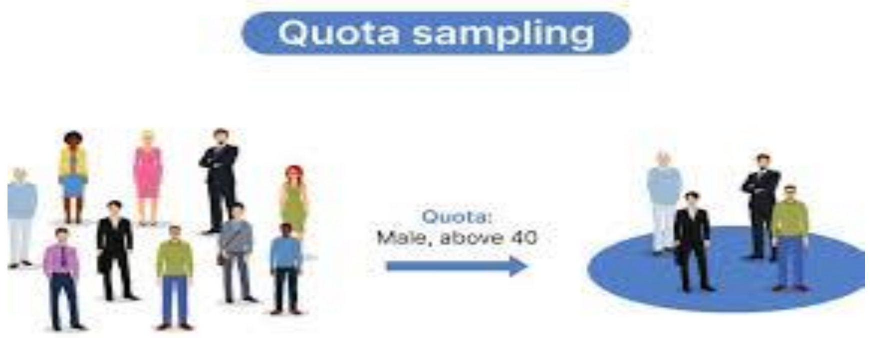
Judgement/Purposive Sampling

- It is also called authoritative sampling, selective sampling, subjective sampling.
- It is a sampling technique in which researcher selects the respondents based on his knowledge & judgement.
- It is easy & cost effective sampling technique. But is vulnerable to sampling bias as it is entirely depends on researcher's judgement.



Quota Sampling

- It is a sampling technique in which entire population is divided in to groups and then quota (no of items to be selected for research) is assigned against each group.
- Groups examples: males, females, employed or unemployed people, age groups, location etc.
- Once the quota is assigned to each group then sample is selected on convenience or personal judgement.



Snowball Sampling

- As the snowball moves further from top to bottom on glacier it gets bigger and bigger.
- It is a sampling technique in which researcher selects one or two respondents first. These respondents refer or identify other respondents.
- Researcher continuously selects respondents based on referral until required sample size is achieved.
- Snowball sampling is also called referral sampling, chain sampling, network sampling, friend to friend sampling.