Subject: Statistics & Probability

Unit 4: Hypothesis Testing

Sampling Distribution, Hypothesis testing, Types of errors, level of significance, Critical value (p-test), Chi-Square test, z test, t-test, ANOVA, Application of hypothesis testing to production control.

Prepared by Mr. Dinesh Kute

Sampling Distribution

- In order to inference about a certain phenomenon, sampling is well accepted tool.

- Entire population cannot be studied due to several reasons. In such a situation sampling is the only alternative.

- **Definition:** A sampling distribution is a **probability distribution** of a **statistic** obtained from a larger number of samples drawn from a specific population.

- **Types of Sampling Method**

  The two different types of sampling methods are::

  - Probability Sampling
  - Non-probability Sampling

Probability Sampling

- The probability sampling method utilizes some form of random selection.
- In this method, all the eligible individuals have a chance of selecting the sample from the whole sample space.
- This method is more time consuming and expensive than the non-probability sampling method.
- The benefit of using probability sampling is that it guarantees the sample that should be the representative of the population.

Types Probability Sampling

1) Simple Random Sampling:

- In simple random sampling technique, every item in the population has an equal and likely chance of being selected in the sample. Since the item selection entirely depends on the chance, this method is known as “Method of chance Selection”.

Example:

- Suppose we want to select a simple random sample of 200 students from a school. Here, we can assign a number to every student in the school database from 1 to 500 and use a random number generator to select a sample of 200 numbers.

Types Probability Sampling

2) Systematic Sampling:

- In the systematic sampling method, the items are selected from a larger population are selected according to a random starting point but with a fixed, periodic interval.

- This interval, called the sampling interval, is calculated by dividing the population size by the desired sample size.

Example:

- Suppose the names of 300 students of a school are sorted in the reverse alphabetical order. We have to choose some 15 students by randomly selecting a starting number, say 5. From number 5 onwards, will select every 15th person from the sorted list. Finally, we can end up with a sample of some students

Types Probability Sampling

3) Stratified Sampling:

- In a stratified sampling method, the total population is divided into subgroups (called strata) to complete the sampling process. The small group is formed based on a few characteristics in the population.
- Simple random sample is taken from each subgroup (or stratum).
- Each subgroup (strata) is homogeneous in nature.

Example:

- One might divide a sample of adults into subgroups by age, like 18–29, 30–39, 40–49, 50–59, and 60 and above.

Types Probability Sampling

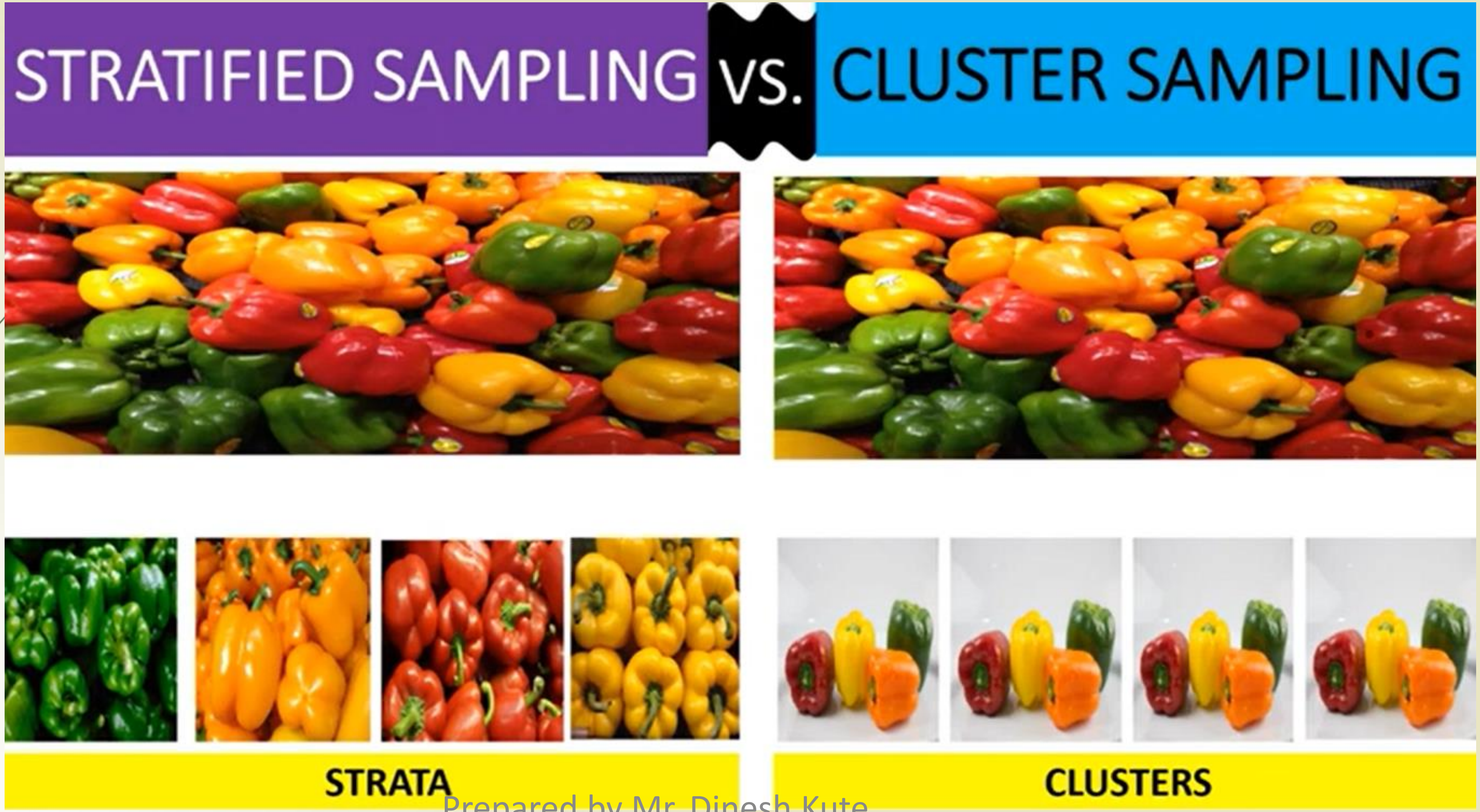
1) Clustered Sampling:

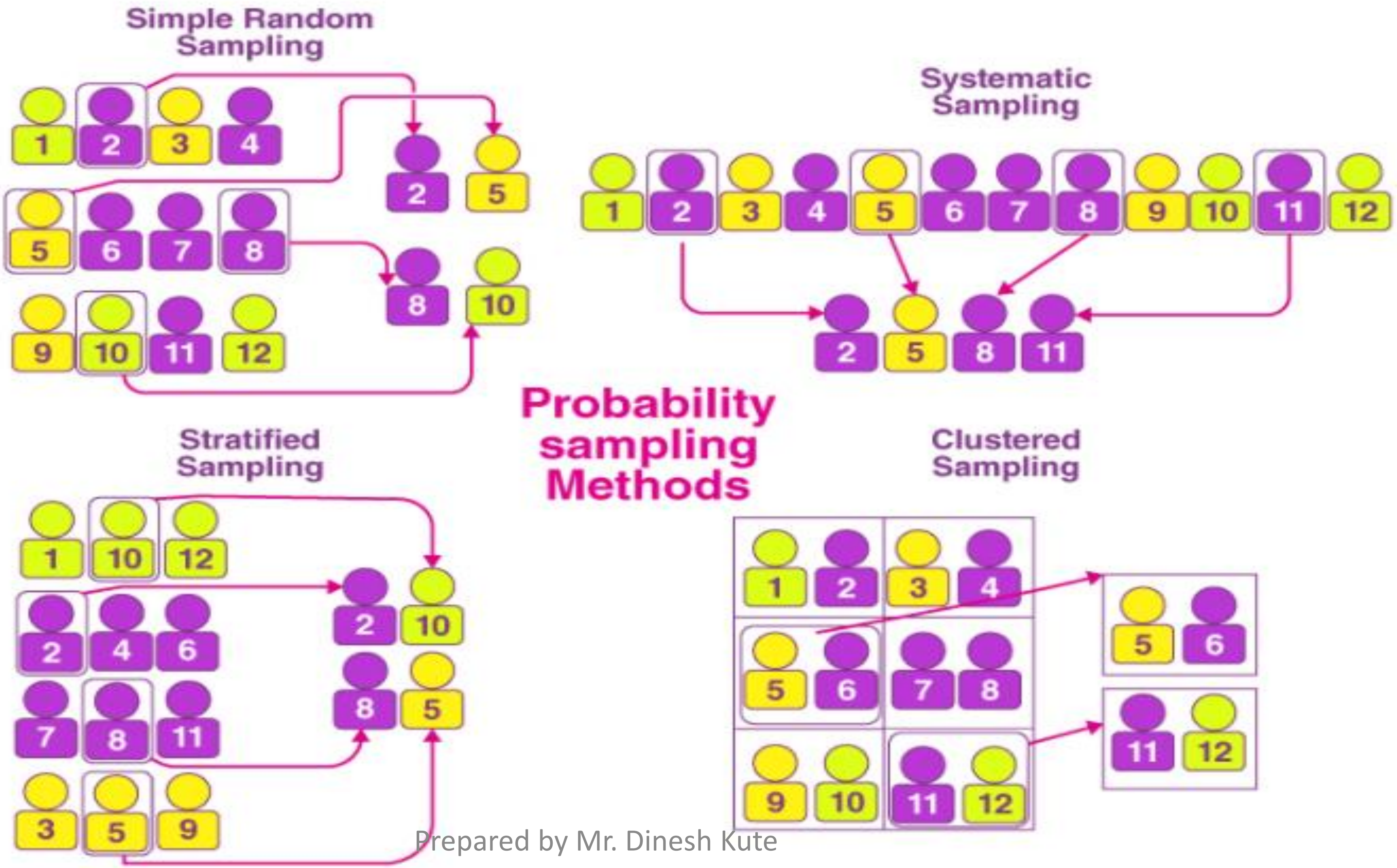
- In the clustered sampling method, the cluster or group of people are formed from the population set.
- Simple random sample of the clusters is taken.
- Each cluster is a representative small scale version of the population, so heterogeneous in nature.

Example:

- An educational institution has ten branches across the country with almost the number of students. If we want to collect some data regarding facilities and other things, we can’t travel to every unit to collect the required data. Hence, we can use random sampling to select three or four branches as clusters.

Difference between Stratified & Cluster Sampling




Simple Random Sampling
1
2
3
4
5
6
7
8
9
10
11
12
2
5
8
10
Systematic Sampling
1
2
3
4
5
6
7
8
9
10
11
12
2
5
8
11
Probability sampling Methods
Stratified Sampling
1
10
12
2
4
6
7
8
11
3
5
9
2
10
8
5
Clustered Sampling
1
2
3
4
5
6
7
8
9
10
11
12
5
6
11
12

Non Probability Sampling

- The non-probability sampling method is a technique in which the researcher selects the sample based on subjective judgment rather than the random selection.

Non Probability Sampling

Types of Non Probability Sampling:

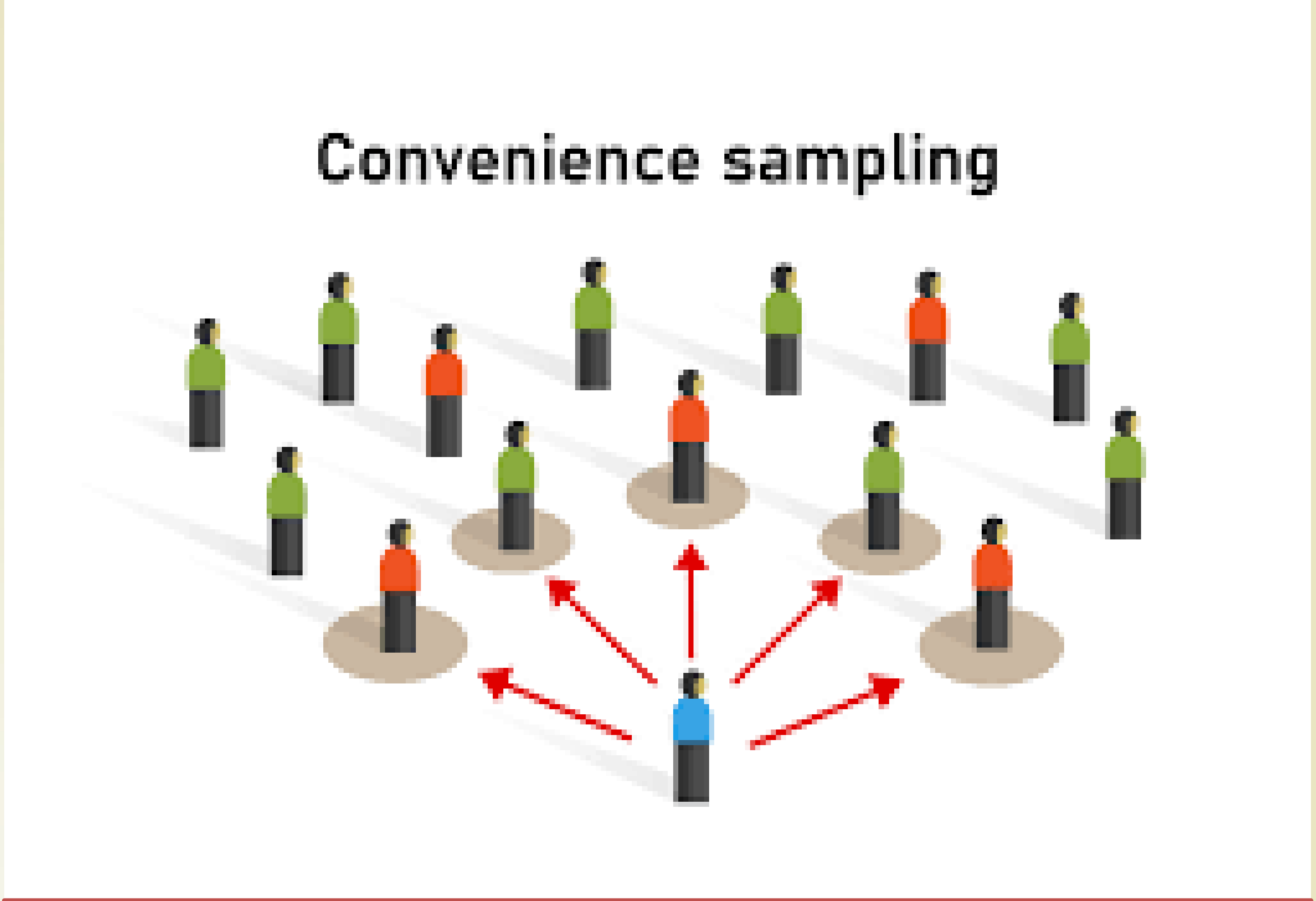
1) Convenience Sampling:

- Convenience sampling is a non-probability sampling technique where samples are selected from the population only because they are conveniently available to the researcher.
- Researchers choose these samples just because they are easy to recruit, and the researcher did not consider selecting a sample that represents the entire population

Example

- A basic example of a convenience sampling method is when companies distribute their promotional pamphlets and ask questions at a mall or on a crowded street with randomly selected participants.
- Businesses use this sampling method to gather information to address critical issues arising from the market.

Non Probability Sampling



I will choose members that are closest to me. That's easy and it doesn't cost anything

Non Probability Sampling

Types of Non Probability Sampling:

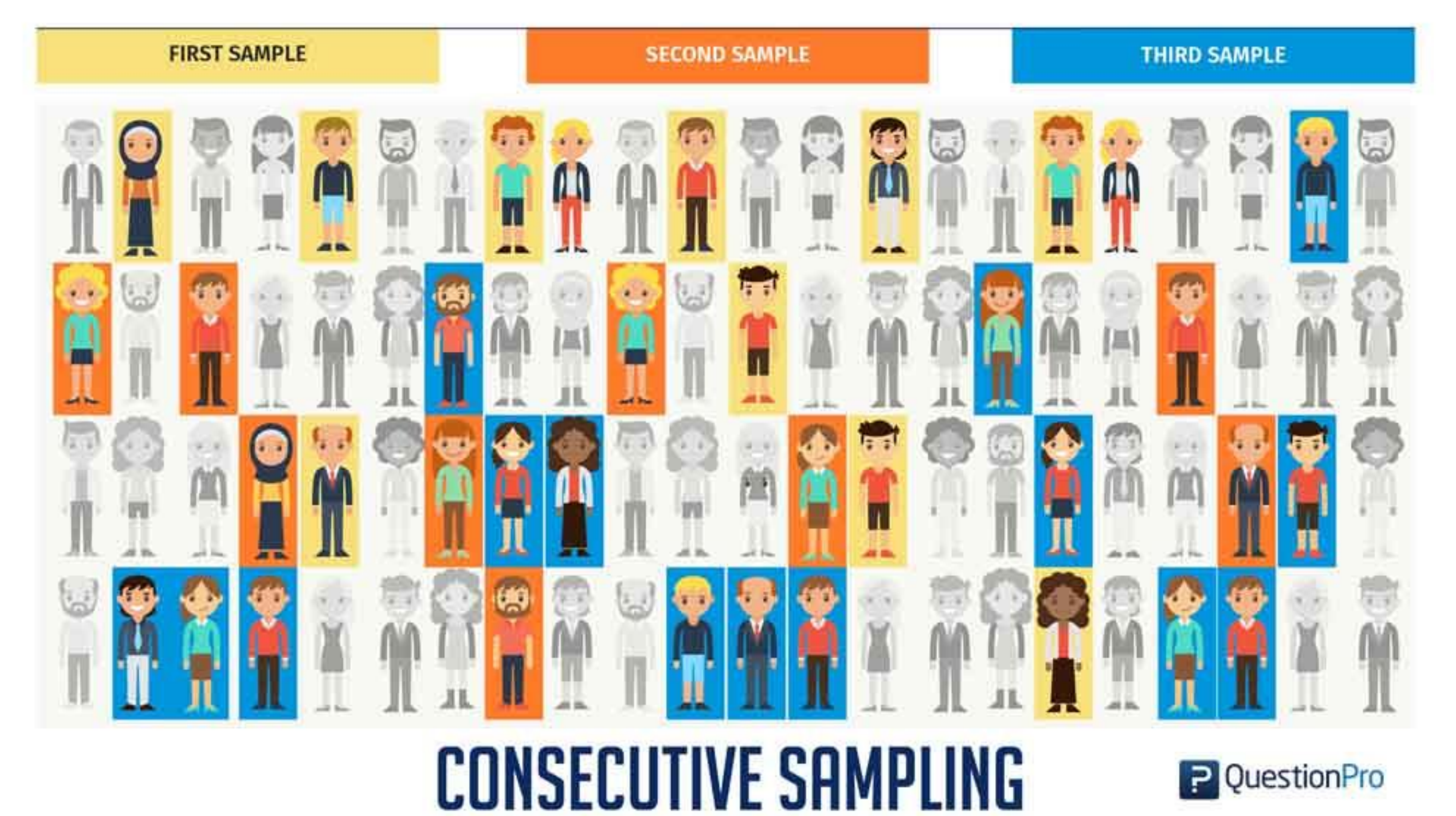
1) Consecutive Sampling:

- This non-probability sampling method is very similar to convenience sampling, with a slight variation. Here, the researcher picks a single person or a group of a sample, conducts research over a period, analyzes the results, and then moves on to another subject or group if needed.

Example

- A basic example of a consecutive sampling method is when companies distribute their promotional pamphlets and ask questions at a mall or on a crowded street with randomly selected participants and then move to the next location.

Consecutive Sampling



Prepared by Mr. Dinesh Kute

Non Probability Sampling

Types of Non Probability Sampling:

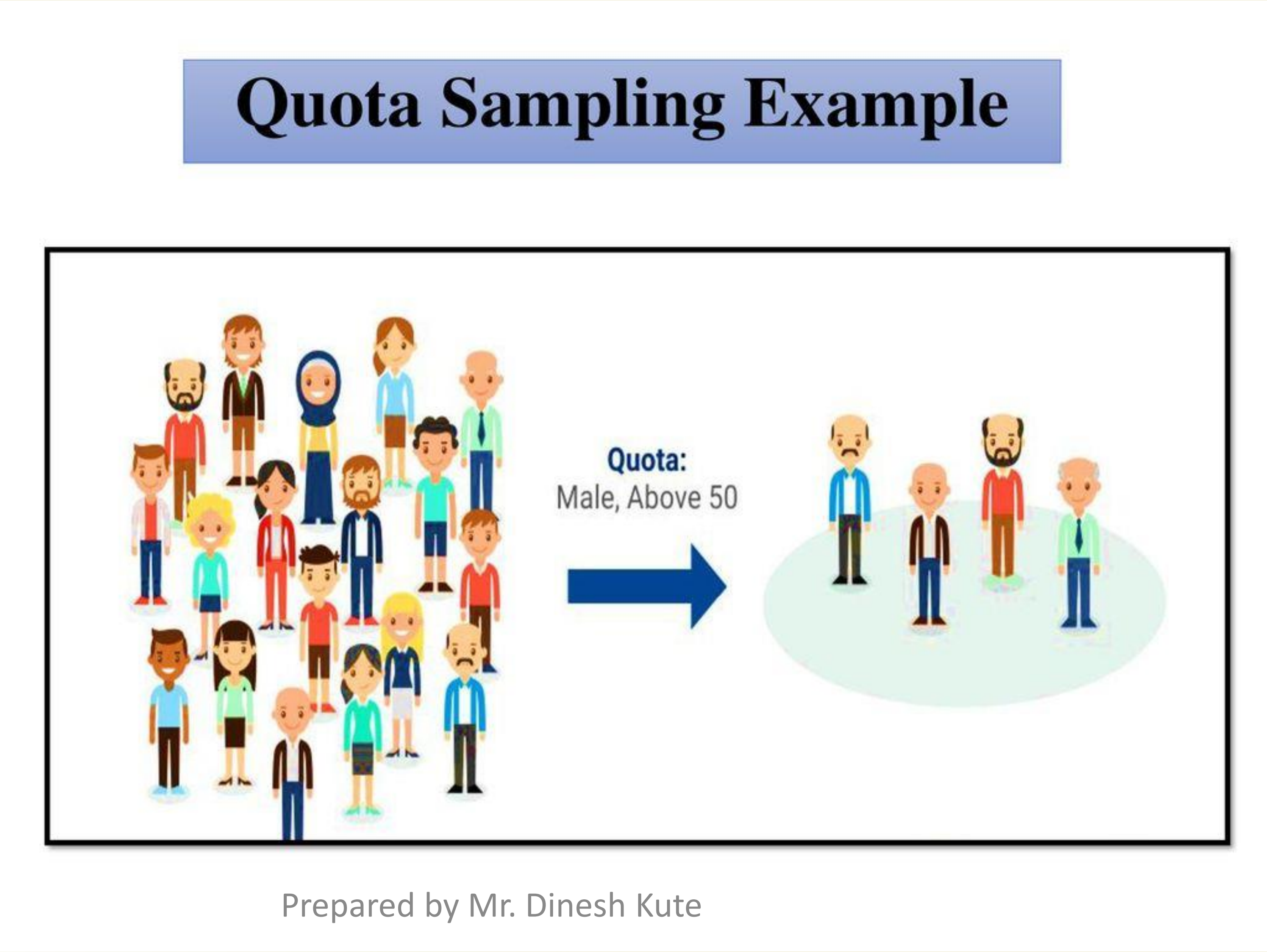
1) Quota Sampling:

- ❑ Quota sampling is defined as a non-probability sampling method in which researchers create a sample involving individuals that represent a population.
- ❑ The main reason why researchers choose quota samples is that it allows the researchers to sample a subgroup that is of great interest to the study

Example

- ❑ Divide the group of students into 30% fresher's, 25% junior & 20% senior. It means that if we sample 1000 students, then we must consider 300 freshers, 250 juniors and 200 seniors.

Quota Sampling



Non Probability Sampling

Types of Non Probability Sampling:

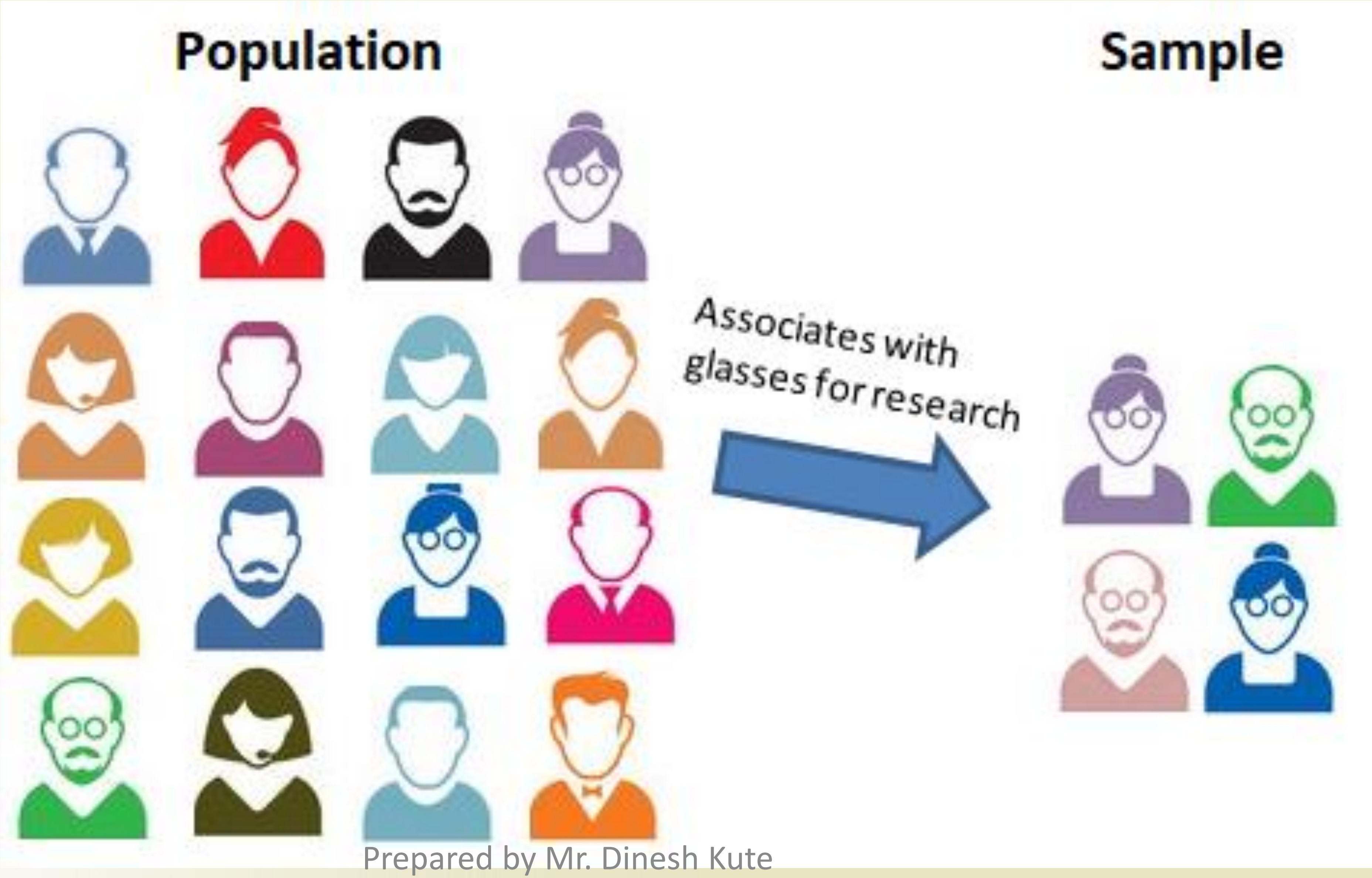
1) Purposive or Judgmental Sampling:

- In the judgmental sampling method, researchers select the samples based purely on the researcher's knowledge and credibility. In other words, researchers choose only those people who are fit to participate in the research study.

Example

- A group of researchers is interested in learning if the reason why people wear eyeglasses is to read books. Common sense tells us that the efforts of the research group should be focused entirely on people that indeed wear eyeglasses. This process is judgment sampling.

Judgmental Sampling



Non Probability Sampling

Types of Non Probability Sampling:

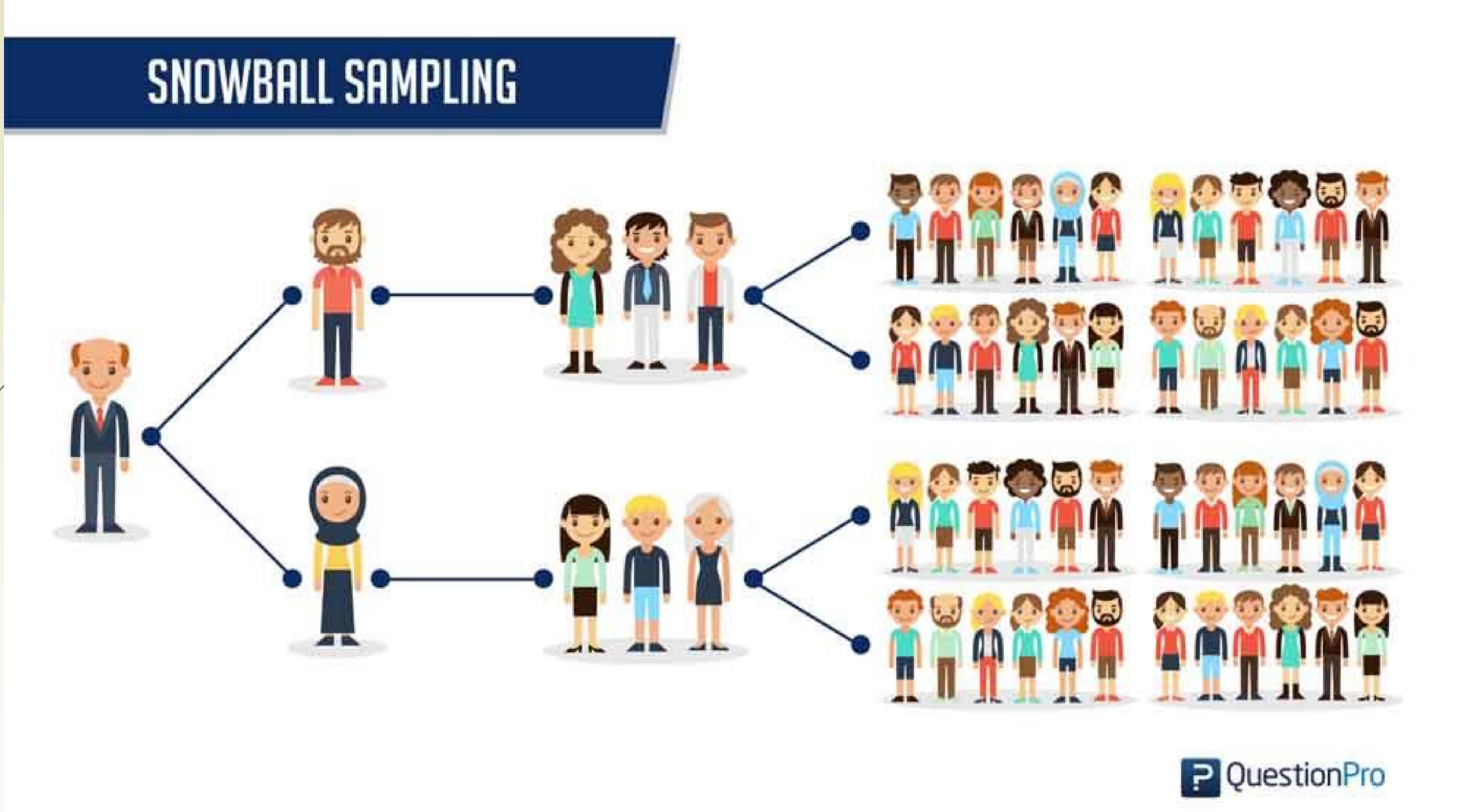
1) Snowball Sampling:

- ❑ Snowball sampling is also known as a chain-referral sampling technique.
- ❑ Snowball sampling is where research participants recruit other participants for a test or study

Example

- ❑ People who have many friends are more likely to be recruited into the sample. When virtual social networks are used, then this technique is called virtual snowball sampling.

Snowball Sampling



What is Hypothesis?

- A Hypothesis is the statement or an assumption about relationships between variables.

- A Hypothesis is a tentative explanation for certain behaviors, phenomenon or events that have occurred or will occur.

- A (statistical) hypothesis is a statement (postulate or assertion) concerning one or more parameters of the population (of variables or models).

- A hypothesis

Research Question?

Is there any relationship between “Use of smartphone” and “Headache”

Hypothesis

Use of smartphone leads to headache

OR

There is direct relationship between use smartphone and Headache.

- To do hypothesis test, set up two contradictory statements which are called as hypotheses.

- The one hypothesis which is assumed to be true is called Null hypothesis and other is called Alternative hypothesis.

- Conduct hypothesis test to check if the data support or does not support null hypothesis.

Types of Hypothesis

a) Null hypothesis (H_0)

b) Alternative Hypothesis (H_a)

Null Hypothesis

It is a hypothesis that says there is no statistically difference between two variables in the hypothesis

It is the hypothesis that researcher is trying to disprove.

Null hypothesis is denoted by H_0

Example:

Hypothesis: Flowers feed with club water is grow faster than flowers that feed with plane water.

H_0 =There is no statistically significant relationship between the type of water feed to the flowers and growth of flowers.

Alternative Hypothesis

It is inverse, or opposite, of the null hypothesis.

H_a =There is some statistically significant relationship between the type of water feed to the flowers and growth of flowers.

$H_0 : \mu = \bar{x}$	$H_a : \mu \neq \bar{x}$
$H_0 : \mu \leq \bar{x}$	$H_a : \mu > \bar{x}$
$H_0 : \mu \geq \bar{x}$	$H_a : \mu < \bar{x}$

Standard Notations

- ❑ μ = Mean of the population.
- ❑ $\bar{x}$ = Mean of the sample

Type I and Type II error

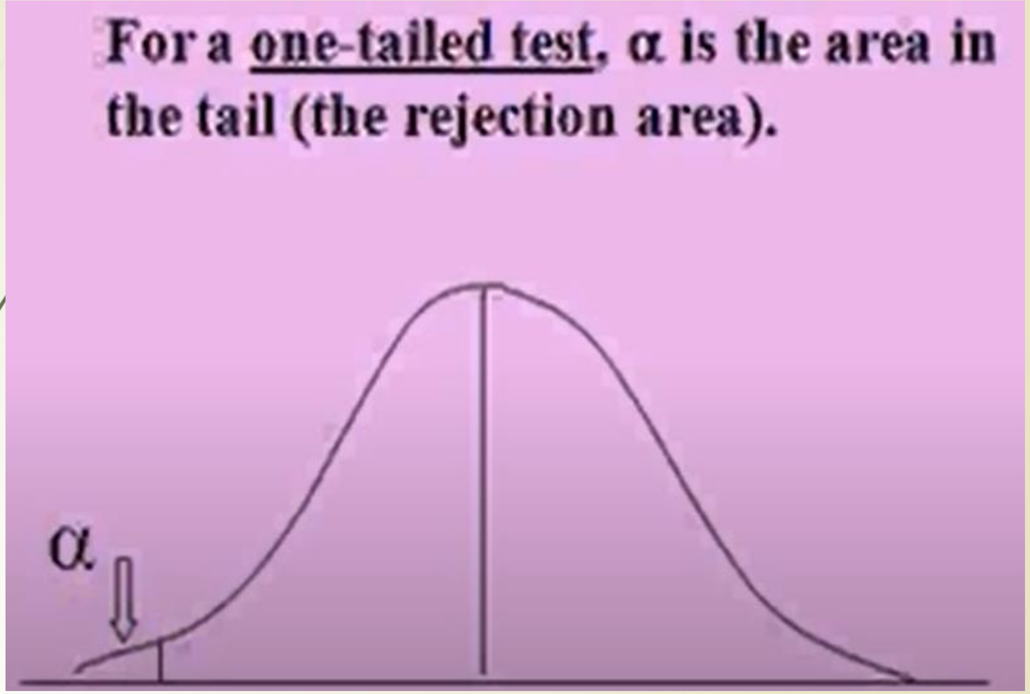
	Decision	
	Accept H_0	Reject H_0
H_0 (true)	Correct decision	Type I error (α error)
H_0 (false)	Type II error (β error)	Correct decision

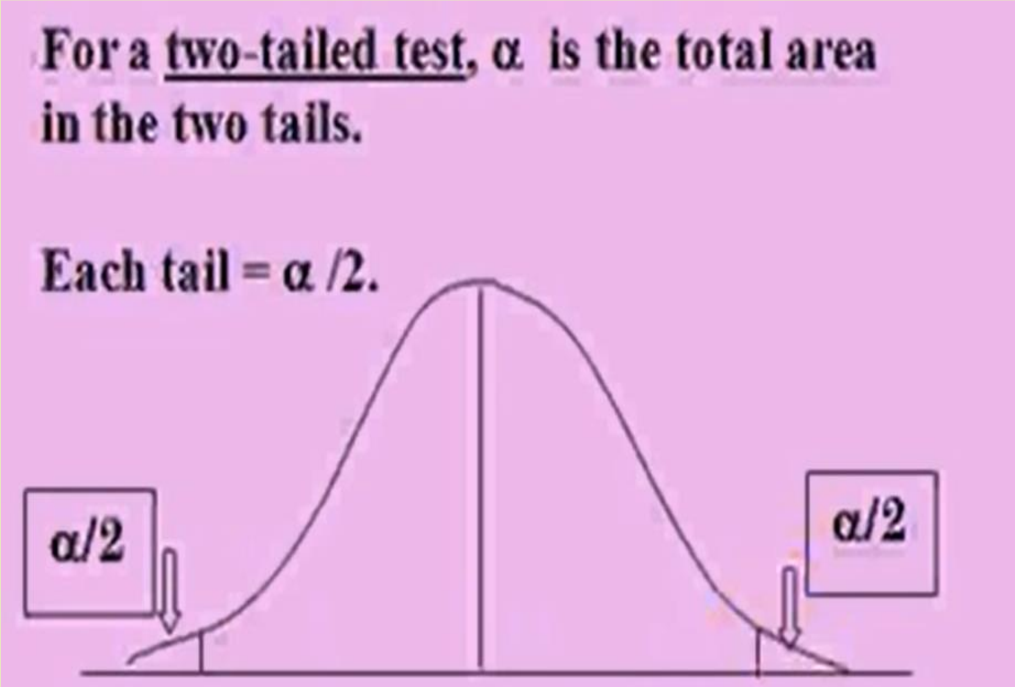
If type I error is fixed at 5 per cent, it means that there are about 5 chances in 100 that we will reject H_0 when H_0 is true. We can control Type I error just by fixing it at a lower level.

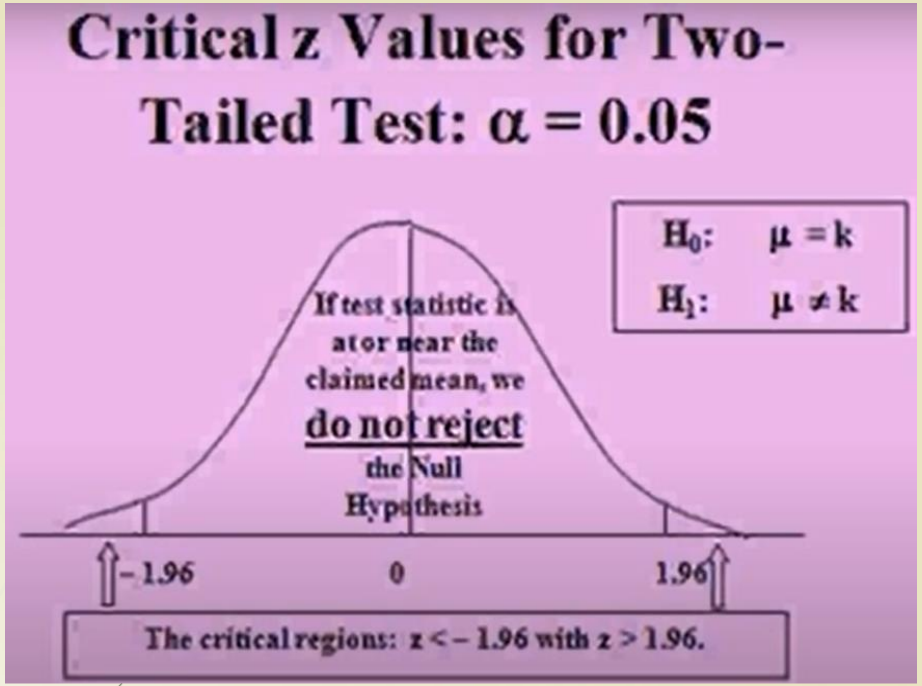
Level of Significance

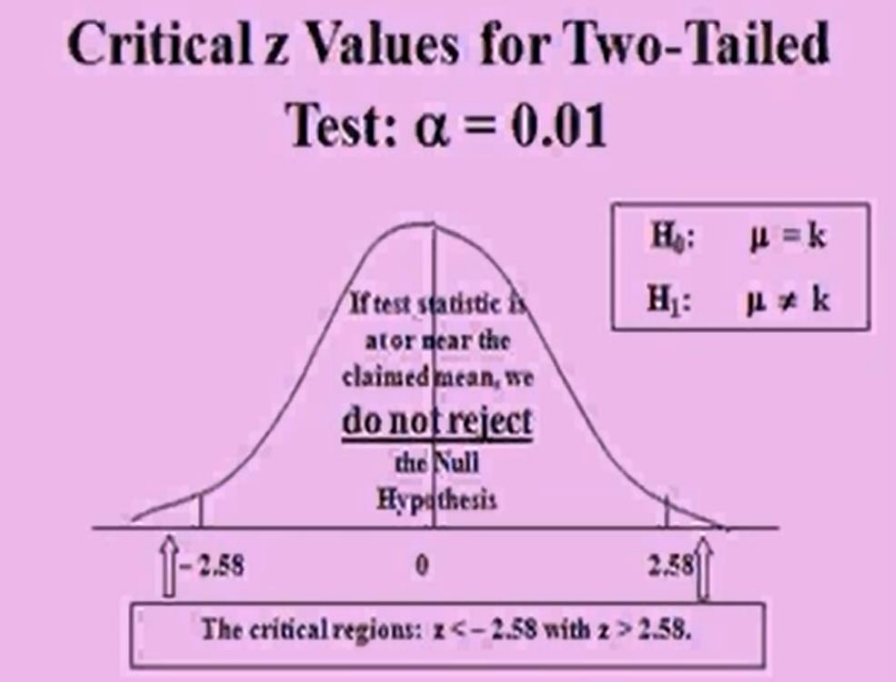
It is denoted by α, is the probability of rejecting null hypothesis when it is true (Type I error).

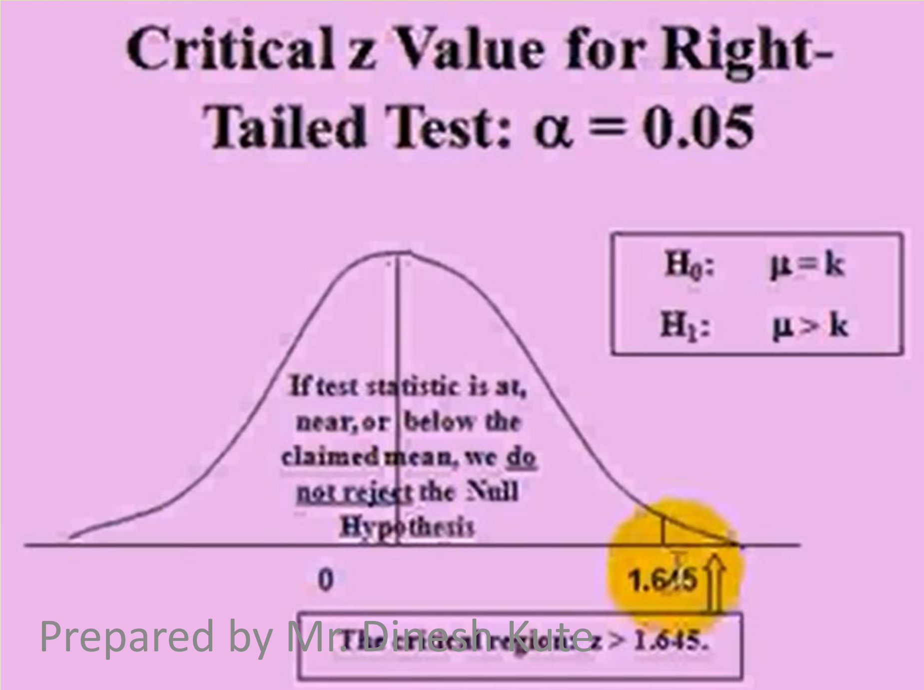
In general 5% or 1% level of significance can be considered.











Decision of Two tailed and one tailed can be taken from Alternative Hypothesis (H_a).

Degrees of Freedom:

Degrees of freedom $(d.f)$ of an estimate is the number of independent pieces of information that went into calculating the estimate.

- For one sample with n number of items:

$d.f. = n - 1$

- For two samples with n_1 and n_2 number of items

$d.f. = (n_1 + n_2) - 2$

Procedure of Testing Hypothesis:

- Set up a Hypothesis
- Set of suitable significance level
- Setting a test criterion
- Doing computations
- Making decisions

Hypothesis Test:

- Chi- square test
- T test
- Z Test
- ANOVA Test

All these tests are based on the assumption of normality i.e., the source of data is considered to be normally distributed.

In some cases the population may not be normally distributed, yet the tests will be applicable on account of the fact that we mostly deal with samples and the sampling distributions closely approach normal distributions.

Test 1 : Chi Square Test

- Chi-Square goodness of fit test is a non-parametric test that is used to find out how the observed value of a given phenomena is significantly different from the expected value.

- If the sample data do not fit the expected properties of the population that we are interested in, then we would not want to use this sample to draw conclusions about the larger population.

- **Example:**

Consider a experiment of tossing a coin 100 times.
Expected Value: 50/50 chance of landing heads or tails.
Observed Value: 50/50 or 60/40 or 90/10

Conclusion: If observed value is farther away from 50/50, the data is less fit to the expected value and we might conclude that this coin is not actually a fair coin.

Procedure:

600 $\downarrow$ 3   100 $\uparrow$ 600

Step 1: Set up the hypothesis

Null hypothesis (H_0): In Chi-Square goodness of fit test, the null hypothesis assumes that there is no significant difference between the observed and the expected value.

Alternative hypothesis (H_a): In Chi-Square goodness of fit test, the alternative hypothesis assumes that there is a significant difference between the observed and the expected value.

Step 2: Compute the value of Chi-Square goodness of fit test using the following formula:

$$\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$$

χ^2

$o_i \rightarrow$ observed

$e_i \rightarrow$ expected

Step 3:

- If $\chi^2_{df,\alpha}(Table\ Value) > \chi^2(Calculated\ Value)$ then we accept Null hypothesis.

- If $\chi^2_{df,\alpha}(Table\ Value) < \chi^2(Calculated\ Value)$ then we reject Null hypothesis.

Where, n =No. of observations, $df = n - 1$: degree of freedom, α : level of significance

$\chi_{df,\alpha} =$

Examples:

Q.1 A nationalized bank utilizes four windows to give fast service to the customers. On a particular day 800 customers were observed. They were given service at the different windows as follows: (Given $\chi^2_{1,0.05} = 3.841$)

Expected	Window Number	Observed Number of customers
200	1	150
200	2	250
200	3	170
200	4	230

Expected

Solution:

There is no significan difference betn obs. & Expt

Let H_0 =~~Customers are uniformly distributed over windows~~.

Under H_0, the expected frequencies are

Window Number	Expected Number of customers
1	200
2	200
3	200
4	200

$\chi^2_{df,\alpha}$

$\chi^2_{3,0.01}$

$df = n-1$

0.01

1%

5%

Not good fit

By Chi square test,

$$\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} = \frac{(-50)^2}{200} + \frac{50^2}{200} + \frac{(-30)^2}{200} + \frac{30^2}{200} = 34 > \chi^2_{1,0.05} = 3.841$$

∴ We reject Null hypothesis (H_0) at 5% level of significance.

Window Number	Observed Number of customers
1	150
2	250
3	170
4	230

Examples:

Q.2 Among 64 offsprings of a certain cross between guinea pigs 34 were red, 10 were black and 20 were white. According generic model, these numbers should be in the ratio 9:3:4. (Given $\chi^2_{2,0.05} = 5.991$)

Solution:

Let H_0 =The offsprings red, black and white in the colour are in the ratio 9:3:4.

Under H_0, the expected frequencies are

Observed Frequencies	34 o_1	10 o_2	20 o_3
Expected Frequencies	$\frac{9}{16} \times 64 = 36$ e_1	$\frac{3}{16} \times 64 = 12$ e_2	$\frac{4}{16} \times 54 = 16$ e_3

By Chi square test, $= \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \frac{(o_3 - e_3)^2}{e_3}$

$$\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} = \underbrace{\frac{(-\cancel{4}\,2)^2}{36} + \frac{(-2)^2}{12} + \frac{(4)^2}{16}} = 1.4444 < \chi^2_{1,0.05} = 5.991$$

$\therefore$ We accept Null hypothesis (H_0) at 5% level of significance.

9 : 3 : 4

Observed Frequencies	34 o_1	10 o_2	20 o_3
Expected Frequencies	$\frac{9}{16} \times 64 = \mathbf{36}$ e_1	$\frac{3}{16} \times 64 = \mathbf{12}$ e_2	$\frac{4}{16} \times 54 = \mathbf{16}$ e_3

A bank utilizes three teller windows to render service to the customer. On a particular day 600 customer were served. If the customers are uniformly distributed over the counters. Expected numbers of customer served on each counter is (2)

(A) 100 (B) 200 (C) 300 (D) 150

600

3 3 3

200 200 200

2. 200 digits are chosen at random from a set of tables. The frequencies of the digits are as follows :

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	18	19	23	21	16	25	22	20	21	15

$\rightarrow$ 200

The expected frequency and degree of freedom for uniform distribution is (2)

(A) 20 and 10 (B) 21 and 9 (C) 20 and 9 (D) 15 and 8

$E \cdot F = 20$

$n = 10$

$df = 10 - 1$

$df = 9$

$\chi_{df,\alpha}$

3. In experiment on pea breeding, the observed frequencies are 222, 120, 32, 150 and expected frequencies are 323, 81, 81, 40, then χ_3^2 has the value (2)

(A) 382.502 ✓ (B) 380.50 (C) 429.59 (D) 303.82

$o_1 = 222, \quad o_2 = 120, \quad o_3 = 32, \quad o_4 = 150$

$e_1 = 323, \quad e_2 = 81, \quad e_3 = 81, \quad e_4 = 40$

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \frac{(o_3 - e_3)^2}{e_3} + \frac{(o_4 - e_4)^2}{e_4}$$

4. If observed frequencies O_1, O_2, O_3 are 5, 10, 15 and expected frequencies e_1, e_2, e_3 are each equal to 10, then χ_2^2 has the value (2)

(A) 20 (B) 10 (C) 15 (D) 5 ✓

$o_1 = 5, \quad o_2 = 10, \quad o_3 = 15$

$e_1 = 10, \quad e_2 = 10, \quad e_3 = 10$

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \frac{(o_3 - e_3)^2}{e_3}$$

5. Number of books issued on six days of the week, excluding Sunday which is holiday are given as 120, 130, 110, 115, 135, 110 and expectation is 120 books on each day, then χ^2_5 is (2)

(A) 2.58 (B) 3.56 (C) 6.56 (D) 4.58

$o_1 = 120,\ o_2 = 130,\ o_3 = 110,\ o_4 = 115,\ o_5 = 135,\ o_6 = 110$

$e_i = 120 \quad \forall\ i = 1 \text{ to } 6$

$\chi^2 =$

6. A coin is tossed 160 times and following are expected and observed frequencies for number of heads

No. of heads	0	1	2	3	4
Observed frequency	17 o_1	52 o_2	54 o_3	31	6
Expected Frequency	10 e_1	40 e_2	60 e_3	40	10

Then χ^2_4 is

(A) 12.72 (B) 9.49 (C) 12.8 (D) 9.00

7. Among 64 offspring's of a certain cross between guinea pig 34 were red, 10 were black and 20 were white. Acceding to genetic model, these number should in the ratio 9 :3 : 4. Expected frequencies in the order (2)

(A) 36, 12, 16 (B) 12, 36, 16 (C) 20, 12, 16 (D) 36, 12, 25

$\frac{9}{16} \times 64 =$ $\frac{3}{16} \times 64 =$ $\frac{4}{16} \times 64 =$

8. A sample analysis of examination results of 500 students was made. The observed frequencies are 220, 170, 90 and 20 and the numbers are in the ratio 4 : 3 : 2 : 1 for the various categories. Then the expected frequencies are

(A) 150, 150, 50, 25 (B) 200, 100, 50, 10 (C) 200, 150, 100, 50 (D) 400, 300, 200, 100

$\frac{4}{10} \times 500 =$ $\frac{3}{10} \times 500,$ $\frac{2}{10} \times 500,$ $\frac{1}{10} \times 500 =$

9. In experiment on pea breeding, the observed frequencies are 222, 120, 32, 150 and the theory predicts that the frequencies should be in proportion 8 : 2 : 2 : 1. Then the expected frequencies are (2)

(A) 323, 81, 40, 81 (B) 81, 323, 40, 81 (C) 323, 81, 81, 40 (D) 433, 81, 81, 35

$e_1 = \frac{8}{13} \times 524$, $\frac{2}{13} \times 524$, $\frac{2}{13} \times 524$

$\frac{1}{13} \times 524 =$

Z test

T test

Chi square test

sample ...

Test 2 : Student’s T distribution (T Test)

- T Test is used for sample size $(n \leq 30)$ and standard deviation of population is unknown

- T Test (Students T Test) is a statistical significance test that is used to compare the means of two groups and determine if the difference in means is statistically significant or not.

Types of T Test

a) One-Sample Student’s t-test:

(Example: Determine whether the average lifespan of a specific town is different from the country average

b) Two Sample Independent T-Test

(Example: Test of the mean weight of mangoes from Farm A equals mean weight of mangoes from Farm B.)

1. T test :

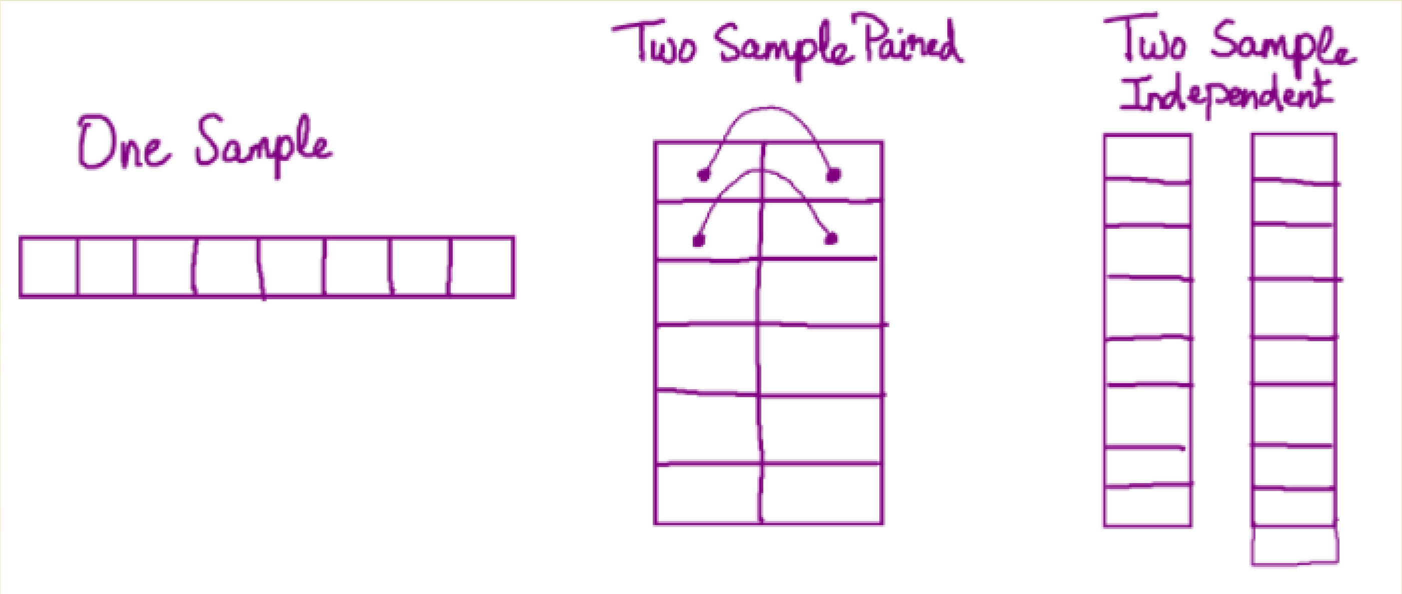
   H_0 = Accept / Reject

   1. $n \leq 30$
   2. Unknown standard deviation / (variance) of population

   H_0 = There is no significant difference betn mean value of sample and population

2. Z test
   1. $n \geq 30$
   2. given value of S.D (variance of population

t Test σ	Z test
① $\underline{\underline{n}} \le 30$ ② S.D is unknown (population)	① $n \geq \underline{\underline{30}}$ ② S.D. is known (population)
H_0 = There is no significant difference betn mean value of sample and population ($H_0 : \bar{x} = \mu$)	H_0 = "same" $\alpha \to$ 5% 1% $H_a = \bar{x} \neq \mu$
① one sample $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$, $s = \sqrt{\dfrac{1}{n-1} \sum (x_i - \bar{x})^2}$	① one sample $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ ✓ $\bar{x} \to$ sample $\mu \to$ popul $s \to$ sample $\sigma \to$ popul
t_α (table) > |t| ⇒ Accept H_0 t_α (table) < |t| ⇒ Reject H_0	for 5% los: $-1.96 \le z \le 1.96$ ⇒ Accept H_0 for 1% los $-2.58 \le z \le 2.58$ ⇒ Accept H_0



a) One-Sample Student's t-test:

Procedure:

Step 1: Set up the hypothesis

Null hypothesis (H_0): Sample Mean ($\bar{x}$) = Population Mean (μ)

Step 2: Compute T statistics

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}, \text{ where } s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

$\bar{x}$: Mean of sample
μ : Mean of the population
n : No. of elements in Sample
s : Standard deviation of Sample

Step 3:

- If $t_\alpha (Table\ Value) > |t|\ (Calculated\ Value)$ then we accept Null hypothesis.

- If $t_\alpha (Table\ Value) < |t|\ (Calculated\ Value)$ then we reject Null hypothesis.

Where, n =No. of observations, $df = \underline{\underline{n-1}}$: degree of freedom, α : level of significance

$n \rightarrow$ no. of elements in sample

$df = \underline{\underline{n-1}}$

Examples:

Q.1 The manufacturer of a certain company make an electric bulbs, claims that his bulbs have a mean life of 25 months with standard deviation of 5 months. A random sample of 6 such bulbs gave the following values,

Life in month	24	26	30	20	20	18

Can you regard the producer's claim to be valid at 1% level of significance? (Given: $t_{0.01} = 4.032$)

$\mu = 25$, $\sigma = 5$, $n = 6$

$\bar{x}$

Solution: 1%

Let H_0 =There is no significantly difference between mean life of population and sample. Im

μ

$H_0 \to$ True

By T test of one sample

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

x	$(x-\bar{x})^2$
24	1
26	9
30	49
20	9
20	9
18	25
138	**102**

$$\bar{x} = \frac{\sum x}{n} = \frac{138}{6} = 23$$

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

1m

$$s = \sqrt{\frac{102}{5}} = 4.517$$

$$\therefore t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{23-25}{4.157/\sqrt{6}} = 1.084 < t_{0.01} = 4.032$$

1m

2m | 3m

1m

∴ We accept Null hypothesis (H_0) at 1% level of significance.

Examples:

Q.2 The life time of item for a random sample of 10 from a large consignment gave the following data;

sample

Item	1	2	3	4	5	6	7	8	9	10
Life in thousand hrs	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average life time of item is 4,000 hours? (Given: $t_{0.01} = 2.226$)

population

Solution:

Let H_0 =There is no significantly difference between mean life of population and sample.

$n = 10 \leq 30$

By T test of one sample

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

x	$(x-\bar{x})^2$
4.2	0.04
4.6	0.04
3.9	0.25
4.1	0.09
5.2	0.64
3.8	0.36
3.9	0.25
4.3	0.01
4.4	0
5.6	1.44
44	3.12

$$\bar{x} = \frac{\sum x}{n} = \frac{44}{10} = 4.4$$

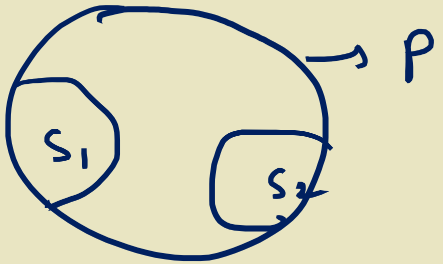
$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{3.12}{9}} = 0.589$$

$$\therefore t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{4.4-4}{0.589/\sqrt{10}} = 2.148$$

$$< t_{0.01} = 2.226$$

$\therefore$ We accept Null hypothesis (H_0) at 1% level of significance.

a) Two Sample Independent T-Test:



$H_0 = \bar{x}_1 = \bar{x}_2$

Procedure:

Step 1: Set up the hypothesis

Null hypothesis (H_0): Sample Mean ($\overline{x_1}$) = Sample Mean ($\overline{x_2}$)

There is no significantly difference between Mean values of two independent samples.

S1 ⟷ S2

10 15 60

Step 2: Compute T statistics

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}}$$

$$= \frac{\bar{x}_1 - \bar{x}_2}{s} \times \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

Where,

$$s = \sqrt{\frac{\sum\left(x_1 - \overline{x}_1\right)^2 + \sum\left(x_2 - \overline{x}_2\right)^2}{n_1 + n_2 - 2}}$$

$\overline{x_1}$: Mean of the first sample

$\overline{x_2}$: Mean of the sample sample

n_1: Size of first sample

n_2: Size of second sample

s : Combined standard deviation

$$\boxed{\textbf{d.f.} = n_1 + n_2 - 2} \rightarrow \text{degree of freedom}$$

Im

Examples:

Q.1 Two types of drugs were used on 5 & 7 patients for reducing their weight. The drug A was imported and Drug B is indigenous. The decrease in weight after using the drugs for six months was as follows;

Drug A	10	12	13	11	14			5
Drug B	8	9	12	14	15	10	9	7

Is there a significant difference in the efficiency of the drugs? If ~~not, which drug should you buy~~. (Given: $t_{0.05} = 2.223$)

$H_0 \rightarrow$ True

Solution:

Let H_0 =There is no significantly difference between efficiency of drug A and drug B

By T test of two independent sample

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Rightarrow t = \frac{\bar{x}_1 - \bar{x}_2}{s} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Given $n_1 = 5, n_2 = 7$

x_1	$(x_1 - \overline{x_1})^2$
10	4
12	0
13	1
11	1
14	4
60	**10**

$$\bar{x}_1 = \frac{60}{5} = 12$$

$$\bar{x}_2 = \frac{77}{7} = 11$$

x_2	$(x_2 - \overline{x_2})^2$
8	9
9	4
12	1
14	9
15	16
10	1
9	4
77	**44**

$$S = \sqrt{\frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

combined S.D

$$= \sqrt{\frac{10+44}{5+7-2}} = \sqrt{5.4} = 2.323$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$= \frac{12 - 11}{2.323} \times \sqrt{\frac{5 \times 7}{5 + 7}}$$

$$= \frac{1}{2.323} \times \sqrt{\frac{35}{12}}$$

$$= 0.731 \quad < t_{0.01} = 2.223$$

∴ We **accept** Null hypothesis (H_0) at 1% level of significance.

As Drug B is indigenous, it should be referred.

1 samp.

2 sm In

t test

1%

Test 3 : Z Test

- Z Test is used for sample size $(n \geq 30)$ and standard deviation of population is known

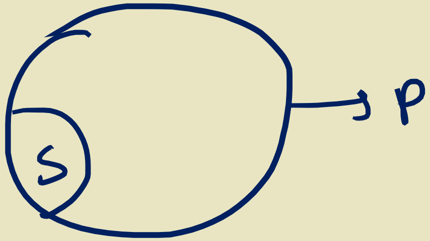
- Z test is statistical hypothesis used to determine whether two samples means are different when variance (or standard deviation) is known.

Types of T Test

a) One-Sample Z Test:

b) Two Sample Z Test

a) One-Sample Z Test:



Procedure:

Step 1: Set up the hypothesis

Null hypothesis (H_0): Sample Mean ($\bar{x}$) = Population Mean (μ)

Step 2: Compute T statistics

$\sigma \to s$

One Sample: $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Two Independent Samples: $z = \dfrac{\overline{x_1}-\overline{x_2}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

$\sigma \to s$

$\bar{x}$: Mean of sample
μ : Mean of the population
n : No. of elements in Sample
σ : Standard deviation of population

Step 3:

1) For 5% level of significance: (Two tailed)

If $-1.96 \le z \le 1.96$ then null hypothesis accepted otherwise rejected.

1) For 1% level of significance: (Two tailed)

If $-2.58 \le z \le 2.58$ then null hypothesis accepted otherwise rejected.

Step 3:

1) For 5% level of significance: (Two tailed)

If $-1.96 \leq z \leq 1.96$ then null hypothesis accepted otherwise rejected.

2) For 1% level of significance: (Two tailed)

If $-2.58 \leq z \leq 2.58$ then null hypothesis accepted otherwise rejected.

3) For 5% level of significance towards right (One tailed)

If $z \leq 1.645$ then null hypothesis is accepted.

4) For 5% level of significance towards left (One tailed)

If $z \geq -1.645$ then null hypothesis is accepted.

Examples:

A sample of 400 male students is found to have a mean height 67.47 inches. Can it be reasonably regarded as a sample from a large population with mean height 67.39 inches and standard deviation 1.30 inches? Test at 5% level of significance. $(-1.96 \le z \le 1.96)$

Sol^n: $n = 400$, $\bar{x} = 67.47$, $\mu = 67.39$, $\sigma = 1.30$

Let H_0 = There is no significancy difference betn mean height of sample and population.

∴ By Z test

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{67.47 - 67.39}{1.30/\sqrt{400}}$$

$$= \frac{0.08}{0.065} = 1.2308 < 1.96$$

∴ We accept the Null hypothesis.

Examples:

Q.1 A herd of 1,500 steer was fed a special high-protein grain for a month. A random sample of 29 were weighed and had gained an average of 6.7 pounds. If the standard deviation of weight gain for the entire herd is 7.1, test the hypothesis that the average weight gain per steer for the month was more than 5 pounds.. (Given: $z_{0.05} = 1.96$)

$\mu = 5$

$\bar{x} > 5$

$H_0: \bar{x} = \mu$

$H_a = \bar{x} \neq \mu = 5$

One

two

Solution:

Let $H_0: \mu = 5$ and $\boxed{\cancel{H_a: \mu > 5}}$

By z test H_0 $\bar{x} = \mu = 5$

$\bar{x} > 5$ ×

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{6.7-5}{7.1/\sqrt{29}} = 1.289$$

$\bar{x} > \mu = 5$ $\quad$ $\bar{x} < \mu = 5$

$\bar{x} > 5$ $\quad$ $\bar{x} < 5$ ×

The calculated of z is between -1.96 and 1.96.

$\therefore$ We accept Null hypothesis (H_0) at 5% level of significance.

Examples:

Q.2 In national use, a vocabulary test is known to have a mean score of 68 and a standard deviation of 13. A class of 19 students takes the test and has a mean score of 65. (Given: $z_{0.05} = 1.96$)

Solution:

Let $H_0: \mu = 68$ and $H_a: \mu \neq 68$

By z test

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad = \frac{65-68}{13/\sqrt{19}} = -1.006$$

The calculated of z is between -1.96 and 1.96

$\therefore$ We accept Null hypothesis (H_0) at 5% level of significance.

Examples:

- The mean of a certain production process is known to be 50 with a standard deviation of 2.5. The production manager may welcome any change is mean value towards higher side but would like to safeguard against decreasing values of mean. He takes a sample of 12 items that gives a mean value of 48.5. What inference should the manager take for the production process on the basis of sample results? Use 5 per cent level of significance for the purpose.

“Thank You”