Mathematics for Computer Applications – Unit 4

Course Code: M23DE0101 – Academic year 2024-2025, I Semester MCA (Odd Semester)

School of Computer Science and Applications

Dr M Vinayaka Murthy

Professor

LECTURE -1

Agenda

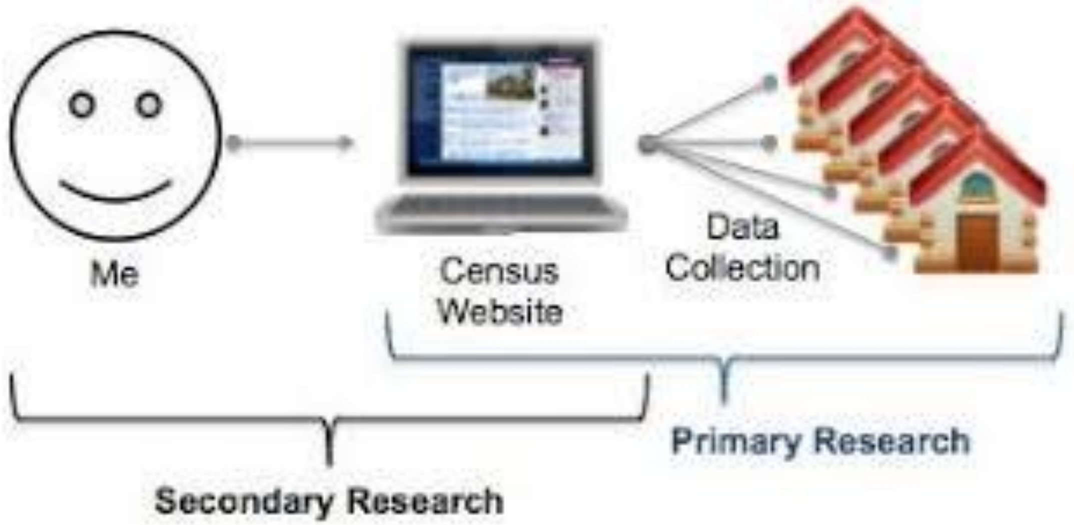
- Unit 4 - Content
- Introduction
- Sampling
- Sample Error
- Types of sampling
- Parameter and Statistics

UNIT 4 - CONTENT

UNIT 4: Hypothesis Testing: Introduction Sampling, Sampling distribution, one and two tailed test, Test of significance, (mean, difference of means), confidence interval 1% and 5% level of significance - Design of Experiments, one way classification, two way classification, ANOVA.

INTRODUCTION- SAMPLING

From the definition of statistics, we know that any statistical investigation is concerned with the study of a collection (objects)



Population or Universe

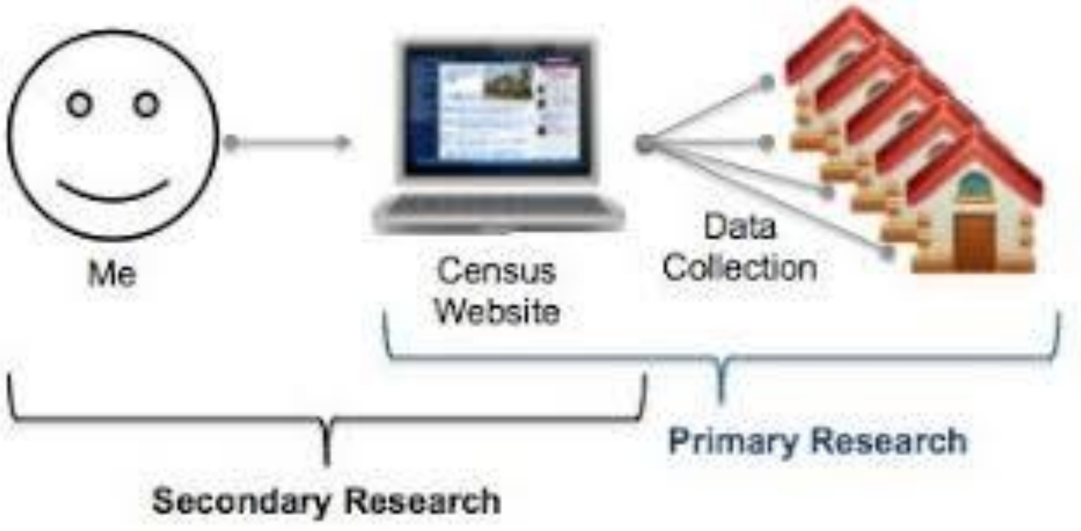
Such a group or collection of objects is called population or Universe.

Example: Heights and weights of the students, marks scored in different subjects etc.,

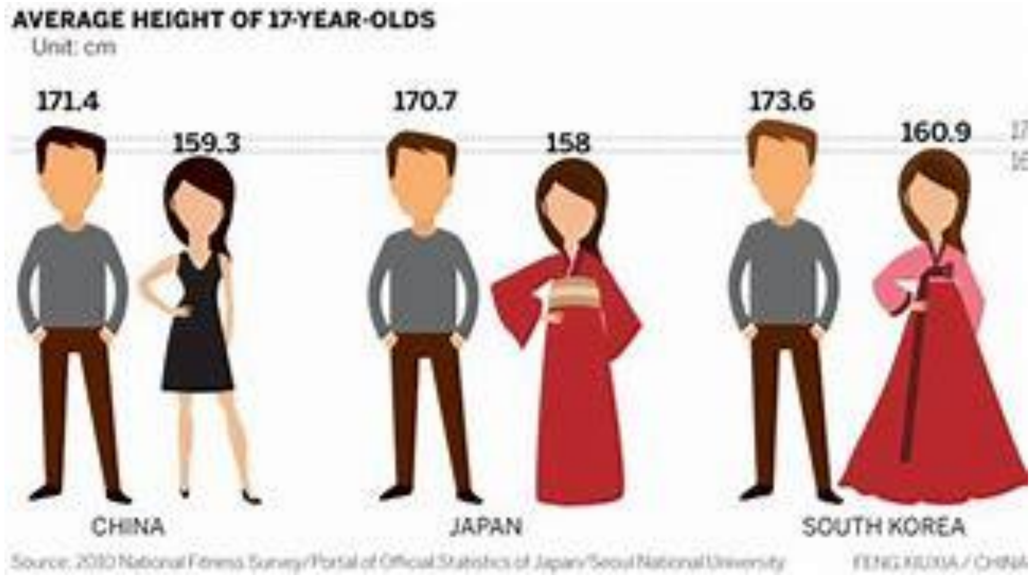
INTRODUCTION- SAMPLING CONTD.

Definition: A finite set of a population is called a sample (Sampling)

Population or Universe



Sample



The number of objects in a sample is called the sample size

SAMPLE ERROR

Definition: In order to determine some population characteristics, the objects in the sample are observed and
the sample characteristics are used to approximately estimate the same for the entire population.
The inherent and unavoidable error in any such approximation is known as sample error

TYPES OF SAMPLING

The investigation based on a sample (a representative portion of the population). is called **sample survey.** Suppose n units are selected from the population, these selected units form a sample of size n. If these units are selected by providing equiprobability to all the units of the population, the sample is called **random sample.**

Some of the important types of sampling are

i) **Purposive sampling**

ii) **Random sampling**

iii) **Simple sampling**

iv) **Stratified sampling**

PARAMETER AND STATISTICS

For a variable in the population, suppose we find constants such as mean, standard deviation, etc., these constants are called **parameters** of the population. On the other hand, if we find mean, standard deviation, etc., of the sample, they are called **statistics.**

Thus, mean height of students of a college is parameter. Whereas, mean height of 50 randomly selected students of the college is a statistic.

PARAMETER AND STATISTICS CONTD.

Parameter is a statistical constant of the population. Statistic is a function of the sample values.

SUMMARY

- Unit 4 - Content
- Introduction
- Sampling
- Sample Error
- Types of sampling
- Parameter and Statistics

QUIZ

1. Within the R A Fisher approach which of the following is not true: (one correct choice)

   a. P value =Probability of the observed data (statistic) or more extreme given that the null hypothesis is true
   b. P value is interpreted on a individual experiment basis
   c. The critical value is specific to a experiment
   d. P value is interpreted as evidence against the null hypothesis, lower values greater strength of evidence
   e. Decision rules form a major component in Fishers approach

8. Which of the following applications is most frequently used to carry out a power analysis? (one correct choice)

   a. SPSS
   b. Epi Info
   c. Gpower
   d. Excel
   e. Word

9. When carrying out a Statistical Power analysis a graph of the following variables is most frequently produced: (one correct choice)

   a. Effect size, power
   b. Sample size, power
   c. P value, power
   d. Effect size, α
   e. β, α

LECTURE -2, SAMPLING DISTRIBUTION

RECAP

- Unit 4 - Content
- Introduction
- Sampling
- Sample Error
- Types of sampling
- Parameter and Statistics

OBJECTIVE

- Sampling Distribution and Standard Error
- Standard Error Problems
- Problems for Practice

SAMPLING DISTRIBUTION AND STANDARD ERROR

Definition: Suppose a sample of size n is drawn from a population and the sample mean $\bar{x}$ is calculated. From the population, many such samples $\bar{x}$ of the same size can be drawn. For each of these samples can be computed. And so, there can be many values of $\bar{x}$.

=AVERAGE(B2:B11)

	B	C
	Yield of potatoes with fertiliser A (kg)	Yield of potatoes wit fertililser B (kg)
	27	28
	20	19
	16	18
	18	21
	22	24
	19	20
	23	25
	21	27
	17	29
	19	21
an	20.2	23.2

Definition: The distribution of values of a statistic for different samples of the same size is called sampling distribution of the statistic.

SAMPLING DISTRIBUTION AND STANDARD ERROR CONTD.

Suppose these different values are tabulated in the form of a frequency distribution, the resulting distribution is called **Sampling distribution of** $\bar{x}$. The standard deviation of this sampling distribution is called **Standard Error (S.E.)**

Definition: Standard Error (S.E.) of a statistic is the standard deviation of the sampling distribution of the statistic.

STANDARD ERROR PROBLEMS CONTD.

Statistic	Mean	Standard Error
$\bar{x}$ (sample mean)	μ	$\frac{\sigma}{\sqrt{n}}$
$\bar{x}_1 - \bar{x}_2$ (difference of means)	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
p (sample proportion)	P	$\sqrt{\frac{PQ}{n}}$
$p_1 - p_2$ (difference of proportions	$P_1 - P_2$	$\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$
$p_1 = p_2 = p$ (equal proportions)	P	$\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

STANDARD ERROR PROBLEMS

Consider a population whose mean is μ and standard deviation is σ. Let a random sample of size n be drawn from this population. Then, the sampling distribution of $\bar{x}$ has mean μ and standard error $\frac{\sigma}{\sqrt{n}}$

Let a random sample of size n_1 be drawn from a population whose mean is μ_1 and standard deviation is σ_1. Also, let a random sample of size n_2 be drawn from another population whose mean is μ_2 and standard deviation is σ_2. Let $\bar{x}_1$ be the mean of the first sample and $\bar{x}_2$ be the mean of the second sample. Then, $(\bar{x}_1 - \bar{x}_2)$

has mean $(\mu_1 - \mu_2)$ and standard error $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

STANDARD ERROR PROBLEMS CONTD.

In a population, let P be the proportion of units which 'posses the attribute'. From such a population, suppose a random sample of size n is drawn. Let x of these n units belong to the class 'possess the attribute'. Then $p = \frac{x}{n}$ is the sample proportion of the attribute.

Here, $p = \frac{x}{n}$ has mean P and standard error $\sqrt{\frac{PQ}{n}}$ where $Q = 1 - P$.

STANDARD ERROR PROBLEMS CONTD.

Let a random sample of size n_1 be drawn from a population with proportion P_1 of an attribute. Let X_1 units in the sample possess the attribute. Then, the sample proportion is $p_i = \frac{x_1}{n_1}$. Let a random simple of size n_2 be drawn from a population with proportion P_2 the attribute. Let x_2 units in this sample possess the attribute. Then, the sample proportion is $p_2 = \frac{x_2}{n_2}$. Here, the difference of the sample proportions $(p_1 - p_2)$ has mean $(p_1 - p_2)$ and standard error

$\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$ where $Q_1 = 1 - P_1$ and $Q_2 = 1 - P_2$.

Here, if $P_1 = P_2 = P$, the standard error is $\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

STANDARD ERROR PROBLEMS CONTD.

Example 1. The mean and standard deviation of weight of boys of a college are 47 kgs. and 3.1 kgs. respectively. The mean and standard deviation of weight of girls of the college are 45 kgs. and 2.8 kgs. From the college, 16 boys and 9 girls are randomly selected.

i) Find the mean and standard deviation (S.E.) of mean weight of the 16 selected boys.

ii) Find the mean and standard deviation of mean weight of the 9 selected girls.

iii) Find the mean and standard deviation of the difference of the mean weight of the selected boys and the mean weight of the selected girls.

STANDARD ERROR PROBLEMS CONTD.

Solution : Here, $\mu_1 = 47$ kgs. $\sigma_1 = 3.1$ kgs., $n_1 = 16$,

$\mu_2 = 45$ kgs., $\sigma_2 = 2.8$ kgs. and $n_2 = 9$

Let $\bar{x}_1$ be the mean weight of the selected boys.

Let $\bar{x}_2$ be the mean weight of the selected girls.

i) Mean of $\bar{x}_1 = \mu_1 = 47$ kgs.

$$\text{S. E } (\bar{x}_1) = \frac{\sigma_1}{\sqrt{n_1}} = \frac{3.1}{\sqrt{16}} = 0.775 \text{ kgs.}$$

ii) Mean of $\bar{x}_2 = \mu_2 = 45$ kgs.

$$\text{S. E } (\bar{x}_2) = \frac{\sigma_2}{\sqrt{n_2}} = \frac{2.8}{\sqrt{9}} = 0.933 \text{ kgs.}$$

iii) Mean of $(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$

$$= 47 - 45 = 2 \text{ kgs.}$$

$$S.E.(x_1 - x_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$= \sqrt{\frac{(3.1)^2}{16} + \frac{(2.8)^2}{9}}$$

$$= 1.213 \text{ kgs.}$$

STANDARD ERROR PROBLEMS CONTD.

Example 3. In city A, 32% voters voted for Congress. In city B, 29% voters voted for Congress.

i) Among 70 randomly selected voters from city A, if p_1 is the proportion of voters who voted for Congress, find the standard error of p_1.

ii) Among 60 randomly selected voters from city B, if p_2 is the proportion of voters who voted for Congress, find the standard error of p_2.

iii) Find the mean and standard error of $(p_1 - p_2)$.

Solution :

Here, $P_1 = \frac{32}{100} = 0.32$ and $P_2 = \frac{29}{100} = 0.29$

$n_1 = 70$ and $n_2 = 60$

i) $$S.E(p_1) = \sqrt{\frac{P_1 Q_1}{n_1}}$$

$$= \sqrt{\frac{0.32 \times 0.68}{70}}$$

$$= 0.05575$$

ii) $$S.E.(p_2) = \sqrt{\frac{P_2 Q_2}{n_2}}$$

$$= \sqrt{\frac{0.29 \times 0.71}{60}}$$

$$= 0.05858$$

STANDARD ERROR PROBLEMS CONTD.

iii) $$S.E.(p_1 - p_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$$

$$= \sqrt{\frac{0.32 \times 0.68}{70} + \frac{0.29 \times 0.71}{60}}$$

$$= 0.08087$$

SUMMARY

- Sampling Distribution and Standard Error
- Standard Error Problems
- Problems for Practice

PROBLEM FOR PRACTICE

Example 5. The poportion of women in a society is 0.48. Among 64 randomly selected people of the society, let p_1 be the proportion of women. In another selection of 86 people, let p_2 be the proportion of women. Find –

i) Standard error of p_1

ii) Standard error of p_2

iii) Standard error of the difference $(p_1 - p_2)$

LECTURE -3, ONE AND TWO TAILED TEST

RECAP

- Sampling Distribution and Standard Error
- Standard Error Problems
- Problems for Practice

OBJECTIVE

- Testing of Hypotheses
- Statistical Hypotheses
- Null Hypotheses
- Critical Region
- Type I and Type II error
- level of significance
- One and Two tailed test

TESTING OF HYPOTHESES

The 'theory of statistical inference' has two branches, namely, 'theory of estimation' and 'theory of testing of hypotheses'. Theory of estimation, is a technique of estimation the population parameters by using simple statistics. Whereas, in testing of hypotheses, validity of presumptions regarding the parameters of the population are verified with the help of sample statics.

Definition : Hypothesis is a statement about the values of the population parameter. It is made on the basis of the information obtained by experimentation. Testing of hypothesis is a procedure for deciding whether to accept or reject the hypothesis. Procedures which enable us to decide whether to accept or reject hypothesis are called tests of hypothesis, also known as tests of significance.

STATISTICAL HYPOTHESES

A statistical hypothesis is an assertion regarding the statistical distribution of the population. It is a statement regarding the parameters of the population.

Statistical hypothesis is denoted by H.

EXAMPLES

1. H: The population has mean $\mu = 25$
2. H: The population is normally distributed with mean $\mu = 25$ and standard deviation $\sigma = 2$.

NULL HYPOTHESES

In a test procedure, to start with, a hypothesis is made. The validity of this hypothesis is tested. If the hypothesis is found to be true, it is accepted. On the other hand, if it is found to be false, it is rejected.

The hypothesis which is being tested for possible rejection its called null hypothesis. The null hypothesis is denoted by H_0.

If the null hypothesis is found to be false another hypothesis which contradicts the null hypothesis is accepted. This hypothesis which is accepted when the null hypothesis is rejected is called alternative hypothesis. The alternative hypothesis is denoted by H_1.

For Example, (i) to test whether there is a difference between two populations we take the null hypothesis as H_0 ; there is no difference between the two populations. Rejecting this null hypothesis H_0 will mean that the two populations are different. Acceptance (non rejection) of the null hypothesis will mean that there is no difference in the population (populations are same).

CRITICAL REGION

samples belonging to the sample space which lead to the rejected of the null hypothesis is called **critical region.** The critical region is denoted by ϖ. The critical region is also called **rejection region.** The set of

samples which lead to the acceptance of the null hypothesis is the **acceptance region.** It is $(S - \varpi)$

In fact, to decide whether the sample under consideration belongs to ϖ, the criteria $|Z| > k$ is adopted. And so, in effect, the critical region is defined by $|Z| > k$.

TYPE I AND TYPE II ERROR

While testing a null hypothesis against an alternative hypothesis, one of the following four situations arise :

Actual fact	Decision based on the sample	Error	
1 H_0 is true	accept H_0	correct decision	
2 H_0 is true	reject H_0	wrong decision	Type I
3 H_0 is not true	accept H_0	wrong decision	Type II
4 H_0 is not true	reject H_0	correct decision	

(1) Error of the first kind (Type I error) is taking a wrong decision to reject the null hypothesis when it is actually true.

(2) Error of the second kind (Type II error) is taking a wrong decision to accept the null hypothesis when it is actually not true.

LEVEL OF SIGNIFICANCE

The probability of occurrence of the first kind of error is denoted by α. It is called level of significance

Definition: The level of significance is the probability of rejection of the null hypothesis when it is actually true. Usually, the level of significance is fixed at 0.05 or 0.01. in other words, the level is fixed at 5% or 1%.

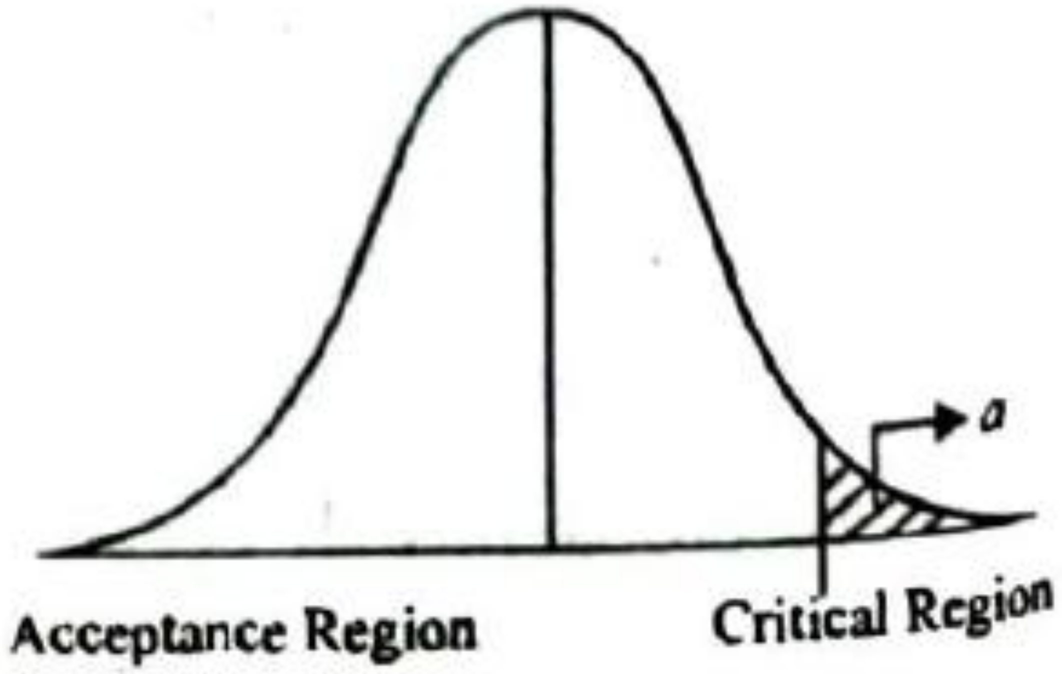
The probability of occurrence of the second kind of error is denoted by β.

The value $(1 - \beta)$ is called power of the test.

Thus it is observed that, the critical value k is based on the level of significance. For tests which are based on normal distribution, if $\alpha = 0.05$, the critical value is $k = 1.96$. If $\alpha = 0.01$, the critical value is $k = 2.58$.

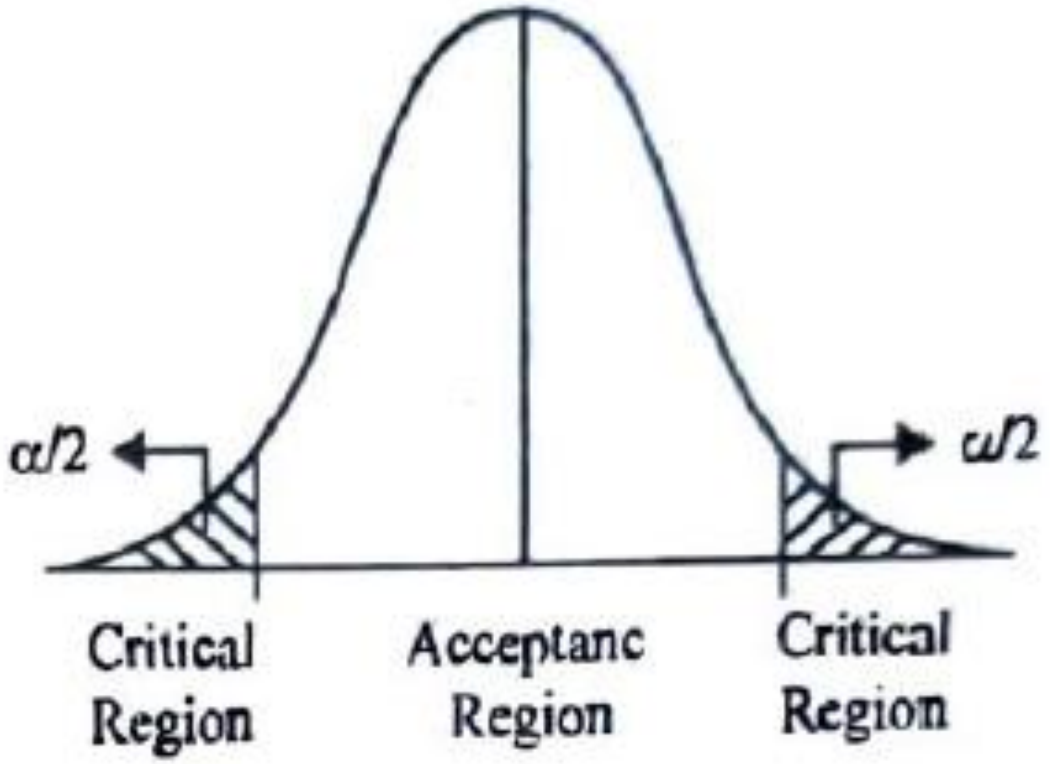
ONE TAILED TEST

While testing H_0, if the critical region is considered at one tail of the sampling distribution of the test statistic, the test is one–tailed test.



TWO TAILED TEST

One the other hand, if the critical region is considered at both the tails of the sampling distribution of the test statistic, the test is two–tailed test.



SUMMARY

- Testing of Hypotheses
- Statistical Hypotheses
- Null Hypotheses
- Critical Region
- Type I and Type II error
- level of significance
- One and Two tailed test

QUIZ

LECTURE -4, TEST OF SIGNIFICANCE, (MEAN)

RECAP

- Testing of Hypotheses
- Statistical Hypotheses
- Null Hypotheses
- Critical Region
- Type I and Type II error
- level of significance
- One and Two tailed test

OBJECTIVE

- Large and Small Samples
- Test for Mean
- Test for Mean Problems
- Problems for practice

LARGE AND SMALL SAMPLES

Large Sample

In the previous sections, we have discussed the means and standard deviations (standard errors) of the sampling distributions sample mean, sample proportion, etc. It can be shown that for large samples, these statistic have normal distribution. And so, while testing hypothesis concerning means and proportions, the critical values can be obtained by using normal distribution.

$n > 30$

Small Sample

If $n \leq 30$ then it is called small sample

Here we will consider all samples are large only

PROCEDURE FOR TESTING A STATISTICAL HYPOTHESIS.

The following is a procedure for testing a statistical hypothesis

Step 1: Set up the null hypothesis H_0.

Step 2: Set up the alternative hypothesis H_1.

Just negation of null hypothesis. i.e., $H_1 = {\sim}H_0$.

Step 3: Compute the test statistic $| Z |$

Here
a) Single mean
b) Difference of two means
c) Single proportion
d) Difference of two proportions
e) Equal of two proportions

Step 4: Choose the critical value 'k', appropriate level of significance α

in general α is 5%, k =1.96 or 1%, k = 2.58

Step 5: Compare $|Z|$ and k

If $|Z| < k$, then H_0 is accepted

or If $|Z| \geq k$, then H_1 is accepted

Conclusion: which is accepted should mention (write)

TEST FOR A MEAN.

Given a sample mean (single mean)

Step 1: Set up the null hypothesis H_0.

Step 2: Set up the alternative hypothesis H_1. i.e., $H_1 = \sim H_0$.

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}}$$

Step 4: Choose the critical value 'k', appropriate level of significance α

in general α is 5%, k =1.96 or 1%, k = 2.58

Step 5: Compare |Z| and k

If $|Z| < k$, then H_0 is accepted

or If $|Z| \geq k$, then H_1 is accepted

Conclusion: which is accepted should mention (write)

TEST FOR MEAN PROBLEMS

Example 1) A machine is designed so as to fill bottles with 200 ml of a medicine. A sample of 100 bottles when measured had a mean content of 201.3 ml. If the standard deviation of the filling is known as to be 5 ml, test whether the machine is functioning properly. Use 5% level of significance

Solution: given $\mu = 200$,

A sample $n = 100$, $\bar{x} = 201.3$ *and* $\sigma = 5$

Step 1: H_0 : the machine is functioning properly.

Step 2: H_1 : the machine is not functioning properly.

Step 3: Test of Statistics $|z| = \frac{|\bar{x}-\mu|}{\frac{\sigma}{\sqrt{n}}}$

$$= \frac{|201.3 - 200|}{\frac{5}{\sqrt{100}}}$$

$= 2.6$

Step 4: 5% level of significance

=> the critical value is $k = 1.96$

Step 5: Compare $|z| = 2.6 > k = 1.96$

$\therefore$ H_1 is accepted

Conclusion: the machine is not functioning properly.

TEST FOR MEAN PROBLEMS CONTD.

Example 2. It is required test the hypothesis that on an average height of Punjabis is 180 cms tall. For this, a random sample containing 50 Punjabis are considered. The mean and standard deviation of heights of these are found to be 178.9 cms and 3.3 cms. Based on this data, that would you conclude? (Use 1% level of significance).

Solution: given average height of Punjabis μ = 180,

A sample n = 50, $\bar{x} = 178.9 \; and \; \sigma = 3.3$

Step 1: H_0 : On average height of Punjabis is 180 cms.

Step 2: H_1 : On average height of Punjabis is not 180 cms.

Step 3: Test of Statistics $|z| = \frac{|\bar{x}-\mu|}{\frac{\sigma}{\sqrt{n}}}$

$$= \frac{|178.9 - 180|}{\frac{3.3}{\sqrt{50}}}$$

= 2.357

Step 4: 1% level of significance

=> the critical value is k = 2.58

Step 5: Compare |z| = 2.357 < k = 2.58

∴ H_0 is accepted

Conclusion: On average Punjabis is 180 cms.

TEST FOR MEAN PROBLEMS CONTD.

Example 3. A firm manufactures resistors which are known to have resistance with standard deviation 0.02 ohms. A random sample of 64 resistors had mean resistance 1.39 ohms. Can you conclude that the mean resistance of the resistors manufactured by the firm is 1.4 ohms?

Solution: given mean resistance μ = 1.4,

A sample n = 64, $\bar{x} = 1.39 \; and \; \sigma = 0.02$

Step 1: H_0 : The mean resistance is 1.4 Ohms.

Step 2: H_1 : The mean resistance is not 1.4 Ohms.

Step 3: Test of Statistics $|z| = \frac{|\bar{x}-\mu|}{\frac{\sigma}{\sqrt{n}}}$

$$= \frac{|1.39 - 1.4|}{\frac{0.02}{\sqrt{64}}}$$

= 4

Step 4: In general consider 5% level of significance

=> the critical value is k = 1.96

Step 5: Compare |z| = 4 > k = 2.58

∴ H_1 is accepted

Conclusion: The mean resistance is not 1.4 Ohms.

TEST FOR MEAN PROBLEMS

Example 8. A random variable has variance 81. Its mean is not known. 144 random observations of the variable have mean 112.6. Test the hypothesis that the mean of variable is 114.5. Use level of significance a = 0.01 and also a = 0.05.

Solution : Here, $\mu_0 = 114.5$ $\sigma = \sqrt{81} = 9$, $n = 144$ and $\bar{x} = 112.6$

H_0 : The mean of the variable is 114.5

H_1 : The mean of the variable differs from 114.5

The test statistic is

$$|Z| = \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}}$$

$$= \frac{|112.6 - 114.5|}{9 / \sqrt{144}} = 2.53$$

i) The level of significance is $\alpha = 0.01$

The critical value is $k = 2.58$

Since $|Z|_{cal} < 2.58$, H_0 is accepted.

ii) The level of significance is $\alpha = 0.05$

The critical value is $k = 1.96$

Since $|Z|_{cal} > 1.96$, H_0 is rejected.

Conclusion : At 1% level of significance, we conclude that the mean of the variable is 114.5 But, at 5% level of significance, we conclude that the mean differs from 114.5

SUMMARY

- Large and Small Samples
- Test for Mean
- Test for Mean Problems
- Problems for practice

PROBLEMS FOR PRACTICE

Example 11. A manufacturer claimed that at least 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipments revealed that 18 were faulty. Test his claim at a significance level of (i) 5% (ii) 1%.

Example 12. From a population with mean 836, a random sample containing 225 observations is taken. The sample had mean 840.5 and standard deviation 45. Test whether the sample mean differs significantly from the population mean.

Note : In this question, the level of significance is not mentioned. In such situations, it is customary to take it as 5%.

LECTURE -5, TEST OF SIGNIFICANCE, (DIFFERENCE OF MEANS)

RECAP

- Large and Small Samples
- Test for Mean
- Test for Mean Problems
- Problems for practice

OBJECTIVE

- Test for Difference of Means
- Test for Difference of Means Problems
- Problems for practice

TEST FOR DIFFERENCE OF MEANS.

Given two samples mean

Step 1: Set up the null hypothesis H_0.

Step 2: Set up the alternative hypothesis H_1. i.e., $H_1 = \sim H_0$.

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Step 4: Choose the critical value 'k', appropriate level of significance α

in general α is 5%, k =1.96 or 1%, k = 2.58

Step 5: Compare |Z| and k

If $|Z| < k$, then H_0 is accepted

or If $|Z| \geq k$, then H_1 is accepted

Conclusion: which is accepted should mention (write)

TEST FOR MEAN PROBLEMS CONTD.

Example 1. It is known that an IQ of boys has SD 10 and that an IQ of girls has SD 12. Mean IQ of 200 randomly selected boys is 99 and Mean IQ of 300 randomly selected girls is 97. Can it be concluded that on an average boys and girls have the same IQ? (Use 1% level of significance)

Solution: given two samples.,

Boys: Mean $\bar{x}_1 = 99$, SD $\sigma_1 = 10$, and $n_1 = 200$

Girls: Mean $\bar{x}_2 = 97$, SD $\sigma_2 = 12$ and $n_2 = 300$

Step 1: H_0 : On average boys and girls have same IQ.

Step 2: H_1 : On average boys and girls have not same IQ

Step 3: Test of Statistics $|z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$$= \frac{|99 - 97|}{\sqrt{\frac{10^2}{200} + \frac{12^2}{300}}} = 2.02$$

Step 4: 1% level of significance

=> the critical value is k = 2.58

Step 5: Compare $|z| = 2.02 < k = 2.58$

$\therefore H_0$ is accepted

Conclusion: On average boys and girls have same IQ.

TEST FOR MEAN PROBLEMS CONTD.

Example 2. Intelligence test on two groups of boys and girls have the following results

	Mean	SD	Size
Girls	75	15	150
Boys	70	20	250

Is there significant difference in the mean marks obtained by the boys and girls?

Solution: given two groups (samples)

Girls: Mean $\bar{x}_1 = 75$, SD $\sigma_1 = 15$, and $n_1 = 150$

Boys: Mean $\bar{x}_2 = 70$, SD $\sigma_2 = 20$, and $n_2 = 250$

Step 1: H_0 : There is significant difference in girls & boys.

Step 2: H_1 : There is no significant difference in girls & boys.

Step 3: Test of Statistics $|z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$$= \frac{|75 - 70|}{\sqrt{\frac{15^2}{150} + \frac{20^2}{250}}} = 2.84$$

Step 4: In general consider, 5% level of significance

=> the critical value is k = 1.96

Step 5: Compare $|z| = 2.84 > k = 1.96$

$\therefore H_1$ is accepted

Conclusion: There is no significant difference in girls & boys.

TEST FOR DIFFERENCE OF MEANS PROBLEMS CONTD.

Example 19. A study of systolic blood pressure of a randomly selected groupo of 36 patients suffering from a disease and another group of 36 persons who do not suffer from the disease gave the following results.

	Suffering	Not suffering
Sample size :	36	36
Mean systolic pressure :	178	161
Standard deviation :	24	12

Test whether the average systolic pressure of the persons of the two groups differ significantly.

Solution : Here, $n_1 = 36$ $\bar{x}_1 = 178$ $s_1 = 24$

$n_2 = 36$ $\bar{x}_2 = 161$ $s_2 = 12$

H_0 : Average systolic pressure of the two groups of persons do not differ significantly.

H_1 : Average systolic pressure of the two groups of differ significantly.

Step 3: Test of Statistics $|z| = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$$= \frac{|178 - 161|}{\sqrt{\frac{24^2}{36} + \frac{12^2}{36}}} = 3.80$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is rejected.

Conclusion : Average systolic pressure of the two groups differ significantly.

TEST FOR MEAN PROBLEMS CONTD.

Example 4. The mean and standard deviation of heights of 100 randomly selected boys are 163 cms and 3 cms respectively. The mean and standard deviation of heights of 80 randomly selected girls are 161 cms and 2 cms respectively. Can it be concluded at 1% level of significance that boys and girls are equally tall?

Solution: given two samples.,

Boys: Mean $\bar{x}_1 = 163$, SD $\sigma_1 = 3$, and $n_1 = 100$

Girls: Mean $\bar{x}_2 = 161$, SD $\sigma_2 = 2$ and $n_2 = 80$

Step 1: H_0 : Boys and girls are equally tall.

Step 2: H_1 : Boys and girls are not equally tall

Step 3: Test of Statistics $|z| = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$$= \frac{|163 - 161|}{\sqrt{\frac{3^2}{100} + \frac{2^2}{80}}} = 5.35$$

Step 4: 1% level of significance

=> the critical value is k = 2.58

Step 5: Compare $|z| = 5.35 > k = 2.58$

$\therefore$ H_1 is accepted

Conclusion: Boys and girls are not equally tall.

SUMMARY

- Test for Difference of Means
- Test for Difference of Means Problems
- Problems for practice

PROBLEM FOR PRACTICE

31) Samples of electric lamps manufactured by two firms gave the following results –

	Firm A	Firm B
Sample size :	80	60
Average buring hours :	1348	1270
Standard deviation :	108	96

Test whether the average life of bulbs manufactured by the firm are the same. [Ans. 4.52, rejected]

32) From the following data test whether the means differ significantly.

	I sample	II sample
Size :	300	200
Mean :	75.4	74.3
Variance :	65.6	57.8

[Ans. 1.54 accepted]

LECTURE -6, TEST OF SIGNIFICANCE, (PROPORTION)

RECAP

- Test for Difference of Means
- Test for Difference of Means Problems
- Problems for practice

OBJECTIVE

- Test for Proportion
- Test for Proportion Problems
- Problems for practice

TEST FOR PROPORTION

Given a proportion (single proportion) here sample $p = \frac{x}{n}$, P = universe Proportion, Q = 1 – P

Step 1: Set up the null hypothesis H_0.

Step 2: Set up the alternative hypothesis H_1. i.e., $H_1 = \sim H_0$.

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|p - P|}{\sqrt{\frac{PQ}{n}}}$$

Step 4: Choose the critical value 'k', appropriate level of significance α

in general α is 5%, k =1.96 or 1%, k = 2.58

Step 5: Compare |Z| and k

If $|Z| < k$, then H_0 is accepted

or If $|Z| \geq k$, then H_1 is accepted

Conclusion: which is accepted should mention (write)

TEST FOR PROPORTION PROBLEMS CONTD.

Example 1. A manufacture claims that only 4% of his supplied by him are defective. A random sample of 600 contained 36 defectives. Test the claim of the manufactures.

Solution: Given 4% are defective, In a sample 600 contained 36 defectives.

Single proportion

A sample n = 600 and defectives x = 36

$$\therefore p = \frac{x}{n} = \frac{36}{600} = 0.06$$

Population proportion P = 4%

$$P = 4\% = \frac{4}{100} = 0.04$$

$$\therefore Q = 1 - P = 1 - 0.04 = 0.96$$

Step 1: H_0 = A manufacture claims that only 4% defectives.

Step 2: H_1 = A manufacture claims that not only 4% defectives

Step 3: Compute the test statistic | Z |

$$|Z| = \frac{|p - P|}{\sqrt{\frac{PQ}{n}}}$$

$$|Z| = \frac{|0.06 - 0.04|}{\sqrt{\frac{0.04 \ x\ 0.96}{600}}}$$

= 2.5

Step 4: Choose the critical value 'k', appropriate level of significance α

let consider α is 5%, $\therefore$ k =1.96

Step 5: Compare |Z| and k

Here |Z| = 2.5 $>$k = 1.96,

$\therefore$ H_1 is accepted

Conclusion: A manufacture claims that not only 4% defectives

TEST FOR PROPORTION PROBLEMS CONTD.

Example 2. Given that on the average 4% of insured men of age 65 die within a year and that 60 of a particular group of 1000 such men died within a year. Can this group be regarded as a representative sample?.

Solution: Given 4% are die, In a sample 1000 contained 60 are died.

Single proportion

A sample n = 1000 and died x = 60

$$\therefore p = \frac{x}{n} = \frac{60}{1000} = 0.06$$

Population proportion P = 4%

$$P = 4\% = \frac{4}{100} = 0.04$$

$$\therefore Q = 1 - P = 1 - 0.04 = 0.96$$

Step 1: H_0 = A group of chosen is a representative of sample.

Step 2: H_1 = A group of chosen is not a representative of sample

Step 3: Compute the test statistic | Z |

$$|Z| = \frac{|p - P|}{\sqrt{\frac{PQ}{n}}}$$

$$|Z| = \frac{|0.06 - 0.04|}{\sqrt{\frac{0.04\,x\,0.96}{1000}}}$$

= 3.23

Step 4: Choose the critical value 'k', appropriate level of significance α

let consider α is 5%, ∴ k =1.96

Step 5: Compare |Z| and k

Here |Z| = 3.23 > k = 1.96,

∴ H_1 is accepted

Conclusion: A group of chosen is not a representative of sample

SUMMARY

- Test for Proportion
- Test for Proportion Problems
- Problems for practice

PROBLEMS FOR PRACTICE

Exercise 23. Doordarshan anticipated 65% viewer ship for the recent India Vs Pakistan cricket match. Among 300 viewers of DD who where contacted. 217 had viewed the match. Does this figure support Doordarshan's anticipated viewership?

Example 27: A survey of 225 randomly selected students of I BCA from all over Bangalore University revealed that 89.4% of them used the book 'Statistics by Srimani + Vinayakamoorthy' as text. Can we conclude at 1% level of significance that the book is being used by 90% of the students.

Example 28. Manufacturer of VM–cars is of the opinion that 35% of the cars plying on the roads of Bangalore are VM–cars. A person decided to verify this by conducting a survey. He counted the cars plying on K.G. Road between 4 p.m. and 4.10 p.m. and noted the make. Among 347 cars which he counted 107 were VM–cars. Does this figure support the manufacturer's claim?

LECTURE -7, TEST OF SIGNIFICANCE, (DIFFERENCE OF PROPORTIONS)

RECAP

- Test for Proportion
- Test for Proportion Problems
- Problems for practice

OBJECTIVE

- Test for Equality Difference of Proportions
- Test for Equality Difference of Proportions Problems
- Problems for practice

TEST FOR DIFFERENCE OF EQUAL PROPORTIONS

Given two equal proportions

Step 1: Set up the null hypothesis H_0.

Step 2: Set up the alternative hypothesis H_1. i.e., $H_1 = \sim H_0$.

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where $p_1 = \frac{x_1}{n_1}, p_2 = \frac{x_2}{n_2}, P = \frac{x_1+x_2}{n_1+n_2}, or \ \ P = \frac{n_1 p_1 + n_2 p_2}{n_1+n_2} \ and \ Q = 1 - P$

Step 4: Choose the critical value 'k', appropriate level of significance α

in general α is 5%, k =1.96 or 1%, k = 2.58

Step 5: Compare |Z| and k

If |Z| < k, then H_0 is accepted, or If $|Z| \geq k$, then H_1 is accepted

Conclusion: which is accepted should mention (write)

TEST FOR DIFFERENCE OF EQUAL PROPORTIONS CONTD.

Suppose there are two populations with unknown proportions, and we wish to test whether the proportions (of certain attribute) in the two populations are equal. The null hypothesis is $H_0: P_1 = P_2$ (the proportions are equal). The alternative hypothesis is $H_1: P_1 \neq P_2$.

Let P be the common proportion. Let a large random sample of size n_1 be drawn from the first population. Among these n_1 units, let x_1 units possess the attribute, so that the sample proportion is $p_1 = \frac{x_1}{n_1}$.

Also, let a large ramp, sample of size n_1 be drawn from the second population. Among the units, let x_2 units possess the attribute, so that the sample proportion is $p_2 = \frac{x_2}{n_2}$

TEST FOR DIFFERENCE OF EQUAL PROPORTIONS CONTD.

And so, the test statistic is

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Generally, the common proportion P will not be known. And so, it is estimated from the samples.

The estimate is $P = \dfrac{x_1 + x_2}{n_1 + n_2}$

Also, $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

For the sample, if $|Z|_{cal} > k$, is rejected

On the other hand, if $|Z|_{cal} \leq k$, H_0 is accepted.

For the level for significance $\alpha = 0.05$, the critical value is $k = 1.96$. However, for $\alpha = 0.01$, the critical value is $k = 2.58$.

TEST FOR DIFFERENCE OF EQUAL PROPORTIONS PROBLEMS

Example 1. 500 articles from a factory are examined and found to be 2% defective. 800 similar articles from a second factory are found to have only 1.5% defectives. Can it be reasonable concluded that the products of the first factory are inferior to those of the second.

Solution: Given

Factory1: $n_1 = 500$ and $p_1 = 2\% = 0.02$

Factory2: $n_2 = 800$ and $p_2 = 1.5\% = 0.015$

Two samples and both factory articles are same.

$\therefore$ Its two equal proportions problems

$$\therefore P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$\Rightarrow P = \frac{500\ x\ 0.02 + 800\ x\ 0.015}{500+800} = 0.0169$$

$\therefore Q = 1 - P = 1 - 0.0169 = 0.9831$

Step 1: H_0= The products of the first factory are inferior to those of the second

Step 2: H_1 = The products of the first factory are not inferior to those of the second

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$|Z| = \frac{|0.02 - 0.015|}{\sqrt{0.0169\ x\ 0.9831\left(\frac{1}{500} + \frac{1}{800}\right)}}$$

$\therefore$ |z| = 0.6804

Step 4: Choose 5% level of significance α

the critical value k = 1.96

Step 5: Compare |Z| and k

Here |Z| = 0.6804 < k = 1.96

$\therefore$ H_0 is accepted,

Conclusion: The products of the first factory are inferior to those of the second

TEST FOR DIFFERENCE OF EQUAL PROPORTIONS PROBLEMS

Example 2. A sample of 1000 products from a factory are examined and found to be 2.5% defective. Another sample of 1500 similar products from another factory are found to have only 2% defective. Can you conclude that the products of the first factory are inferior to those of the second?.

Solution: Given

Factory1: $n_1 = 1000$ and $p_1 = 2.5\% = 0.025$

Factory2: $n_2 = 1500$ and $p_2 = 2\% = 0.02$

Two samples and both factory articles are same.

∴ Its two equal proportions problems

$$\therefore P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$\Longrightarrow P = \frac{1000\ x\ 0.025 + 1500\ x\ 0.02}{1000 + 1500} = 0.022$$

∴ Q = 1 – P = 1 – 0.022 = 0.978

Step 1: H_0= The products of the first factory are inferior to those of the second

Step 2: H_1 = The products of the first factory are not inferior to those of the second

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$|Z| = \frac{|0.025 - 0.02|}{\sqrt{0.022\ x\ 0.978\left(\frac{1}{1000} + \frac{1}{1500}\right)}}$$

∴ |z| = 0.834958

Step 4: Choose 5% level of significance α

the critical value k = 1.96

Step 5: Compare |Z| and k

Here |Z| = 0.835 < k = 1.96

∴ H_0 is accepted,

Conclusion: The products of the first factory are inferior to those of the second

TEST FOR DIFFERENCE OF EQUAL PROPORTIONS PROBLEMS

Example 3. During a country wide investigation the incidence of TB was found to be 1%. In a college of 400 strength 5 were reported to be affected whereas in another college of 1200 strength 10 were reported to be affected. Does this indicate any significant difference?

Solution: Given

College1: n_1 = 400 and x_1 = 5

College2: n_2 = 1200 and x_2 = 10

Two samples and both affected TB.

∴ Its two equal proportions problems

Country wide TB was found 1%

i.e., P = 1% = 0.01

∴ Q = 1 – P = 1 – 0.01 = 0.99

Here $p_1 = \frac{x_1}{n_1} = \frac{5}{400} = 0.0125$

$p_2 = \frac{x_2}{n_2} = \frac{10}{1200} = 0.0083$

Step 1: H_0= This is indicate any significant difference

Step 2: H_1 = This is not indicate any significant difference

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$|Z| = \frac{|0.0125 - 0.0083|}{\sqrt{0.01\ x\ 0.99\left(\frac{1}{400} + \frac{1}{1200}\right)}}$$

∴ |z| = 0.731126

Step 4: Choose 5% level of significance α

the critical value k = 1.96

Step 5: Compare |Z| and k

Here |Z| = 0.731 < k = 1.96

∴ H_0 is accepted,

Conclusion: This is indicate any significant difference

TEST FOR DIFFERENCE OF EQUAL PROPORTIONS PROBLEMS

Example 4. A survey of men and women who were aged 50 or more was conducted. Among the 80 men, 8 had high B.P. Among the 80 women, 6 had high B.P. Test whether the proportion of men with high B.P. differs from the proportion of women with high B.P.?

Solution: Given

Men: $n_1 = 80$ and $x_1 = 8$

Women: $n_2 = 80$ and $x_2 = 6$

Two samples and both affected high B.P.

$\therefore$ Its two equal proportions problems

Here $P = \frac{x_1+x_2}{n_1+n_2} = \frac{8+6}{80+80} = 0.0875$

$\therefore$ Q = 1 – P = 1 – 0.0875 = 0.9125

Here $p_1 = \frac{x_1}{n_1} = \frac{5}{400} = 0.0125$

$p_2 = \frac{x_2}{n_2} = \frac{10}{1200} = 0.0083$

Step 1: H_0= Men with high BP is differs from women BP

Step 2: H_1 = Men with high BP is not differs from women BP

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$|Z| = \frac{|0.0125 - 0.0083|}{\sqrt{0.0875\ x\ 0.9125\left(\frac{1}{80}+\frac{1}{80}\right)}}$$

$\therefore$ |z| = 0.094006703

Step 4: Choose 5% level of significance α

the critical value k = 1.96

Step 5: Compare |Z| and k

Here |Z| = 0.094 < k = 1.96

$\therefore$ H_0 is accepted,

Conclusion: Men with high BP is differs from women BP

TEST FOR DIFFERENCE OF EQUAL PROPORTIONS PROBLEMS

Example 5. Among 37 educated youths, 3 were smokers. Among 26 uneducated youths, 4 were smokers. Test whether the proportion of smokers is the same among the educated and uneducated.

Solution: Given

Educated youths : $n_1 = 37$ and $x_1 = 3$

Uneducated youths: $n_2 = 26$ and $x_2 = 4$

Two samples and both smokers data.

∴ Its two equal proportions problems

Here $P = \frac{x_1+x_2}{n_1+n_2} = \frac{3+4}{37+26} = 0.1111$

∴ $Q = 1 - P = 1 - 0.1111 = 0.8889$

Here $p_1 = \frac{x_1}{n_1} = \frac{3}{37} = 0.0811$

$p_2 = \frac{x_2}{n_2} = \frac{4}{26} = 0.1538$

Step 1: H_0= Smokers is the same among the educated & uneducated.

Step 2: H_1 = Smokers is not the same among the educated & uneducated

Step 3: Compute the test of statistic is | Z |

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$|Z| = \frac{|0.0811 - 0.1538|}{\sqrt{0.1111 \ x \ 0.8889\left(\frac{1}{37}+\frac{1}{26}\right)}}$$

∴ $|z| = 0.903999498$

Step 4: Choose 5% level of significance α

the critical value k = 1.96

Step 5: Compare |Z| and k

Here $|Z| = 0.904 < k = 1.96$

∴ H_0 is accepted,

Conclusion: Smokers is the same among the educated & uneducated.

SUMMARY

- Test for Equality Difference of Proportions
- Test for Equality Difference of Proportions Problems
- Problems for practice

PROBLEMS FOR PRACTICE

Example 25. A machine puts out 16 imperfect articles in a sample of 500 articles. After the machine is overhauled it puts out 3 defective articles in a sample of 100. Has the machine improved?

Example 30. In a year there are 956 births in a town A of which 52.5% were males while in town A and B combined this proportion in total of 1406 births was .496. Is there any significant difference in the proportion of male births in the two towns?

Example 34. From the following data, test whether the difference between the proportions in the two samples is significant.

	Size	Proportion
Sample I	1000	0.02
Sample II	1200	0.01