

Hypothesis Testing



Odisha Social Science
Research & Consultancy

CONTENTS

- **Introduction**
- **Hypothesis Testing : A single population mean**
- **Hypothesis Testing: The difference between two population means**
- **Paired comparison**
- **Hypothesis Testing: A single population Proportion**
- **Hypothesis Testing: The difference between two population Proportion**
- **Hypothesis Testing: A single population Variance**
- **Hypothesis Testing: The ratio of two population variance**

Learning Objective

- **The trainees will be able to take the decision/reach a conclusion concerning a population by examining a sample from that population.**

1. Introduction

➤ Hypothesis testing- Purpose

- To aid the clinician, researcher , or administrator in reaching a conclusion concerning a population by examining a representative sample from that population.

➤ Hypothesis Defined:- A statement about one or more population.

➤ Types of Hypothesis:

- Research hypothesis and statistical hypothesis

➤ Research hypothesis

- It is the conjecture or supposition that motivates the research.
- Research project emanates from the desire of such health practitioners to validate their supposition by rigors of scientific investigation.

Statistical Hypothesis

- That are stated in such a way that they may be evaluated by appropriate statistical techniques.

Hypothesis Testing

- Hypothesis testing is the Procedures which enable researcher to decide whether to accept or reject hypothesis or whether observed samples differ significantly from expected results. The results can't be attributed to chance variation due to sampling

Hypothesis Testing Steps

1. Data

The nature of the data should be understood as it determines the particular test to be employed.

2. Assumption

Assumptions regarding the distribution of the data decide the testing procedure.

3. Hypotheses

- Two statistical hypotheses involved in hypothesis testing- Null (H_0) & Alternative (H_a) hypotheses.
- H_0 = Hypothesis of no difference. This is set up with the express purpose of being discredited.
- H_a = A set of alternative to H_0 . A statement of what the researcher believe to be true if our sample data cause us to reject H_0 .

Hypothesis Testing Steps. Contd.

Rules for stating statistical hypotheses:

- a) What you hope or expect to be able to conclude as a result of test is usually placed in the alternative hypothesis.
- b) The H_0 should contain a statement of equality, either $=$, $\leq$, $\geq$.
- c) H_0 is the hypothesis that is tested.
- d) The H_0 and H_a hypothesis are complimentary. That is the two together exhaust all possibilities regarding the value that the hypothesized parameter can assume.

Hypothesis testing does not lead to the proof of a hypothesis. It merely indicates whether the hypothesis is supported or not by the available data.

4. Test Statistics

- It is some statistic computed from the data of the sample.
- The decision to reject or not to reject the null hypothesis depends on the magnitude of the test statistic
- The general formula for test statistic is

$$\text{Test Statistics} = \frac{\text{Relevant Statistics} - \text{Hypothesized parameter}}{\text{Standard error of the Relevant Statistics}}$$

- An example of test statistic is

where μ_0 is a hypothesized value of a population mean.

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

5. Distribution of test statistic

- Key to statistical inference is the sampling distribution. This plays an important role in testing hypothesis.
- For example the distribution of the test statistic

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

follows the standard normal distribution if the null hypothesis is true and the assumptions are made.

6. Decision rule

- All possible values that the test statistic can assume are points on the horizontal axis of the graph of the distribution of the test statistic and are divided into two groups- one group constitute rejection region and the other makes up the non rejection region.

- If the test statistic falls in the rejection region the H_0 is rejected and not rejected otherwise.

Errors in test of hypothesis:

- The error committed when a true null hypothesis is rejected is called a type I error and the probability of type-I error is designated by α

	Decision	Decision
Condition Of Null Hypothesis	Accept H_0	Reject H_0
$H_0(\text{true})$	Correct decision	Type I error (α error)
$H_0(\text{false})$	Type II error (β error)	Correct decision

- The type II error is committed when a false null hypothesis is not rejected. The probability of type II error is designated by β .
- Error is determined in advance as level of significance for a given sample size
- If we try to reduce type I error, the probability of committing type II error increases
- Both type errors cannot be reduced simultaneously
- Decision maker has to strike a balance / trade off examining the costs & penalties of both type errors

Level of significance (α):

- Some percentage (usually 5%) chosen with great care, thought & reason so that H_0 will be rejected when the sampling result (observed evidence) has a probability of < 0.05 of occurring if H_0 is true
- Researcher is willing to take as much as a 5% risk of rejecting H_0
- Significance level is the maximum value of the probability of rejecting H_0 when it is true
- It is usually determined in advance, i.e., the probability of type I error (α) is assigned in advance and hence nothing can be done about it.

7. Calculation of test statistic

From the data we compute a value of the test statistic and compare it with the rejection and non rejection regions.

8. Statistical decision

H_0 is rejected if the computed value of test statistic falls in the rejection region or it is not rejected if it falls in the non rejection region.

9. Conclusion

If H_0 is rejected, we conclude that H_A is true. If H_0 is not rejected we conclude that H_0 may be true.

10. p values

The p value is a number that tells us how unusual our sample results are, given H_0 is true. It is the prob. that the test statistic will fall in the rejection region when H_0 is true. If ' p ' $\leq \alpha$, we reject H_0 . If ' p ' $> \alpha$, H_0 is not rejected.

STEPS IN HYPOTHESIS TESTING PROCEDURE

EVALUATE DATA

REVIEW ASSUMPTIONS

STATE HYPOTHESIS

SELECT TEST STATISTICS

DETERMINE DISTRIBUTION OF TEST STATISTICS

State decision rule

CALCULATE TEST STATISTICS

Do not
Reject H_0

Conclude H_0
May be true

MAKE STATISTICAL
DECISION

Reject
 H_0

Conclude H_1
Is true

2.Hypothesis Testing : A single population mean

The testing of a hypothesis about a population mean under three different conditions:

➤ When sampling is from:

2.1 a normally distributed population of values with known variance

2.2 a normally distributed population with unknown variance

2.3 a population that is not normally distributed.

2.1 sampling is from a normally distributed population and the population variance is known

The test statistic for testing $H_0 : \mu = \mu_0$ is

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

One tail test

$H_0 : \mu \leq \mu_0$ ($H_0 : \mu \geq \mu_0$)

$H_a : \mu > \mu_0$ (or $H_a : \mu < \mu_0$)

Rejection region

$z > z_\alpha$ ($z < -z_\alpha$)

Two tail test

$H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$

Rejection region

$z < -z_{\alpha/2}$ (or $z > z_{\alpha/2}$)

Example :1

Researchers are interested in the mean age of a certain population. Data available to the researchers are ages of a simple random sample of 10 individuals. From this sample a mean of $\bar{x} = 27$ and let us assume that the population has a known variance $\sigma^2 = 20$ and it is normally distributed.

Can we conclude that

- a. The mean age of this population is different from 30 years.
- b. The mean age of the population is less than 30 years

2.2 Sampling from a normally distributed population: Population variance unknown.

When sampling is from a normally distributed population with an unknown variance, the test statistic for testing $H_0 : \mu = \mu_0$ is

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

When H_0 is true, it follows a Student's 't' with $n-1$ d.f.

One tailed test

■ $H_0 : \mu \leq \mu_0$ ($H_0 : \mu \geq \mu_0$)

■ $H_a : \mu > \mu_0$ (or $H_a : \mu < \mu_0$)

■ Rejection region

$t > t_{\alpha}$ (or $t < -t_{\alpha}$)

(t_{α} is the t-value
such that $P(t > t_{\alpha}) = \alpha$)

Two tail test

$H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$

Rejection region

$t < -t_{\alpha/2}$ (or $t > t_{\alpha/2}$)

($t_{\alpha/2}$ is the t-value such that
 $P(t > t_{\alpha/2}) = \alpha/2$)

Example: 2

The investigators' subjects were 14 healthy adult males representing a wide range of body weight. One of the variable on which measurements were taken was body mass index (BMI) = $\text{weight}(\text{kg})/\text{height}^2 (\text{m}^2)$. The results are shown in the table. If we can conclude that the mean BMI of the population from which the sample was drawn is not 35.

Table

Subject	BMI	Subject	BMI	Subject	BMI
1	23	6	21	11	23
2	25	7	23	12	26
3	21	8	24	13	31
4	37	9	32	14	45
5	39	10	57		

2.3 Sampling from a population that is not normally distributed.

(How to test the normality of the population? What is central Limit Theorem)

In this situation if the sample is large (greater than or equal to 30), we can take advantage of central limit theorem and use as the test statistic. $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$

If the population standard deviation is not known, we can use the sample standard deviation as an estimate.

The test statistic for testing $H_0 : \mu = \mu_0$, then, is $z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$ which is distributed approximately as the standard normal distribution if n is large when $H_0 : \mu = \mu_0$ is true.

One tail test

$$H_0 : \mu = \mu_0$$

$$H_a : \mu > \mu_0 \text{ (or } H_a : \mu < \mu_0)$$

Rejection region

$$z > z_\alpha \text{ (or } z < -z_\alpha)$$

Two tail test

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0$$

Rejection region

$$z < -z_{\alpha/2} \text{ (or } z > z_{\alpha/2})$$

Example: 3

The objective of a study by Wilbur et al.(A-2)were to describe the menopausal symptoms, energy expenditure and aerobic fitness of healthy midlife women and to determine relationships among these factors. Among the variables measured was maximum oxygen uptake (Vo_{2max}). The mean Vo_{2max} score for a sample of 242 women was 33.3 with a standard deviation of 12.14.(*Source:family and community health*, Vol. 13:3, p. 73, Aspen publisher , Inc.,1990.)We wish to know if, on the basis of these data we may conclude that the mean score for a population of such women is greater than 30.

3.Hypothesis Testing: The difference between two population means

- Employed to determine whether or not it is reasonable to conclude that the two population means are unequal.
- In such cases, one or the other of the following hypothesis may be formulated:
 1. $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$
 2. $H_0: \mu_1 - \mu_2 \geq 0$, $H_a: \mu_1 - \mu_2 < 0$
 3. $H_0: \mu_1 - \mu_2 \leq 0$, $H_a: \mu_1 - \mu_2 > 0$

Hypothesis testing involving the difference between two population means will be discussed in three different contexts:

3.1. When sampling is from a normally distributed population of values with known variance

3.2 when sampling is from a normally distributed population with unknown variance

3.3 when sampling is from a population that is not normally distributed.

3.1 Sampling from normally Distributed population variance known

When each of two independent simple random samples has been drawn from a normally distributed population with a known variance, the test statistic for testing the null hypothesis of equal population means is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

When H_0 is true the test statistics is distributed as the standard normal variate.

Example 4:

Researchers wish to know if the data they have collected provide sufficient evidence to indicate a difference in mean serum uric acid levels between normal individuals and individuals with down's syndrome. The data consist of serum uric acid readings on 12 individuals with Down's syndrome and 15 normal individuals. The means are

$$\bar{x}_1 = 4.5\text{mg}/100\text{ml} \text{ and } \bar{x}_2 = 3.4 \text{ mg}/100\text{ml}$$

3.2 Sampling from normally Distributed population variance unknown

Variances may be equal or unequal.

Population variance equal: when the population variances are unknown but assumed to be equal.

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

When each of two independent simple random sample has been drawn from a normally distributed population and the two populations have equal but unknown variances, the test statistic for testing $H_0 :$

$$\mu_1 = \mu_2 \text{ is } t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

when H_0 is true, is distributed as Student's 't' with $n_1 + n_2 - 2$ degrees of freedom.

Example 5:

The purpose of a study by Eidelman et al(A-6) was to investigate the nature of lung destruction in cigarette smokers before the development of marked emphysema. Three lung destructive index measurements were made on the lungs of lifelong non smokers and smokers who died suddenly outside the hospitals of non respiratory causes. A larger score indicates greater lung damage. For one of the indexes the score yielded by the lungs of a sample of nine non smokers and a sample of 16 smokers are in the table. If we may conclude that in general, have greater lung damage as measured by this destructive index than do nonsmokers.

Table

Nonsmoker	18.1	6.0	10.8	11.0	7.7	17.9	8.5
	13.0	18.9					
Smokers	16.6	13.9	11.3	26.5	17.4	15.3	15.8
	12.3	18.6	12.0	24.1	16.5	21.8	16.3
	23.4	18.8					

Population variance unequal:

$$t' = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The critical value of t' for an α level of significance and a two sided test is approximately

$$t'_{1-\alpha/2} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$$

Where $w_1 = s_1^2/n_1$, $w_2 = s_2^2/n_2$, $t_1 = t_{1-(\alpha/2)}$ for n_1-1 degree of freedom, and $t_2 = t_{1-(\alpha/2)}$ for n_2-1 degrees of freedom. The critical value of t' for a one sided test is found by computing $t'_{1-\alpha}$ using $t_1 = t_{1-\alpha}$ for n_1-1 d.f. & $t_2 = t_{1-\alpha}$ for $n_2 - 1$ d.f. For a two sided test reject H_0 if computed value of $t' \geq$ to the critical value given by the above equation or $\leq$ to negative of that.

Example 6:

Researchers wish to know if two populations differ with respect to the mean value of total serum complement activity (C_{H50}). The data consist of C_{H50} determination on $n_2 = 20$ apparently normal subjects and $n_1 = 10$ subjects with disease. The sample means and standard deviation are

$$\bar{x}_1 = 62.6, s_1 = 33.8, \bar{x}_2 = 47.2, s_2 = 10.1$$

Sampling from populations that are not normally Distributed

Say, each of the two large independent SRS has been drawn from a population i.e. not normally distributed.

We want to test $H_0: \mu_1 = \mu_2$

On the basis of central limit theorem if sample sizes are large (≥ 30) the distribution of difference between sample means will be approximately normal.

The test statistic,
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

follows a normal distribution when H_0 is true. σ_1^2 and σ_2^2 are population variance respectively. If they are not known sample variances are used as estimates.

Example 7:

An article by Becker et al. in the American journal of health Health Promotion (A-7) describes the development of a tool to measures barriers to health promotion among persons with disabilities. The authors state that the issue of barriers is especially salient for disabled persons who experience barriers in such contexts as employment, transportation, housing, education, insurance, architectural access, entitlement programs, and society's attitudes. Studies suggest that measurement of barriers can enhance health workers' understanding of the likelihood of the people engaging in various health promoting behaviors and may be a relevant construct in assessing the health behaviors of disabled persons.

Sample	Mean score	Standard deviation
D	31.83	7.93
ND	25.07	4.80

4. Paired comparison

- A method frequently employed for assessing the effectiveness of a treatment or experimental procedure is one that makes use of related observations resulting from nonindependent samples. A hypothesis test based on this type of data is known as a paired comparisons.

Reasons for Pairing

- While testing the H_0 of no difference based on two independent samples H_0 may be rejected when it is true because of extraneous factors.
- On the other hand true difference may be masked by the presence of extraneous factors.
- The objective in paired comparisons tests is to eliminate a maximum number of sources of extraneous variation by making the pairs similar with respect to as many variables as possible.

- In paired comparison we use d_i , the difference between pairs of observations, as the variable of interest.
- When the 'n' sample differences computed from the 'n' pairs of measurements constitute a simple random sample from a normally distributed population differences, the test statistics for testing hypothesis about the population mean difference μ_d is

$$t = \frac{\bar{d} - \mu_{d0}}{S_{\bar{d}}}$$

- Where $\bar{d}$ is the sample mean difference, μ_{d0} is the hypothesized population mean difference, $S_{\bar{d}} = S_d/\sqrt{n}$, n is the number of sample differences, and S_d is the standard deviation of the sample differences. When H_0 is true, the test statistic is distributed as Student's t with n-1 degrees of freedom.

Example 8:

Nancy Stearns Burgess(A-11) conducted a study to determine weight loss, body composition, body fat distribution , and resting metabolic rate in obese subjects before and after 12weeks of treatment with a very low caloric diet (VLCD) and to compare hydrodensitometry with bioelectrical impedance analysis. The 17 subjects(9 women and 8 men) participating in the study were from an outpatients, hospital based treatment program for obesity. The women's weights before and after the 12 week VLCD treatment are in Table. We wish to know if these data provide sufficient evidence to allow us to conclude that the treatment is effective in causing weight reduction in obese woman.

B	117.3	111.4	98.6	104.3	105.4	100.4	81.7	89.5	78.2
A	83.3	85.9	75.8	82.9	82.3	77.7	62.7	69.0	63.9

5. Hypothesis Testing: A single population Proportion

- Testing hypothesis about population proportions is same as for means when the condition necessary for using the normal curve are met.
- When a sample sufficiently large for application of the central limit theorem, the test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

- When H_0 is true, it is distributed approximately as the standard normal.

Example 8: In a survey of injection drug users in a large city found that 18 out of 423 were HIV positive. Whether we conclude that fewer than 5% of the injection drug users in the sampled population are HIV positive.

6. Hypothesis Testing : The difference between two population proportion

- ❑ Let p_1 and p_2 are the population proportions of some character under study from two populations.
- ❑ When the null hypothesis to be tested is $p_1 - p_2 = 0$, we are hypothesizing that the two population proportion are equal.
- ❑ This is the justification of combining the result of the two samples to come up with a estimate of hypothesized common
- ❑ $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ where x_1 and x_2 are the numbers in the first and second samples.

The estimated standard error of the estimator,

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_1} + \frac{\bar{p}(1 - \bar{p})}{n_2}}$$

The test statistic $z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}}$ which is distributed approximately as the standard normal if the null hypothesis is true.

Example:

In a study of nutrition care in nursing homes among 55 patients with hyper tension, 24 were on sodium restricted diets. Of 149 patients without hypertension, 36 were on sodium restricted diets. May we conclude that in the sampled population the proportion of patients on sodium restricted diets is higher among patients with hypertension than among patients without hypertension?

7.Hypothesis Testing: A single population Variance

- When the data available for analysis consists of a simple random sample drawn from a normally distributed population, the test statistic for testing hypothesis about a population variance is

when H_0 is true, is distributed as χ^2 with $n-1$ degrees of freedom.

$$\chi^2 = (n-1)s^2/\sigma^2$$

Example:

The release of preformed and newly generated mediators in the immediate response to allergen inhalation in allergic primates. Subjects were 12 wild-caught, adult male cynomolgus monkeys meeting certain criteria of the study. Among the data reported by the investigator was a standard error of the sample mean of 0.4 for one of the mediator received from the subject by bronchoalveolar lavage(BAL). We wish to know if we may conclude from these data that the population variance is not 4.

For $H_a: \sigma^2 > \sigma_0^2$ (reject H_0 if χ^2 computed.)

For $H_a: \sigma^2 < \sigma_0^2$ (reject H_0 if χ^2 computed.)

8. Hypothesis Testing: The ratio of two population variance

Variance Ratio Test:

- Decision regarding the comparability of two population variances are usually based on the variance ratio test.
- When two population variances are equal, testing the hypothesis that their ratio is equal to 1.
- If $\sigma_1^2 = \sigma_2^2$, then the hypothesis is true, and the two variances cancel out in the above expression leaving s_1^2 / s_2^2 , which follows the same F distribution.
- The ratio s_1^2 / s_2^2 will be designated V.R. for variance ratio.

Say σ_1^2 and σ_2^2 be the population variance of two population. n_1 and n_2 be the sample from population 1 and 2.

If $H_0: \sigma_1^2 = \sigma_2^2$

$H_a: \sigma_1^2 \neq \sigma_2^2$

$$\frac{s_1^2}{s_2^2} \approx F, n_1 - 1$$

numerator d.f and $n_2 - 1$ denominator d.f.

For a two sided test we put larger sample variance in the numerator and obtain the critical value for $\alpha/2$ and appropriate d.f.

One sided test

$H_0: \sigma_1^2 \leq \sigma_2^2$

$H_a: \sigma_1^2 > \sigma_2^2$

Appropriate test statistics V.R

$$\frac{s_1^2}{s_2^2}$$

F_{α} is obtained for appropriate d.f.

$$H_0: \sigma_1^2 \geq \sigma_2^2$$

$$H_a: \sigma_1^2 < \sigma_2^2$$

$$\text{Test statistic V.R} = \frac{s_1^2}{s_2^2}$$

F_{α} is obtained for α and appropriate d.f.

Decision rule in all these cases is that reject H_0 if computed.

Example:

To investigate alterations of thermoregulation in patients with certain pituitary adenomas (P) the standard deviation of the weights of a sample of 12 patients was 21.4kg. The weights of a sample of five control subjects yielded a standard deviation of 12.4kg. Whether we conclude that the weights of the population represented by the sample of patients are more variable than the weights of the population represented by the sample of control subjects.

9. The Type II Error and The Power of a test

- α is a measure of the acceptable risk of rejecting a true null hypothesis.
- But β may assume one of many values.
- To test the null hypothesis for some population parameter is equal to some specified value. If H_0 is false and we fail to reject it then it is a type- II error.
- If the hypothesized value of parameter is not the true value, the value of β depends on several factors.
 - the true value of the parameter of interest
 - hypothesized value of the parameter
 - the value of α
 - the sample size n
- For fixed α and n we compute many values of β by postulating many values for the parameter.

- If H_0 is in fact false , we would like to know probability of its rejection. The power of test, designated $1-\beta$ this information.
- The quantity $1-\beta$ is the probability of reject a null hypothesis ; it may be computed for any alternative value of the parameter .
- For a given test number of possible values of the parameter of interest and for each compute the value of $1-\beta$. The result is called a power function.
- The graph of a power function is called a power curve, helpful device for quickly assessing the nature of the power of a given test.

The background is a solid blue color with a subtle grid pattern of thin, lighter blue lines. The grid consists of both horizontal and vertical lines, creating a series of small squares across the entire image.

THANK YOU