

Test of Hypothesis

Lecture 1

Statistical Hypothesis:

A statistical hypothesis is some statement or assertion about a population or equivalently about the probability distribution characterizing a population which we want to verify on the basis of information available from a sample.

Test of Significance:

Test of significance is a statistical procedure to arrive at a conclusion or decision on the basis of samples and to test whether the formulated hypothesis can be accepted or rejected in probability sense. The aim of test of significance is to reject the null hypothesis.

Null Hypothesis:

According to Prof. R. A. Fisher, the hypothesis which we are going to test for possible rejection under the assumption that it is true is called the null hypothesis. Usually it is denoted by H_0 .

Alternative Hypothesis:

Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis. Usually it is denoted by H_1 .

Example:

If we want to test the null hypothesis that the population has a specified mean μ_0 (say) i.e. $H_0 : \mu = \mu_0$ then the alternative hypothesis could be $H_1 : \mu \neq \mu_0$ or $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$.

Simple Hypothesis:

If the hypothesis specifies the population completely then it is termed as a simple hypothesis.

Example:

If $x_1, x_2, \dots, x_n$ is a random sample of size n from a normal population with mean μ and variance σ^2 , then the hypothesis $H_0 : \mu = \mu_0, \sigma^2 = \sigma_0^2$ is a simple hypothesis.

Composite Hypothesis:

If the hypothesis does not specify the population completely then it is termed as a composite hypothesis.

Example:

If $x_1, x_2, \dots, x_n$ is a random sample of size n from a normal population with mean μ and variance σ^2 then each of the following hypothesis is a composite hypothesis.

$\rightarrow \mu = \mu_0$	$\rightarrow \sigma^2 = \sigma_0^2$
$\rightarrow \mu < \mu_0, \sigma^2 = \sigma_0^2$	$\rightarrow \mu > \mu_0, \sigma^2 = \sigma_0^2$
$\rightarrow \mu = \mu_0, \sigma^2 < \sigma_0^2$	$\rightarrow \mu = \mu_0, \sigma^2 > \sigma_0^2$
$\rightarrow \mu < \mu_0, \sigma^2 > \sigma_0^2$	$\rightarrow \mu > \mu_0, \sigma^2 < \sigma_0^2$

Note: A hypothesis which does not specify completely r parameters of a population is termed as a composite hypothesis with r degrees of freedom.

Parametric Hypothesis:

When the hypothesis concerning the parameters of the distribution, provided the form of the distribution is called parametric hypothesis.

Example:

If $X \sim N(\mu, \sigma^2)$ then $H_0 : \mu = \mu_0$ is a parametric hypothesis.

Non-parametric Hypothesis:

While the hypothesis regarding the form of the distribution with specified or unspecified parameters is called non-parametric hypothesis.

Example:

$H_0 : F_X(x) = F_Y(y)$ is a non-parametric hypothesis, where $F_X(x)$ and $F_Y(y)$ are the distribution function of two population.

Type I Error:

If the null hypothesis is actually true but on the basis of sample we reject H_0 (accepting H_1), this type of error is known as type I error.

The probability of type I error is denoted by α . Thus

$$\begin{aligned}\alpha &= \text{Probability of type I error} \\ &= \text{Probability of rejecting } H_0 \text{ when } H_0 \text{ is true}\end{aligned}$$

Symbolically, $P(x \in w | H_0) = \alpha$; where $x = (x_1, x_2, \dots, x_n)$

Example: In a murder case suppose that

H_0 : The man is not guilty

H_1 : The man is guilty

On the basis of sample if we reject the H_0 but it is true this type of error is known as type I error.

Type II Error:

If the null hypothesis is false but on the basis of sample we accept H_0 (H_1 is true), this type of error is known as type II error.

The probability of type II error is denoted by β . Thus

$$\begin{aligned}\beta &= \text{Probability of type II error} \\ &= \text{Probability of accepting } H_0 \text{ when } H_0 \text{ is false}\end{aligned}$$

Symbolically, $P(x \in \bar{w} | H_1) = \beta$; where $x = (x_1, x_2, \dots, x_n)$

Degrees of freedom

Degrees of freedom are the number of independent values that a statistical analysis can estimate. Degrees of freedom encompasses the notion that the amount of independent information you have limits the number of parameters that you can estimate. Typically, the degrees of freedom equal your sample size minus the number of parameters you need to calculate during an analysis. It is usually a positive whole number.

One Tailed Test:

A test of any statistical hypothesis where the alternative hypothesis is one tailed (right or left tailed) is called a one tailed test. In such a case the critical region is given by the portion of the area lying in the first or last tails of the probability curve of the test statistic.

Example:

A test for testing the mean of a population $H_0 : \mu = \mu_0$ against the alternative hypothesis $H_1 : \mu > \mu_0$ (*Right tailed*) or $H_1 : \mu < \mu_0$ (*left tailed*) is a one tailed test. In the right tailed test $H_1 : \mu > \mu_0$, the critical region lies entirely in the right tail of the sampling distribution of $\bar{x}$, while for the left tail test $H_1 : \mu < \mu_0$, the critical region is entirely in the left tail of the distribution.

Two Tailed Test:

A test of statistical hypothesis where the alternative hypothesis is two tailed as called two tailed test. In such a case the critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic.

Example:

Suppose that there are two population brands of bulbs, one manufactured by standard process (with mean life μ_1) and the other manufactured by some new technique (with mean life μ_2). If we want to test if the bulbs differ significantly then our null hypothesis is $H_0 : \mu_1 = \mu_2$ and the alternative will be $H_1 : \mu_1 \neq \mu_2$ this giving us a two tailed test.

Steps in Solving Testing of Hypothesis Problem:

The major steps involved in the solution of a “testing of hypothesis” problem may be outlined as follows:

- ❖ Explicit knowledge of the nature of the population distribution and the parameter(s) of interest, i.e., the parameter(s) about which the hypothesis are set up.
- ❖ Setting up of the null hypothesis H_0 and the alternative hypothesis H_1 in terms of the range of the parameter values each one embodies.

- ❖ Choose the appropriate level of significance (α) depending on the reliability of the estimates and permissible risk.
- ❖ Choose the suitable test statistic and compute the test statistic under the null hypothesis.
- ❖ We compare the computed value of test statistic with the significant value (tabulated value) at the given level of significance α .
 - If the computed value of test statistic (in modulus value) is greater than the significant value (tabulated value) then we reject the null hypothesis at level of significance α and we say that it is significant.
 - If the computed value of test statistic is less than the tabulated value then we accept the null hypothesis at level of significance α and we say that it is not significant.

Best Critical Region:

Let C denote a subset of the sample space. Then C is called a best critical region of size α for testing the simple hypothesis $H_0: \theta = \theta'$ against the alternative simple hypothesis $H_1: \theta = \theta_n$ if for every subset A of the sample space for which

$$\begin{aligned}
 &P\{(x_1, x_2, \dots, x_n) \in A; H_0\} = \alpha \\
 a) &P\{(x_1, x_2, \dots, x_n) \in C; H_0\} = \alpha \\
 b) &P\{(x_1, x_2, \dots, x_n) \in C; H_1\} \geq P\{(x_1, x_2, \dots, x_n) \in A; H_1\}
 \end{aligned}$$

The best critical region is abbreviated as BCR.

Most Powerful Test (MP test):

The test based on a BCR is called most powerful (MP) test. Let us consider the problem of testing a simple hypothesis $H_0: \theta = \theta_0$ against a simple alternative hypothesis $H_1: \theta = \theta_1$.

The critical region ω is the most powerful (MP) critical region of size α and the corresponding test is a most powerful test of level α for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$ if

$$P(x \in \omega | H_0) = \int_{\omega} L_0 dx = \alpha \quad \dots \quad \dots \quad \dots \quad (1)$$

$$and \quad P(x \in \omega | H_1) \geq P(x \in \omega_1 | H_1) \quad \dots \quad \dots \quad \dots \quad (2)$$

for every other critical region ω_1 satisfying (1).

Example of BCR:

Consider the random variable X that has a binomial distribution with $n=5$ and $p=\theta$. Let

$f(x; \theta)$ denote the *p.d.f.* of X and let $H_0: \theta = \frac{1}{2}$ against $H_1: \theta = \frac{3}{4}$ then the following

tabulation, at points of positive probability density, the values of $f\left(x; \frac{1}{2}\right)$, $f\left(x; \frac{3}{4}\right)$ and the

$$\text{ratio } \frac{f\left(x; \frac{1}{2}\right)}{f\left(x; \frac{3}{4}\right)}.$$

x	0	1	2	3	4	5
$f\left(x; \frac{1}{2}\right)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$
$f\left(x; \frac{3}{4}\right)$	$\frac{1}{1024}$	$\frac{15}{1024}$	$\frac{90}{1024}$	$\frac{270}{1024}$	$\frac{405}{1024}$	$\frac{243}{1024}$
$\frac{f\left(x; \frac{1}{2}\right)}{f\left(x; \frac{3}{4}\right)}$	32	$\frac{32}{3}$	$\frac{32}{9}$	$\frac{32}{27}$	$\frac{32}{81}$	$\frac{32}{243}$

$$\text{Where, } f\left(x; \frac{1}{2}\right) = \binom{5}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} ; \quad x=0, 1, \dots, 5$$

$$\text{and } f\left(x; \frac{3}{4}\right) = \binom{5}{x} \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{5-x} ; \quad x=0, 1, \dots, 5$$

Now, we first assign the significance level of the test to be $\alpha = \frac{1}{32}$.

We seek a best critical region of size $\alpha = \frac{1}{32}$. Now, we have,

$$P\{X=0\} = \frac{1}{32} \quad \text{and} \quad P\{X=5\} = \frac{1}{32}$$

So, we have two critical region $X=0$ and 5 . Now we find out which is best. From the table we have,

$$P\{X=0|H_1\} = \frac{1}{1024} \quad \text{and} \quad P\{X=5|H_1\} = \frac{243}{1024}$$

$$\text{Since } P\{X=5|H_1\} = \frac{243}{1024} > P\{X=0|H_1\} = \frac{1}{1024}$$

So, the best critical region is at $x=5$.

Another way, in which point the ratio $\frac{f\left(x; \frac{1}{2}\right)}{f\left(x; \frac{3}{4}\right)}$ is minimum that point is a BCR. From the

table we have the ratio $\frac{f\left(x; \frac{1}{2}\right)}{f\left(x; \frac{3}{4}\right)}$ is minimum at $x=5$. So, the best critical region is at $x=5$ for

test the hypothesis $H_0: \theta = \frac{1}{2}$ against $H_1: \theta = \frac{3}{4}$. Again, let $\alpha = \frac{6}{32}$ then we have,

$$P\{X = 0, 1\} = \frac{6}{32}$$

$$P\{X = 0, 4\} = \frac{6}{32}$$

$$P\{X = 1, 5\} = \frac{6}{32}$$

$$P\{X = 4, 5\} = \frac{6}{32}$$

By direct computation it is found that the best critical region of this size is $\{X = 4, 5\}$. Since

the ratio $\frac{f\left(x; \frac{1}{2}\right)}{f\left(x; \frac{3}{4}\right)}$ is minimum at $x = 4, 5$.

The power of this test, which has $\alpha = \frac{6}{32}$ is:

$$P\{X = 4, 5 | H_1\} = \frac{405}{1024} + \frac{243}{1024} = \frac{648}{1024} \text{ which is maximum.}$$

The p -Value Approach to Hypothesis Testing:

The p -value is the probability of obtaining a test statistic equal to or more extreme than the result obtained from the sample data, given that the null hypothesis H_0 is really true.

The p -value is often referred to as the observed level of significance, which is the smallest level at which H_0 can be rejected for a given set of data. The decision rule for rejecting H_0 in the p -value approach is:

- If the p -value greater than or equal to α , the null hypothesis is not rejected.
- If the p -value is smaller than α , the null hypothesis is rejected.

Steps in Determining the p -value :

- ❖ State the null hypothesis H_0
- ❖ State the alternative hypothesis H_1 .
- ❖ Choose the level of significance α .
- ❖ Choose the sample size n .
- ❖ Determine the appropriate statistical technique and corresponding test statistic to use.
- ❖ Collect the data and compute the sample value of the appropriate test statistic.
- ❖ Calculate the p -value based on the test statistic. This involves
 - Sketching the distribution under the null hypothesis H_0 .
 - Placing the test statistic on the horizontal axis.
 - Shading in the appropriate area under the curve, on the basis of the alternative hypothesis H_1 .
- ❖ Compare the p -value to α .
- ❖ Make the statistical decision. If the p -value is greater than or equal to α , the null hypothesis is not rejected. If the p -value is smaller than α , the null hypothesis is rejected.
- ❖ Express the statistical decision in terms of the particular situation.

Confidence Interval

A confidence interval is the probability that a value will fall between an upper and lower bound of a probability distribution. For example, given a 99% confidence interval, stock XYZ's return will fall between -6.7% and +8.3% over the next year. In layman's terms, you are 99% confident that the returns of holding XYZ stock over the next year will fall between -6.7% and +8.3%.

Statisticians use confidence intervals to measure uncertainty. A higher probability associated with the confidence interval means that there is a greater degree of certainty that the

parameter falls within the bounds of the interval. Therefore, a higher confidence level indicates that the parameters must be broader to ensure that level of confidence.

Calculating the Confidence Interval

For example, suppose a group of researchers is studying the heights of high school basketball players. The researchers take a [random sample](#) from the population and establish a mean height of 74 inches. The mean of 74 is a point estimate of the population mean. A point estimate by itself is of limited usefulness because it does not reveal the uncertainty associated with the estimate; you do not have a good sense of how far this sample mean may be from the population mean. What they are missing at this point is the degree of uncertainty in this single sample.

Confidence intervals provide more information than point estimates. By establishing a 95% confidence interval using the sample's [mean](#) and [standard deviation](#), and assuming a [normal distribution](#) as represented by the [bell curve](#), the researchers arrive at an upper and lower bound that contains the true mean 95% of the time. Assume the interval is 72 inches to 76 inches. If the researchers take 100 random samples from the population of high school basketball players as a whole, the mean should fall between 72 and 76 inches in 95 of those samples.