Testing of Hypothesis

A statistical test of hypothesis is a procedure for assessing the compatibility of data with the claim about the parameter value. Keeping the objective of the study in mind, we formulate a statement (called hypothesis) regarding the value of the population parameter and see if the data are consistent or inconsistent with that statement.

Basic Concepts and Terminology:

This section introduces the basic terminology and the concepts needed for studying the tests of hypothesis.

Statistical hypothesis:

A statistical hypothesis (or simply hypothesis) is a conjecture about the probability distribution of one or more random variables. It often involves one or more parameters of the distribution. Mostly, the hypothesis are stated in terms of population parameter.

Since the conjecture may be true or false, we may think of two types of hypothesis that are complementary of each other, viz, null and alternative hypothesis.

Null and alternative hypothesis :

A statement or conjuncture about the value of a parameter which we adopt until it is proven false is called null hypothesis. It is denoted by ~~Ho (read~~ (pronounced as 'H-naught'), followed by a colon (:). It is a statement about the parameter which is supported unless the data provide convincing evidence against it.

The word 'null' is to be interpreted as 'no difference' between the true value of the parameter and its hypothesized value.

A hypothesis against which the null hypothesis is tested is known as an ~~alter~~ alternative hypothesis. It is usually denoted by H or H_a or H_1 etc. It contradicts the null hypothesis. An alternative hypothesis is also known as a research hypothesis, because it is ~~At~~ often the hypothesis of researcher's concern. The alternative hypothesis are of two types viz. one-sided alternative (directional) and two-sided (non-directional) alternative.

Suppose we wish to test a null hypothesis (H_0) about an unknown population mean μ, given by

$H_0 : \mu = \mu_0$

against alternative hypothesis

$H_1 : \mu > \mu_0$

or, $H_2 : \mu < \mu_0$

or, $H_3 : \mu \neq \mu_0$

Here H_1 and H_2 are called one-sided ~~alter~~ alternative and H_3 is two-sided. To be more specific, H_1 is a right-sided alternative and H_2 is a left-sided alternative.

Here μ_0 is the value of the parameter μ specified by the null hypothesis. It is called a hypothesized value. The value of μ_0 is known.

Simple and Composite hypothesis:

If the population distribution is completly specified by the hypothesis, then it is called a simple hypothesis; composite otherwise.

Test of hypothesis :

A test of hypothesis or a test of significance is a procedure for assessing the compatibility of the data with the null hypothesis. It provides us with the ways of using sample data ~~o~~to decide between the two competing hypothesis, null and alternative. It is a rule for ~~e~~deciding whether to reject or not to reject H_0, on the basis of the sample data.

At the outset, we assume that the null hypothesis (H_0) is true. Then we want to see if the data contradict the assumption. We reject H_0 if there is enough convincing evidence against H_0. (H_0) is not rejected, otherwise. Two possible conclusions of any test could be: 'reject H_0' or 'fail to reject H_0'. In ~~the~~ latter case, we say "H_0 cannot be rejected", but we should not say, 'H_0 is accepted'. According to the astronomer Carl Sagan, "absence of evidence is not evidence of absence". In loose sense the term 'acceptance' in ...

'non-rejection' of H_0.

From the above discussion it is clear that we must be careful in setting up the hypothesis for a test. A test is only capable of ~~demonstratin~~ demonstrating a strong support for the alternative hypothesis by rejection of H_0. But remember that when H_0 is not rejected, it does not mean a strong support for H_0; it only indicates insufficient evidence or lack of strong evidence against H_0. Consider the example of contents of softdrinks bottles. There we test $H_0: \mu = 300$ against $H_a: \mu < 300$. If H_0 is rejected, that indicates a strong support for '$\mu < 300$', i.e, the mean content is less than 300 ml, whereas non-rejection of H_0 indicates that there is insufficient evidence against $H_0: \mu = 300$, i.e, we do not have enough evidence to say that mean content differs from (less than) 300 ml. In latter case we do not say that there is a strong support for $\mu = 300$ (ml).

Errors in hypothesis testing:

In hypothesis testing 'rejection of H_0' does not mean that 'H_0 is false' (only means that the data are incompatible with H_0). Neither does 'non-rejection of H_0' mean that 'H_0 is true' (only means that the data provide insufficient evidence against H_0). It may happen that the data lead to the rejection of a true null hypothes, H_0, or non-rejection of false H_0. Thus two types of errors may creep ~~into~~ in to the statistical test.

Decision	Actual situation	
	H_0 is true	H_0 is false
Reject H_0	Type I error	Correct decision
Do not reject H_0	Correct decision	Type II error

Type I and Type II error:

Type I error is the error of rejecting the null hypothesis H_0, when it is true. Type II error is the error of not rejecting the null hypothesis H_0, when it is false.

The following decision table shows all the possibilities:

To understand the idea of type I and type II errors consider the following example of an accused person who is tried in a court of law being indicted for committing crime.

At the beginning of the trial the person is presumed not guilty. Here the appropriate hypothesis are:

H_0: The accused person is innocent

H_a: The accused person is guilty.

The person, actually, may or may not be guilty, which is not known. The null hypothesis H_0 that the person is innocent should be rejected if there is enough evidence against it. otherwise H_0 should not be rejected.

On the basis of the evidence provided, if the guilty person is set free, then that indicates that the judge committed a type II error. On the other hand, if the person is actually innocent, but the evidence goes against ~~them~~

him, i.e., he is wrongly proven to be guilty and is punished, then that would be a type I error. To guard against the situation that the innocent person will be punished by mistake a small probability of type I ~~and~~ error should be considered.

Level of Significance (α):

The maximum probability of type I error is called the level of significance or the size of the test, and is denoted by α. Usually we take $\alpha = 0.05$ or 0.01.

Actually, $P(\text{type I error}) \leq \alpha$,

i.e, α represents the largest tolerable risk of incorrectly rejecting H_0.

The probability of type II error is denoted by β.

Power of the test:

The probability of rejecting H_0 when it is false is known as the power of a test. It is the power of a test in detecting the falsity of H_0.

power of a test = 1 − P(type II error)
$= 1 - \beta$

We try to design the test rule in such a way that the error ~~probability~~ probabilities are minimized. But it is not possible to reduce both the error probabilities, α and β, simultaneously. Therefore, we try to increase the power of the test subject to a fixed level α, i.e, we try to minimize β ~~(i.e. maximize~~ (i.e, maximize power) for a fixed α.

Test Statistic

A statistic, a function of sample observations used to test the null hypothesis is known as a test statistic. A conclusion to reject or not to reject H_0 is based on the value of the test statistic, observed in a sample ~~so~~ selected.

There are two approaches for making a decision regarding rejecting or non-rejection of H_0, viz.,

- p-value approach
- critical value approach.

Both the approaches lead to the same conclusion. Let us first define p-value and critical value.

<u>p-value:</u>

p-value is the probability that a test statistic will take the value as extreme as or more extreme than the one actually observed, when H_0 is true. It is also known as observed level of significance.

The p-value is the smallest value of the level of significance α for which H_0 can be rejected. It is the actual risk of committing a type I error if H_0 is rejected based on the observed value of the test statistic. The p-value measures the strength of the evidence against H_0. We reject H_0 at level α if p-value $\leq \alpha$. Then we can report that the results are statistically significant at level α.

Critical region, non-rejection ~~rejt~~ region and critical value:

The entire set of possible values of test statistic is divided into two sets or regions. One region consisting of the values that support the alternative hypothesis and leads to the rejection of H_0 is called the critical region or rejection region. The ~~oth~~ other consisting of the values that support H_0 is called the non-rejection region (often called acceptance region). The value of the test statistic which lies at the boundary of critical region and acceptance region is known as critical value. It separates the critical region and acceptance region. The critical value depends on the level of significance, α, the risk of making an incorrect decision regarding rejection of true H_0.

One-tailed and Two-tailed test:

When the critical region is in one tail of the distribution of test statistic, the test is called a one-tailed test. If, in particular, it is in the left tail (right tail) of the distribution, then it is called a left-tailed (right-tailed) test. When the

Scanned by CamScanner

critical region is in both tails of the ~~distribu~~ distribution, the test is called a two-tailed test. Left-tailed (right-tailed) test corresponds to the left-sided (right-sided) alternative and two-tailed test corresponds to the two-sided alternative.

Randomized and non-randomized test:

A test is called a non-randomized test if we reject H_0 if and only if the observed value of the test statistic falls in the critical region.

A test is called a randomized test if 'rejection of H_0' depends on the outcome of a random experiment. For instance, reject H_0 if and only if an even number appears in a ~~die-thr~~ die-throwing experiment.