

TESTING OF HYPOTHESIS

In applied investigations or in experimental research, one may wish to estimate the yield of a new hybrid line of corn, but ultimate purpose will involve some use of this estimate. One may wish, for example, to compare the yield of new line with that of a standard line and perhaps recommend that the new line replaces the standard line if it appears superior. This is the common situation in research. One may wish to determine whether a new method of sealing light bulbs will increase the life of the bulbs, whether a new germicide is more effective in treating a certain infection than a standard germicide, whether one method of preserving foods is better than the other so far as the retention of vitamin is concerned, which one among the six available varieties of any crop is best in terms of yield per hectare.

Using the light bulb example as an illustration, let us suppose that the average life of bulbs made under a standard manufacturing procedure is about 1400 hours. It is desired to test a new procedure for manufacturing the bulbs. Here, we are dealing with two populations of light bulbs: those made by the standard process and those made by the proposed process. From the past investigations, based on sample tests it is known that the mean of the first population is 1400 hours. The question is whether the mean of the second population is greater than or less than 1400 hours? This we have to decide on the basis of observations taken from a sample of bulbs made by second process.

In making comparisons of above type, one cannot rely on the mere numerical magnitudes of the index of comparison such as mean, variance, etc. This is because each group is represented only by a sample of observations and if another sample were drawn, the numerical value would change. This variation between samples from the same population can at best be reduced in a well designed experiment but can never be eliminated. One is forced to draw inference in the presence of the sampling fluctuations which affect the observed differences between the groups, clouding the real differences. Hence, we have to devise some statistical procedure, which can test whether those difference are due to chance factors or really due to treatment.

The tests of hypothesis are such statistical procedures which enable us to decide whether the differences are attributed to chance or fluctuations of sampling. In order to fix the ideas, let us define some concepts involved in testing of hypothesis.

Sample space: The set of all possible outcomes of an experiment is called sample space. It is denoted by S . For example in an experiment of tossing two coins simultaneously, the sample space is $S = \{HH, HT, TH, TT\}$; where 'H' denotes the head and 'T' denotes the tail outcomes. In testing of hypothesis, we are concerned with drawing inferences about the population based on a random sample. Let there are 'N' units in a population and we

have to draw sample of size 'n'. Then the set of all possible samples of size 'n' is the sample space and any sample $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is the point of the sample space.

Parameter: A function of population values is known as parameter. For example, population mean (μ) and population variance (σ^2).

Statistic: A function of sample values say, (x_1, x_2, \dots, x_n) is called a statistic. For example, sample mean (\bar{x}), sample variance (s^2), where

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}; \quad s^2 = \frac{1}{n-1} [(x_1^2 + x_2^2 + \dots + x_n^2) - nx^{-2}].$$

A statistic does not involve any unknown parameter.

Sampling distribution: The distribution of a statistic computed for all the possible samples is known as sampling distribution of that statistic.

Statistical Hypothesis: A Statistical Hypothesis is an assertion or conjecture (tentative conclusion) either about the form or about the parameter of the distribution. For example

- i) The normal distribution has mean 20.
- ii) The distribution of process is Poisson.
- iii) Effective life of a bulb is 1400 hours.
- iv) A given detergent cleans better than any washing soap.

In a statistical hypothesis, all the parameters of a distribution may be specified completely or partly. *A statistical hypothesis in which all the parameters of a distribution are completely specified is called simple hypothesis, otherwise, it is known as composite hypothesis.* For example, in case of normal population, the hypothesis

- i) Mean (μ) = 20, variance (σ^2) = 5 (Simple hypothesis)
- ii) $\mu = 20, \sigma^2 > 1$ (composite hypothesis)
- iii) $\mu = 20$ (composite hypothesis)

Null Hypothesis: The statistical hypothesis under sample study is called null hypothesis. It is usually that the observations are the result purely of chance. It is usually denoted by H_0 .

Alternative Hypothesis: In respect of every null hypothesis, it is desirable to state, what is called an alternative hypothesis. It is complementary to the null hypothesis. Or "The desirable attitude of the statistician about the hypothesis is termed as alternative hypothesis". It is taken usually that the observations are the result of real effect plus chance variation. It is usually denoted by H_1 . For example, if one wishes to compare the yield per hectare of a new line with that of standard line, then, the null hypothesis:

H_0 : Yield per hectare of new lie (μ_1)=Yield per hectare of standard line (μ_2)

The alternative hypothesis corresponding to H_0 , can be the following:

- i) $H_1 : \mu_1 > \mu_2$ (Right tail alternative)
- ii) $H_1 : \mu_1 < \mu_2$ (left tail alternative)
- iii) $H_1 : \mu_1 \neq \mu_2$ (Two tailed alternative)

(i) and (ii) are called one tailed test and (ii) is a two tailed test. Whether one sets up a one tailed test or a two-tailed test depends upon the conclusion to be drawn if H_0 is rejected.

The location of the critical region will be decided only after H_1 has been stated. For example in testing a new drug, one sets up the hypothesis that it is no better than similar drugs now on the market and tests this hypothesis against the alternative hypothesis that the new drug is superior. Such an alternative hypothesis will result in a one tailed test (right tail alternative).

If we wish to compare a new teaching technique with the conventional classroom procedure, the alternative hypothesis should allow for the new approach to be either inferior or superior to the conventional procedure. Hence the test is two-tailed.

To fix up the idea consider the following examples.

Example 1: The manufacturer of a certain brand of cigarettes claims that the average nicotine content does not exceed 2.5 mg. State the null hypothesis and alternative hypothesis to be used in testing this claim.

Solution: The manufacturer's claim should be rejected only if μ is greater than 2.5 mg and should be accepted if μ is less than or equal to 2.5 mg. Therefore, $H_0 : \mu = 2.5$ and $H_1 : \mu > 2.5$

Example 2: A real-estate agent claims that 60% of all private residences being built today are 3-bedroom homes. To test this claim, a large sample of new residences are inspected; the proportion of these homes with 3 bedrooms is recorded and used as our test statistic. State the null and alternative hypotheses to be used in this test.

Solution: If the test statistic is substantially higher or lower than $p = 0.6$, we should reject the agent's claim. Hence, we should make the test $H_0 : p = 0.6$ and $H_1 : p \neq 0.6$.

The framing of null hypothesis (H_0) versus alternative hypothesis (H_1), the four possibilities in order of preference are

S.No.	H_0	H_1
1	S(simple)	S
2	S	C(Composite)
3	C	S
4	C	C

The case when both H_0 and H_1 are simple is used sparingly. Since 'S versus C' is preferable to 'C versus S', we should always structure the problem in such a way that H_0 is simple one in the parameter of interest.

After framing the null hypothesis and alternative hypothesis, we would like to find a rule or procedure that will enable us to decide whether to reject null hypothesis (H_0) or not. Such a rule or procedure is called **test of hypothesis**.

For making a decision, the whole sample space 'S' is partitioned into, two disjoint and complementary regions C and $\bar{C} = S - C$ such that if sample point x belongs to C, we reject H_0 otherwise, accept H_0 .

Critical Region: It is region of rejecting null hypothesis when it is true if sample point belongs to it. Hence 'C' is the critical region.

Suppose that if the test is based on a sample of size 2, then the outcome set or sample space is the first quadrant in a two dimensional space and a test criterion will enable us to separate our outcome set into two complementary subsets, C and \bar{C} . If the sample point falls in the subset C, H_0 is rejected, otherwise, H_0 is accepted. This is shown in the following figure.

The terms acceptance and rejection are, however, not to be taken in their literal senses. Thus acceptance of H_0 does not mean that H_0 has been proved true. It means only that so far as the given observations are concerned, there is no evidence to believe otherwise. Similarly, rejection of H_0 does not disprove the hypothesis; it merely means that H_0 does not look plausible in the light of given observations.

It is now known that, in order to establish the null hypothesis, we have to study the sample instead of entire population. Hence, whatever, decision rule we may employ, there is every chance of committing errors in the decision for rejecting or accepting the hypothesis. Four possible situations, which can arise in any test procedure, are given in the following table.

True situation → Decision made ↓	H_0 is true	H_0 is false or H_1 is true
Reject H_0	Reject H_0 / H_0 is true (Type I error)	Reject H_0 / H_1 is true
Do not reject H_0 Or Accept H_1	Do not reject H_0 / H_0 is true	Accept H_0 / H_1 is true (Type II error)

From the table, it is clear that the errors committed in making decisions are of two types.

Type I error: Reject H_0 when H_0 is true.

Type II error: Accept (does not reject) H_0 when H_0 is false.

For example, a judge, who has to decide whether the person has committed the crime. The statistical hypothesis in this case is:

H_0 : Person is innocent; H_1 : Person is criminal.

In this situation, two types of errors which the judge may commit are:

Type I error: Innocent person is found guilty and punished.

Type II error: A guilty person is set free.

Since, it is more serious to punish an innocent than to set free a criminal. Therefore, Type I error is more serious than the Type II error.

Probabilities of the errors:

Probability of Type I error = $P(\text{Reject } H_0 / H_0 \text{ is true}) = \alpha$

Probability of Type II error = $P(\text{Accept } H_0 / H_1 \text{ is true}) = \beta$

In quality control terminology, Type I error amounts to rejecting a lot when it is good and Type II error may be regarded as accepting a lot when it is bad.

$P(\text{Reject a lot when it is good}) = \alpha$ (producer's risk)

$P(\text{Accept a lot when it is bad}) = \beta$ (consumer's risk)

Level of significance: the probability of Type I error (α) is called the level of significance. It is also known as size of the critical region.

Although 5% level of significance has been taken as a rough line demarcation in which deviations due to sampling fluctuations alone will be interpreted as real ones in 5% of the cases. Hence, yet the inferences about the population based on samples are subject to some degree of uncertainty. It is not possible to remove this uncertainty completely, but it can be reduced by choosing the level of significance still lesser like 1%, in which chances of interpreting the deviations due to sampling fluctuations, as real one, is only one in 100 cases.

Power function of a Test: The probability of rejecting H_0 when H_1 is true is called power function of the test.

$$\begin{aligned}\text{Power function} &= P(\text{Reject } H_0 / H_1 \text{ is true}) \\ &= 1 - P(\text{Accept } H_0 / H_1 \text{ is true}) \\ &= 1 - \beta.\end{aligned}$$

The value of this function plays the same role in hypothesis testing as the mean square error plays in estimation. It is usually used as our standard in assessing the goodness of a

test or in comparing two tests of same size. The value of this function for a particular point is called the power of the test.

In testing of hypothesis, the ideal procedure is to minimize the probabilities of both type of errors. Unfortunately, for a fixed sample size n , both the probabilities cannot be controlled simultaneously. A test which minimizes the one type of error, maximizes the other type of error. For example, if there is a critical region which makes the probability of Type I error zero, will be of the form always accept H_0 and that the probability of Type II error will be one. It is, therefore, desirable to fix the probability of one of the error and choose a critical region which minimizes the probability of the other error. As Type I error is consider to be more serious than Type II error. Therefore, we fix the probability of Type I error (α) and minimize the probability of Type II error (β) and hence, maximize the power function of test.

Steps in solving testing of hypothesis problem:

- i) Explicit knowledge of the nature of the population distribution and the parameters of interest, i.e., about which the hypothesis is set up.
- ii) Setting up of null hypothesis H_0 and alternative hypothesis H_1 in terms of the range of parameter values, each one embodies.
- iii) The choice of a suitable test statistic, say, $t = t(x_1, \dots, x_n)$ called the test statistic which will best reflect on H_0 and H_1 .
- iv) Partition the simple space (set of possible values of test statistic, t) into two disjoint and complementary subsets C and $\bar{C} = A$ (say) and framing be test, such as
 - (a) Reject H_0 if value $t \in C$.
 - (b) Accept H_0 if value $t \in A$.

After framing the above test, obtain the experimental sample observations, compute the test statistic and take action accordingly.

Example 3: let 'p' be the probability that coin will fall as head is a single toss in order to test; $H_0 : p = 1/2$ against $H_1 : p = 3/4$. The coin is tossed 5 times and H_0 is rejected if more than three heads are obtained. Find the probability of Type I and Type II errors and also find the power of test.

Solution: Given that $H_0 : p = 1/2$ versus $H_1 : p = 3/4$.

The probability of heads in n tosses is given by

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x};$$

where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$; $n! = n(n-1)(n-2)\dots 1$ and $0! = 1$.

For $n = 5$, the probability is given by

Testing of Hypothesis

$$P[X = x] = \binom{5}{x} p^x (1-p)^{5-x}.$$

We reject H_0 , if number of heads are more than three, i.e., four or more than four. The sample space of heads in 5 coin tosses is $S = \{0, 1, 2, 3, 4, 5\}$. Therefore, the critical region is given by: $C = \{x : x \geq 4\}$, i.e., $C = \{4, 5\}$ and the acceptance region A is given by: $A = \{x : x \leq 3\}$, i.e., $A = \{0, 1, 2, 3\}$.

Hence,

$$\begin{aligned} \text{Probability of Type I error } (\alpha) &= P(\text{reject } H_0 / H_0 \text{ is true}) \\ &= P(x \geq 4 / p = 1/2) \\ &= P(x = 4 / p = 1/2) + P(x = 5 / p = 1/2) \\ &= \binom{5}{4} (1/2)^4 (1/2)^1 + \binom{5}{5} (1/2)^5 (1/2)^0 \\ &= \binom{5}{4} (1/2)^5 + \binom{5}{5} (1/2)^5 \\ &= (1/2^5) \left\{ \binom{5}{4} + \binom{5}{5} \right\} \\ &= (1/2)^5 (5+1) \\ &= (1/32) \times 6 \\ &= 3/16 \end{aligned}$$

$$\begin{aligned} \text{Probability of Type II error } (\beta) &= P(\text{Accept } H_0 / H_1 \text{ is true}) \\ &= P(x \leq 3 / p = 3/4) \\ &= 1 - P(x \geq 4 / p = 3/4) \\ &= 1 - [P(x = 4 / p = 3/4) + P(x = 5 / p = 3/4)] \\ &= 1 - \left[\binom{5}{4} (3/4)^4 (1/4)^1 + \binom{5}{5} (3/4)^5 \right] \\ &= 1 - \{5(3/4)^4 (1/4) + 1(3/4)^5\} \\ &= 1 - \{(5 \times 81)/1024 + (1 \times 243)/1024\} \\ &= 1 - \{(405 + 243)/1024\} \\ &= 1 - (648/1024) \\ &= 1 - 81/128 \\ &= 47/128 \end{aligned}$$

$$\begin{aligned} \text{Power of the test} &= 1 - \beta \\ &= 1 - 47/128 \\ &= 81/128. \end{aligned}$$