1. Introduction

Welcome to the series of E-learning modules on Hypothesis. In this module we are going cover the basic concept of null and alternative hypothesis and also simple and composite hypothesis, the role of hypothesis in statistical tests and various types of hypotheses.

At the end of this session, you will be able to:
- Understand Statistical Hypothesis
- Understand the role and importance of hypothesis in statistical tests
- Describe various types of hypothesis
- Explain the characteristics of a good hypothesis

Very often in practice we are called upon to make decisions about populations on the basis of sample information. Such decisions are called statistical decisions.

For example we may wish to decide on the basis of sample data whether a new serum is really effective in curing a disease, whether one educational procedure better than another or whether there is increase in consumption of tea after excise duty of tea is reduced.

In practical situations statistical inference can involve either estimating a population parameter or making decisions about the value of the parameter. For example if a pharmaceutical company is fermenting a vat of antibiotic , samples from the vat can be used to estimate the mean potency mu for all of the antibiotic in the vat.

In contrast, suppose the company is not concerned about the exact mean potency of the antibiotic but is concerned only that it meet the minimum government policy standards, then, the company can use the samples from the vat to decide between these two possibilities
- The mean potency mu does not exceed the minimum allowable potency
- The mean potency mu exceeds the minimum allowable potency

The pharmaceutical company's problem illustrates a statistical test hypothesis

The reasoning used in statistical test hypothesis is similar to the process in a court trial.
In trying a person for theft, the court must decide between innocence and guilt. As the trial begins the accused person is assumed to be innocent.
The prosecution collects and presents all possible evidences in an attempt to contradict the innocent hypothesis and hence obtain a conviction.

If there is enough evidence against the defendant, the court will reject the innocence hypothesis and declare that the defendant guilty.
If the prosecution does not present enough evidence to prove the defendant guilty, the court will find him not guilty.

Notice that this does not prove that the defendant is innocent but merely that there was not enough evidence to conclude that the defendant was guilty.
We use this same type of reasoning to explain the basic concepts of hypothesis testing.

2. Simple and Composite Hypothesis

Hypothesis:
A hypothesis is a theoretical proposition capable of empherical verification. It may be viewed as a statement of an event, which may or may not be true.

A distinction is often made between maintained and testable hypothesis.
Assumptions that we normally use in theory as a simplifying device are not testable empirically. They are called maintained hypothesis.
For example: when we formulate the demand theory we assume that the taste and preferences of the consumer remains constant. We will not test this hypothesis in general. It is used in theory merely as a simplifying device.

The testable theoretical hypothesis on the other hand states that there is “no difference” between the sample statistics and population parameter.

Statistical Hypothesis:
In attempting to reach decisions it is useful to make assumptions or guesses about the population involved. Such assumptions which may or may not be true are called statistical hypothesis. They are generally statements about the probability distributions of the populations. By statistical hypothesis we mean a statement about a population from which samples have been drawn.

The assumption may be about the form of the population or about the parameters of the population. A statistical hypothesis may be defined as a tentative conclusion logically drawn concerning any parameter or form of the distribution of the population. It is a systematic and formal statement usually in the form of a proposition.

For example:

- A statement that for a Normal population mu equal to mu naught and sigma is equal to sigma naught
- For a Normal population with an unknown mean and variance sigma square is equal to twelve
- For a binomial variable ‘p’ is equal to half

Simple Hypothesis
A hypothesis may be simple or composite. If a hypothesis is concerning the population completely (such as functional form and parameter) and completely specifies a distribution, it is called a simple hypothesis and otherwise a composite hypothesis. For example a statement that the population is Normal with mean equal to twenty five and a standard deviation equal to ten, is a simple hypothesis, since the population is completely specified.

Examples:

- A statement that for a Normal population mu equal to mu naught and sigma equal to sigma naught and

- If X is a Poisson population, with parameter lambda then the hypothesis H:lambda equal to two

Composite hypothesis: if the hypothesis is not simple it is composite. That is, if the hypothesis does not include the complete population, it is called a composite hypothesis.
For example: The statement that the population follows Normal distribution with mean equal to twenty five is a composite hypothesis, since only one parameter of the population is specified and the other parameter namely standard deviation is not specified.

- For a Normal population with an unknown mean and variance sigma square is equal to twelve
- For a binomial variable p equal to half
- If X is a Poisson population with parameter lambda then the hypothesis that lambda lies between seven and eight are composite hypotheses

While testing whether a new method of cultivation is more efficient than the old method, the statement, the hypothesis that mu one is equal to mu two is equal to fifty is a simple hypothesis.
Whereas a statement, the hypothesis that mu one is not equal to mu two is a composite hypothesis where mu one is the average yield of crop when the old method is used and mu two is the average yield of crop when the new method is used.

If in a composite hypothesis ‘k’ parameters, of a population under consideration is not specified then k is said to be the degrees of freedom of the composite hypothesis.

3. Parametric, Non Parametric and Null Hypothesis

Parametric hypothesis:
A hypothesis which specifies only the parameters of the probability density function of the population is called a parametric hypothesis.
For example: the hypothesis that the population follows Normal distribution with mean equal to twenty five, that is, we want to test whether the mean is equal to twenty five or not.

A hypothesis which specifies only the form of the density function in the population is called the non parametric hypothesis.
For example: Hypothesis that the population is Normal is a nonparametric hypothesis (that is we want to test whether the population is Normal).

Setting up Hypothesis:
In the first step of statistical tests, hypothesis about a population parameter is set. Information from the sample data is used to decide how likely is to accept or to reject the set hypothesis.
Conventionally two hypotheses are set in such a way that if one is accepted the other automatically gets rejected. The two hypotheses that are normally used are null hypothesis and alternative hypothesis.

Null Hypotheses:
In many instances, we formulate the statistical hypothesis for the sole purpose of rejecting or nullifying it.
In any test there are two hypothesis and they are so constructed that if one is accepted the other is rejected.
A statistical hypothesis which is to be tested is called a null hypothesis.

It is a neutral hypothesis or an unbiased hypothesis and is denoted by H naught. For example if we want to test whether the given coin is loaded, we formulate the hypothesis that the coin is fair (that is p equal to zero point five, where p is the probability of heads).

Similarly, if we want to decide whether one procedure is better than the other, we formulate the hypothesis that the there is no difference between the procedures.
That is our observed differences are merely due to fluctuations in sampling from the same population.

The null hypothesis asserts that there is no significant difference between population parameter and sample statistic and the difference if any is due to the sampling fluctuations.
Such hypothesis are often called null hypothesis and are denoted by H naught.

The null hypothesis constitutes a challenge and the function of the experiment is to give the facts a chance to refute or fail to refute this challenge.

For example, if we want to find out whether the new vaccine has benefited the people or not,

the null hypothesis shall be setting up saying that “the new vaccine has not benefitted the people”.
The rejection of the null hypothesis indicates that the differences have statistical significance and the acceptance of the null hypothesis indicates that the differences are due to chance.

4. Alternative Hypothesis

Alternative Hypothesis:
The alternative hypothesis specifies those values that the researcher believes to hold true and hopes that the sample data would lead to acceptance of this hypothesis to be true.
The alternate hypothesis may embrace the whole range of values rather than a single point.
As per the definition it is very difficult to identify the null and alternative hypothesis. However for statistical convenience, the following definitions are used.

A statistical hypothesis which is complementary to the null hypothesis is called an alternative hypothesis.
Any hypothesis that differs from a given hypothesis is called an alternate hypothesis. Hence in testing null hypothesis an alternative hypothesis is also made so that if the null hypothesis is to be rejected based on the samples and tests there is an alternative hypothesis to be accepted.

For example if one hypothesis is p is equal to zero point five, alternate hypothesis might be p is equal to zero point seven or p is not equal to zero point five or p is greater than zero point five.
A hypothesis alternate to the null hypothesis is denoted by H one.

Hence when the null hypothesis is that mu equal to mu not is rejected, we are forced to accept the alternative hypothesis that mu is not equal to mu not.
Both null and alternative hypothesis may be simple or both composite or one simple and the other composite.

For example:

- The null hypothesis H naught, that lambda is equal to three against the alternative hypothesis that lambda is equal to four, here both are simple
- The null hypothesis that lambda is equal to three against the alternative hypothesis that lambda is greater than three. Here the Null hypothesis is simple but the alternative hypothesis is composite
- The null hypothesis that lambda is less than or equal to three against the alternative hypothesis that lambda is greater than three. Here both are Composite

In a Normal population with known sigma square we have a null hypothesis H naught as: mu is equal to mu naught.
The alternative hypotheses here could be mu is not equal to mu naught or mu is greater than mu naught or mu is less than mu naught.

The two competing hypotheses are the null and alternate hypothesis. . Generally the hypothesis that the researcher wishes to support is the H one and the null hypothesis is a contradiction of the alternate hypothesis.
As we will soon see it is easier to show support for the H one by proving that H naught is false. Hence the statistical researcher always begins by assuming that the null hypothesis is true.

The researcher then uses the sample data to decide whether the evidence favours H one rather than H naught and draws one of the two conclusions.
Reject H naught and conclude that H one is true. Accept H naught (do not reject) as true.

Example one: You wish to show that average hourly wage of carpenters in the state of California is different from fourteen dollars, which is the national average.
This is the alternative hypothesis written as H one, colon, mu not equal to fourteen.
The null hypothesis is H naught, colon, mu is equal to fourteen. You would like to reject the H naught, thus concluding that the California mean is not equal to fourteen dollars.

A milling process currently produces an average of three percent defectives. You are interested in showing that the sample adjustment on a machine will decrease p, the proportion of defectives produced in the milling process. Thus the alternative hypothesis is ‘p’ is less than zero point zero three and the null hypothesis is ‘p’ is equal to zero point zero three.
If you can reject H naught you can conclude that the adjusted process produces fewer than three percent defectives.

The purpose of the study is to assess the effect of the lactation nurse on attitudes towards breast feeding among women.
Research question: Does the lactation nurse have an effect on attitudes towards breast feeding?
The null hypothesis is that, the lactation nurse has no effect on attitudes towards breast feeding.
The alternative hypothesis is that, the lactation nurse has an effect on attitudes towards breast feeding.

5. Importance of Null and Alternative Hypothesis

Importance of Null and Alternative hypothesis:
The following examples illustrate the importance of null and alternative hypothesis. Suppose a coin is tossed hundred times and fifty two heads were observed. It would not be correct to jump the conclusion that the said coin is biased to yield more number of heads than the tails. Because the experiment yielded two heads more than the expected fifty heads. In fact fifty two heads is consistent with the hypothesis that the coin is unbiased.

Thus it would not be surprising to flip the coin hundred times and observe fifty two heads. On the other hand flipping eighty or ninety heads in hundred flips would seem to be contradicting the hypothesis that the coin is unbiased. In this case the coin is a biased one.

Example 2: An advertising department of a leading farm newspaper believes that the farmer who subscribes to this newspaper earns higher average income than the state average income of all farmers including the subscribers.

In support of this claim a manager of an advertising department collected a sample of three thousand six hundred subscribers from its mailing list and estimated the average income of four thousand two ninety rupees.
From the government sources, the average income of all the farmers in the state was obtained as four thousand one sixty two rupees. Since the difference between the two observations is one twenty eight rupees, the newspapers claim that the subscribers have more income than the state average because of their accessibility to the farm newsletters.

Now the real problem is to assess whether the difference of this magnitude is too small to be ignored or too big to be taken care off.
If we prove that the difference is too small and hence should be ignored then the newspaper's claim too is to be ignored.

On the other hand if there is strong evidence to state that the observed difference is too big, then we accept the farm paper's claim that its subscriber's income is more than the state average. The problem here is enough to illustrate the procedure for setting the hypothesis.

By inductive reasoning, we try to prove a more general result by using the sample.
To begin with, we do not magnify the difference; instead we say that the difference between the sample mean and the population mean is insignificant.

So in our illustrative case we start by saying that the difference of one hundred and twenty eight rupees is very small and can be ignored.
Since the difference between the sample value and the corresponding population value is hypothesized as zero, the so framed hypothesis is often called null hypothesis.

The null hypothesis is that, mu is equal to four thousand one hundred and sixty two. That is,

there is no change in the average income of the farming community in the state.
On the other hand, if there is enough evidence to tell that the difference of one hundred and twenty eight rupees is very high then we simply reject the null hypothesis. By simply rejecting the null hypothesis we move to an alternative called alternative hypothesis.

However an alternative hypothesis here can take three distinct forms. It may be either “not equal to” or “greater than” that is, as the farm paper claims in our illustration or “less than” type.
The hypothesis in our illustrative example , may be expressed as a one tailed test in the following form:
H naught or the null hypothesis is that, mu is equal to four thousand one sixty two against the alternative hypothesis that is, H one is, mu is greater than four thousand one sixty two.

Characteristics of a good hypothesis:
A good hypothesis must be based on a good research question.
- Hypothesis should be simple
- Hypothesis should be specific
- Hypothesis should be stated in advance

Here’s a summary of our learning in this session where we have:
- Understood the basic concept of statistical hypothesis
- Understood the role and importance of hypothesis in statistical tests
- Explained the types of hypotheses that is Simple and Composite, Null and Alternative
- Understood the characteristics of a good hypothesis