1. Introduction

Welcome to the series of E-learning modules on level of significance or size of the test. In this module we are going cover the chances for occurrence of type one error, level of significance, its role in testing of hypothesis and illustrative examples.

By the end of this session, you will be able to:

- Understand the Level of Significance or Size of the test
- Explain its importance
- Understand the chance of occurrence Type one error
- Demonstrate the procedure to obtain size of the test

To test the statistical hypothesis we must decide whether that hypothesis appears to be consistent with the data of the sample. For instance, a tobacco firm claims that it has discovered a new way of curing tobacco leaves that will result in a mean nicotine content of a cigarette of one point five milligrams or less. A null hypothesis is formed which states that the mean nicotine content of a cigarette is less than or equal to one point five milligrams.

The null hypothesis is rejected if it appears to be inconsistent with the sample data. Of course the decision of whether to reject the null hypothesis is based on the value of a test statistic.
In an experiment where we retain or reject the null hypothesis, our goal is to maximize the chances of making a correct decision and minimize the chances of making an incorrect decision.

Thus, what we see here is the possibility of 4 different consequences of our decisions:
Two good and correct choices and also two bad and incorrect choices.
You will make a correct decision by:

- Retaining the null hypothesis when the null hypothesis is true, or
- Rejecting the null hypothesis when the null hypothesis is false

Your decision will be incorrect if you:

- Retain the null hypothesis when the null hypothesis is false
- Reject the null hypothesis when the null hypothesis is true

The result of the test may be negative or positive, relative to the null hypothesis.
If the result of the test corresponds with real situation, then a correct decision has been made. However, if the result of the test does not correspond with reality, then an error has occurred.

Due to the statistical nature of a test, the result is never, except in very rare cases, free of error. Two types of error are distinguished: type one error and type two error.
Type one and Type two errors are also called errors of the first **and** second kind respectively.

A type one error or error of first kind is when you reject the null hypothesis when the null hypothesis is true.
A type two error or error of second kind is when you retain the null hypothesis when the null

hypothesis is not true.
In many practical applications type one errors are more delicate than type two errors. In these cases, care is usually focused on minimizing the occurrence of this statistical error.

Generally speaking, researchers are not willing to reject the null hypothesis if the chance of rejecting it wrongly is greater than about five percent.
We usually want to reject the null hypothesis but we only want to do this if the chance that we do it incorrectly is rather low.

Unfortunately, setting some criterion about how much Type one error rate we are willing to tolerate does little to regulate or establish what level of Type two errors we are willing to tolerate. The literature seems to be fixated only on the rate of Type one errors without much attention to Type two error.

Now it must be understood that the objective of statistical test of null hypothesis is not to determine whether the null hypothesis is true but rather to determine if its truth is consistent with the resultant data.
Therefore given this objective it is reasonable that a null hypothesis should be rejected only if the sample data are very unlikely when the null hypothesis is true.

The classical way of accomplishing this is to specify a small value of alpha and then require that the test have the property that whenever the null hypothesis is true its probability of being rejected is less than or equal to alpha.
The value of alpha is called the level of significance of the test and is usually set in advance with commonly chosen values being alpha is equal to zero point one zero, zero point zero five and zero point zero one.

2. Size or Significance Level of a Test or Alpha

Size or Significance level of a test or alpha
For simple hypotheses, this is the test's probability of **incorrectly** rejecting the null hypothesis or **the false positive** rate.
For composite hypotheses this is the upper bound of the probability of rejecting the null hypothesis over all cases covered by the null hypothesis.

One interpretation of the significance level alpha, based in **decision theory**, is that alpha corresponds to the value for which one chooses to reject or accept the null hypothesis H naught. In decision theory, this is known as a **Type one error**. The probability of a Type one error is equal to the significance level alpha.

Hence the probability with which we may reject the null hypothesis when it is true is called a **level of significance** and it is also called as a **size of the test**.
Deciding over suitable significance level is the third step in testing of hypothesis. In this step the validity of null hypothesis against that of the alternate hypothesis is decided, to be ascertained at certain level of significance.

The confidence with which the experimenter rejects or accepts a null hypothesis depends upon the significance level adopted. This significance level is expressed as a percentage.
For example: five percent is the probability of rejecting the null hypothesis, if it is true.
When the hypothesis in question is accepted at five percent level, the statistician is running a risk that in the long run he may be making a wrong decision about five percent of the time.

Similarly by rejecting the null hypothesis at the same level he is running the risk of rejecting the true hypothesis five times out of every one hundred occasions. When the level of significance is zero point zero five. It implies in the long run a statistician is rejecting the null hypothesis five times out of every one hundred times.

Suppose, the probability for a type one error is one percent, then there is a one percent chance that the observed variation is not true. This is called the level of significance, denoted with the Greek letter alpha.
While one percent or five percent might be an acceptable level of significance for one application, a different application can require a very different level.

Computation of the size of the test
Size of the test is equal to probability of type one error. This is equal to probability of [H naught rejected given that H naught is true]
Which is equal to probability of ‘x’ belongs to C provided H naught is true which is equal to integral of L naught of (x one, x two, etc, till x n) d x one, d x two, etc, till d x n which is equal to alpha.

Where L naught of (x one, x two, etc, till x n) is the likelihood function under the null hypothesis. And C is the rejection region of the null hypothesis. In case of simple null hypothesis the probability of first kind of error is called the size of the test or level of significance.

3. Operating Characteristic

Operating Characteristic of a test (O.C)
If alpha is the size of the test or probability of Type one error then (1 minus alpha) is called a Operating Characteristic of a test
That is we have alpha is equal to probability of x belongs to C provided H naught is true. This is equal to one minus probability of x belongs to A provided H naught is true.

This implies, one minus alpha is equal to Probability of x belongs to A provided H naught is true which implies one minus alpha is equal to Probability of Accept H naught when H naught is true.
Where A is the acceptance region of the null hypothesis
Operating Characteristic can be defined as a probability of accepting H naught when it is true.

Note that the complement of the false positive rate, (one minus alpha), is termed **specificity** in biostatistics. ("This is a specific test. Because the result is positive, we can confidently say that the patient has the condition.").
The probability of rejecting the null hypothesis when it is in fact false (that is, a correct decision) is equal to (one minus alpha). To minimize the probability of Type one error, the significance level is generally chosen to be small.

While we are close to being able to take decision either to accept or reject the null hypothesis and so either we reject or accept the alternative hypothesis, rigour requires that we first set the significance level for our study.
Statistical significance is about probability. It asks the question: What is the probability that a score could have arisen by chance?

Let us consider two distributions, a seminar only distribution and a lectures only distribution.
Statistically analysing the differences in the distribution may, for example, suggest that there are no differences in the two distributions; hence we should accept the null hypothesis. However, how confident are we that there really exists no difference between the distributions?

Typically, if there was a five percent or less chance (that is, five times in one hundred or less) that a score from the focal distribution (that is, the “seminar” distribution”) could not have come from the comparison distribution (the “lectures only” distribution); we would **accept** the null hypothesis.

Alternately, if the chance was greater than five percent (that is six times in one hundred or more) we would reject the null hypothesis.
We do not reject the null hypothesis because our statistical analysis did not show that the two distributions were the same.
We reject it because, at a significance level of zero point zero five (that is, five percent or less chance) we could not be confident enough that this result did not simply happen by chance.

While there is relatively little justification why a significance level of zero point zero five is used rather than zero point zero four or zero point one zero, for example, it is widely accepted

in academic research.
However, if we want to be particularly confident in our results, we set a more stringent level of zero point zero one (that is a one percent chance or less or one in one hundred chance or less).

4. Pros and Cons of Significance Level

Pros and Cons of setting a significance level:
Setting a significance level (that is, before doing inference) has the advantage that the analyst is not tempted to choose a cut-off on the basis of what he or she hopes is true.

It has the disadvantage that it neglects that some significant values might best be considered borderline. This is one reason why it is important to report significant values when reporting results of hypothesis tests.

It is also good practice to include confidence intervals corresponding to the hypothesis test.
For example, if a hypothesis test for the difference of two means is performed, also give a confidence interval for the difference of those means. If the significance level for the hypothesis test is zero point zero five, then use confidence level ninety five percent for the confidence interval.

Deciding what significance level to use:

This should be done before analyzing the data, preferably before gathering the data. The choice of significance level should be based on the consequences of Type one and Type two errors.
If the consequences of a type one error are serious or expensive, then a very small significance level is appropriate.

Example one:
Two drugs are being compared for effectiveness in treating the same condition.
Drug one is very affordable, but Drug two is extremely expensive.
The null hypothesis is "both drugs are equally effective," and the alternate is "Drug two is more effective than Drug one."
In this situation, a Type one error would be deciding that Drug two is more effective, when in fact it is no better than Drug one, but would cost the patient much more money.

That would be undesirable from the patient's perspective, so a small significance level is warranted.
If the consequences of a Type one error are not very serious (and especially if a Type two error has serious consequences), then a larger significance level is appropriate.

5. Illustrations and Application

Example two:
Two drugs are known to be equally effective for a certain condition. They are also equally affordable.
However, there is some suspicion that Drug two causes a serious side effect in some patients, whereas Drug one has been used for decades with no reports of the side effect.

The null hypothesis is "the incidence of the side effect in both drugs is the same", and the alternate is "the incidence of the side effect in Drug two is greater than that in Drug one."

Falsely rejecting the null hypothesis when it is in fact true (that is, a Type one error) would have no great consequences for the consumer.
A Type two error that is , failing to reject the null hypothesis when in fact the alternate is true, would result in deciding that Drug two is no more harmful than Drug one when it is in fact more harmful, could have serious consequences from a public health standpoint.
So setting a large significance level is appropriate.

Minimizing errors of decision is not a simple issue.
For any given sample size, the effort to reduce one type of error generally results in increasing the other type of error. The only way to minimize both types of error, without just improving the test, is to increase the sample size, and this may not be feasible.

An investigator's conclusion may at times be wrong.
Sometimes, by chance alone, a sample is not representative of the population. Thus the results in the sample do not reflect reality in the population, and the random error leads to an erroneous inference.

Although type one and type two errors can never be avoided entirely, the investigator can reduce their likelihood by increasing the sample size.
The larger the sample, the lesser is the likelihood that it will differ substantially from the population.

The classical procedure for testing a null hypothesis is to fix a small significance level or alpha, and then requiring that the probability of rejecting the null hypothesis when H naught is true is less than or equal to alpha.
Because of asymmetry in the test regarding the null and alternative hypothesis, it follows that the only time in which an hypothesis can be regarded as having been proved by the data is when the null hypothesis is rejected (thus 'proving' that the alternative is true).

Application:
A coin is thrown eight times. Null Hypothesis H naught is p equal to one by two and H one is p equal to two by three. Test procedure is, a null hypothesis is rejected if six or more tosses give heads. P is the probability for getting head in each trial. Then determine the level of significance.

Solution:

Tossing of a coin, getting head or tail follows Binomial distribution.
Hence suppose x is the number of heads , its probability mass function is given by, F of x is equal to (n C x) into p to the power x into q to the power (n minus x)

Rejection region is six or more heads.
f of x given H naught is equal to f of x given p is equal to half which is equal to eight C x into half to the power x (into half to the power) (eight minus x).
f of x given H one is equal to f of x given p is equal to two by three which is equal to eight C x into (two by three) to the power x into (one by three) to the power (eight minus x)

Size of the test or Level of significance is equal to Probability of Type one error which is equal to Probability of rejecting H naught when $_H$ naught is true.
This is equal to Probability of getting six or more heads given p is equal to one by two.
Which is equal to Probability of x greater than or equal to six given p is equal to half which is equal to ,eight C six into (half to the power six) into (half to the power) (eight minus six) plus eight C seven into (half to the power seven) into half to the power (eight minus seven) plus eight C eight into (half to the power eight) into half to the power (eight minus eight)
This is equal to half to the power eight into [eight C six plus eight C seven plus eight C eight] which equals to thirty seven by two fifty six which is equal to zero point one four four five.
Hence the level of significance is equal to zero point one four four five.

Here's a summary of our learning in this session where we have:

- Understood the concept of probability of Type one error
- Understood level of significance or size of the test
- Explained the Operating Characteristic function
- Understood the pros and cons of significance level
- Described the application of significance level