1. Introduction

Welcome to the series of E-learning modules on power of the test. In this module we are going cover the chances for occurrence of type two error and power of the test. Its interpretation and role in testing of hypothesis and illustrative examples.

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

- Understand Power of the test and Power function
- Interpret the importance of Power of the test
- Understand the probability of Type two error
- Explain the procedure to obtain Power of the test

In the previous module, discussed about two types of errors that we make in the process of deciding either to accept or reject the null hypothesis.
Recall that Type one error occurs when we reject the null hypothesis when it is actually true.
Type two error occurs when we accept the null hypothesis when it is actually false.

A Type two error occurs when we don't reject a false null hypothesis that is, accept the null hypothesis. This occurs when a guilty defendant is acquitted.
In practice, this type of error is by far the most serious mistake we normally make.

For example, if we test the hypothesis that the amount of medication in a heart pill is equal to a value which will cure your heart problem and accept the hull hypothesis that "the amount is ok".
Later on we find out that the average amount is way too large and people die from "too much medication".
This could be a grave mistake to make as it could be too late and the pills would have been shipped the public.

The probability with which we may reject the null hypothesis when it is true is called level of significance and is denoted by alpha.
The probability with which we may accept the null hypothesis when it is false is denoted by beta.

Our object is to determine a test procedure for which both the errors are minimised.
However, we find that as one error decreases the other increases. Therefore we fix one of the errors at a low permissible level and try to minimize the other.
Usually the probability of type one error is fixed at a certain level that is, alpha is fixed at zero point zero one, zero point zero two, zero point zero five or zero point one depending on the situation or requirements .

2. Power of the Test

When we fix the level of significance at alpha, the problem is just to minimize beta, the probability of type two error.
This is equivalent to maximize one minus beta. The quantity one minus beta is called power of the test and is also called as power of a rejection region.

The power of a statistical test is the probability that the test will reject the null hypothesis when the null hypothesis is false (that is the probability of not committing a type two error.
The power is in general a function of the possible distributions, often determined by a parameter, under the alternative hypothesis.
As the power increases, the chances of a type two error occurring decrease.

The probability of making a type one error or alpha, can be decreased by altering the level of significance.
It will be more difficult to find a significant result if the power of the test will be decreased because the risk of Type two error will be increased.
The probability of making a Type two error or beta can be decreased by increasing the level of significance which will increase the chance of a Type one error.

Rejecting a null hypothesis when it is false is what every good hypothesis test should do.
Having a high value for one minus beta (that is near one) means it is a good test, and having a low value (near zero) means it is a bad test.

Hence, one minus beta is a measure of how good a test is, and it is known as the "power of the test."
Hence the power of the test is the probability that the test will reject H naught when in fact it is false.
Conventionally, a test with a power of zero point eight is considered good.

Computation of the Power of the test
Beta is equal to Probability of Type two error
Which is equal to Probability of [H naught is accepted given that H one is true]
Which is equal to Probability of x belongs to A provided H one is true which is equal to integral of L one of (x one, x two, etc till x n) d x one, d x two, etc, d x n which is equal to beta
Where L one of (x one, x two, etc, x n) is the likelihood function under the alternative hypothesis. And A is the acceptance region of the null hypothesis.

If beta is the probability of Type two error, then (one minus beta) is called Power of a test.
That is, we have Probability of x belongs to A provided H one is true which is equal to one minus Probability of x belongs to C provided H one is true.
This implies one minus beta is equal to Probability of x belongs to C provided H one is true.
Which implies one minus beta is equal to Probability of, reject H naught when H one is true.
Where C is the rejection region of the null hypothesis.
Power of a test can be defined as a probability of rejecting H naught when it is false.

Power function:

With each test for testing a null hypothesis against an alternative hypothesis a function is associated called the power function of a test.

The power function of a test is that function which gives the probability of rejecting the null hypothesis.

The Power function of a test T is denoted by PT, that is, P subscript T and power at point theta is denoted by P T of theta.

PT of theta is equal to probability of reject H naught given theta.

If H naught and H one are the functions of the parameters theta and if power of the test, PT of theta is equal to one minus beta is regarded as a function of theta then it is known as a power function.

3. Factors on Which a Power of a Test is Dependent Upon

Statistical power may depend on a number of factors. Some of these factors may be particular to a specific testing situation, but at a minimum, power nearly always depends on the following three factors:

- The statistical significance criterion used in the test
- The magnitude of the effect of interest in the population
- The sample size used to detect the effect

One easy way to increase the power of a test is to carry out a less conservative test by using a larger significance criterion.
This increases the chance of rejecting the null hypothesis (that is obtaining a statistically significant result) when the null hypothesis is false, that is, reduces the risk of a Type two error.

However it also increases the risk of obtaining a statistically significant result (that is rejecting the null hypothesis) when the null hypothesis is not false; that is, it increases the risk of a type one error.

The magnitude of the effect of interest in the population can be quantified in terms of an effect size, where there is greater power to detect larger effects.
An effect size can be a direct estimate of the quantity of interest, or it can be a standardized measure that also accounts for the variability in the population.

The sample size determines the amount of sampling error inherent in a test result.
Other things being equal, effects are harder to detect in smaller samples. Increasing sample size is often the easiest way to boost the statistical power of a test.

The precision with which the data are measured also influences statistical power. Consequently, power can often be improved by reducing the measurement error in the data.
The design of an experiment or observational study often influences the power.

For example, in a two-sample testing situation with a given total sample size n, it is optimal to have equal numbers of observations from the two populations being compared (as long as the variances in the two populations are the same).

4. Interpretation

Although there are no formal standards for power (sometimes referred as PT or pi), most researchers assess the power of their tests using PT is equal to pi is equal to zero point eight zero as a standard for adequacy.
This convention implies a (four to one) trade off between beta risk and alpha risk.
Beta is the probability of a Type two error; alpha is the probability of a Type one error, zero point two and zero point zero five are conventional values for beta and alpha.

Power analysis is appropriate when the concern is with the correct rejection, or not, of a null hypothesis.
In many contexts, the issue is less about determining if there is or is not a difference but rather with getting a more refined estimate of the population effect size.

For example, if we were expecting a population correlation between intelligence and job performance of around point five zero, a sample size of twenty will give us approximately eighty percent power to reject the null hypothesis of zero correlation (that is, alpha equal to zero point zero five)

However, in doing this study we are probably more interested in knowing whether the correlation is zero point three zero or zero point six zero or zero point five zero.
In this context we would need a much larger sample size in order to reduce the confidence interval of our estimate to a range that is acceptable for our purposes.

Techniques similar to those employed in a traditional power analysis can be used to determine the sample size required for the width of a confidence interval to be less than a given value.

Many statistical analyses involve the estimation of several unknown quantities. In simple cases, all but one of these quantities is a nuisance parameter.
In this setting, the only relevant power pertains to the single quantity that will undergo formal statistical inference.
In some settings, particularly if the goals are more "exploratory", there may be a number of quantities of interest in the analysis.

For example, in a multiple regression analysis we may include several covariates of potential interest. In situations such as this where several hypotheses are under consideration, it is common that the powers associated with the different hypotheses differ.

For instance, in multiple regression analysis, the power for detecting an effect of a given size is related to the variance of the covariate.
Since different covariates will have different variances, their powers will differ as well.

Any statistical analysis involving multiple hypotheses is subject to inflation of the type one error rate if appropriate measures are not taken.
Such measures, typically involve applying a higher threshold of stringency to reject a hypothesis, in order to compensate for the multiple comparisons being made.

In this situation, the power analysis should reflect the multiple testing approach to be used.

Factors that affect the power of a test:
The power of a hypothesis test is affected by three factors.

a. Sample size (n). Other things being equal, the greater the sample size, the greater the power of the test.

b. Significance level (alpha). The higher the significance level, the higher the power of the test.
If you increase the significance level, you reduce the region of acceptance.
As a result, you are more likely to reject the null hypothesis. This means you are less likely to accept the null hypothesis when it is false. That is, less likely to make a Type two error.
Hence, the power of the test is increased.

c. The "true" value of the parameter being tested.
The greater the difference between the "true" value of a parameter and the value specified in the null hypothesis, the greater the power of the test.
That is, the greater the effect size, the greater the power of the test.

5. Uses and Applications

Uses of Power:
Power analysis can be used to calculate the minimum sample size required so that one can be reasonably likely to detect an effect of a given size.
Power analysis can also be used to calculate the minimum effect size that is likely to be detected in a study using a given sample size.

In addition, the concept of power is used to make comparisons between different statistical testing procedures: for example, between a parametric and a nonparametric test of the same hypothesis.

The goodness of a statistical test is measured by the size of the two error rates, alpha and beta.
A good test is the one for which both of these error rates are small.
The experimenter begins by selecting alpha, the probability of Type one error. If she or he also decides to control the value of beta, then an appropriate sample size is chosen.

Another way of evaluating a test is to look at the complement of a type two error.
That is, rejecting null hypothesis when the alternate is true which has the probability one minus beta which is equal to probability of reject H naught when H one is true.
The quantity (one minus beta) is called the power of the test because it measures the probability of taking the right action that we always wish to take.
That is rejecting the null hypothesis when it is false and alternative is true.

The power of the statistical test given as (one minus beta) is equal to probability of reject null hypothesis when alternate hypothesis is true, measures the ability of the test to perform as required.

A graph of (one minus beta), the probability of rejecting null hypothesis when in fact it is false as a function of the true value of the parameter of interest is called the power curve of the statistical test.
Ideally, one would like alpha to be small and the power (one minus beta) to be large.

The experimenter must decide on the values of alpha and beta measuring the risks of the possible errors he or she can tolerate.
He or she must also decide how much power is needed to detect the differences that are practically important in the experiment.
Once these decisions are made the sample size can be chosen by consulting the power curves corresponding to various sample sizes for the chosen test.

Application:
A coin is thrown eight times. Null Hypothesis H naught is P equal to one by two and H one is, P is 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. Determine the Power of the test.

Solution:

Tossing of a coin, getting head or tail follows Binomial distribution. If 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. This 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. This is equal to eight C x into (two by three) to the power x into (one by three) to the power (eight minus x).
Power of the test is equal to one minus probability of type two error which is equal to one minus probability of accepting H naught when H one is true.

This is equal to probability of rejecting H naught when H one is true.
Which is equal to probability of [getting six or more heads given p is equal to two by three]
Which is equal to Probability of x greater than or equal to six given p is equal to two by three.
This is equal to eight C six into (two by three) to the power six, into (one by three) to the power (eight minus six), plus eight C seven into (two by three) to the power seven into (one by three) to the power (eight minus seven) plus eight C eight into (two by three) to the power eight, into (one by three) to the power (eight minus eight).
This is equal to (two by three) to the power six, into [eight C six into (one by three) square, plus eight C seven into (two by three) into (one by three) plus eight C eight into (two by three) square into (one by three) to the power zero]. This is equal to three thousand and seventy two by six thousand five hundred and sixty one which is equal to zero point four six eight two.
Hence the power of the test is equal to zero point four six eight two.

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

- Understood the concepts of Power of the test and Power function
- Described the importance of Power of the test
- Understood the probability of Type two error
- Explained the procedure to obtain Power of the test