

Testing Hypothesis

The objective of statistical inference is to make inference about unknown parameters based on information contained in sample data. It can be achieved in two ways: Estimation and Testing Hypothesis.

Example 1: Non-statistical Hypothesis Testing

A criminal trial is an example of hypothesis testing without statistics.

In a trial a jury must decide between two hypotheses.

The null hypothesis is H_0 : The defendant is innocent

The alternative hypothesis or research hypothesis is H_1 : The defendant is guilty

The jury does not know which hypothesis is true. They must make a decision on the basis of evidence presented.

In the language of statistics convicting(declare to be guilty) the defendant is called *rejecting the null hypothesis in favor of the alternative hypothesis*. That is, the jury is saying that there is enough evidence to conclude that the defendant is guilty (i.e., there is enough evidence to support the alternative hypothesis).

If the jury acquits (free (someone) from a criminal charge) it is stating that there is not enough evidence to support the alternative hypothesis. Notice that the jury is not saying that the defendant is innocent, only that there is not enough evidence to support the alternative hypothesis. That is why we never say that we accept the null hypothesis, although most people in industry will say “We accept the null hypothesis”

Using sample data, when unknown parameter is estimated, there are many questions to be addressed: how we decide whether the sample disagrees with proposed hypothesis? When should we reject the hypothesis? Or when should we accept the hypothesis? What is the probability that we will make the wrong decision and consequently be led to a loss? And, particularly, what function of the sample measurements should be used to reach a decision? The answers to all these questions are contained in a study of testing hypothesis.

Hypothesis testing or testing hypothesis is a way of systematically quantifying how certain you are of the result of a statistical experiment. It is a process to check the validity of the assumption made about the population on the basis of the sample study. So, through this process, we compute the statistic on the sample drawn from the population. Then we compare the value of the statistic with the population parameter. This comparison is done through a suitable statistic usually known as **Test Statistic** say Z test defined as $Z = \frac{(T - \theta)}{\sigma_T}$, where T is the sample statistic, θ is the parameter of the population and σ_T is the standard error(deviation) of the sample statistic.

Example 2: Suppose that a manufacturer, of a certain type of products, observes that probability of failures is about $P = .05$. An engineer and a statistician working together suggest some changes that might improve the design of the product. To test this new procedure, it was agreed that $n = 200$ circuits would be produced using the proposed method and then checked. Let $X = \text{no. of failures}$.

If $(X/n \cong 0.05)$ then it seems that the new procedure has not resulted in an improvement. On the other hand, if $(X/n \cong 0.01 \text{ or } 0.02)$ we might believe that the new method is better than the old one.

What we need to establish is a formal rule that tells us when to accept the new procedure as an improvement. In addition, we must know the consequences of this rule. For example, we could accept the new procedure as an improvement if $(X/n < 0.025)$

Statistical hypothesis: A statistical hypothesis is an assumption about a population parameter. This assumption may or may not be true, Or, it is a verification of the character of the population on the basis of sample statistic.

The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Since that is often impractical, researchers typically examine a random sample from the population. If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected. The hypothesis can be simple or composite. $H_0 : \theta = \theta_0$ is a simple hypothesis. $H_0 : \theta < > \leq \geq = \theta_0$ is a composite hypothesis.

There are two types of statistical hypotheses.

- **Null hypothesis.** The null hypothesis, denoted by H_0 , is usually the hypothesis that sample observations result purely from chance, it is .
- **Alternative hypothesis.** The alternative hypothesis, denoted by H_1 or H_0 is the hypothesis that sample observations are influenced by some non-random cause.

For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in Heads and half, in Tails. The alternative hypothesis might be that the number of Heads and Tails would be very different. Symbolically, these hypotheses would be expressed as

$$H_0: P = 0.5$$

$$H_a: P \neq 0.5$$

Suppose we flipped the coin 50 times, resulting in 40 Heads and 10 Tails. Given this result, we need to reject the null hypothesis. We would conclude, based on the evidence, that the coin was probably not fair and balanced.

Can We Accept the Null Hypothesis?

Some researchers say that a hypothesis test can have one of two outcomes: you accept the null hypothesis or you reject the null hypothesis. Many statisticians, however, take issue with the notion of

"accepting the null hypothesis." Instead, they say: you reject the null hypothesis or you fail to reject the null hypothesis.

Why the distinction between "acceptance" and "failure to reject?" Acceptance implies that the null hypothesis is true. Failure to reject implies that the data are not sufficiently persuasive for us to prefer the alternative hypothesis over the null hypothesis.

The main goal of hypothesis testing is to tell us whether we have enough evidence to reject the null hypothesis.

Test Statistic (as an estimator) is a function of the sample measurements upon which the statistical decision will be depend.

Rejection Region or Critical Region

This region specifies the values of the test statistic for which the null hypothesis is to be rejected in favor of the alternative hypothesis.

For example, if a sample gives the computed value of the test statistic falls in the RR, we reject H_0 and accept H_1 .

Sampling errors : In testing hypothesis, we generally face with one of the following possibilities:

1. The null hypothesis is true but test rejects it.
2. The null hypothesis is false but test accepts it.
3. The null hypothesis is true and test accepts it.
4. The null hypothesis is false and the test reject it.

Decision Errors

Two types of errors can result from a hypothesis test.

- **Type I error.** A Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the significance level. This probability is also called alpha, and is often denoted by α .
- **Type II error.** A Type II error occurs when the researcher fails to reject a null hypothesis that is false. The probability of committing a Type II error is called Beta, and is often denoted by β . The probability of not committing a Type II error is called the **Power of the test**.

Significance level : Level of significance is the maximum probability for taking risk of type I error. Generally, it is expressed as percentage, say 5%, 1%, etc. as an explanation, 5% level of significance

means that we are 95% confident that the true H_0 is accepted, there are 5% chance for true H_0 to be rejected.

Errors in Example 1:

There are two possible errors.

A *Type I error* occurs when we reject a true null hypothesis. That is, a Type I error occurs when the jury convicts an innocent person. We would want the probability of this type of error [maybe 0.001 – beyond a reasonable doubt] to be very small for a criminal trial where a conviction results in the death penalty, whereas for a civil trial, where conviction might result in someone having to “pay for damages to a wrecked auto”, we would be willing for the probability to be larger [0.49 – preponderance of the evidence]

$$P(\text{Type I error}) = \alpha \text{ [usually 0.05 or 0.01]}$$

A *Type II error* occurs when we don't reject a false null hypothesis [accept the null hypothesis]. That 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 null 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” [I wish we had rejected the hypothesis and threw the pills in the trash can], it's too late because we shipped the pills to the public.

The probability of a Type I error is denoted as α (Greek letter *alpha*). The probability of a type II error is β (Greek letter *beta*).

The two probabilities are inversely related. Decreasing one increases the other, for a fixed sample size. In other words, you can't have α and β both real small for any old sample size. You may have to take a much larger sample size, or in the court example, you need much more evidence.

To start with testing hypothesis, these steps are followed:

1. State H_0 and H_a
2. Select the level of significant
3. Choose a statistic whose distribution under the null hypothesis is known
4. Specify the rejection region in terms of the selected statistic
5. Using the observed random sample, compute the value of the selected statistic
6. Decide whether or not the value computed in step 5 is included in the CR and then draw an appropriate conclusion.

Decision Rules: The analysis plan includes decision rules for rejecting the null hypothesis. In practice, statisticians describe these decision rules in two ways - with reference to a P-value or with reference to a region of acceptance.

- P-value. The strength of evidence in support of a null hypothesis is measured by the P-value.
- Region of acceptance. The region of acceptance is a range of values. If the test statistic falls within the region of acceptance, the null hypothesis is not rejected. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level.

The set of values outside the region of acceptance is called the region of rejection. If the test statistic falls within the region of rejection, the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the α level of significance. These approaches are equivalent. Some statistics texts use the P-value approach; others use the region of acceptance approach

What is a p-value?

A p-value is a **measure of how much evidence we have against the null hypothesis**. The null hypothesis, traditionally represented by the symbol H_0 , represents the hypothesis of no change or no effect.

The smaller the p-value, the more evidence we have against H_0 . It is also a measure of how likely we are to get a certain sample result or a result “more extreme,” assuming H_0 is true. The type of hypothesis (right tailed, left tailed or two tailed) will determine what “more extreme” means.

Much research involves making a hypothesis and then collecting data to test that hypothesis. In particular, researchers will set up a null hypothesis, a hypothesis that presumes no change or no effect of a treatment. Then these researchers will collect data and measure the consistency of this data with the null hypothesis.

The p-value measures consistency by calculating the probability of observing the results from your sample of data or a sample with results more extreme, assuming the null hypothesis is true. The smaller the p-value, the greater the inconsistency.

Traditionally, researchers will reject a hypothesis if the p-value is less than 0.05. Sometimes, though, researchers will use a stricter cut-off (e.g., 0.01) or a more liberal cut-off (e.g., 0.10). The general rule is that **a small p-value is evidence against the null hypothesis while a large p-value means little or no evidence against the null hypothesis**. Note that little or no evidence against the null hypothesis is not the same as a lot of evidence for the null hypothesis.

It is easiest to understand the p-value in a data set that is already at an extreme. Suppose that a drug company alleges that only 50% of all patients who take a certain drug will have an adverse event of some kind. You believe that the adverse event rate is much higher. In a sample of 12 patients, all twelve have an adverse event.

The data supports your belief because it is inconsistent with the assumption of a 50% adverse event rate. It would be like flipping a coin 12 times and getting heads each time.

The p-value, the probability of getting a sample result of **12 adverse events in 12 patients** assuming that the adverse event rate is 50%, is a measure of this inconsistency. **The p-value, 0.000244, is small enough that we would**

reject the hypothesis that the adverse event rate was only 50%.

A large p-value should not automatically be construed as evidence in support of the null hypothesis. Perhaps the failure to reject the null hypothesis was caused by an inadequate sample size. When you see a large p-value in a research study, you should also look for one of two things:

The main goal of hypothesis testing is to tell us whether we have enough evidence to reject the null hypothesis. In our case we want to know whether the coin is biased or not, so our null hypothesis should be “the coin is fair.” If we get enough evidence that contradicts this hypothesis, say, by flipping it 100 times and having it come up heads only once, then we can safely reject it.

Example: For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in Heads and half, in Tails. The alternative hypothesis might be that the number of Heads and Tails would be very different. Symbolically, these hypotheses would be expressed as

$$H_0: P = 0.5$$

$$H_1: P \neq 0.5$$

Power of a Hypothesis Test: The probability of not committing a Type II error is called the **power** of a hypothesis test.

Effect Size: To compute the power of the test, one offers an alternative view about the "true" value of the population parameter, assuming that the null hypothesis is false. The effect size is the difference between the true value and the value specified in the null hypothesis. **Effect size = True value - Hypothesized value.**

For example, suppose the null hypothesis states that a population mean is equal to 100. A researcher might ask: What is the probability of rejecting the null hypothesis if the true population mean is equal to 90? In this example, the effect size would be 90 - 100, which equals -10.

Factors That Affect Power : The power of a hypothesis test is affected by three factors.

- Sample size (n). Other things being equal, the greater the sample size, the greater the power of the test.
- Significance level (α). 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; i.e., less likely to make a Type II error. Hence, the power of the test is increased.

- 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.

How to select the C.R:

The C.R. is selected in a way that the prob. of type I error (α) and prob. of type II error (β) are as minimum as possible. In practice, we find that reducing simultaneously the values of these two probabilities is not possible without changing the sample size.

Consider an extreme case: reduce prob. of type I error (α) to zero ($\alpha \rightarrow 0$); it is achieved by including the whole range of the test statistic in acceptance region., i.e., always accepting H_0 , or, always rejecting H_1 , and prob. of type II error (β) =1.

Therefore, minimize α implies that maximize β . To overcome this, we fix an upper bound on the value of one of these two probabilities (basically an upper bound on prob. of type I error) and the probability is reduced to a minimum if possible.

The pre-assigned upper bound on the level of prob. of type I error (α) is called the Level of Significant of the C.R. while the maximum attained by this probability is called the Size of the C.R.

Example: Consider the following hypothesis :

$$H_0: P = 0.05$$

$H_1: P = 0.2$, where p is the probability of success for a Binomial Experiment. Find the best C.R. in terms of the number of successes in a R.S. of size $n=3$, subject to the restriction that prob. of type I error should not exceeds 0.1.

All possible C.R. together with the associated prob. Of type I error and prob. Of type II error are in the following table.

C.R	prob. Of type I error	prob. Of type II error
{0,1,2,3}	1.000	0.000
{0,1,2}	0.999	0.008
{0,1,3}	0.9929	0.0096
{1,2,3}	0.1426	0.5120
{0,1}	0.9928	0.1040
{0,2}	0.8645	0.3920
{0,3}	0.8575	0.4800
{1,2}	0.1425	0.5200

{1,3}	0.1355	0.6080
{2,3}	0.0072	0.8960
{0}	0.1354	0.9040
{1}	0.0071	0.9040
{2}	0.0001	0.9920
{3}	0.0000	1.0000

The prob. Of type I error should not exceeds 0.1 includes the following:

{2,3}	0.0072	0.8960
{2}	0.0001	0.9920
{3}	0.0000	1.0000

The upper bound $\alpha = 0.1$ is called the level of significance of the C.R.

The size of the actual region = 0.0072.

Conclusions:

1. More than one C.R. may yield the same prob. Of type I error.
2. The size of a test may exactly equals the given level of significance.

P-value for a test statistic is the smallest level of significance α for which the observed data indicates that the null hypothesis should be rejected. The smallest the p - value becomes the more evidence that the null hypothesis should be rejected.

The null hypothesis should be rejected for any value of α down to and including the P-value . (if $\alpha < P$ -value, then H_0 cannot be rejected).