Type I and Type II Errors

The statistical practice of hypothesis testing is widespread not only in statistics, but also throughout the natural and social sciences. While we are conducting a hypothesis test there a couple of things that could go wrong. There are two kinds of errors, which by design cannot be avoided, and we must be aware that these errors exist.

The errors are given the quite pedestrian names of type I and type II errors. At first glance it may seem that we would want to make the probability of both of these errors as small as possible. And if it is not possible to reduce the probabilities of these errors, we may wonder which of the two errors is more serious to make.

What are type I and type II errors, and how we distinguish between them?

Hypothesis Testing

The process of hypothesis testing can seem to be quite varied with a multitude of test statistics. But the general process is the same. Hypothesis testing involves the statement of a null hypothesis, and the selection of a level of significance. The null hypothesis is either true or false, and represents the default claim for a treatment or procedure. For example, when examining the effectiveness of a drug, the null hypothesis would be that the drug has no effect on a disease.

After formulating the null hypothesis and choosing a level of significance, we acquire data through observation. Statistical calculations tell us whether or not we should reject the null hypothesis.

In an ideal world we would always reject the null hypothesis when it is false, and we would not reject the null hypothesis when it is indeed true. But there are two other scenarios that are possible, each of which will result in an error.

In our statistical test, the null hypothesis is a statement of no effect. Many times the null hypothesis is a statement of the prevailing claim about a population. The alternative hypothesis is the statement that we wish to provide evidence for in our hypothesis test.

For tests of significance there are four possible results:

1. We reject the null hypothesis and the null hypothesis is true. This is what is known as a Type I error.
2. We reject the null hypothesis and the alternative hypothesis is true. In this situation the correct decision has been made.
3. We fail to reject the null hypothesis and the null hypothesis is true. In this situation the correct decision has been made.
4. We fail to reject the null hypothesis and the alternative hypothesis is true. This is what is known as a Type II error.
**Type I Error**

The first kind of error that is possible involves the rejection of a null hypothesis that is actually true.

This kind of error is called a type I error, and is sometimes called an error of the first kind.

Type I errors are equivalent to false positives. Let’s go back to the example of a drug being used to treat a disease. If we reject the null hypothesis in this situation, then our claim is that the drug does in fact have some effect on a disease. But if the null hypothesis is true, then in reality the drug does not combat the disease at all. The drug is falsely claimed to have a positive effect on a disease.

Type I errors can be controlled. The value of alpha, which is related to the level of significance that we selected has a direct bearing on type I errors.

Alpha is the maximum probability that we have a type I error. For a 95% confidence level, the value of alpha is 0.05. This means that there is a 5% probability that we will reject a true null hypothesis. In the long run, one out of every twenty hypothesis tests that we perform at this level will result in a type I error.

**Type II Error**

On the other hand, a Type II error occurs when the alternative hypothesis is true and we do not reject the null hypothesis. It is also referred to as an error of the second kind.

Type II errors are equivalent to false negatives. If we think back again to the scenario in which we are testing a drug, what would a type II error look like? A type II error would occur if we accepted that the drug had no effect on a disease, but in reality it did.

The probability of a type II error is given by the Greek letter beta. This number is related to the power or sensitivity of the hypothesis test, denoted by 1 – beta.

The probability of a type II error is typically indicated by the Greek letter beta. This is the complement event to the power of a test. By the use of the complement rule, we have that the power of a test has value of 1 – beta. Due to the fact that power is a probability, the value for power can range anywhere from 0 to 1.

**Typical Values for Power**

Although the power of a test can have values from 0 to 1, some values are better than others.

Since the power indicates the probability of a correct decision being made, the higher that the power is, the better. This corresponds to the desire to have a lower value of beta.

Much as we have control over the value of alpha, we can control the value of beta in a hypothesis test as well. The main way of doing this is through our choice of sample size.
One rule of thumb regarding the size of beta is helpful in determining which value should be used for the power of a test. A general guideline is that beta should be no more than four times the size of alpha. A common value of statistical significance is 95%, corresponding to a value of alpha of 0.05.

Thus a typical value for beta is 0.2 or less. This means that power should be 1 - 0.2 = 0.8 or more.

In some situations, such as medical testing, we would want the power of the test to be as close to 1 as possible. The reason for this is that there are some applications when we would want to be sure that the probability of a false negative is very low. A positive result in a medical test can always be supplemented with a follow up test. So even if this is a false positive, the follow up test will reveal this. However a false negative could result in a patient mistakenly thinking that he or she is healthy, when there is actually a problem.

**How to Avoid Errors**

Type I and type II errors are part of the process of hypothesis testing. Although the errors cannot be completely eliminated, we can minimize one type of error.

Typically when we try to decrease the probability one type of error, the probability for the other type increases. We could decrease the value of alpha from 0.05 to 0.01, corresponding to a 99% level of confidence. However, if everything else remains the same, then the probability of a type II error will nearly always increase.

Many times the real world application of our hypothesis test will determine if we are more accepting of type I or type II errors. This will then be used when we design our statistical experiment.

**Is a Type I Error or a Type II Error More Serious?**

The short answer to this question is that it really depends upon the situation.

In some cases a Type I error is preferable to a Type II error. In other applications a Type I error is more dangerous to make than a Type II error. We need to carefully consider the consequences of both of these kinds of errors, then plan our statistical test procedure accordingly.

**Which Error Is Better**

By thinking in terms of false positive and false negative results, we are better equipped to consider which of these errors are better. Suppose you are designing a medical screening for a disease. Is a Type I or a Type II error better? A false positive may give our patient some anxiety, but this will lead to other testing procedures. Ultimately our patient will discover that the initial test was incorrect.

Contrasted to this, a false negative will give our patient the incorrect assurance that he does not have a disease when he in fact does. As a result of this incorrect information, the disease will not be treated. If we could choose between these two options, a false positive is more desirable than a false negative.
Now suppose that you have been put on trial for murder. The null hypothesis here is that you are not guilty. Which of the two errors is more serious? Again, it depends. A Type I error occurs when you are found guilty of a murder that you did not commit. This is a very dire outcome for you. A Type II error occurs when you are guilty but are found not guilty. This is a good outcome for you, but not for society as a whole. Here we see the value in a judicial system that seeks to minimize Type I errors.

**What Level of Alpha Determines Statistical Significance?**

Not all results of hypothesis tests are equal. A hypothesis test or test of statistical significance typically has a level of significance attached to it. This level of significance is a number that is typically denoted with the Greek letter alpha. One question that comes up in statistics class is, “What value of alpha should be used for our hypothesis tests?”

The answer to this question, as with many other questions in statistics is, “It depends on the situation.” We will explore what we mean by this.

Many journals throughout different disciplines define that statistically significant results are those for which alpha is equal to 0.05 or 5%. But the main point to note is that there is not a universal value of alpha that should be used for all statistical tests.

**Commonly Used Values Levels of Significance**

The number represented by alpha is a probability, so it can take a value of any nonnegative real number less than one. Although in theory any number between 0 and 1 can be used for alpha, when it comes to statistical practice this is not the case. Of all levels of significance the values of 0.10, 0.05 and 0.01 are the ones most commonly used for alpha.

As we will see, there could be reasons for using values of alpha other than the most commonly used numbers.

**Level of Significance and Type I Errors**

One consideration against a “one size fits all” value for alpha has to do with what this number is the probability of. The level of significance of a hypothesis test is exactly equal to the probability of a Type I error. A Type I error consists of incorrectly rejecting the null hypothesis when the null hypothesis is actually true.

The smaller the value of alpha, the less likely it is that we reject a true null hypothesis.

There are different instances where it is more acceptable to have a Type I error. A larger value of alpha, even one greater than 0.10 may be appropriate when a smaller value of alpha results in a less desirable outcome.

In medical screening for a disease, consider the possibilities of a test that falsely tests positive for a disease with one that falsely tests negative for a disease. A false positive will result in anxiety for our patient, but will lead to other tests that will determine that the verdict of our test was indeed incorrect.
A false negative will give our patient the incorrect assumption that he does not have a disease when he in fact does. The result is that the disease will not be treated. Given the choice we would rather have conditions that result in a false positive than a false negative.

In this situation we would gladly accept a greater value for alpha if it resulted in a tradeoff of a lower likelihood of a false negative.

**Level of Significance and P-Values**

A level of significance is a value that we set to determine statistical significance. This is ends up being the standard by which we measure the calculated p-value of our test statistic. To say that a result is statistically significant at the level alpha just means that the p-value is less than alpha. For instance, for a value of alpha = 0.05, if the p-value is greater than 0.05, then we fail to reject the null hypothesis.

There are some instances in which we would need a very small p-value to reject a null hypothesis. If our null hypothesis concerns something that is widely accepted as true, then there must be a high degree of evidence in favor of rejecting the null hypothesis. This is provided by a p-value that is much smaller than the commonly used values for alpha.

**Conclusion**

There is not one value of alpha that determines statistical significance. Although numbers such as 0.10, 0.05 and 0.01 are values commonly used for alpha, there is no overriding mathematical theorem that says these are the only levels of significance that we can use. As with many things in statistics we must think before we calculate and above all use common sense.