7.3: Outcomes and the Type I and Type II Errors

When you perform a hypothesis test, there are four possible outcomes depending on the actual truth (or falseness) of the null hypothesis H_0 and the decision to reject or not. The outcomes are summarized in the following table:

Table 9.2

Statistical Decision	H_0 is actually...	
	True	**False**
Cannot accept H_0	Type I error	Correct outcome
Cannot reject H_0	Correct outcome	Type II error

The four possible outcomes in the table are:

1. The decision is **cannot reject H_0** when **H_0 is true (correct decision).**
2. The decision is **cannot accept H_0** when **H_0 is true** (incorrect decision known as a **Type I error**). This case is described as "rejecting a good null". As we will see later, it is this type of error that we will guard against by setting the probability of making such an error. The goal is to NOT take an action that is an error.
3. The decision is **cannot reject H_0** when, in fact, **H_0 is false** (incorrect decision known as a **Type II error**). This is called "accepting a false null". In this situation you have allowed the status quo to remain in force when it should be overturned. As we will see, the null hypothesis has the advantage in competition with the alternative.
4. The decision is **cannot accept H_0** when **H_0 is false** (**correct decision**).

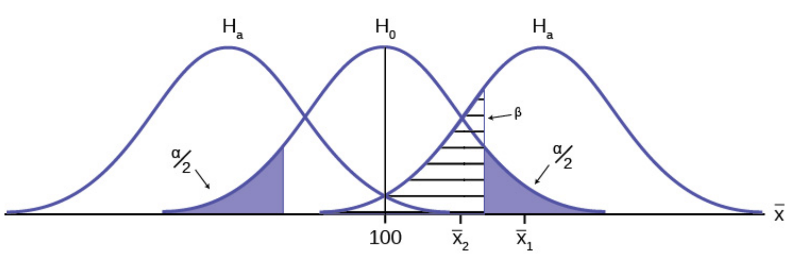
Each of the errors occurs with a particular probability. The Greek letters α and β represent the probabilities.

- α = probability of a Type I error = **$\mathbf{P}$(Type I error)** = probability of rejecting the null hypothesis when the null hypothesis is true: rejecting a good null.
- β = probability of a Type II error = **$\mathbf{P}$(Type II error)** = probability of not rejecting the null hypothesis when the null hypothesis is false. $(1-\beta)$ is called the **Power of the Test**.

α and β should be as small as possible because they are probabilities of errors.

Statistics allows us to set the probability that we are making a Type I error. The probability of making a Type I error is α. Recall that the confidence intervals in the last unit were set by choosing a value called Z_α (or t_α) and the alpha value determined the confidence level of the estimate because it was the probability of the interval failing to capture the true mean (or proportion parameter p). This alpha and that one are the same.

The easiest way to see the relationship between the alpha error and the level of confidence is with the following figure.



H_0: $\mu = 100$
H_a: $\mu \neq 100$

Figure 9.2

In the center of Figure 9.2 is a normally distributed sampling distribution marked H_0. This is a sampling distribution of $\overline{X}$ and by the Central Limit Theorem it is normally distributed. The distribution in the center is marked H_0 and represents the distribution for the null hypotheses H_0: $\mu = 100$. This is the value that is being tested. The formal statements of the null and alternative hypotheses are listed below the figure.

The distributions on either side of the H_0 distribution represent distributions that would be true if H_0 is false, under the alternative hypothesis listed as Ha. We do not know which is true, and will never know. There are, in fact, an infinite number of distributions from which the data could have been drawn if Ha is true, but only two of them are on Figure 9.2 representing all of the others.

To test a hypothesis we take a sample from the population and determine if it could have come from the hypothesized distribution with an acceptable level of significance. This level of significance is the alpha error and is marked on Figure 9.2 as the shaded areas in each tail of the H_0 distribution. (Each area is actually \alpha/2 because the distribution is symmetrical and the alternative hypothesis allows for the possibility for the value to be either greater than or less than the hypothesized value--called a two-tailed test).

If the sample mean marked as $\overline{X}_1$ is in the tail of the distribution of H_0, we conclude that the probability that it could have come from the H_0 distribution is less than alpha. We consequently state, "the null hypothesis cannot be accepted with (\alpha) level of significance". The truth **may** be that this $\overline{X}_1$ did come from the H_0 distribution, but from out in the tail. If this is so then we have falsely rejected a true null hypothesis and have made a Type I error. What statistics has done is provide an estimate about what we know, and what we control, and that is the probability of us being wrong, α.

We can also see in Figure 9.2 that the sample mean could be really from an Ha distribution, but within the boundary set by the alpha level. Such a case is marked as $\overline{X}_2$. There is a probability that $\overline{X}_2$ actually came from Ha but shows up in the range of H_0 between the two tails. This probability is the beta error, the probability of accepting a false null.

Our problem is that we can only set the alpha error because there are an infinite number of alternative distributions from which the mean could have come that are not equal to H_0. As a result, the statistician places the burden of proof on the alternative hypothesis. That is, we will not reject a null hypothesis unless there is a greater than 90, or 95, or even 99 percent probability that the null is false: the burden of proof lies with the alternative hypothesis. This is why we called this the tyranny of the status quo earlier.

By way of example, the American judicial system begins with the concept that a defendant is "presumed innocent". This is the status quo and is the null hypothesis. The judge will tell the jury that they can not find the defendant guilty unless the evidence indicates guilt beyond a "reasonable doubt" which is usually defined in criminal cases as 95% certainty of guilt. If the jury cannot accept the null, innocent, then action will be taken, jail time. The burden of proof always lies with the alternative hypothesis. (In civil cases, the jury needs only to be more than 50% certain of wrongdoing to find culpability, called "a preponderance of the evidence").

The example above was for a test of a mean, but the same logic applies to tests of hypotheses for all statistical parameters one may wish to test.

The following are examples of Type I and Type II errors.

? Example 9.4

Suppose the null hypothesis, H_0, is: Frank's rock climbing equipment is safe.

Type I error: Frank thinks that his rock climbing equipment may not be safe when, in fact, it really is safe.

Type II error: Frank thinks that his rock climbing equipment may be safe when, in fact, it is not safe.

$\alpha =$ **probability** that Frank thinks his rock climbing equipment may not be safe when, in fact, it really is safe. $\beta =$ **probability** that Frank thinks his rock climbing equipment may be safe when, in fact, it is not safe.

Notice that, in this case, the error with the greater consequence is the Type II error. (If Frank thinks his rock climbing equipment is safe, he will go ahead and use it.)

This is a situation described as "accepting a false null".

? Example 9.5

Suppose the null hypothesis, H_0, is: The victim of an automobile accident is alive when he arrives at the emergency room of a hospital. This is the status quo and requires no action if it is true. If the null hypothesis cannot be accepted then action is required and the hospital will begin appropriate procedures.

Type I error: The emergency crew thinks that the victim is dead when, in fact, the victim is alive. **Type II error**: The emergency crew does not know if the victim is alive when, in fact, the victim is dead.

$\alpha =$ **probability** that the emergency crew thinks the victim is dead when, in fact, he is really alive = P(Type I error). $\beta =$ **probability** that the emergency crew does not know if the victim is alive when, in fact, the victim is dead = P(Type II error).

The error with the greater consequence is the Type I error. (If the emergency crew thinks the victim is dead, they will not treat him.)

Exercise 9.5

Suppose the null hypothesis, H_0, is: a patient is not sick. Which type of error has the greater consequence, Type I or Type II?

Example 9.6

It's a Boy Genetic Labs claim to be able to increase the likelihood that a pregnancy will result in a boy being born. Statisticians want to test the claim. Suppose that the null hypothesis, H_0, is: It's a Boy Genetic Labs has no effect on gender outcome. The status quo is that the claim is false. The burden of proof always falls to the person making the claim, in this case the Genetics Lab.

Type I error: This results when a true null hypothesis is rejected. In the context of this scenario, we would state that we believe that It's a Boy Genetic Labs influences the gender outcome, when in fact it has no effect. The probability of this error occurring is denoted by the Greek letter alpha, \alpha.

Type II error: This results when we fail to reject a false null hypothesis. In context, we would state that It's a Boy Genetic Labs does not influence the gender outcome of a pregnancy when, in fact, it does. The probability of this error occurring is denoted by the Greek letter beta, \beta.

The error of greater consequence would be the Type I error since couples would use the It's a Boy Genetic Labs product in hopes of increasing the chances of having a boy.

Exercise 9.6

"Red tide" is a bloom of poison-producing algae–a few different species of a class of plankton called dinoflagellates. When the weather and water conditions cause these blooms, shellfish such as clams living in the area develop dangerous levels of a paralysis-inducing toxin. In Massachusetts, the Division of Marine Fisheries (DMF) monitors levels of the toxin in shellfish by regular sampling of shellfish along the coastline. If the mean level of toxin in clams exceeds 800 μg (micrograms) of toxin per kg of clam meat in any area, clam harvesting is banned there until the bloom is over and levels of toxin in clams subside. Describe both a Type I and a Type II error in this context, and state which error has the greater consequence.

Example 9.7

A certain experimental drug claims a cure rate of at least 75% for males with prostate cancer. Describe both the Type I and Type II errors in context. Which error is the more serious?

Type I: A cancer patient believes the cure rate for the drug is less than 75% when it actually is at least 75%.

Type II: A cancer patient believes the experimental drug has at least a 75% cure rate when it has a cure rate that is less than 75%.

In this scenario, the Type II error contains the more severe consequence. If a patient believes the drug works at least 75% of the time, this most likely will influence the patient's (and doctor's) choice about whether to use the drug as a treatment option.
