

## 7.3: Outcomes and 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 7.3.1

Statistical Decision	$H_0$ is actually...	
	True	False
Reject $H_0$	Type I error	Correct outcome
Do not reject $H_0$	Correct outcome	Type II error

The four possible outcomes in the table are:

1. The decision is **reject  $H_0$**  when  **$H_0$  is true** (incorrect decision known as a **Type I error**). This can be thought of as a "false positive" or "false alarm". As we will see later, it is this type of error that we will guard against by setting a small probability of making such an error.
2. The decision is **reject  $H_0$**  when  **$H_0$  is false** (**correct decision**).
3. The decision is **do not reject  $H_0$**  when  **$H_0$  is true** (**correct decision**).
4. The decision is **do not reject  $H_0$**  when, in fact,  **$H_0$  is false** (incorrect decision known as a **Type II error**). This can be thought of as a "false negative" or "miss".

Each of the errors occurs with a particular probability. The Greek letters  $\alpha$  and  $\beta$  represent the probabilities.

- $\alpha$  = probability of a Type I error =  **$P(\text{Type I error})$**  = probability of rejecting the null hypothesis when the null hypothesis is true: rejecting a good null.
- $\beta$  = probability of a Type II error =  **$P(\text{Type II error})$**  = probability of not rejecting the null hypothesis when the null hypothesis is false.  $(1 - \beta)$  is called the **power of the test**, or **statistical power**.

$\alpha$  and  $\beta$  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  $\alpha$ . Recall that the confidence intervals in the last unit were set by choosing a value called  $z_\alpha$  (or  $t_\alpha$ ) 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 true proportion  $P$ ). 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.



Figure 7.3.1

In the center of Figure 7.3.1 is a normally distributed sampling distribution marked  $H_0$ . This is a sampling distribution of  $\bar{x}$  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  $H_a$ . 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  $H_a$  is true, but only two of them are on Figure 7.3.1 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 7.3.1 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 in this example 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  $\bar{x}_1$  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 is rejected at  $\alpha$  level of significance". The

truth **may** be that this  $\bar{x}_1$  did come from the  $H_0$  distribution, but from out in the tail. If this is so then we have erroneously 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,  $\alpha$ .

We can also see in Figure 7.3.1 that the sample mean could be really from an  $H_a$  distribution, but within the boundary set by the alpha level. Such a case is marked as  $\bar{x}_2$ . There is a probability that  $\bar{x}_2$  actually came from  $H_a$  but shows up in the range of  $H_0$  between the two tails. This probability is the beta error, the probability of failing to reject 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 call this the tyranny of the status quo.

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 rejects the null, innocence, then action will be taken, jail time. The null hypothesis is the "default", so 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 single 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 7.3.1

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

**Type I error:** Frank thinks that his rock climbing equipment is not safe (that is, rejecting the null) when, in fact, it really is safe (that is, the null is really true).

**Type II error:** Frank thinks that his rock climbing equipment is safe (that is, failing to reject the null) when, in fact, it is not safe (that is, the null is really false).

$\alpha$  = **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. (Frank will be using unsafe climbing equipment!)

#### Example 7.3.2

Suppose the null hypothesis,  $H_0$ , is: A particular work-from-home training program helps remote workers be more productive, as shown by an employee survey after the training.

**Type I error:** The employee survey shows that the training program is effective when, in fact, the training program is not effective.

**Type II error:** The employee survey shows that the training program is not effective when, in fact, the training program is effective.

$\alpha$  = **probability** that the employee survey shows that the training program is effective when, in fact, the training program is not effective =  $P(\text{Type I error})$ .

$\beta$  = **probability** that the employee survey shows that the training program is not effective when, in fact, the training program is effective =  $P(\text{Type II error})$ .

In this case, one could argue that both a Type I error and a Type II error could have important consequences for the company. If we commit a Type I error, we could be using an ineffective training program, which may waste company and worker resources.

On the other hand, if we commit a Type II error, we could decide to cut the program, thinking it is not working, but we would be missing out on a truly effective training program for our workers.

#### Exercise 7.3.1

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 7.3.3

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

**Type II error:** 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.

---

This page titled [7.3: Outcomes and Type I and Type II Errors](#) is shared under a [CC BY](#) license and was authored, remixed, and/or curated by .