Marta Olazabal Chias

Type I vs Type II errors.

With hypothesis testing you can see if your data support or refute your research question or prediction, using the null and alternative hypothesis (1). As all decisions are based on probabilities measures, there is always a risk of making a wrong conclusion. Depending on whether the null hypothesis is true or false, and assuming that the study is free of bias, 4 situations are possible (2):

The null hypothesis is…	**True**	**False**
Rejected	Type I error (false positive) Significance level α	Correct decision Power of the test 1 - β
Not rejected	Correct decision	Type II error β (false negative)

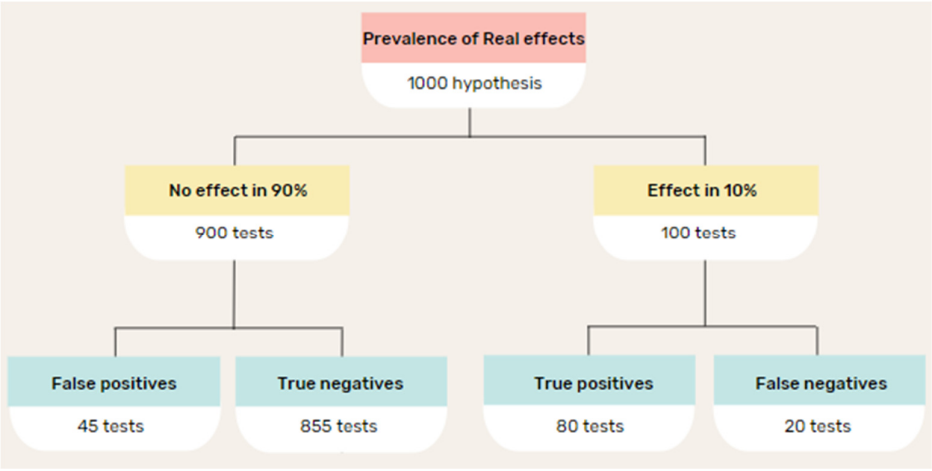
TYPE I ERROR.

Type I error, also known as false-positive, occurs when the investigator rejects a null hypothesis that is true. The risk of committing the error is the significance level (α).

There are some factors that influence this type of error (3):

- Prevalence of real effects: Indicates the probability that an effect exists in the population before conducting your study.
- Power
- Significance level

In the next figure you can see the amount of false positives (type I error), with preestablished conditions: perform 1000 test with a prevalence of 0.1, significance level of 0,05 (5% of the test will incorrectly be significant) and a power of 80% (80% of these tests will correctly detect the effect being true positives),



We can also make ourselves two questions:

1) If a result is statistically significant, what is the probability that the null hypothesis is really true?

2) Of all experiments that reach a statistically significant conclusion, what fraction are false positives (Type I errors)?

To answer these two questions, we have the False Positive Report Probability (FPRP) or the False Positive Rate (FPR); that is not the same as the significance level. The FPRP is higher when the prior probability is low and when the power is low (4).

Why do type I error occur? Type I error can be due to bad luck, or because we didn't respect the duration of the experiment or had a wrong sample size. False-positive findings can easily arise when statistical methods are applied incorrectly or when P-values are interpreted without sufficient understanding.

And also, it can appear due to bias in the experiment, but they might be difficult to detect and cannot usually be quantified.

How can we reduce type I error? We can reduce type I error reducing the significance level. Also, lower p-values indicate stronger evidence against the null hypothesis and a lower probability of a false positive (5).

TYPE II ERROR.

The type II error, or false negative, occurs when the investigator fails to reject a null hypothesis that is false. This means that your study may not have had enough statistical power to detect an effect of a certain size.

We can't set the Type II error rate for your analysis; but we can estimate it before we begin the study by approximating properties of the alternative hypothesis that we're studying. This type of estimation, it's called power analysis (6). Power is the extent to which a test can correctly detect a real effect when there is one, so the risk of type II error is inversely related to the statistical power (1). Analysts typically estimate power rather than type II error directly.

There are 3 factors that affect the power and consequently type II errors:

- Sample size
- Variability in population
- Effect size

Why do type II error occur? Small effect sizes, small sample sizes, high data variability, small power of the test.

How can we reduce it? We need to ensure the statistical power is 80% or above and also, increasing the sample size of the experiment (7).

SOME EXAMPLES TO UNDERSTAND TYPE I AND TYPE II ERROR.

Example 1: You went out with some friends to have drinks; and in your way back you get stopped by the police and take a breathalyser test. *Nobody wants to be stopped for a breath alcohol test, but then nobody wants to be killed by a drunk driver either.*

The null hypothesis: "You are below the alcohol limit."

A **false positive** would show that you are over the limit when you haven't even touched an alcoholic drink, and a **false negative** would register you as sober when you are drunk, or at least over the limit.

Example 2: You test whether a new drug intervention can alleviate symptoms of an autoimmune disease.

The null hypothesis (H0) is that the new drug has no effect on symptoms of the disease. The alternative hypothesis (H1) is that the drug is effective for alleviating symptoms of the disease.

A **Type I error or false positive** happens when you conclude that the drug intervention improved symptoms when it actually didn't. These improvements could have arisen from other random factors or measurement errors. A **Type II error or false negative** happens when you conclude that the drug intervention didn't improve symptoms when it actually did. Your study may have missed key indicators of improvements or attributed any improvements to other factors instead.

Example 3: A biotechnology company wants to compare how effective two of its drugs are for treating diabetes. The biotech company implements a large clinical trial of 3,000 patients with diabetes to compare the treatments. The company randomly divides the 3,000 patients into two equally sized groups, giving one group one of the treatments and the other group the other treatment.

The null hypothesis states the two medications are equally effective.

In this situation if the company concludes that the two medications are equally effective when they are not, we are committing a **type I error**; and if they state that the drugs are not equally effective even though they really are, we are committing a **type II error**.

TYPE I VS TYPE II ERROR.

In the following graph we can see a comparison between the two errors. The distribution on the left represents the null hypothesis; if this hypothesis is true, we only need to worry about type I error (α). And the distribution on the right is that the alternative hypothesis is true, and here we need to worry about Type II errors. The rest of the alternative distribution represents the probability of detect an effect correctly (the power of the test).

We see there is a critical value that will change by changing the significance level. If you decrease the significance level the line will move to the right and you will reduce the Type I error and increase the Type II error.

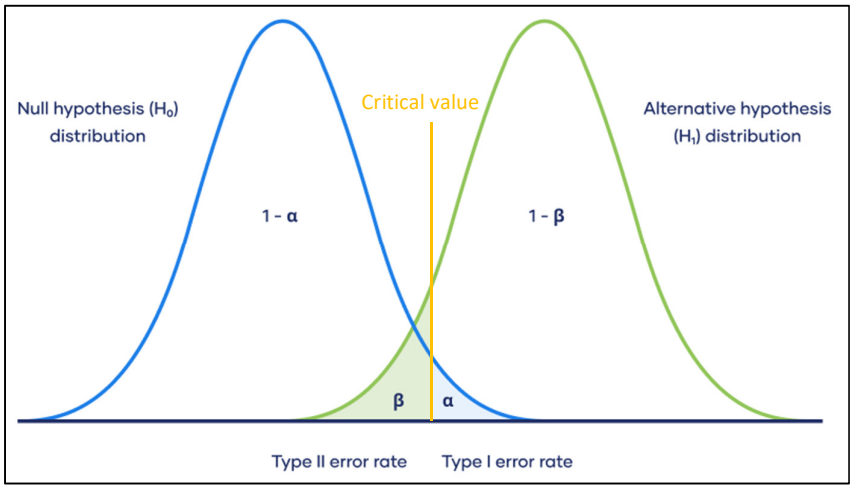


Figure 1. Type I vs Type II error. Bhandari, P. (2022, November 11). Type I & Type II Errors | Differences, Examples, Visualizations. Scribbr. Retrieved November 17, 2022, from https://www.scribbr.com/statistics/type-i-and-type-ii-errors/

We see there is a critical value that will change by changing the significance level. If you decrease the significance level the line will move to the right and you will reduce the Type I error and increase the Type II error.

We can conclude that one error influence the other as the significance level affects the statistical power that is related with type II error.

- If we decide to lower the significance level you will decrease the Type I error, but increase the Type II error.
- If you increase the power, you decrease type II error, but increase type I error.

So, it is important to maintain a balance between both errors. Although type I and type II errors can never be avoided entirely, the investigator can reduce their likelihood, and establish the maximum chance of making them in advance of the study (2).

In the next links you can see a few videos where all these concepts are explained in a more visual way:

https://www.youtube.com/watch?v=a_l991xUAOU

https://www.youtube.com/watch?v=Hdbbx7DIweQ

FALSE POSITIVE AND FALSE NEGATIVE RESULTS APPLIED TO DIAGNOSTICS TESTS.

Type I and Type II error can be applied on our daily basis with the used of diagnostics tests, like a pregnancy test, or a COVID-19 test. But it is also used in clinics to look at other many diagnoses like a HIV test, cancer screening, prenatal test for Down's Syndrome, etc.

All tests have a chance of resulting in false positive and false negative errors, they are an unavoidable problem in scientific testing. So, scientist have to be careful when they make decisions; try to minimize errors, collect additional information or perform a test multiple time.

For example, a blood test can be used to screen for a number of diseases, including diabetes. To test for diabetes, doctors look at the sugar level in blood when a person has not eaten recently. High blood sugar while fasting is an indicator of diabetes (null hypothesis).

If a patient did not fast before their blood test, the test may show high levels of blood sugar, and the patient may be diagnosed with diabetes when they actually do not have the disease. This is a **false positive** and can lead to unnecessary medical treatment.

On the other hand, a **false negative** is when the test shows that a patient does not have diabetes when they actually do. In this case the patient may not get treatment and get worse because their disease was undetected.

In a study carried out by Kuniaki Terato et al. they tried to prevent intense false positive and negative reactions attributed to the principle of ELISA to re-investigate antibody studies in autoimmune diseases. (https://doi.org/10.1016/j.jim.2014.03.013).

They show that in an ELISA, there are four types of false positive reactions:

1) Non-specific reaction caused by the secondary antibody.
2) Hydrophobic binding of immunoglobulin components in sample specimens to plastic surfaces.
3) Ionic interaction between immunoglobulin in sample specimens and antigen.
4) Immune recognition of blocking agents by antibodies in serum specimens

And two false negative reactions:

1) The competitive inhibition of test antibodies by relevant antibodies presents in animal serum
2) Denaturation of enzymes conjugated to detection antibodies

So, they review a new buffer to decrease this false negative and false positive reactions to assay a correct level of antibodies in human sera; because to study the possible involvement of potential environmental pathogens in the pathogenesis of autoimmune diseases, it is essential to investigate antibody responses to a variety of environmental agents and autologous components in a more effective way.

CONCLUSION.

From all of this we can conclude that it is important to understand the difference between type I and type II errors when performing hypothesis testing, so that you can reduce them previously to the study in relation to what you are looking for. As there are certain situations where is more important to reduce one of the errors rather than the other.

BIBLIOGRAPHY.

1. Bhandari, P. (2022, November 11). *Type I & Type II Errors | Differences, Examples, Visualizations.* Scribbr. Retrieved November 21, 2022, from https://www.scribbr.com/statistics/type-i-and-type-ii-errors/

2. Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial psychiatry journal*, *18*(2), 127–131. https://doi.org/10.4103/0972-6748.62274

3. Frost J. [P-Values, Error Rates, and False Positives] [Internet]. [cited 2022 Nov 21]. Available from: https://statisticsbyjim.com/hypothesis-testing/p-values-error-rates-false-positives/

4. Motulsky H. (2018). *Intuitive biostatistics: a nonmathematical guide to statistical thinking* (Fourth). Oxford University Press.

5. Frost J. [Five P Value Tips to Avoid Being Fooled by False Positives and other Misleading Hypothesis Test Results] [Internet]. [cited 2022 Nov 21]. Available from: https://statisticsbyjim.com/hypothesis-testing/p-values-error-rates-false-positives/

6. Frost J. [Types of Errors in Hypothesis Testing] [Internet]. [cited 2022 Nov 21]. Available from: https://statisticsbyjim.com/hypothesis-testing/p-values-error-rates-false-positives/

7. Blog F. [Type I vs Type II Errors: Causes, Examples & Prevention] [Internet]. [cited 2022 Nov 21]. Available from: https://www.formpl.us/blog/type-errors