**Statistics Assignment 5**

**Name = Khushal tondon**

**Ans no. 1**

[https://drive.google.com/file/d/1lKwGf-ySvrxPRVRoObW61VAs2fFbjV6I/view?usp=drivesdk](https://drive.google.com/file/d/1lKwGf-ySvrxPRVRoObW61VAs2fFbjV6I/view?usp=drivesdk" \t "_blank)

**Ans no. 2**

Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data. The test provides evidence concerning the plausibility of the hypothesis, given the data. Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed.

In hypothesis testing, an analyst tests a statistical sample, with the goal of providing evidence on the plausibility of the null hypothesis.

Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed. All analysts use a random population sample to test two different hypotheses: the null hypothesis and the alternative hypothesis.

The null hypothesis is usually a hypothesis of equality between population parameters; e.g., a null hypothesis may state that the population mean return is equal to zero. The alternative hypothesis is effectively the opposite of a null hypothesis (e.g., the population mean return is not equal to zero). Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true.

Hypothesis Testing Steps ( how does this hypothesis testing works )

Hypothesis testing can be easily performed in five simple steps. The most important step is to correctly set up the hypotheses and identify the right method for hypothesis testing. The basic steps to perform hypothesis testing are as follows:

Step 1: Set up the null hypothesis by correctly identifying whether it is the left-tailed, right-tailed, or two-tailed hypothesis testing.

Step 2: Set up the alternative hypothesis.

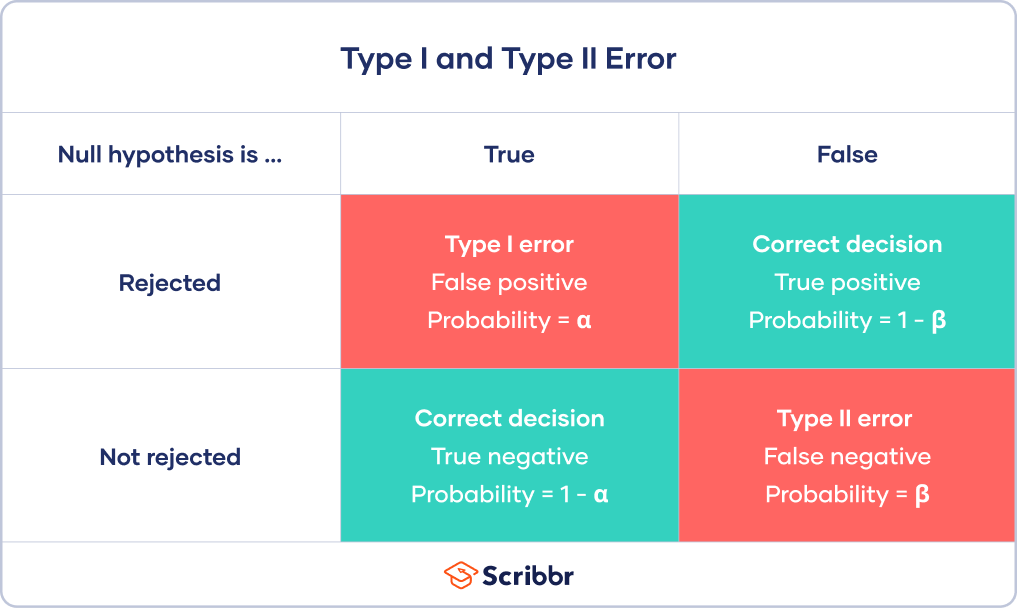
Step 3: Choose the correct significance level, α, and find the critical value. Basically state our decision rule .

Step 4: Calculate the correct test statistic (z, t or) and p-value.

Step 5: Compare the test statistic with the critical value or compare the p-value with α to arrive at a conclusion. In other words, decide if the null hypothesis is to be rejected or not.

**Ans no. 3**

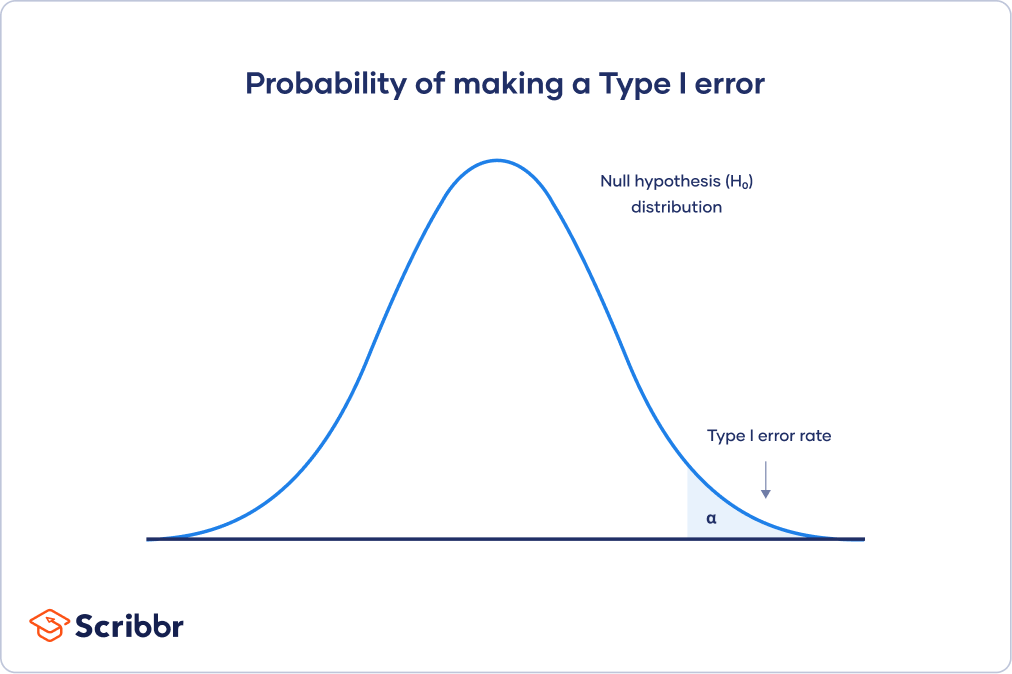
The probability of making a Type I error is the significance level, or alpha (α), while the probability of making a Type II error is beta (β).



Type I error

A Type I error means rejecting the null hypothesis when it’s actually true. It means concluding that results are statistically significant when, in reality, they came about purely by chance or because of unrelated factors.

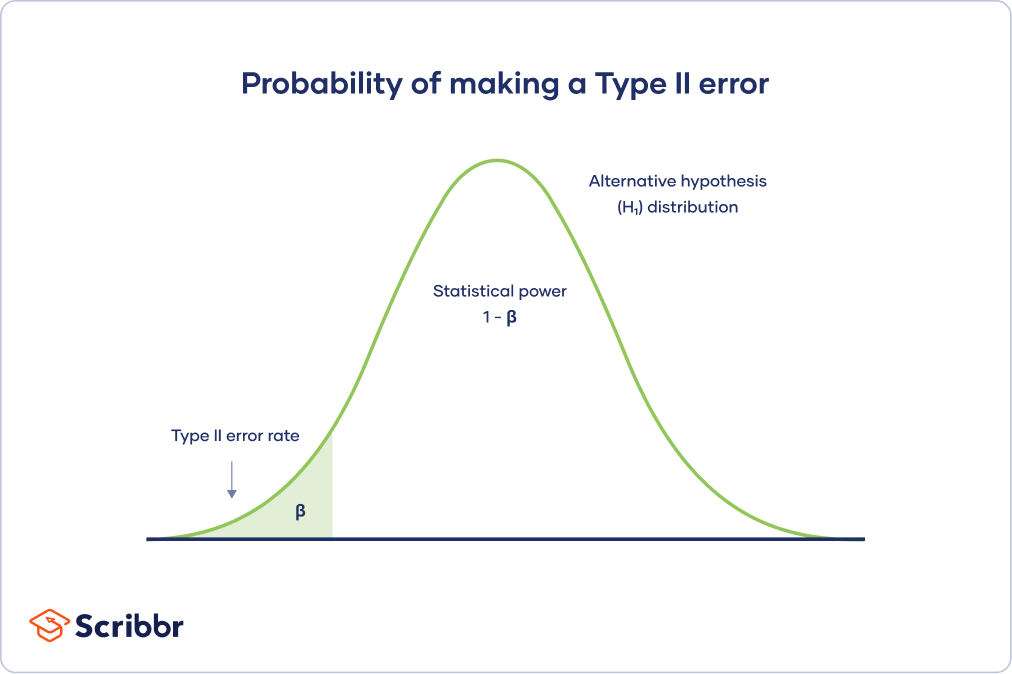
The risk of committing this error is the significance level (alpha or α) you choose. That’s a value that you set at the beginning of your study to assess the statistical probability of obtaining your results (p value).



Type II error

A Type II error means not rejecting the null hypothesis when it’s actually false. This is not quite the same as “accepting” the null hypothesis, because hypothesis testing can only tell you whether to reject the null hypothesis.

Instead, a Type II error means failing to conclude there was an effect when there actually was. In reality, your study may not have had enough statistical power to detect an effect of a certain size.



**Type 1 error is most hazardous**

For statisticians, a Type I error is usually worse. In practical terms, however, either type of error could be worse depending on your research context.

A Type I error means mistakenly going against the main statistical assumption of a null hypothesis. This may lead to new policies, practices or treatments that are inadequate or a waste of resources.

**Ans no. 4**

A p-value, or probability value, is a number describing how likely it is that your data would have occurred by random chance (i.e. that the null hypothesis is true). P-values are used in hypothesis testing to help decide whether to reject the null hypothesis. The smaller the p-value, the more likely you are to reject the null hypothesis.

The level of statistical significance is often expressed as a p-value between 0 and 1. The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

P-values are most often used by researchers to say whether a certain pattern they have measured is statistically significant.

Statistical significance is another way of saying that the p-value of a statistical test is small enough to reject the null hypothesis of the test.

**Ans no. 5**

Probability distribution is a function that gives the relative likelihood of occurrence of all possible outcomes of an experiment. There are two important functions that are used to describe a probability distribution. These are the probability density function or probability mass function and the cumulative distribution function.

The probability distribution function is also known as the cumulative distribution function (CDF). If there is a random variable, X, and its value is evaluated at a point, x, then the probability distribution function gives the probability that X will take a value lesser than or equal to x. It can be written as F(x) = P (X ≤ x). Furthermore, if there is a semi-closed interval given by (a, b] then the probability distribution function is given by the formula P(a < X ≤ b) = F(b) - F(a). The probability distribution function of a random variable always lies between 0 and 1. It is a non-decreasing function.

Perhaps the most common probability distribution is the normal distribution, or "bell curve," although several distributions exist that are commonly used. Typically, the data generating process of some phenomenon will dictate its probability distribution. This process is called the probability density function.

Probability distributions can also be used to create cumulative distribution functions (CDFs), which adds up the probability of occurrences cumulatively and will always start at zero and end at 100%.