**Notes – Ch9 Hypothesis Tests**

The goal of hypothesis testing is to determine the likelihood that a population parameter, such as the mean, is likely to be true. There are four steps of hypothesis testing:

**Step 1:** State the hypotheses.

**Step 2:** Set the criteria for a decision.

**Step 3:** Compute the test statistic.

**Step 4:** Make a decision.

**Step 1: Formulation of hypothesis:**

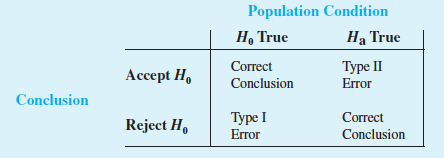
State the hypotheses. We begin by stating the value of a population mean in a null hypothesis, which we presume is true. This is a starting point so that we can decide whether this is likely to be true, similar to the presumption of innocence in a courtroom. When a defendant is on trial, the jury starts by assuming that the defendant is innocent. The basis of the decision is to determine whether this assumption is true. Likewise, in hypothesis testing, we start by assuming that the hypothesis or claim we are testing is true. This is stated in the null hypothesis. The basis of the decision is to determine whether this assumption is likely to be true.

**Null Hypothesis:** The hypothesis tentatively assumed true in the hypothesis testing procedure. It is the assumption about a population parameter we wish to test, usually, an assumption of the status quo. Symbolized by H0.

**Alternative hypothesis**: The hypothesis concluded to be true if the null hypothesis is rejected. The alternative hypothesis is often what the test is attempting to establish. Symbolized by Ha.

The purpose of hypothesis testing is not to question the computed value of the sample statistics but to make a judgement about the difference between that sample statistics and a hypothesized population parameter. The null and alternative hypotheses are competing statements about the population. Either the null hypothesis H0 is true or the alternative hypothesis Ha is true, but not both. Ideally the hypothesis testing procedure should lead to the acceptance of H0 when H0 is true and the rejection of H0 when Ha is true. Unfortunately, the correct conclusions are not always possible. Because hypothesis tests are based on sample information, we must allow for the possibility of errors.

* If H0 is true, the conclusion to accept H0 is correct.
* If Ha is true (in other words H0 is false), and we conclude to accept H0 we make a **Type II error**; that is, we accept H0 when it is false.
* If H0 is true, and we conclude to reject H0 we make a **Type I error**; that is, we reject H0 when it is true.
* However, if Ha is true (in other words H0 is false), rejecting H0 is correct.



|  |  |  |
| --- | --- | --- |
| **Null Hypothesis**  (Examples) | **Type I Error / False Positive** (H0 true, but rejected as false) | **Type II Error / False Negative** (H0 false, but accepted as true) |
| Person is not guilty of the crime | Person is judged as guilty when the person actually did not commit the crime (convicting an innocent person) | Person is judged not guilty when they actually did commit the crime (letting a guilty person go free) |
| Medicine A cures Disease B | Medicine A cures Disease B, but is rejected as false | Medicine A does not cure Disease B, but is accepted as true |
| Display Ad A is effective in driving conversions | Display Ad A is effective in driving conversions, but is rejected as false | Display Ad A is not effective in driving conversions, but is accepted as true |
| Wolf is not present | Shepherd thinks wolf is present (shepherd cries wolf) when no wolf is actually present | Shepherd thinks wolf is NOT present (shepherd does nothing) when a wolf is actually present |

**Step 2: Set the criteria for a decision:**

We also need to set a criterion for deciding whether to accept or reject the null hypothesis. This criterion can be set by setting the significance level.

**Significance level:** A value indicating the percentage of sample values that is outside certain limits, assuming the null hypothesis is correct, that is, the probability of rejecting the null hypothesis when it is true. In other words, significance level is α - the probability of making Type I error when the null hypothesis is true. The common choices for α are 0.05 and 0.01.

**Step 3: Compute the test statistic.**

The **test statistic** is a mathematical formula that allows researchers to determine the likelihood of obtaining sample outcomes if the null hypothesis were true. The value of the test statistic is used to make a decision regarding the null hypothesis.

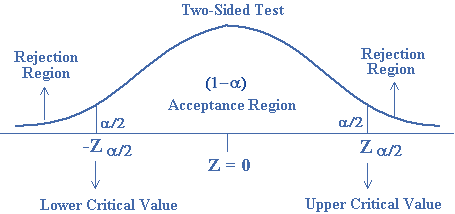
Specifically, a test statistic tells us how far, or how many standard deviations, a sample mean is from the population mean. The larger the value of the test statistic, the further the distance, or number of standard deviations, a sample mean is from the population mean stated in the null hypothesis. The value of the test statistic is used to make a decision in Step 4.

The test statistic we use depends largely on what we know about the population. There arise 2 cases for availability population variance:

1. σ known: When we know the mean and standard deviation in a single population, we can use the *z* test
2. σ unknown: In this case, we use students-t distribution and calculate t as test statistics

**One tailed and Two-tailed test**

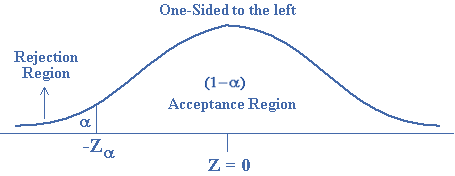
|  |  |  |  |
| --- | --- | --- | --- |
|  | **Two-tailed test** | **Left-tailed test** | **Right-tailed test** |
| **Sign in H0** |  |  |  |
| **Sign in Ha** | = | < | > |
| **Rejection Region** | Both sides | Left | Right |

**Two tailed test:** A two-tailed test of a hypothesis will reject the null hypothesis if the sample mean is significantly higher than or lower than the hypothesized population mean. Thus, it has two rejection regions.

A two-tailed test is appropriate when the null hypothesis and alternative hypothesis is given as:

H0 : μ = μ0 is

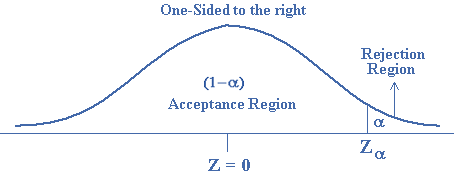
Ha : μ μ0, μ0 is the assumed population mean

**Left-tailed test**: A left-tailed test of a hypothesis will reject the null hypothesis in favour of alternative hypothesis if the sample mean is significantly lower than the hypothesized population mean. Thus, it has one rejection regions in the lower tail or left side.

A left-tailed test is appropriate when the null hypothesis and alternative hypothesis is given as:

H0 : μ μ0

Ha : μ < μ0, μ0 is the assumed population mean

**Right-tailed test:** A right-tailed test of a hypothesis will reject the null hypothesis in favour of alternative hypothesis if the sample mean is significantly higher than the hypothesized population mean. Thus, it has one rejection regions in the upper tail or right side.

A right-tailed test is appropriate when the null hypothesis and alternative hypothesis is given as:

H0 : μ μ0

Ha : μ > μ0, μ0 is the assumed population mean

**Step 4: Make a decision.**

Make a decision. We use the value of the test statistic to make a decision about the null hypothesis. There are 2 approaches for decision making:

1. p-value Approach
2. Critical Value Approach

**1. p-value Approach:** The decision is based on the probability of obtaining a sample mean, given that the value stated in the null hypothesis is true. If the probability of obtaining a sample mean is less than 5% when the null hypothesis is true, then the decision is to reject the null hypothesis. If the probability of obtaining a sample mean is greater than 5% when the null hypothesis is true, then the decision is to retain the null hypothesis. In sum, there are two decisions a researcher can make:

* Reject the null hypothesis. The sample mean is associated with a low probability of occurrence when the null hypothesis is true.
* Retain the null hypothesis. The sample mean is associated with a high probability of occurrence when the null hypothesis is true.

***p-value:*** The probability of obtaining a sample mean, given that the value stated in the null hypothesis is true, is stated by the *p*-value. The *p*-value is a probability, so it varies between 0 and 1 and can never be negative.

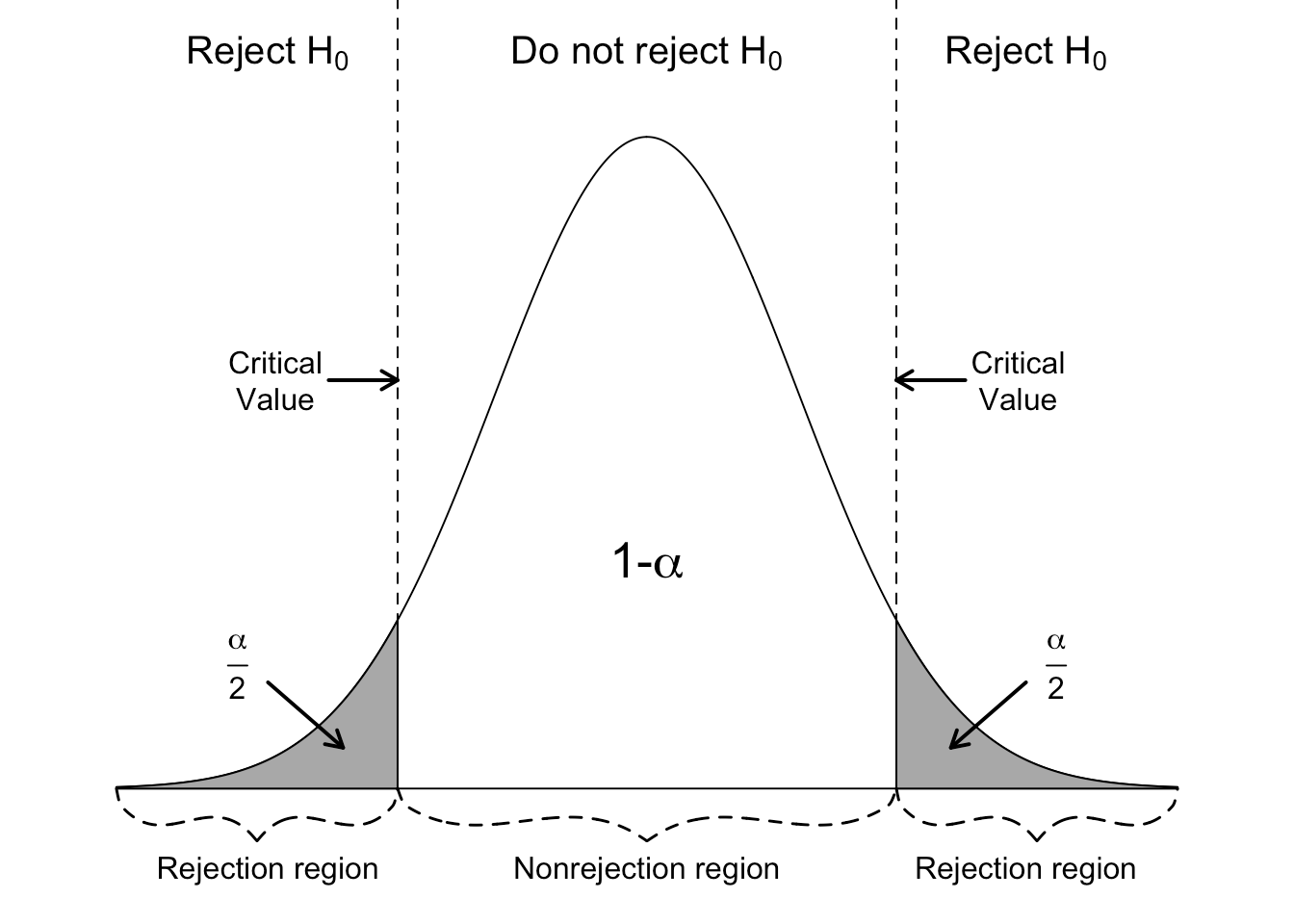
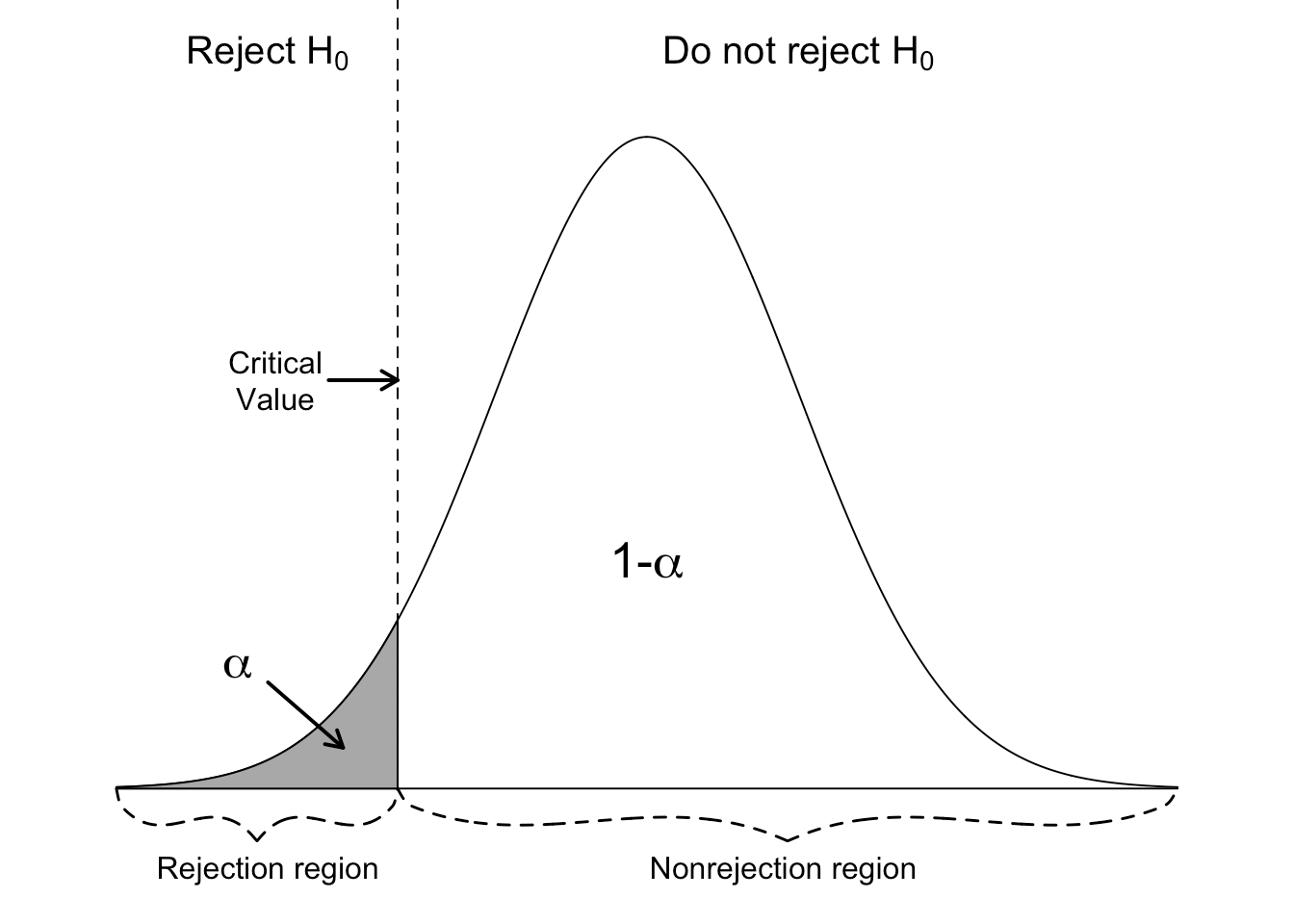
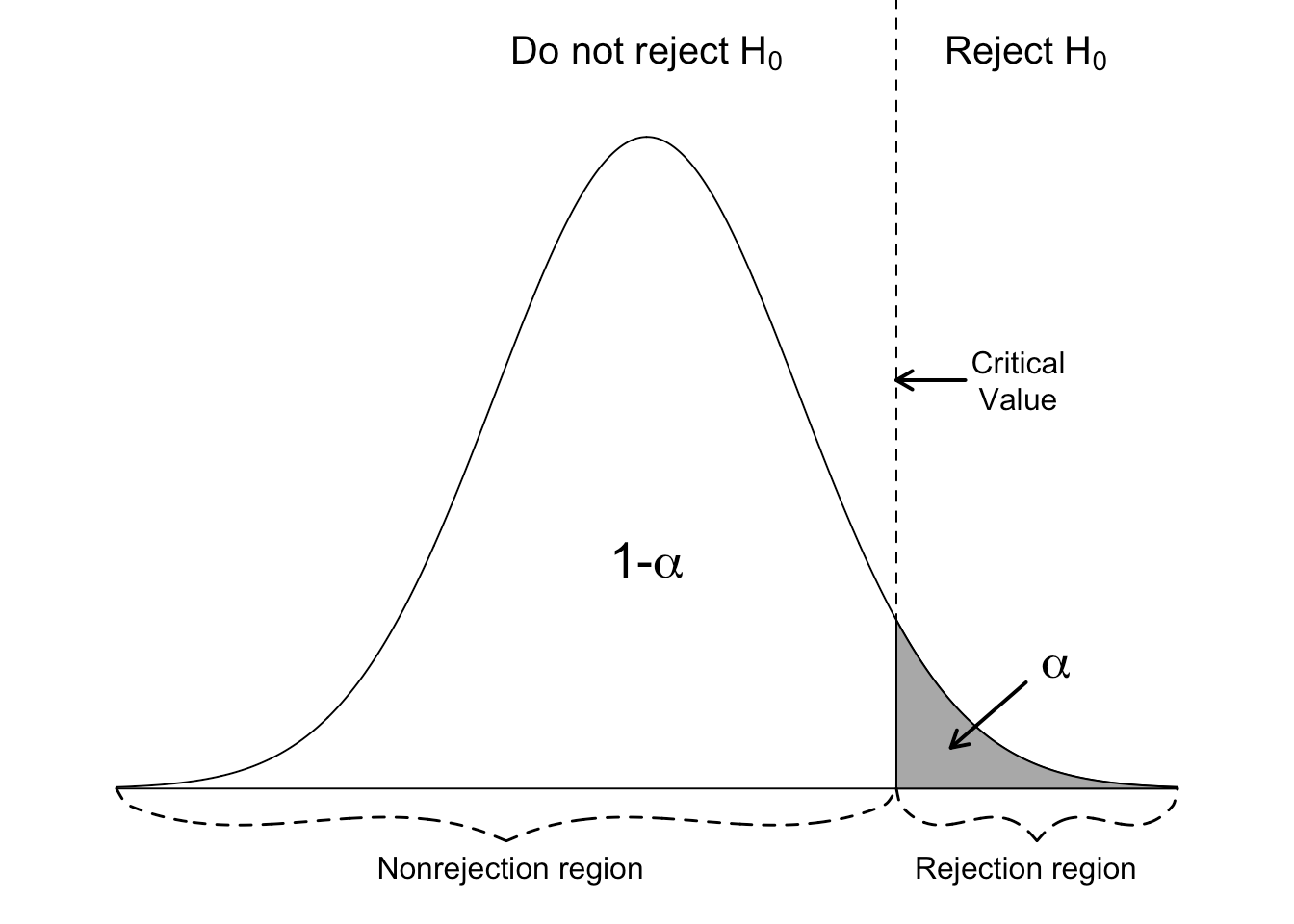
In Step 2, we stated the criterion or probability of obtaining a sample mean at which point we will decide to reject the value stated in the null hypothesis(α), which is typically set at 5% in behavioural research. To make a decision, we compare the *p*-value to the α we set in Step 2.

When the *p*-value is less than or equal to 5% (*p* α), we reject the null hypothesis. When the p value is greater than 5% (*p* > α), we retain the null hypothesis. The decision to reject or retain the null hypothesis is called significance. When the *p*-value is less than or equal to α, we reach significance; the decision is to reject the null hypothesis. When the *p*-value is greater than α, we fail to reach significance; the decision is to retain the null hypothesis.

**2.Critical Value Approach:**

The critical value approach requires that we first determine a value for the test statistic called the **critical value**. For a lower tail test, the critical value serves as a benchmark for determining whether the value of the test statistic is small enough to reject the null hypothesis. It is the value of the test statistic that corresponds to an area of *α* (the level of significance) in the lower tail of the sampling distribution of the test statistic. In other words, the critical value is the largest value of the test statistic that will result in the rejection of the null hypothesis.

In a two­tailed test, we divide the alpha value in half so that an equal proportion of area is placed in the upper and lower tail. The **rejection region** is the region beyond a critical value in a hypothesis test. When the value of a test statistic is in the rejection region, we decide to reject the null hypothesis; otherwise, we retain the null hypothesis.

****

**DECISION: RETAIN THE NULL HYPOTHESIS**

When we decide to retain the null hypothesis, we can be correct or incorrect. The correct decision is to retain a true null hypothesis. This decision is called a null result or null finding. This is usually an uninteresting decision because the decision is to retain what we already assumed: that the value stated in the null hypothesis is correct. For this reason, null results alone are rarely published in behavioural research.

The incorrect decision is to retain a false null hypothesis. This decision is an example of a **Type II error.** With each test we make, there is always some probability that the decision could be a Type II error. In this decision, we decide to retain previous notions of truth that are in fact false. While it’s an error, we still did nothing; we retained the null hypothesis. We can always go back and conduct more studies.

**Type II error**, or **beta** , is the probability of retaining a null hypothesis that is actually false.

**DECISION: REJECT THE NULL HYPOTHESIS**

When we decide to reject the null hypothesis, we can be correct or incorrect. The incorrect decision is to reject a true null hypothesis. This decision is an example of a **Type I error.** With each test we make, there is always some probability that our decision is a Type I error. A researcher who makes this error decides to reject previous notions of truth that are in fact true. Making this type of error is analogous to finding an innocent person guilty. To minimize this error, we assume a defendant is innocent when beginning a trial. Similarly, to minimize making a Type I error, we assume the null hypothesis is true when beginning a hypothesis test.

**Type I error** is the probability of rejecting a null hypothesis that is actually true. Researchers directly control for the probability of committing this type of error.

An **alpha level** is the level of significance or criterion for a hypothesis test. It is the largest probability of committing a Type I error that we will allow and still decide to reject the null hypothesis.

Since we assume the null hypothesis is true, we control for Type I error by stating a level of significance. The level we set, called the **alpha level** (symbolized as *α*), is the largest probability of committing a Type I error that we will allow and still decide to reject the null hypothesis. This criterion is usually set at 0.05 (*α* = 0.05), and we compare the alpha level to the *p* value. When the probability of a Type I error is less than 5% (*p* 0.05), we decide to reject the null hypothesis; otherwise, we retain the null hypothesis.

**Significance Level:** In practice, the person responsible for the hypothesis test specifies the level of significance. By selecting α, that person is controlling the probability of making a Type I error. If the cost of making a Type I error is high, small values of α are preferred. If the cost of making a Type I error is not too high, larger values of α are typically used. Applications of hypothesis testing that only control for the Type I error are called significance tests. Many applications of hypothesis testing are of this type.

Although most applications of hypothesis testing control for the probability of making a Type I error, they do not always control for the probability of making a Type II error. Hence, if we decide to accept H0, we cannot determine how confident we can be with that decision. Because of the uncertainty associated with making a Type II error when conducting significance tests, statisticians usually recommend that we use the statement “do not reject H0” instead of “accept H0.” Using the statement “do not reject H0” carries the recommendation to withhold both judgment and action. In effect, by not directly accepting H0, the statistician avoids the risk of making a Type II error. Whenever the probability of making a Type II error has not been determined and controlled, we will not make the statement “accept H0.” In such cases, only two conclusions are possible: do not reject H0 or reject H0.

Although controlling for a Type II error in hypothesis testing is not common, it can be done. If proper controls have been established for this error, action based on the “accept H0” conclusion can be appropriate.

If the sample data are consistent with the null hypothesis H0, we will follow the practice of concluding “do not reject H0.” This conclusion is preferred over “accept H0,” because the conclusion to accept H0 puts us at risk of making a Type II error.

**Deciding what significance level to use:**

The choice of significance level is based on the consequences of Type I and Type II errors.

If the consequences of a type I error are serious or expensive, then a very small significance level is appropriate.

Example: Two drugs are being compared for effectiveness in treating the same condition. Drug 1 is very affordable, but Drug 2 is extremely expensive.  The null hypothesis is "both drugs are equally effective," and the alternate is "Drug 2 is more effective than Drug 1." In this situation, a Type I error would be deciding that Drug 2 is more effective, when in fact it is no better than Drug 1, but would cost the patient much more money. That would be undesirable from the patient's perspective, so a small significance level is warranted.

If the consequences of a Type I error are not very serious (and especially if a Type II error has serious consequences), then a larger significance level is appropriate.

Example: Two drugs are known to be equally effective for a certain condition. They are also each equally affordable. However, there is some suspicion that Drug 2 causes a serious side-effect in some patients, whereas Drug 1 has been used for decades with no reports of the side effect. The null hypothesis is "the incidence of the side effect in both drugs is the same", and the alternate is "the incidence of the side effect in Drug 2 is greater than that in Drug 1." Falsely rejecting the null hypothesis when it is in fact true (Type I error) would have no great consequences for the consumer, but a Type II error (i.e., failing to reject the null hypothesis when in fact the alternate is true, which would result in deciding that Drug 2 is no more harmful than Drug 1 when it is in fact more harmful) could have serious consequences from a public health standpoint. So setting a large significance level is appropriate.

**Power of the test:**

The correct decision of hypothesis testing is to reject a false null hypothesis. There is always some probability that we decide that the null hypothesis is false when it is indeed false. This decision is called the **power** of the decision making process. It is called power because it is the decision we aim for. Remember that we are only testing the null hypothesis because we think it is wrong. Deciding to reject a false null hypothesis, then, is the power, inasmuch as we learn the most about populations when we accurately reject false notions of truth. This decision is the most published result in behavioural research.

The **power** in hypothesis testing is the probability of rejecting a false null hypothesis. Specifically, it is the probability that a randomly selected sample will show that the null hypothesis is false when the null hypothesis is indeed false.

For any particular value of μ, the power is 1 - β, that is, the probability of correctly rejecting the null hypothesis is 1 minus the probability of making Type II error.

Power curve is a graph of the probability of rejecting H0 for all possible values of the population parameter not satisfying the null hypothesis. The power curve provides the probability of correctly rejecting the null hypothesis.

