Drawing and testing assumptions

Hypothesis testing 101

What is a hypothesis?

> A hypothesis is an **assumption** about a particular phenomenon or a relationship between variables. The hypothesis is what we are testing **explicitly** while the assumption is being tested **implicitly**.

We restate a hypothesis as a **null** and an **alternative hypothesis** to make sure that it is **testable** and **falsifiable**:

H_0 Null hypothesis

- An assumption of no effect, no difference, or no relationship.
- The assumption **we are trying to reject**.
- Considered to be the opposite of the result we are hypothesising or the absence of it.

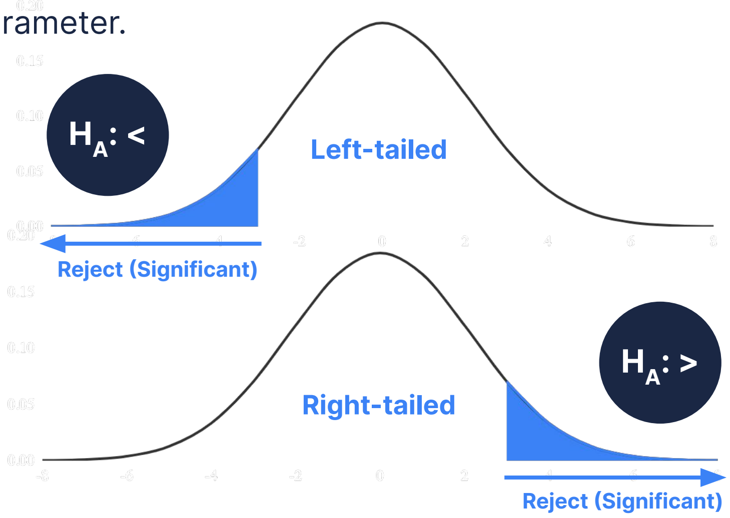
Alternative hypothesis H_A

- An assumption of true effect, difference, or relationship.
- An assumption that **contradicts the null** hypothesis.

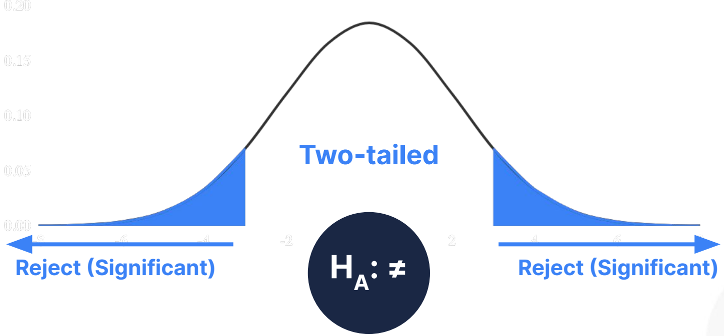
One-tailed versus two-tailed tests

> The **possibility of an effect in a specific direction or not** in hypothesis testing indicates whether we are considering a one-tailed or two-tailed hypothesis, i.e. is the value bigger or smaller or not equal.

One-tailed tests look for an **increase** or **decrease** in a parameter.



Two-tailed tests look for **change** in a parameter.



One-tailed versus two-tailed tests

Example

Suppose we want to test the poverty rates between two countries, namely Country A and Country B.

For both a **one-tailed** and **two-tailed** test our null hypothesis would be:

H_0: The rates of poverty between Country A and Country B are the same.

If we are interested in knowing whether the one country has a higher poverty rate than the other, we would perform a **one-tailed test**.

The alternative hypothesis is:

H_A: The rate of poverty in Country A is higher than that of Country B, i.e.

$$H_A: A_{Poverty} > B_{Poverty}$$

If we are interested in knowing whether the one country has a higher *OR* lower poverty rate than the other, we would perform a **two-tailed test**.

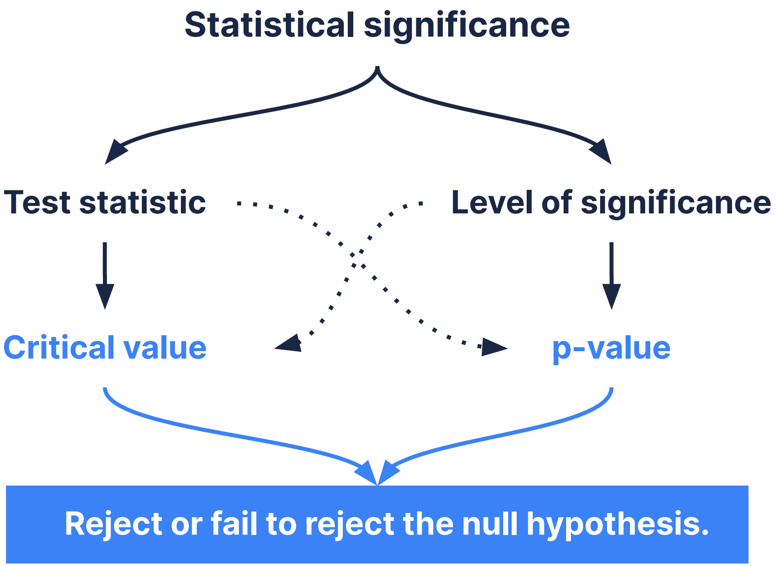
The alternative hypothesis is:

H_A: There is a difference in poverty rates between Country A and Country B, i.e.

$$H_A: A_{Poverty} \neq B_{Poverty}$$

Hypothesis testing 101

In hypothesis testing, there are several important concepts that are used to make decisions about the **statistical significance of a test**.



The **critical value** and the **p-value** are two **different ways of determining statistical significance** in hypothesis testing.

The test statistic is used to determine the p-value and the level of significance the critical value.

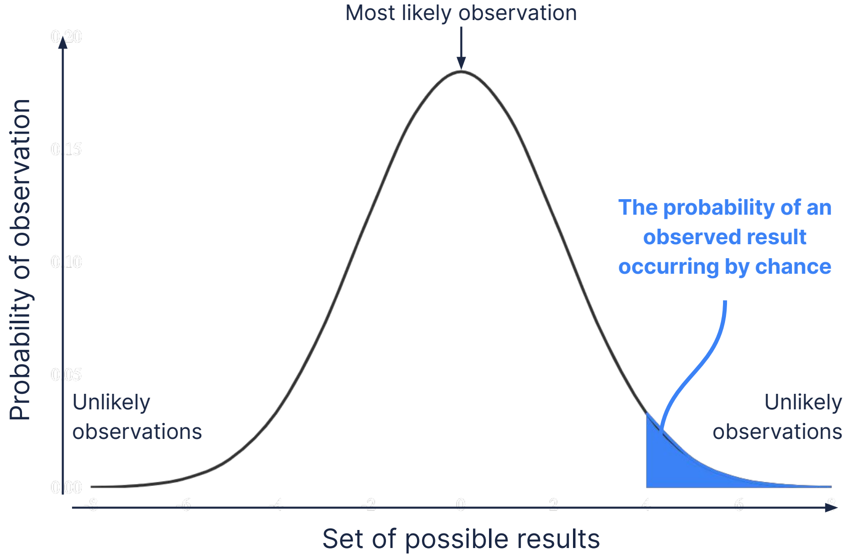
So, the critical value and the p-value provide complementary information about the statistical significance of the results.

What is statistical significance?

> **Statistical significance** describes the **likelihood** that the results observed in a sample are **unlikely to have occurred by chance**.

In other words, it is a measure of the probability that the observed effect in the sample is **not due to random variation**, but rather reflects a **true effect** in the population.

We determine statistical significance by either comparing the **level of significance** to a **probability value** or comparing the **critical value** to a **test statistic**.



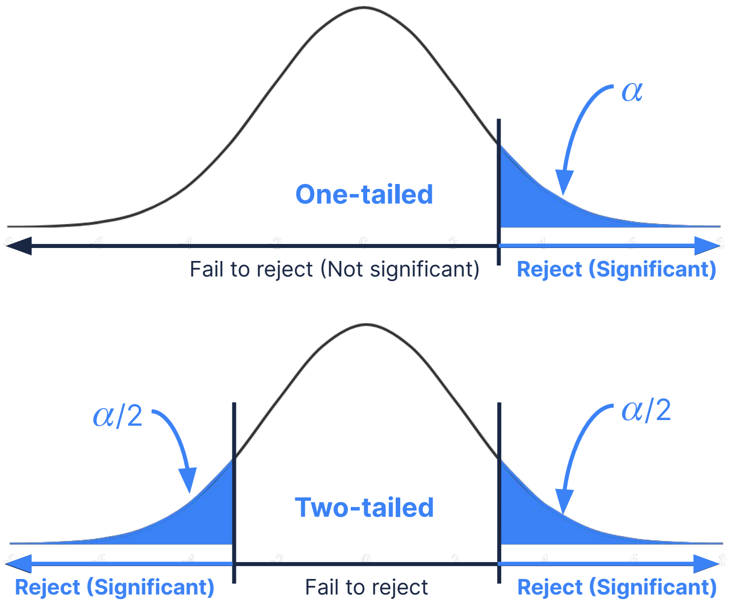
Level of significance

> The **level of significance**, denoted by alpha (α), is a predetermined threshold for deciding whether to reject or fail to reject the null hypothesis.

We could also say that the **level of significance** is the probability of rejecting the null hypothesis when it is actually true. In other words, it is the **threshold** at which a **result is considered statistically significant**.

It is typically set before conducting a hypothesis test to 0.05, or **5%**. This means we have a 95% threshold of significance.

It also means that we are **willing to accept a 5% chance of rejecting** the null hypothesis when it is actually **true**.



Test statistic

> A test statistic is a numerical value calculated in a hypothesis test used to determine whether the observed data **provides evidence for or against** a hypothesis.

A test statistic is often **calculated from a sample** of data that is used to make **inferences** about an **unknown population parameter**.

For example, the z-score is commonly used in hypothesis testing to determine whether a sample mean is significantly different from a hypothesised population mean.

We compare a test statistic to a **critical value** to determine whether to reject or fail to reject the null hypothesis.

How it is calculated in hypothesis testing **depends on the type of hypothesis test** being performed.

The most common test statistics include the **t-score**, **z-score**, and **chi-square statistic**.

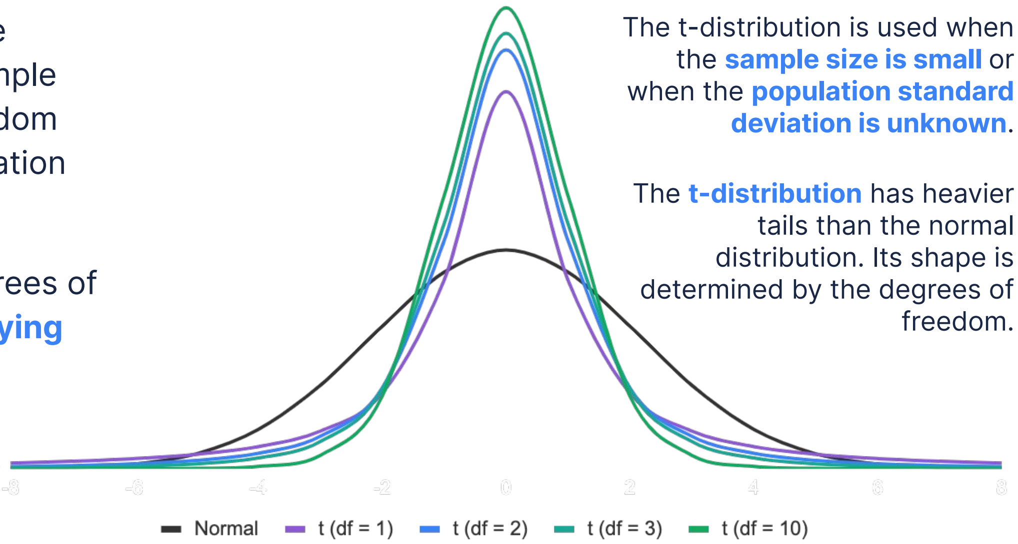
Degrees of freedom

> Degrees of freedom (df) refers to the number of independent observations in a sample that can vary and influences the statistical significance of a test.

The degrees of freedom are calculated from the **sample size**. This means that the larger the sample size, the greater the number of degrees of freedom and the more precise the estimate of the population parameter.

In other words, the sample size affects the degrees of freedom which defines the shape of the **underlying** hypothesis test **t-distribution**.

df = sample size - 1*



The t-distribution is used when the **sample size is small** or when the **population standard deviation is unknown**.

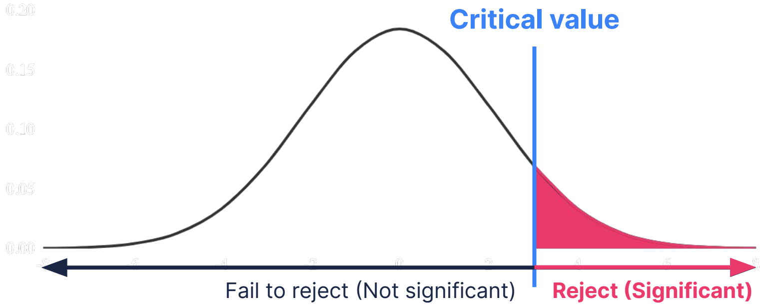
The **t-distribution** has heavier tails than the normal distribution. Its shape is determined by the degrees of freedom.

*When our test includes more than one sample or relationships, then df = (sample size) - (number of different parameters)

The critical value

> The critical value is a threshold that is used to **compare with the absolute value of the test statistic** to determine whether to reject or fail to reject the null hypothesis.

The **critical value** is determined based on the level of significance (α) and degrees of freedom (df).



critical value ≤ |test statistic|: the observed data is unlikely to have occurred by chance and therefore provides evidence to **reject the null** hypothesis.

critical value > |test statistic|: we **fail to reject the null** hypothesis since we have insufficient evidence to reject it.

In practice, we **determine the critical value** using tables of probability distributions or using software.

The p-value

> The p-value is the **probability value**, which is the likelihood of an outcome resulting from randomness rather than true effect.

In hypothesis testing, we often use the **p-value** (probability value) to determine the **strength of evidence against the null** hypothesis.

It represents the probability of observing a **test statistic** as **extreme** (unlikely), or more extreme than the one determined from the sample data.

The **type of hypothesis test** we perform and the **distribution** of the test statistic will determine how we **calculate the p-value**.

p-value ≤ α: the observed data is unlikely to have occurred by chance and therefore provides evidence to **reject the null** hypothesis.

p-value > α: we **fail to reject the null** hypothesis since we have insufficient evidence to reject it.

In practice, we **determine the p-value** using tables of probability distributions or using software.

Steps to the critical value and p-value

Critical value

01. Determine the level of significance (α) for the hypothesis test.
02. Determine the degrees of freedom (df) based on the sample size.
03. Look up the critical value from a table of probability distributions based on α and df.
04. Calculate the test statistic for the sample data.
05. Compare the test statistic to the critical value to determine if the results are statistically significant.

critical value ≤ |test statistic|: reject the null
critical value > |test statistic|: fail to reject the null

p-value

01. Determine the level of significance (α) for the hypothesis test.
02. Determine the degrees of freedom (df) based on the sample size.
03. Calculate the test statistic from the sample data using the appropriate formula based on the chosen test.
04. Determine the p-value from a probability table.
05. Compare the p-value to the level of significance to determine if the results are statistically significant.

p-value ≤ α: reject the null
p-value > α: fail to reject the null