Reference card

Hypothesis testing

Formulating hypotheses

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**.

H_0 Null hypothesis

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

H_A Alternative hypothesis

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

Formulating hypotheses

In hypothesis testing, there are several important concepts that are used to make decisions about the **statistical significance of a test**, i.e. the **likelihood** that the results observed in a sample are **unlikely to have occurred by chance**.

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

Statistical significance

Test statistic — **Critical value**

1. Determine the **level of significance (α)** for the hypothesis test.
2. Determine the **degrees of freedom (df)** based on the sample size.
3. Look up the **critical value** from a table of probability distributions based on α and df.
4. Calculate the **test statistic** for the sample data.
5. Compare the **test statistic** to the **critical value** to determine if the results are statistically significant.

|test statistic| ≥ critical value: reject the null

|test statistic| < critical value: fail to reject the null

Level of significance (α) — **p-value**

1. Determine the **level of significance (α)** for the hypothesis test.
2. Determine the **degrees of freedom (df)** based on the sample size.
3. Calculate the **test statistic** from the sample data using the appropriate formula based on the chosen test.
4. Determine the **p-value** from a probability table.
5. 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

We can either use a **parametric** or **non-parametric** test to calculate the **test statistic** we need for both the **critical value** and **p-value**.

Parametric tests

Assume that the data follow a **specific distribution.**

Depending on the test, assumptions include: random sampling, normality, independence, and homoscedasticity.

Tests include: t-test, z-test, f-test, ANOVA (Analysis of Variance).

Non-parametric tests

No assumptions about the distribution of the data.

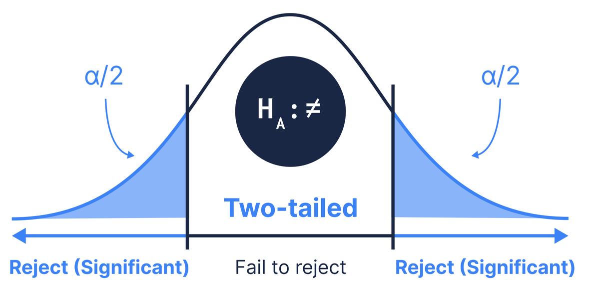
Often preferred when the **sample size is small.**

Tests include: Kolmogorov-Smirnov test, Mann-Whitney U-test, Chi-square, Spearman's rank correlation coefficient, Wilcoxon signed rank test, Friedman test, Kruskal-Wallis H test.

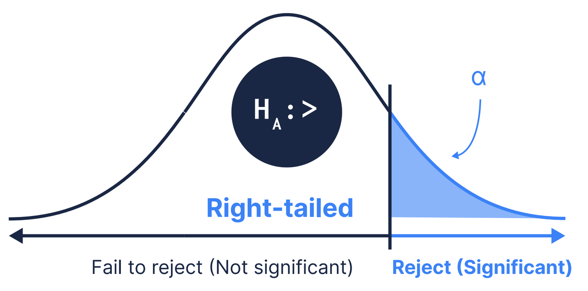
One-tailed versus two-tailed tests

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.

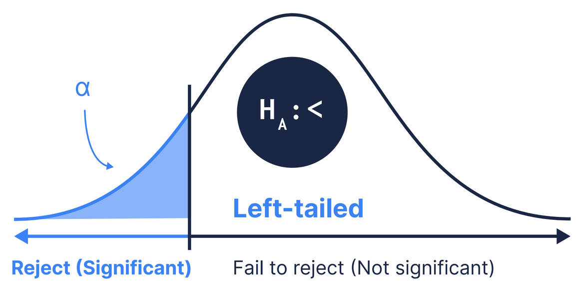
Two-tailed tests look for **change** in a parameter.



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



The **number of tails influences** the **level of significance** (α) and therefore **how we reject or fail to reject the null** since the **critical value** size and **p-value comparison** depend on α.

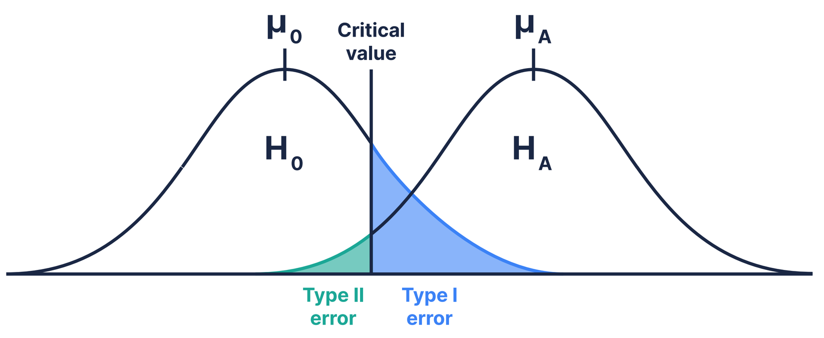


Errors and estimates

Errors and estimates are important because they allow us to **properly interpret** the results of a **hypothesis test** and draw **accurate conclusions** about the **population**.

The uncertainty resulting from the estimates we use means that there is a **chance of making an incorrect decision** in our hypothesis tests. These incorrect decisions are called **type I** and **type II** errors.

Decision \ Truth	H_0 is true	H_0 is false
Fail to reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision



$H_0: \mu = \mu_0$

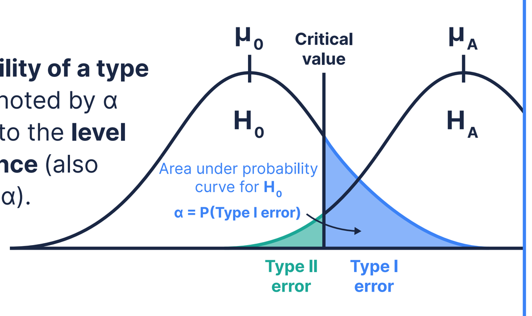
$H_A: \mu \neq \mu_0$

where **μ** is the population mean, μ_0 the hypothesised mean, and μ_A the alternative mean.

Type I error

A type I error occurs when the null hypothesis is **rejected when it is actually true**. It is a false positive error.

The **probability of a type I error** is denoted by α and relates to the **level of significance** (also denoted by α).

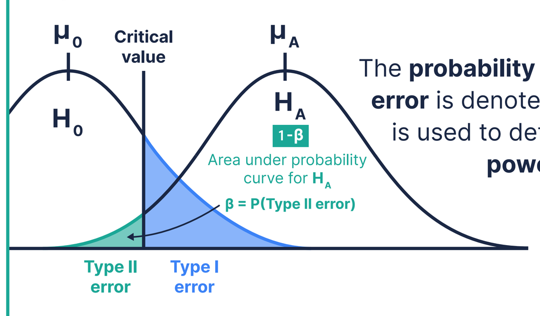


To **decrease the probability of a type I error**, we need to **decrease the level of significance**, but changing the sample size has no effect on the probability of a type I error.

Type II error

A type II error occurs when the null hypothesis is **accepted when it is actually false**. It is a false negative error.

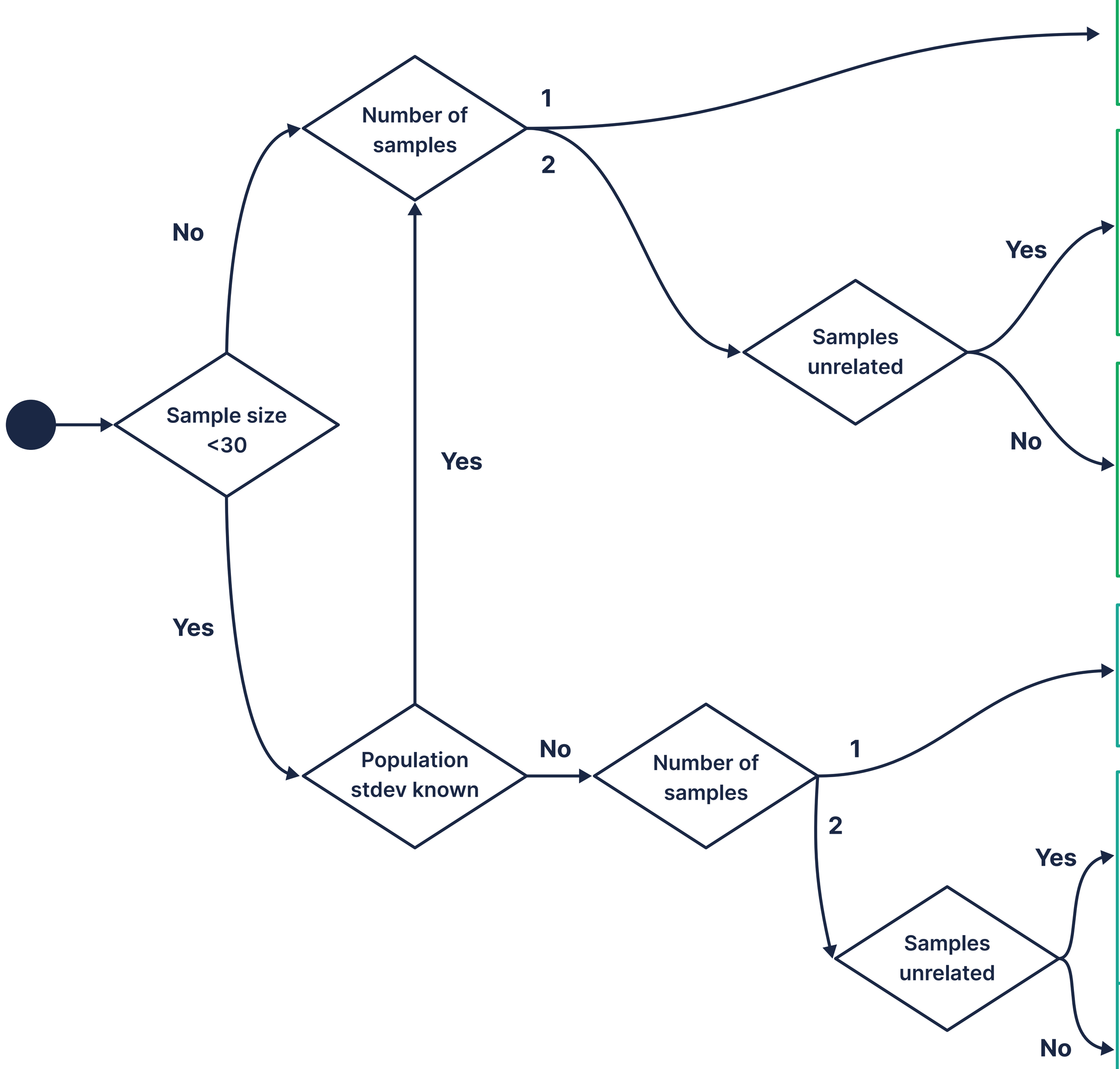
The **probability of a type II error** is denoted by **β** and is used to determine the **power** of a test.



To **decrease the chances of a type II error**, we can either take a **larger sample** or we can **increase the power** by increasing the level of significance. However, if we do the second, we **increase the probability of a type I error**.

Parametric: t-test and z-test

If our data follow a **normal distribution**, then:



|z-score| ≥ critical value: reject the null

|z-score| < critical value: fail to reject the null

The **z-test** is a parametric test based on the **normal distribution** (a.k.a. the z-distribution) and is similar to the t-test. However, the z-test is used when the **sample is large** and the **population standard deviation is known**.

p-value for z-test in Google Sheets

```
=Z.TEST(data, value, [standard_deviation])
```

Test	Formula	Definitions	Assumptions
One-sample z-test	$z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$	$\bar{x}$ is the sample mean; σ is the population standard deviation; n is the sample size; μ is the population mean	**Assumptions:** random sampling, normality, independence, large sample size, known population standard deviation
Two-sample independent z-test	$z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$\bar{x}_1$ is sample one's mean; $\bar{x}_2$ is sample two's mean; σ_1 is population one's standard deviation; σ_2 is population two's standard deviation; n_1 is sample one's size; n_2 is sample two's size	**Assumptions:** normality, independence, homoscedasticity, large sample size, known population standard deviation
Two-sample paired z-test	$z = \dfrac{\bar{d} - D}{\sqrt{\dfrac{\sigma^2}{n}}}$	$\bar{d}$ is the mean of the differences between the samples; D is the hypothesised mean of the differences (usually equal to zero); σ is the standard deviation of the differences; n is the sample size	**Assumptions:** normality, independence, homoscedasticity, large sample size, known population standard deviation
One-sample t-test	$t = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$	$\bar{x}$ is the sample mean; s is the sample standard deviation; n is the sample size; μ is the population mean	**Assumptions:** random sampling, normality, independence, homoscedasticity
Two-sample independent t-test	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	$\bar{x}_1$ is sample one's mean; $\bar{x}_2$ is sample two's mean; s is the pooled standard deviation; n_1 is sample one's size; n_2 is sample two's size. The pooled standard deviation is: $s = \sqrt{\dfrac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{n_1 + n_2 - 2}}$; df = $n_1 + n_2 - 2$	**Assumptions:** normality, independence, homoscedasticity
Two-sample paired t-test	$t = \dfrac{\bar{d}}{\sqrt{\dfrac{s^2}{n}}}$	$\bar{d}$ is the mean of the differences between the samples; s is the standard deviation of the differences; n is the sample size	**Assumptions:** normality, independence

|t-score| ≥ critical value: reject the null

|t-score| < critical value: fail to reject the null

The **t-test** is a parametric test based on the **t-distribution** and is used to test hypotheses on the **mean of a single population**, or the **difference between the means of two samples**, when the sample size is smaller.

p-value for t-test in Google Sheets

```
=T.TEST(range1, range2, tails, type)
```

Non-parametric: Kolmogorov-Smirnov (KS)

Kolmogorov-Smirnov (KS) is a non-parametric test based on the **empirical cumulative distribution function (ECDF)**, which is a way to visually represent how data are **distributed**.

Steps to performing KS:

1. State the **null** and **alternative** hypotheses:
   a. H_0 is that the sample is drawn from a population with a specific distribution, e.g. a normal distribution.
   b. H_A is that the sample is not drawn from a population with the specified distribution.
2. Specify the **level of significance** (α).
3. Calculate the **test statistic**, D, using the Kolmogorov-Smirnov test statistic formula.
4. Determine the **critical value** using the KS table, level of significance, and sample size.
5. Compare the test statistic (D) to the critical value.

|D| ≥ critical value: reject the null

|D| < critical value: fail to reject the null

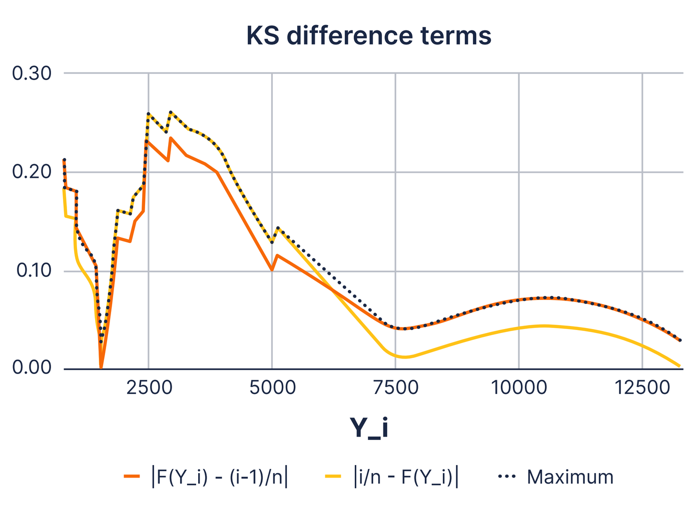
The Kolmogorov-Smirnov test statistic:

$$D = \max_{1 \leq i \leq n} \left(\left| F(Y_i) - \frac{i-1}{n} \right|, \left| \frac{i}{n} - F(Y_i) \right| \right)$$

where

- i is the index of the ordered sample $Y_1, Y_2, \ldots, Y_n$, i.e. the rank
- n is the sample size
- Y_i is the ith ordered value in the sample
- $F(Y_i)$ is the hypothesised cumulative distribution function (CDF) evaluated at the ith ordered value of the sample data Y_i

Both **(1-1)/n** and **i/n** represent the empirical cumulative distribution function (**ECDF**)*.



The test statistic D is a single value which is the maximum across both difference terms, |F(Yi) - (i-1)/n| and |i/n - F(Yi)|, for all sample values, Yi.

*(i-1)/n represents the cumulative proportion observations that are expected to be strictly less than the ith ordered value, while i/n represents the proportion that is less than or equal to the ith.

The Kolmogorov-Smirnov test in Google Sheets:

Sorted average (Y_i)

```
=SORT(range, sort_column, is_ascending)
=SORT(the_data, the_data, TRUE)
```

Index (i)

```
=RANK(value, data, [is_ascending])
=RANK(Y_i, Y_i_range, TRUE)
```

CDF hypothesised (F(Y_i))

(based on the hypothesised distribution)

```
=NORM.DIST(x, mean, standard_deviation, cumulative)
=NORM.DIST(Y_i, sample_mean, sample_standard_deviation, TRUE)
```

i/n (ECDF)

```
=i/sample_size
```

(i-1)/n (ECDF)

```
=(i-1)/sample_size
```

|i/n - F(Y_i)|

```
=ABS(i/n (ECDF) - F(Y_i))
```

(i-1)/n (ECDF)

```
=ABS((i-1)/n (ECDF) - F(Y_i))
```

```
D = MAX(range_of_both_|i/n - F(Y_i)|_and_(i-1)/n(ECDF))
```

EXPLORE AI ACADEMY