# How to choose the suitable statistical test

## When to perform a statistical test:

You can perform statistical tests on data that have been collected in a statistically valid manner – either through an experiment, or through observations made using probability sampling methods.

For a statistical test to be valid, your sample size needs to be large enough to approximate the true distribution of the population being studied.

To determine which statistical test to use, you need to know:

- Whether your data meets certain assumptions.
- The types of variables that you're dealing with.

## Statistical assumptions:

Statistical tests make some common assumptions about the data they are testing:

1. **Independence of observations (a.k.a. no autocorrelation):** The observations/variables you include in your test are not related (for example, multiple measurements of a single test subject are not independent, while measurements of multiple different test subjects are independent).
2. **Homogeneity of variance:** the variance within each group being compared is similar among all groups. If one group has much more variation than others, it will limit the test's effectiveness.
3. **Normality of data:** the data follows a normal distribution (a.k.a. a bell curve). This assumption applies only to quantitative data.

## Choosing a parametric test: regression, comparison, or correlation:

Parametric tests usually have stricter requirements than nonparametric tests, and are able to make stronger inferences from the data. They can only be conducted with data that adheres to the common assumptions of statistical tests.

The most common types of parametric test include regression tests, comparison tests, and correlation tests.

**Regression tests:**

Regression tests look for cause-and-effect relationships. They can be used to estimate the effect of one or more continuous variables on another variable.

**Comparison tests:**

Comparison tests look for differences among group means. They can be used to test the effect of a categorical variable on the mean value of some other characteristic.

T-tests are used when comparing the means of precisely two groups (e.g., the average heights of men and women). ANOVA and MANOVA tests are used when comparing the means of more than two groups (e.g., the average heights of children, teenagers, and adults).

**Correlation tests:**

Correlation tests check whether variables are related without hypothesizing a cause-and-effect relationship.

These can be used to test whether two variables you want to use in (for example) a multiple regression test are autocorrelated.


## Choosing a nonparametric test:

Non-parametric tests don't make as many assumptions about the data, and are useful when one or more of the common statistical assumptions are violated. However, the inferences they make aren't as strong as with parametric tests.

**Non-Parametric T-Test**

Whenever a few assumptions in the given population are uncertain, we use non-parametric tests, which are also considered parametric counterparts. When data are not distributed normally or when they are on an ordinal level of measurement, we have to use non-parametric tests for analysis. The basic rule is to use a parametric t-test for normally distributed data and a non-parametric test for skewed data.

**Non-Parametric Paired T-Test**

The paired sample t-test is used to match two means scores, and these scores come from the same group. Pair samples t-test is used when variables are independent and have two levels, and those levels are repeated measures.