# Statistical Tests and Assumptions

Tools



Here we'll describe research questions and the corresponding statistical tests, as well as, the test assumptions.


# Research questions and corresponding statistical tests

The most popular research questions include:

1. whether **two variables** (n = 2) are **correlated** (i.e., associated)
2. whether **multiple variables** (n > 2) are **correlated**
3. whether **two groups** (n = 2) of samples **differ** from each other
4. whether **multiple groups** (n >= 2) of samples **differ** from each other
5. whether the **variability** of two samples differ

Each of these questions can be answered using the following statistical tests:

1. **Correlation test** between two variables
2. **Correlation matrix** between multiple variables
3. **Comparing the means of two groups**:
   - **Student's t-test** (parametric)
   - **Wilcoxon rank test** (non-parametric)

4. **Comparing the means of more than two groups**
    - **ANOVA test** (analysis of variance, parametric): extension of t-test to compare more than two groups.
    - **Kruskal-Wallis rank sum test** (non-parametric): extension of Wilcoxon rank test to compare more than two groups
5. **Comparing the variances**:
    - Comparing the variances of two groups: **F-test** (parametric)
    - Comparison of the variances of more than two groups: **Bartlett's test** (parametric), **Levene's test** (parametric) and **Fligner-Killeen test** (non-parametric)

# Statistical test requirements (assumptions)

Many of the statistical procedures including correlation, regression, t-test, and analysis of variance assume some certain characteristic about the data. Generally they assume that:

- the data are **normally distributed**
- and the **variances** of the groups to be compared are **homogeneous** (equal).

> These assumptions should be taken seriously to draw reliable interpretation and conclusions of the research.

> These tests - correlation, t-test and ANOVA - are called **parametric tests**, because their validity depends on the distribution of the data.

Before using parametric test, we should perform some **preleminary tests** to make sure that the test assumptions are met. In the situations where the assumptions are violated, **non-paramatric** tests are recommended.

# How to assess the normality of the data?

1. With **large enough sample sizes** (n > 30) the violation of the normality assumption should not cause major problems (central limit theorem). This implies that we can ignore the distribution of the data and use parametric tests.
2. However, to be consistent, we can use **Shapiro-Wilk's significance test** comparing the sample distribution to a normal one in order to ascertain whether data show or not a serious deviation from normality.

# How to assess the equality of variances?

The standard **Student's t-test** (comparing two independent samples) and the ANOVA test (comparing multiple samples) assume also that the samples to be compared have equal variances.

If the samples, being compared, follow normal distribution, then it's possible to use:

- **F-test** to compare the variances of two samples
- **Bartlett's Test** or **Levene's Test** to compare the variances of multiple samples.

# Infos

This analysis has been performed using **R software** (ver. 3.2.4).

Enjoyed this article? I'd be very grateful if you'd help it spread by emailing it to a friend, or sharing it on Twitter, Facebook or Linked In.

Show me some love with the like buttons below... Thank you and please don't forget to share and comment below!!