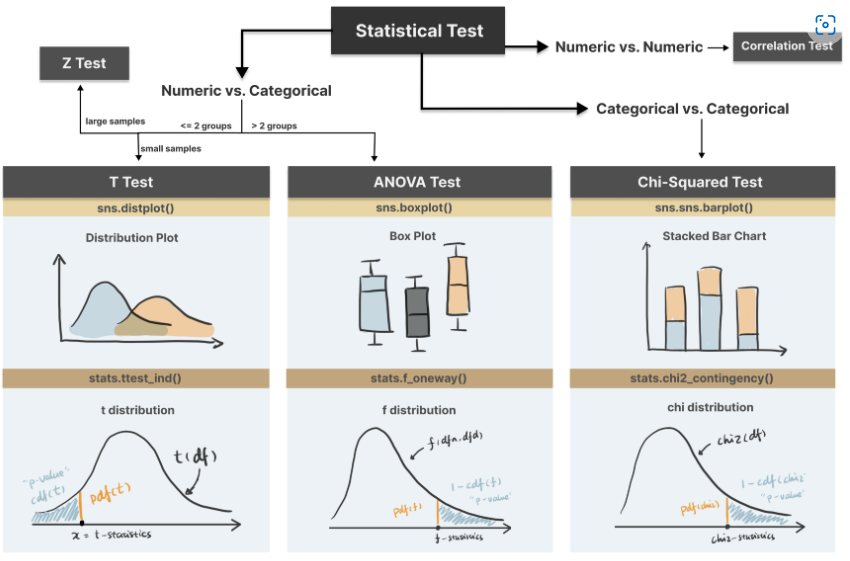
<https://towardsdatascience.com/an-interactive-guide-to-hypothesis-testing-in-python-979f4d62d85>



**Hypothesis testing :** we use observed data in a sample to draw conclusions about unobserved data.

Implication of hypothesis testing:

Clinical research: widely used in psychology, biology and healthcare research to examine the effectiveness of clinical trials

A/B testing: can be applied in business context to improve conversions through testing different versions of campaign incentives, website designs etc

Feature selection in machine learning: filter-based feature selection methods use different statistical tests to determine the feature importance

College or university: well, if you major in statistics or data science, it is likely to appear in your **exams**

4 Steps in Hypothesis testing

Step 1. Define null and alternative hypothesis

**Null hypothesis (H0):** There is no difference, no relationship or no dependency exists between two or more variables.

**Alternative hypothesis (H1**) is contradictory to the null hypothesis and it claims that relationships exist.

**Step 2. Choose the appropriate statistical test**

**T-test:** compare two groups/categories of numeric variables with small sample size

**Z-test:** compare two groups/categories of numeric variables with large sample size

**ANOVA test:** compare the difference between two or more groups/categories of numeric variables

**Chi-Squared test**: examine the relationship between two categorical variables

**Correlation test:** examine the relationship between two numeric variables

**z-test:** test the difference between two categories of numeric variables — when sample size is LARGE

**correlation**: test the relationship between two numeric variables

<https://towardsdatascience.com/multi-collinearity-in-regression-fe7a2c1467ea>

Multi-Collinearity : independent variables are highly correlated, change in one variable would cause change to another

**First simple** method is to plot the correlation matrix of all the independent variables

**The second method** to check multi-collinearity is to use the Variance Inflation Factor(VIF) for each independent variable

* If the VIF value is higher than 10, it is usually considered to have a high correlation with other independent variables

**How to fix the Multi-Collinearity issue?**

1. Variable Selection

2. Variable Transformation

3. Principal Component Analysis