



ANOVA, T-test and other statistical tests with Python



Francesco Dallanoce · Follow

Published in Towards Data Science · 6 min read · Aug 18, 2021



752



3



Analysis of the main statistical tests (ANOVA, T-test, MANOVA etc.) and their characteristics, applying them in Python.



GIF by Crazy Ex-GIFs on gyrfy.com

Has your boss asked you to do some statistical tests and you don't know where to start? 🤔

Are you already going crazy thinking that you have to read tons of statistics books that you have been able to avoid until now?

This is a quick guide that will help you with the basics, choose which test is right for you, and how to implement them in Python with examples and libraries. 😎

What is a Statistical Test

Open in app ↗

Medium

🔍 Search

✍ Write



- determine whether an input variable has a statistically significant relationship with an output (target) variable.
- estimate the difference between two or more groups.

The idea of how a statistical test works is quite simple: they assume that a particular situation occurs, and then they estimate how likely this assumption is to be false.

Statistical tests assume a null hypothesis. The null hypothesis corresponds to the hypothesis that there is no relationship or difference between the groups (sets) considered. Then they determine whether the observed data come outside or inside of the range of values predicted by the null hypothesis.

There are **multiple statistical tests**, which vary according to the problem to be faced and the characteristics of the data available. Now we're going to go through them in detail, and then apply some of them in **Python**.

How a Statistical Test Works

The output value of a statistical test is defined (with enormous imagination 😏) a test statistic — a number that describes how much the relationship between input and output variables in your test differs from the null hypothesis.

But this is not the result that interests us.

What allows us to understand whether the null hypothesis is true or not is the p-value (probability value). The *p*-value estimates how plausible it is that you would see the difference reported by the test statistic if the null hypothesis were true.

To say whether the *p*-value is significant or not, we need a significance threshold called the significance level. This threshold is usually set at 0.05, but we won't get into this boring mathematician discussion. 😊

If the *p*-value is **BELOW** the threshold (meaning smaller than), then you can infer a **statistically significant relationship** between the input and target variables.

Otherwise, then you can infer **no statistically significant relationship** between the predictor and outcome variables.

Let's take a quick example. I apply T-test between two groups. If the T-test's corresponding *p*-value is .03, then a statistically significant relationship would be implied. There is only a 3% probability the null hypothesis is correct

(and the results are random). Therefore, we reject the null hypothesis, and accept the alternative hypothesis.

Decide which Test is right for us

As we have said, there are many statistical tests, and each of them can only be applied in specific scenarios. To determine which statistical test to use, you need to know:

- the types of variables that you're dealing with (categorical, quantitative etc.)

whether your data meets certain assumptions (Independence of observations, Homogeneity of variance, Normality of data)

Statistic tests are divided into two main categories:

- parametric
- non-parametric

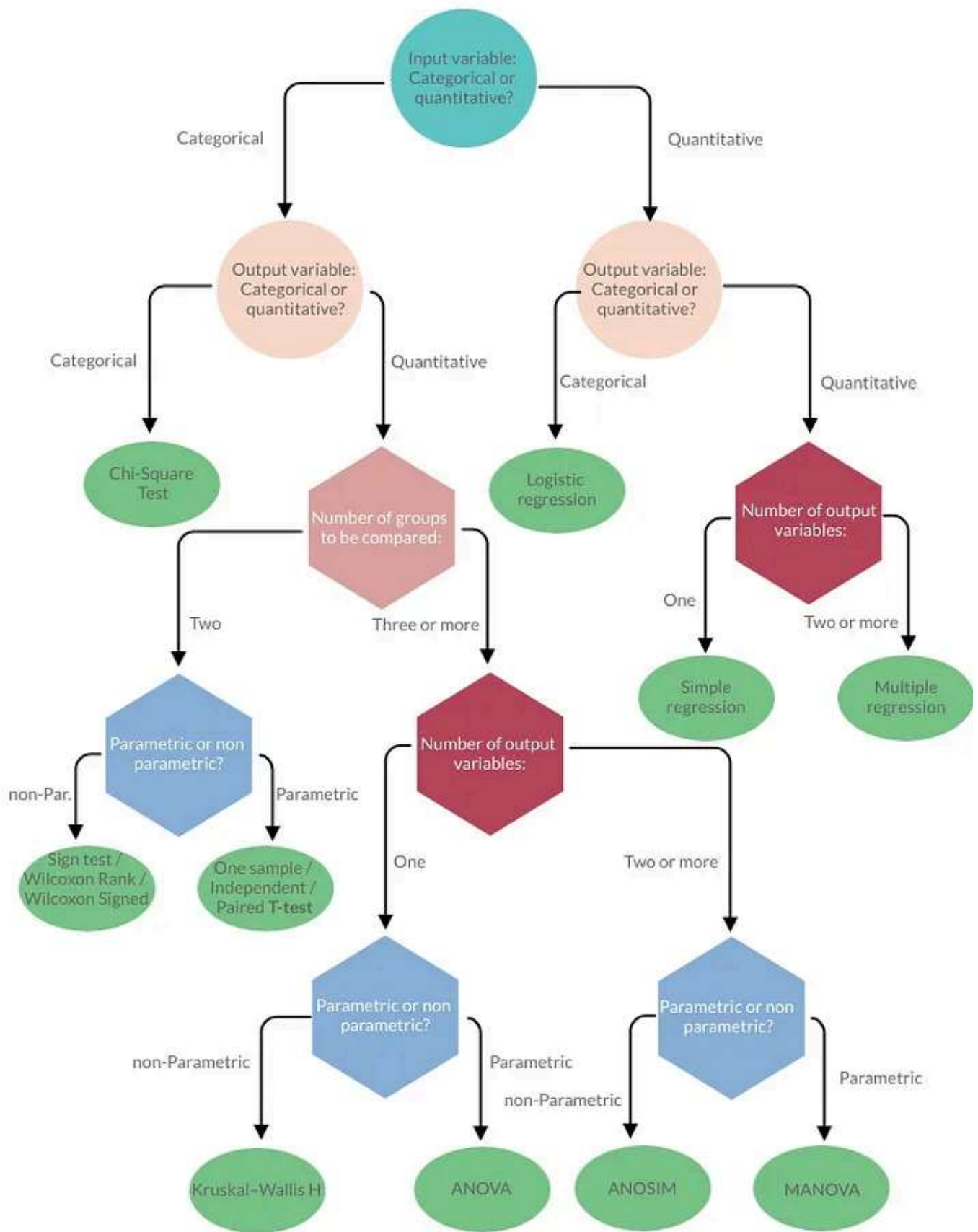
The former are the most “powerful” and are therefore those recommended to be used, but they must respect the abovementioned assumptions. Let's see them in detail:

1. **Independence of observations:** the individual observations (each entry of the variables) are independent of each other (for instance, repeating the same test on a single patient generates non-independent measurements, that is, repeated measurements).

2. **Normality of data:** the data follows a normal distribution. This assumption is required only for quantitative data. (For more details, see also here)
3. **Homogeneity of variance:** the variance (i.e., the distribution, or “spread,” of scores around the mean) within each group being compared is similar among all groups. If one group has much more variance than the others, this will reduce the “power” of the test in identifying differences.

If your data do not meet the assumption of independence of observations, you can alternatively use tests that take this situation into account (i.e., repeated-measures tests).

If your data, instead, do not satisfy the assumptions of normality or homogeneity of variance, you may be able to perform a **non-parametric statistical test**, which allows you to make comparisons without these two assumptions.



Flowchart to guide the choose of the correct Statistical Test. [Image by author]

Now that we know the criteria for choosing between the various tests, you can use the flowchart below to select the one that's right for you. Now that the **boring part is over**, we can look at some code. 🙄

P.S. There are also other statistical tests as alternatives to those proposed, which perform the same functions. To avoid making the article too long, they have been omitted.

Python libraries for statistical tests

The most famous and supported python libraries that collect the main statistical tests are:

- **Statsmodel**: a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration.
- **Pingouin**: an open-source statistical package written in Python 3 and based mostly on Pandas and NumPy.
- **Scipy**: a Python-based ecosystem of open-source software for mathematics, science, and engineering.

Testing the assumptions

As for the independence assumption, this must be known a priori by you, there is no way to extrapolate it from the data. For the other two assumptions instead: we can use **Scipy** (data can be downloaded [here](#)):


```

from scipy import stats
import pandas as pd

# import the data
df= pd.read_csv("Iris_Data.csv")
setosa = df[(df['species'] == 'Iris-setosa')]
versicolor = df[(df['species'] == 'Iris-versicolor')]

# homogeneity
stats.levene(setosa['sepal_width'], versicolor['sepal_width'])

# Shapiro-Wilk test for normality
stats.shapiro(setosa['sepal_width'])
stats.shapiro(versicolor['sepal_width'])

```

Output: *LeveneResult(statistic=0.66, pvalue=0.417)*

The test is not significant (huge p-value), meaning that there is homogeneity of variances and we can proceed.

Output: *(0.968, 0.204)(0.974, 0.337)*

Neither test for normality was significant, so neither variable violates the assumption. Both tests were successful. As for independence, we can assume it a priori knowing the data. We can proceed as planned.

T-test

To conduct the Independent t-test, we can use the *stats.ttest_ind()* method:

```
stats.ttest_ind(setosa['sepal_width'], versicolor['sepal_width'])
```

Output: *Ttest_indResult(statistic=9.282, pvalue=4.362e-15)*

The Independent t-test results are significant (p-value very very small)! Therefore, we can reject the null hypothesis in support of the alternative hypothesis.

If you want to use the non-parametric version, just replace `stats.ttest_ind` with `stats.wilcoxon`.

ANOVA

To apply ANOVA, we rely on **Pingouin**. We use a dataset included in the library:

```
import pingouin as pg

# Read an example dataset
df = pg.read_dataset('mixed_anova')

# Run the ANOVA
aov = pg.anova(data=df, dv='Scores', between='Group', detailed=True)
print(aov)
```

Output

Source	SS	DF	MS	F	p-unc	np2
Group	5.460	1	5.460	5.244	0.023	0.029
Within	185.343	178	1.041	nan	nan	nan

Result from ANOVA [from [Pingouin](#)]

As we can see we have a p-value below the threshold, so there is a significant difference between the various groups! Unfortunately, having more than two groups, we cannot know which of them is the difference. To find out, you need to apply T-test in pairs. It can be done via the method `pingouin.pairwise_ttests`.

In case you cannot ensure independency, use **repeated measure ANOVA**:

```
pg.rm_anova(data=df, dv='Scores', within='Time', subject='Subject',
detailed=True)
```

Output

Source	SS	DF	MS	F	p-unc	np2	eps
Time	7.628	2	3.814	3.913	0.023	0.062	0.999
Error	115.027	118	0.975	nan	nan	nan	nan

Result from repeated-measure ANOVA [from [Pingouin](#)]

MANOVA

In this example, we go back to using the initial dataset. We are going to use the width and length columns as dependent variables. Besides, the species column is used as the independent variable.

MANOVA is currently only implemented by the **Statsmodel** library. One of the main features of this library is that it uses **R-style formulas** to pass parameters to models.

```
from statsmodels.multivariate.manova import MANOVA
maov = MANOVA.from_formula('Sepal_Length + Sepal_Width + \
                            Petal_Length + Petal_Width ~ Species',
data=df)
print(maov.mv_test())
```

Multivariate linear model						
Intercept	Value	Num DF	Den DF	F Value	Pr > F	
Wilks' lambda	0.0170	4.0000	144.0000	2086.7720	0.0000	
Pillai's trace	0.9830	4.0000	144.0000	2086.7720	0.0000	
Hotelling-Lawley trace	57.9659	4.0000	144.0000	2086.7720	0.0000	
Roy's greatest root	57.9659	4.0000	144.0000	2086.7720	0.0000	
Species	Value	Num DF	Den DF	F Value	Pr > F	
Wilks' lambda	0.0234	8.0000	288.0000	199.1453	0.0000	
Pillai's trace	1.1919	8.0000	290.0000	53.4665	0.0000	
Hotelling-Lawley trace	32.4773	8.0000	203.4024	582.1970	0.0000	
Roy's greatest root	32.1919	4.0000	145.0000	1166.9574	0.0000	

Result from MANOVA [Image by author]

The p-value to consider in this case is that of Wilks' lambda, relative to the output variable (Species). As we can see, even in this case it is significant.

We can consider this short guide to statistical tests finished. I hope it helped to clarify the concepts and avoided unnecessary headaches. 😊

See you soon,
Francesco

Anova

T Test

Python

Data Science

Statistics



Published in Towards Data Science

769K Followers · Last published 1 hour ago

Follow

Your home for data science and AI. The world's leading publication for data science, data analytics, data engineering, machine learning, and artificial intelligence professionals.



Written by Francesco Dallanocce

400 Followers · 760 Following

Follow

Computer Science Engineer 🖥️ | Data Scientist 🧠 | Python lovers 🐍 | Excited to learn every day

Responses (3)



What are your thoughts?

Respond



Joe San Pietro

over 3 years ago



First, thank you for taking the time to explain Statistical Tests. I think it takes courage to put together a concise post on a topic. I hope you don't mind my suggesting a correction in your p-value example. I think I know what you were aiming.....

[Read More](#)



1



1 reply

Reply



Jeff the Content Profit Coach
about 3 years ago



The idea of how a statistical test works is quite simple: they assume that a particular situation occurs, and then they estimate how likely this assumption is to be false.

Thanks for that...

Thanks also for the follow - followed you back :-). Since you're into Content Creation and I run SmartContentIncome.com I'm inviting you to visit and get your 3 Fresh Content Ideas on me, at the very top of the page.

See you on the inside, ~ Jeff



Reply



Kevin Connolly
over 3 years ago



Well done!



Reply

More from Francesco Dallanocce and Towards Data Science



 Francesco Dallanoce

Explainable AI: A Comprehensive Review of the Main Methods

In recent years, with the spread of AI models of various types (trees, neural networks, etc.)...

Jan 5, 2022  462  3  




 In Towards Data Science by Javier Marin

A Story of Long Tails: Why Uncertainty in Marketing Mix...

“Details matter. It’s worth waiting to get it right.” — Steve Jobs

 1d ago  45  2  



 In Towards Data Scie... by Jonathan R. Williford, P...

Neuromorphic Computing—an Edgier, Greener AI

There are periodic proclamations of the coming neuromorphic computing revolution...

6d ago  222  4  

 Francesco Dallanoce

Pandas Profiling: Full Guide on Complete Data Exploration in one...

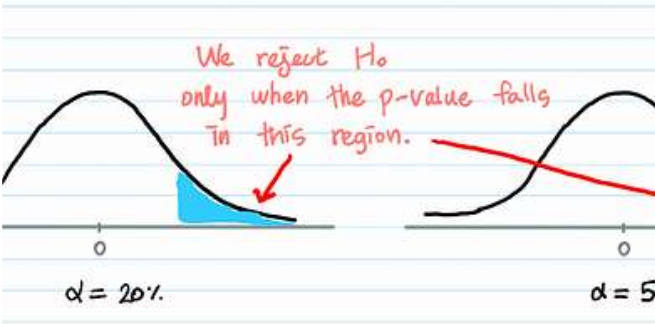
A comprehensive guide on how to generate an exhaustive data exploration via pandas...

Aug 17, 2022  256  1  

See all from Francesco Dallanoce

See all from Towards Data Science

Recommended from Medium



 In IntuitionMath by Aerin Kim

Chi Square Test — Intuition, Examples, and Step-by-Step...

The best way to see if two variables are related.

★ Feb 13, 2023 👤 654 💬 4 📌 ⋮



 Irina (Xinli) Yu, Ph.D.

Understanding Hypothesis Testing: T-Test, Z-Test, Chi-Square Test, an...

Hypothesis testing is a critical component of statistical analysis, allowing researchers to...

★ Jun 1 👤 13 💬 2 📌 ⋮

Lists



Predictive Modeling w/ Python

20 stories · 1689 saves



Practical Guides to Machine Learning

10 stories · 2057 saves



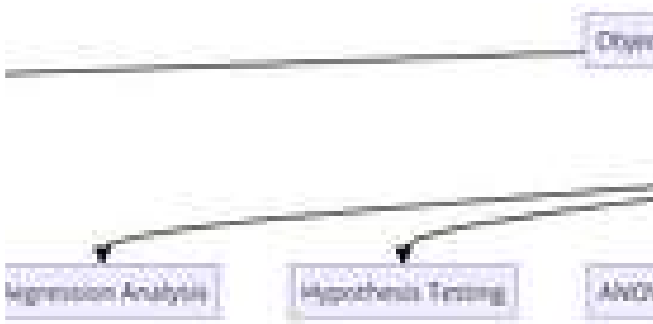
Coding & Development


11 stories · 920 saves



ChatGPT prompts

50 stories · 2294 saves



 Prem Vishnoi(cloudvala)

Probability and Statistics in Machine Learning

Introduction:

★ Aug 19  6



 Diogo Santos

Why Correlation Isn't Enough: Unlocking Causality with...

How Randomized Controlled Trials Transform Data into Actionable Insights and Drive...

★ Nov 20  31




 JABERI Mohamed Habib

Hypothesis Testing in Python Part 2: Two-Sample and ANOVA Tests

Hypothesis testing is a crucial concept in statistics, used to make inferences about...

Aug 3

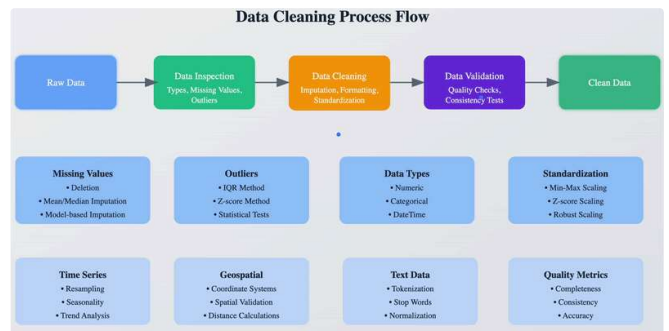


 In Level Up Coding by Md Arman Hossen

The Art of Data Cleaning in Python: From Chaos to Clarity—A...

A Data Scientist's Guide to Taming Wild Datasets

★ 6d ago  61



See more recommendations