New to Plotly?

Plotly's Python library is free and open source! Get started (https://plot.ly/python/getting-started/) by dowloading the client and reading the primer (https://plot.ly/python/getting-started/).
You can set up Plotly to work in online (https://plot.ly/python/getting-started/#initialization-for-online-plotting) or offline (https://plot.ly/python/getting-started/#initialization-for-offline-plotting) mode, or in jupyter notebooks (https://plot.ly/python/getting-started/#start-plotting-online).
We also have a quick-reference cheatsheet (https://images.plot.ly/plotly-documentation/images/python_cheat_sheet.pdf) (new!) to help you get started!

Imports

The tutorial below imports NumPy (http://www.numpy.org/), Pandas (https://plot.ly/pandas/intro-to-pandas-tutorial/), SciPy (https://www.scipy.org/), and Statsmodels (http://statsmodels.sourceforge.net/stable/).

In [1]:
```
import plotly.plotly as py
import plotly.graph_objs as go
from plotly.tools import FigureFactory as FF

import numpy as np
import pandas as pd
import scipy

import statsmodels
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

One-Way ANOVA

An `Analysis of Variance Test` or an `ANOVA` is a generalization of the t-tests to more than 2 groups. Our null hypothesis states that there are equal means in the populations from which the groups of data were sampled. More succinctly:

$$\mu_1 = \mu_2 = ... = \mu_n$$

for n groups of data. Our alternative hypothesis would be that any one of the equivalences in the above equation fail to be met.

In [2]:
```
moore = sm.datasets.get_rdataset("Moore", "car", cache=True)

data = moore.data
data = data.rename(columns={"partner.status" :"partner_status"})  # make name pythonic

moore_lm = ols('conformity ~ C(fcategory, Sum)*C(partner_status, Sum)', data=data).fit()
table = sm.stats.anova_lm(moore_lm, typ=2) # Type 2 ANOVA DataFrame

print(table)
```

```
                                       sum_sq    df          F  \
C(fcategory, Sum)                   11.614700   2.0   0.276958
C(partner_status, Sum)             212.213778   1.0  10.120692
C(fcategory, Sum):C(partner_status, Sum)  175.488928   2.0   4.184623
Residual                           817.763961  39.0        NaN

                                      PR(>F)
C(fcategory, Sum)                   0.759564
C(partner_status, Sum)              0.002874
C(fcategory, Sum):C(partner_status, Sum)  0.022572
Residual                                 NaN
```

In this ANOVA test, we are dealing with an `F-Statistic` and not a `p-value`. Their connection is integral as they are two ways of expressing the same thing. When we set a `significance level` at the start of our statistical tests (usually 0.05), we are saying that if our variable in question takes on the 5% ends of our distribution, then we can start to make the case that there is evidence against the null, which states that the data belongs to *this particular distribution*.

The F value is the point such that the area of the curve past that point to the tail is just the p-value. Therefore:

$$Pr(>F) = p$$

For more information on the choice of 0.05 for a significance level, check out this page (http://www.investopedia.com/exam-guide/cfa-level-1/quantitative-methods/hypothesis-testing.asp).

Let us import some data for our next analysis. This time some data on tooth growth:

In [3]:
```
data = pd.read_csv('https://raw.githubusercontent.com/plotly/datasets/master/tooth_growth_csv')
df = data[0:10]

table = FF.create_table(df)
py.iplot(table, filename='tooth-data-sample')
```

Out[3]:

Two-Way ANOVA

In a `Two-Way ANOVA`, there are two variables to consider. The question is whether our variable in question (tooth length `len`) is related to the two other variables `supp` and `dose` by the equation:

$$len = supp + dose + supp \times dose$$

In [4]:
```
formula = 'len ~ C(supp) + C(dose) + C(supp):C(dose)'
model = ols(formula, data).fit()
aov_table = statsmodels.stats.anova.anova_lm(model, typ=2)
print(aov_table)
```

```
                    sum_sq    df          F        PR(>F)
C(supp)          205.350000   1.0  15.571979  2.311828e-04
C(dose)         2426.434333   2.0  91.999965  4.046291e-18
C(supp):C(dose)  108.319000   2.0   4.106991  2.186027e-02
Residual         712.106000  54.0        NaN           NaN
```

In [1]:
```
from IPython.display import display, HTML

display(HTML('<link href="//fonts.googleapis.com/css?family=Open+Sans:600,400,300,200|Inconsolata|Ubuntu+Mono:400,700" rel="stylesheet" type="text/css" />'))
display(HTML('<link rel="stylesheet" type="text/css" href="http://help.plot.ly/documentation/all_static/css/ipython-notebook-custom.css">'))

! pip install git+https://github.com/plotly/publisher.git --upgrade
import publisher
publisher.publish(
    'python-Anova.ipynb', 'python/anova/', 'Anova | plotly',
    'Learn how to perform a one and two way ANOVA test using Python.',
    title='Anova in Python | plotly',
    name='Anova',
    language='python',
    page_type='example_index', has_thumbnail='false', display_as='statistics', order=8,
    ipynb= '~notebook_demo/108')
```

```
Collecting git+https://github.com/plotly/publisher.git
  Cloning https://github.com/plotly/publisher.git to /var/folders/ld/6cl3s_l50wd40tdjq2b03jxh0000gp/T/pip-KRjKqE-build
Installing collected packages: publisher
  Found existing installation: publisher 0.10
    Uninstalling publisher-0.10:
```