



# 1-Way ANOVA from scratch - dissecting the ANOVA table

Link to the article:



**1-way ANOVA from scratch — dissecting the ANOVA table with a worked example**

What really happens in ANOVA!

**M** Medium | Joos Korstanje | 11 nov. 2019



# Step-by-Step 1-Way ANOVA from scratch notebook

## 1. Data Creation

```
In [0]: import pandas as pd
```

```
In [0]: A = [12.6, 12, 11.8, 11.9, 13, 12.5, 14]
B = [10, 10.2, 10, 12, 14, 13]
C = [10.1, 13, 13.4, 12.9, 8.9, 10.7, 13.6, 12]
```

```
In [0]: all_scores = A + B + C
company_names = (['A'] * len(A)) + (['B'] * len(B)) + (['C'] * len(C))
```

```
In [0]: data = pd.DataFrame({'company': company_names, 'score': all_scores})
```

```
In [25]: data
```

```
Out[25]:
```

	company	score
0	A	12.6
1	A	12.0
2	A	11.8
3	A	11.9
4	A	13.0
5	A	12.5
6	A	14.0
7	B	10.0
8	B	10.2
9	B	10.0
10	B	12.0
11	B	14.0
12	B	13.0
13	C	10.1
14	C	13.0
15	C	13.4
16	C	12.9



19	C	13.6
20	C	12.0

## 2. Descriptives

In [26]: `data.groupby('company').mean()`

Out[26]:

	score
company	
A	12.542857
B	11.533333
C	11.825000

## 2. A 1-Way ANOVA Using StatsModels

In [0]: `import statsmodels.api as sm  
from statsmodels.formula.api import ols`

In [28]: `lm = ols('score ~ company',data=data).fit()  
table = sm.stats.anova_lm(lm)  
print(table)`

	df	sum_sq	mean_sq	F	PR(>F)
company	2.0	3.606905	1.803452	0.821297	0.455683
Residual	18.0	39.525476	2.195860	NaN	NaN

## 3. 1-Way ANOVA by hand (from scratch)

In [29]: `# compute overall mean  
overall_mean = data['score'].mean()  
overall_mean`

Out[29]: 11.980952380952381

In [30]: `# compute Sum of Squares Total  
data['overall_mean'] = overall_mean  
ss_total = sum((data['score'] - data['overall_mean'])**2)  
ss_total`

Out[30]: 43.132380952380956



```
group_means = group_means.rename(columns = {'score': 'group_mean'})
group_means
```

Out[31]:

	group_mean	overall_mean
company		
A	12.542857	11.980952
B	11.533333	11.980952
C	11.825000	11.980952

```
In [0]: # add group means and overall mean to the original data frame
data = data.merge(group_means, left_on = 'company', right_index = True)
```

```
In [33]: # compute Sum of Squares Residual
ss_residual = sum((data['score'] - data['group_mean'])**2)
ss_residual
```

Out[33]: 39.52547619047619

```
In [14]: # compute Sum of Squares Model
ss_explained = sum((data['overall_mean'] - data['group_mean'])**2)
ss_explained
```

Out[14]: 3.6069047619047776

```
In [14]: # compute Mean Square Residual
n_groups = len(set(data['company']))
n_obs = data.shape[0]
df_residual = n_obs - n_groups
ms_residual = ss_residual / df_residual
ms_residual
```

Out[14]: 2.1958597883597886

```
In [15]: # compute Mean Square Explained
df_explained = n_groups - 1
ms_explained = ss_explained / df_explained
ms_explained
```

Out[15]: 1.8034523809523888

```
In [16]: # compute F-Value
f = ms_explained / ms_residual
f
```

Out[16]: 0.8212966923081592

```
In [17]: # compute p-value
import scipy.stats
p_value = 1 - scipy.stats.f.cdf(f, df_explained, df_residual)
```



Link to the article:



**1-way ANOVA from scratch — dissecting the ANOVA table with a worked example**

What really happens in ANOVA!

**M** Medium | Joos Korstanje | 11 nov. 2019



Joos Korstanje, Data Scientist

Serris, Ile de France, France

Don't hesitate to send me a message or add me on [LinkedIn](#) !.

Credits: Lightbulb icon made by and adapted from Freepik from [www.flaticon.com](http://www.flaticon.com)