Tutorial 13 - Crosstab Wikipedia: In statistics, a contingency table (also known as a cross tabulation or crosstab) is a type of table in a matrix format that displays the (multivariate) frequency distribution of the variables. They are heavily used in survey research, business intelligence, engineering and scientific research. They provide a basic picture of the interrelation between two variables and can help find interactions between them. The term contingency table was first used by Karl Pearson in "On the Theory of Contingency and Its Relation to Association and Normal Correlation",[1] part of the Drapers' Company Research Memoirs Biometric Series I published in 1904.

In [1]:
```
import pandas as pd
```

In [2]:
```
# Observe que no arquivo excel há 2 níveis do título ou 2 linhas de título (linha 1
 e 2 = index 0 e 1 -> header[0,1])
# O arquivo precisa ser .xlsx caso contrário ele não verá as células mescladas e o
 título em 2 níveis
df = pd.read_excel('sample_data_tutorial_13.xlsx')
df
```

Out[2]:

	Name	Nationality	Sex	Age	Handedness
0	Kathy	USA	Female	23	Right
1	Linda	USA	Female	18	Right
2	Peter	USA	Male	19	Right
3	John	USA	Male	22	Left
4	Fatima	Bangadesh	Female	31	Left
5	Kadir	Bangadesh	Male	25	Left
6	Dhaval	India	Male	35	Left
7	Sudhir	India	Male	31	Left
8	Parvir	India	Male	37	Right
9	Yan	China	Female	52	Right
10	Juan	China	Female	58	Left
11	Liang	China	Male	43	Left

In [3]:
```
# Crosstab gera a frequência do parâmetro. Quando se utiliza o comando "margins=Tru
e" habilita-se a coluna de total
pd.crosstab(df.Nationality, df.Handedness, margins=True)
```

Out[3]:

Handedness	Left	Right	All
Nationality			
Bangadesh	2	0	2
China	2	1	3
India	2	1	3
USA	1	3	4
All	7	5	12

In [4]:
```
# É possível também gerar crosstab com múltiplas variáveis nas linhas e colunas:
pd.crosstab([df.Sex],[df.Nationality, df.Handedness])
```

Out[4]:

Nationality	Bangadesh	China		India		USA	
Handedness	Left	Left	Right	Left	Right	Left	Right
Sex							
Female	1	1	1	0	0	0	2
Male	1	1	0	2	1	1	1

In [5]:
```
# O crosstab tambem normaliza pelas linhas "index" ou colunas "columns"
pd.crosstab(df.Nationality, df.Handedness, normalize='index')
```

Out[5]:

Handedness	Left	Right
Nationality		
Bangadesh	1.000000	0.000000
China	0.666667	0.333333
India	0.666667	0.333333
USA	0.250000	0.750000

In [6]:
```
pd.crosstab(df.Nationality, df.Handedness, normalize='columns')
```

Out[6]:

Handedness	Left	Right
Nationality		
Bangadesh	0.285714	0.0
China	0.285714	0.2
India	0.285714	0.2
USA	0.142857	0.6

In [7]:
```
# É possível também inserir os valores médios do parâmetro "Age"
import numpy as np
pd.crosstab([df.Sex],[df.Handedness], values=df.Age, aggfunc=np.average)
```

Out[7]:

Handedness	Left	Right
Sex		
Female	44.5	31.0
Male	31.2	28.0