

In [22]:

```
import pandas as pd
import numpy as np
import scipy.stats as sps
import matplotlib.pyplot as plt
from statsmodels.sandbox.stats.multicomp import multipletests

%matplotlib inline
```

In [5]:

```
data = []
```

In [6]:

```
wine = pd.read_csv("wine.data", header=None)
for c in [1,3,4,8,9]:
    data.append(wine[c].values)
```

In [15]:

```
vowel = pd.read_csv("vowel-context.data", sep=" ", header=None)
data.append(vowel[7].values)
```

In [21]:

```
slump = pd.read_csv("slump_test.data")
data.append(slump.iloc[:, -1].values)
```

Проведём проверки нормальности для данных выборок.

In [24]:

```
p_values = []
for X in data:
    p_values.append(sps.kstest(X, cdf='norm')[1])
    p_values.append(sps.normaltest(X)[1])
    p_values.append(sps.shapiro(X)[1])
```

Для поправки на множественную проверку гипотез воспользуемся методом Бенджамини-Иекутиели, т.к. он контролирует FDR и является мощнейшей процедурой при отсутствии информации о зависимости между статистиками (в нашем случае некоторые статистики считаются по одной и той же выборке, поэтому они могут быть зависимы).

Нисходящей процедурой Бенджамини-Хохберга воспользоваться нельзя, т.к. она требует независимости статистик либо выполнения условия FDRS, которое я не понял.

In [26]:

```

multipletests_results = multipletests(p_values, method='fdr_by')
print multipletests_results

```

```

(array([ True,  True, False,  True, False, False,  True, False, False,
        True,  True,  True,  True, False, False,  True, False, False,
        True, False, False], dtype=bool), array([ 0.          ,  0.002
90587,  0.11807768,  0.          ,  0.19743032,
        0.19743032,  0.          ,  0.89329463,  1.          ,  0.
        ,
56,      0.02039897,  0.00100979,  0.          ,  0.05568023,  0.092203
        0.          ,  0.62660259,  0.33904126,  0.          ,  1.
        ,  1.          ]), 0.0024395572596688231, 0.002380952380952381)

```

In [27]:

```

multipletests_results[0].reshape((-1,3))

```

Out[27]:

```

array([[ True,  True, False],
       [ True, False, False],
       [ True, False, False],
       [ True,  True,  True],
       [ True, False, False],
       [ True, False, False],
       [ True, False, False]])

```

Т.о. единогласно было отклонена лишь одна гипотеза: о нормальности Зей (с нуля) выборки. Для остальных выборок по KS тесту гипотезы отклоняются, а по Шапиро принимаются, что говорит о необходимости дальнейшего исследования этих выборок.