In [67]:
```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as sps
import pandas as pd


from statsmodels.distributions.empirical_distribution import ECDF
from sklearn.neighbors import KernelDensity
import statsmodels.api as sm
from statsmodels.sandbox.stats.multicomp import multipletests
from skidmarks import wald_wolfowitz

%matplotlib inline
```

In [70]:
```
wine_data = pd.read_csv('wine.data', header = None, sep=',')
wine_data.head()

sample = np.array(wine_data.values[:,1])
```

In [71]:
```
def chi_sq(distr, sample, K):
    K = 6
    frequance = np.array([np.sum([1 for x in sample if
                  ((x > np.percentile(sample, j * 100 / K))
                   and (x <= np.percentile(sample, (j+1) * 100 / K))) ]) fo
    bounds = np.array([np.percentile(sample, q=i * 100 / K) for i in range(
    theory_fr = np.array([distr(bounds[i + 1]) - distr(bounds[i]) for i in
    print (frequance)
    print (theory_fr)
    return sps.chisquare(frequance, theory_fr, ddof=K)


chi_sq(sps.norm(np.mean(sample),np.var(sample)).cdf,sample, 8)
```

```
[29 30 33 26 29 30]
[ 17.6691722   23.10248704  52.81049983  42.07723087  23.1451848
  17.4974407 ]
```

Out[71]:
```
Power_divergenceResult(statistic=33.314424049484806, pvalue=nan)
```

In [72]:
```

def normility(sample, alpha = 0.05, method='bonferroni'):
    num_tests = 6
    p_val = np.zeros(num_tests)
    sequence = np.sign(sample - np.mean(sample))
    p_val[0] = wald_wolfowitz(sequence)['p']
    p_val[1] = sps.kstest(sample, sps.norm(np.mean(sample),np.var(sample)).
    p_val[2] = sps.jarque_bera(sample)[1]
    p_val[3] = sps.skewtest(sample)[1]
    p_val[4] = sps.shapiro(sample)[1]
    p_val[5] = sps.normaltest(sample)[1]
    return multipletests(p_val, alpha=alpha, method=method)

for method in ['bonferroni', 'holm']:
    reject, pvals_corrected =  normility(sample, alpha=0.05, method=method)
    print("method %s" % (method))
    print(reject)
```

```
method bonferroni
[ True False False False False  True]
method holm
[ True  True False False False  True]
```

Наиболее мощный без предположений о независимости - метод Холма, отвергает гипотезу о том, что эта выборка распределена нормально. Но зато это выборка:)