一、数据摘要

1、对标称属性，给出每个可能取值的频数

标称变量“季节”的频数：

{'autumn': 40, 'spring': 53, 'winter': 62, 'summer': 45}

标称变量“河流大小”的频数：

{'small': 71, 'large': 45, 'medium': 84}

标称变量“河流速度”的频数：

{'high': 84, 'medium': 83, 'low': 33}

2、数值属性，最大、最小、均值、中位数、四分位数及缺失值的个数

'max'表示最大值

'min'表示最小值

'mean'表示最均值

'mid'表示中位数

'quantile'表示四分位数

'max'表示最大值

'lost\_num'表示缺失值的个数

数值属性'mxPH'：

('mxPH', {'lost\_num': 1, 'min': 5.6, 'max': 9.7, 'quantile': {'Q1': 7.7, 'Q3': 8.4, 'Q2': 8.06}, 'mid': 8.06, 'mean': 8.0117336683417086})

数值属性'mnO2'：

('mnO2', {'lost\_num': 2, 'min': 1.5, 'max': 13.4, 'quantile': {'Q1': 7.7, 'Q3': 10.8, 'Q2': 9.8}, 'mid': 9.8, 'mean': 9.1177777777777713})

数值属性'CI'：

('CI', {'lost\_num': 10, 'min': 0.222, 'max': 391.5, 'quantile': {'Q1': 10.96525, 'Q3': 57.96875, 'Q2': 32.730000000000004}, 'mid': 32.730000000000004, 'mean': 43.636278842105256})

数值属性'NO3'：

('NO3', {'lost\_num': 2, 'min': 0.05, 'max': 45.65, 'quantile': {'Q1': 1.28775, 'Q3': 4.4712499999999995, 'Q2': 2.675}, 'mid': 2.675, 'mean': 3.2823888888888906})

数值属性'NH4'：

('NH4', {'lost\_num': 2, 'min': 5.0, 'max': 24064.0, 'quantile': {'Q1': 38.02775, 'Q3': 227.8875075, 'Q2': 103.1665}, 'mid': 103.1665, 'mean': 501.2958283838384})

数值属性'oPO4'：

('oPO4', {'lost\_num': 2, 'min': 1.0, 'max': 564.59998, 'quantile': {'Q1': 15.36675, 'Q3': 100.5, 'Q2': 40.15}, 'mid': 40.15, 'mean': 73.590595959595973})

数值属性'PO4'：

('PO4', {'lost\_num': 2, 'min': 1.0, 'max': 771.59998, 'quantile': {'Q1': 40.50025, 'Q3': 214.375, 'Q2': 103.2855}, 'mid': 103.2855, 'mean': 137.88210095959593})

数值属性'Chla'：

('Chla', {'lost\_num': 12, 'min': 0.2, 'max': 110.456, 'quantile': {'Q1': 2.0, 'Q3': 18.322499999999998, 'Q2': 5.475}, 'mid': 5.475, 'mean': 13.971196808510641})

数值属性'a1'：

('a1', {'lost\_num': 0, 'min': 0.0, 'max': 89.8, 'quantile': {'Q1': 1.5, 'Q3': 24.8, 'Q2': 6.95}, 'mid': 6.95, 'mean': 16.923500000000001})

数值属性'a2：

('a2', {'lost\_num': 0, 'min': 0.0, 'max': 72.6, 'quantile': {'Q1': 0.0, 'Q3': 11.525, 'Q2': 3.0}, 'mid': 3.0, 'mean': 7.4585000000000017})

数值属性'a3：

('a3', {'lost\_num': 0, 'min': 0.0, 'max': 42.8, 'quantile': {'Q1': 0.0, 'Q3': 4.975, 'Q2': 1.55}, 'mid': 1.55, 'mean': 4.3095000000000008})

数值属性'a4：

('a4', {'lost\_num': 0, 'min': 0.0, 'max': 44.6, 'quantile': {'Q1': 0.0, 'Q3': 2.4, 'Q2': 0.0}, 'mid': 0.0, 'mean': 1.9925000000000004})

数值属性'a5：

('a5', {'lost\_num': 0, 'min': 0.0, 'max': 44.4, 'quantile': {'Q1': 0.0, 'Q3': 7.5, 'Q2': 1.9}, 'mid': 1.9, 'mean': 5.0644999999999989})

数值属性'a6：

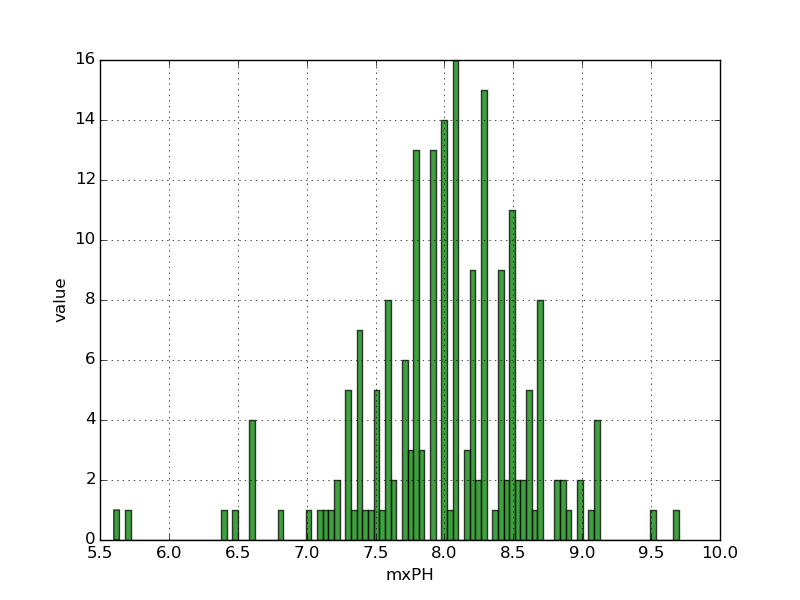
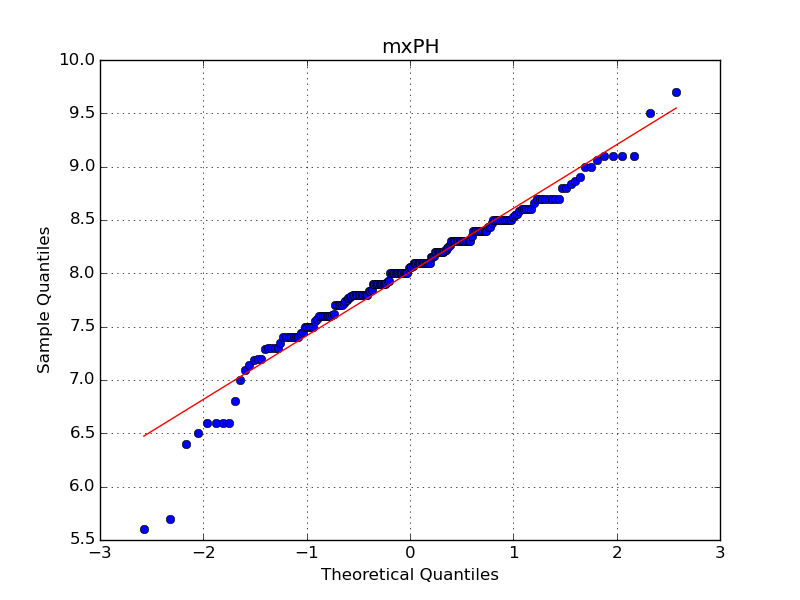
('a6', {'lost\_num': 0, 'min': 0.0, 'max': 77.6, 'quantile': {'Q1': 0.0, 'Q3': 6.975, 'Q2': 0.0}, 'mid': 0.0, 'mean': 5.9639999999999995})

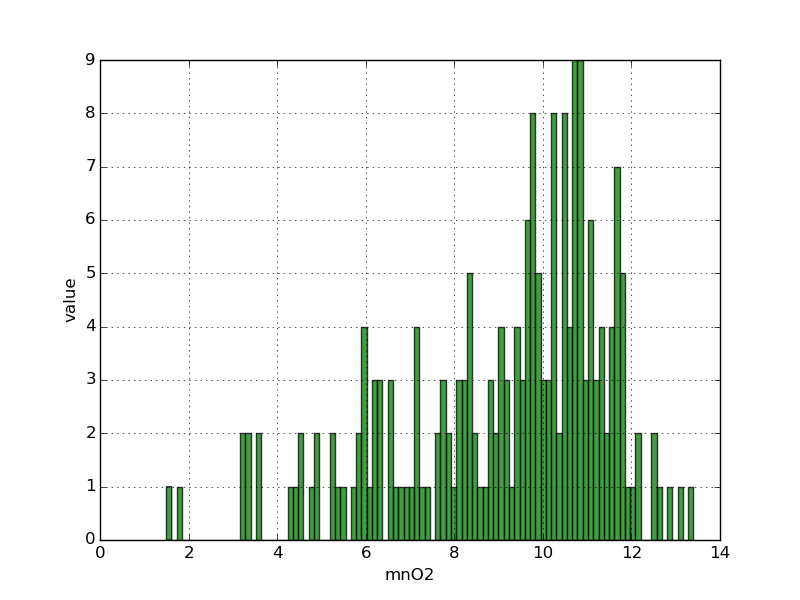
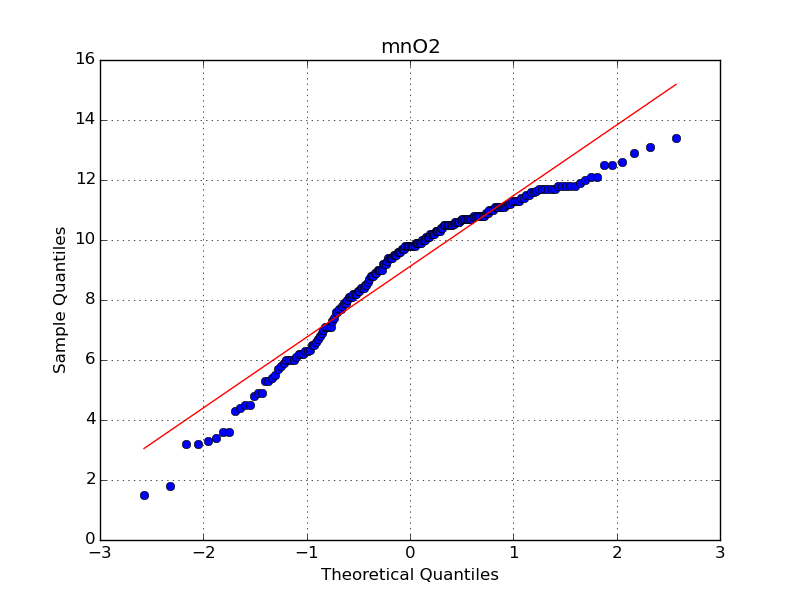
数值属性'a7：

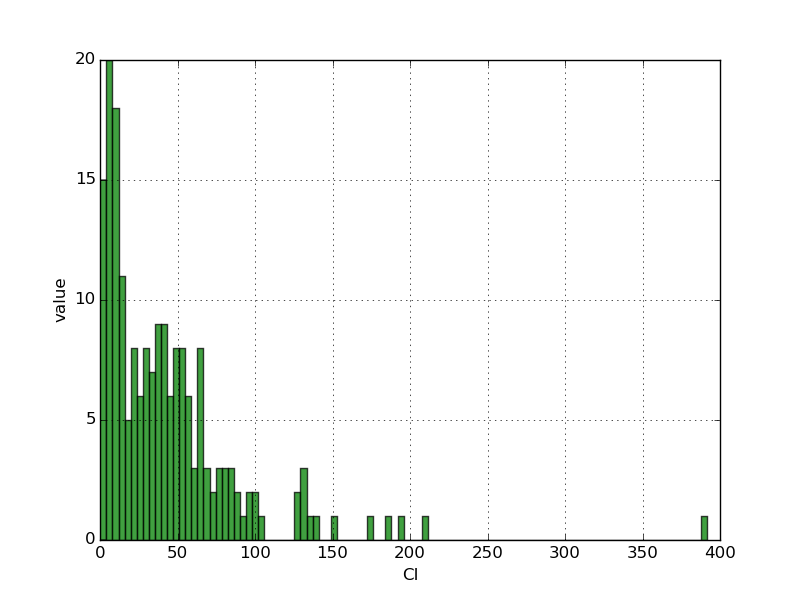
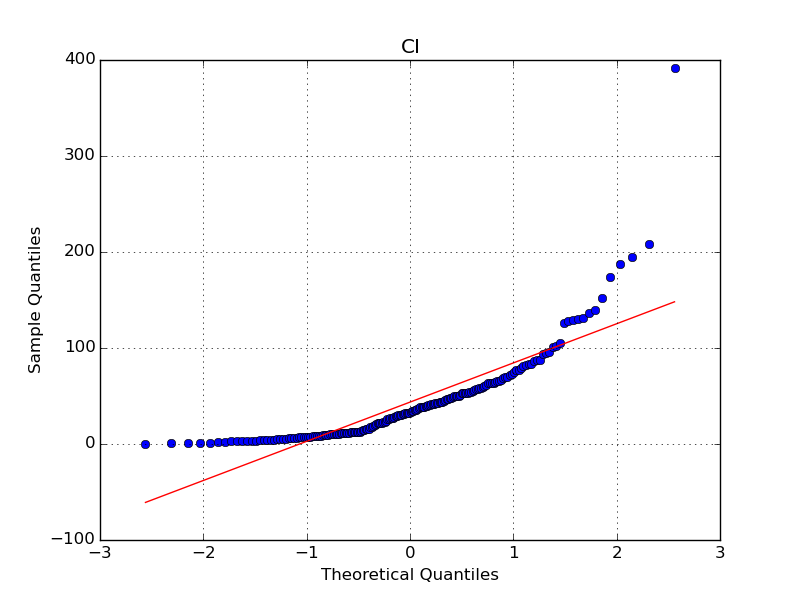
('a7', {'lost\_num': 0, 'min': 0.0, 'max': 31.6, 'quantile': {'Q1': 0.0, 'Q3': 2.4, 'Q2': 1.0}, 'mid': 1.0, 'mean': 2.4955000000000007})

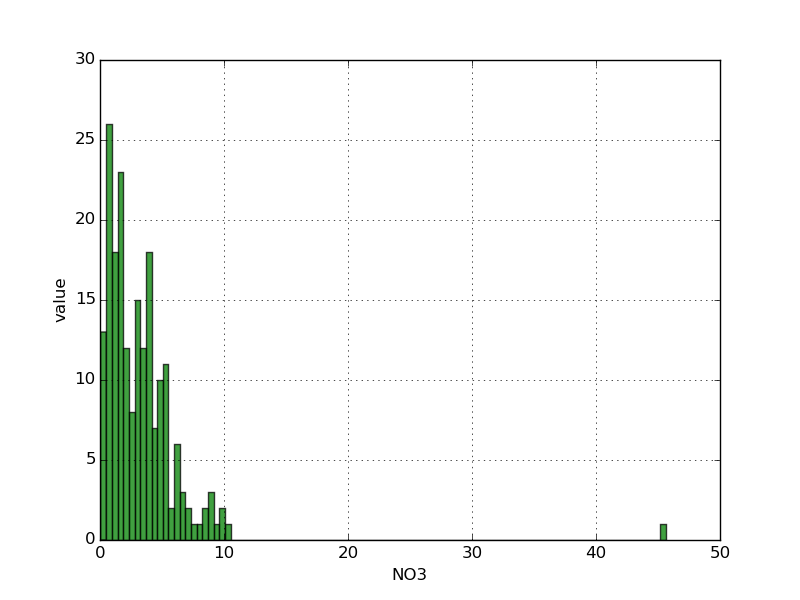
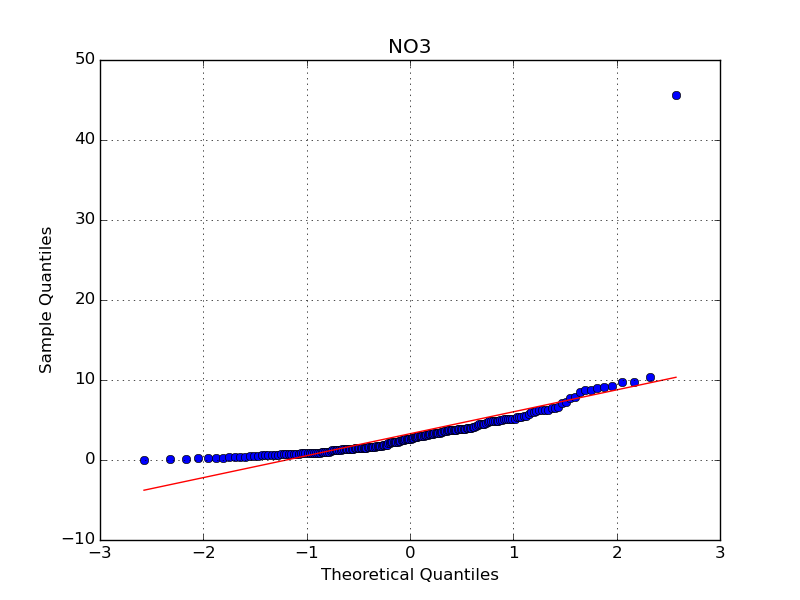
二、数据的可视化

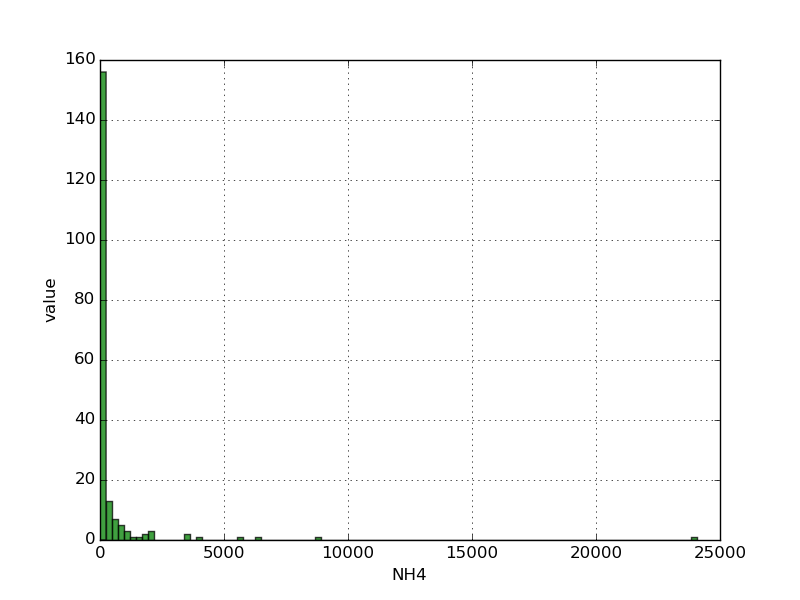
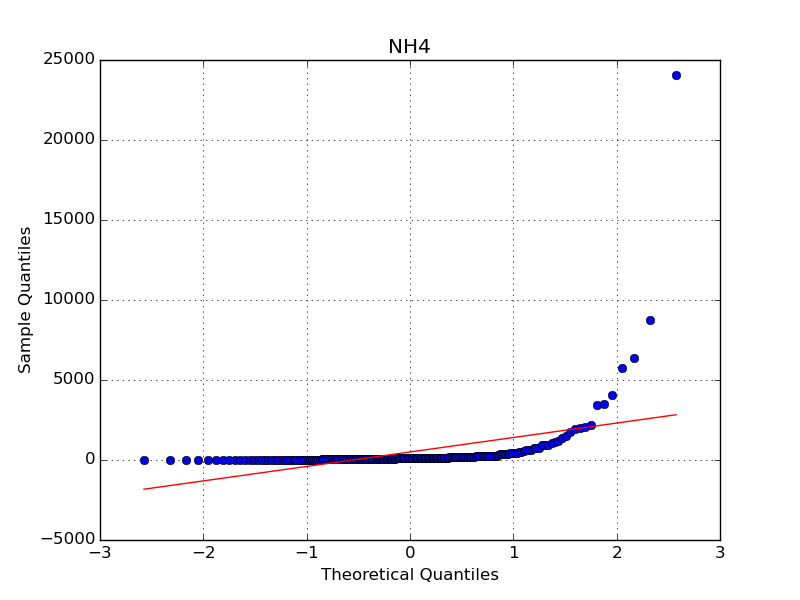
1、直方图及qq图检验

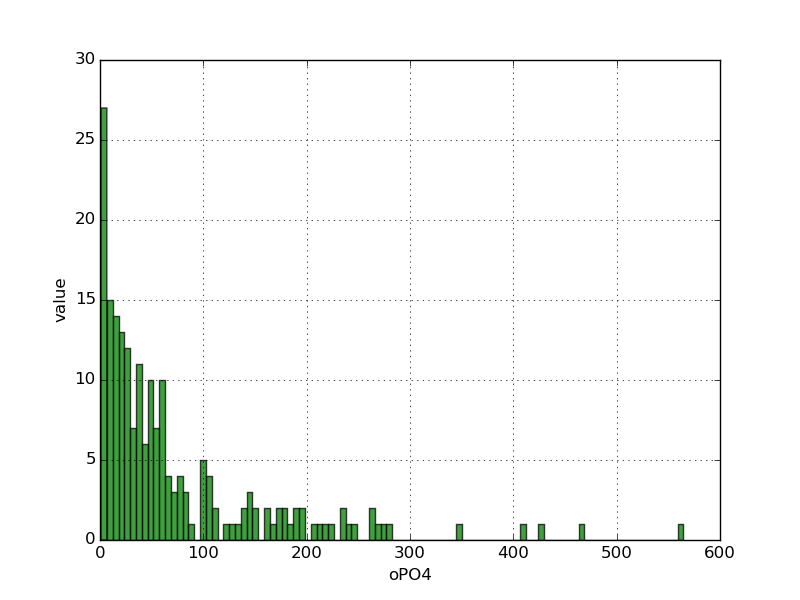
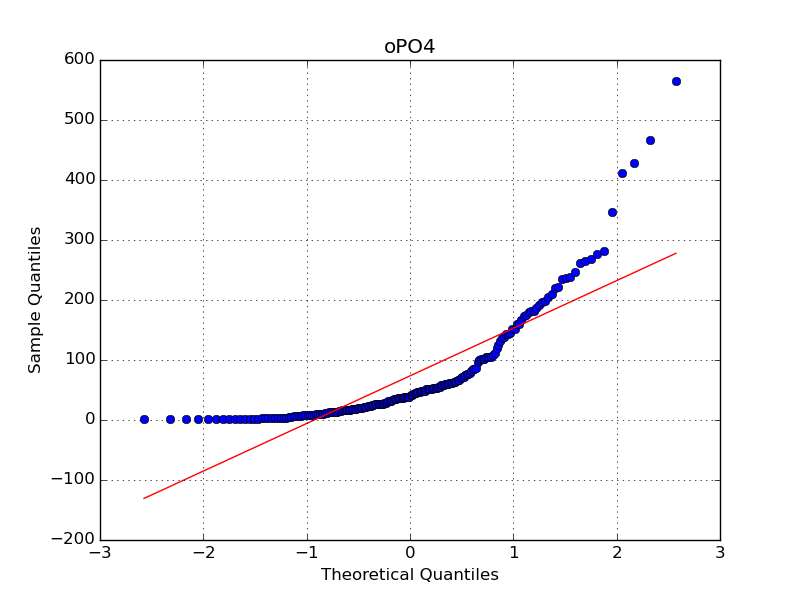
 

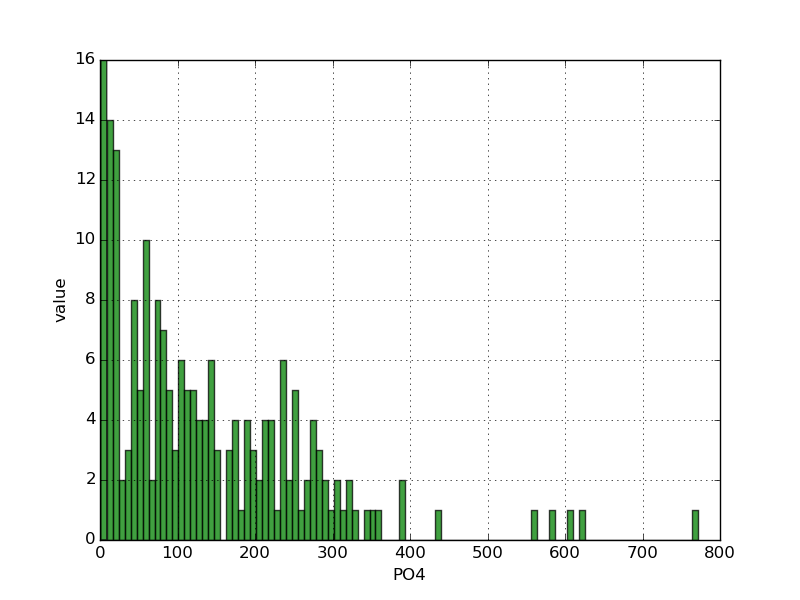
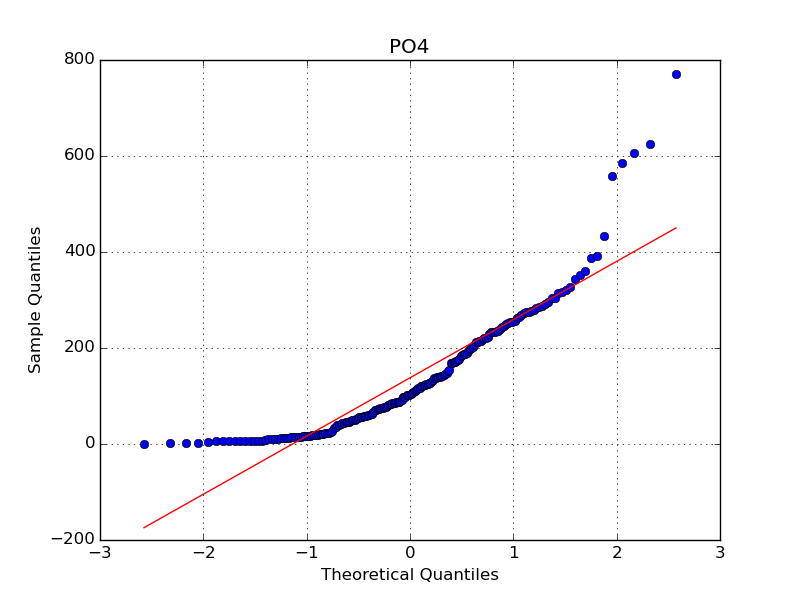
 

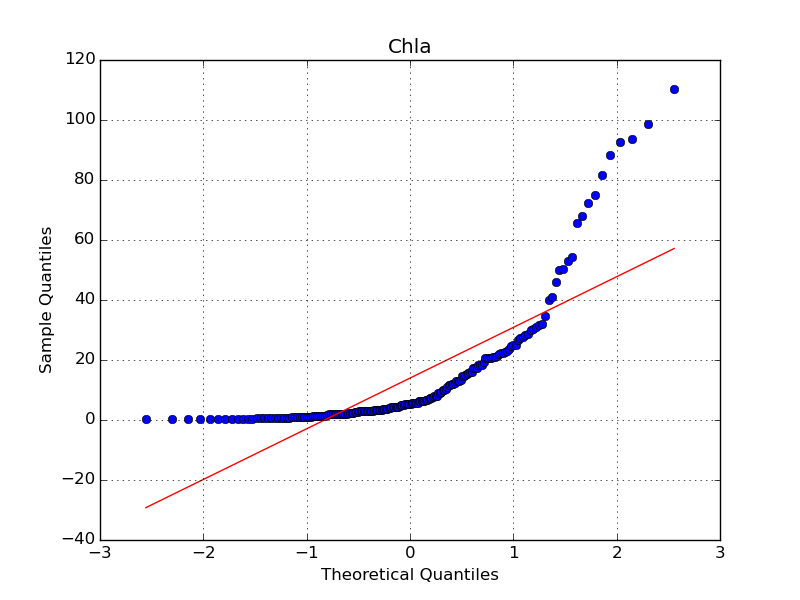
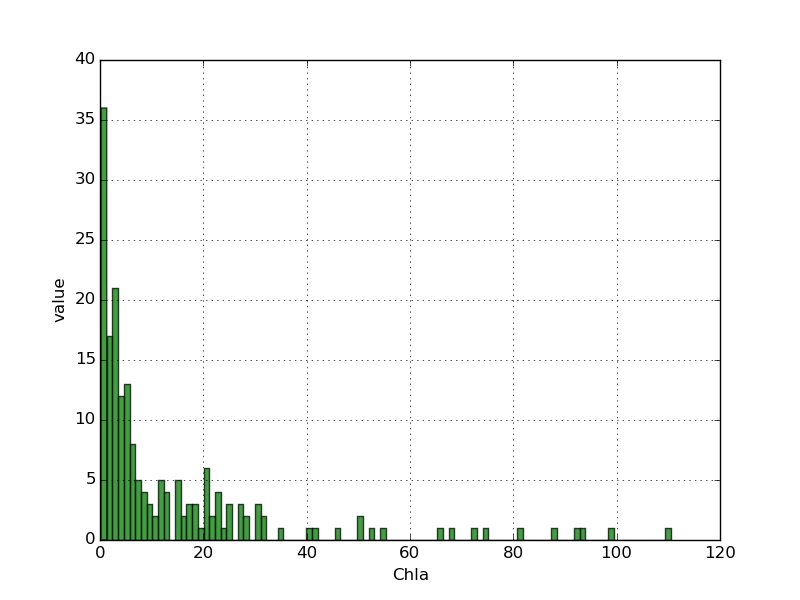
 

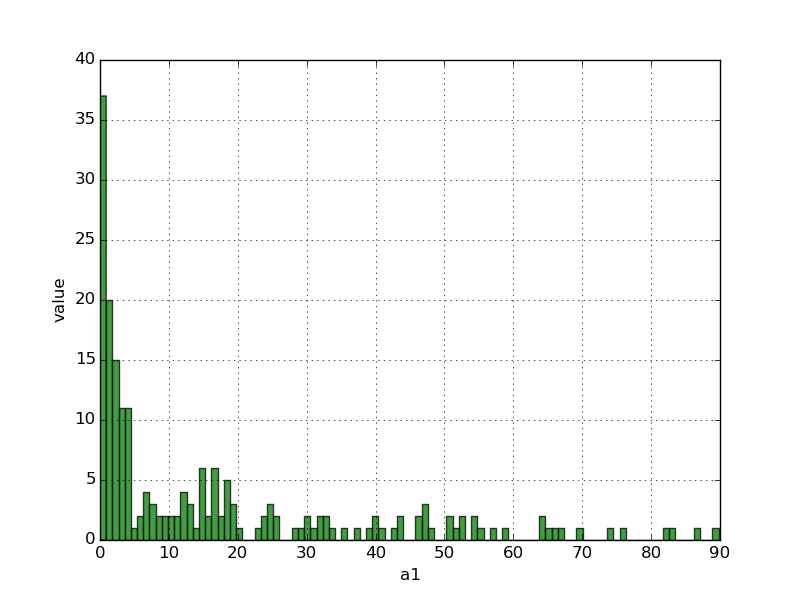
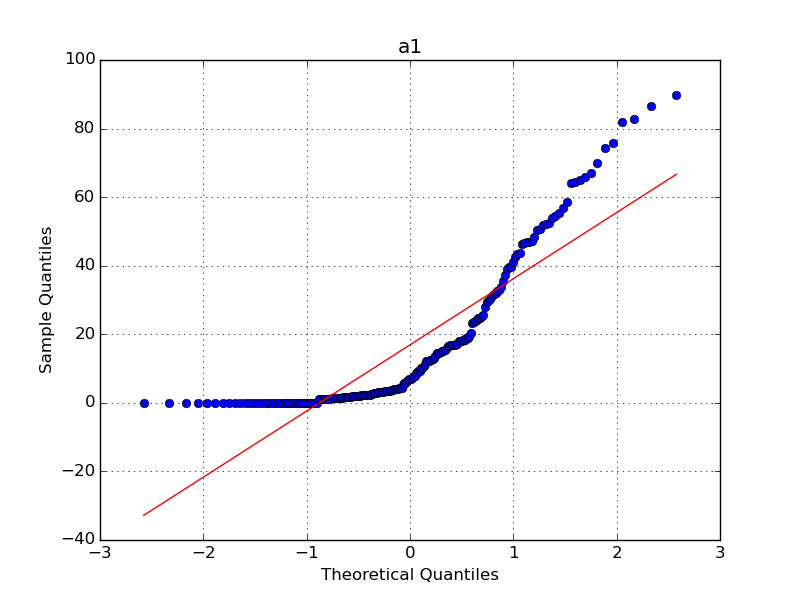
 

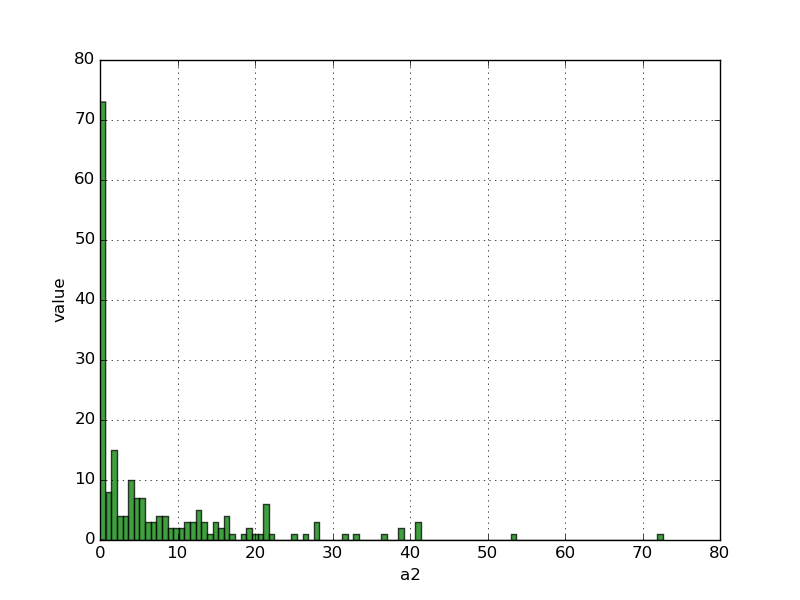
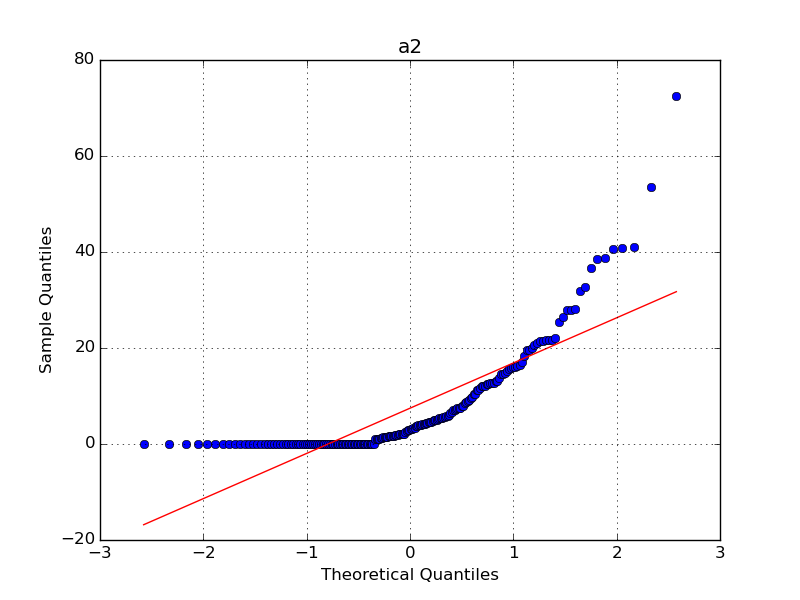
 

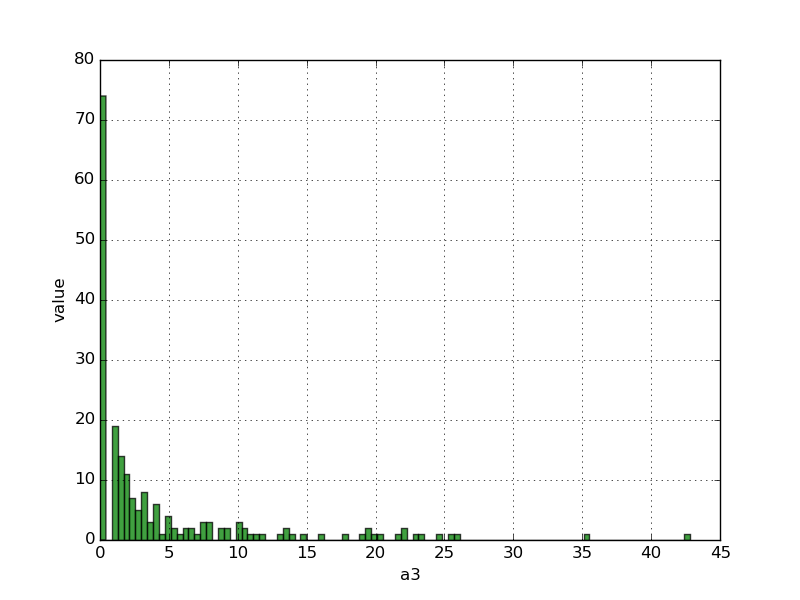
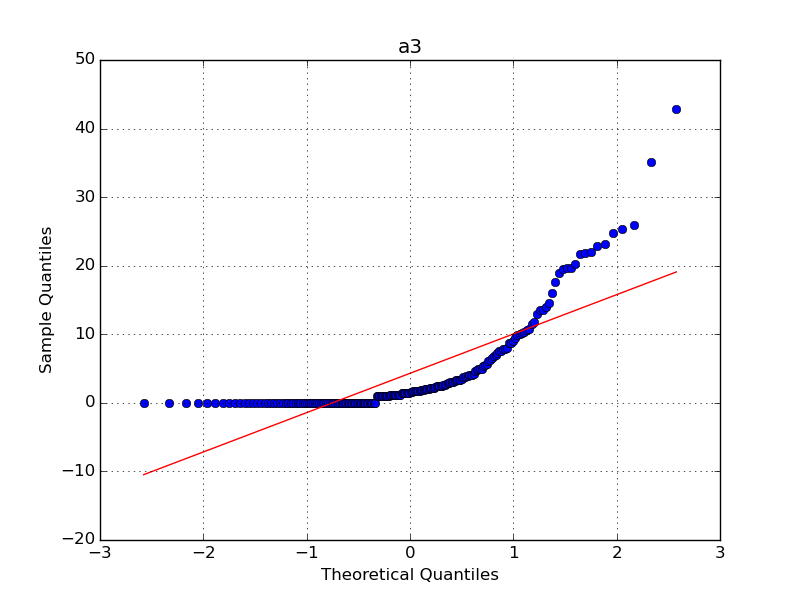
 

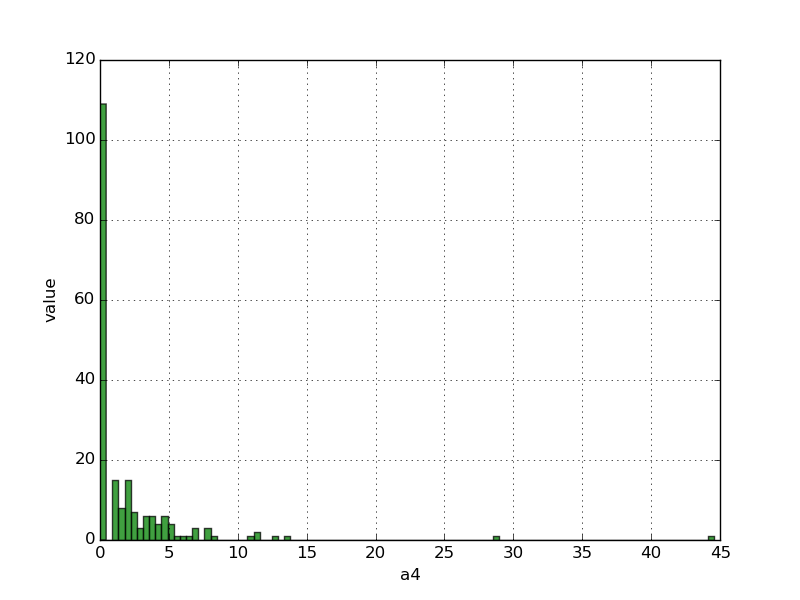
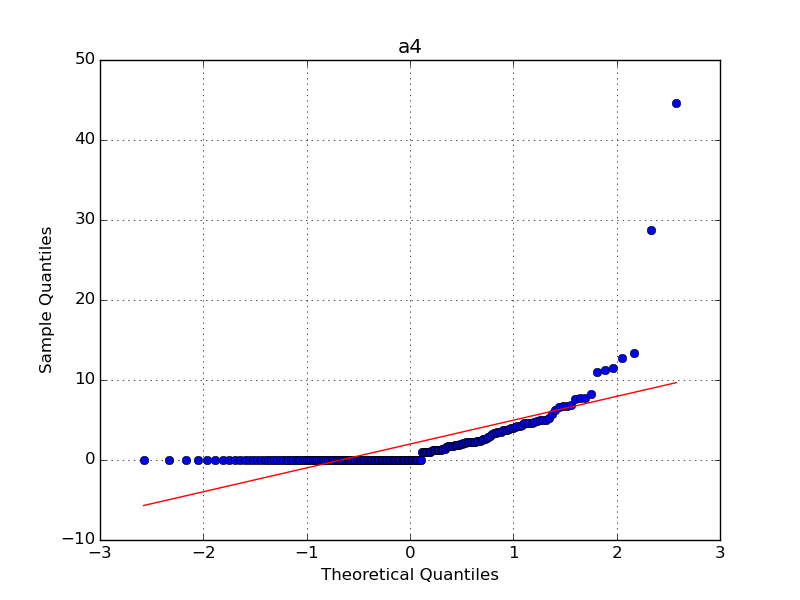
 

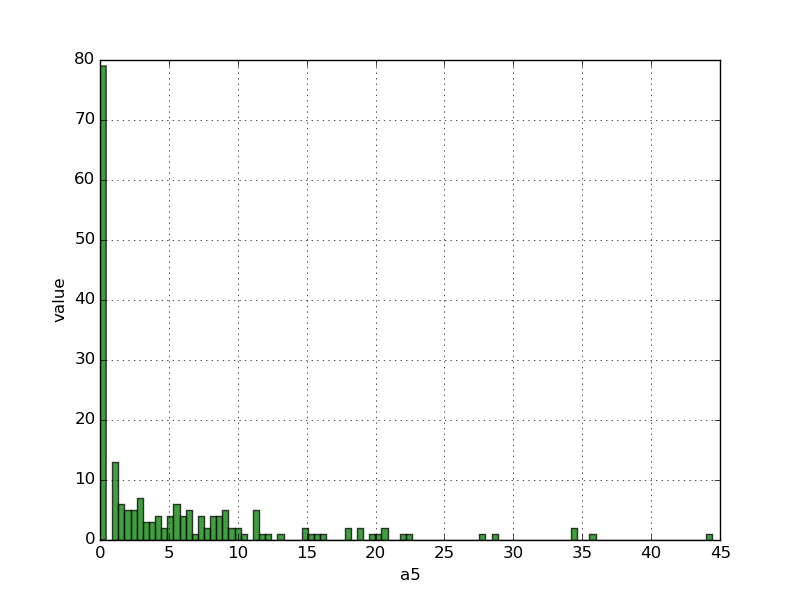
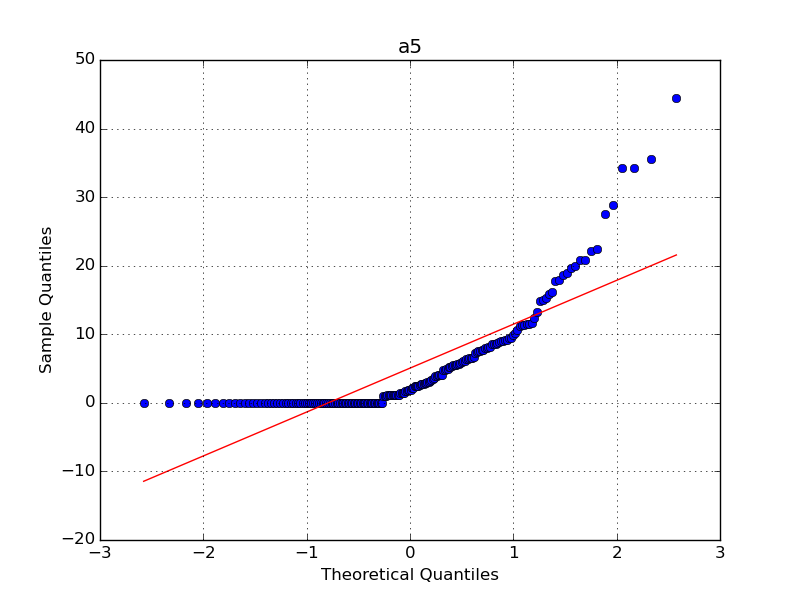


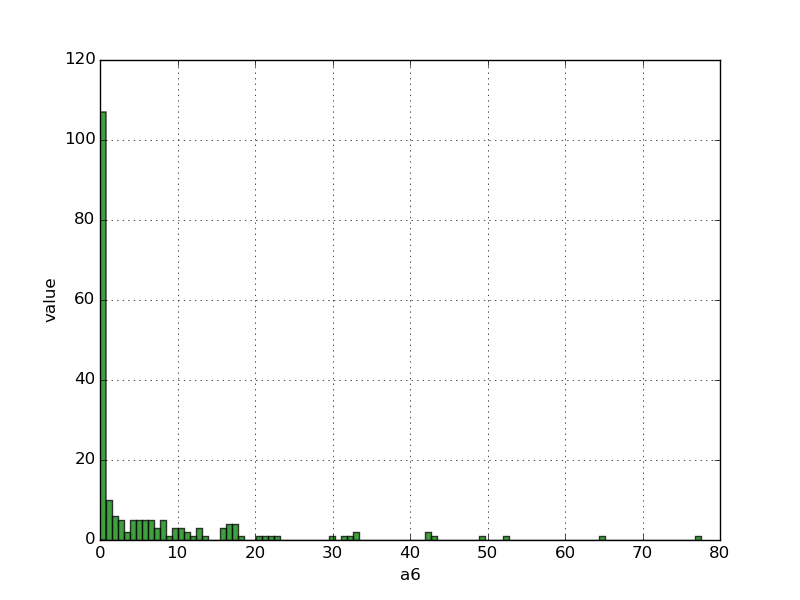
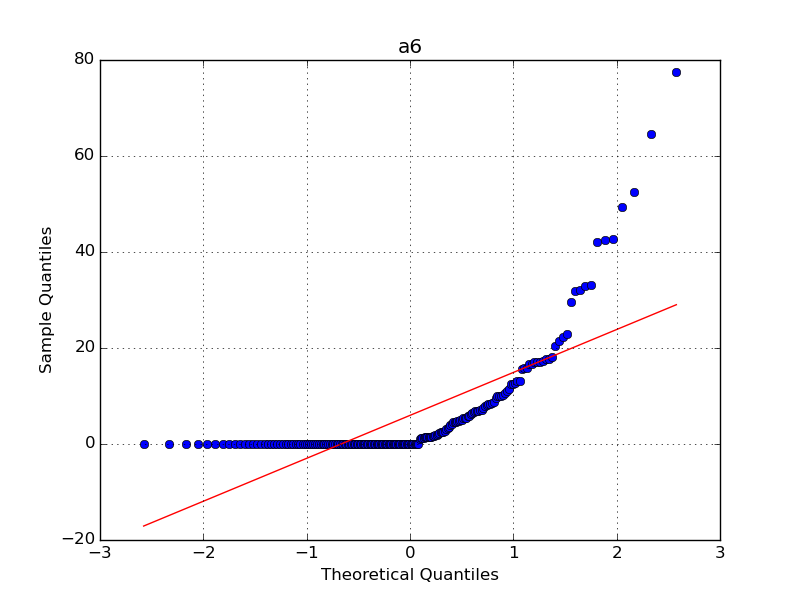
 

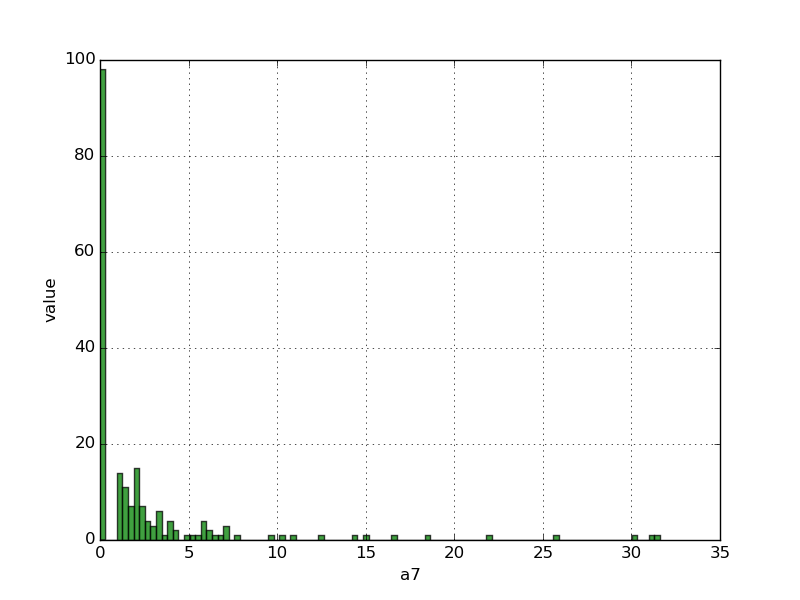
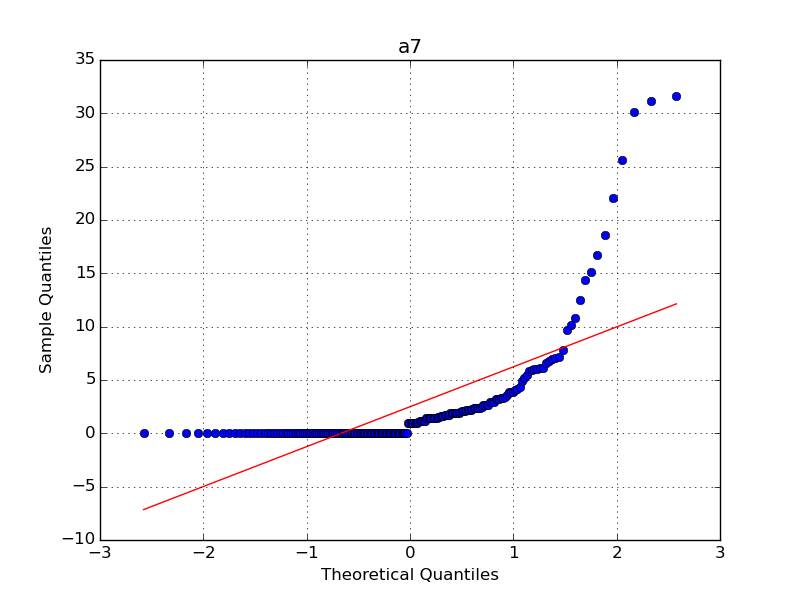
 

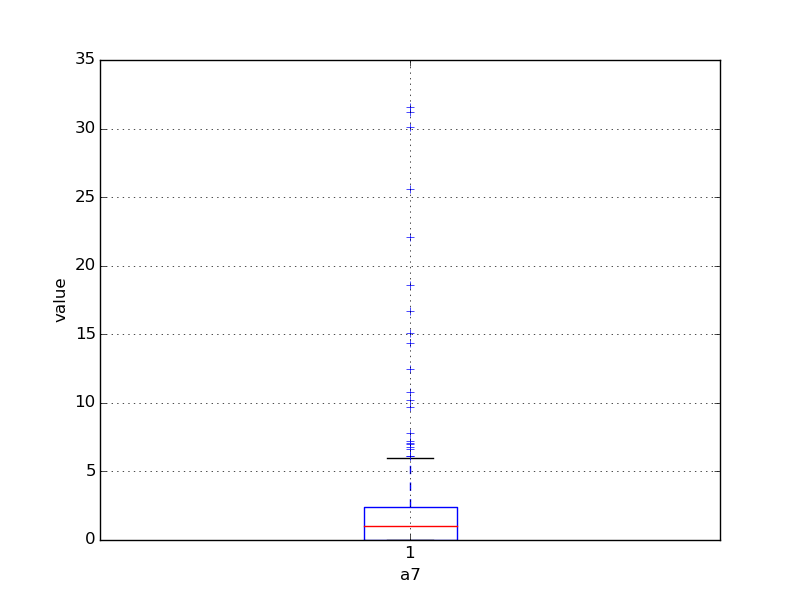
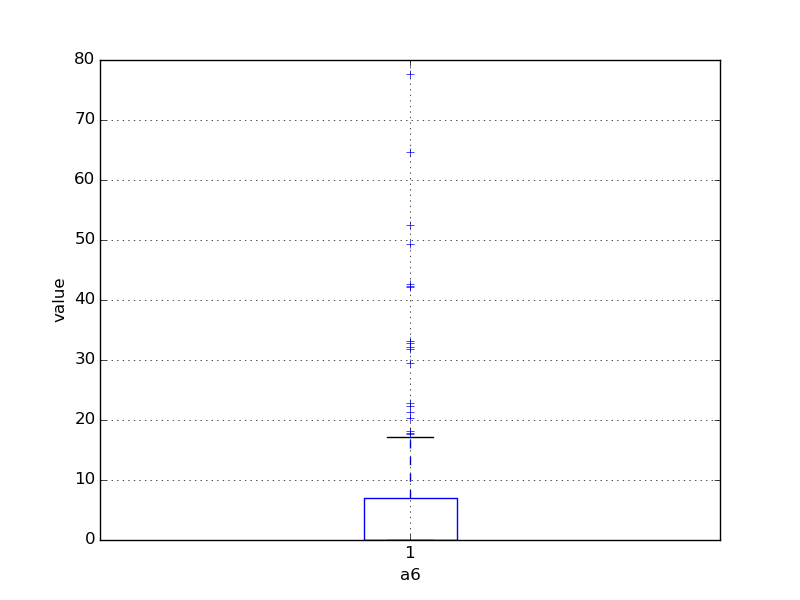
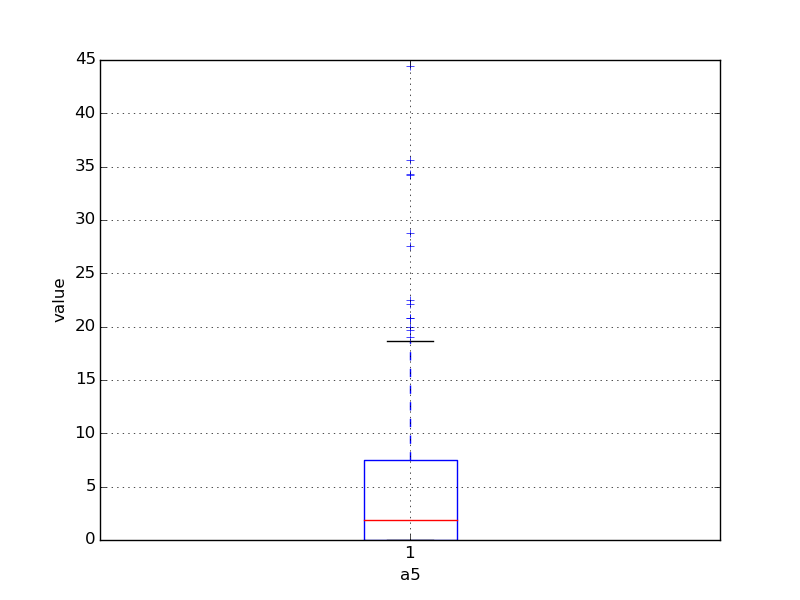
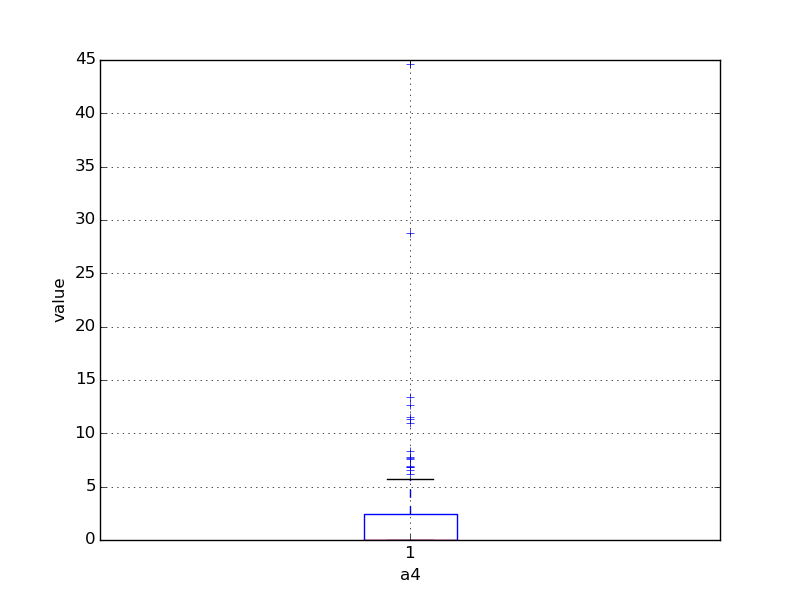
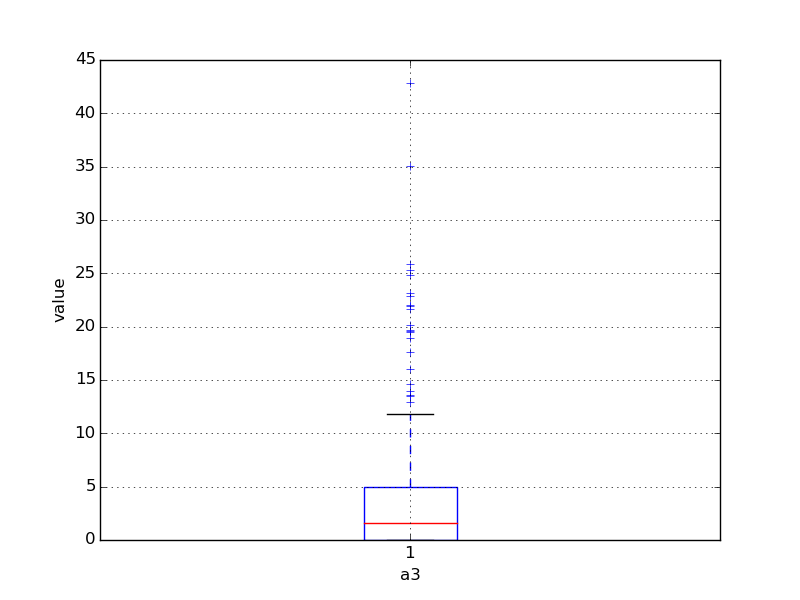
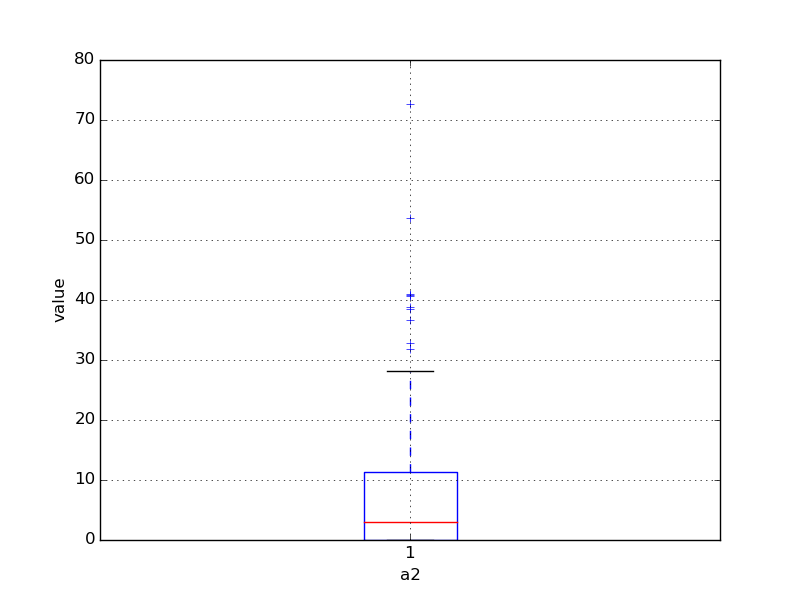
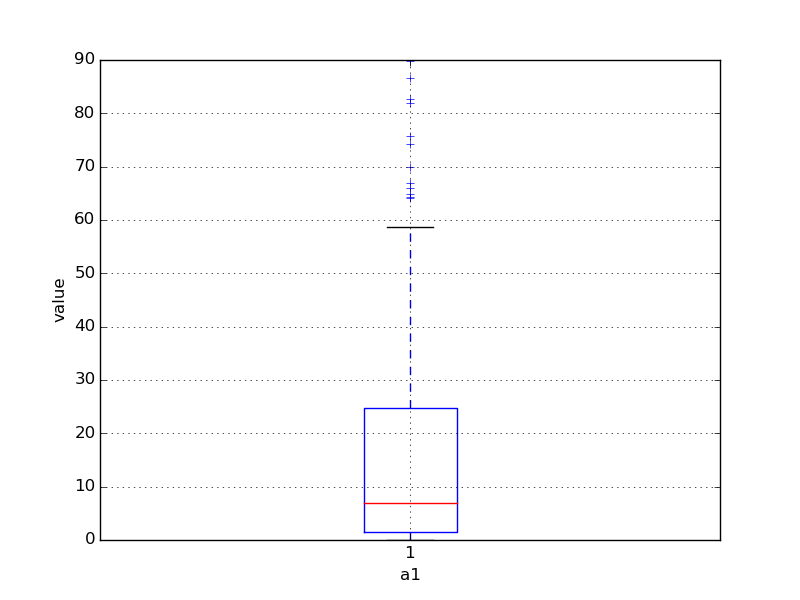
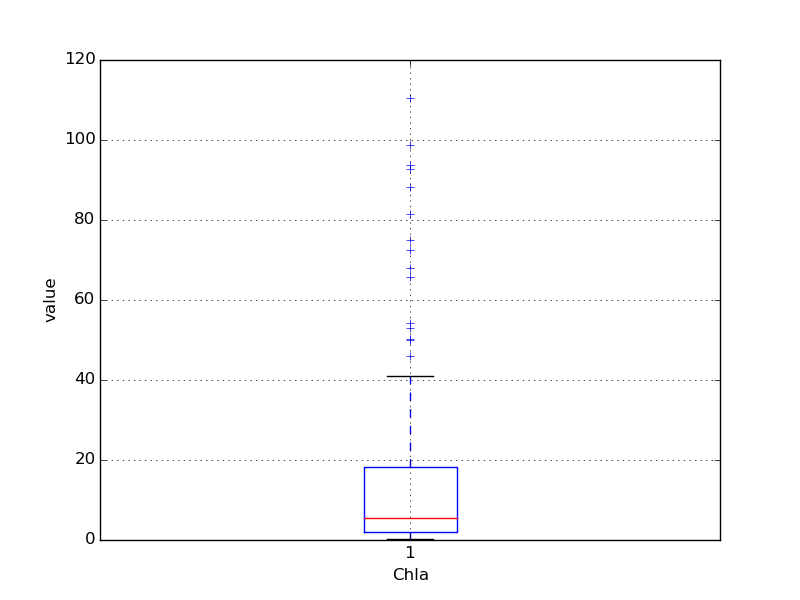
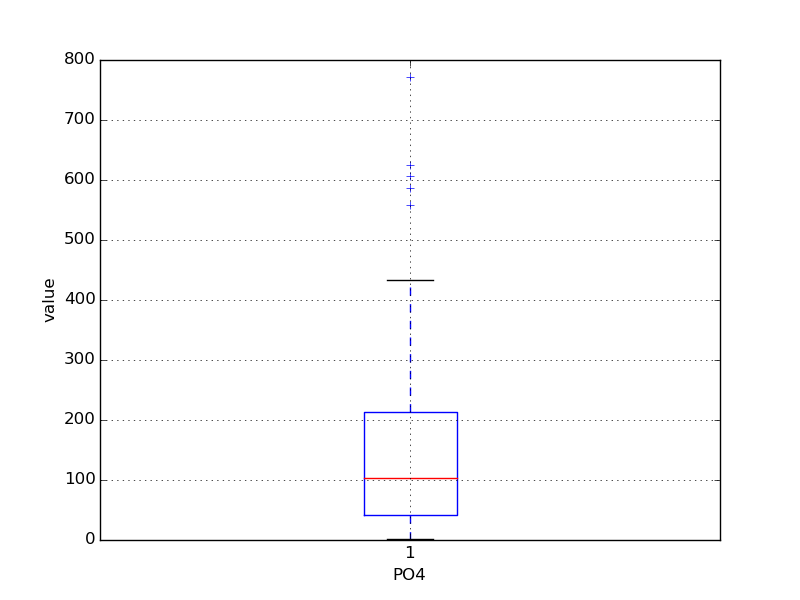
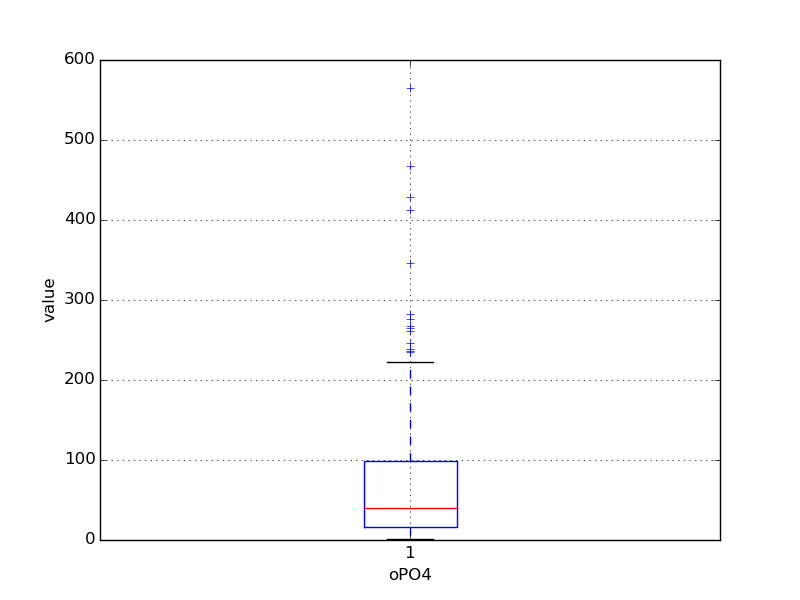
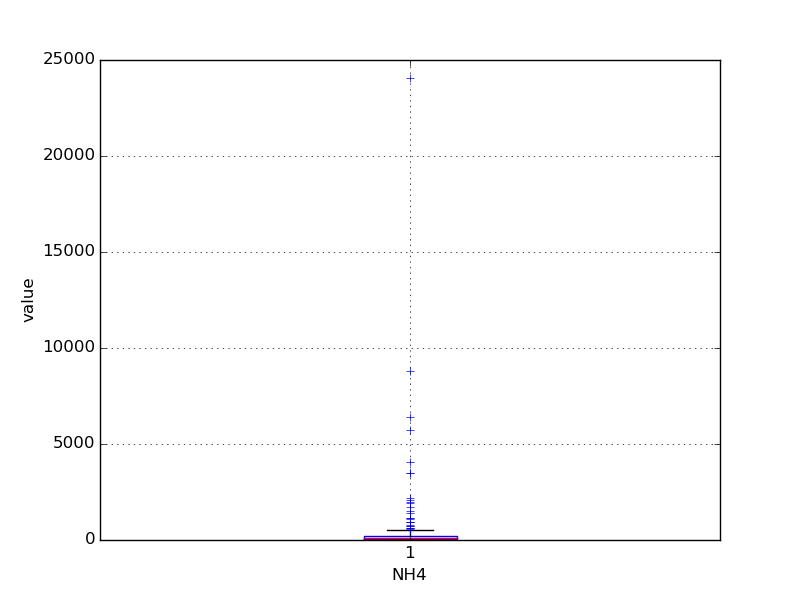
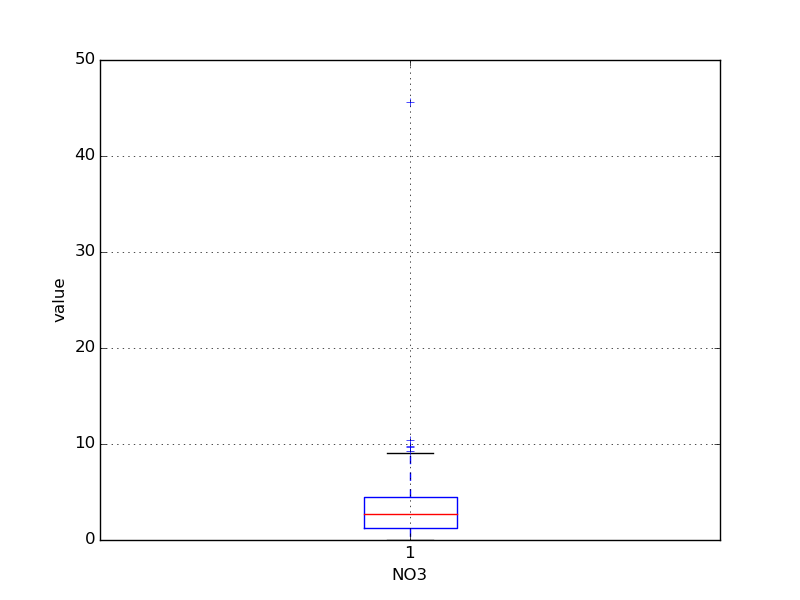
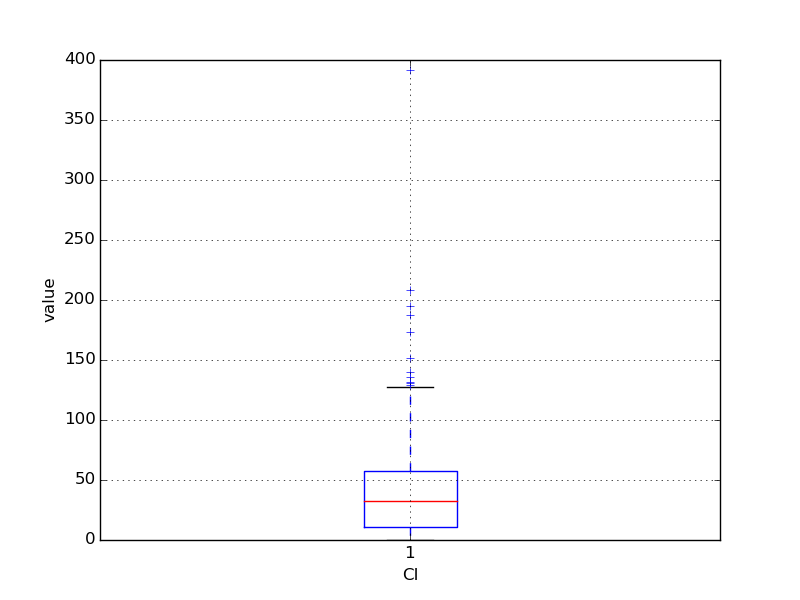
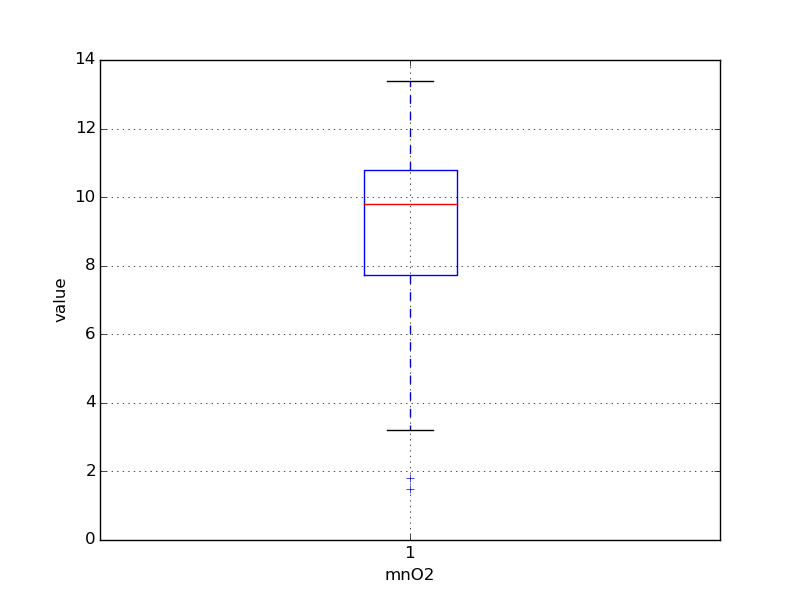
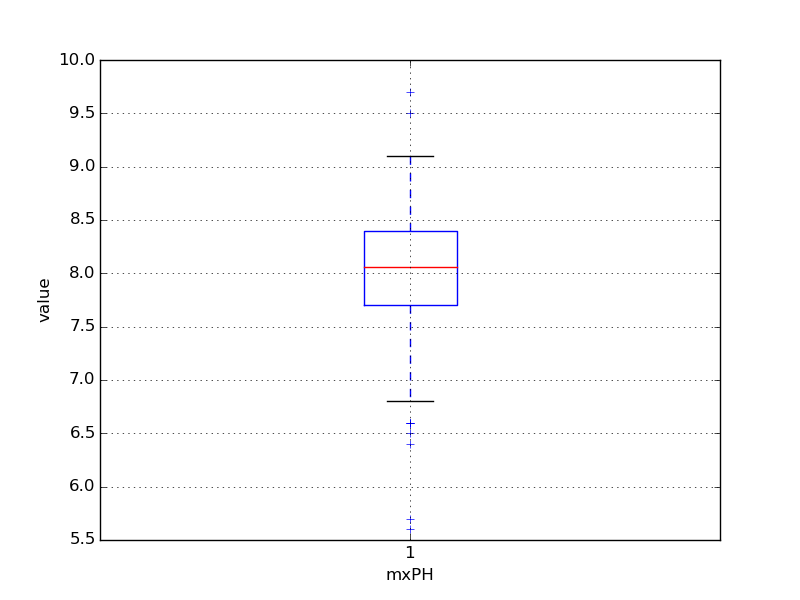
 

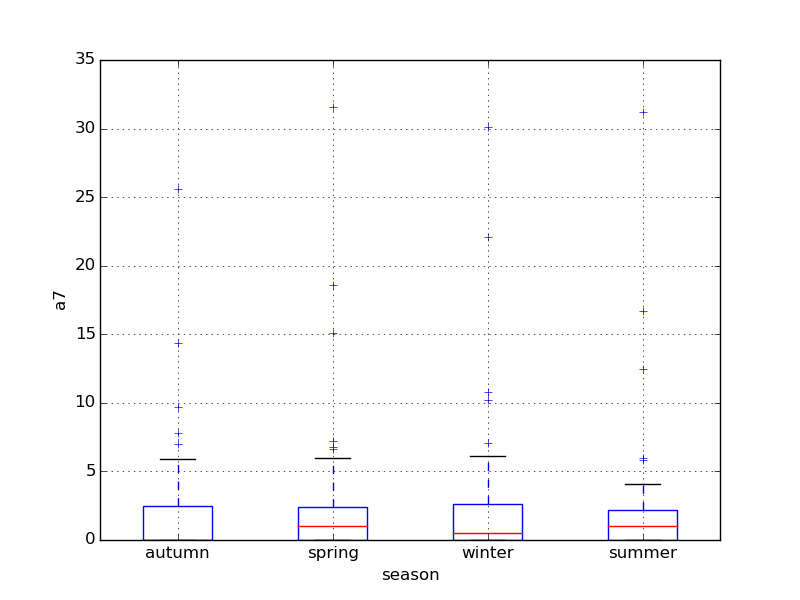
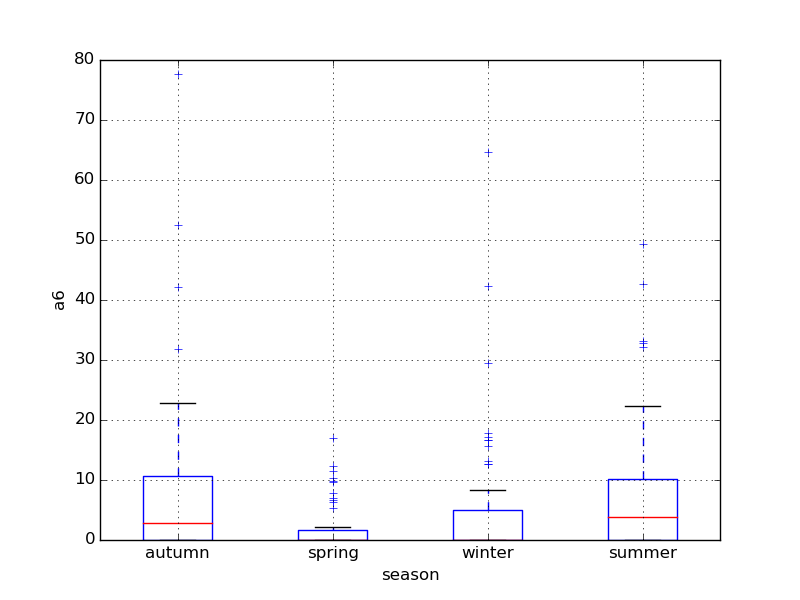
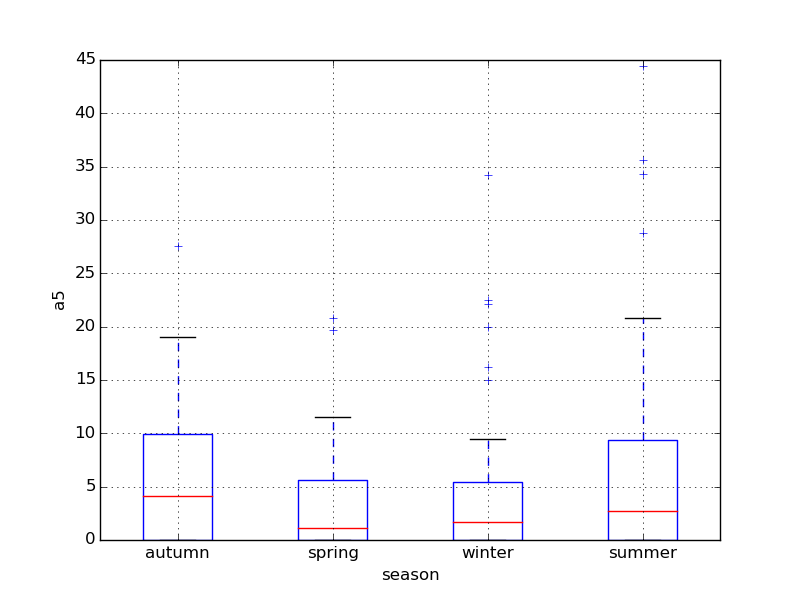
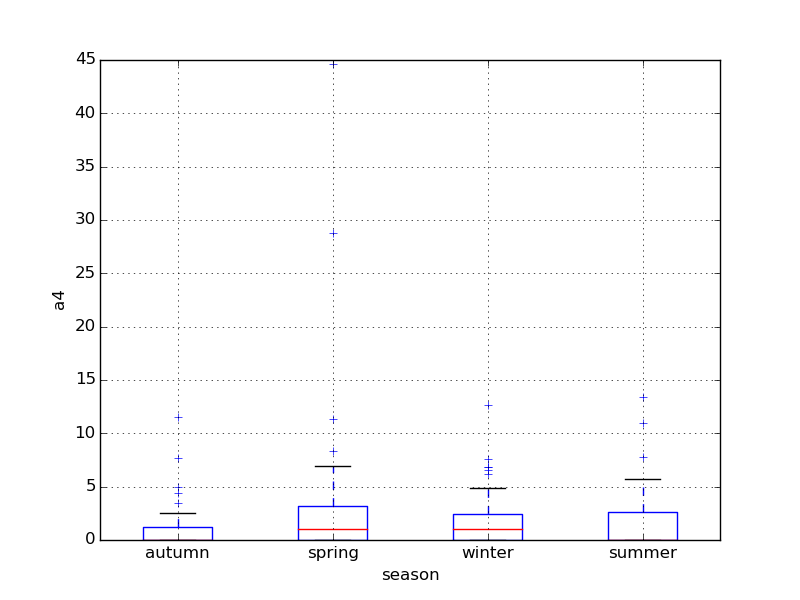
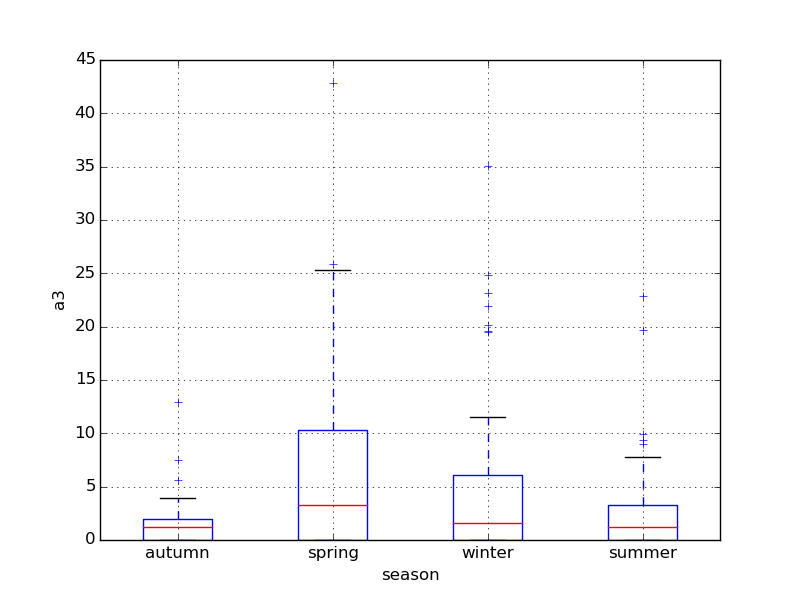
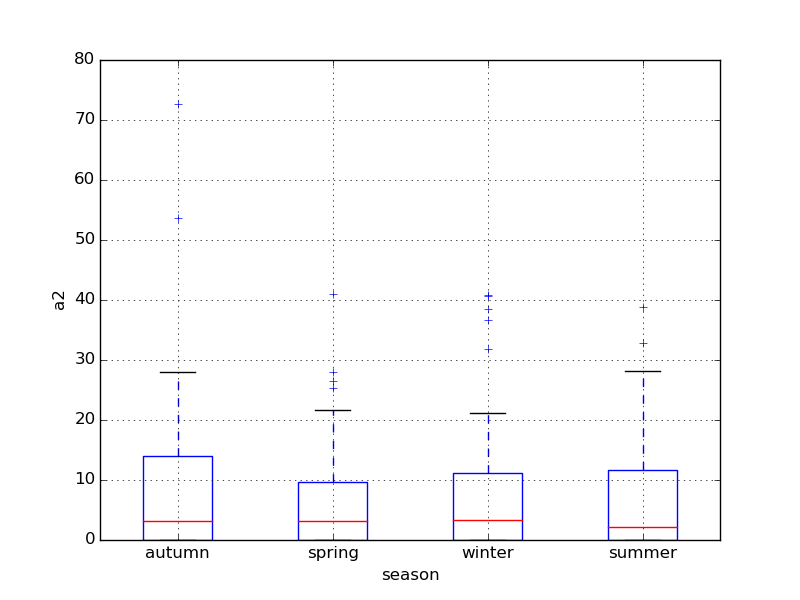
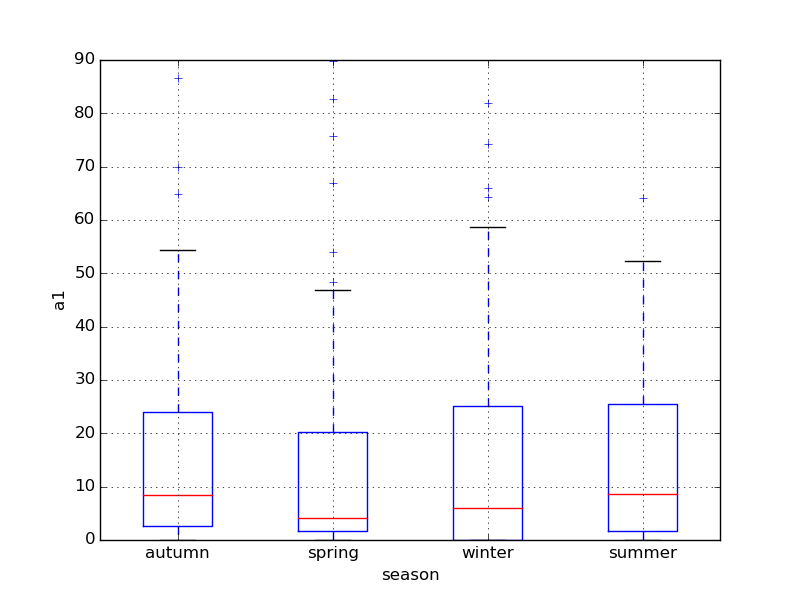
 

2、盒图

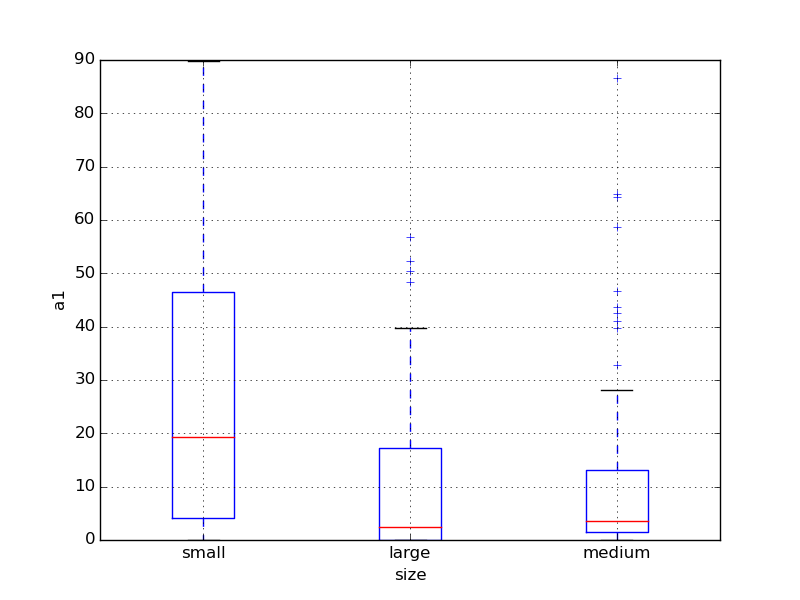
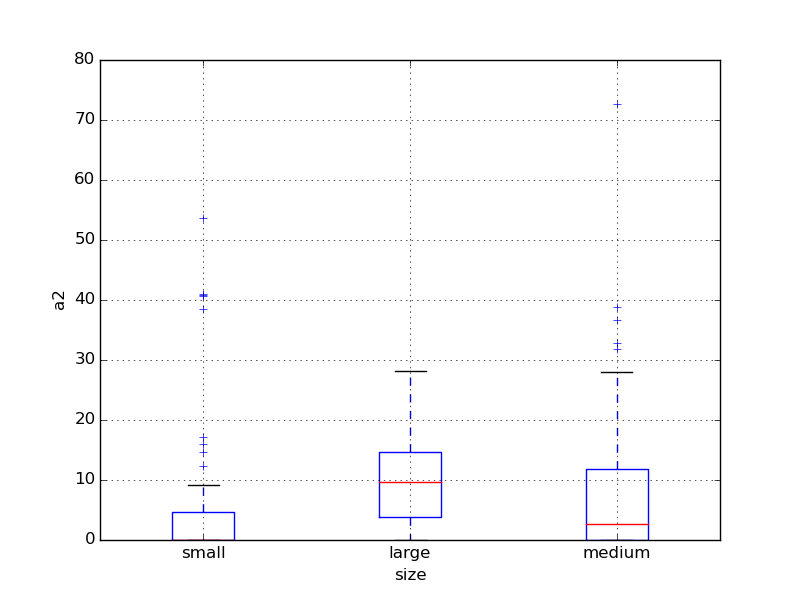
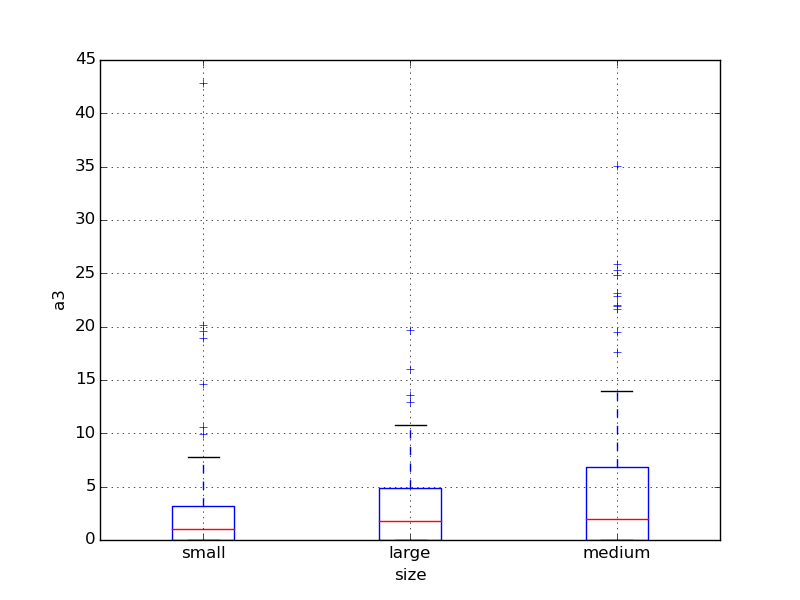
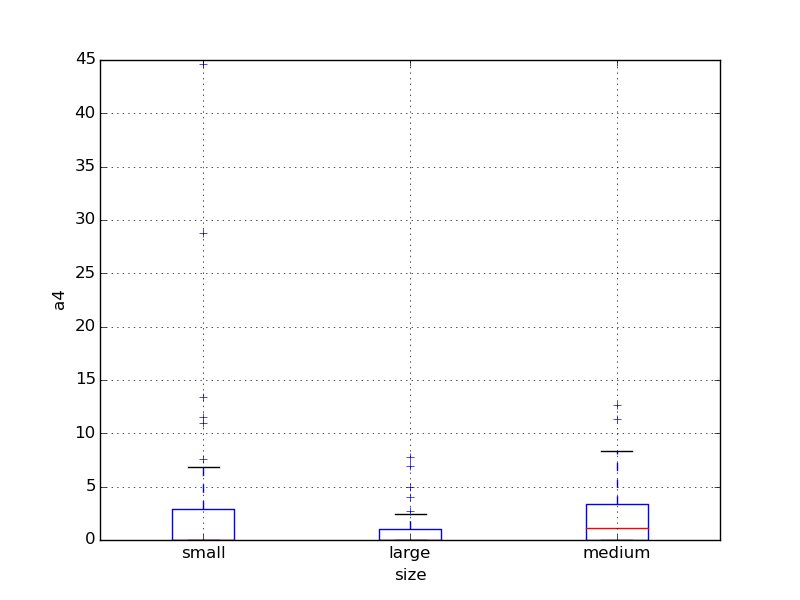
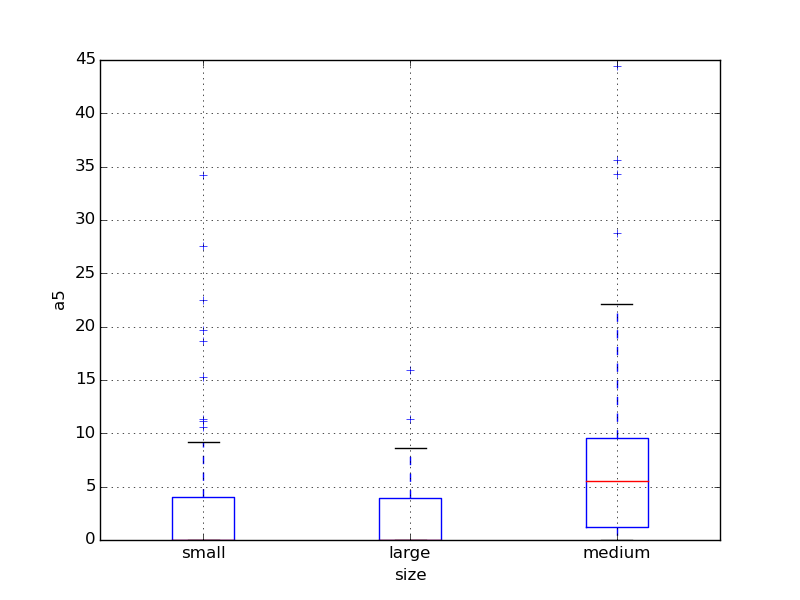
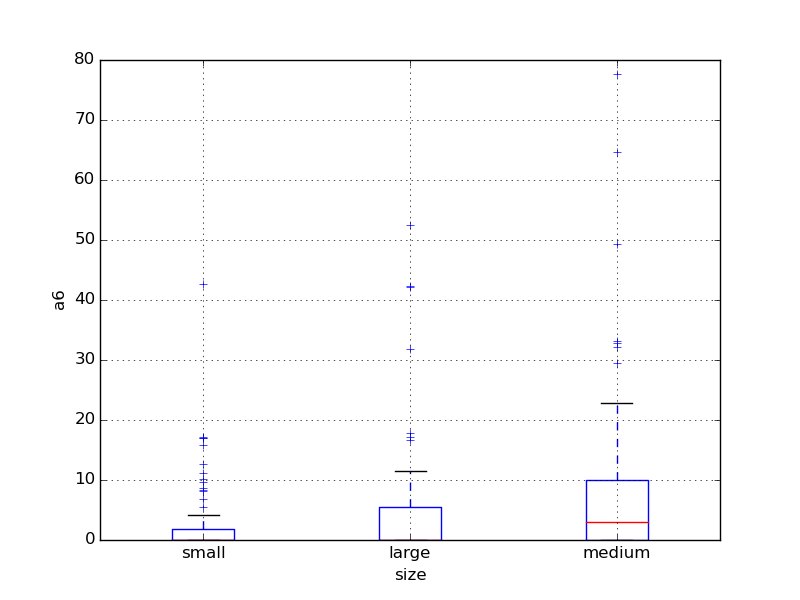
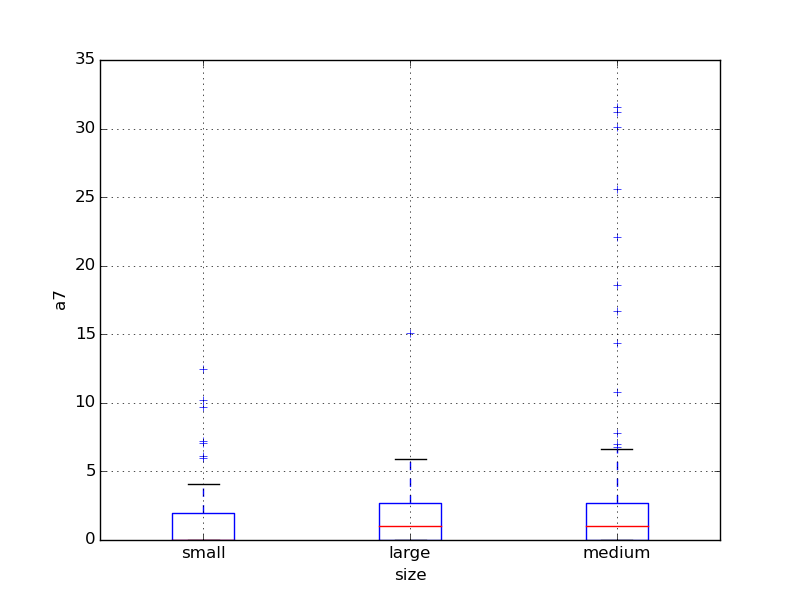


3、对7种海藻，其数量与标称变量的条件盒图

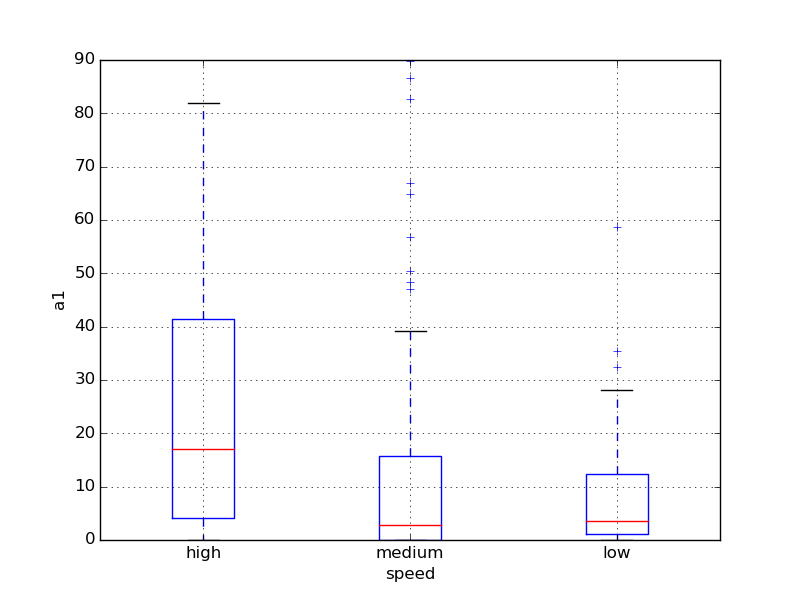
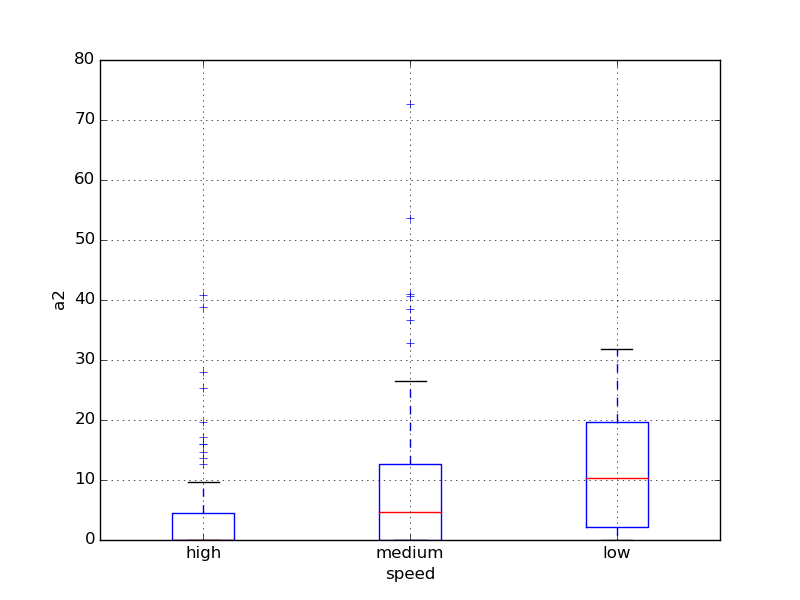
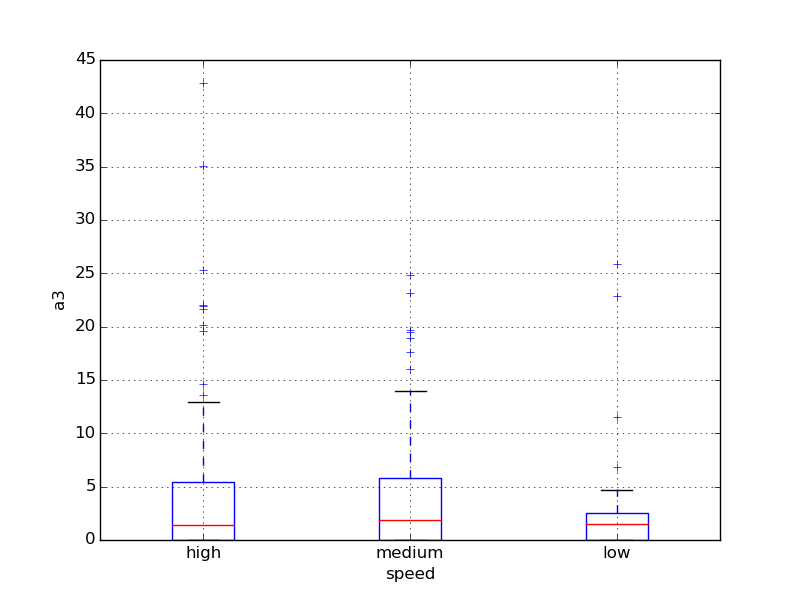
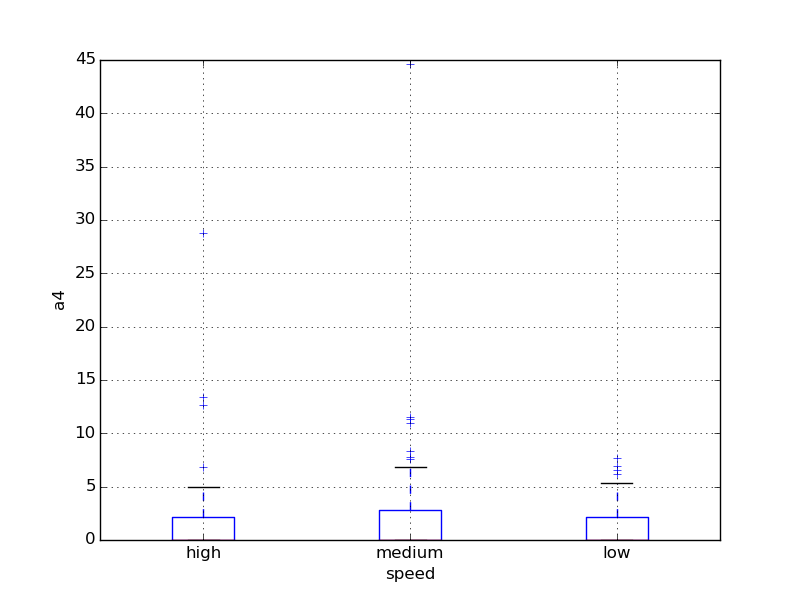
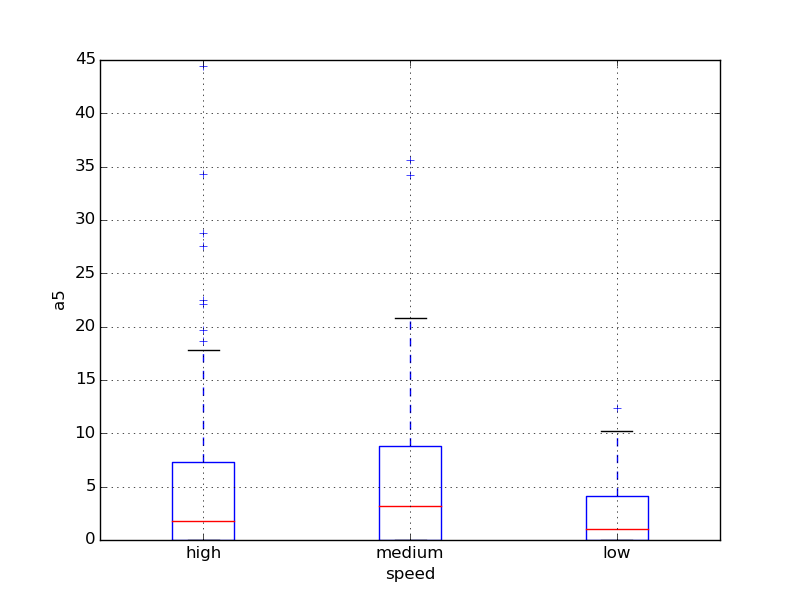
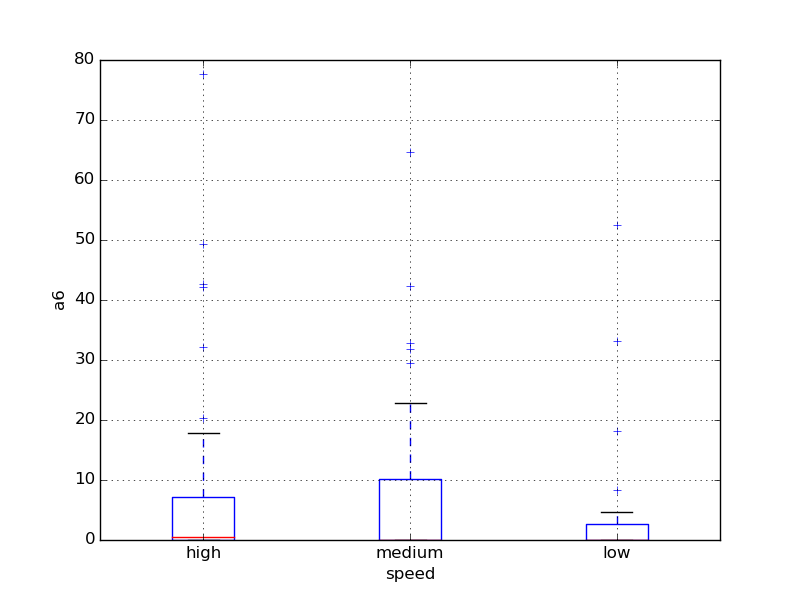
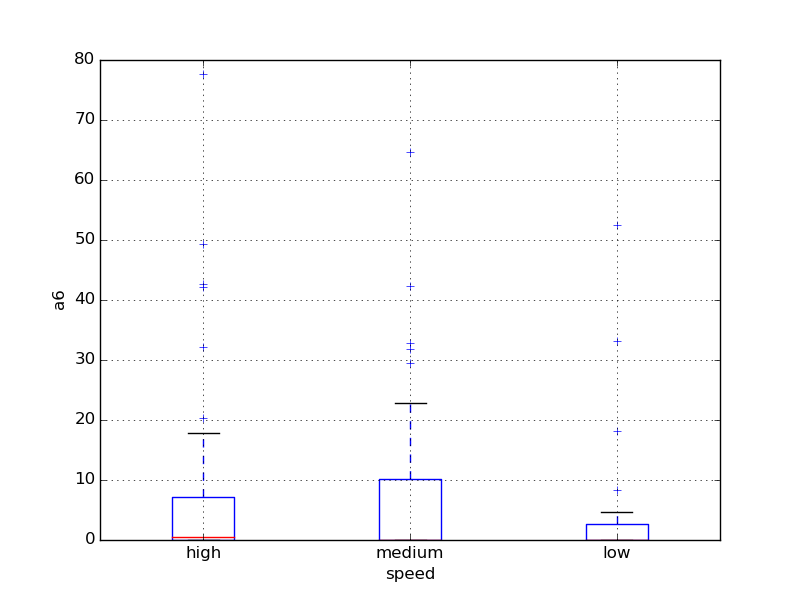
标称变量“季节”：



标称变量“河流大小”：

标称变量“河流速度”：

三、数据缺失的处理

1、将缺失部分剔除

2、用最高频率值来填补缺失值