**海藻数据分析报告**

1. 数据摘要：

Season：（autumn:40， spring:53，summer:45，winter:62）

Size：（large :45，medium:84，small :71）

Speed：（high  :84 ，low   :33，medium:83）

mxPH:（缺失值：1，平均值：8.01，最大值：9.7，最小值：5.6，四分位数：7.7,，中位数：8.06）

mnO2：（缺失值：2，平均值：9.11，最大值：13.4，最小值：1.5，四分位数：7.7,，中位数：9.8）

Cl： （缺失值：10，平均值：43.63，最大值：13.4，最小值：1.5，四分位数：7.7,，中位数：9.8）

NO3：（缺失值：2，平均值：3.28，最大值：45.65，最小值：0.05，四分位数：1.29,，中位数：2.67）

NH4：（缺失值：2，平均值：501.29，最大值：24064，最小值：5，四分位数：38.33,，中位数：103.16）

oPO4：（缺失值：2，平均值：73.59，最大值：564.59，最小值：1，四分位数：15.7,，中位数：40.15）

PO4：（缺失值：2，平均值：137.88，最大值：771.59，最小值：1，四分位数：41.37，中位数：103.28）

Chla：（缺失值：12，平均值：13.97，最大值：110.45，最小值：0.2，四分位数：2.0,，中位数：5.47）

海藻1频数：（平均值：16.92，最大值：89.8，最小值：0，四分位数：1.5,，中位数：6.95）

海藻2频数：（平均值：7.45，最大值：72.6，最小值：0，四分位数：0,，中位数：3）

海藻3频数：（平均值：4.30，最大值：42.0，最小值：0，四分位数：0,，中位数：1.55）

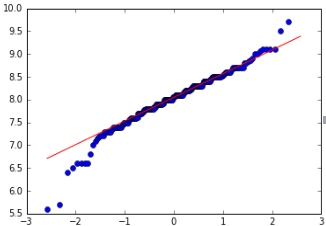
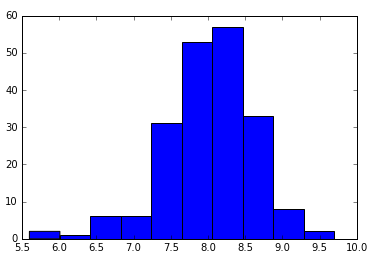
海藻4频数：（平均值：1.99，最大值：44.6，最小值：0，四分位数：0,，中位数：0）

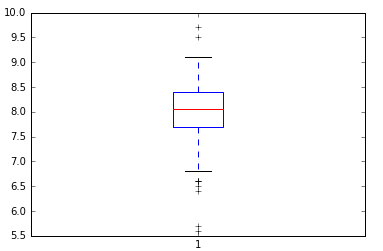
海藻5频数：（平均值：5.06，最大值：44.4，最小值：0，四分位数：0,，中位数：1.9）

海藻6频数：（平均值：5.96，最大值：77.6，最小值：0，四分位数：0,，中位数：0）

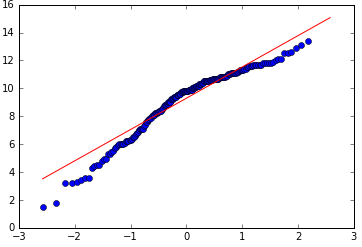
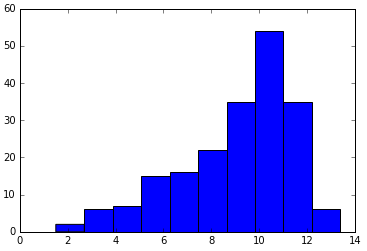
二、数据可视化

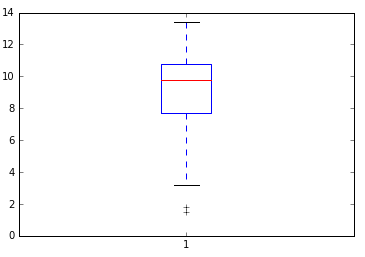
mxPH（直方图，qq图，盒图）：



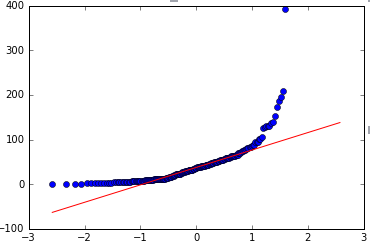
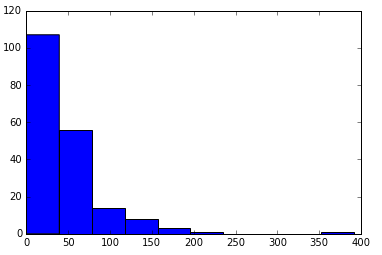


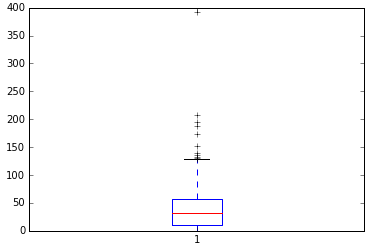
mnO2（直方图，qq图，盒图）：



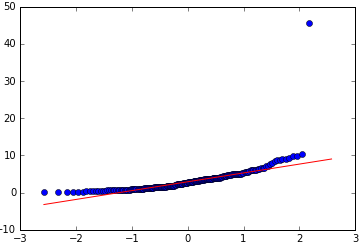
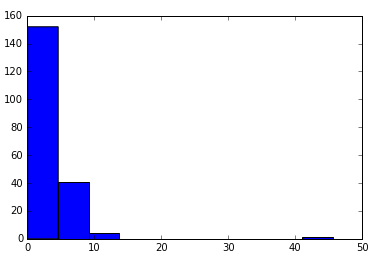


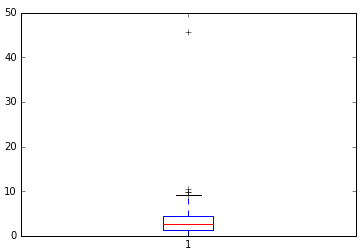
Cl:（直方图，qq图，盒图）：



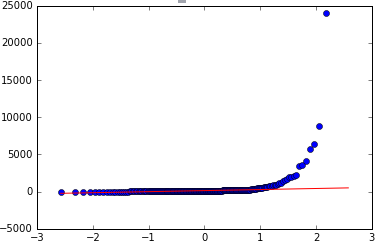
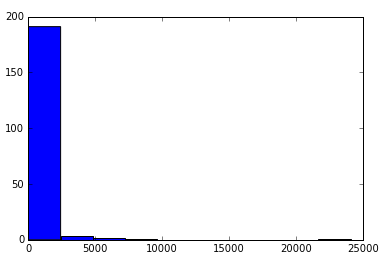


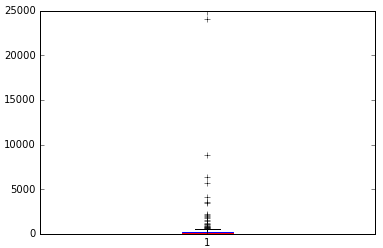
NO3:（直方图，qq图，盒图）：



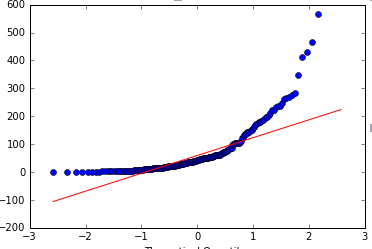
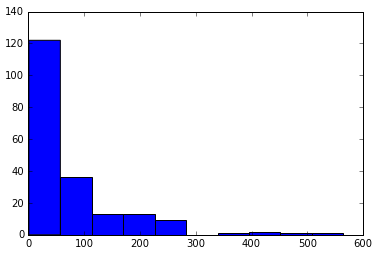


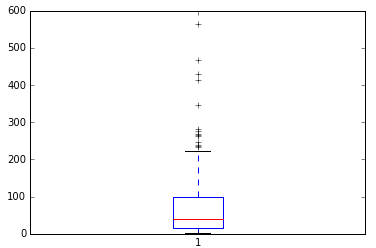
NH4:（直方图，qq图，盒图）：



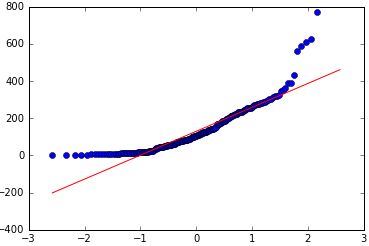
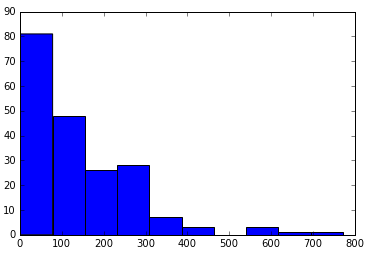


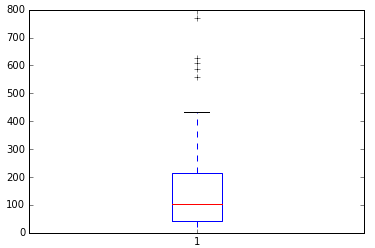
oPO4:（直方图，qq图，盒图）：



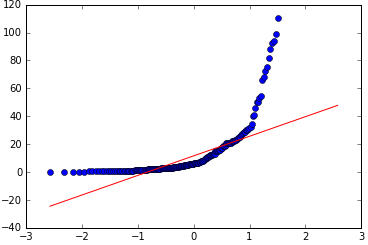
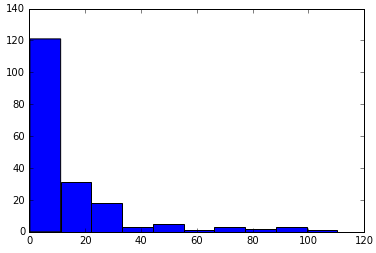


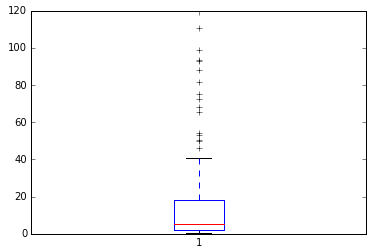
PO4：（直方图，qq图，盒图）：



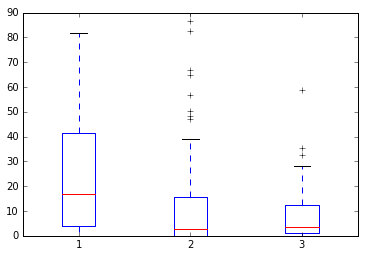


Chla：（直方图，qq图，盒图）：

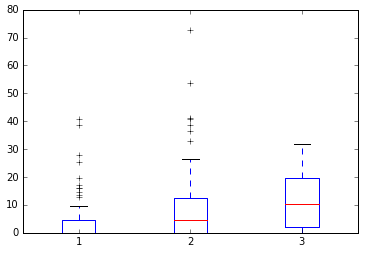




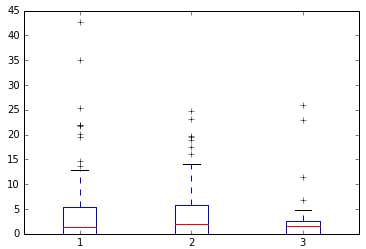
条件盒图（海藻1 河水流速 ，从左到右由高到低）



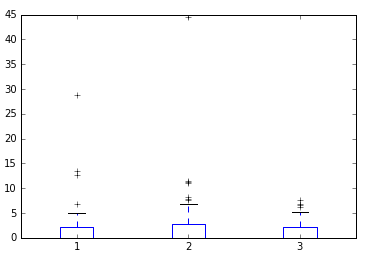
海藻2



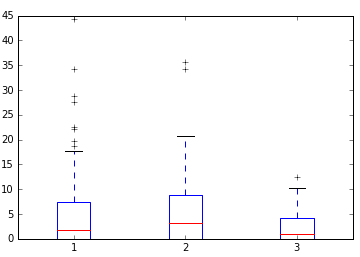
海藻3



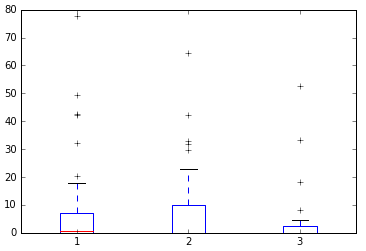
海藻4



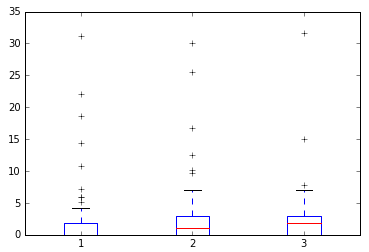
海藻5



海藻6



海藻7



三、缺失值处理

1. 对数据进行了去除缺失值处理，写入了delnull.xls中。
2. 用最高频数代替缺失值，写入了fillwithfreq.xls中。
3. 用属性的相关关系处理缺失值，大致步骤如下：

1、算出当前属性相关系数最高的属性。

2、通过相关性最高的属性进行插值。

文件最后写入了fillwithattr.xls中

（4）通过对象的相关关系处理缺失值，主要是用欧氏距离寻找最接近的对象，然后用该对象的中位数填补，最后数据写入了fillwithobj.xls