### 描述性分析

对PM2.5、PM10、SO2、CO、NO2、O3\_8h、AQI进行描述性分析：



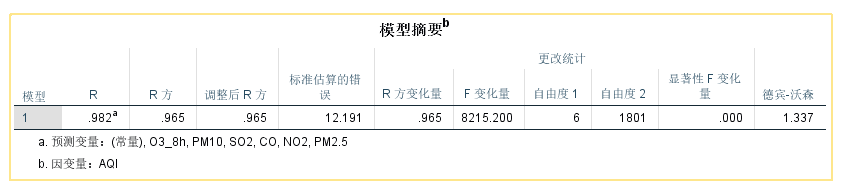
可以看出AQI、PM2.5、PM10、O3\_8h最小值有0，不符合，使用open refine对于不符合的数据进行删除，将删除后的数据继续进行描述性分析：



通过描述性分析可以看到不同变量的区间范围以及平均值，通过方差数值能够看出，除了co、SO2变量的每日的差异较小之外，其他变量每日的差异较大，通过偏度数据可以看出各个变量呈现左偏分布状态，通过峰度能看出AQI、PM2.5 、SO2、PM10呈现高峰态 CO、NO2、O3\_8h呈现低峰态。

### 线性回归分析

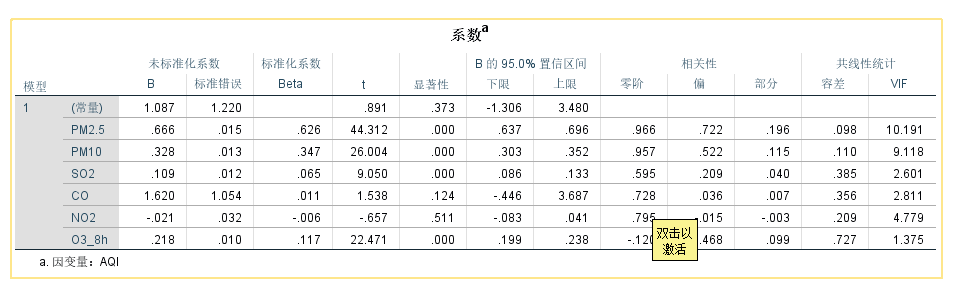
1. 将AQI 定义为因变量，PM2.5、PM10、SO2、CO、NO2、O3\_8h定义为自变量，进行线性回归分析
2. 根据调整后的R方为1.337>0.6，初步判断模型拟合效果良好

* 

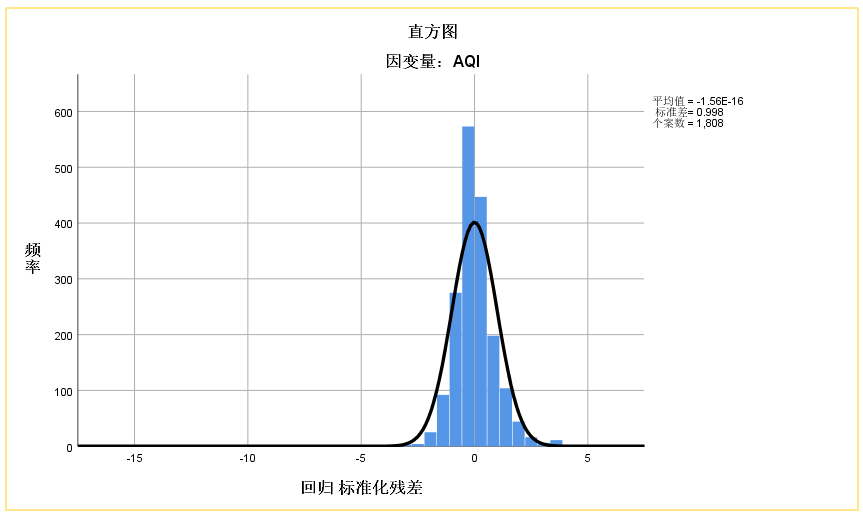
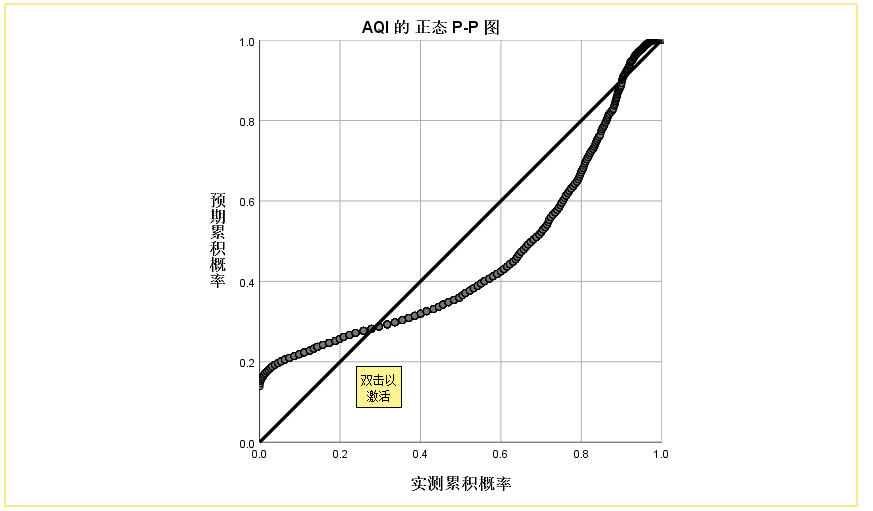
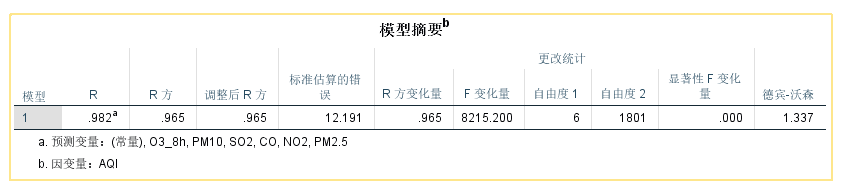
1. 方差分析的显著性值=0.000<0.01<0.05,表明由自变量PM2.5、PM10、SO2、CO、NO2、O3\_8h和因变量AQI建立的线性回归模型具有极显著的统计学意义，若自变量PM2.5、PM10、SO2、CO、NO2、O3\_8h的数值同时增加，可增加AQ的值的线性关系显著



1. 由下图为标准化系数可知模型表达式A=1.087+0.666B+0.328C+0.109D+1.620E-0.021F+0.218G,t检验原假设回归系数没有意义，PM2.5、PM10、 SO2、O3\_8h的显著性小于或等于0.01，说明这些自变量与因变量之间是正比相关，并且显著

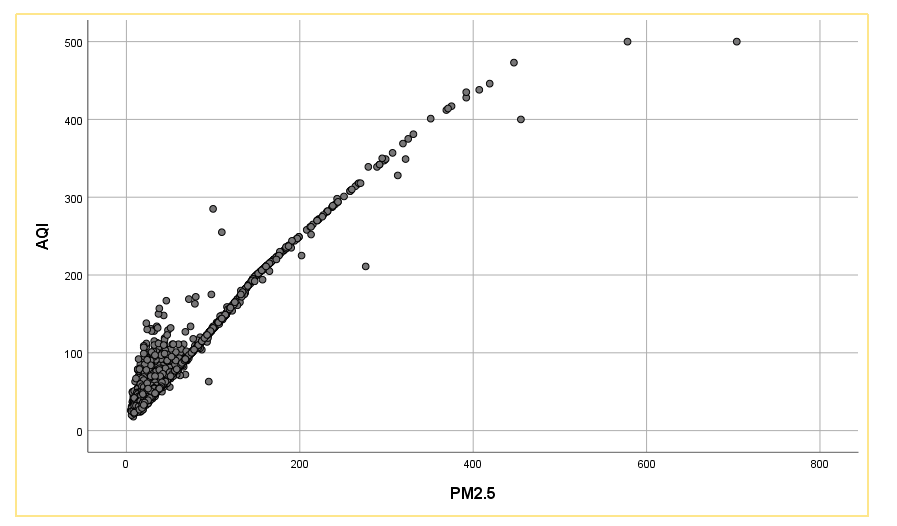
* 

1. 从标准化残差直方图来看，左右基本对称，从标准化残差的P-P图来看，两条线并不完全靠近，综合而言数据做回归分析虽不理想但可以接受

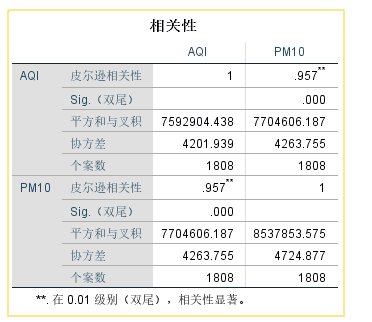
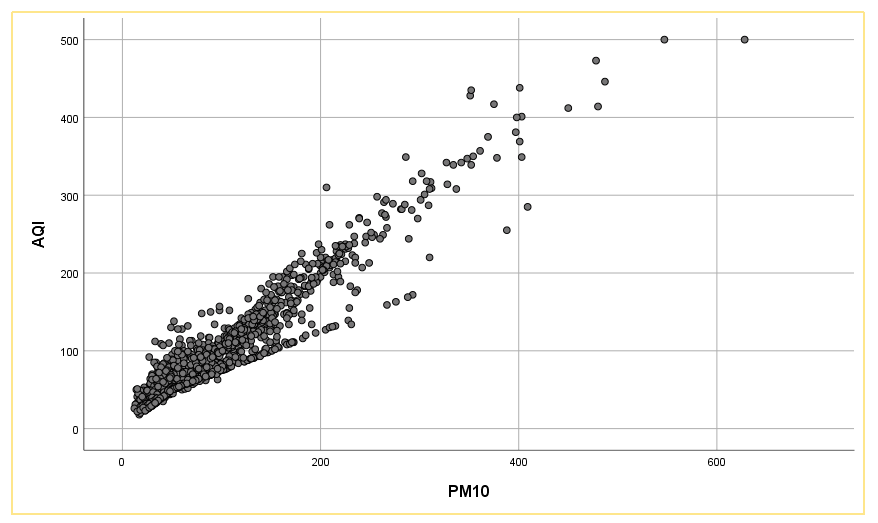
* 
* 
* 6.模型残差独立性检验。DW=1.337，查询Durbin Watson Table可知，认定残差独立，通过检验
* 

### 相关性分析

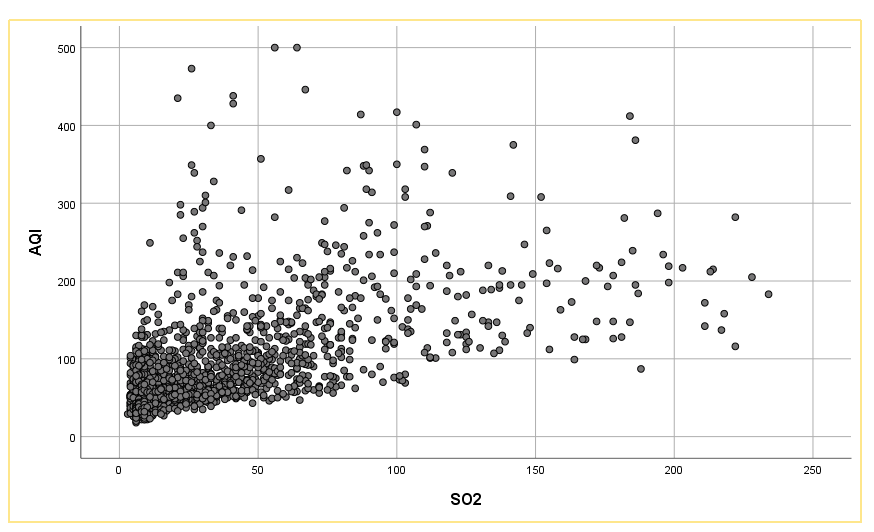
1. 对AQI与PM2.5进行双变量相关性分析，相关性显著，两个变量的值在100以内时先相关性要略小，变量的值大于100时，基本呈正相关

* 
* 

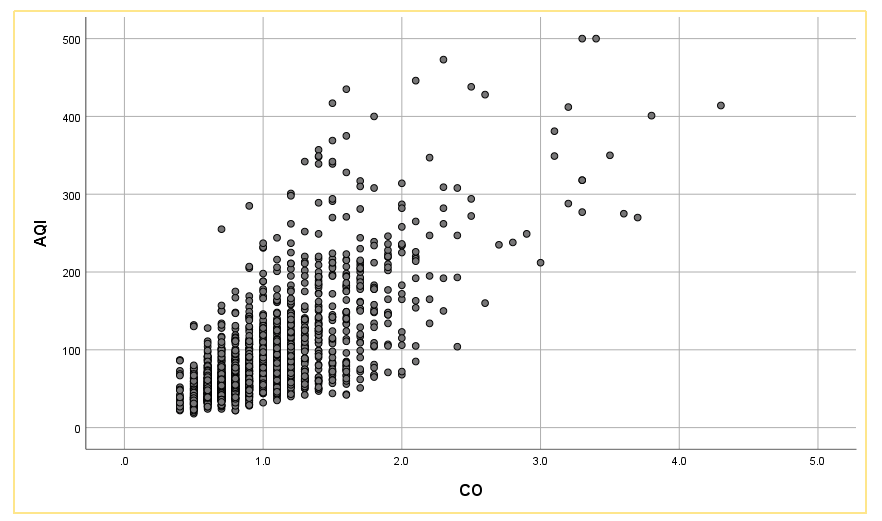
1. 对AQI与PM10进行双变量相关性分析，变量值在200以内，相关性显著，变量值超过200后相关性下降

* 
* 

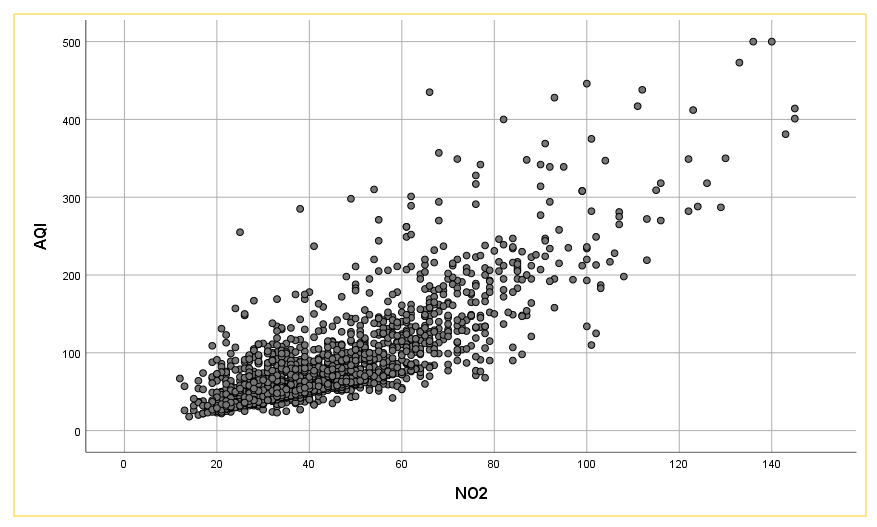
1. 对AQI与SO2进行双变量相关性分析，变量值小于200时相关性显著

* 
* 

1. 对AQI与CO进行双变量相关性分析,相关性显著

* 
* 

1. 对AQI与NO2进行双变量相关性分析,相关性显著

* 
* 

1. 对AQI与O3\_8h进行双变量相关性分析,相关性显著

* 
* 