数据文件Beijing.csv包含了北京某大气污染监测站点从2014年1月1日0点至2015年12月31日23时所收集的六个大气污染物和天气变量的逐小时数据。变量名所代表的具体含义请参看下面的code book。

请用这份数据做一份详细的数据分析报告。请使用你在“时间序列分析”这门课上所学的技术，除此以外也可以加上你在回归分析、多元分析、统计计算、机器学习、非参数统计等课上所学得的分析与建模技术。你的报告应至少包括以下几部分内容：1、简介（请自行查找六大污染物，即PM2.5，PM10，SO2，NO2，CO和O3的含义，以及它们与天气变量之间可能的关系）；2、数据探索分析及讨论，主要讨论你所使用模型的动机；3、使用ARIMA模型拟合数据的结果及分析；4、至少10个时间点的预报及预报效果分析（可以是10个小时，也可以是10天，也可以是10周）；5、结论及讨论。

注：1、你可以使用全部两年的数据，也可以只提取其中一部分来分析，但时间跨度不能少于3个月。

2、 你可以将所有变量都放到一个模型中分析，也可以在一个模型中只分析一部分变量。

2、时间尺度可以是原始数据的hour，也可以是day或者week，但不能用比week更粗的尺度。

3、不论你使用何种技术对数据进行分析，需要写明使用该技术的理由/动机。

4、你可以使用R，也可以使用任何你熟悉的统计软件。你的报告必须附上你的程序以及相应的输出结果。程序的关键步骤需要注释，调用的程序包必须写明。我们会随机抽查程序，如果结果不能重现（reproducible），则视为抄袭。

5、你可以使用中文或者英文撰写你的报告，报告首页必须注明小组各成员的具体分工。

Code Book：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year  年 | Month 月 | Day  天 | Hour  时 | PM2.5  细颗粒物 | PM10  颗粒物 | SO2  二氧化硫 | NO2  二氧化氮 | O3  臭氧 |  | TEMP | PRES | RAIN | HUMI | wd | DEWP |
| CO | TEMP | PRES | RAIN | HUMI | DEWP | WD | WS |

一氧化碳 地表温度 地表气压 降水量 湿度 露点温度 风向 风速