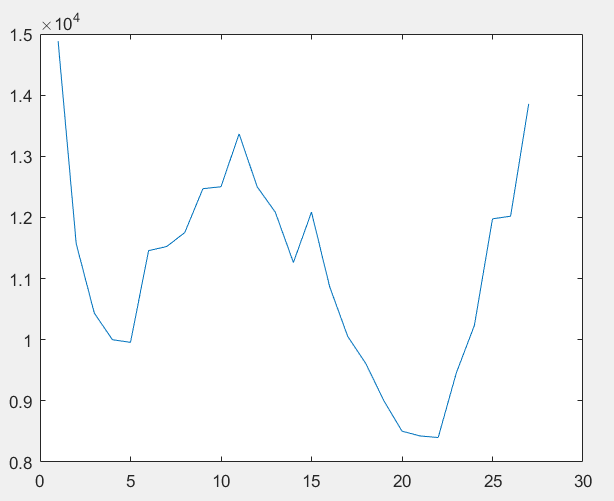
### ARIMA时间序列预测销售额算法

## 时间序列的获取

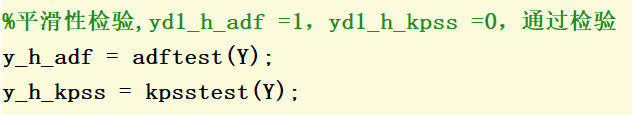
为了预测销售额，第一步就是需要进行数据筛选和获取，需要的数据是餐厅的每日总销售额和相应的时间节点，并将此做为时间序列数据。以下为源数据的销售额随时间变化图：



**图1 销售额变化图**

## 时间序列平稳化

使用ARIMA时间序列模型要求时间序列必须是平稳的，所以第一步是对原始数据进行平稳性检验。检验方法这里采用ADF检验和KPSS检验。



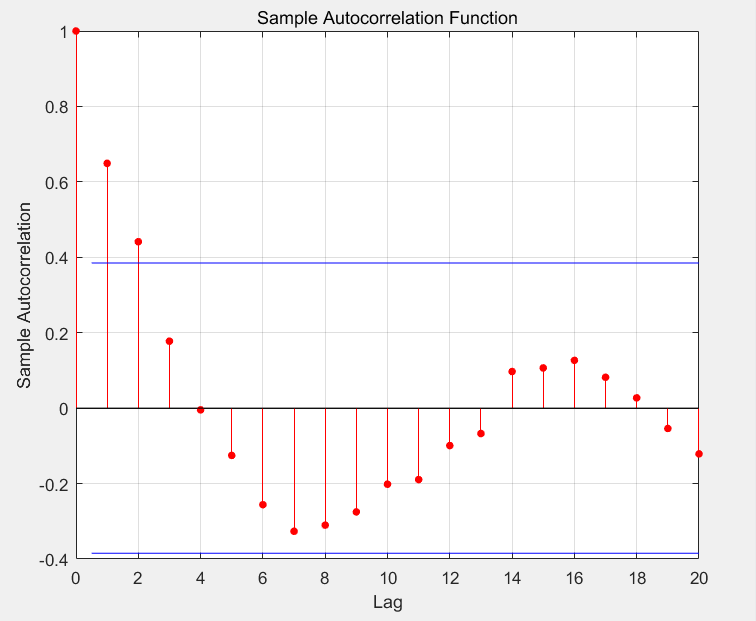
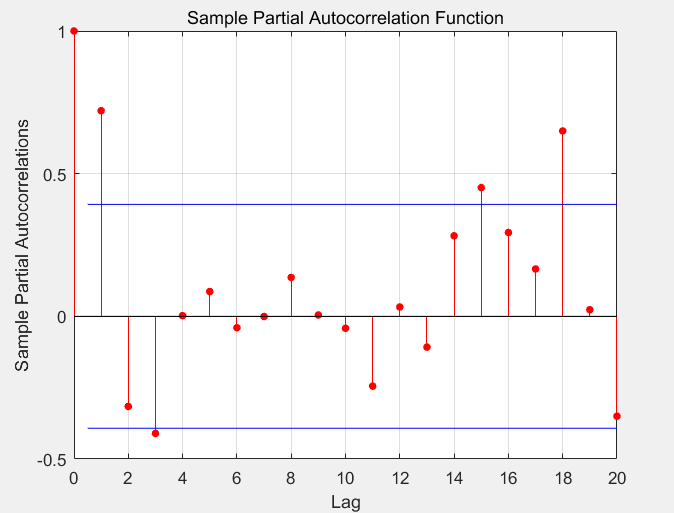
如果没有通过检验则需要对Y进行去差分，一阶差分后如果平稳则确定数据，如果不平稳则再取二阶差分，如此反复直至数据达到平稳。该步骤确定了模型ARIMA（p ,q , i）中的i一阶则对应i=1。

由此将非平稳的时间序列源数据转化成了平稳的时间序列源数据。

## 模型识别

模型识别主要是为了确定ARIMA模型的阶数，对应ARIMA（p ,q , i）中的p和q。这里采用的模型阶数获取方法是ACF和PACF法，即自相关函数和偏自相关函数。

根据数据画出的ACF和PACF图如下所示：

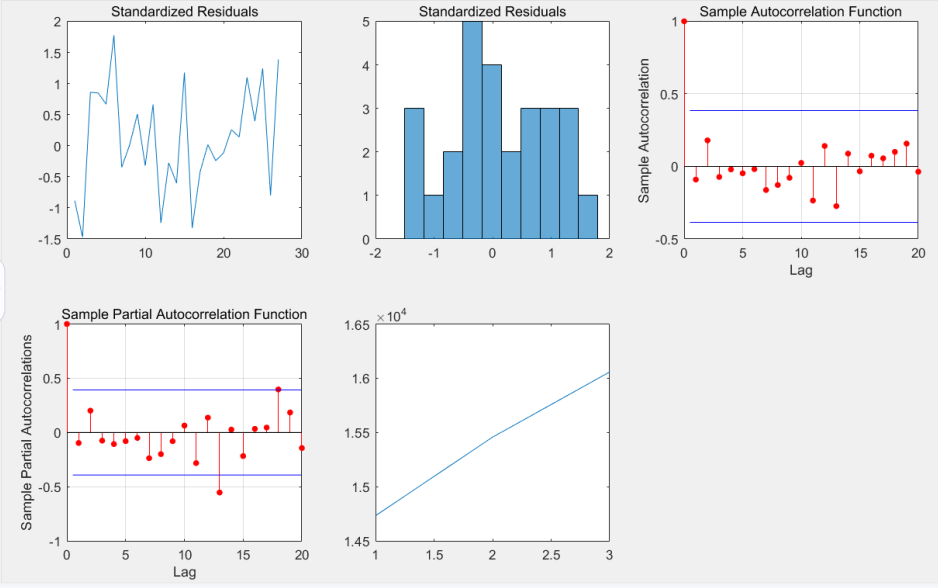
 

**图2 ACF自相关图**  **图3 PACF偏自相关图**

目测可以发现其中q等于2，p等于18，相对应对应ARIMA模型中的ARIMA（p，q，i）中的p和q。

## 残差检验

为了确保确定的阶数合适，还需要进行残差检验。残差即原始信号减掉模型拟合出的信号后的残余信号。如果残差是随机正态分布的、不自相关的，这说明残差是一段白噪声信号，也就说明有用的信号已经都被提取到ARMA模型中了。以下为得出的残差检验图：

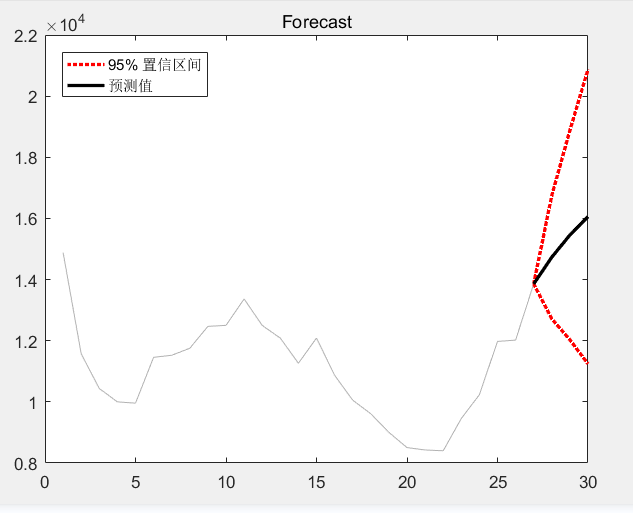


**图4 残差检验图**

其中Standardized Residuals图是查看残差是否接近正态分布，理想的残差要接近正态分布，除此之外还画出了直方图进行展示；ACF和PACF检验残差的自相关和偏自相关，理想的结果应该在图中不存在超出蓝线的点；最后一张QQ图是检验残差是否接近正太分布的。

## 模型预测

这里对餐厅销售额的前27天做为源数据，预测餐厅最后3天的销售额，即设置预测步数为3步，最终的结果图如下所示：



**图5 ARIMA预测结果图**

上图中灰线为用来训练的27个数据，黑线为未来值的预测，红线为95%置信区间上下限。也就是说未来真实值有95%的概率落在这个范围内。除此之外，可以看到使用ARIMA方法进行长期预测的结果是趋势性的。

由此便完成了基于ARIMA时间序列的餐厅销售额预测，预测餐厅销售额最后三天数据分别是14733、15457和16058，观察可发现与源数据相差不大，符合价格趋势。（具体可见transaction\_data.csv文件）。