**异常值：**

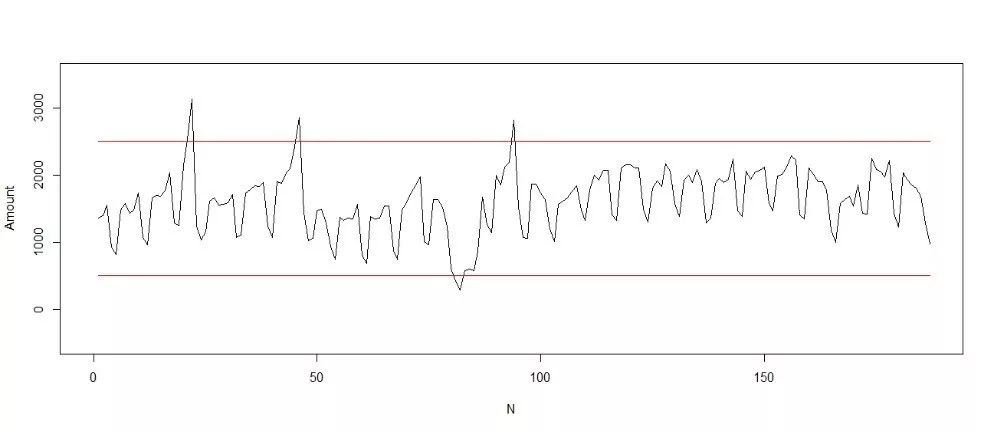
1. **判断**
   * 判断方法
2. 恒定阈值
3. 动态阈值（移动窗口）
4. 迭代阈值（每次选偏移值最大的一个异常值，剩下的重新计算阈值）
5. 差分序列取阈值
6. 时间序列分解法，对残差序列取阈值

* 阈值

1. （近似符合正态分布）在异常值检测时，一般认为偏离3倍标准差以上为异常值
2. （不符合正态分布）箱线图检测异常值，异常值的标准以四分位数和四分位距为基础
3. **处理:**
   * 删除
   * 替换
     + 众数，平均数，阈值，预测函数补齐....?
   * 分离，‘ 对异常值数据集进行进一步的数据挖掘
     + 周期性
     + 趋势性

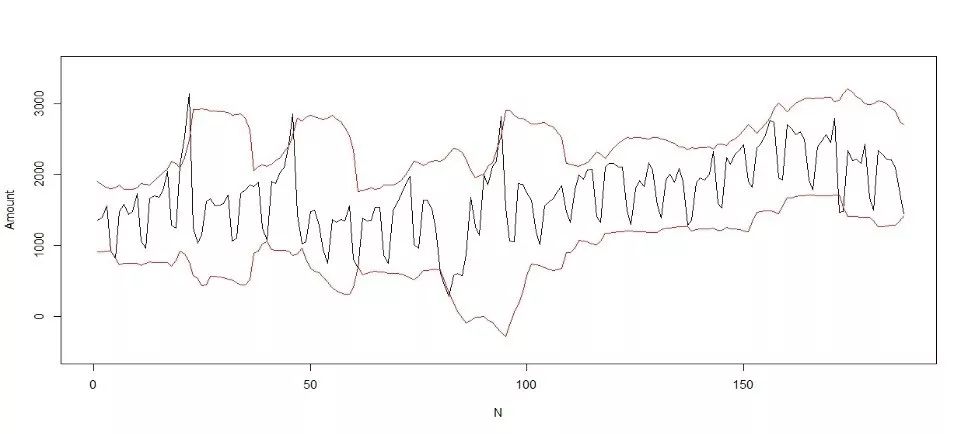
a配置恒定阈值

数据无趋势性，近似平稳，可配置简单的恒定阈值。时刻T配置的恒定阈值是对历史数据的恒定，在T+1时刻，这个阈值会被新加入的数据更新。



b、配置动态阈值

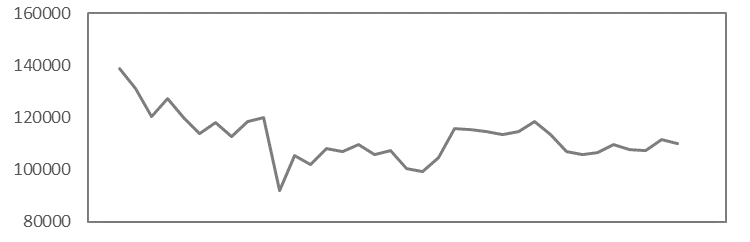
如果时间序列含有趋势性，但无明显周期性，可以配置动态阈值。比如基于固定的移动窗口计算移动平均值和移动标准差，基于两者给出监控的上下界。动态阈值会受到移动窗口大小设定的影响，对判定当前数据异常有一定的延迟性。



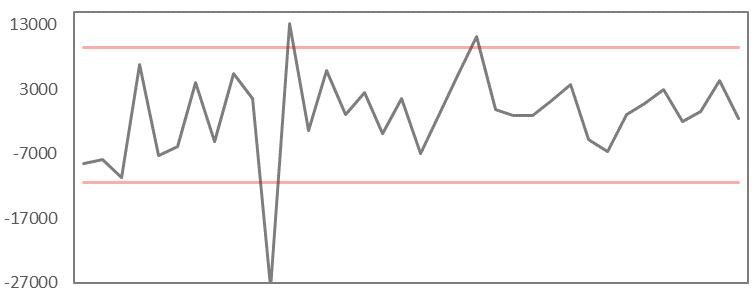
c、监控差分序列

对原始序列作一阶差分，如果差分序列稳定，可对差分序列配置恒定阈值，从而判定原序列的异常情况。

原始序列：

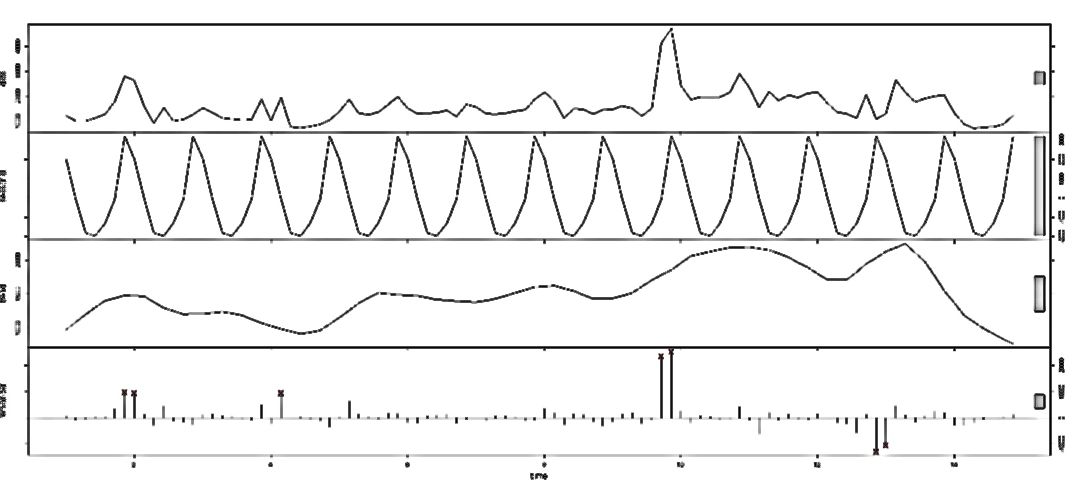


差分序列：



d、时间序列分解法

如果业务数据既有趋势性又有周期性，可将时间序列模型运用于监控任务中，如Arima，STL，TBATS等时间序列模型。在STL鲁棒加权回归时间序列分解法中，模型通过加权最小二乘回归将原始序列分解成周期序列，趋势序列和残差序列。下图从上到下依次是原始序列，周期序列，趋势序列和残差序列。



可对残差序列配置恒定阈值