**销售额预测用到的ARIMA模型**

ARIMA模型（英语：Autoregressive Integrated Moving Average model）， 差分整合移动平均自回归模型，又称整合移动平均自回归模型（移动也可称作滑动），[时间序列](https://baike.baidu.com/item/%E6%97%B6%E9%97%B4%E5%BA%8F%E5%88%97" \t "https://baike.baidu.com/item/ARIMA%E6%A8%A1%E5%9E%8B/_blank)预测分析方法之一。ARIMA（p，d，q）中，AR是"自回归"，p为自回归项数；MA为"滑动平均"，q为滑动平均项数，d为使之成为平稳序列所做的差分次数（阶数）。“差分”一词虽未出现在ARIMA的英文名称中，却是关键步骤。

ARIMA（p，d，q）模型是[ARMA](https://baike.baidu.com/item/ARMA" \t "https://baike.baidu.com/item/ARIMA%E6%A8%A1%E5%9E%8B/_blank)（p，q）模型的扩展。ARIMA（p，d，q）模型可以表示为：



其中*L*是滞后算子（Lag operator），

## **ARIMA模型原理**

**自回归模型AR**

描述当前值与历史值之间的关系，用变量自身的历史时间数据对自身进行预测，自回归模型必须满足平稳性的要求。

p阶自回归过程的公式定义：



yt是当前值 u是常数项 P是阶数 ri是自相关系数 et是误差。（P当前值距p天前的值的关系）

**自回归模型的限制**

1、自回归模型是用自身的数据进行预测。

2、必须具有平稳性。

3、必须具有相关性，如果自相关系数（φi）小于0.5，则不宜采用。

4、自回归只适用于预测与自身前期相关的现象。

**移动平均模型MA**

移动平均模型关注的是自回归模型中的误差项的累加。

q阶自回归过程的公式定义：



移动平均法能有效地消除预测中的随机波动。



I是差分模型，需要确定P和Q，d是做几阶差分,一般1阶就可以了。

**选择P与Q的方法：**

**自相关函数ACF（autocorrelation function）**

有序的随机变量序列与其自身相比较自相关函数反映了同一序列在不同时序的取值的相关性

公式：

****

变量与自身的变化，yt和yt-1到yt和yt-k的相关系数，k阶滞后点，Pk的取值范围[-1，1]

1、对于一个平稳AR（p）模型，求出滞后k自相关系数p（k）时，实际上得到的并不是x（t）与x（t-k）之间单纯的相关关系。

2、x（t）同时还会受到中间k-1个随机变量x（t-1）、x（t-2）……、x（t-k+1）的影响，而这k-1个随机变量又都和x（t-k）具有相关关系，所以自相关系数p（k）里实际掺杂了其他变量对x（t）与x（t-k）的影响。

3、剔除了中间k-1个随机变量x（t-1）、x（t-2）、……、x（t-k+1）的干扰之后。x（t-k）对x（t）影响的相关程度。

4、ACF还包含了其他变量的影响，而偏自相关系数PACF是严格这两个变量之间的相关性。

## **时间序列分析**

时间序列分析是根据系统观察得到的时间序列数据，通过曲线拟合和参数估计来建立数学模型的理论和方法。时间序列分析常用于国民宏观经济控制、市场潜力预测、气象预测、农作物害虫灾害预报等各个方面。

## **建模基本步骤**

1. 获取被观测系统时间序列数据；
2. 对数据绘图，观测是否为平稳时间序列；对于非平稳时间序列要先进行d阶差分运算，化为平稳时间序列；
3. 经过第二步处理，已经得到平稳时间序列。要对平稳时间序列分别求得其自相关系数ACF 和偏自相关系数PACF ，通过对自相关图和偏自相关图的分析，得到最佳的阶层 p 和阶数 q。
4. 由以上得到的d、q、p ，得到ARIMA模型。然后开始对得到的模型进行模型检验。

## **模型检验与预测**

1. 在指数平滑模型下，观察ARIMA模型的残差是否是平均值为0且方差为常数的正态分布（服从零均值、方差不变的正态分布），同时也要观察连续残差是否（自）相关。
2. 我们对ARMA模型所产生的残差做自相关图。
3. 德宾-沃森（Durbin-Watson）检验。德宾-沃森检验,简称D-W检验，是目前检验自相关性最常用的方法，但它只使用于检验一阶自相关性。
4. 用于直观验证一组数据是否来自某个分布，或者验证某两组数据是否来自同一（族）分布。在教学和软件中常用的是检验数据是否来自于正态分布。
5. Ljung-Box test是对randomness的检验,或者说是对时间序列是否存在滞后相关的一种统计检验。
6. 模型确定之后，就可以开始进行预测了，我们对未来的数据进行预测。