目录

[1. 背景介绍 1](#_Toc14453)

[2. 数据说明 1](#_Toc12340)

[2.1 数据说明 1](#_Toc31490)

[2.2 案例目标 2](#_Toc18789)

[3. 数据预处理 2](#_Toc17164)

[4. 数据初步分析 2](#_Toc5140)

[5. 数据可视化 4](#_Toc25008)

[6. 售货机画像 11](#_Toc1714)

[6.1 定义概念 11](#_Toc15961)

[6.2 贴标签 11](#_Toc29390)

[6.3售货机画像 11](#_Toc5349)

[7. 业务预测 13](#_Toc15150)

[8. 小结 15](#_Toc471)

# 背景介绍

自动售货机以线上经营的理念，提供线下的便利服务，以小巧、自助的经营模式节省人工成本，让实惠、高品质的商品触手可及，成为当下零售经营的又一主流模式。自动售货机内商品的供给频率、种类选择、供给量、站点选择等是自动售货机运营者需要重点关注的问题。

因此，科学的商业数据分析能够帮助经营者了解用户需求，掌握商品需求量，为用户提供精准贴心的服务，是掌握经营方向的重要手段，对自动售货机这一营销模式的发展有着非常重要的意义。

# 数据说明

## 2.1 数据说明

某商场在不同地点安放了5台自动售货机，编号分别为A、B、C、D、E。附件1提供了从2017年1月1日至2017年12月31日每台自动售货机的商品销售数据，附件2提供了商品的分类。

附件1共70680条数据；附件2共315条数据。

## **2.2 案例目标**

1） 根据自动售货机的经营特点，对经营指标数据、商品营销数据及市场需求进行分析，完成对销量、库存、盈利三个方面各项指标的计算，按要求绘制对应图表，分析每台售货机2018年1月商品销量的预测问题。

2）为每台售货机所销售的商品贴上标签，使其能够很好地展现销售商品的特征。

# 数据预处理

1.先将数据导入jupyter notebook中，分别检查两个表的异常值，缺失值，重复值等，结果发现附件1第70679行的时间为“2017/2/29”，为异常值，予以删除；附件1的支付时间列存在缺失值，予以删除并设为索引列；附件1的“支付时间”改为datetime日期类型。

2.数据处理完之后附件1共70678条数据；附件2共315条数据

3.导入excel的Power Pivot进行处理：

1) 将附件1和附件2的“商品”字段作为连接两张表的字段;

2) 将设备ID，应付金额，状态，提现列都进行隐藏,以防止字段过多；

3) 附件1利用公式提取支付时间的月份，天数，时间，分别记为：month,day,hour；

4）新建4个度量值：“订单量”，“交易额”，“平均交易额”，“订单量排名”；

5）统计附件1中1-12月每月的天数，以计算每个月的日均订单量，以制成表格；

6）生成各售货机1-12月订单量，交易额，总交易额，平均交易额等数据透视表；

# **数据初步分析**

从图1可以5台售货机的交易额随订单量的增加而增加，并且售货机E的订单量和交易额最高，售货机D的订单量最低，售货机ABC的销量居于中间。



**表1 售货机交易总额和订单总量和交易总额**

图2中2017年5月订单量和交易额也符合全年的规律，售货机E订单量最高。



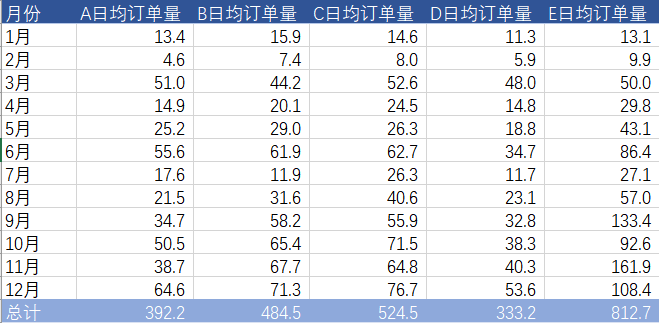
**表2 售货机2017年5月的交易额、订单量**

从图对比5 台售货机1-12月平均交易额分布各有不同，平均交易额最高的月份分别为1月，7月，5月，3月和1月。



**表3 售货机每月的每单平均交易额**

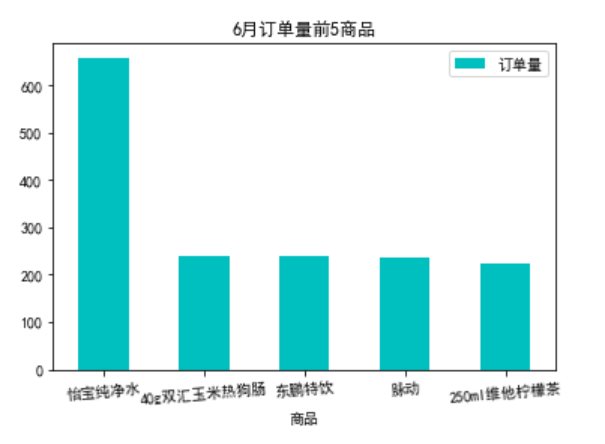
再对比5台售货机日均订单量，可以看到，2月日均订单量在所有月份中是最低的，而5台售货机最高日均订单量月份在11月和12月两个月。



**表4 售货机每月日均订单量**

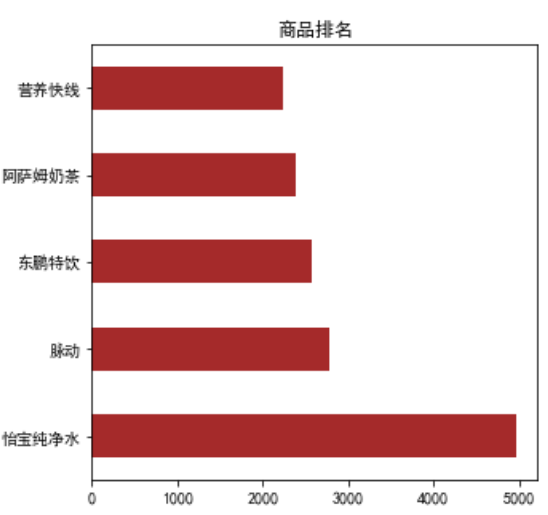
# **数据可视化**

从excel数据透视表中得到2017年6月订单量排名情况，可以看到6月订单量最高的商品是怡宝纯净水，其他4种商品订单量基本持平。值得注意的是，订单量排名前5的商品中，4中商品是水或者饮料。



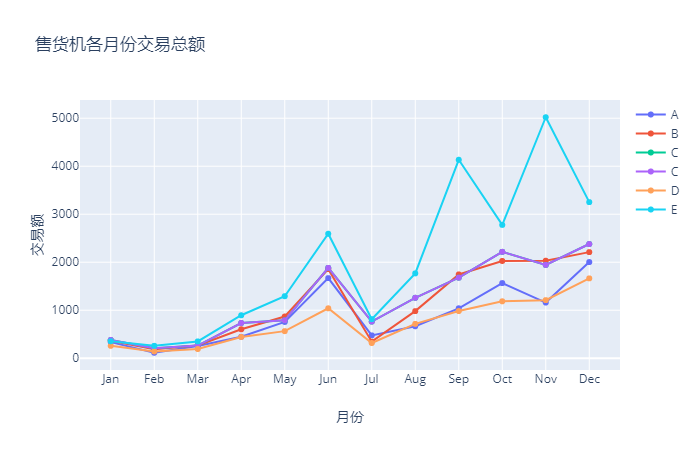
**图1 6月订单量前5商品**

图1，图2对比，可以看到6月订单量前5的商品中，有3种商品：怡宝纯净水，脉动，东鹏特饮，在所有商品订单量排名中也是位列前5。由此可见，这3种商品深受消费者欢迎。



**图2 订单量前5商品**

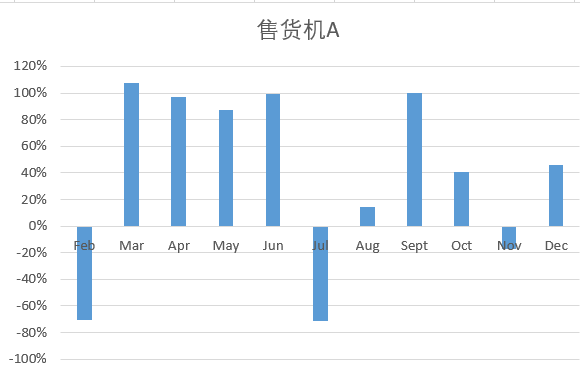
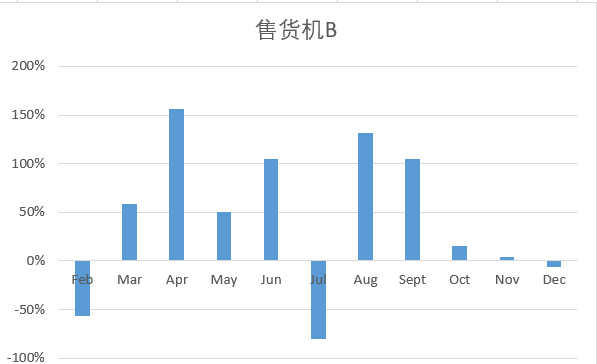
根据图3售货机各月份交易总额可以看出售货机E的交易额从6月份开始呈现波动上升趋势，且上升幅度远大于其他四台售货机。其他四台售货机交易额趋势基本相同，呈现上升趋势，且幅度较小。



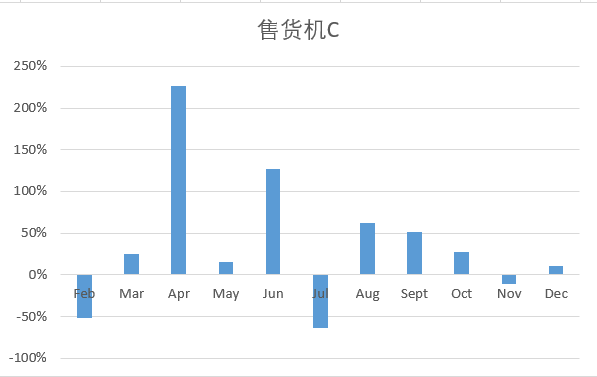
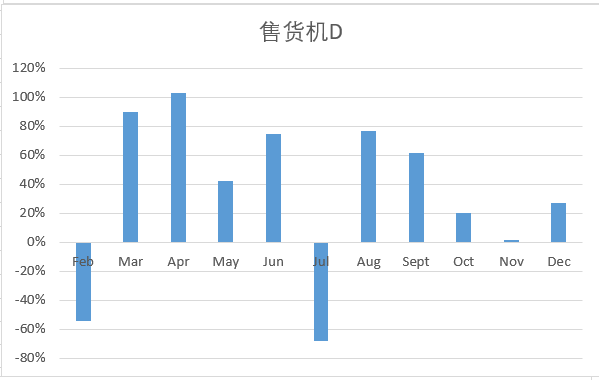
**图3 售货机各月份交易总额**

对比观察5台售货机交易额月环比增长率柱状图，可以知道，售货机A的增长率最稳定，与之相反，最不稳定的是售货机C，售货机E呈现负增长率的月份最多。

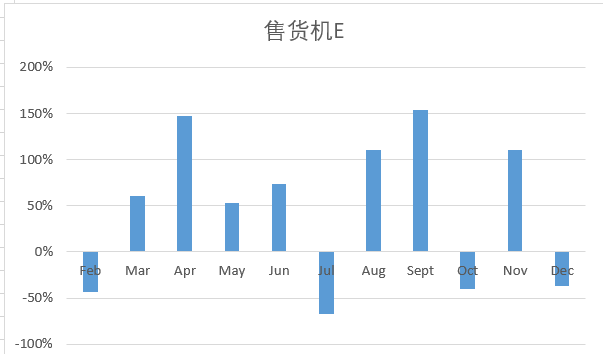
售货机A与售货机D的月环比增长率基本相同：售货机A 有3个月呈现负增长率，且保持在80%以内，售货机D有2个月呈现负增长率，同样在80%以内；售货机A有5个月的增长率在80%以上，售货机D有2个月保持80%的增长率。售货机B和售货机E都有三个月的负增长率，而正增长率的月份表现却不尽相同。

**图4 售货机A交易额月环比增长率 图5 售货机B交易额月环比增长率**

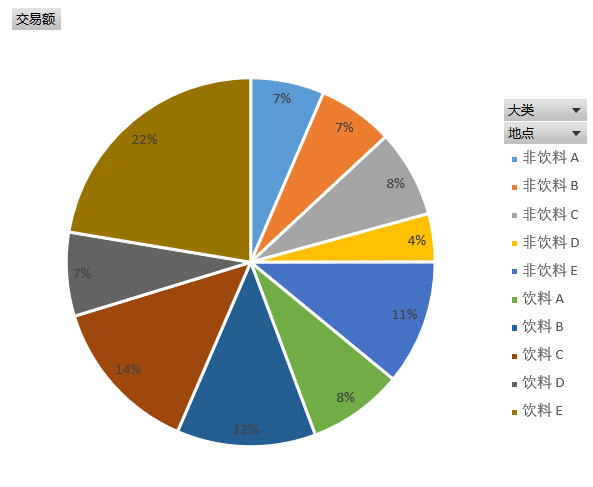
 

**图6 售货机C交易额月环比增长率 图7 售货机D交易额月环比增长率**



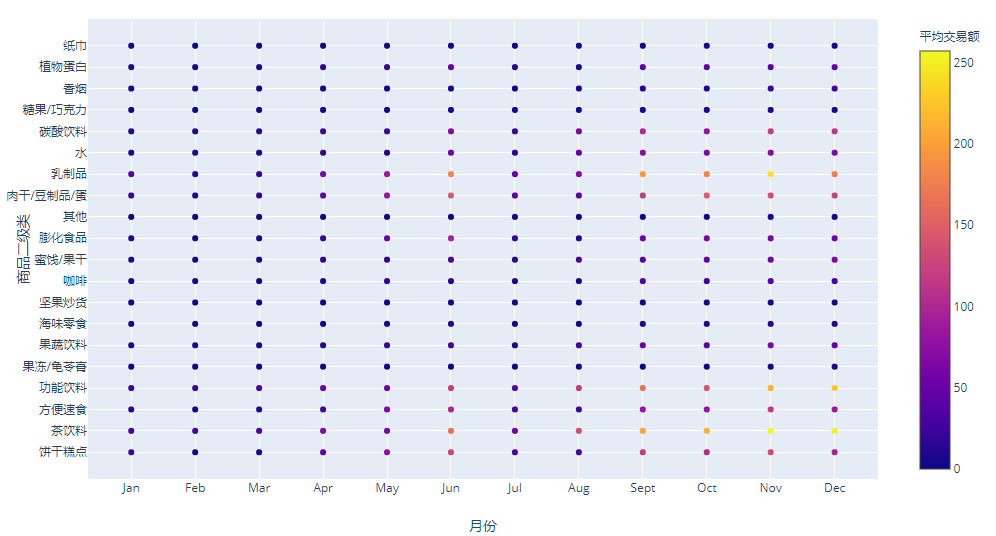
**图8 售货机E交易额月环比增长率**

从每台售货机毛利润占总毛利润比例饼图可以看出，饮料类产品毛利润比例明显大于非饮料类产品。说明售货机的饮料产品对利润贡献率较大，进而可以知道饮料类产品深受消费者喜欢，因此，在运营者考虑售货机内商品种类时应着重考虑饮料类商品。



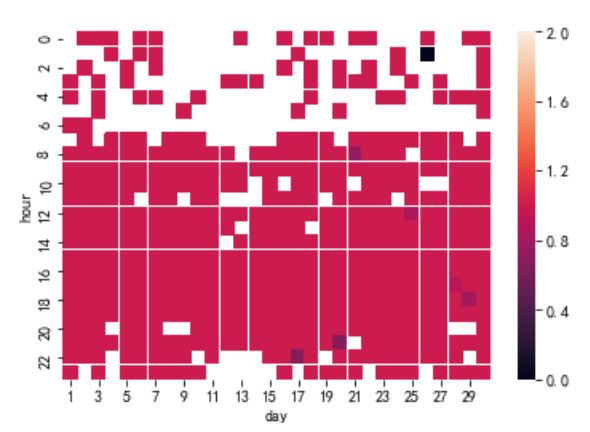
**图9 每台售货机毛利润占总毛利润比例**

图10可以清楚看到，乳制品，肉干/豆制品/蛋，功能饮料和茶饮料平均交易额较高，并且这几类产品在9-12月份平均交易额最高。而其他产品平均交易额相差不大。

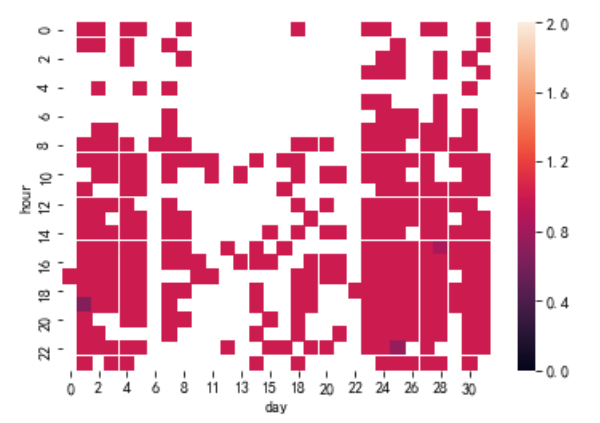


**图10 每月交易额均值**

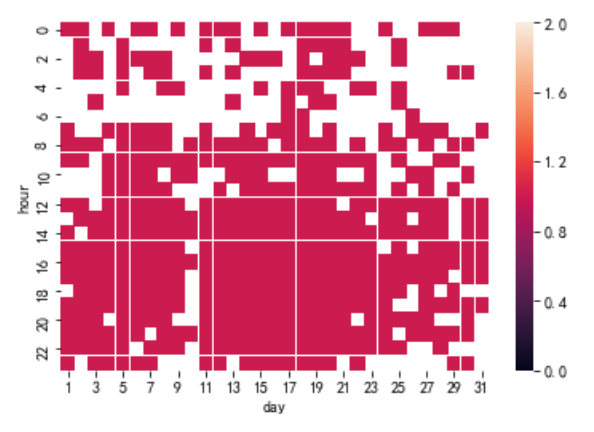
从以下三个图，可以看到6，7，8三个月份每一天的订单量随时间的变化。6月全月的订单量基本集中在8时至21时这段时间内；7月订单量有两段较为集中的时段，一段是7月1日至7月5日的8时至20时，另一段是7月22日至7月31日的9时至22时；8月订单量基本集中在8月1日至8月23日的9时至22时。因此，可以基本推测消费者在这三个月中的消费时间在8时至22时。运营者需要定期检查，保证在这段时间前将受欢迎的商品补充到售货机中。



**图11 6月订单量**

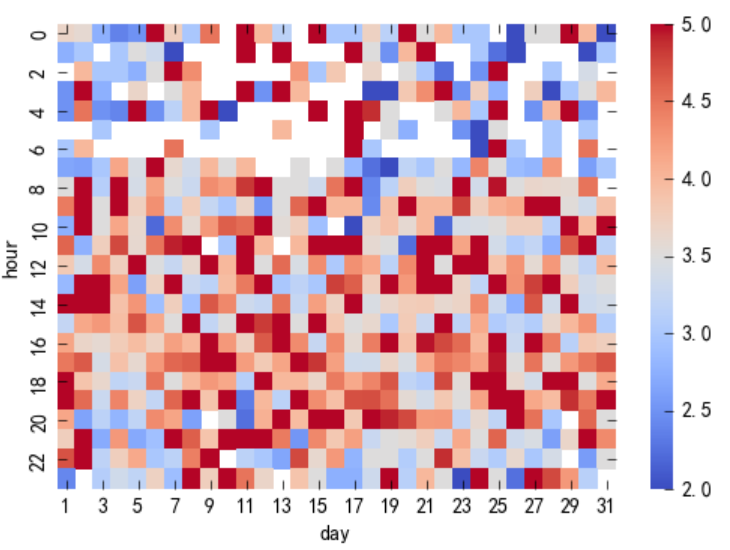


**图12 7月订单量**

****

**图13 8月订单量**

图显示了6-8月订单量的总体情况，可以看出，订单量的情况基本跟每月订单量情况一致。基本集中在10时至21时之间，且在一个月的上中旬订单量最多。



**图14 售货机C 6、7、8三个月订单量**

# **售货机画像**

## **6.1 定义概念**

由于商品订单量与交易额并不成正比，因此无法任选两者之一作为划分的依据。在查阅了资料后，滞销一般是订单量为零或者据具体数据而言进行定义。因此，对商品标签化时选择订单量进行分析，以划分畅销/正常/滞销的订单量区间。

## **6.2 贴标签**

对附件2所有商品的订单量和交易额进行描述性分析，得到下表，可以看到订单量25%，50%，75%分位数对应的订单量，因数据极值差异化较大，故不宜用平均数进行划分，所以按三大分位数将订单量分为：畅销区间；正常区间；滞销区间。

将5台售货机的商品订单量按以上区间各自贴上标签。



**表5 商品订单量描述性分析**

## **6.3售货机画像**

将5台售货机的商品按照附件2的大类和二级类完成标签拓展。主要借助Power BI Desktop进行标签拓展，再将各自的标签存为.txt文件，导入Jupyter Notebook进行词云图的绘制，需注意的是需要将词云图的背景和字体提前导入数据所在文件夹中。

运行程序得到5台售货机各自的画像，结果如下所示。



**图16 售货机A画像 图17 售货机B画像**



**图18 售货机C画像 图19 售货机D 画像**



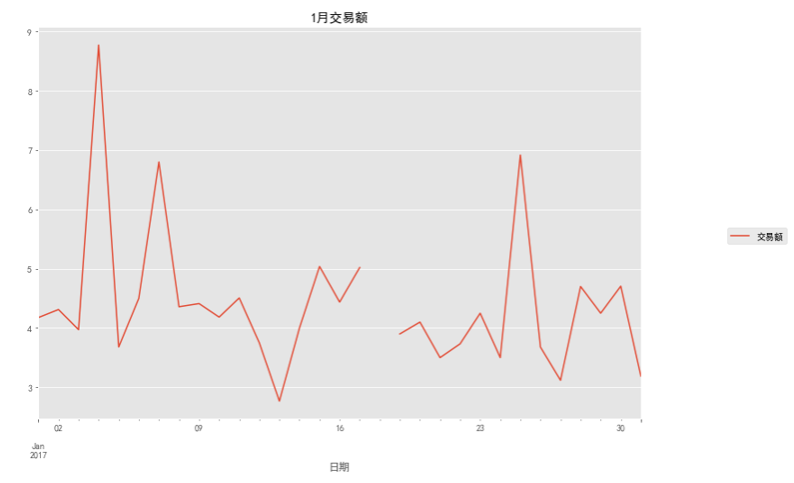
**图20 售货机E画像**

可以看到，5台售货机的标签主要是正常和滞销商品，销量较大的商品大部分是饮料类的乳制品，茶饮料等；以及非饮料的膨化食品，饼干糕点，肉干，豆制品等。

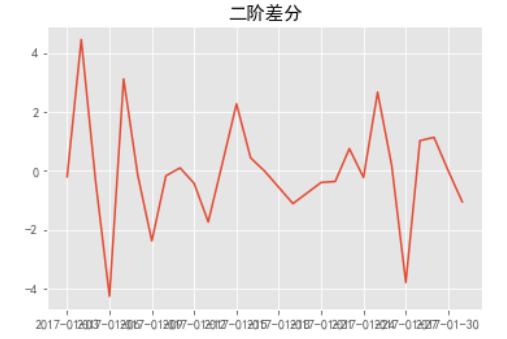
5台售货机中，售货机A、B和D的滞销商品较多，正常商品次之，售货机C 滞销商品和正常商品数量基本相同，而售货机E正常商品数量多于滞销商品。

# **业务预测**

将附件1中2017年1月的数据整理成文档，导入Jupyter Notebook进行时间序列分析，先初步对数据进行折线图的展示，可以看到，数据分布差异较大，故对数据进行1阶差分，得到的折线图仍不理想，进一步进行二阶差分，得到图22，数据分布基本在零上下波动。

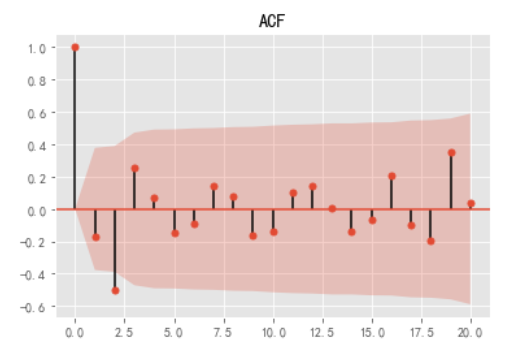


**图21 1月交易额**

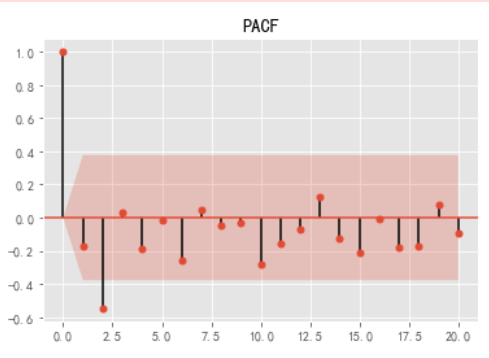


**图22 二阶差分**

图23和24可以看到ACF图和PACF图，从3阶开始以后均落在2倍标准差内，因此，将此模型定为ARIMA（3，2，3）。



**图23 自相关系数图**



**图24 偏自相关图**

# **小结**

根据以上分析，建议售货机运营者在选择售货机产品时按照受消费者欢迎度进行商品购置，如饮料类商品，乳制品和饼干糕点类商品，并且注意在一天的10时至21时这段时间内增大受欢迎商品数量，特别是对于表现一般的售货机。以达到更好的增长率。