### 分析流程 数据源： Task 2机器学习.xlsx 算法配置： 算法： 时间序列分析ARIMA 分析结果： 暂无数据

### 分析步骤 1. ARIMA模型要求序列满足平稳性，查看ADF检验结果，根据分析t值，分析其是否可以显著性地拒绝序列不平稳的假设(P<0.05)。 2. 查看差分前后数据对比图，判断是否平稳（上下波动幅度不大），同时对时间序列进行偏（自相关分析），根据截尾情况估算其p、q值。 3. ARIMA模型要求模型具备纯随机性，即模型残差为白噪声，查看模型检验表，根据Q统计量的P值(P>0.05)对模型白噪声进行检验，也可以结合信息准则AIC和BIC值进行分析（越低越好），也可以通过模型残差ACF/PACF图进行分析根据模型参数表，得出模型公式结合时间序列分析图进行综合分析，得到向后预测的阶数结果。 Tips：采用ARIMA模型预测时序数据，必须是稳定的，如果不稳定的数据，是无法捕捉到规律的。比如股票数据用ARIMA无法预测的原因就是股票数据是非稳定的，常常受政策和新闻的影响而波动，可以使用ADF检验，该检验用于稳定性检验，使用差分分析对数据进行稳定性处理。

### 详细结论

**输出结果1：ADF检验表**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ADF检验表 | | | | | | | |
| 变量 | 差分阶数 | t | P | AIC | 临界值 | | |
| 1% | 5% | 10% |
| 3 tries | 0 | -7.637 | 0.000\*\*\* | 2362.272 | -3.449 | -2.87 | -2.571 |
| 1 | -9.689 | 0.000\*\*\* | 2386.908 | -3.449 | -2.87 | -2.571 |
| 2 | -9.386 | 0.000\*\*\* | 2428.463 | -3.45 | -2.87 | -2.571 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | |

**图表说明：**

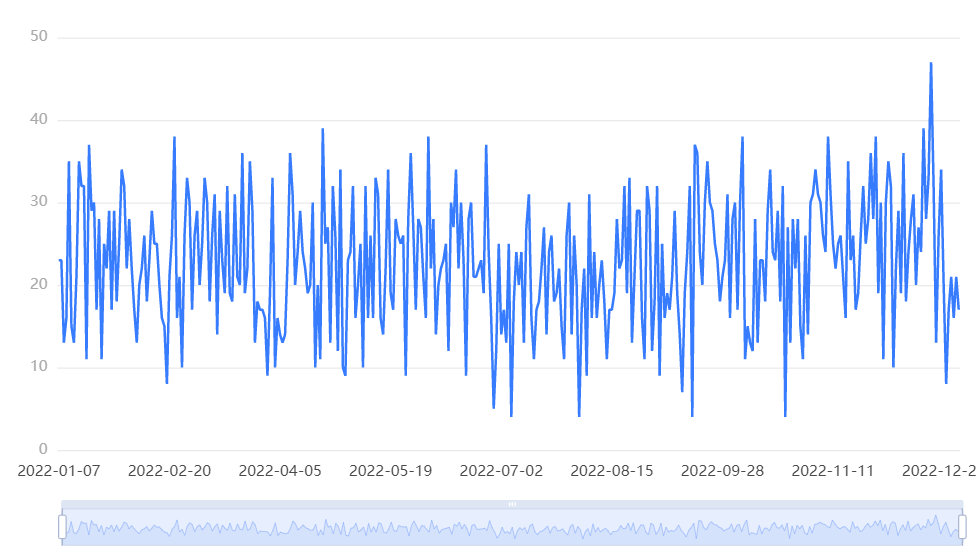
上表格为ADF检验的结果，包括变量、差分阶数、T检验结果、AIC值等，用于检验时间序列是否平稳。  
● 该模型要求序列必须是平稳的时间序列数据。通过分析t值，分析其是否可以显著地拒绝序列不平稳的原假设。  
● 若呈现显著性(P<0.05)，则说明拒绝原假设，该序列为一个平稳的时间序列，反之则说明该序列为一个不平稳的时间序列。  
● 临界值1%、5%、10%不同程度拒绝原假设的统计值和ADF Test result的比较，ADF Test result同时小于1%、5%、10%即说明非常好地拒绝该假设。  
● 差分阶数：本质上就是下一个数值 ，减去上一个数值，主要是消除一些波动使数据趋于平稳，非平稳序列可通过差分变换转化为平稳序列。  
● AIC值：衡量统计模型拟合优良性的一种标准，数值越小越好。  
● 临界值：临界值是对应于一个给定的显着性水平的固定值。

**智能分析：**

该序列检验的结果显示，基于变量3 tries:  
在差分为0阶时，显著性P值为0.000\*\*\*，水平上呈现显著性，拒绝原假设，该序列为平稳的时间序列。  
在差分为1阶时，显著性P值为0.000\*\*\*，水平上呈现显著性，拒绝原假设，该序列为平稳的时间序列。  
在差分为2阶时，显著性P值为0.000\*\*\*，水平上呈现显著性，拒绝原假设，该序列为平稳的时间序列。

**输出结果2：最佳差分序列图**

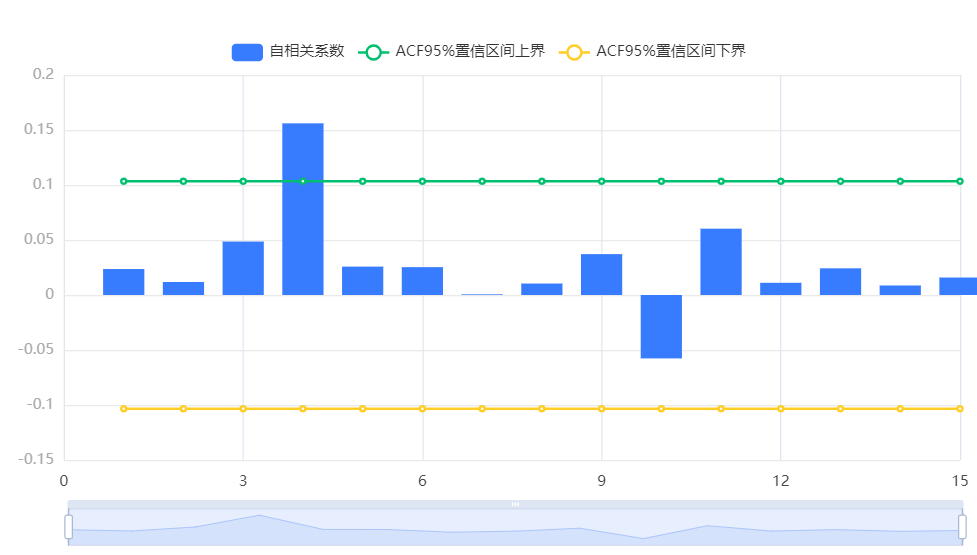
3 tries



**图表说明：**

上图展示了原始数据0阶差分后的时序图。

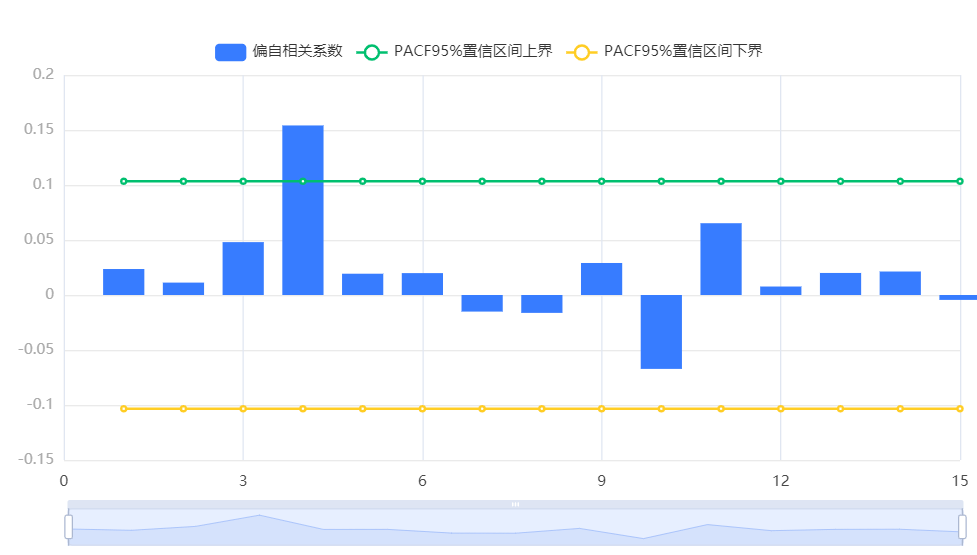
**输出结果3：最终差分数据自相关图(ACF)**



**图表说明：**

上图展示了自相关图(ACF)，包括系数，置信上限和置信下限。  
● 横轴代表延迟数目，纵轴代表自相关系数。  
● 自相关(ACF)图在q阶进行截尾，偏自相关(PACF)图拖尾，ARMA模型可简化为MA(q)模型。  
● 倘若自相关与偏自相关图均拖尾，可结合PACF、ACF图中最显著的阶数(最小值)作为p、q值。  
● 倘若自相关与偏自相关图均截尾，可以选择更换更高的差分，或则不适合建立ARMA模型。  
● 截尾是在置信区间内，ACF或PACF在某阶后就恒等于零(或在0附近随机波动)。  
● 拖尾是在置信区间内，ACF或PACF始终有非零取值，不呈现在某阶后就恒等于零(或在0附近随机波动)。

**输出结果4：最终差分数据偏自相关图(PACF)**



**图表说明：**

上图展示了偏自相关图(PACF)，包括系数，置信上限和置信下限。  
● 偏自相关(PACF)图在p阶进行截尾，自相关(ACF)图拖尾，ARMA模型可简化为AR(P)模型。  
● 倘若自相关与偏自相关图均拖尾，可结合PACF、ACF图中最显著的阶数(最小值)作为p、q值。  
● 倘若自相关与偏自相关图均截尾，可以选择更换更高的差分，或则不适合建立ARMA模型。  
● 截尾是在置信区间内，ACF或PACF在某阶后就恒等于零(或在0附近随机波动)。  
● 拖尾是在置信区间内，ACF或PACF始终有非零取值，不呈现在某阶后就恒等于零(或在0附近随机波动)。

**输出结果5：模型参数表**

|  |  |  |
| --- | --- | --- |
| ARIMA模型（0,0,0）检验表 | | |
| 项 | 符号 | 值 |
|  | Df Residuals | 358 |
| 样本数量 | N | 359 |
| Q统计量 | Q6(P值) | 0.202(0.653) |
| Q12(P值) | 10.48(0.106) |
| Q18(P值) | 13.664(0.323) |
| Q24(P值) | 15.263(0.644) |
| Q30(P值) | 18.87(0.759) |
| 信息准则 | AIC | 2494.915 |
| BIC | 2502.681 |
| 拟合优度 | R² | 0 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | |

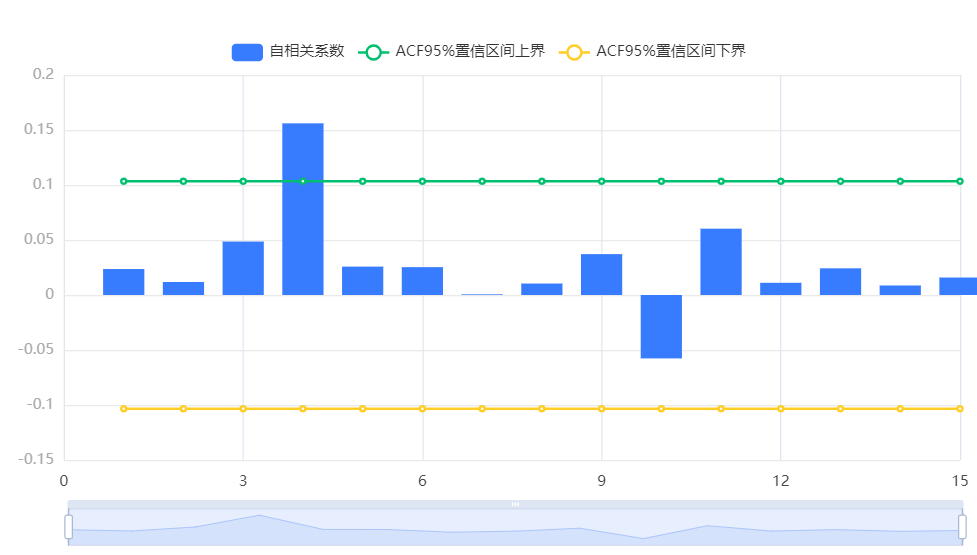
**图表说明：**

上表格展示本次模型检验结果，包括样本数、自由度、Q统计量和信息准则模型的拟合优度。  
● ARIMA模型要求模型的残差不存在自相关性，即模型残差为白噪声，查看模型检验表，根据Q统计量的P值（P值大于0.1为白噪声）对模型白噪声进行检验。  
● 根据信息准则AIC和BIC值用于多次分析模型对比（越低越好）。  
● R²代表时间序列的拟合程度，越接近1效果越好。

**智能分析：**

系统基于AIC信息准则自动寻找最优参数，模型结果为ARIMA模型（0,0,0）检验表，基于变量：3 tries，从Q统计量结果分析可以得到：Q6在水平上不呈现显著性，不能拒绝模型的残差为白噪声序列的假设，同时模型的拟合优度R²为0.0，模型表现较差，模型基本满足要求。

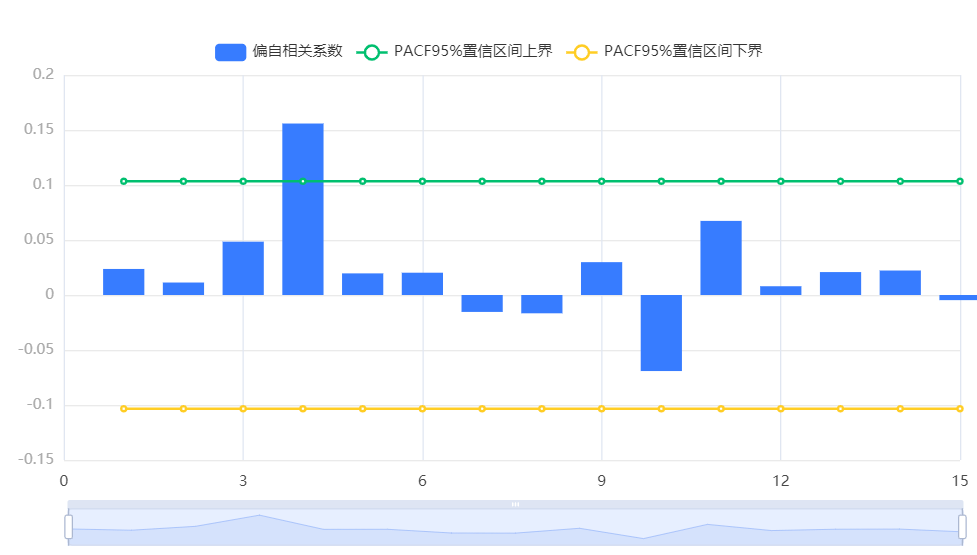
**输出结果6：模型残差自相关图(ACF)**



**图表说明：**

上图展示了模型的残差自相关图(ACF)，包括系数，置信上限和置信下限。  
● 横轴代表延迟数目，纵轴代表自相关系数。  
● 若相关系数均在虚线内，自回归模型(AR)残差为白噪声序列，时间序列要求模型残差为白噪声序列。

**输出结果7：模型残差偏自相关图(PACF)**



**图表说明：**

上图展示了模型的残差偏自相关图(PACF)，包括系数，置信上限和置信下限。  
● 若相关系数均在虚线内，滑动平均模型(MA)残差为白噪声序列，时间序列要求模型残差为白噪声序列。

**输出结果8：模型检验表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 模型参数表 | | | | | | |
|  | 系数 | 标准差 | t | P>|t| | 0.025 | 0.975 |
| 常数 | 22.727 | 0.41 | 55.419 | 0 | 21.923 | 23.531 |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | |

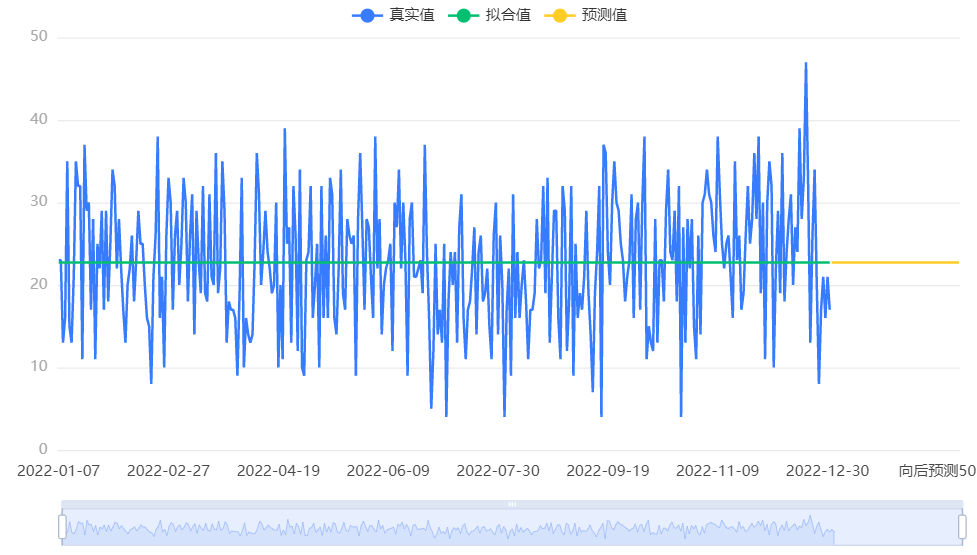
**图表说明：**

上表格展示本次模型参数结果，包括模型的系数、标准差，T检验结果等，用于分析模型公式。

**智能分析：**

基于变量3 tries，系统基于AIC信息准则自动寻找最优参数，模型结果为ARIMA模型（0,0,0）检验表且基于0差分数据，模型公式如下：  
y(t)=22.727

**输出结果9：时间序列图**



**图表说明：**

上图表示了该时间序列模型的原始数据图、模型拟合值、模型预测值。

**输出结果10：时间序列预测表**

|  |  |
| --- | --- |
| 预测值 | |
| 阶数（时间） | 预测结果 |
| 1 | 22.727019498607245 |
| 2 | 22.727019498607245 |
| 3 | 22.727019498607245 |
| 4 | 22.727019498607245 |
| 5 | 22.727019498607245 |
| 6 | 22.727019498607245 |
| 7 | 22.727019498607245 |
| 8 | 22.727019498607245 |
| 9 | 22.727019498607245 |
| 10 | 22.727019498607245 |
| 11 | 22.727019498607245 |
| 12 | 22.727019498607245 |
| 13 | 22.727019498607245 |
| 14 | 22.727019498607245 |
| 15 | 22.727019498607245 |
| 16 | 22.727019498607245 |
| 17 | 22.727019498607245 |
| 18 | 22.727019498607245 |
| 19 | 22.727019498607245 |
| 20 | 22.727019498607245 |
| 21 | 22.727019498607245 |
| 22 | 22.727019498607245 |
| 23 | 22.727019498607245 |
| 24 | 22.727019498607245 |
| 25 | 22.727019498607245 |
| 26 | 22.727019498607245 |
| 27 | 22.727019498607245 |
| 28 | 22.727019498607245 |
| 29 | 22.727019498607245 |
| 30 | 22.727019498607245 |
| 31 | 22.727019498607245 |
| 32 | 22.727019498607245 |
| 33 | 22.727019498607245 |
| 34 | 22.727019498607245 |
| 35 | 22.727019498607245 |
| 36 | 22.727019498607245 |
| 37 | 22.727019498607245 |
| 38 | 22.727019498607245 |
| 39 | 22.727019498607245 |
| 40 | 22.727019498607245 |
| 41 | 22.727019498607245 |
| 42 | 22.727019498607245 |
| 43 | 22.727019498607245 |
| 44 | 22.727019498607245 |
| 45 | 22.727019498607245 |
| 46 | 22.727019498607245 |
| 47 | 22.727019498607245 |
| 48 | 22.727019498607245 |
| 49 | 22.727019498607245 |
| 50 | 22.727019498607245 |
| 51 | 22.727019498607245 |
| 52 | 22.727019498607245 |
| 53 | 22.727019498607245 |
| 54 | 22.727019498607245 |
| 55 | 22.727019498607245 |
| 56 | 22.727019498607245 |
| 57 | 22.727019498607245 |
| 58 | 22.727019498607245 |
| 59 | 22.727019498607245 |
| 60 | 22.727019498607245 |

**图表说明：**

上表显示了时间序列模型最近60期数据预测情况。

### 参考文献 [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com. [2] 王燕．应用时间序列分析[M]．北京：中国人民大学出版社 2005.