

# 时间序列数据分析

- 时间序列数据分析，这一步的必要性体现在，**可解释性**  
(解释模型为什么这样选？解释数据是否达到模型的要求等)
- 时间序列趋势分解✔——与数据建立最直接的认识
  - 意义
    - **理解数据本质**：分离出趋势和季节性，更清晰地观察数据规律
    - **提高预测精度**：对趋势和季节性单独建模（如ARIMA、指数平滑）比直接建模原始序列更有效
    - **异常检测**：通过残差部分识别异常点（如突发的销量暴跌）
  - 时间通常由**趋势性、季节性、周期性、随机性**构成，时间序列分解就是把时间序列分解为这四个性质  
(有季节性和无季节性对ARIMA的时序数据处理的影响大)
  - 分解方法：移动平均分解、STL分解
  - 代码函数（示例使用的是statsmodels 库中seasonal\_decompose 函数，原理是移动平均分解）见附件
- Hurst指数✔——模型选择依据
  - 作用：衡量时间序列的长期记忆性（时间序列的过去行为对未来有多大程度的影响？）**指导模型选择**
    - 若序列有强趋势性 ( $H > 0.5$ )，需选择能捕捉趋势的模型。
      - 如果Hurst指数显示序列具有**趋势性**，适合以下模型
        - 趋势跟随模型：
          - Holt-Winters三指数平滑：捕捉趋势和季节性。
          - ARIMA(p,d,q)（需结合差分）：通过差分消除趋势后建模。
          - SARIMA：加入季节性参数（如周、月周期性）。
        - 长记忆模型：
          - ARFIMA (Fractional ARIMA)：允许分数阶差分 ( $d \in (0, 0.5)$ )，直接建模长期记忆性。
        - 机器学习模型：
          - Prophet (Meta)：内置趋势、季节性和节假日效应。
          - LSTM神经网络：捕捉非线性长期依赖关系。
      - 若序列均值回复 ( $H < 0.5$ )，需考虑均值回复模型。
    - 计算方法  
写代码可以用包，但是简单的原理要懂
      -

Hurst 指数计算:

① 分割序列:  
将长度为  $n$  的序列分割为  $n$  个子区间 (每个子区间 <sup>位置标记</sup> 长度为  $k$ )

② 计算每个子区间累积离差

$$X_{t,k} = \sum_{i=1}^k (y_i - \bar{y}_k) \quad (t=1, 2, \dots, k)$$

子区间所有数据均值.

③ 计算极差

$$R = \max(X_{t,k}) - \min(X_{t,k})$$

④ 计算标准差

$$S_k = \sqrt{\frac{1}{k} \sum_{i=1}^k (y_i - \bar{y}_k)^2}$$

⑤ 计算重标极差 (R/S)

$$(R/S)_k = \frac{R_k}{S_k}$$

⑥ 拟合幂律关系 (线性回归拟合)

$$\log\left(\frac{R_k}{S_k}\right) = \log(C) + \underbrace{H}_{\text{Hurst 指数}} \cdot \log(k)$$

- 判断条件

- 如果  $H=0.5$ , 表明时间序列接近随机游走 (无记忆);
- 如果  $0.5 < H < 1$ , 表明时间序列趋势持续 (存在长期记忆性);
- 如果  $0 \leq H < 0.5$ , 表明时间序列均值回复过程。

- 代码 (见hurst指数附件)

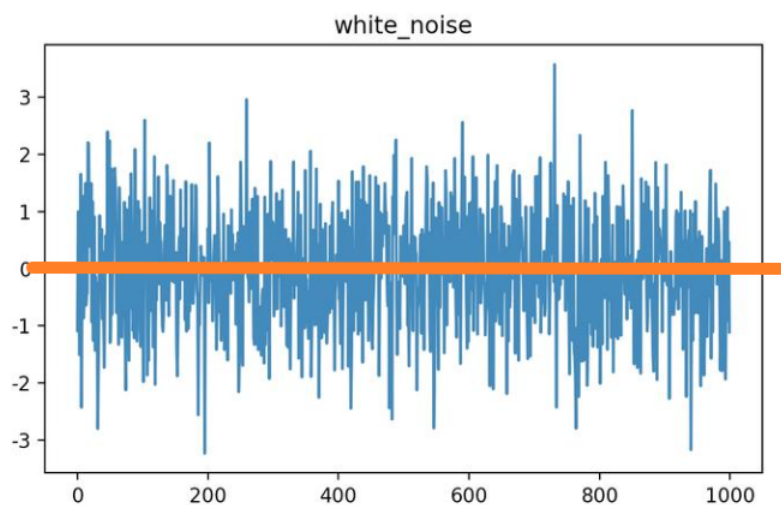
- 平稳性检验 (ADF检验) ✓ —— 数据条件检验

在对任何时间序列进行分析前, 我们都要判断该序列是否为平稳序列

- 所谓平稳: 在均值附近无趋势、无周期性、无季节性地波动

一般平稳序列和白噪声序列 (纯随机序列)

- 例如: 图



- 白噪声序列 (纯随机序列) —— 特殊的平稳序列

特点: 任意两项的协方差/相关性系数都是零, 白噪声序列作为纯随机序列, 不具有分析意义

- 平稳性检验方法

序列平稳表示均值、方差、协方差不随时间变化；ARMA等模型要求序列平稳，非平稳序列需差分处理

- ADF检验（单位根检验）

最普适，用的最多的方法；该讲不含ADF检验统计学原理

- 基本逻辑：ADF检验就是判断序列是否存在单位根：如果序列平稳，就不存在单位根；否则，就会存在单位根（基于假设检验）

- $H_0$  假设：存在单位根

如果得到的显著性检验统计量小于三个置信度（10%，5%，1%），则对应有（90%，95，99%）的把握来拒绝原假设

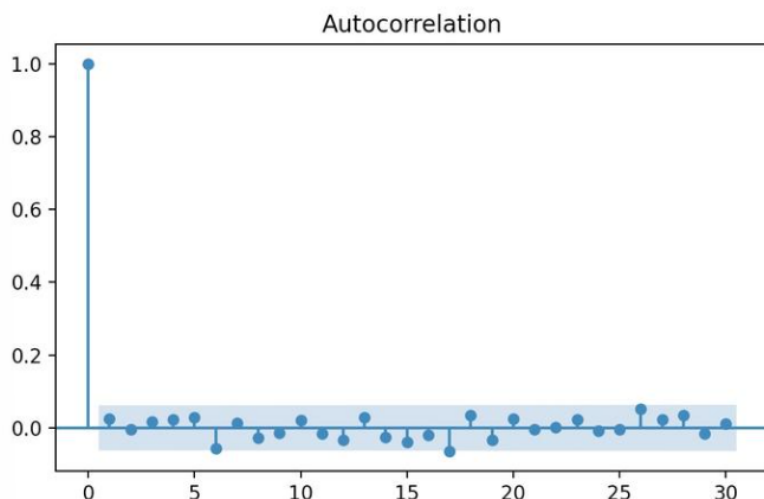
- statsmodels包中有adf检验函数，用法见附件

- 白噪声检验（ACF检验）✔——数据条件检验

任意两项的协方差/相关性系数都是零，也就是说任意不同的两项之间不存在相关性关系

- 自相关图（acf）

- 判定标准：除0阶自相关系数为1外，延迟k阶自相关系数均为0（接近于0）



- ACF代码见附件