# Assumptions and Justifications

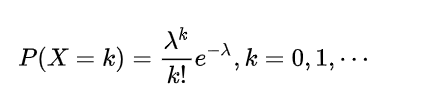
1. 上传数据大体上是真实的，符合游戏规则的
2. 数据代表的都是真实的人，他们解题的思路没有发生重大的变化

# model 1

报告的结果数量每天都有所不同。开发一个模型来解释这种变化，*并使用您的模型创建一个关于2023年3月1日报告结果数量的预测区间*。

1报告总人数和时间的关系比例

尝试1：



利用泊松分布模型的概率函数，创建线性回归（）

尝试1可能的问题：倍数相差太小（？2w与35w相比也有17倍）

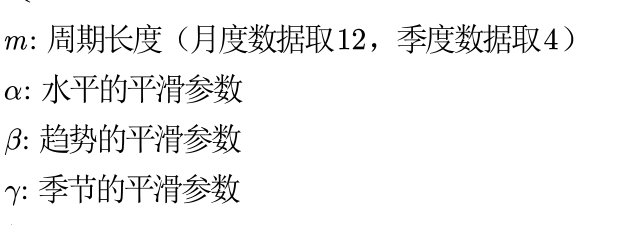
不够拟合

非线性回归

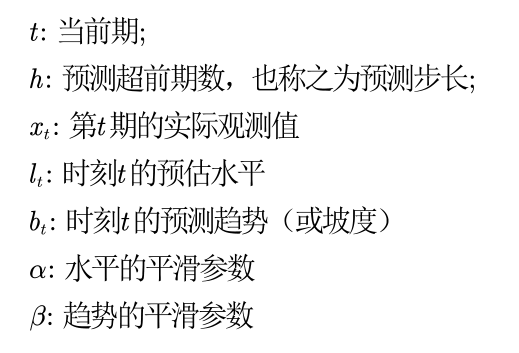
**暂时确定：**

我们使用温特乘法模型建模

记号：



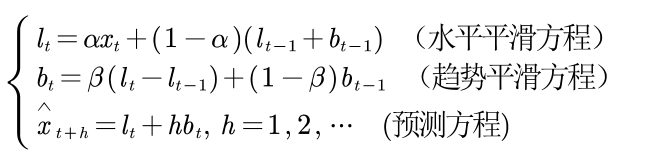
m取7



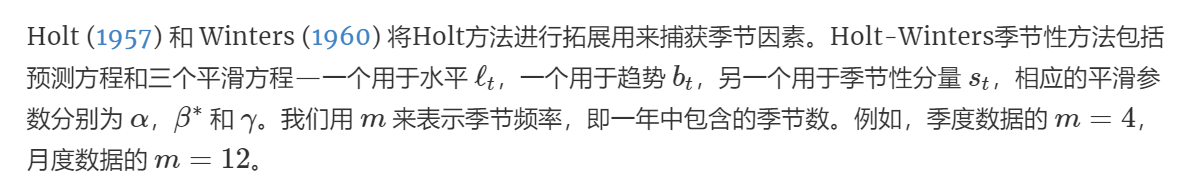
（判断应该用时间序列预测）

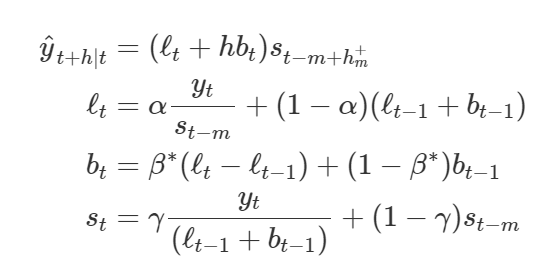
Holt 在1957年把简单的指数平滑模型进行了延伸，能够预测包含趋势的数据，

该方法包含一个预测方程和两个平滑方程（一个用于水平，另一个用于趋势）



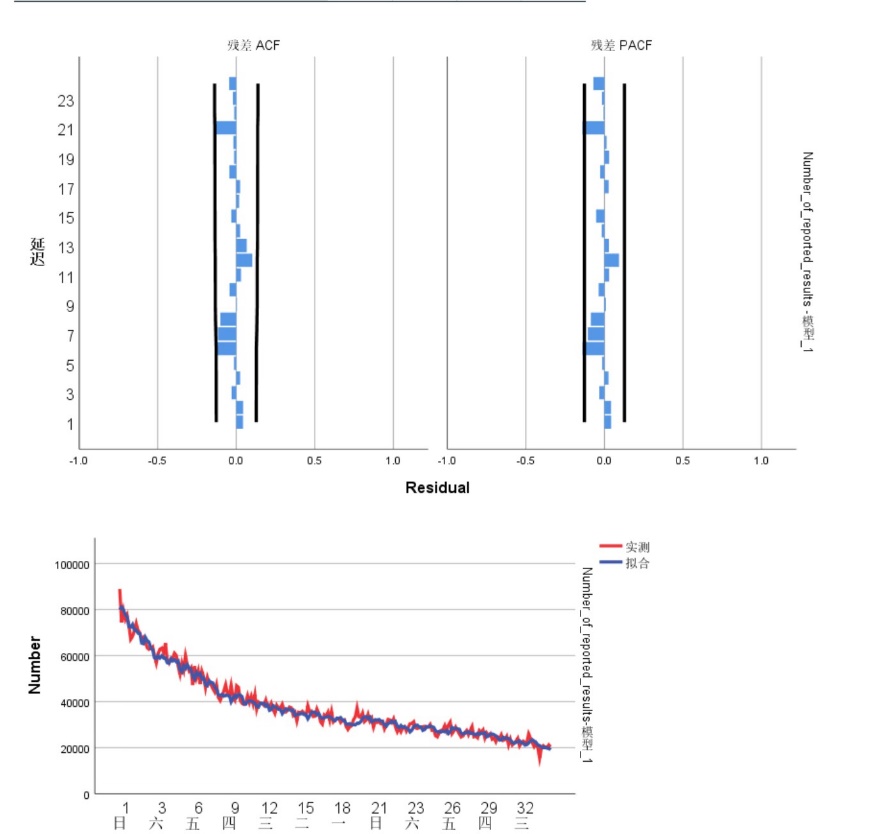
（判断其中可能存在的季节性）





h\_m暂未确定意义，暂不表

实际数据：问cy



区间确定：

暂未完成，比较简陋的思路是再次用对数线性回归，确定上下界

直接拿e^（kx+b）[+x]拟合（k为负数）

在确定预测模型后，我们根据每日的波动，确定预测区间

# model 2

是否有单词的属性会影响报告结果的比例？

对于未来日期的**给定解决方案单词**，开发一个模型，使您可以**预测报告结果的分布**。换句话说，预测未来日期的相关百分比（1、2、3、4、5、6、X）的分布。

*您的模型和预测有哪些不确定性？请举一个关于2023年3月1日单词EERIE的预测的具体例子。您对您模型的预测有多自信？*

*——需要我们自己寻找确定所谓的词的属性，所以一定存在某种属性影响这个比例*

注意到这里与时间无关，通过单词与报告结果分布，训练出单词预测报告分布的模型。

那我们需要对单词做出一个分析，大概是单词->特征向量->报告分布

# model 3

开发并总结一个模型，通过难度**分类解决方案单词**。**确定**与每个分类**相关**联的给定单词的**属性**。*使用您的模型，单词EERIE有多难？讨论您的分类模型的准确性。列出并描述该数据集的其他有趣特征。*

1 难是如何定义的

2 在模型2的基础上聚类分析即可