**实验报告**

## 一、实验目的

通过数据收集与整理进行统计分析。

## 二、实验要求

可以采用社会调查的数据，也可以自行在网络收集数据或使用Python第三方数据接口包调用数据。在数据分析中不能仅罗列数据和指标，应包含具体的分析内容，形成相对完整的统计分析报告。

## 三、实验内容与步骤

## （一）提出问题

哪些数据会影响系统的局部稳定性，哪些数据影响因子会比较大？

## （二）数据整理

**1、数据来源**

网站**：**[UCI Machine Learning Repository: Electrical Grid Stability Simulated Data Data Set](http://archive.ics.uci.edu/ml/datasets/Electrical+Grid+Stability+Simulated+Data+)

**2、数据与样本的基本情况**

(a) 数据的基本属性

(1) Tanu[x]:参与者的反应时间(从0.5s到10s不等)

(2) p[x]:消耗的额定功率

(3) g[x]:与弹性价格成正比的系数(gamma)

(4) stab:特征方程根的最大实部(如果为正，则不稳定)

(5) stabf:判定系统是否稳定

(b) 数据用途:实施分散智能电网控制概念的4节点星系统（发电站位于中心）的局部稳定性分析。

### 3、研究方法

(a) 通过pandas导入数据:调用read\_csv函数csv文件数据，并进行查看前十行数据内容。

(b) 描述数据集的数学属性：调用describe函数查看数据集数学属性(平均值、最大最小值)

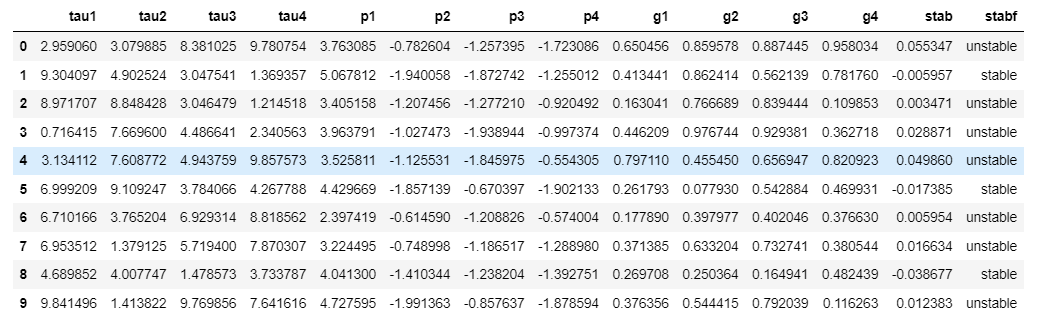
(c) 进行空值检测：调用info函数查看数据集信息：个列数据类型数据是否存在空值。

(d)将文本数据转化为数字类型：从前几步我们可以发现，scorestabf列为object类型，我们将其转换为int类型

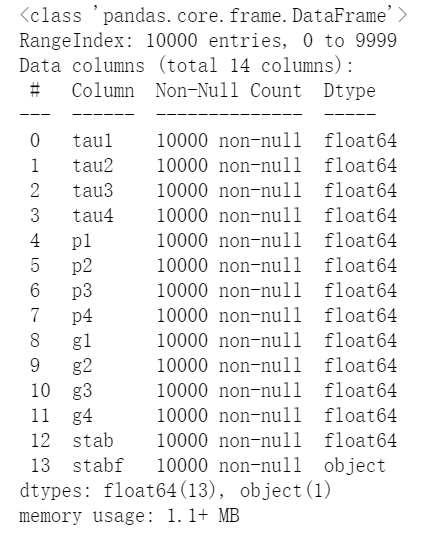
(e) 收集可用数据并进行可视化：将所有数据可视化，横轴为数据大小，纵轴为数据数量，并绘出数据属性之间相关性热力图

## （三）数据的分析

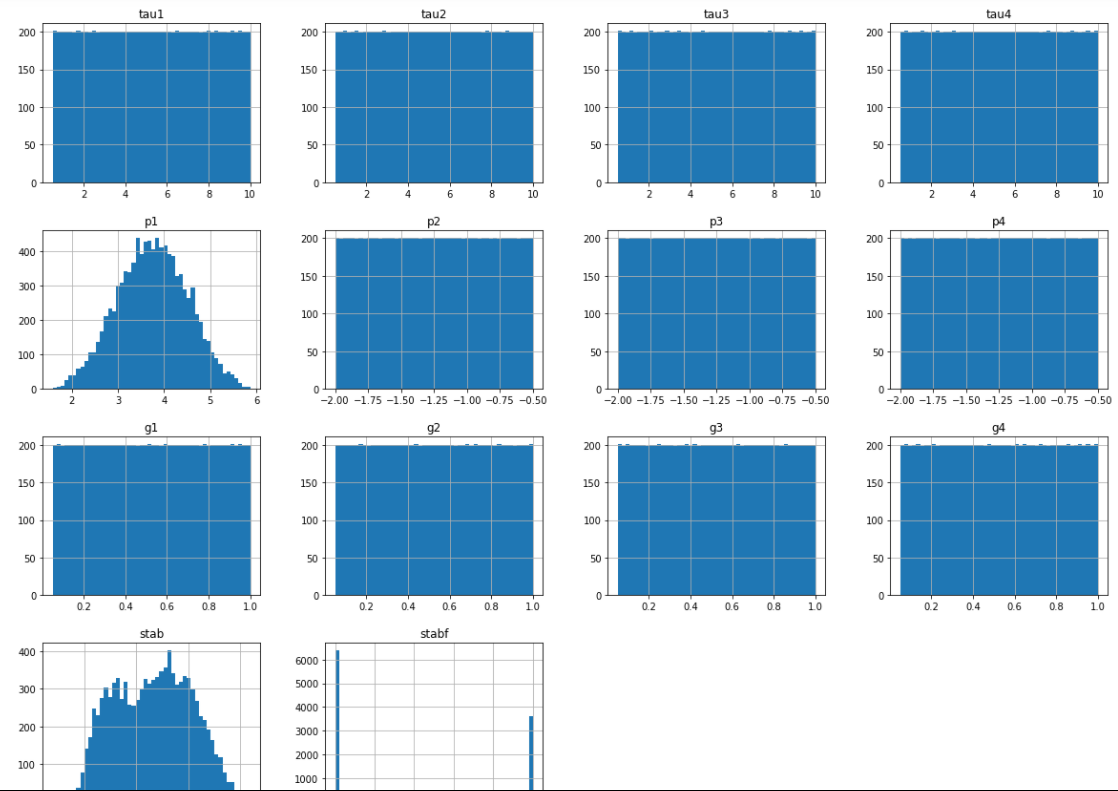
查看数据的前几行，了解样本的基本情况。



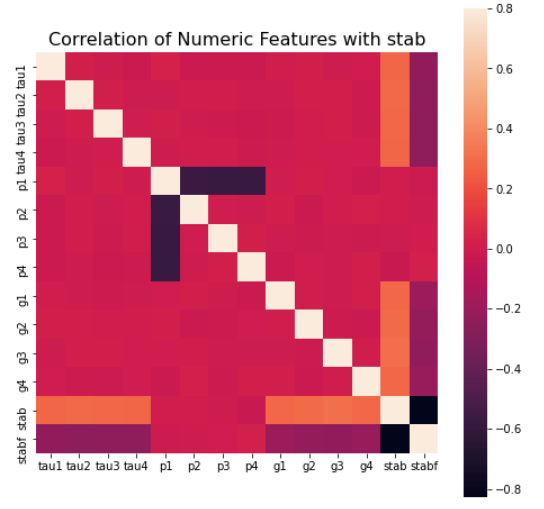
调用info函数后发现数据集比较干净，并没有空值



对所有属性分别绘制直方图，我们可以清晰地看出除了stab和p1大部分属性分布都比较均匀。这些图像也验证了第二步的结果。



通过相关性热力图，我们可以发现 p2 p3 列与stab的相关性并不强，在后续稳定性研究上我们对其可以不用过多考虑。



## 四、实验总结

在实验的强化下，感觉我对于Python语法、numpy、pandas、matplotlib还有列表，类方法，类变量等知识使用得更加熟练了，希望自己可以在不断得加强学习中掌握相关的技能。继续加油！