

# 数据分析方法

## 统计方法建模应注意的问题

- 了解统计方法的适用范围，检验条件是否满足
- 利用**方差分析**建模时，注意其基本假设是两组数据均满足**正态分布**，检验均值是否相等时是在方差相等的假设下做出的，应用时应作出检验
- 对葡萄酒的**质量**, **影响因素**---专家，**两组实验**---两个水平-----**单因素方差分析**
- 若同时考虑各组专家在4个一级指标或10个二级指标的差异-----**多因素方差分析**

- **方差分析**

原假设

用 `matlab` 程序

步骤

1. 先验证条件（满足齐次性）
2. 选取方法
  - **单因素方差分析**就是用来检验同一个影响因素的不同水平对因量是否有影响的一种方法。
  - **协方差**反映是否具有线性关系
  - **多因素方差分析**还有因素之间的协调关系
  - **双因素方差分析**
3. 制表
4. 根据表得结

- 秩和检验
- t检验
- f检验

## 建模方法

- 回归模型：质量得分与理化指标进行多因素回归分析
- 相关性分析：指标因子相关分析，降维；或者质量得分与理化指标之间作相关分析，去掉无关的、或低相关的因素，再做回归分析
- 聚类分析法，定义一些距离（欧氏距离，Minkowski距离），利用聚类分析法按不同的原则进行聚类，根据评分决定类别的高低；（先指标聚类、再样本聚类，再根据得分结果划分等级）
- 主成分分析法、因子分析-----降维
- 其他：多元判别分析，多因素综合评判，隶属度函数进行分级；SOM神经网络的聚类；等等

激活  
转到

## 问题一：建模

- 第一问：两组专家对同一组样本打分，分析两组数据是否有显著性差异-----直观理解：方差分析，或t检验，F检验，秩和检验，Wilcoxon符号秩检验，构造一些统计量，计算它们的值，是否通过检验，给出结果。
- 第二问：确定哪一组专家更可信，直观理解：总体方差过小的组别区分度不好，因此可以通过比较方差大小来确定，也可以采用Cronbach可信度系数，比较酒样F值和评酒员F值的相对大小，Spearman秩相关系数，肯德尔和谐系数法。