

常见统计问题讨论 (1)

姜晓东

2015-04-03

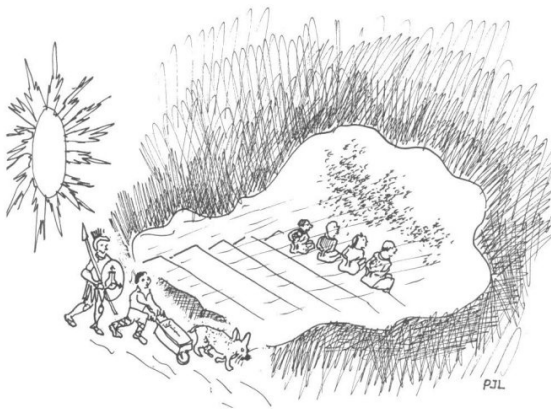
讨论的问题

- 为什么要统计？
- “假设检验”是什么？ p 值又是什么？
- 两组数据比较用什么方法？
- 三组及以上数据比较用什么方法？

为什么要统计?

柏拉图的理想国

我们研究现实世界的“样本”，推测理想国中的“总体”。



Plato's Allegory of the Cave
(The Republic)

“假设检验”是什么？ p 值又是什么？

假设检验的基本原理

Once you've ruled out the impossible, whatever remains, however improbable, must be true.



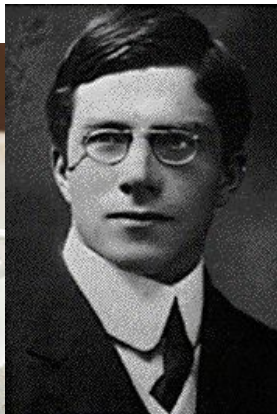
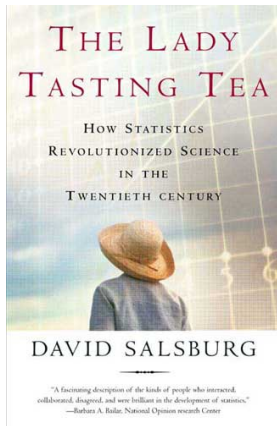
假设检验的基本原理

- H_0 : 原假设，对所研究的“总体”做的假设。常规的，公认的，一般的假设。
- H_1 : 备择假设，“你期望得到的阳性结论”。

如同福尔摩斯破案，我们通常无法直接研究**备择假设**，而是通过排除**原假设**来间接认定**备则假设**。

在统计学上，假设检验的过程中的 p 值是**原假设 (H_0)** 成立的概率。

女士品茶



女士品茶

- H_0 : 原假设，两种茶水无差别，不能分辨，只能靠随机乱猜的。
- H_1 : 备择假设，两种茶水有区别，可以分辨出来。

女士品茶

- H_0 : 原假设, 两种茶水无差别, 不能分辨, 只能靠随机乱猜的。

$$p = \underbrace{\frac{1}{2} \times \frac{1}{2} \times \cdots \times \frac{1}{2}}_{12} = \left(\frac{1}{2}\right)^{12} = 2.4414062 \times 10^{-4}$$

- H_1 : 备择假设, 两种茶水有区别, 可以分辨出来。

假设检验

● 假设检验过程

- ① 建立原假设与备择假设（你来做）
- ② 计算原假设的概率 p （软件自动计算）
- ③ 根据 α 水平（0.05 或 0.01）接受或拒绝原假设（你来做）

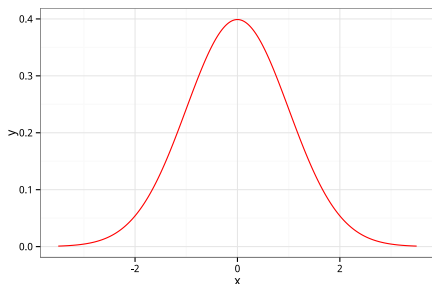
● 注意事项

- 统计推断与逻辑推断的区别
- 统计推断与专业推断的区别
- p 值大小的含义

两组数据比较用什么方法？

数据类型与表示

- 连续数据
 - 均值
 - 方差
- 分类数据
 - 比例



两组连续数据的假设检验

- 正态分布
 - t 检验
 - 配对 t 检验
- 偏态分布或其他分布 (频率、百分率、复和单位) 或方差不齐
 - 秩和检验 (Wilcoxon 检验)
 - 配对秩和检验 (Wilcoxon 检验)

三组及以上连续数据比较用什么方法

多组连续数据的假设检验

- 正态分布
 - 方差分析
- 非正态同分布
 - Kruskal-Wallis 检验

方差分析

- 数学模型

- $y_{ij} = \mu + \alpha_i + \varepsilon_{i,j}$

- 数据格式

- 测量值作为因变量
 - 分组信息作为自变量

多重比较

- p 值校正的必要性
- 校正方法
 - 避免使用 LSD 法

进一步的阅读

- 孙振球,《医学统计学》, 人民卫生出版社
- M.R. 斯皮格尔,《统计学》, 科学出版社
- 薛毅, 陈立萍,《统计建模与 R 软件》, 清华大学出版社

课件：<https://github.com/JiangXD/Course>