

R 统计模型

谢益辉

中国人民大学统计学院



生物多样性数据分析和生物统计培训班
2008年11月17至20日于中国科学院植物研究所

- 1 统计推断
 - 统计分布
 - 参数估计
 - 假设检验
- 2 统计模型
 - 线性回归分析
 - 方差分析
 - 广义线性模型
 - 生存分析
 - 多元统计分析
 - 非参数统计
 - 树模型

- Tinn-R Is Not Notepad
- Rprofile.site文件的配置
- 自定义快捷键
- 与R交互

目录

- 1 统计推断
 - 统计分布
 - 参数估计
 - 假设检验

- 2 统计模型
 - 线性回归分析
 - 方差分析
 - 广义线性模型
 - 生存分析
 - 多元统计分析
 - 非参数统计
 - 树模型

统计分布相关函数

- 分布函数 $F(x) = P(X < x)$; R函数: `p*()`
- 密度函数 $f(x) = F'(x)$; R函数: `d*()`
- 分位数函数 $P(X < \alpha) = p$; R函数: `q*()`
- 经验分布函数; R函数: `ecdf()` ¹

¹http://en.wikipedia.org/wiki/Empirical_distribution_function

参数估计

- 什么是参数估计
- 矩估计：推公式、写代码²
- 极大似然估计³
- 区间估计：推公式、写代码（或偷懒模拟）

²[http://en.wikipedia.org/wiki/Method_of_moments_\(statistics\)](http://en.wikipedia.org/wiki/Method_of_moments_(statistics))

³http://en.wikipedia.org/wiki/Maximum_likelihood

假设检验

- 什么是假设检验，什么是P值（什么是小的P值）
- 单样本、配对样本和两样本 t 检验: `t.test()`⁴
- 相关系数的检验: `cor.test()`
- 列联表的独立性检验: `fisher.test()`、`chisq.test()`

⁴http://en.wikipedia.org/wiki/T_test

目录

- 1 统计推断
 - 统计分布
 - 参数估计
 - 假设检验
- 2 统计模型
 - 线性回归分析
 - 方差分析
 - 广义线性模型
 - 生存分析
 - 多元统计分析
 - 非参数统计
 - 树模型

线性回归分析

- 线性模型: `lm()`⁵
- 逐步回归: `step()`
- 局部平滑: `lowess()`

⁵http://en.wikipedia.org/wiki/Linear_regression

方差分析

- `aov()`

- 广义线性模型 `glm()`

- survival 包

聚类分析

- 层次聚类 `hclust()` (谱系图)
- K-Means 聚类 `kmeans()`

主成分分析

- `princomp()`

线性判别分析

- MASS 包 `lda()`

非参数检验

- `wilcox.test()`
- `kruskal.test()`
- `friedman.test()`

核密度估计

- `density()`

分类与回归树

- rpart 包 `rpart()`