

统计数据分析及可视化（讲义）

周旷

2024-02-04

Table of contents

目录	3
1 课程简介	4
2 假设检验	5
2.0.1 第一类错误	5
2.0.2 第二类错误	5
2.0.3 p 值	5
2.0.4 样本量	5
3 课程总结	7
参考资料	8

目录

- Tidy 数据处理框架
 - R base
 - R Tidy
 - ggplot
- 统计数据分析
 - 参数估计
 - 假设检验
 - 回归分析
- 可视化
 - 案例 1: 商业数据分析
 - 案例 2: 体育数据分析

1 课程简介

2 假设检验

2.0.1 第一类错误

- 拒真概率：当原假设为真时，检验统计量落入拒绝域的概率

$$P(X \in W | H_0 = TRUE)$$

- α

2.0.2 第二类错误

- 取伪概率：当备择假设为真时，检验统计量未落入拒绝域的概率

$$P(X \notin W | H_1 = TRUE)$$

- β
- 假验的效能：正确接受备择假设的概率

$$(1 - \beta) \times 100\%$$

2.0.3 p 值

2.0.4 样本量

- 效能分析：决定统计学检验的效能和计算揭示一定大小效应所需要的最小样本量
 - α : 第一类错误概率
 - β : 第二类错误概率

- d : 效应的大小，即所研究的效应相对样本的标准差 σ 的大小
 - n : 样本大小
- 上面四个参数，只要固定三个，就可以确定剩余参数的值

3 课程总结

参考资料 (Wickham and Grolemund 2023)

参考资料

Wickham, J. J., H. W. Jr., Hadley, and G. G. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. R for Data Science: Import, Tidy, Transform, Visualize,; Model Data.