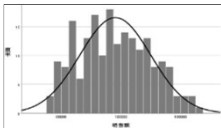
	应用数理统计方法
	<div>第二章 比较总体大小特征的假设检验</div> <div>2.1 假设检验与假设检验方法</div> <div>2.2 比较一个或两个总体均值的参数检验</div> <div>2.3 比较两个独立总体大小的非参数检验</div> <div>2.4 比较两个对应总体大小的非参数检验</div> <div>2.5 比较多个总体大小的非参数检验</div> <div>应用举例</div>

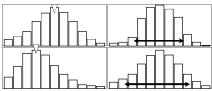
2.1.1
<div>2.1.1 假设检验</div> <div>2.1.2 假设检验思路与方法</div> <div>2.1.3 假设检验的风险和功效</div> <div>2.1.4 参数与非参数方法</div> <div>2.1.5 假设检验方法选择</div>

统计推断 复习	p.2-3
<div>▪ 统计推断</div> <div>经典数理统计的基本方法</div> <div>在概率论的基础上,通过观察样本,对总体特征做出判断</div> <div>包括参数估值和假设检验</div> <div>▪ 参数估值</div> <div>描述总体统计特征,即根据样本统计量估计总体统计量</div> <div>如 根据样本均值估计总体均值,根据样本分布判断总体分布特征</div> <div>▪ 假设检验</div> <div>据样本观测结果,针对总体的假设成立与否作出判断</div> <div>如 两总体大小是否相同,某因素是否有显著影响,两变量是否相关</div>	

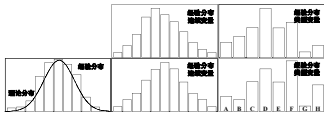
随机变量的特征 复习	p.20
<div>▪ 随机变量的特征</div> <div>个体取值具随机性,全部个体取值则有特定规律</div> <div>数理统计方法即研究这些规律的宏观特征</div> <div>许多学科的主要研究对象是随机变量 环境、生态、地学、生物、化学、社科、经济 ...</div> <div>▪ 统计特征</div> <div>分布 个体在不同取值区间出现的概率</div> <div>大小 分布中心的 - 在数轴上的位置</div> <div>离散 个体对分布中心的偏离程度 - 聚集/分散</div>	



假设检验问题 大小与离散比较举例
<div>▪ 个别方法已介绍</div> <div>异常剔除,独立性检验 都是假设检验方法</div> <div>▪ 总体大小特征比较</div> <div>湖泊 A 和 湖泊 B 中鱼的平均重量是否有差别? 两总体大小比较</div> <div>大,中,小三类湖泊湖水 pH 值是否相同? 三总体大小比较</div> <div>▪ 总体离散特征比较</div> <div>湖泊中草食性鱼与杂食性鱼的重量变幅是否一致? 两总体离散程度比较</div> <div>湖水 pH 值和天然有机物的变异是否有差别? 两总体离散程度比较</div>

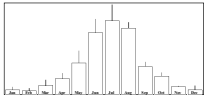


假设检验问题 分布检验举例
<div>▪ 经验分布与理论分布比较</div> <div>某鱼种大小是否服从正态分布,或对数正态分布? 正态检验</div> <div>湖泊丰水期水面面积是否服从正态分布? 正态检验</div> <div>某种鱼的雌雄比是否为 1? 二项检验</div> <div>▪ 经验分布与经验分布比较</div> <div>两种鱼的重量分布是否一致? 连续变量的拟合度检验</div> <div>两个湖泊中八种鱼的比例是否有差异? 类型变量的拟合度检验</div>



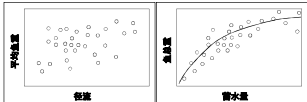
假设检验问题 影响因素研究举例

- 单因素
 - 不同月份湖泊水面面积有差别吗？具体哪些月不同？ 季节对面积的影响
 - 径流量不同的湖泊中某种鱼的重量是否有差别？ 径流对鱼重的影响
- 多因素
 - 汇流面积和深度不同湖泊水面面积有差别吗？ 汇流面积与深度对面积的影响
 - 不同水化学类型湖泊鱼的生长速率是否不同？ 类型/亚类对生长速率的影响
- 方差分析
 - 研究影响因素



假设检验问题 变量关系研究举例

- 两变量共变关系
 - 蓄水量较大的湖泊中的鱼是否也比较大？ 蓄水量与鱼的正相关关系
 - 藻类比较多的湖泊中的鱼是否也比较多？ 藻类与鱼的相关关系
- 预测模型
 - 能否/如何根据湖泊蓄水量估算鱼的总重量？ 一元回归
 - 能否/如何根据湖泊多种参数预测鱼的总数？ 多元回归
- 相关与回归
- 目的和用法不同
- 简单讨论
 - 方差, 相关和回归分析的关系



假设检验课程内容的安排 p. 63

- 特征比较
 - 两总体或多总体大小比较 2.2 – 2.5
 - 两总体或多总体离散程度比较 3.1 – 3.2
 - 总体分布特征比较 3.3 – 3.4
- 影响因素
 - 方差分析及补充分析 4.1 – 4.4
- 变量关系
 - 相关分析 5.1 – 5.2
 - 回归分析 5.3 – 5.4

2.1.2

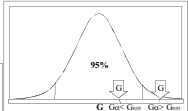
- 2.1.1 假设检验
- 2.1.2 假设检验思路与方法
- 2.1.3 假设检验的风险和功效
- 2.1.4 参数与非参数方法
- 2.1.5 假设检验方法选择

假设检验 p. 64-66

- 假设检验
 - 在概率论的基础上, 对是否接受特定统计假设进行判断的手段
- 方法
 - 据样本观测结果, 对针对总体的假设成立与否作出判断
 - 仅是在是与否两种可能性之间做出判断

假设检验的基本思路 p. 64-65


- 基本思路
 - 建立: 总体没有显著差别的原假设
 - 计算: 相伴概率 p , 即原假设成立条件下得到实际观察结果及更极端结果的概率
 - 更极端结果: 比实际观测结果更偏离原假设
 - 判断: 相伴概率太小则拒绝原假设, 否则接受原假设
- 更极端结果举例
 - 异常 更极端情况包括比可疑值更偏离分布中心的所有取值
 - 独立 观察游程为 28 或 3, 更极端情况: 2, 29, 30



假设检验思路举例 黑白棋子

▪ 问题

袋中有200粒棋子, 据称黑白各 100 粒, 混匀后随机取出 20 粒
A 白色 18 粒, 黑色 2 粒, 相信黑白一样多吗? 判断有把握吗?
B 白色 11 粒, 黑色 9 粒, 相信黑白一样多吗? 判断有把握吗?




假设检验思路举例 黑白棋子

▪ 问题

袋中有200粒棋子, 据称黑白各 100 粒, 混匀后随机取出 20 粒
A 白色 18 粒, 黑色 2 粒, 相信黑白一样多吗? 判断有把握吗?
B 白色 11 粒, 黑色 9 粒, 相信黑白一样多吗? 判断有把握吗?

▪ 猜测

前者黑白一样多的可能性 极小, 有很大把握
后者黑白一样多的可能性 比前者大的多, 但无把握
不能得出显著不同的判断




假设检验思路举例 黑白棋子

▪ 问题

袋中有200粒棋子, 据称黑白各 100 粒, 混匀后随机取出 20 粒
A 白色 18 粒, 黑色 2 粒, 相信黑白一样多吗? 判断有把握吗?
B 白色 11 粒, 黑色 9 粒, 相信黑白一样多吗? 判断有把握吗?

▪ 检验

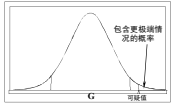
假设 袋中黑白棋子相等 无差别的原假设
计算 A 和 B 的概率 $9.4 \times 10^{-5}, 0.167$ 判断依据, 更极端情景
判断概率太小则判定黑白不同 拒绝原假设
否则 接受原假设
两种判断的可靠性完全不同 参见 2.1.2 两类风险



假设检验思路举例 异常值剔除

▪ 问题

正态分布总体中的可疑值, 判断是异常值吗? 是否属于该总体?
如果可疑值出现在分布的尾端, 会是异常吗? 判断有把握吗?
如果可疑值出现在分布的中部, 会是异常吗? 判断有把握吗?



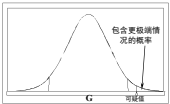
假设检验思路举例 异常值剔除

▪ 问题

正态分布总体中的可疑值, 判断是异常值吗? 是否属于该总体?
如果可疑值出现在分布的尾端, 会是异常吗? 判断有把握吗?
如果可疑值出现在分布的中部, 会是异常吗? 判断有把握吗?

▪ 猜测

前者是异常的可能性 很大, 有很大把握
后者是异常的可能性 较小, 但仍可能不属于这个总体,
不能得到有把握的判断



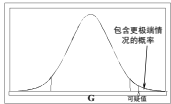
假设检验思路举例 异常值剔除

▪ 问题

正态分布总体中的可疑值, 判断是异常值吗? 是否属于该总体?
如果可疑值出现在分布的尾端, 会是异常吗? 判断有把握吗?
如果可疑值出现在分布的中部, 会是异常吗? 判断有把握吗?

▪ 检验

假设 可疑值属于总体, 不是异常 无差别的原假设
计算 可疑值及更极端值的出现概率 判断依据
判断 概率太小则判定为异常 拒绝原假设
否则 接受原假设
两种判断的可靠性完全不同 参见 2.1.2 两类风险



检验假设举例

▪ 日常生活中基于概率的决策

排队, 路径, 食物, 课程, 工作 ...

优化

▪ 黑白棋子

H₀ 黑白棋子个数相同; H₁ 黑白棋子个数不同

黑白棋子个数是否相同

▪ 异常值剔除

H₀ 可疑值不是异常; H₁ 可疑值是异常

可疑值是否属于该总体

▪ 独立性检验

H₀ 个体具有独立性; H₁ 个体不具独立性

与独立总体无差别

检验假设举例

▪ 主要特征比较

两类湖泊湖水 pH 是否一致 ?

H₀ $\mu_1 = \mu_2$, H₁ $\mu_1 \neq \mu_2$

总体均值

湖水硬度和 pH 变幅相同吗 ?

H₀ $V_{p_1} = V_{p_2}$, H₁ $V_{p_1} \neq V_{p_2}$

总体变异系数

湖泊面积是否为正态分布 ?

H₀ $\gamma_1, \gamma_2 = 0$, H₁ $\gamma_1, \gamma_2 \neq 0$

总体偏峰系数

▪ 影响因素研究

汇流面积不同湖泊面积是否相同 ?

H₀ μ_1 都相同, H₁ μ_1 都不相同

总体均值

▪ 变量关系研究

湖泊径流量与面积有关吗 ?

H₀ $\rho = 0$, H₁ $\rho \neq 0$

总体相关系数

能否据湖泊蓄水量预测鱼产量 ?

$Y = a + bX$, H₀ $\beta = 0$, H₁ $\beta \neq 0$

总体斜率

原假设与对立假设

p.64-65

▪ 两者关系

包括所有可能

相互矛盾

对立假设远比原假设包含更多可能性, 两者不能互换

▪ 举例

黑白棋子 相等或不等

H₀ 100:100 vs. H₁ 0:200 - 99:101, 101:99 - 200:0

大小比较 相同或不同

H₀ $\mu_1 = \mu_2$ 差别为 0, H₁ $\mu_1 \neq \mu_2$ 差别为 0 以外所有可能性

▪ 例外

大小比较

H₀ $\mu_1 = \mu_2$ 差别为 0

H₁ $\mu_1 = \mu_2 + C$ 差别为 C

假设检验的基本思路

复习

p.64-65

▪ 基本思路

建立: 总体没有显著差别的原假设

计算: 相伴概率 p 原假设成立条件下得到实际观察结果及更极端结果的概率

– 更极端结果: 比实际观测结果更偏离原假设

判断: 相伴概率太小则拒绝原假设, 否则接受原假设

描述统计量与检验统计量

p.67

▪ 描述统计量

用于参数估值

根据样本观察结果获得样本参数, 据此估计总体参数

点估计与区间估计

如: 算术均值, 标准差, 几何均值, 相关系数, 斜率 等 及其置信区间

▪ 检验统计量

用于假设检验

基于样本观察结果及相应检验方法

如 异常值剔除 Grubbs 检验中的 G, t-检验中的 t, 方差分析中的 F 等

用于计算相伴概率 p 统计软件直接给出计算结果

少数情况可直接计算 如新冠阳性概率计算, 黑白棋子判断, 随机化检验 参见 2.3.4

个体分布与检验统计量分布

p.67

▪ 总体中个体的分布

随机变量中个体出现规律 总体特征之一

检验方法选择的依据

参数方法与非参数方法

▪ 检验统计量的分布

用于假设检验

定义检验值与 p 的关系 相伴概率的计算基础

如 Grubbs 检验中的标准正态分布

t-检验中的 t-分布

方差分析中的 F-分布

相伴概率

p.71

▪ 概率密度函数

取值区间与概率有确切的对应关系

▪ 相伴概率

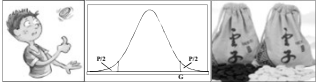
原假设成立条件下,得到实际观察结果及更极端结果的概率

如: 抛 100 次硬币, 65 次正, 更极端情况 66, 67, 68, ... 100 假设: 50:50

如: 异常值剔除 Grubbs 法中, 可疑值及更偏离均值的取值 假设: 属于该总体

如: 抽取 20 粒棋子, 18 粒白, 更极端情况为 19 和 20 假设: 50:50

大部分情况下统计软件直接计算 p 值



假设检验的基本思路 复习

p.64-65

▪ 基本思路

建立: 总体没有显著差别的原假设

计算: 相伴概率 p 原假设成立条件下得到实际观察结果及更极端结果的概率

– 更极端结果: 比实际观测结果更偏离原假设

判断: 相伴概率太小则拒绝原假设, 否则接受原假设

显著性水平

p.71

▪ 定义

拒绝客观上正确原假设的最大概率, 又称拒绝水平, 记为 α

主观选择, 在检验前确定以确保检验的客观性 得到 p 值前

拒绝原假设: 表达为在 α 水平下显著 即有显著差异

一般取值: 0.01, 0.05, 0.10

▪ 与相伴概率的关系

p 为实际概率

α 是拒绝原假设的概率边界

假设检验的判断

p.71-72

▪ 直接判断

计算相伴概率, 若 $p < \alpha$, 则拒绝原假设

拒绝原假设时, 可了解确切错误率

▪ 间接判断

计算检验统计量 如 Grubbs 检验中的 G

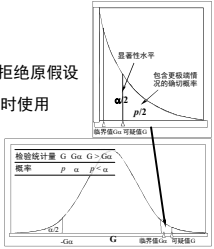
查临界值, 若统计量大于临界值 如 $G \geq G_{\alpha}$, 则拒绝原假设

此时仅知道错误率小于 α , 只在不能计算 p 值时使用

▪ 直接判断与间接判断的对应关系

计算值与相伴概率对应 $G \sim p$

临界值与显著性水平对应 $G_{\alpha} \sim \alpha$



临界值查取

p.73, 378

▪ 临界值表

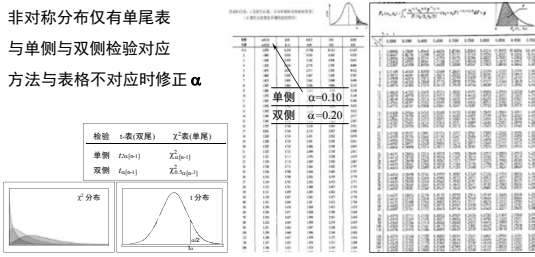
在无法直接获得 p 的情况下使用临界值, 与检验统计量比较 教材附表

▪ 单尾表与双尾表

非对称分布仅有单尾表

与单侧与双侧检验对应

方法与表格不对应时修正 α



双侧检验与单侧检验

p.68

▪ 双侧检验

如果有差异, 可能出现在相反方向上, α 均分在分布的两端

如 黑棋多或白棋多, 高异常或低异常, 游程太大或太小

▪ 单侧检验

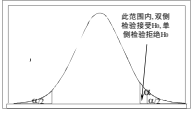
若有差异, 只可能出现在一个方向上 根据已有知识已排除另一半可能

对立假设包含的可能为双侧检验的一半 α 在分布的一端

如 黑棋只可能比白旗多, 只可能高异常, 只可能出现极小游程

在较少可能性中选择, 拒绝 H_0 的可能性增加

参见下节



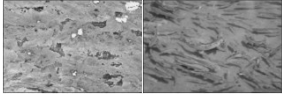
双侧与单侧检验问题举例

双侧检验

大型湖泊和小型湖泊湖水硬度是否有差别？差别可以是双向的
两种鱼大小是否有差异？若有差别，两种鱼都可能偏大

单侧检验

流域降水量增加是否导致湖泊面积增加？ 降水增加，湖泊面积不可能减少
鱼体内农药含量与湖水农药含量是否相关？不可能负相关



假设检验的步骤 p.73

基本方法

建立: 总体没有显著差别的原假设 及相应的对立假设
计算: 相伴概率 p 原假设成立条件下得到实际观察结果及更极端结果的概率
判断: 相伴概率太小则拒绝原假设 比较检验统计量与临界值

步骤

直接判断和间接判断 差别在于拒绝 H_0 时能否获得确切的错误概率

提出
检验假设

计算
相伴概率

与显著性
水平比较

作出判断

确定显著
性水平 α

计算
检验统计量

与检验临
界值比较

2.1.3

2.1.1 假设检验

2.1.2 假设检验思路与方法

2.1.3 假设检验的风险和功效

2.1.4 参数与非参数方法

2.1.5 假设检验方法选择

统计推断的可能错误 p.70

参数估值的可能错误

根据样本参数估计总体参数
点估计的错误率表达: 置信区间与相应概率

假设检验的可能错误

基于样本数据, 针对总体作出判断
判断结果可能不对 不能穷尽所有个体
接受和拒绝原假设的错误率不同 原假设和对立假设包含的可能性不同
结论的可靠性完全不同


假设检验思路举例 黑白棋子 复习

问题

袋中有200粒棋子, 据称黑白各100粒, 混匀后随机取出20粒
若 白色18粒, 黑色2粒, 相信黑白一样多吗? 判断有把握吗?
若 白色11粒, 黑色9粒, 相信黑白一样多吗? 判断有把握吗?

检验

假设 袋中黑白棋子相等 无差别的原假设
计算 白色 ≥ 18 或 ≥ 11 的概率 判断依据, 更极端情景
判断 概率太小则判定黑白不同 拒绝原假设
否则 接受原假设
两种判断的可靠性完全不同 见 2.1.2 两类风险



假设检验的不确定性 p.70

两类错误

原假设有两种可能: 正确或错误 检验者不了解的客观事实
主观判断也有两种可能: 接受或拒绝 都可能错, 但错误率不同
共四种可能, 两种是错的
第 I 类错误: 拒绝客观上正确的原假设 错误率 α , 主观设定
第 II 类错误: 接受客观上错误的原假设 错误率 β , 客观存在

举例

黑白棋子 接受与拒绝的风险
异常值剔除 接受与拒绝的风险

	接受原假设	拒绝原假设
假设客观上正确	正确: $1 - \alpha$	错误 I: α
假设客观上错误	错误 II: β	正确: $1 - \beta$

第 II 类错误率

p.70

▪ 特点

不可计算 已知分布的检验统计量是基于原假设构建的

不可控制 α 控制

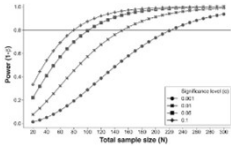
远高于 α 对立假设包含的可能性一般远多于原假设

▪ 第 II 类错误率的变化趋势

尽管不能计算,但可以表征其变化的一般特征

构建假设总体,抽样模拟 β 随 α 和 n 变化

方法选择的依据 不能用于实际计算



判断的不确定性和结论

p.70

▪ 判断的不确定性

任何判断都没有绝对把握 由于不能穷尽全部个体,类似参数估值

拒绝和接受原假设的不确定性不同 由于两种假设包含的可能性不同

例: 从 2 种或 256 种颜色中猜

拒绝原假设 不确定性低,可控,已知且较小的 p

接受原假设 不确定性高,不可控,未知且较大的 p

▪ 假设检验的结论

拒绝原假设,可以有明确的结论 已知,且可控的错误率 p 或 $\leq \alpha$

接受原假设,不能下明确的结论 未知,且很高的错误率 β

应用实例 药物疗效比较

▪ 关于假设检验结论的例子

强调结论表述,具体方法在后续各节中介绍

▪ 研究

问题 对比两种药物治疗某疾病的效果

方法 募集两组各 100 名患者,分别用两种药物治疗


数据 采集两组人群的疗效

▪ 结果与结论

假设 $H_0: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2$ 两种药物疗效有/无差别

结果 $p < 0.05$, 拒绝原假设 两种药物有显著差别

结果 $p > 0.05$, 接受原假设 没有观察到差别



应用实例 鱼重与径流量间关系

▪ 研究

问题 湖泊中鱼的重量与径流量有关吗?

方法 随机抽 50 个湖泊,每个湖采集 50 条鱼,

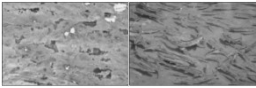
数据 分别测量湖泊年均径流量和鱼重

▪ 结果与结论

假设 $H_0: \rho = 0; H_1: \rho \neq 0$ 两者相关/不相关

结果 $p < 0.05$ 拒绝原假设 重量与径流量显著相关

结果 $p > 0.05$ 接受原假设 看不到鱼重与径流量之间有相关关系



应用实例 湖泊流量的正态检验

▪ 研究

问题 某区域湖泊径流量是否服从正态分布 使用参数方法的前提

方法 随机抽取 100 个湖泊

数据 测量湖泊的年均径流量

▪ 结果与结论

假设 $H_0: \gamma_1 = 0, \gamma_2 = 0; H_1: \gamma_1 \neq 0, \gamma_2 \neq 0$ 偏度系数 γ_1 和峰态系数 γ_2

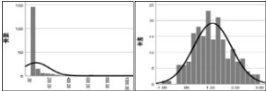
结果 $p_1 < 0.05, p_2 < 0.05$, 结果显著 拒绝原假设,分布偏离正态,否则接受

结果 若对原始数据做对数变换,检验结果均不显著

结论 不是正态分布 此结论可靠

结论 是对数正态分布 此结论不可靠

讨论 现实与妥协,经验,误判风险



假设检验的功效

p.74

▪ 假设检验的功效

拒绝客观上错误原假设的可能性 $1 - \beta$ 对比第二类错误率 β

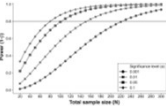
不能就具体问题计算功效 不能计算 β

▪ 应用

可模拟获得特定非参数方法功效随 α 和 n 变化的理论趋势

是方法选择的依据 基于 n

	接受原假设	拒绝原假设
假设客观上正确	正确: $1 - \alpha$	错误 I: α
假设客观上错误	错误 II: β	正确: $1 - \beta$



影响检验功效的因素

p.74-75

不同方法的差别

参数方法的功效总是不低于相应的非参数方法 同类问题, 如总体大小比较
同类非参数方法的功效与样本量有关 如 U-检验和随机化检验

其他因素

样本量大小 样本量越大, 置信区间越小, 功效越高 代表性好
显著性水平 α 越高, 拒绝客观上错误原假设机会越大 不应为提高功效改变 α
对立假设 包含可能性越少, 功效越高, 如:
 $H_0: \mu = 10, H_1: \mu \neq 10$ 双侧检验
 $H_0: \mu = 10, H_1: \mu < 10$ 单侧检验, 排除一半可能
 $H_0: \mu = 10, H_1: \mu = 30$ 原假设与对立假设包括的可能性相当

2.1.4

2.1.1 假设检验

2.1.2 假设检验思路与方法

2.1.3 假设检验的风险和功效

2.1.4 参与与非参数方法

2.1.5 假设检验方法选择

总体分布与参数估值 回顾

p.36-39

总体大小表征

正态分布总体, 用算术均值 代表集中趋势
非正态分布, 可做正态变换, 仍用算术均值 变换与逆变换
非正态分布, 不能或不便变换, 用中位数 算术均值不能代表集中趋势

总体其它特征表征

离散, 偏度, 峰态, 相关, 截距, 斜率等计算均以算术均值为基础
总体参数 $\mu, \sigma, \sigma^2, \gamma_1, \gamma_2, \rho, \beta, \alpha, \dots$

总体分布与特征比较

p.76-78

总体大小比较

正态分布总体, 比较均值 比较集中趋势
非正态分布, 可做正态变换, 用类似方法比较 数据变换, 如对数变换
非正态分布, 不能或不便变换, 不能比较均值 非参数方法

总体其它特征比较

正态分布总体可用基于均值的其它参数 $\mu, \sigma, \sigma^2, \gamma_1, \gamma_2, \rho, \beta, \alpha, \dots$
非正态分布总体不能依赖总体参数比较 非参数方法

假设检验的参数与非参数方法

p.76

参数方法

基于参数的经典方法 类似参数估值, 基于以均值为基础的参数
如 异常值的 Grubbs 检验

非参数方法

不依赖参数的无分布方法 基于秩数据或精确概率计算
如 个体独立性波动游程检验
如 黑白棋子判断
如 随机化检验 参见 2.3.3

参数与非参数检验方法对比

p.76

参数与非参数方法的差别

本质差别: 是否基于总体参数
其它差别: 源自总体参数/非总体参数的差别 附表

方法选择

正态分布: 优先选择功效高的参数方法 如总体大小比较的 t-检验
非正态分布: 正态变换后用参数方法, 或直接用非参数方法 如 U-检验
类比参数估值的非参数方法

	参数	非参数
检验方法	经典方法	无分布方法
总体参数	基于总体参数	不依赖总体参数
变量类型	连续量, 离散量	顺序量, 类型量
信息利用	充分	仅利用顺序信息
检验功效	相对高	相对低
样本要求	高	低
异常干扰	强	弱

非参数方法的功效

p.77

非参数方法的检验功效

非参数方法的选择依据

对同类问题 仅有唯一的参数方法 如两总体大小比较 t-检验

常有不止一种非参数方法 如大小比较的 U-检验与随机化检验

不同非参数方法的功效不同 如 U-检验与随机化检验

参数方法与非参数方法的检验功效

非参数方法的检验功效不可能高于相应的参数方法 信息利用不充分

	接受原假设	拒绝原假设
假设客观上正确	正确: $1 - \alpha$	错误 I: α
假设客观上错误	错误 II: β	正确: $1 - \beta$

非参数方法的功效效率

p.77

功效效率

非参数方法对应于响应参数方法的相对功效 相对大小

达到相同的检验功效, 非参数方法需要更大的样本量 对应方法

定义: 达到同等功效的参数与非参数方法样本量比 $100 N_p/N_n$

举例

相同功效, 参数方法 $N_p = 20$, 非参数方法 $N_n = 25$, $1 - \beta = 100\% \times 20/25 = 80\%$

非参数方法的功效效率 = 功效效率的应用

选择方法的依据, 但不能直接计算 因为 β 未知

模拟获得一般性趋势

如 n 趋向无穷时, U-检验功效效率趋向 95%

2.1.5

2.1.1 假设检验

2.1.2 假设检验思路与方法

2.1.3 假设检验的风险和功效

2.1.4 参数与非参数方法

2.1.5 假设检验方法选择

假设检验方法的前提

p.77-78

基本前提

采样随机性 适用于所有统计方法, 不可检验

个体独立性 适用于所有统计方法, 波动游程检验功效很低 第二类错误率很高

通过可靠的实验设计和采样方法保证

参数方法的前提

分布正态性 适用于所有参数方法, 有时可用正态变换获得

方差同质性 方差分析的不同水平, 回归分析的不同自变量 一般 n 太小, 很难检验

方差加和性 方差分析的不同因子, 方差分析的基础 可用于考察交互作用

数据中个体的对应关系

p.95

个体对应关系

受特定因素影响, 两个或多个总体中个体具有对应且相同的变化规律

有对应关系的两总体数据构成成对数据 不限于两个总体

这样的对应关系可体现在个体, 时间, 空间等方面

对应总体中个体的数据构成

无对应关系: 总体差异 + 随机波动 个体间差异也是随机的

有对应关系: 总体差异 + 个体间差异 + 随机波动 两总体个体间差异一致

例: 蓝色和棕色两样本比较

重要概念

下章重复

数据的对应关系

p.95

对应关系的利用

用于大小比较, 排除规律性个体差异的影响, 提高检验功效

大小比较基于差异/波动相对大小 更多例子参见 2.2.5

个体对应关系举例

时间 某河流上下游 BOD 差别: 按时间随机采样 vs. 按时间同步采样

空间 北京两年 PM2.5 差别: 两年分别随机布点 vs. 两年在相同点位

个体 下页

更多例子参见 2.1

成对数据举例

▪ 据称某药物能控制体重, 征集志愿者测试药效

方案一 200 名志愿者, 分两个组, 分别服用药物和安慰剂 随机分组

方案二 100 名志愿者, 分两阶段, 分别服用药物和安慰剂 个体随机分阶段

▪ 影响体重的因素

药物作用 可能的药物作用, 关注的问题


个体差异 志愿者个体间差异, 方案二关注每人服药与否的差别, 排除个体差异影响

随机效应 纯随机, 所有个体自身的随机波动

▪ 数据分析

方案一 相对个体差异和随机波动, 药效是否显著

方案二 相对随机波动, 药效是否显著, 排除了个体差异影响



检验方法选取原则

▪ 检验目的

总体 根据研究目的确定总体及总体个数 特征, 变量 大小, 离散, 分布

▪ 检验功效

原则上选择功效高的方法 功效差别不大时例外, 如: 大样本量 U-检验

▪ 其他因素

数据特点: 分布特征, 个体对应关系 如 非参数检验, 成对数据比较

方法假设: 对立假设覆盖的可能性 如 单侧检验

样本量: 影响非参数方法的功效 如 U-检验和随机化检验的选择

保守性: 保守性/严格性, 相当于改变 α 如 异常值 Grubbs 检验和 t-检验

其它: 等价方法, 无概率意义的简便方法 如 拟合度卡方检验和 G-检验

统计方法汇总表

▪ 检验目的

特征比较 大小, 离散, 分布

影响因素 方差分析

变量关系 相关, 回归

其他检验 如数据预处理

▪ 其他因素

总体个数 一个, 两个或多个

个体关系 是否有对应关系

假设范畴 单侧或双侧

其它 如样本量大小

▪ 参数与非参数方法

参数方法	非参数方法
正态分布检验	秩和检验
方差分析	秩和检验
回归分析	秩和检验
相关分析	秩和检验
多元回归	秩和检验
主成分分析	秩和检验
判别分析	秩和检验
聚类分析	秩和检验
因子分析	秩和检验
路径分析	秩和检验
结构方程模型	秩和检验
贝叶斯统计	秩和检验
马尔可夫链蒙特卡罗	秩和检验
深度学习	秩和检验
图神经网络	秩和检验
强化学习	秩和检验
生成对抗网络	秩和检验
变分自编码器	秩和检验
自监督学习	秩和检验
迁移学习	秩和检验
联邦学习	秩和检验
知识蒸馏	秩和检验
模型压缩	秩和检验
模型解释	秩和检验
模型鲁棒性	秩和检验
模型泛化能力	秩和检验
模型性能评估	秩和检验
模型超参数调优	秩和检验
模型集成	秩和检验
模型融合	秩和检验
模型迁移	秩和检验
模型部署	秩和检验
模型维护	秩和检验
模型更新	秩和检验
模型退役	秩和检验

要点

▪ 假设检验

问题: 大小, 离散, 分布, 影响因素, 相关, 回归

思路: 建立原假设-对立假设, 计算相伴概率, 判断

定义: 相伴概率-显著性水平; 检验值-临界值

▪ 假设检验方法

参数方法与非参数方法

双侧检验与单侧检验

成对数据检验

▪ 判断的风险

两类错误, 功效与功效效率

谢谢

