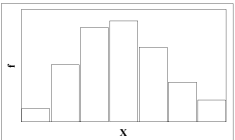
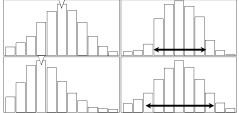


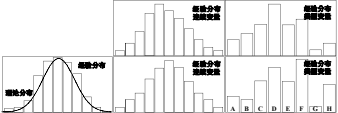
	应用数理统计方法 2024.7
	第二章 比较总体大小特征的假设检验 2.1 假设检验与假设检验方法 2.2 比较一个或两个总体均值的参数检验 2.3 比较两个独立总体大小的非参数检验 2.4 比较两个对应总体大小的非参数检验 2.5 比较多个总体大小的非参数检验 应用举例

2.1.1
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 复习	p.2-3
■ 统计推断 在概率论的基础上, 根据对样本的观测判断总体的某些特征 包括参数估值和假设检验 ■ 参数估值 描述总体统计特征, 即根据样本统计量对总体统计量进行估计 如 根据样本均值估计总体均值 ■ 假设检验 据样本观测结果, 对针对总体的假设成立与否作出判断 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	

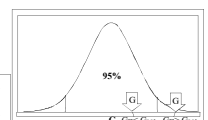
随机变量的主要特征 复习	p.20
■ 随机变量的特征 宏观规律的表现 ■ 统计学特征 分布 不同取值 (范围) 出现的概率 大小 分布中心在数轴上的位置, 取值大小 离散 个体到分布中心的距离, 聚集/分散程度	
	

假设检验问题 大小与离散比较举例
■ 个别方法之前已介绍 如 异常剔除, 独立性检验 ■ 总体大小特征比较 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 大, 中, 小三类湖泊湖水 pH 值是否有差别? 三总体大小比较 ■ 总体离散特征比较 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 湖水 pH 值和湖水天然有机物的变异是否有差别? 两总体离散程度比较


假设检验问题 分布检验举例
■ 经验分布与理论分布比较 某鱼种大小是否服从正态分布, 抑或对数正态分布? 正态检验 湖泊丰水期水面面积是否服从正态分布? 某种鱼的雌雄比是否为 1? 二项检验 ■ 经验分布与经验分布比较 两种鱼的重量分布是否一致? 拟合度检验 两个湖泊中八个主要鱼种的比例是否有差异?


Month	Number of cases (approx.)
Jan	10
Feb	15
Mar	30
Apr	40
May	60
Jun	100
Jul	110
Aug	100
Sep	50
Oct	30
Nov	15
Dec	10

28	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
14	+	+	+	+	-	+	+	-	+	+	-	+	-	-	+	-	+	+	-
3	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-



假设检验思路举例 黑白棋子

问题

袋中有棋子共 200 粒, 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, 相信黑白一样多的说法吗? 判断有把握吗?
若 白色 11 粒, 黑色 9 粒, 相信黑白一样多的说法吗? 判断有把握吗?



假设检验思路举例 黑白棋子

问题

袋中有棋子共 200 粒, 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, 相信黑白一样多的说法吗? 判断有把握吗?
若 白色 11 粒, 黑色 9 粒, 相信黑白一样多的说法吗? 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, 有把握吗?



假设检验思路举例 黑白棋子

问题

袋中有棋子共 200 粒, 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, 相信黑白一样多的说法吗? 判断有把握吗?
若 白色 11 粒, 黑色 9 粒, 相信黑白一样多的说法吗? 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	若原假设成立的前提
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	

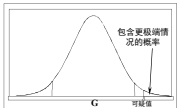


假设检验思路举例 异常值剔除

p.52

问题

从正态分布总体中采样, 判断是否有异常值 可疑值为最大和最小值
如果可疑值出现在分布的尾端, 会是异常吗? 远离均值, 属于总体概率小
如果可疑值出现在分布的中部, 会是异常吗? 接近均值, 属于概率概率大



假设检验思路举例 异常值剔除

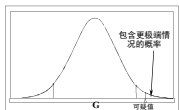
p.52

问题

从正态分布总体中采样, 判断是否有异常值 可疑值为最大和最小值
如果可疑值出现在分布的尾端, 会是异常吗? 远离均值, 属于总体概率小
如果可疑值出现在分布的中部, 会是异常吗? 接近均值, 属于概率概率大

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, 仍可能不属于这个总体, 有把握吗?



假设检验思路举例 异常值剔除

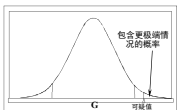
p.52

问题

从正态分布总体中采样, 判断是否有异常值 可疑值为最大和最小值
如果可疑值出现在分布的尾端, 会是异常吗? 远离均值, 属于总体概率小
如果可疑值出现在分布的中部, 会是异常吗? 接近均值, 属于概率概率大

检验

假设 可疑值属于总体, 不是异常	无差别的原假设
计算 可疑值及更偏值的出现概率	如果原假设成立: 可疑值属于该总体
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		
▪ 日常生活中的判断	排队选择, 路径选择, 食物选择 ...	
▪ 黑白棋子	H ₀ 黑白棋子个数相同; H ₁ 黑白棋子个数不同 黑白棋子个数是否相同	
▪ 异常值剔除	H ₀ 可疑值不是异常; H ₁ 可疑值是异常	可疑值是否来自该总体
▪ 独立性检验	H ₀ 个体具有独立性; H ₁ 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? H ₀ $\mu_1 = \mu_2$, H ₁ $\mu_1 \neq \mu_2$ 总体均值 湖水硬度和 pH 变异程度相同吗? H ₀ $V_{p1} = V_{p2}$, H ₁ $V_{p1} \neq V_{p2}$ 总体变异系数 湖泊面积是否为正态分布? H ₀ $\gamma_1 = 0, \gamma_2 = 0$; H ₁ $\gamma_1 \neq 0, \gamma_2 \neq 0$ 总体偏峰系数	
▪ 影响因素研究	汇流面积不同湖泊面积是否有差异 H ₀ μ_1 都相同, H ₁ μ_1 不都相同, 总体均值	
▪ 变量关系研究	湖泊径流量与面积有关吗? H ₀ $\rho = 0$, H ₁ $\rho \neq 0$ 总体相关系数 能否据湖泊蓄水量预测鱼产量? $Y = a + bX$, H ₀ $\beta = 0$, H ₁ $\beta \neq 0$ 总体斜率	

原假设与对立假设	<i>p.64-65</i>
▪ 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 H ₀ 100:100 vs. H ₁ 0:200 - 99:101, 101:99 - 200:0 大小比较 相同或不同 H ₀ $\mu_1 = \mu_2$, 差别为 0, H ₁ $\mu_1 \neq \mu_2$ 差别为 0 以外所有可能值
▪ 例外	大小比较 H ₀ $\mu_1 = \mu_2$ 差别为 0 H ₁ $\mu_1 = \mu_2 + C$ 差别为 C

假设检验思路 复习	<i>p.64-65</i>
▪ 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 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	<i>p.67</i>
▪ 描述统计量	用于参数估值 根据样本获得样本参数, 据此估计总体参数 点估计与区间估计 如: 算术均值, 标准差, 几何均值等 及其置信区间
▪ 检验统计量	用于假设检验 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

总体分布与检验统计量分布	<i>p.67</i>
▪ 总体中个体的经验分布	随机变量中个体出现规律 总体特征之一 决定检验方法选择 参数方法与非参数方法
▪ 检验统计量的理论分布	用于假设检验, 定义检验值与 p 的关系 相伴概率计算的基础 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

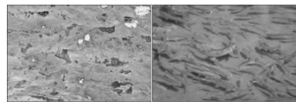
双侧与单侧检验问题举例

■ 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
杂食性鱼和草食性鱼大小是否有差异？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



假设检验步骤

p.73

■ 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

判断：相伴概率太小则拒绝原假设

■ 步骤

直接判断和间接判断 差别在于能否获得准确的相伴概率



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

区间估计：将点估计的可能错误表达为置信区间

■ 假设检验的可能错误

检验假设基于样本，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

结论的把握不同

假设检验思路举例 黑白棋子

■ 问题

袋中有棋子共 200 粒，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，相信黑白一样多的说法吗？判断有把握吗？
若 白色 11 粒，黑色 9 粒，相信黑白一样多的说法吗？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 若原假设成立的前提
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

原假设客观上有两种可能：正确或错误 检验者不了解的事实

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
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 ，客观事实

■ 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

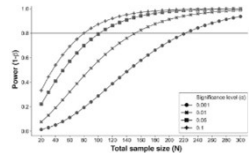
p.70

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 控制了 α
远高于 α 对立假设包含的可能性远多于原假设

第 II 类错误率的理论表征

可构建假设总体, 模拟 β 随 α 和 n 变化的一般趋势 方法选择的依据
不能用于实际计算



假设检验判断的不确定性

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
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完全不同
例: 从 2 种或 256 种颜色中猜

拒绝原假设 不确定性低, 可控, p 或 $\leq \alpha$

接受原假设 不确定性高, 不可控, 未知, β

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

实例 药物疗效大小比较

关于假设检验结论的例子

强调结论, 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 $H_0: \mu_1 = \mu_2$; $H_1: \mu_1 \neq \mu_2$ 两种药物疗效无/有差别

结果 $p > 0.05$, 接受原假设

结论 没有观察到两种药物的差别



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?

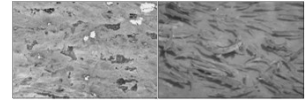
方法 随机抽 50 个湖泊, 每个湖采集 50 条鱼, 测量湖泊年均径流量和鱼重

结果与结论

假设 $H_0: \rho = 0$; $H_1: \rho \neq 0$ 总体相关系数 ρ 为 0 则不相关, 否则相关

结果 $p = 0.15, > 0.05$ 接受原假设, 重量与径流量不相关

结论 看不到鱼的重量与径流量之间有相关关系 无明确结论



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, 测量湖泊年均径流量

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

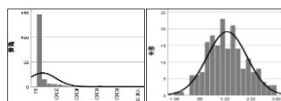
结果 $p_1 < 0.05, p_2 < 0.05$, 结果显著, 拒绝原假设, 分布偏离正态

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 现实和妥协, 误判的风险



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

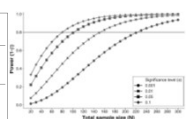
不能计算 β

应用

可模拟获得非参数方法功效随 α 和 n 的变化趋势

根据样本量大小选择方法依据

	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: $1 - \beta$



参数与非参数检验方法对比		p.76
<ul style="list-style-type: none"> 参数与非参数方法的差别 <p>本质差别: 是否基于总体参数</p> <p>其它差别: 与总体参数有关的差别 附表</p> 方法选择 <p>正态分布总体: 优先选择功效高的参数方法 不绝对, 如大样本量 U-检验</p> <p>非正态分布总体: 正态变换后用参数方法, 或直接用非参数方法 类似参数估值</p> 		
	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

非参数方法的功效效率	p.77
<ul style="list-style-type: none"> 非参数方法的检验功效 $1 - \beta$ <p>同一问题, 一般仅有唯一的参数方法, 但可能有不止一个非参数方法</p> <p>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</p> <p>对同类问题, 不同非参数检验方法的检验功效不同 如两个总体大小比较</p> 功效效率 <p>非参数方法对应于参数方法的相对功效 $\leq 100\%$ 相同功效, 非参数方法需较大 n</p> <p>计算: 达到同等功效的参数与非参数方法样本量比 $100 N_p/N_n$</p> <p>举例 实现相同功效, 参数方法 $N_p = 20$, 非参数方法 $N_n = 25$</p> <p>非参数方法功效效率 = $100\% \times 20/25 = 80\%$</p> <p>选择方法的依据, 不能直接计算 因为 β 未知</p> <p>只可模拟获得<u>一般性趋势</u> 如 n 趋向无穷时, U-检验功效效率趋向 95%</p> 	

2.1.5
<ul style="list-style-type: none"> 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
<ul style="list-style-type: none"> 基本要求 <p>采样随机性 所有统计方法的前提, 不可检验</p> <p>个体独立性 所有统计方法的前提, 波动游程检验功效很低, <u>第二类错误率</u>很高</p> <p>通过可靠的实验设计和采样方法保证</p> 参数方法的要求 <p>分布正态性 所有参数方法的前提, 有时可用正态变换获得</p> <p>方差同质性 方差分析的不同水平, 回归分析的不同自变量, <u>一般 n 太小, 很难检验</u></p> <p>方差加和性 方差分析的不同因子, 方差分析的基础, <u>可用来考察交互作用</u></p> 	

数据中个体的对应关系	p.95
<ul style="list-style-type: none"> 个体对应关系 <p>受特定因素影响, 两个或多个总体中个体具有对应的共同变化规律</p> <p>有对应关系的两总体数据构成<u>成对数据</u> 不限于两个总体</p> <p>这样的对应关系可体现在<u>个体, 时间, 空间</u>等方面</p> 对应总体的数据构成 <p>有对应关系: 总体差异 + 个体间差异 + 随机波动 个体间差异影响一致</p> <p>无对应关系: 总体差异 + 随机波动 个体差异包含在随机波动中</p> 	

数据的对应关系	p.95
<ul style="list-style-type: none"> 对应关系的利用 <p>用于大小比较, 排除规律性个体差异的影响, 提高检验功效</p> <p>大小比较是基于<u>差异/波动</u>相对大小</p> <p>在成对数据大小比较方法中介绍更多实例 参见 2.2.5, 2.4, 2.5.3</p> 个体对应关系举例 <p>个体 100 只服药小鼠和 100 只未服药小鼠对比; 100只小鼠服药前后的差别</p> <p>时间 比较某年某河流上下游 BOD, 按时间随机采样或同步采样</p> <p>空间 比较两年各随机抽 100 个城市的 PM2.5; 随机抽 100 个城市比较两年 PM2.5</p> 	

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 分组随机
方案二 100 名志愿者, 分两阶段, 分别服用药物和安慰剂 两阶段随机
- 影响体重的因素
药物作用 药物可能的作用, 关注的问题
个体差异 志愿者个体差异, 方案二考察个体用药前后关系, 可排除个体差异影响
随机效应 纯随机, 各组各阶段所有个体之间
- 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

p.79-80

- 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 如选择大样本量 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 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

方法汇总表

p. 81-83 表 2-4

- 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 如预处理等
- 其他因素
总体个数 1个, 2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

表 2-4 参数与非参数假设检验方法一览表		
检验目的	参数方法	非参数方法
样本均值比较	独立样本 t-检验 方差分析	秩和检验 中位数上下侧检验 Von Neumann 试验 Wilcoxon 符号秩检验
方差和协方差比较	F-检验 卡方检验 Levene 检验 Bartlett 检验	Wilcoxon 符号秩检验
相关性检验	Pearson 相关系数 Spearman 秩相关系数 Kendall 秩相关系数	Wilcoxon 符号秩检验
总体均值与特定值比较	t-检验 F-检验	Wilcoxon 符号秩检验 Mann-Whitney U-检验 Sign 检验 Kolmogorov-Smirnov 检验 游程检验
两独立总体大小比较	卡方检验 F-检验	Wilcoxon 符号秩检验 Mann-Whitney U-检验 Sign 检验 Kolmogorov-Smirnov 检验 游程检验
两相关总体大小比较	配对 t-检验 F-检验	Wilcoxon 符号秩检验 Mann-Whitney U-检验 Sign 检验 Kolmogorov-Smirnov 检验 游程检验
多个独立总体大小比较	方差分析 (F-检验) ANOVA Scheffé 多重比较 Tukey 多重比较 Duncan 多重比较 LSD 法多重比较 Bonferroni 法多重比较 GT2 多重比较 一般 SNK 多重比较 新 SNK 多重比较	Kruskal-Wallis 检验 Mann-Whitney U-检验 Sign 检验 Kolmogorov-Smirnov 检验 游程检验

谢谢

