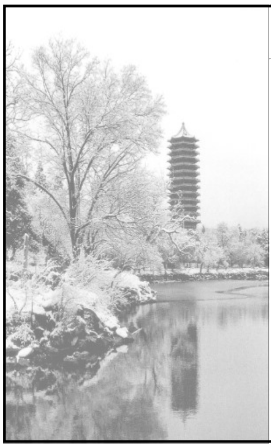


应用数理统计方法
2024.7

第四章 方差分析

4.1 方差分析方法
4.2 单因子方差分析及其补充分析
4.3 双因子和多因子方差分析
4.4 二级与多级因子方差分析
应用举例

假设检验问题 复习

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
两总体分布特征比较 3.4
总体分布是否服从特定理论分布 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

大小比较检验方法 复习

p.84-87

- 大小比较假设检验方法

参数方法与非参数方法
涉及一个, 两个或多个总体比较 两个以上总体可能涉及对应关系
其中多总体大小比较的参数方法第四章介绍 方差分析不限于简单大小比较

总体	总体关系	参数方法		非参数方法举例	
一个		t-检验, 正态检验	2.2.3		
二个	非对应	t-检验	2.2.4	Mann-Whitney U 检验	2.3.1
				随机化检验	2.3.3
	对应	成对数据 t-检验	2.2.5	Wilcoxon 加符秩检验	2.3.2
				成对数据随机化检验	2.3.4
多个	非对应	方差分析	4	Kruskal Wallis 检验	2.5.1
	对应	随机区组设计方差分析		Friedman 秩方差分析	2.5.3

从大小比较到方差分析

p.230

- 从总体大小比较到方差分析

大小比较中不同总体的构成 季节: 温度, 国家: 发展水平
从简单比较到影响因素分析 聚焦区分总体的因素, 如温度, 发展水平 等

- 方差分析

从大小比较到影响因素研究 多总体大小比较, 得到关于因素的结论和细节
从单一因素到多因素 两个或多个影响因素, 如温度与湿度
从独立影响到因素交互作用 多因素交互, 如温-湿度交互作用
从独立因素到从属关系 多级系统中有从属关系的因素, 如类别与亚类
从一般检验到逐对比较 用专门方法逐对比较, 不能用简单t-检验

4.1.1

4.1.1 方差分析及方差分析模型
4.1.2 方差分析方法

大小比较检验思路 复习

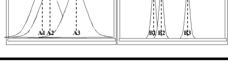
p.84

- 比较两总体大小的假设检验

假设: H_0 两总体大小没有显著差别; H_1 两总体大小有显著差别
计算: 总体大小无显著差别时, 得到实际观察结果和差别更大结果的概率 p
判断: 若可能性太小, 则判定两总体大小有显著差异

- 参数方法

原假设成立条件下, 影响相伴概率的因素包括:
样本均值差距越小, p 越大 A_1-A_3 比 A_1-A_2 更倾向显著
标准差越大, p 越大 B_1-B_2 比 A_1-A_2 更倾向显著
即 是否显著取决于均值差/标准差相对关系 两个分布重叠程度



方差分析的思路 *p.230*

- 两总体大小比较

基于两总体均值差别与离散程度的相对大小关系

- 多总体大小比较 – 方差分析

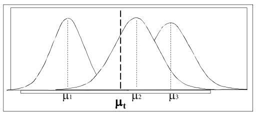
从两总体比较推广到多总体间关系　根据影响因素划分总体

影响因素是否显著同样取决于多总体间均值差和个体离散程度相对关系

总体均值差别越小, 总体越离散, 越不容易区分差异, 检验结果越倾向不显著

总体	数据	均值
1	X1i	μ_1
2	X2i	μ_2

总体	数据	均值	总均值
1		μ_1	
⋮		⋮	
i	Xij	μ_i	μ_t
⋮		⋮	
a		μ_a	



多总体中个体偏离的定义和拆解 *p.230*

- 均值

总均值 μ_t:　全部 a 个总体中所有个体的均值

单一总体均值 μ_i:　第 i 个总体中所有个体的均值, i = 1 .. a

- 多总体比较中的个体对均值的偏离 – 变异的构成

随机波动:　个体 x_{ij} 对所属总体均值 μ_i 的偏离

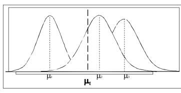
影响因素的作用:　μ_i 对 μ_t 的偏离

个体总偏离:　所有个体 x_{ij} 对总均值 μ_t 的偏离

总偏离 = 影响因素作用 + 随机波动

- 两总体大小比较与方差分析

两总体大小比较是 a=2 的特例



总体	数据	均值	总均值
1		μ_1	
⋮		⋮	
i	Xij	μ_i	μ_t
⋮		⋮	
a		μ_a	

平方和分解与加和性 *p.230*

- 平方和

上述偏离用平方和表达　$SS = \sum (x_i - \bar{x})^2$

- 平方和可分解为

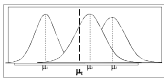
随机平方和　所有个体 x_{ij} 对所属总体均值 μ_i 偏离的平方和　$SS_W = \sum (x_{ij} - \bar{x}_i)^2$

总体间平方和　待比较总体均值 μ_i 对总均值 μ_t 偏离的平方和　$SS_a = \sum (\bar{x}_i - \bar{x})^2$

总平方和　所有个体 x_{ij} 对总均值 μ_t 偏离的平方和　$SS_t = \sum (x_{ij} - \bar{x})^2$

- 平方和加和性

总平方和可拆解为总体间平方和及随机平方和　$SS_t = SS_a + SS_W$



总体	数据	均值	总均值
1		μ_1	
⋮		⋮	
i	Xij	μ_i	μ_t
⋮		⋮	
a		μ_a	

方差分析 *22*

- 方差分析

根据平方和加和性原理, 将总平方和拆解为影响因素和随机波动的独立贡献

用F-检验判断影响因素是否显著　影响因素导致的差异, 相对于随机波动

与两总体大小类比　是否显著取决于均值差/离散程度相对关系

- 因子与水平

考察影响因素对随机变量的影响　如温度和湿度对作物生长的影响

因子: 研究的影响因素　温度与湿度

水平: 因子的不同取值　温度与湿度的不同取值, 如 10°C, …50 ° C, 10%, … 70%

湿度(%) \ 温度(°C)	10	20	30	40	50
10					
30			Xij		
50					
70					

4.1.2

4.1.1　**方差分析及方差分析模型**

4.1.2　**方差分析方法**

变量属性 复习 *p.15*

- **影响因素的变量属性**

与方差分析模型选择密切相关　类似回归分析 参见 **5.3-5.4** 回归

- **变量属性**

观测水平 { 连续变量, 离散变量, 顺序变量 } 定量变量; 类型变量　定性变量

取值性质 { 固定变量, 随机变量 }

获取方式 { 观测变量, 衍生变量 }

根据取值性质分类 复习
p.15-19

- **随机变量**

  个体随机出现, 大量取值表现出宏观规律

  数理统计方法的直接研究对象或影响因素
- **固定变量**

  人为控制

  不是数理统计方法的直接研究对象, 可以是方法中的影响因素

随机变量与固定变量实例 复习
p.15-19

- **连续变量 – 温度对微生物生长的影响**

  固定 实验室内或装置内人为控制的温度

  随机 非控制条件下实际观测到的温度
- **属性变量 – 不同城市降水酸度的差别**

  固定 覆盖所有待研究城市

  随机 从所有待研究城市中随机抽取部分城市
- **检验**

  固定 可重复

  随机 不可重复

方差分析方法
p.235

- 单因子方差分析

  单一影响因素, 若干水平　　多总体大小比较
- 双因子和多因子方差分析

  两个或多个影响因素　　独立影响及可能存在的交互作用
- 二级因子与多级因子方差分析

  两个或多个有从属关系的影响因素　因素间无交互作用

  多层结构数据, 下层从属上层　　如 大学-学院-系, 省-市-县, 土类-亚类

方差分析模型
p.233

- 方差分析中的变量

  研究变量, 研究对象　　必须是随机变量

  影响因素, 非研究对象　　随机变量 (随机效应) 或固定变量 (固定处理)

  例: 温度对作物生长的影响　作物生长速率为直接研究对象, 温度为影响因素
- 方差分析模型

  取决于影响因素的变量属性 影响因素的水平是否随机出现

  不同模型的差异　　检验假设, 补充分析

  固定模型, 模型 I

  随机模型, 模型 II

  混合模型, 模型 III

	模型	特点	检验对象
I	固定模型	所有因子均为固定处理	均值
II	随机模型	所有因子均为随机效应	方差
III	混合模型	既有固定处理也有随机效应	均值与方差

单因子方差分析
p.235

- 单因子方差分析

  两总体大小比较的直接推广　两个总体大小比较相当于两水平单因子方差分析

  影响因素 A, 包含 a 个水平　　两个总体大小比较相当于 a = 2

  每个水平下有若干重复数据　X_{ij}, i = 1..a, j = 1.. n_a, 提供离散信息
- 单因子方差分析的模型

  I 固定模型　影响因素为固定处理

  II 随机模型　影响因素为随机效应

水平	数据
1	
⋮	
i	Xij
⋮	
a	

水平	数据
1	2.1 4.3 5.0 8.0 3.3 5.4 3.3
2	7.2 0.4 5.4 8.2 1.2 9.8
3	6.5 4.0 2.0 4.4 2.1 7.5 2.1 3.3 3.3 4.9 5.0 4.4 6.6 2.7 6.3 5.5
4	5.2 1.4 8.1 1.2 2.2 1.1 0.3 3.9

双因子及多因子方差分析
p.235

- 双因子方差分析

  两个独立影响因素 A 和 B, 分别包含 a 和 b 个水平, 每个水平下有若干重复

  x_{ijk}, i = 1..a, j = 1.. b, k = 1..n_{ab}　重复提供随机波动信息

  可同时研究两个因子的独立影响及两个因子的交互作用
- 双因子方差分析模型

  I 固定模型, II 随机模型, III 混合模型 取决于 A 和 B 的属性
- 多因子方差分析

  双因子方差分析的推广　三因子, 四因子等

		B 1	…	j	…	b
A	1					
	⋮					
	i			Xijk		
	⋮					
	a					

		B 1	2	3
A	1	5.2,1.4,8.1,1.2	2.2,4.8,1.1,4.3	5.1,4.2,1.3,0.2
	2	1.2,2.1,2.4,5.2	9.5,5.4,4.5,0.1	3.1,7.1,2.3,1.6
	3	2.4,1.5,8.2,2.4	3.2,6.6,9.1,4.5	5.2,4.2,1.3,4.2
	4	0.6,6.6,8.8,8.1	2.1,1.4,1.2,6.0	9.1,4.2,1.3,0.2

二级因子及多级因子方差分析

p.236

- 二级因子方差分析

两个有<u>从属</u>关系的影响因素 $\mathbf{A}, \mathbf{B} \in \mathrm{A}$

一级因子的每个水平下有若干二级因子水平 x_{ijk}, $i = 1..a, j = 1..bi, k = 1..n_{ab}$

不同一级水平下的二级水平相互独立 如 x_{121} 和 x_{a21}

研究两级因子的独立影响 不存在交互作用

- 二级因子方差分析模型

二级因子为<u>随机</u>效应, 模型取决于一级因子属性

II 随机模型, III 混合模型

- 多级因子方差分析

二级因子方差分析的推广

一级因子	二级因子	数据
1	1	$x_{111}, \ldots x_{11n_{11}}$
	2	$x_{121}, \ldots x_{12n_{12}}$
	…	
	b1	$x_{1b_11}, \ldots x_{1b_1n_{1b1}}$
…	…	
a	1	$x_{a11}, \ldots x_{a1n_{a1}}$
	2	$x_{a21}, \ldots x_{a2n_{a2}}$
	…	
	ba	$x_{ab_11}, \ldots x_{ab_an_{aba}}$

方差分析的一般步骤

p.238

- 选择方法和模型

目的: 研究特定因素对某随机变量的影响 随机变量在不同水平间的差异

选择: 方法和模型 因子个数, 因子关系, 因子类型

- 假设检验

确定显著性水平 α

假设: 不同水平间有显著差异 引申为: <u>因素</u>是否有显著影响

具体假设与方法与模型有关

计算相伴概率 p 或检验统计量 F 与 α 或检验临界值 $F_{\alpha[v_1,v_2]}$ 比较

方差分析的检验假设

p.202

- 假设

对每一影响因素及交互作用有独立的假设 原假设与对立假设

- 固定因子

不显著表现为不同水平总体均值无显著差异 影响因素无显著作用

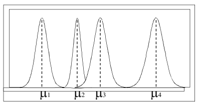
H_0 μ_i 无显著差异, H_1 μ_i 有显著差异（不<u>完全</u>相同）类似两总体大小比较, a=2

- 随机因子

不显著表现为不同水平<u>均值的方差</u>为零, 无显著差异 与单总体方差比较

H_0 $\sigma^2 = 0$, H_1 $\sigma^2 \neq 0$

因子水平随机出现, 比较均值没有意义



方差分析的检验假设

p.202

- 假设

对每一影响因素及交互作用有独立的假设 原假设与对立假设

- 固定因子

不显著表现为不同水平总体均值无显著差异 影响因素无显著作用

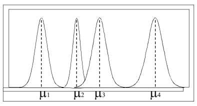
H_0 μ_i 无显著差异, H_1 μ_i 有显著差异（不<u>完全</u>相同）类似两总体大小比较, a=2

- 随机因子

不显著表现为不同水平<u>均值的方差</u>为零, 无显著差异 与单总体方差比较

H_0 $\sigma^2 = 0$, H_1 $\sigma^2 \neq 0$

因子水平随机出现, 比较均值没有意义



方差分析的检验假设

p.202

- 假设

对每一影响因素及交互作用有独立的假设 原假设与对立假设

- 固定因子

不显著表现为不同水平总体均值无显著差异 影响因素无显著作用

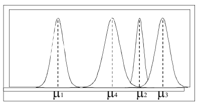
H_0 μ_i 无显著差异, H_1 μ_i 有显著差异（不<u>完全</u>相同）类似两总体大小比较, a=2

- 随机因子

不显著表现为不同水平<u>均值的方差</u>为零, 无显著差异 与单总体方差比较

H_0 $\sigma^2 = 0$, H_1 $\sigma^2 \neq 0$

因子水平随机出现, 比较均值没有意义



方差分析的检验假设

p.202

- 假设

对每一影响因素及交互作用有独立的假设 原假设与对立假设

- 固定因子

不显著表现为不同水平总体均值无显著差异 影响因素无显著作用

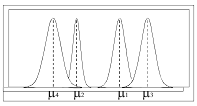
H_0 μ_i 无显著差异, H_1 μ_i 有显著差异（不<u>完全</u>相同）类似两总体大小比较, a=2

- 随机因子

不显著表现为不同水平<u>均值的方差</u>为零, 无显著差异 与单总体方差比较

H_0 $\sigma^2 = 0$, H_1 $\sigma^2 \neq 0$

因子水平随机出现, 比较均值没有意义



方差分析的检验假设

p.202

- 假设

  对每一影响因素及交互作用有独立的假设　　原假设与对立假设

- 固定因子

  不显著表现为不同水平总体均值无显著差异　　影响因素无显著作用

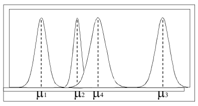
  H_0 μ_i 无显著差异, H_1 μ_i 有显著差异（不完全相同）类似两总体大小比较, a=2

- 随机因子

  不显著表现为不同水平均值的方差为零, 无显著差异　与单总体方差比较

  H_0 $\sigma^2 = 0$, H_1 $\sigma^2 \neq 0$

  因子水平随机出现, 比较均值没有意义



方差分析的补充分析

p.238

- 方差分析的补充分析

  如果方差分析检验结果显著, 可进一步做补充分析

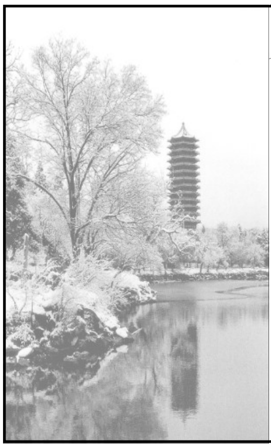
- 补充分析目的

  取决于方差分析模型

  固定处理: 逐对比较　　考察不同水平之间的显著性

  不能用 t-检验逐对比较　　水平多次使用会改变 p 值

  随机效应: 方差分量计算　　无实质应用价值



应用数理统计方法
2024.7

第四章 方差分析

4.1 方差分析方法
4.2 单因子方差分析及其补充分析
4.3 双因子和多因子方差分析
4.4 二级与多级因子方差分析
应用举例

4.2.1

4.2.1 单因子方差分析模型与显著性检验

4.2.2 模型 I 和 模型 II 单因子方差分析的补充分析

单因子方差分析 复习

p.235

- 单因子方差分析

  两总体大小比较的直接推广　两个总体大小比较相当于两水平单因子方差分析

  影响因素 A, 包含 a 个水平　两个总体大小比较相当于 a = 2

  每个水平下有若干重复数据　X_{ij}, i = 1..a, j = 1.. $n_{a,}$ 提供离散信息

- 单因子方差分析的模型

  I 固定模型　影响因素为固定处理

  II 随机模型　影响因素为随机效应

水平	数据
1 ⋮ i ⋮ a	Xij

水平	数据
1	2.1 4.3 5.0 8.0 3.3 5.4 3.3
2	7.2 0.4 5.4 8.2 1.2 9.8
3	6.5 4.0 2.0 4.4 2.1 7.5 2.1 3.3 3.3 4.9 5.0 4.4 6.6 2.7 6.3 5.5
4	5.2 1.4 8.1 1.2 2.2 1.1 0.3 3.9

单因子方差分析数据构成

p.243

- 模型 I 单因子方差分析

  $x_{ij} = \mu + \mathbf{a_i} + \varepsilon_{ij}$　　i = 1..n, j = 1..m_i

  μ 为总体均值　　图中 μ_t

  $\mathbf{a_i}$ 为影响因素 A 在水平 i 对μ_t的偏离　　图中 μ_i 对 μ_t 的偏离, 固定影响

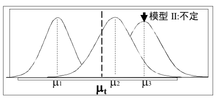
  ε_{ij} 为个体随机波动　　图中个体所属水平对 μ_i 的偏离

- 模型 II 单因子方差分析

  $x_{ij} = \mu + \mathbf{A_i} + \varepsilon_{ij}$　　i = 1..n, j = 1..m_i

  $\mathbf{A_i}$ 为影响因素 A 在水平 i对μ_t的偏离　　图中 μ_i 对 μ_t 的偏离, 随机影响

  μ 与 ε_{ij} 同上



课堂练习

- 模型 I 单因子方差分析

  $x_{ij} = \mu + \mathbf{a_i} + \varepsilon_{ij}$　　$i = 1..n, j = 1..m_i$

  μ 为总体均值　　图中 μ_t

  $\mathbf{a_i}$ 为影响因素 A 在水平 i 对μ_t的偏离　　图中 μ_i 对 μ_t 的偏离, <u>固定影响</u>

  ε_{ij} 为个体随机波动　　图中个体所属水平对 μ_i 的偏离

- 模型 II 单因子方差分析

  $x_{ij} = \mu + \mathbf{A_i} + \varepsilon_{ij}$　　$i = 1..n, j = 1..m_i$

  $\mathbf{A_i}$ 为影响因素 A 在水平 i对μ_t的偏离　　图中 μ_i 对 μ_t 的偏离, <u>随机影响</u>

  μ 与 ε_{ij} 同上

- 两总体大小比较 t-检验的数据构成？

课堂练习

- 两总体大小比较 t-检验的数据构成？

  相当于: 固定处理, 两水平的单因子方差分析

  $x_{ij} = \mu + a_i + \varepsilon_{ij}$　$i = 1..\mathbf{2};\ j = 1..m_i$

  进一步理解大学比较与方差分析的关系

单因子方差分析的假设

p.244

- 固定模型方差分析

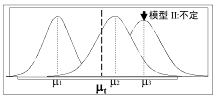
  H_0 μ_i <u>都不</u>显著, H_1 μ_i <u>不都</u>显著

  任意两个水平显著即为显著

- 随机模型方差分析

  H_0 $\sigma^2 a = 0$, H_1 $\sigma^2 a \neq 0$　$\sigma^2_a = \sum(\mu_i - \mu_t)^2/(a-1)$, 与总体方差计算类似

  不同水平均值的方差与零无显著差异



单因子方差分析计算与判断

p.245

- 计算

  <u>组间</u>平方和与均方 SSa, Msa　反映影响因素的作用

  <u>组内</u>平方和与均方 SSw, MSw 代表个体随机波动, 或变异

  组间: 水平间

  组内: 水平内

- 检验

  F =MSa/MSw

  类比两总体大小比较的 t-检验　t_s = 均值差/方差与样本量的函数

方差分析表

方差来源	平方和	自由度	均方	检验值	临界值
组间 (水平间, a)	SS_a	v_a	MS_a	F	$F_{\alpha[va,vw]}$
组内 (水平内, w)	SS_w	v_w	MS_w		

4.2.2

4.2.1 单因子方差分析模型与显著性检验

4.2.2 模型 I 和 模型 II 单因子方差分析的补充分析

方差分析的补充分析

p.246-248

- 补充分析

  检验结果显著, 可进一步做补充分析

- 模型 I 方差分析的补充分析 – 均值比较

  后决比较=无计划比较, <u>多重比较</u>, 水平间逐对比较

  先决比较=有计划比较, 每个水平出现不超过一次　t-检验或方差分析–有限用途

- 模型 II 方差分析的补充分析 – 方差分量计算

  影响因素水平取值随机出现, 不能做均值比较

  方差分量无实质用途

- 模型 III 方差分析的补充分析

  均值比较与方差分量计算　取决于待比较因素的属性

多重比较方法 *p.250*

- 多重比较方法<u>举例</u>

适用条件不同 水平数, 样本量 等

<u>保守性</u>有差别 保守性顺序越小, 方法拒绝原假设的可能越小

方法	保守性顺序	适用范围
Scheffe法	1	水平数较小的数据
T法	2	等样本量数据
T' 法	2	样本量相差不大的不等样本量数据
Welsch法	3	等样本量数据
GT2法	3	样本量相差较大的不等样本量数据
一般SNK法	4	等样本量数据
近似SNK法	4	不等样本量数据

多重比较的一般步骤 *p.250*

- 一般步骤

水平按均值从大到小依次排序

数据: 计算所有水平间的<u>均值差</u> <u>逐对</u>计算

假设: H_0 两均值不显著, H_1 两均值显著 逐对检验

计算: p 值或检验值 对比t-检验, 修正了<u>拒绝概率</u>

判断: $p < \alpha$ 或 检验值 < 均值差, 拒绝原假设 临界值可以是一个或多个

水平	1	2	…	a-1
2	x_1-x_2			
3	x_1-x_3	x_2-x_3		
…	…	…	…	
a	x_1-x_a	x_2-x_a	…	X_{a-1}-x_a

多重比较的检验值 *p.251-253*

- 检验值举例

T 法: 单一检验值 MSD

一般 SNK 法: a-1 个检验值 SNK_i a 为水平数

近似 SNK 法: a(a-1)/2 个检验值 SNK_{ij} 每个均值对有独立的检验值

T法

	1	2	3
2	x_{12}		
3	x_{13}	x_{23}	
4	x_{14}	x_{24}	x_{34}

↓

MSD

SNK法

	1	2	3
2	x_{12}		
3	x_{13}	x_{23}	
4	x_{14}	x_{24}	x_{34}

↓

1	2	3
SNK_4		
SNK_3	SNK_4	
SNK_2	SNK_3	SNK_4

近似SNK法

	1	2	3
2	x_{12}		
3	x_{13}	x_{23}	
4	x_{14}	x_{24}	x_{34}

↓

	1	2	3
2	SNK_{12}		
3	SNK_{13}	SNK_{23}	
4	SNK_{14}	SNK_{24}	SNK_{34}

多重比较结果图示 *p.251*

- 不显著组

根据多重比较结果, 将不同水平划分为若干不显著组

不显著组 组内水平没有显著差异, 水平按大小次序排列

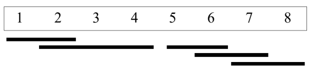
- 不显著组的表达

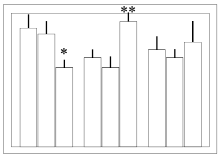
符号 用 *, **, *** 表示在 0.10, 0.05, 0.01 水平显著

线段 用直线表示不显著组

- 不显著组的重叠

<u>讨论</u>





不显著组的解释 *p.251*

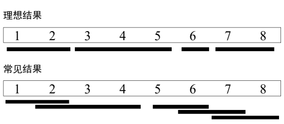
- 不显著组划分结果

<u>理想</u>结果, 不显著组不重叠 可以确切划分水平类别

<u>非理想</u>结果, 不显著组重叠 不能确切划分水平类别

- 讨论

为什么会有貌似矛盾的结果？ <u>逻辑</u>矛盾与<u>统计</u>的合理性



不显著组划分的特例

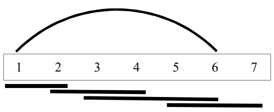
- 特例

水平均值递增排列, 均值差从上至下递增 均值从 x_{12} 到 x_{17} 递增

水平 **1** 与 3, 4, 5, 7 均显著☺, 但与 **6** 无显著差异☹ x_{16}

- 讨论若

<u>为什么</u>会有貌似矛盾的结果？ <u>逻辑</u>矛盾与<u>统计</u>的合理性



	1	2	3	4	5	6
1						
2	x_{12}					
3	x_{13}☺	x_{23}				
4	x_{14}☺	x_{24}	x_{34}			
5	x_{15}☺	x_{25}☺	x_{35}	x_{45}		
6	x_{16}	x_{26}☺	x_{36}	x_{46}	x_{56}	
7	x_{17}☺	x_{27}☺	x_{37}☺	x_{47}☺	x_{57}	x_{67}

不显著组划分的特殊情况

- 特例

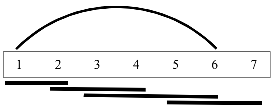
  水平均值递增排列, 均值差从上至下递增　　均值从 x_{12} 到 x_{17} 递增

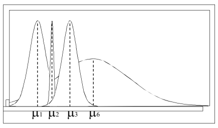
  水平 **1** 与 3, 4, 5, 7 均显著☺, 但与 **6** 无显著差异☹　x_{16}

- 讨论

  为什么会有貌似矛盾的结果？ 逻辑矛盾与统计的合理性

  数据解释　　推论1-6也显著





模型 II 方差分析的补充分析

p.257

- 模型 II 方差分析的补充分析

  不能做多重比较　水平随机出现, 两两比较无意义

  如果检验结果显著, 可计算方差分量

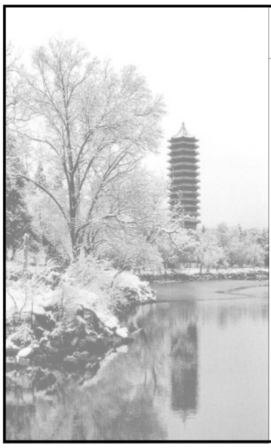
- 方差分量的分解

  总方差分解为组间和误差方差　表达为相对贡献 %

  影响因素和随机波动相对贡献的定量表达

  可进一步计算方差分量的置信区间

  相对贡献不影响是否显著的基本结论　　无实质意义



应用数理统计方法
2024.7

第四章 方差分析

4.1 方差分析方法
4.2 单因子方差分析及其补充分析
4.3 双因子和多因子方差分析
4.4 二级与多级因子方差分析
应用举例

双因子方差分析

p.260

- 双因子方差分析

  同时研究两个影响因素 A 和 B 以及两者间交互作用

  两因素分别包含 a 个和 b 个水平, 每个水平下有若干重复　　提供随机波动

  x_{ijk}, i = 1..a, j = 1.. b, k = 1..n_{ab},

- 双因子方差分析模型

  I 固定模型, II 随机模型, III 混合模型　　取决于 A 和 B 的属性

		B				
		1	…	j	…	b
A	1					
	⋮					
	i			**Xijk**		
	⋮					
	a					

		B 1	2	3
A	I	5.2,1.4,8.1,1.2	2.2,4.8,1.1,4.3	5.1,4.2,1.3,0.2
	II	1.2,2.1,2.4,5.2	9.5,5.4,4.5,0.1	3.1,7.1,2.3,1.6
	III	2.4,1.5,8.2,2.4	3.2,6.6,9.1,4.5	5.2,4.2,1.3,4.2
	IV	0.6,6.6,8.8,8.1	2.1,1.4,1.2,6.0	9.1,4.2,1.3,0.2

课堂练习

- 单因子方差分析的数据构成

  任一个体 x_{ij} 的数据构成

  模型 I　$x_{ij} = \mu + \mathbf{a_i} + \varepsilon_{ij}$

  模型 II　$x_{ij} = \mu + \mathbf{A_i} + \varepsilon_{ij}$

- 问题

  双因子方差分析呢

  模型 I ?

  模型 II ?

  模型 III ?

课堂练习

- 单因子方差分析的数据构成

  任一个体 x_{ij} 的数据构成

  模型 I　$x_{ij} = \mu + \mathbf{a_i} + \varepsilon_{ij}$

  模型 II　$x_{ij} = \mu + \mathbf{A_i} + \varepsilon_{ij}$

- 双因子方差分析

  模型 I　$x_{ijm} = \mu + a_i + b_j + ab_{ij} + \varepsilon_{ijm}$

  模型 II　$x_{ijm} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{ijm}$

  模型 III　$x_{ijm} = \mu + a_i + B_j + aB_{ij} + \varepsilon_{ijm}$

双因子方差分析的检验假设 *p.262*

- 固定模型 I

  H0 $\mu_{i(A)}$无显著差别 $\mu_{i(B)}$无显著差别 $\mu_{ij(AB)} = 0$ H1 略

- 随机模型 II

  H0 $\sigma^2_{a(A)} = 0$ $\sigma^2_{a(B)} = 0$ $\sigma^2_{a(AB)} = 0$ H1 略

- 混合模型 III

  H0 $\mu_{i(A)}$无显著差别 $\sigma^2_{a(B)} = 0$ $\sigma^2_{a(AB)} = 0$ H1 略

		B: 1 … j … b
A	1 ⋮ i ⋮ a	Xijk

双因子方差分析计算与检验 *p.264*

- 双因子方差分析计算

  计算两个因子的随机和总平方和 <u>平方和加和性</u>适用于每个<u>独立</u>因子

  可用差减方式计算交互作用平方和 $SS_{交互作用} = SS_{总} - SS_A - SS_B - SS_{误差}$

- 双因子方差分析检验

  先检验交互作用 $F_{ab} = MS_{ab}/MS_w$

  若交互作用不显著, 分别检验独立因子 $F_a = MS_a/MS_{w,}$ $F_b = MS_b/MS_w$

  若交互作用显著, 合并均方, 检验独立因子 $F_a = MS_a/MS_{p,}$ $F_b = MS_b/MS_p$

  $MS_p = (SS_{ab}+SS_w)/(v_{ab}+v_w)$

  拒绝条件

  $F > F_{\alpha[v,v]}$ 或 $p < \alpha$

方差来源	平方和	自由度	均方
因子 A 组间 (A)	SS_a	v_a	MS_a
因子 B 组间 (B)	SS_b	v_b	MS_b
AB 交互作用 (AB)	SS_{ab}	v_{ab}	MS_{ab}
组内 (W)	SS_w	v_w	MS_w

双因子方差分析的补充分析 *p.238*

- 补充分析

  <u>分别</u>对两因子做补充分析 将另一因子并入重复, 相当于两个单因子方差分析

  模型 I 两因子各自做均值比较

  模型 II 两因子各自计算方差分量

  模型 III 分别比较均值和计算方差分量

		B: 1 … j … b
A	1 ⋮ i ⋮ a	Xijk

多级因子方差分析 *p.*

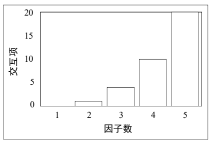
- 多因子方法

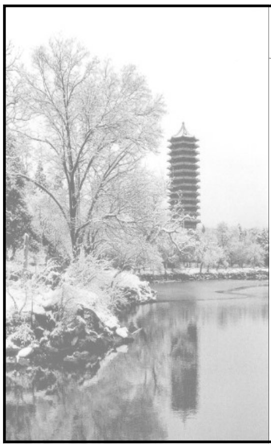
  双因子方差分析的推广

- 限定

  交互作用项不宜太多 随因子数增加, 交互作用项迅速增多, 干扰独立因子检验

  少见三因子以上的方差分析





应用数理统计方法

2024.7

第四章 方差分析

4.1 方差分析方法

4.2 单因子方差分析及其补充分析

4.3 双因子和多因子方差分析

4.4 二级与多级因子方差分析

应用举例

二级因子方差分析 *p.238*

- 双因子和二级因子

  双 因 子: 两个相互独立的因子 A 和 B

  二级因子: 一个独立因子及一个<u>从属</u>因子 A 和 B∈A

- 二级因子方差分析

  与双因子方差分析数据结构不同, 不存在<u>交互作用</u>

  二级因子为随机效应 水平<u>不对应</u>, 不存在固定水平取值

  模型取决于一级因子 II 或 III

  模型 II $x_{ijm} = \mu + A_i + B_j + \varepsilon_{ijm}$

  模型 III $x_{ijm} = \mu + a_i + B_j + \varepsilon_{ijm}$

		B: 1 … j … b
A	1 ⋮ i ⋮ a	x_{ijk}

A	B	数据
1	1	X_{111}, … $X_{11n_{11}}$
	2	X_{121}, … $X_{12n_{12}}$
	…	
	b1	X_{1b_11}, … $X_{1b_1n_{1b1}}$
…	…	
a	1	X_{a11}, … $X_{a1n_{a1}}$
	2	X_{a21}, … $X_{a2n_{a2}}$
	…	
	ba	X_{ab_a1}, … $X_{ab_an_{aba}}$

二级因子方差分析的假设与检验

p.282

- 检验假设

无交互作用, 仅检验一级和二级因子　方差分析表

一级因子 A　$H_{0(A)}$ μi 无差异(固定处理)), H_1 略

$H_{0(A)}$ $\sigma^2_a = 0$ (随机效应),　H_1 略

二级因子 $B_{\in A}$　$H_{0(B\in A)}$ $\sigma^2_{b\in a} = 0$,　H_1 略

- 补充分析

类似双因子方差分析, 分别检验　另一因子并入误差项

方差来源	平方和	自由度	均方	检验值	临界值
一级因子（A）	SS_a	v_a	MS_a	F_a	$F_{\alpha[v_a,v_{b\in a}]}$
二级因子（B∈A）	$SS_{b\in a}$	$v_{b\in a}$	$MS_{b\in a}$	$F_{b\in a}$	$F_{\alpha[v_{b\in a},v_w]}$
误差（w）	SS_w	v_w	MS_w		

多级因子方差分析及其应用

p.279,287

- 多级因子方差分析

二级因子方差分析的推广 无交互作用, 级数限制比多因子方差分析小

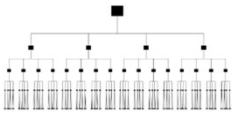
次级因子均为随机效应, 模型取决于一级因子类型

- 应用

多层次结构, 判断各层次是否显著

如: 多步程序中各步的误差　采样 – 保存 – 前处理 – 分析误差

如: 多级系统中各级的差异　土类 – 亚类;　大学-学院-系



课堂练习

- 举例

举一个可以用方差分析研究的例子, 至少两个因子

- 要求

就近随意分组, 每组讨论一个例子

可以针对随机变量的任何特征

要求　简述问题

给出 原假设和对立假设

假定的 p 值, 结论

字大些

应用数理统计方法
2024.7

第四章 方差分析

应用实例 茶叶对阿尔茨海默症小鼠行为的影响

姜炫旭等, 浙大学报 2022

- 问题与方法

问题: 茶叶及多奈哌对阿尔茨海默症小鼠空间记忆力影响　茶叶主要有效成分

方法: Y 迷宫实验测定空间记忆力　六水平: 正常对照, 病鼠, 四种处理的病鼠

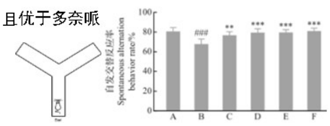
统计: <u>单因子方差分析</u>及<u>多重比较</u>

- 结果和讨论

假设: H_0 无显著差别, H_1 有显著差别

结果: 单因子方差分析结果显著, 多重比较部分显著

结论: 均显著, 茶叶间无差别, 且优于多奈哌



应用实例 不同品牌口罩对颗粒物去除率的差异

Shen et al. Environ Int. 2021

- 问题与方法

问题: 比较八种主流品牌口罩对颗粒物的去除效果

方法: 模拟实验测定颗粒物去除率　百分数据做角变换

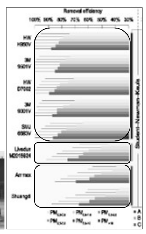
统计: 针对品牌的<u>单因子方差分析</u>及<u>补充分析</u>　不同粒径

- 结果和讨论

假设: H_0 无显著差别, H_1 有显著差别

结果: 显著, 六粒径结果相似, 品牌分三个不显著组

<u>讨论</u>: 可考虑<u>双因子</u>方差分析, 探讨<u>交互作用</u> 品牌, 粒径



应用实例 品种和地域对小麦铁锌含量的影响

龚瀚颖等, 核农学报, 2019

- 问题与方法

  问题: 探讨三个地区的三个品种小麦 Fe 含量的差别

  方法: 连续 4 年重复采样, 测定小麦籽粒 Fe 含量

  统计: 分别对区域和品种差异做单因子方差分析和多重比较

- 结果和讨论

  假设: H_0 无显著差别, H_1 有显著差别 两组检验分别针对区域和品种

  结果: 品种 杨凌和辉县的衡5228的Fe含量显著低于其它品种

  区域 周麦16无区域差异, 衡5229在杨凌和赵县间, 邯6172在辉县和赵县间显著

  讨论: 用双因子方差分析, 并探讨交互作用

应用实例 影响蚊子翅膀长度的因素

Sokal and Rohlf, Biometry 1969

- 问题与方法

  问题: 研究不同培养箱, 不同采样个体对蚊子翅膀长度测定结果的影响

  方法: 从 **3** 组培养箱中各随机采集 **4** 只蚊子测量左翅长度 每只测 2 个重复

  统计: 二级分组方差分析 随机模型

- 结果和讨论

  假设: H_0 培养箱间差异显著, H_0 个体间差异显著

  结果: 培养箱间无显著差异 $p > 0.05$

  个体间差异显著 $p < 0.001$

  结论: 增加个体采样量可提高测量精度

培养箱	采样	测量1	测量2
1	1	58.5	59.5
	2	77.8	80.9
	3	84.0	83.6
	4	70.1	68.3
2	1	69.8	69.8
	2	56.0	54.5
	3	50.7	49.3
	4	63.8	65.8
3	1	56.6	57.5
	2	77.8	79.2
	3	69.9	69.2
	4	62.1	64.5

应用实例 影响纳米纤维滤膜直径的工艺参数

Cao et al., Sci. Total Environ. 2019

- 问题与方法

  问题: 六种工艺条件对纳米纤维直径的影响 浓度, 电压, 转速, 距离, 流量, 温度

  方法: 每种工艺设置三个水平 正交设计

  统计: 六因子方差分析, 未检验交互作用

- 结果和讨论

  假设: H_0 特定因子没有显著影响, H_1 特定因子有显著影响 六组假设

  结果: 仅发现溶液浓度有显著影响 图: 右图加标准差

  讨论: 工艺优化的正交试验 ?

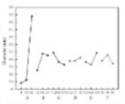
  方差分析 ?

  可能的交互作用

Analysis of variance for the orthogonal experiment.

Source	SS	DF	MS	F[a]	P[b]
PAN solution concentration	46,084.825	2	23,042.412	26.720	<0.001***
Voltage	2755.444	2	1377.722	1.598	0.237
Rotating speed	1283.519	2	641.760	0.744	0.495
Distance	123.187	2	61.593	0.071	0.931
Flow rate	1233.696	2	616.848	0.715	0.506
Temperature	658.370	2	329.185	0.382	0.690
Error	12,073.192	14	862.371		
Corrected total	64,212.233	25			

[a] $F_{0.1}(2, 14) = 2.73$, $F_{0.05}(2, 14) = 3.74$, $F_{0.01}(2, 14) = 6.51$.
[b] Statistical significance level was set as follows: * if P < 0.05, ** if P < 0.01, *** if P < 0.001.



应用实例 影响深圳土壤微量元素含量的主导因素

- 问题与方法

  问题: 确定影响土壤微量元素含量的主导因素 土壤类型还是母岩类型 ?

  方法: 采集 83 个表土样品, 测定了 12 种微量元素含量

  因子 1: 土壤类型 红壤, 赤红壤, 黄壤, 水稻土, 海积土

  因子 2: 母岩类型 花岗-火山岩, 砂页岩, 变质岩, 冲积母质, 海积母质

- 结果

  数据: 不完整的数据集 双因子方差分析 ? 数据补缺的可能 ?

  重构: 16 种母质–土类组合 如花岗岩残积, 坡积和冲积母质上发育的红壤

  转换: 基于新组合的单因子方差分析

  结果: 11 种元素结果显著

  Cd 例外, 测定方法误差大

母质\土类	赤红壤	红壤	黄壤	水稻土	滨海砂土	滨海粘土
花岗岩残、坡积	21	3				
砂页岩残、坡积	13	4	1			
变质岩残、坡积	8					
火山岩残、坡积	4	2	2			
花岗岩冲积				4		
砂页岩冲积				10		
冲积 - 海积				4		1
海积砂				2	2	
海积粘土						4

应用实例 影响深圳土壤微量元素含量的主导因素

- 讨论 非统计方法

  根据多重比较结果, 标出显著的元素 ** 为多于三个, 16 个单元矩阵

  确定相似组合, 均不显著或个别例外 1-2-10, 3-4-5, 7-8-9, 12-13, 14-15

  16 类并为 8 类 合并的多为土类, 母质基本分开

- 结论

  母质是主导因素, 与常量元素不同 成土过程继承了母质微量元素特性

1															
2															
3	**	Cr													
4	Cr,V	Cr													
5	Pb	Pb	Cu												
6	**	Cu,Cr	Cu	Cu	Cu,Pb										
7					Pb	Cu									
8			Zn,Cr	Cr	Pb,Zn	Cu,Cr									
9			Pb,Cr	Pb,Cr	Pb	Cu,Cr									
10			Cr,V	Cr,V	Pb	**	V								
11	**	Cr,F	Hg,F		**	Cu,Hg	Hg,F	Cr,F	Cr,F	Cr,V					
12	**	Hg	Hg	Hg	**	Cu,Hg	Hg		Hg	V	F				
13	**	Cu,Cr			Pb,Co,Cu			Cr	Co,Cr	Cr,V	Hg,F				
14	Zn	Zn	V	**		**	V	Zn	Pb,Zn		Cr,V	V	**		
15			**	**	Pb,Zn	**	Cu,V				**	**	**		
16	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**
单元	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

1	花岗岩残、坡积	赤红壤
2	花岗岩残、坡积	红壤
3	砂页岩残、坡积	赤红壤
4	砂页岩残、坡积	红壤
5	砂页岩残、坡积	黄壤
6	变质岩残、坡积	赤红壤
7	火山岩残、坡积	赤红壤
8	火山岩残、坡积	红壤
9	火山岩残、坡积	黄壤
10	花岗岩冲积	水稻土
11	砂页岩冲积	水稻土
12	冲、海积	水稻土
13	冲、海积	滨海粘土
14	海积砂	水稻土
15	海积砂	滨海沙土
16	海积粘土	滨海粘土

单元	母质-土类组合
U1	花岗岩及花岗岩冲积母质上的赤红壤、红壤和水稻土
U2	砂页岩母质上发育的赤红壤、红壤和黄壤
U3	变质岩母质上发育的赤红壤
U4	火山岩母质上发育的赤红壤、红壤和黄壤
U5	砂页岩地区冲积母质上发育的水稻土
U6	冲积一海积母质上发育的水稻土和滨海沼泽土
U7	海积砂母质上发育的水稻土和滨海砂土
U8	海积粘土母质上发育的滨海粘土

谢谢