

心理统计学第六讲

讲师 司马紫衣



高途学院APP下载



高途学院公众号





第十章 方差分析

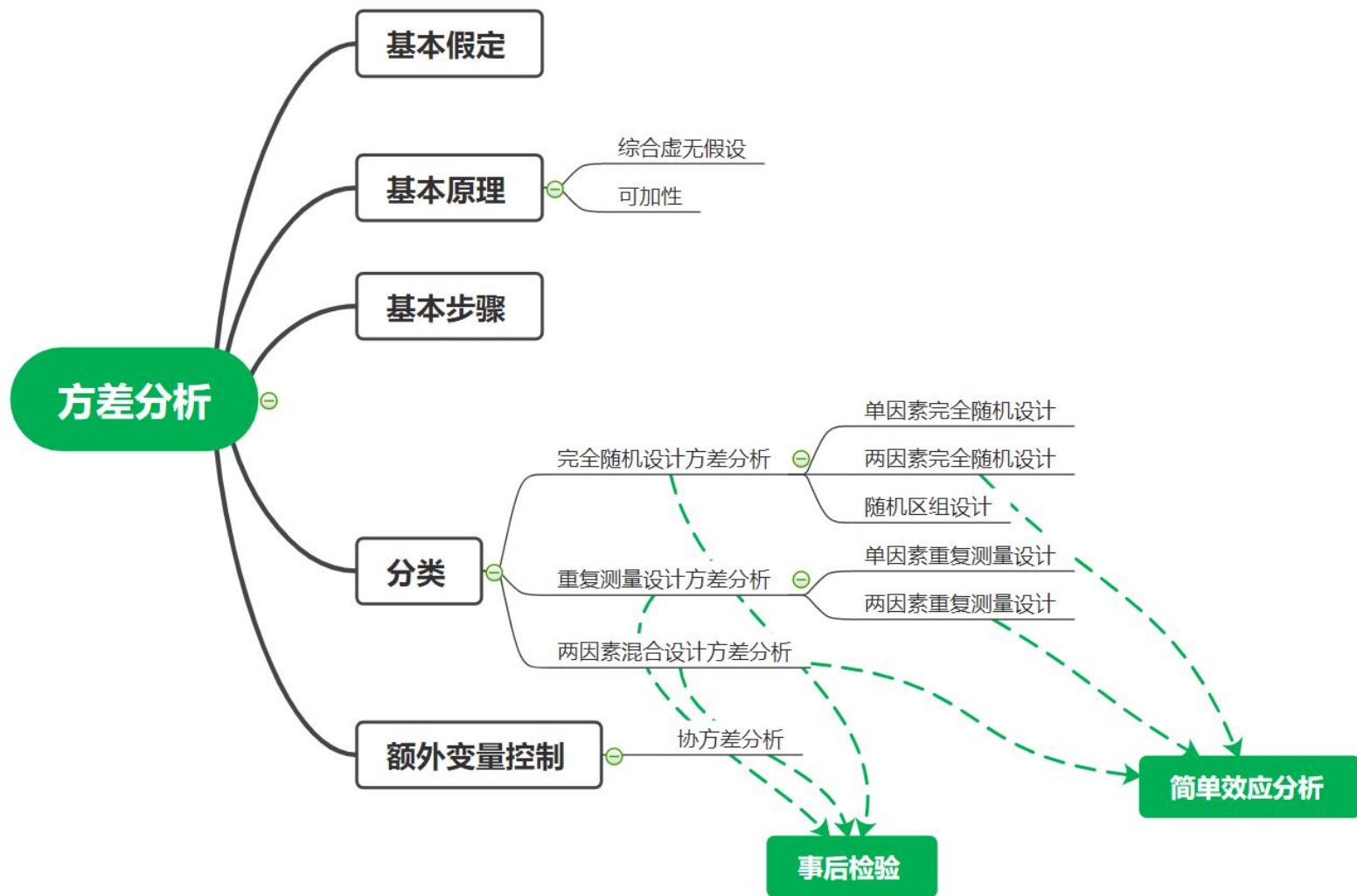


概览

- ◆ 方差分析的基本原理（选择）
- ◆ 完全随机设计的方差分析（单选、简答、综合）
- ◆ 重复测量设计的方差分析（单选、简答、综合）
- ◆ 两因素混合设计的方差分析（单选、简答、综合）
- ◆ 协方差分析（单选）
- ◆ 事后检验和简单效应分析（选择、简答、综合）



本章结构





一、方差分析的基本原理

方差分析，又称变异分析，是探讨一个因变量和一个或多个自变量之间的关系的一种假设检验方法。

方差分析的功能在于分析实验数据中不同来源的变异对总变异的贡献大小，从而确定自变量是否对因变量有重要影响。

从统计上讲，方差分析主要处理两个以上的平均数之间的差异检验问题。



一、方差分析的基本原理

(一) 假设

假设有三个样本，每个样本的平均数是 \bar{X}_1 、 \bar{X}_2 、 \bar{X}_3 。

$H_0: \mu_1 = \mu_2 = \mu_3$ ——综合虚无假设

$H_1: \mu_1$ 、 μ_2 、 μ_3 至少存在一组是有差异的

如果 H_0 被拒绝了，我们需要知道到底是哪两个组的平均值有显著差异，需要做“部分虚无假设”：

$\mu_1 = \mu_2$ $\mu_1 = \mu_3$ $\mu_2 = \mu_3$



一、方差分析的基本原理

(二) 方差的可分解性 (可加性)

方差分析把实验数据的总变异分解为若干个不同来源的分量，用平方和 (SS) 来表示。

- 探讨噪音对解决数学问题的影响作用。噪音是自变量，划分为三个强度水平：强、中、无。因变量是解决数学问题时产生的错误频数。随机抽取12名被试，再随机把他们分到强、中、无三个实验组。

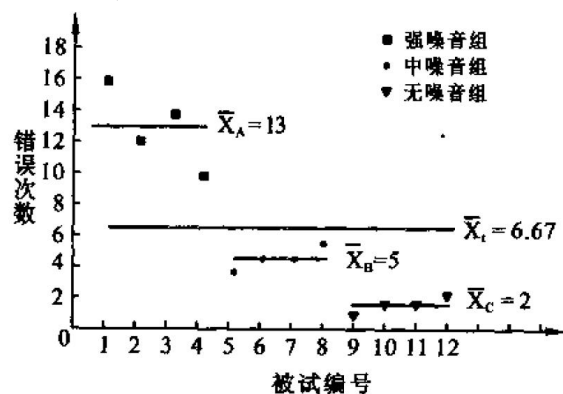


图 9-1 数据变异示意图

表 9-1 不同强度噪音下解数学题犯错误频数

	噪音(分贝)			k = 3
	强(100)(A)	中(50)(B)	无(C)	
n = 4	16	4	1	$\bar{X}_t = 6.67$
	14	5	2	
	12	5	2	
	10	6	3	
\bar{X}_j	13	5	2	



一、方差分析的基本原理

表 9-1 不同强度噪音下解数学题犯错误频数

	噪音(分贝)			k=3
	强(100)(A)	中(50)(B)	无(C)	
n=4	16	4	1	
	14	5	2	
	12	5	2	
	10	6	3	
\bar{X}_j	13	5	2	$\bar{X}_t = 6.67$

方差合成公式:

$$S_T^2 = \frac{\sum N_i S_i^2 + \sum N_i d_i^2}{\sum N_i}$$

$$\sum N_i S_T^2 = \sum N_i S_i^2 + \sum N_i d_i^2$$

总变异 组内变异 组间变异

$$SS_T = SS_W + SS_B$$



一、方差分析的基本原理

表 9-1 不同强度噪音下解数学题犯错误频数

	噪音(分贝)			k=3
	强(100)(A)	中(50)(B)	无(C)	
n=4	16	4	1	
	14	5	2	
	12	5	2	
	10	6	3	
\bar{X}_j	13	5	2	$\bar{X}_t = 6.67$

方差合成公式:

$$S_T^2 = \frac{\sum N_i S_i^2 + \sum N_i d_i^2}{\sum N_i}$$

【例9-5】

$$\sum N_i S_T^2 = \sum N_i S_i^2 + \sum N_i d_i^2$$

总变异

组内变异

组间变异

$$SS_T = SS_W + SS_B$$



一、方差分析的基本原理

总变异（ SS_T ）由两部分组成：

- ① 组间差异（ SS_B ）：主要由于接受不同实验处理而造成的各组之间的变异，主要指处理效应。
- ② 组内变异（ SS_W ）：由于实验中一些希望加以控制的非实验因素和一些未被有效控制的未知因素造成的变异，主要指个体差异、随机误差。



一、方差分析的基本原理

怎么检验？

在检验的时候，方差分析主要关心组间均方（ MS_B ）是否显著大于组内均方（ MS_W ）。

$$MS_B = \frac{SS_B}{df_B}$$

$$MS_W = \frac{SS_W}{df_W}$$

$$F = \frac{MS_B}{MS_W}$$

方差分析做的是F检验。

目的：检验各自代表的总体的方差是否一致，即变量带来的影响是否还在随机误差带来的影响范围内。

如果F值落在置信区间（95% or 99%）内，那么就意味着实验处理造成的差异没有超出误差的范围，接受综合虚无假设；否则，拒绝综合虚无假设。



一、方差分析的基本原理

(三) 方差分析的步骤

- (1) 求平方和SS;
- (2) 计算自由度df;
- (3) 计算均方MS (SS / df);
- (4) 计算F值;
- (5) 查单侧F表进行F检验;
- (6) 陈列方差分析表。

变异来源	平方和	自由度	均方	F	p
组间	258.67	2	129.34	48.44	0.01
组内	24	9	2.67		
总变异	282.67	11			



一、方差分析的基本原理

(四) 基本假定

1. **总体正态分布**：要求样本必须来自正态分布的总体。
2. **变异来源独立**：不同来源的变异在意义上必须明确，而且彼此要相互独立。
3. **各处理内方差齐性**：即各处理内的方差彼此应无显著差异，这是方差分析中最为重要的基本假定。
 - ✓ 方差齐性检验的方法采用**哈特莱 (Hartley) 的最大F比率法**：

$F_{max} = \frac{S_{max}^2}{S_{min}^2}$ ，求出各样本中方差最大值与最小值的比，通过查表判断。



一、方差分析的基本原理

(五) 基本概念

1. 因素/变量
2. 水平
3. 处理
4. 主效应
5. 交互效应
6. 简单效应
7. 完全随机设计/组间设计
8. 重复测量设计/组内设计
9. 混合设计



真题演练

方差分析需要满足的前提条件有（ ）。(多选) (312, 2008)

- A. 总体正态分布
- B. 各处理方差齐性
- C. 总体方差已知
- D. 各组样本容量相同

重复测量实验设计方差分析的假设是（ ）。(多选) (312, 2015)

- A. 不同处理水平下的总体方差相等
- B. 每个处理水平下的总体服从正态分布
- C. 不同处理水平下的总体服从正态分布
- D. 因变量的方差-协方差矩阵符合球形假设



真题演练

方差分析需要满足的前提条件有（ ）。(多选) (312, 2008)

- A. 总体正态分布
- B. 各处理方差齐性
- C. 总体方差已知
- D. 各组样本容量相同

答案：AB

重复测量实验设计方差分析的假设是（ ）。(多选) (312, 2015)

- A. 不同处理水平下的总体方差相等
- B. 每个处理水平下的总体服从正态分布
- C. 不同处理水平下的总体服从正态分布
- D. 因变量的方差-协方差矩阵符合球形假设



真题演练

方差分析需要满足的前提条件有（ ）。(多选) (312, 2008)

- A. 总体正态分布
- B. 各处理方差齐性
- C. 总体方差已知
- D. 各组样本容量相同

答案：AB

重复测量实验设计方差分析的假设是（ ）。(多选) (312, 2015)

- A. 不同处理水平下的总体方差相等
- B. 每个处理水平下的总体服从正态分布
- C. 不同处理水平下的总体服从正态分布
- D. 因变量的方差-协方差矩阵符合球形假设

答案：ABCD



真题演练

在方差分析中，拒绝综合虚无假设 $H_0: \mu_1 = \mu_2 = \mu_3$ ，则表明（ ）。
(北航，2018)

- A. μ_1 、 μ_2 、 μ_3 两两均不相等
- B. μ_1 、 μ_2 、 μ_3 两两均相等
- C. μ_1 、 μ_2 、 μ_3 的两两组合中至少有一对不相等
- D. μ_1 、 μ_2 、 μ_3 的两两组合中至少有一对相等



真题演练

在方差分析中，拒绝综合虚无假设 $H_0: \mu_1 = \mu_2 = \mu_3$ ，则表明（ ）。
(北航，2018)

- A. μ_1 、 μ_2 、 μ_3 两两均不相等
- B. μ_1 、 μ_2 、 μ_3 两两均相等
- C. μ_1 、 μ_2 、 μ_3 的两两组合中至少有一对不相等
- D. μ_1 、 μ_2 、 μ_3 的两两组合中至少有一对相等

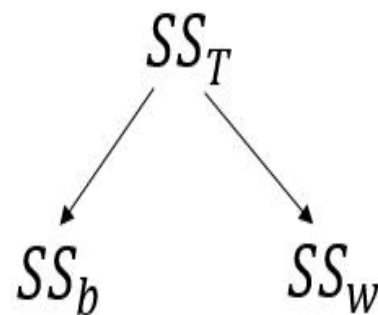
答案：C



二、完全随机设计的方差分析

(一) 单因素完全随机设计

1. 组间设计、被试间设计
2. 1个自变量，水平为k
3. $k \geq 2$
4. 实验组数=实验处理数=自变量水平数
5. 变异来源有2个：
 - ✓ 自变量 SS_b
 - ✓ 其他（测量误差、被试差异、偶然因素造成的差异） SS_w
6. 观测值=群体均值+自变量影响+误差





二、完全随机设计的方差分析

单因素完全随机设计方差分析表

变异来源	平方和	自由度	均方	F	p
组间	SS_b	$df_b = k - 1$	SS_b / df_b	MS_b / MS_w	$< \text{or} > \alpha$
组内	SS_w	$df_w = N - k$	SS_w / df_w		
总变异	SS_t	$df_t = N - 1$			



二、完全随机设计的方差分析

数据关系：

1. 总自由度 $df = N - 1$ ， N 是数据量，1个处理在1个被试身上产生一个数据。
2. $df_t = df_b + df_w$
3. 自变量的自由度=自变量水平-1
4. F检验的分母永远是误差项的均方
5. 组内变异是误差项

结果：

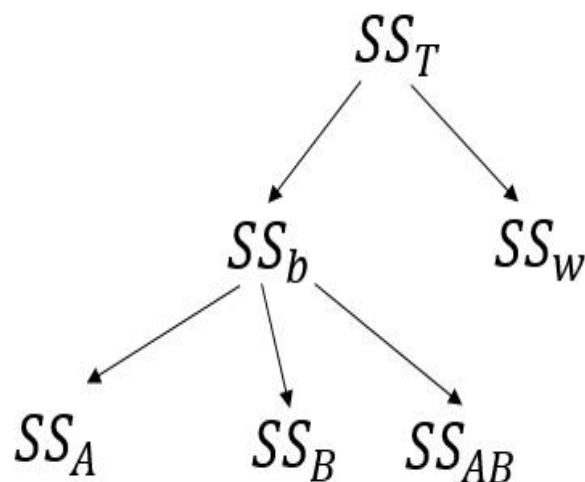
1. 1个主效应
2. “主效应显著&自变量水平 $k > 2$ ”，需要做事后检验。



二、完全随机设计的方差分析

(二) 两因素完全随机设计

1. 组间设计、被试间设计
2. 2个自变量A、B，水平数分别为a、b
3. $a \geq 2$ 、 $b \geq 2$
4. 该实验被描述为a*b的实验
5. 实验的组数=实验的处理数=a*b
6. 变异来源有2类4个：
 - ✓ a*b个组之间的变异
 - ✓ A变量 SS_A
 - ✓ B变量 SS_B
 - ✓ 两变量的交互作用 SS_{A*B}
 - ✓ 其他 SS_w
7. 观测值=群体均值+A变量影响+B变量影响+A*B影响+误差





二、完全随机设计的方差分析

两因素完全随机设计方差分析

变异来源	平方和	自由度	均方	F	p
组间	SS_b	$k - 1 = ab - 1$			
A因素	SS_A	$a - 1$	MS_A	MS_A / MS_w	$< \text{or} > \alpha$
B因素	SS_B	$b - 1$	MS_B	MS_B / MS_w	$< \text{or} > \alpha$
A×B	SS_{AB}	$(a - 1)(b - 1)$	MS_{AB}	MS_{AB} / MS_w	$< \text{or} > \alpha$
组内	SS_w	$N - k$	MS_w		
总变异	SS_t	$N - 1$			



二、完全随机设计的方差分析

数据关系：

1. 总自由度 $df=N-1$ ， N 是数据量
2. $df_t = df_b + df_w = df_A + df_B + df_{AB} + df_w$
3. 自变量的自由度=自变量水平-1
4. 交互作用的自由度=自变量A的自由度*自变量B的自由度
5. F检验的分母永远是误差项的均方
6. 组内变异是误差项

结果：

1. 2个主效应，1个交互作用
2. “主效应显著&自变量水平 $k > 2$ ”，需要做事后检验。
3. 交互作用显著，需要做简单效应分析。



二、完全随机设计的方差分析

(三) 单因素随机区组设计

1. 1个自变量
2. 自变量的水平 $k \geq 2$
3. 处理数=自变量水平数
4. 1个区组变量, 属于控制变量/无关变量
5. 区组变量的水平 $n \geq 2$
6. 区组数=区组变量水平数
7. 区组内的被试数量=自变量水平数 $\times m$ (≥ 1 的整数)



二、完全随机设计的方差分析

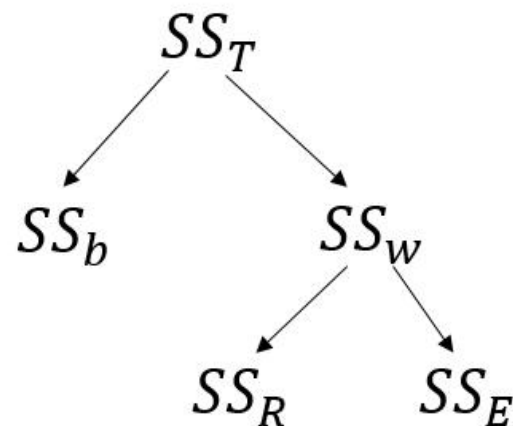
8. 设计原则：区组内尽量同质，区组间异质。
9. 优点：考虑到个体差异（区组效应）并将这种差异从组内变异中分离，提高实验效率，可获得对处理效应的更加精确的估计；
10. 缺点：区组划分较困难，若不能保证区组内同质，则有更大的误差。
11. 跟重复测量设计的差异：区组设计仅在区组变量这一项上有所控制；重复测量设计在所有被试变量上都控制了。



二、完全随机设计的方差分析

12. 变异来源有2类3个:

- ✓ 组间变异 SS_b (自变量造成的变异)
- ✓ 组内变异 SS_w
 - ✓ 区组间变异 SS_R
 - ✓ 其他变异 SS_E



13. 观测值=群体均值+区组差异+自变量影响+误差



二、完全随机设计的方差分析

单因素随机区组设计方差分析表 (m=1)

变异来源	平方和	自由度	均方	F	p
组间	SS_b	$k - 1$	MS_b	MS_b / MS_E	$< \text{or} > \alpha$
组内	SS_w				
区组	SS_R	$n - 1$	MS_R		
误差	SS_E	$N - n - k + 1$	MS_E		
总变异	SS_t	$N - 1$			



二、完全随机设计的方差分析

数据关系：

1. 总自由度 $df=N-1$ ， N 是数据量
2. $df_t = df_b + df_w = df_b + df_R + df_E$
3. 自变量的自由度=自变量水平-1
4. F检验的分母永远是误差项的均方

结果：

1. 1个主效应
2. “主效应显著&自变量水平 $k > 2$ ”，需要做事后检验。



真题演练

一个实验有3组被试，方差分析的组内自由度为27，则该实验的被试总数为（ ）。(312, 2011)

- A. 24 B. 28 C. 30 D. 81

某研究中共有3组被试，每组12人，进行方差分析时，误差自由度为（ ）。(312, 2016)

- A. 2 B. 11 C. 33 D. 35



真题演练

一个实验有3组被试，方差分析的组内自由度为27，则该实验的被试总数为（ ）。(312, 2011)

- A. 24 B. 28 C. 30 D. 81

答案：C

某研究中共有3组被试，每组12人，进行方差分析时，误差自由度为（ ）。(312, 2016)

- A. 2 B. 11 C. 33 D. 35



真题演练

一个实验有3组被试，方差分析的组内自由度为27，则该实验的被试总数为（ ）。(312, 2011)

- A. 24 B. 28 C. 30 D. 81

答案：C

某研究中共有3组被试，每组12人，进行方差分析时，误差自由度为（ ）。(312, 2016)

- A. 2 B. 11 C. 33 D. 35

答案：C



真题演练

为了评定某种抗抑郁药的效果，某公司进行了临床试验，被试随机分配到药物组、安慰剂组和控制组，最后完成研究的人数分别为10人、12人和15人。使用方差分析来检验实验处理是否有效果，确定 F 值的临界值时，分母的自由度为（ ）。(华中师大, 2017)

A. 2

B. 24

C. 34

D. 35



真题演练

为了评定某种抗抑郁药的效果，某公司进行了临床试验，被试随机分配到药物组、安慰剂组和控制组，最后完成研究的人数分别为10人、12人和15人。使用方差分析来检验实验处理是否有效果，确定 F 值的临界值时，分母的自由度为（ ）。(华中师大, 2017)

A. 2

B. 24

C. 34

D. 35

答案：C



真题演练

研究者欲研究某城市居民环境保护的动机类型（A因素：短暂动机、长远动机）和环保行为（B因素：宣传、批评、治理）对所在社区环境质量的影响。研究者欲采用组间设计，每种处理方式下研究了10位成年居民，下面是尚未填写完毕的研究结果的方差分析表。请给出方差分析表中①②③④⑤对应的数值（312, 2013）

变异源	SS	df	MS	F
组间				
A	280	1	280	⑤
B	48	2	24	
A×B	120	②	60	
组内	540	③	④	
总和	①			



真题演练

研究者欲研究某城市居民环境保护的动机类型（A因素：短暂动机、长远动机）和环保行为（B因素：宣传、批评、治理）对所在社区环境质量的影响。研究者欲采用组间设计，每种处理方式下研究了10位成年居民，下面是尚未填写完毕的研究结果的方差分析表。请给出方差分析表中①②③④⑤对应的数值（312，2013）

变异源	SS	df	MS	F
组间				
A	280	1	280	28
B	48	2	24	
A×B	120	2	60	
组内	540	54	10	
总和	988			



真题演练

请完成下表的方差分析结果，并说明实验设计的类型以及被试的人数和分组情况。（华中师范大学，2016）

变异来源		SS	df	MS	F
组间					
	因子A	60	2	()	()
	因子B	40	2	()	()
	交互作用	100	()	()	()
组内		()	()	()	
总体		308	62		



真题演练

请完成下表的方差分析结果，并说明实验设计的类型以及被试的人数和分组情况。（华中师范大学，2016）

变异来源		SS	df	MS	F
组间					
	因子A	60	2	(30)	(15)
	因子B	40	2	(20)	(10)
	交互作用	100	(4)	(25)	(12.5)
组内		(108)	(54)	(2)	
总体		308	62		

THANKS

- 期待下次相遇 -



高途学院APP下载



高途学院公众号

