

课件

第二章 常用试验设计方法

张文庆

lsszwq@mail.sysu.edu.cn

试验设计的总目标：用较少的人力、物力和时间获得较多且可靠的数据资料。

主要内容

- 随机区组设计
- 平衡不完全区组设计
- 裂区设计
- 拉丁方设计
- 正交设计
- 二进制在试验设计中的应用
- 课堂练习

第二节 随机区组设计

随机区组设计

- 随机区组设计（**randomized blocks design**）：指根据局部控制和随机原则进行的，将试验单位按性质不同分成与重复数一样多的区组（窝组），使区组内非试验因素差异最小而区组间内非试验因素差异最大，每个区组均包括全部的处理。区组内各处理随机排列，各区组独立随机排列。
- 随机区组设计**通过划分区组降低对试验单元的整体要求。**

试验排列示意图

区组2

3	8	4	6	2	5	9	1	7
---	---	---	---	---	---	---	---	---

区组1

5	6	2	3	7	9	1	8	4
---	---	---	---	---	---	---	---	---

区组3

2	4	7	8	6	3	9	5	1
---	---	---	---	---	---	---	---	---

随机区组设计的优点

- (1) 设计简单
- (2) 富于弹性，单因素、多因素以及综合性的试验都可应用
- (3) 能提供无偏的误差估计
- (4) 对试验区的形状要求不严，不同区组亦可分散设置在不同地段上。

随机区组设计的不足

不允许处理数太多，至多不超过**20**个，最好**10**个左右。

随机区组设计的基本步骤

- (1) 划分区组。根据局部控制原则，将试验材料和其他可控的非处理因素一致性较好的划为同一组，进而将整个试验单元划分为若干组。
- (2) 区组内所有处理随机化。
- (3) 区组随机排列。

随机区组设计试验结果的统计分析

- 把区组（窝组）看作一个因素，和试验因素一起做方差分析。

单因素随机区组设计试验结果的统计分析

- 统计分析方法

把区组（窝组）看作一个因素，和试验因素一起作为二因素的试验，按二因素无重复观测值的方差分析方法进行。

案例2-2

- 利用注射法RNAi技术研究5个基因（A、B、C、D和E）对褐飞虱产卵量的影响，设阳性对照1个（卵黄原蛋白基因Vg），阴性对照一个（GFP）。由于工作量较大，每次完成1个试验重复（共3个重复）。数据如下表。试比较5个基因对褐飞虱产卵量的影响。

案例2-2

	产卵量（粒）						
	GFP	Vg	A	B	C	D	E
区组1	415	132	215	262	375	406	333
区组2	463	153	231	260	392	432	351
区组3	432	150	243	284	401	460	312

加载Excel数据分析模块的步骤

- (1) 在“文件”主菜单点击“选项”，
- (2) 在最左列，点击“加载项”，
- (3) 在最下方，点击“转到”，出现分析工具库选择，
- (4) 勾选分析工具库、分析工具库-VBA，点击“确定”，
即完成加载。

案例2-2：方差分析

本题中，3个重复在不同的时间完成，所使用的试验材料也有所不同，因此，这是一个**单因素随机区组设计**。在Excel数据分析模块中，选择“无重复双因素方差分析”，结果表明：褐飞虱基因这一因素的计算F值为**156.48**， **$p < 0.0001$** ，说明7个褐飞虱基因对其产卵量的影响有极显著差异。

（见第2章excel文件）

二因素随机区组设计试验结果的统计分析(1)

- 区组作为一个因素，所以需要做多因素的方差分析
- 但是，**Excel**只有二因素的方差分析，所以需要自己计算部分结果

二因素随机区组设计试验结果的统计分析(2)

- 平方和和自由度的分解

$$SS_T = SS_t + SS_r + SS_e$$

SS_t : 处理间的平方和, SS_r : 区组间的平方和

SS_e : 试验误差的平方和

$$SS_t = SS_A + SS_B + SS_{AB}$$

$$df_T = df_t + df_r + df_e$$

$$df_t = df_A + df_B + df_{AB}$$

二因素随机区组设计试验结果的统计分析(3)

- 各项的方差

$$s_A^2 = SS_A / df_A$$

$$s_B^2 = SS_B / df_B$$

$$s_{AB}^2 = SS_{AB} / df_{AB}$$

$$s_r^2 = SS_r / df_r$$

$$s_e^2 = SS_e / df_e$$

二因素随机区组设计试验结果的统计分析(4)

- F检验（以固定模型为例）

$$F_A = S_A^2 / S_e^2$$

$$F_B = S_B^2 / S_e^2$$

$$F_{AB} = S_{AB}^2 / S_e^2$$

$$F_r = S_r^2 / S_e^2$$

例题1

- 微肥种类（A）与施用方式（B）对小麦产量的影响，二因素随机区组设计，试验结果见Excel文件。试作方差分析，并进行多重比较。

例题1—方差分析(1)

- 重新整理数据（参照有重复数据的二因素方差分析）
- 利用Excel中有重复数据的二因素方差分析进行
- 获得 SS_A 、 SS_B 、 SS_{AB} ， df_A 、 df_B 、 df_{AB} 以及 S_A^2 、

$$S_B^2、S_{AB}^2$$

但是， SS_r 、 SS_e 、 df_r 、 df_e 、 S_r^2 、 S_e^2 需要重新计算

例题1—方差分析(2)

$$SS_r = \frac{\sum T_r^2}{ab} - C \quad C = \frac{T^2}{abn}$$

SS_e = Excel计算的内部平方和 - SS_r

$$df_r = n - 1$$

df_e = Excel计算的内部自由度 - df_r

$$s_r^2 = SS_r / df_r$$

$$s_e^2 = SS_e / df_e$$

例题1—方差分析(3)

$$F_A=827.48$$

$$F_B=816.82$$

$$F_{AB}=123.08$$

$$F_r=25.32$$

查F表，以A因素为例， $F(1,14)_{0.05}=4.60$

A因素（施肥方式）间差异显著

例题1—多重比较

以AxB互作为例（SSR检验）

$$S_x^- = \sqrt{\frac{S_e^2}{n}} = 0.9906$$

LSR_{0.05}

M=2	3.00
M=3	3.15
M=4	3.28

完全随机设计

- 每个处理随机地选取实验单元，这种方式适用于实验的例数较大或实验单元差异很小的情况。
- 要求所有实验单元的条件尽可能一致，已知可控制的非处理因素尽可能一致
- 所有实验单元（含重复）随机排列
- 要求同时完成所有实验（含重复）。

完全随机设计与随机区组设计的区别

(1) 非处理因素。随机区组设计：每个区组内尽可能一致；完全随机设计：所有45个处理组合尽可能一致。

(2) 实验材料。随机区组设计：每个区组内尽可能一致；完全随机设计：所有45个处理组合尽可能一致。

(3) 实验时间。随机区组设计：每个区组在同一时间开展；完全随机设计：所有组合在同一时间开展。

(4) 随机排列。随机区组设计：在同一区组内，15个处理组合随机排列；完全随机设计：所有45个组合随机排列。

随机区组设计

区组2	A ₃ B ₁	A ₁ B ₄	A ₁ B ₁	A ₂ B ₅	A ₂ B ₄	A ₁ B ₅	A ₁ B ₂	A ₂ B ₁	A ₃ B ₅	A ₃ B ₂
	A ₂ B ₃	A ₁ B ₃	A ₂ B ₂	A ₃ B ₄	A ₃ B ₃					
区组3	A ₂ B ₃	A ₂ B ₄	A ₁ B ₄	A ₁ B ₂	A ₁ B ₁	A ₃ B ₃	A ₂ B ₂	A ₃ B ₂	A ₃ B ₁	A ₁ B ₃
	A ₃ B ₄	A ₁ B ₅	A ₂ B ₁	A ₂ B ₅	A ₃ B ₅					
区组1	A ₃ B ₄	A ₁ B ₃	A ₁ B ₂	A ₁ B ₅	A ₂ B ₃	A ₃ B ₅	A ₃ B ₃	A ₂ B ₅	A ₂ B ₄	A ₃ B ₁
	A ₁ B ₄	A ₂ B ₁	A ₁ B ₁	A ₃ B ₂	A ₂ B ₂					

完全随机设计

A ₁ B ₃	A ₂ B ₂	A ₁ B ₄	A ₂ B ₃	A ₂ B ₄	A ₃ B ₁	A ₁ B ₅	A ₂ B ₄	A ₂ B ₁	A ₂ B ₂	A ₃ B ₂	A ₁ B ₁
A ₁ B ₃	A ₂ B ₂	A ₃ B ₄	A ₃ B ₅	A ₂ B ₃	A ₂ B ₃	A ₁ B ₂	A ₁ B ₄	A ₁ B ₁	A ₃ B ₃	A ₃ B ₅	A ₃ B ₁
A ₃ B ₂	A ₁ B ₃	A ₃ B ₄	A ₂ B ₁	A ₂ B ₄	A ₁ B ₂	A ₂ B ₅	A ₃ B ₅	A ₂ B ₅	A ₃ B ₄	A ₂ B ₁	A ₁ B ₅
A ₁ B ₂	A ₁ B ₅	A ₃ B ₃	A ₃ B ₂	A ₂ B ₅	A ₃ B ₁	A ₁ B ₄	A ₁ B ₁	A ₃ B ₃			

成对试验

- 同一对的2个试验单位的条件尽可能一致
- 在同一对中，2个试验单位随机分配到处理组合对照组

配对	处理组	对照组
1		
2		
3		
4		
5		

思考

- 成对数据的t检验中：试验方法就是成对试验
- 思考：成对试验与随机区组试验的关系？

第三节 平衡不完全区组设计

平衡不完全区组设计 (1)

- 基本思想：不要求每一区组包含全部处理
- 特点：
 - (1) 每个处理在每一区组中至多出现一次
 - (2) 每个处理在全部试验中出现的次数均等
 - (3) 任何两个处理都有机会出现于同一区组中，且在全部试验中任意两个处理出现于同一区组中的次数均等

平衡不完全区组设计 (2)

$$rv = bk$$

$$\lambda = \frac{r(k-1)}{v-1}$$

$$b \geq v$$

v: 试验处理数

k: 每一区组包含的处理数

r: 每一处理的重复数

b: 区组数

λ : 任意两个处理在同一区组中相遇的次数 (整数)

平衡不完全区组设计 (3)

- 对于不同的参数值，有不同的设计方案
- 附录3：实用价值较大的平衡不完全区组设计方案

设计7: $v=6, k=3, r=5, b=10, \lambda=2$

1 2 5	2 3 4
1 2 6	2 3 5
1 3 4	2 4 6
1 3 6	3 5 6
1 4 5	4 5 6

平衡不完全区组设计的优点和不足

- 优点：不要求每一区组包含全部处理
- 不足：
 - (1) 区组数必须严格按照规定数目设计，否则会失去均衡
 - (2) 平衡不完全区组的实际总小区数往往比随机区组的多
 - (3) 平衡不完全区组试验设计中两个处理间比较的精确度较低

$$E = \frac{1 - \frac{1}{k}}{1 - \frac{1}{v}}$$

只有当难以进行随机区组设计时才用

平衡不完全区组设计的统计分析(1)

$$SS_T = SS_t \text{ (调整的)} + SS_r + SS_e$$

$$SS_T = \sum \sum x_{ij}^2 - C \qquad C = \frac{T_{..}^2}{N} \qquad N = rv$$

$$SS_t \text{ (调整的)} = \frac{\sum Q_i^2}{\lambda kv}$$

$$SS_r = \frac{\sum R_{.j}^2}{k} - C$$

$$df_T = df_t + df_r + df_e \qquad df_T = N - 1, \quad df_t = v - 1, \quad df_r = b - 1$$

平衡不完全区组设计的统计分析(2)

- **F检验**

$$F = S_{t(\text{调整的})}^2 / S_e^2$$

- 进行多重比较时, 需要调整的平均数 \bar{v}_i

$$\bar{v}_i = \frac{Q_i}{\lambda v} + \frac{T_{..}}{N}$$

其标准误为:

$$s_v = \sqrt{S_e^2 \times \frac{k}{\lambda v}}$$

案例2-4

- 在案例2-2中，每个区组需要注射280头褐飞虱成虫，共3个区组。如果在第一次注射时，出现突发情况（如仪器故障），最后只完成了3个基因的dsRNA注射。这时，操作者面临两个选择：（1）如果继续采用随机区组设计，本次注射只能作废（因为该区组没有包括所有的处理和对照）；（2）采用其他试验设计方法补救。如果选择方案2，则可以采用平衡不完全区组设计。

选择设计12

- 本题中， $v=7$ （试验处理数）， $k=3$ （每一区组包含的处理数）
- 查附件3，选择设计12（ $r=3$ （每一处理的重复数） $b=7$ （区组数）， $\lambda=1$ （任意两个处理在同一区组中相遇的次数））：

1	2	4
2	3	5
3	4	6
4	5	7
5	6	1
6	7	2
7	1	3

案例2-4—方差分析

- 计算平方和、自由度（第2章D2_2.xlsx文件）
- 计算F值
 $F=58.42$
- 查表 $F(6, 8)_{0.01}=6.63$
 $df_t=v-1=6, df_e=8$

各基因对褐飞虱产卵量的影响存在极显著的差异

多重比较

调整的平均数 $\bar{v}_i = \frac{Q_i}{\lambda_v} + \frac{T_{..}}{N}$

SSR检验:

$S_v = 14.2668$

比较

单因素随机区组设计 单因素平衡不完全区组设计 的异同

第四节 裂区设计

裂区设计

- 裂区设计是先将每一区组按第一因素的处理数划分小区，称为主区，在主区里随机安排主处理；然后在主区内引进第二个因素的各个处理（即副处理），就是主处理的小区内分设与副处理数相等的更小的小区，称为副区或裂区，在副区里随机排列副处理。
- 裂区设计的特点：**主处理分设**在主区，**副处理则分设**于**主区内的副区**，副区之间比主区之间的试验空间更为接近，即副区的比较比主区的比较更为精确。

随机区组的裂区试验设计步骤

区组2

区组1

区组3



各区组内随机排列主处理 (A)



主处理的小区内引入随机排列的副处理 (B)



裂区设计的常用范围

- (1) 一个因素的各处理比另一个因素的各处理需要更大区域时
- (2) 试验中某一因素的主效比另一因素的主效更为重要，而且要求更精确时，或两个因素的交互作用比其主效更为重要时。将要求精度高的因素作为副处理
- (3) 某一因素的效应比另一因素的效应更大时。将可能表现较大差异的因素作为主处理
- (4) 试验设计需要临时改动再加入一个试验因素时。可在原来的小区（主区）中再设计副区。

裂区设计试验结果的统计分析(1)

- 平方和的分解

$$SS_T = SS_t + SS_r + SS_e$$

SS_t : 处理间的平方和, SS_r : 区组间的平方和,

$SS_e = SS_{e_a} + SS_{e_b}$: 试验误差的平方和

$$SS_t = SS_A + SS_B + SS_{AB}$$

$$SS_{e_a} = SS_m - SS_A - SS_r$$

$$SS_m = \frac{1}{b} \sum T_m^2 - C$$

$$SS_r = \frac{1}{ab} \sum T_r^2 - C$$

裂区设计试验结果的统计分析(2)

- 自由度的分解

$$df_T = df_t + df_r + df_e \quad df_T = abn - 1, df_r = n - 1$$

$$df_t = df_A + df_B + df_{AB} \quad df_A = a - 1, df_B = b - 1, df_{AB} = (a - 1)(b - 1)$$

$$df_e = df_{ea} + df_{eb}$$

$$df_{ea} = (a - 1)(n - 1)$$

$$df_{eb} = a(b - 1)(n - 1)$$

裂区设计试验结果的统计分析(3)

- 各项的方差

$$s_A^2 = SS_A / df_A$$

$$s_B^2 = SS_B / df_B$$

$$s_{AB}^2 = SS_{AB} / df_{AB}$$

$$s_r^2 = SS_r / df_r$$

$$s_{ea}^2 = SS_{ea} / df_{ea}$$

$$s_{eb}^2 = SS_{eb} / df_{eb}$$

裂区设计试验结果的统计分析(4)

- F检验（以固定模型为例）

$$F_A = S_A^2 / S_{ea}^2 \quad F_r = S_r^2 / S_{ea}^2$$

$$F_B = S_B^2 / S_{eb}^2$$

$$F_{AB} = S_{AB}^2 / S_{eb}^2$$

裂区设计试验结果的统计分析(5)

- 多重比较

标准误的计算:

(1) 主区因子间比较: $SX = \sqrt{\frac{2s_{ea}^2}{bn}}$

(2) 副区因子间比较: $SX = \sqrt{\frac{2s_{eb}^2}{an}}$

(3) 同一主区因素水平下两个副区间的比较: $SX = \sqrt{\frac{2s_{eb}^2}{n}}$

(见第2章Excel文件)

例题2

- 研究绿肥耕翻时期（**A**因素）与施用氮肥量（**B**因素）对甜菜产量的影响，采用二裂式裂区设计。**A1**、**A2**（主区），**B1**、**B2**、**B3**、**B4**（副区），重复3次。数据见Excel文件。试作方差分析，并作多重比较。

区组1

A1				A2			
B2	B3	B1	B4	B1	B4	B3	B2

例2—方差分析(1)

- 重新整理数据（参照有重复数据的二因素方差分析）
- 利用Excel中有重复数据的二因素方差分析进行
- 获得 SS_A 、 SS_B 、 SS_{AB} ， df_A 、 df_B 、 df_{AB} 以及 S_A^2 、

$$S_B^2、S_{AB}^2$$

但是， SS_r 、 SS_{ea} 、 SS_{eb} 、 df_r 、 df_{ea} 、 df_{eb} 、 S_r^2 、 S_{ea}^2 、 S_{eb}^2

需要重新计算（见Excel文件）

方差分析中获得的误差项平方和= $SS_r+SS_{ea}+SS_{eb}$

例2—方差分析(2)

$$F_A=104.06$$

$$F_B=118.96$$

$$F_{AB}=10.33$$

$$F_r=1.56$$

查F表，以A因素为例， $F(1,2)_{0.05}=18.51$

A因素间差异显著

例2—多重比较

(见第2章Excel文件)

比较

二因素随机区组设计
二因素裂区设计
的异同

第五节 拉丁方设计

拉丁方设计

- 拉丁方设计（**latin square design**），就是在行和列两个方向上都进行局部控制，**使行、列两向皆成完全随机区组或重复**，是比随机区组设计多一个区组的设计。
- 拉丁方设计的特点：处理数、重复数、行数、列数均相等。
- 可控制试验误差、提高试验精确度。据研究，拉丁方设计的误差约为随机区组设计的**73%**。
- 试验处理数：**5-10**个为宜。
- 在动物试验中，如要控制来自两个方面的系统误差，且在试验动物头数较少的情况下，常采用这种方法。

例题3

- 研究5种不同饲料（分别用1， 2， 3， 4， 5表示）对乳牛产乳量的影响。试用拉丁方设计进行试验， 并做统计分析。

拉丁方试验设计的步骤(1)

1、选择标准方（附录4）

处理数 $k=5$ ，所以选择 5×5 的标准方

泌乳时间		一月	二月	三月	四月	五月
牛号	I	A	B	C	D	E
	II	B	A	E	C	D
	III	C	D	A	E	B
	IV	D	E	B	A	C
	V	E	C	D	B	A

A、B、C、D、E：分别代表5种饲料

拉丁方试验设计的步骤(2)

2、列随机

3、行随机

4、饲料（处理）随机

泌乳时间		一月	二月	三月	四月	五月
牛号	I	2	5	1	3	4
	II	4	3	2	1	5
	III	1	2	4	5	3
	IV	5	4	3	2	1
	V	3	1	5	4	2

拉丁方设计试验结果的统计分析(1)

- 平方和的分解

$$SS_T = SS_t + SS_r + SS_e$$

SS_t : 处理间的平方和, SS_r : 区组间的平方和,

$$SS_r = SS_{\text{行}} + SS_{\text{列}}$$

$$SS_{\text{行}} = \frac{1}{k} \sum T_r^2 - C$$

$$SS_{\text{列}} = \frac{1}{k} \sum T_c^2 - C$$

$$SS_t = \frac{1}{k} \sum T_t^2 - C$$

$$SS_T = \sum \sum x_{ij}^2 - C$$

$$C = \frac{T^2}{k \times k}$$

拉丁方设计试验结果的统计分析(2)

- 自由度的分解

$$df_T = k(k-1)$$

$$df_t = k-1$$

$$df_{\text{列}} = k-1$$

$$df_{\text{行}} = k-1$$

$$df_e = df_T - df_t - df_{\text{列}} - df_{\text{行}}$$

拉丁方设计试验结果的统计分析(3)

- 各项的方差

$$s_{\text{行}}^2 = SS_{\text{行}} / df_{\text{行}}$$

$$s_{\text{列}}^2 = SS_{\text{列}} / df_{\text{列}}$$

$$s_t^2 = SS_t / df_t$$

$$s_e^2 = SS_e / df_e$$

拉丁方设计试验结果的统计分析(4)

- F检验

$$F_t = S_t^2 / S_e^2$$

拉丁方设计试验结果的统计分析(5)

- 多重比较（以q检验为例）

$$S_x^- = \sqrt{\frac{S_e^2}{k}}$$

例3-方差分析

- 见第2章Excel文件

思考并比较

- 研究5种不同饲料（分别用1， 2， 3， 4， 5表示）对乳牛产乳量的影响。试分别用随机区组设计和拉丁方设计进行试验，请比较两种试验设计的异同。

比较 拉丁方设计 随机区组设计 的异同

第六节 正交设计

正交设计

- 正交设计(orthogonal design): 是利用规格化的表格—正交表来科学合理地安排试验的设计方法, 其特点是在全部试验处理组合中, **挑选部分有代表性的水平组合** (处理组合) 进行试验。
- 例: 4因素3水平的多因素试验, 需要 $3^4=81$ 个处理组合。采用正交表 $L_9(3^4)$ 安排试验, 只需要9个处理组合就可以了。其中, 9表示试验次数; 3表示水平数; 4表示最多可以安排的因素的个数 (含互作)。

正交试验的步骤

- 1、确定试验因素数和水平数
- 2、选择合适的正交表（附录5）
- 3、进行表头设计，列出试验方案
- 4、开展试验

没有交互作用的情况(1)

2、选用合适的正交表

(1) 计算最少试验次数n

$$n = \sum(\text{水平数} - 1) + 1$$

如4因素2水平, $n = 4 * (2 - 1) + 1 = 5$

3因素3水平, $n = 3 * (3 - 1) + 1 = 7$

(2) 根据因素数、水平数和最少试验次数, 选择正交表

如4因素2水平, 可选用 $L_8(2^7)$

3因素3水平, 可选用 $L_9(3^4)$

没有交互作用的情况(2)

3、进行表头设计，列出试验方案

表头设计：就是把试验中确定研究的各因素填到正交表的表头各列。其原则是：（1）不要让主效应间、主效应与交互作用间有混杂现象。由于正交表中一般都有交互列，因此当试验因素数少于列数时，尽量不在交互列安排试验因素。（2）当存在交互作用时，需查交互作用表，将交互作用安排在合适的列上。

没有交互作用的情况 (3)

$L_8(2^7)$ 两列间的交互作用

列号	1	2	3	4	5	6	7
1		3	2	5	4	7	6
2			1	6	7	4	5
3				7	6	5	4
4					1	2	3
5					(5)	3	2
6						(6)	1

没有交互作用的情况(4)

表头设计

列号	1	2	3	4	5	6	7
因素	A	B	(AxB)	C	(AxC)	(BxC)	D

注：在此不考虑交互作用

列号	1	2	3	4
因素	A	B	C	

没有交互作用的情况 (5)

安排试验

试验号	A	B	3	C	5	6	D
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

没有交互作用的情况(6)

开展试验

做实验的顺序要依照随机化的原则。即8个试验号的开展顺序要随机

有交互作用的情况(1)

2、选用合适的正交表

(1) 计算最少试验次数n

$$n = \sum(\text{水平数}-1) + 1$$

如4因素2水平，考虑AxB、AxC互作，

$$n = 4*(2-1) + (2-1)*(2-1) + (2-1)*(2-1) + 1 = 7$$

5因素3水平，考虑AxB、BxC互作，

$$n = 5*(3-1) + (3-1)*(3-1) + (3-1)*(3-1) + 1 = 19$$

(2) 选择正交表

如4因素2水平，可选用 $L_8(2^7)$

5因素3水平，可选用 $L_{27}(3^{13})$

有交互作用的情况(2)

3、进行表头设计，列出试验方案

列号	1	2	3	4	5	6	7
因素	A	B	AxB	C	AxC		D

二水平时，交互作用只占1列

有交互作用的情况(3)

3、进行表头设计，列出试验方案

列号	1	2	3	4	5	6	7
因素	A	B	(AxB)1	(AxB)2	C	D	AxD

列号	8	9	10	11	12	13
因素	(BxC)1	E		(BxC)2		

三水平时，交互作用占2列

注：表中红色字体不安排，仅作说明用

正交试验结果的统计分析

- 直观分析
- 方差分析

例题4

- 某工厂生产一种产品，采收率低且不稳定，一般在60%-80%之间。现希望通过试验设计，找到好的生产方案，提高采收率。考虑3个因素：反应温度（A）、加碱量（B）、催化剂种类（C），每个因素分别设3个水平。

解：计算最少试验次数n

$$n=3*(3-1)+1=7$$

选用 $L_9(3^4)$ 正交表

表头设计

列号	1	2	3	4
因素	A	B	C	

正交试验结果的直观分析

- 见第2章Excel文件

正交试验结果的方差分析(1)

- 计算平方和

$$SS_T = \sum (y_i - \bar{y})^2$$

\bar{y} 为 y_i 的平均数

$$SS_A = \sum n_i (\bar{T}_{ai} - \bar{y})^2$$

n_i : 在第 i 水平下所做实验的次数

$$SS_B = \sum n_i (\bar{T}_{bi} - \bar{y})^2$$

\bar{T}_{ai} : (A因素第 i 水平下各 y_i 值之和)/ n_i

$$SS_C = \sum n_i (\bar{T}_{ci} - \bar{y})^2$$

$$SS_e = SS_T - SS_A - SS_B - SS_C$$

正交试验结果的方差分析(2)

- 计算自由度

$df_T = n - 1$, n 为实验数据个数

$df_A = a - 1$, a 为A因子的水平数

$df_B = b - 1$

$df_C = c - 1$

$df_e = df_T - df_A - df_B - df_C$

正交试验结果的方差分析(3)

- 方差计算

$$s_A^2 = SS_A / df_A$$

$$s_B^2 = SS_B / df_B$$

$$s_C^2 = SS_C / df_C$$

$$s_e^2 = SS_e / df_e$$

正交试验结果的统计分析(4)

- F检验

$$F_A = S_A^2 / S_e^2$$

$$F_B = S_B^2 / S_e^2$$

$$F_C = S_C^2 / S_e^2$$

正交试验结果的统计分析(5)

- 如果所有因素对实验结果均没显著影响，可将F值最小的因素的平方和与自由度合并到误差项，再重新做方差分析
- 如有交互作用，只需要把交互作用也作为因素看待，其余步骤相同。三水平时，交互作用占2列，计算平方和与自由度时，需要把两列的值相加。

(第2章Excel文件)

重复实验的方差分析 (1)

- 对k次重复实验，计算 K_i 等值时需要统计重复值
- 空白列也作为一个因素参与计算

重复实验的方差分析(2)

- 计算平方和

$$SS_T = \sum (y_i - \bar{y})^2$$

y_i : 包括重复实验的值

$$SS_{\text{因}} = \sum n_i (\bar{T}_{\text{因}i} - \bar{y})^2$$

n_i : 在第*i*水平下所做实验的次数($n_i \cdot k$)

$\bar{T}_{\text{因}i}$: (该因素第*i*水平下各 y_i 值之和)/ n_i

纯误差平方和:

$$SS_{\text{ep}} = SS_T - \sum SS_{\text{列}}$$

$\sum SS_{\text{列}}$ 是正交表各列的平方和, 包括因素、交互作用和空白列

重复实验的方差分析(3)

- 计算自由度

$df_T = n - 1$, n 为实验数据个数 (含重复数, 即 $n*k$)

$df_{\text{因}} = a - 1$, a 为该因子的水平数

$$df_{eP} = df_T - \sum df_{\text{列}}$$

重复实验的方差分析(4)

- SA^2 、 F_A 等的计算与没有重复数据的情况相同。

例题5

- 对某中药进行提取工艺改进实验，以提取率为实验指标，采用 $L_9(3^4)$ 正交表做**重复**实验，三个因素分别为溶剂量、提取时间和提取次数。数据见Excel，试做方差分析。

(见第2章Excel文件)

第七节 二进制在试验设计中的应用

案例2-8

- 有8瓶水，其中1瓶有毒。拟用小鼠做试验，确定哪瓶水有毒。假设仅进行一次试验（所有小鼠在同一时间开始试验），试问最少需要多少只小鼠？写出简略实验过程。

二进制和十进制

- 二进制: 0,1
- 十进制: 0,1,2...9
- 十进制转二进制的方法

十进制数除2取余法，即十进制数除2，余数为权位上的数，得到的商值继续除2，依此步骤继续向下运算直到商为0为止。

十进制转换成二进制

以6为例，把十进制转换成二进制

- 第一步： $6/2$: 商为3，余0
- 第二步： $3/2$: 商为1，余1
- 第三步： $1/2$: 商为0，余1

- 所以，6的二进制编码为110

8
瓶
水
的
编
号

十进制

二进制

0

000

1

001

2

010

3

011

4

100

5

101

6

110

7

111

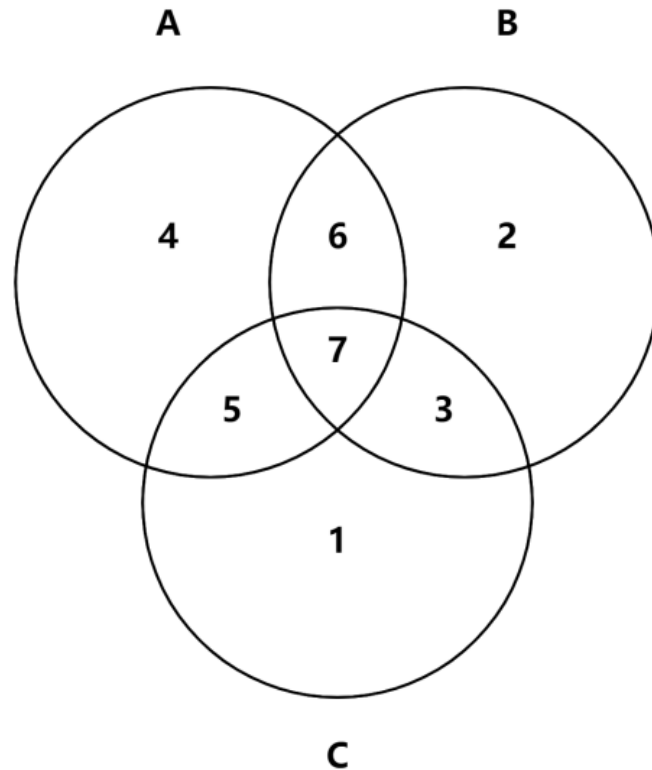
利用二进制--最少需要3只小鼠

第A只小鼠：饲喂100, 101, 110, 111共4瓶水的混合物

第B只小鼠：饲喂010, 011, 110, 111共4瓶水的混合物

第C只小鼠：饲喂001, 011, 101, 111共4瓶水的混合物

- 根据死亡（或表现中毒迹象）小鼠的**编号及数量**（**1只、2只或3只**），即可判断哪瓶水有毒



思考与讨论

- 要求只能进行一次试验时，最少小鼠数量：**1只**

案例2-9

- 某小麦品种在田间出现自然变异植株的概率为**0.0045**。（1）期望有**0.99**的概率获得**1株或1株以上**变异植株，至少应调查多少株？（2）假设变异植株的判定标准是某个基因片段的一个已知的单核苷酸多态性（**SNP**），如何找出变异植株？

解：（1）出现变异植株的概率 **$p=0.0045$** ， **$q=1-p=0.9955$**

（二项分布）

$P(x \geq 1) = 1 - P(0) = 0.99$ ，所以 **$P(0) = 0.01$** ，即： **$C_n^0 p^0 q^n = 0.01$**

$n=1021$

案例2-9

解：

(2) 利用二进制找出变异植株的初步思路。首先，考虑到 $2^{10}=1024$ ，因此在田间随机选取**1024株小麦植株**（十进制编号**0-1023**）；其次，把**1024株按二进制编号**；第三，分**10组**，其中，第**1组**是二进制第**10位数为1**的植株（**1000000000, 1000000001,, 1111111111**），第**2组**是二进制第**9位数为1**的植株（**0100000000, 0100000001,, 1111111111**），依次类推。

案例2-9

解：

(2) 第四，针对每一组，取每一植株某组织（基因在该组织高表达）的一定数量提取DNA，然后把该组所有植株（共512株）的DNA等量混合后进行高通量测序。最后，根据测序结果即可判断那一植株有SNP变化（变异），判别方法可参考案例2-8。

Box 2-8: 确定理论样本量的意义和作用

在案例2-9中，明白至少要调查1021株，有何意义或作用？

- (1) 根据试验方法，可估算所需要的条件和大致时间。
- (2) 树立信心。在植物学研究中，有时需要变异植株作为试验材料。如果调查了几百株还没有获得突变株是正常的，不要怀疑试验方法，因为调查的数量还不够。

课堂练习

1. 比较6种不同类型饲料对奶牛产乳量的影响，如果试验用奶牛数量只有6头，且个体间差异较大，该如何做试验设计？
2. 现有A、B和C三个因素，考虑B与C的交互作用。试完成正交试验的表头设计。

讨论

在新冠病毒的PCR检测中，为什么没有采用二进制的方法？

谢谢！