

试验设计类型之无法考察交互作用的多因素设计： 随机区组设计与拉丁方设计

谷恒明¹ 胡良平^{1 2*}

(1. 军事医学科学院生物医学统计学咨询中心, 北京 100850;

2. 世界中医药学会联合会临床科研统计学专业委员会, 北京 100029

* 通信作者: 胡良平, E-mail: lphu812@sina.com)

【摘要】 本文目的是介绍两种无法考察交互作用的多因素设计类型, 即随机区组设计与拉丁方设计。通过详细介绍与前述两种设计类型对应的“主要内容”“操作方法”和“设计变形”, 全面展示了这两种设计类型的核心内容、要领和注意事项, 为用户正确、合理地选用前述的两种设计类型创造了有利的条件。

【关键词】 交互作用; 随机区组设计; 拉丁方设计; 交叉设计; 嵌套设计

中图分类号: R195.1

文献标识码: A

doi: 10.11886/j.issn.1007-3256.2017.01.003

Types of the multifactor experimental designs with no interaction: randomized block design and Latin square design

Gu Hengming¹, Hu Liangping^{1 2*}

(1. Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China;

2. Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing 100029, China

* Corresponding author: Hu Liangping, E-mail: lphu812@sina.com)

【Abstract】 The purpose of this paper is to introduce two types of the multifactor experimental designs with no interaction—randomized block design and Latin square design. Through the detailed introduction of the "main contents", "operation method" and "design deformation" corresponding to the above two types of designs, the three aspects, such as the core contents, essentials and matters needing attention, are displayed comprehensively in this article, which results to provide better conditions for users choose a correct and reasonable design type from the two designs.

【Keywords】 Interaction; Randomized block design; Latin square design; Crossover design; Nested design

1 概 述

1.1 携带式与统筹式多因素设计的区别

携带式多因素设计, 就是以单因素 K (≥ 2) 水平设计为“骨架”, 记录各组受试对象所有重要非试验因素的取值。临床科研中绝大多数都采用此种多因素设计类型^[1-4]。

所谓统筹式多因素设计, 就是在试验研究开始之前, 就决定了多个试验因素及其水平, 并依据一定原理(如因素水平全面组合或正交性或均匀性或优良性)将全部拟同时考察的试验因素进行合理布局, 以表格的形式呈现出各试验点(即各因素分别取什么水平的一种特定试验条件)是如何组合的。如析因设计、正交设计、均匀设计、组合设计、最优设计等^[5-10]。

上述两类多因素设计之间的区别在于:

携带式多因素设计属于“顺水推舟”, 其结果准

确与否, 取决于样本含量。当样本含量足够大时, 其结论的可信度比较高。其优点是: 可以同时记录的影响因素非常多且不需要事先为其设定具体的水平, 灵活方便、适用面宽; 其缺点是: 需要的样本含量非常大且要求样本对于总体的代表性非常好, 不能明确得出哪些非试验因素或区组因素与充当“骨架”的那个单因素在什么样的水平组合下, 试验效果最好。

统筹式多因素设计属于“主动作为”, 其结果准确与否, 取决于设计者对试验因素及其交互作用有关信息把握的精准程度以及所选取的设计类型是否合适。其优点是: 能主动驾驭多个试验因素及其交互作用, 所需要的样本含量相对较少, 结果及预测结果的精确度较高; 其缺点是: 在样本含量相对较小的情况下, 不能精准估计试验因素之间高阶交互作用的效应大小。

1.2 何为交互作用

所谓因素之间的交互作用, 就是一个因素各水

项目基金: 国家高技术研究发展计划课题资助(2015AA020102)

平对评价指标影响不是恒定的,而是随着另一个因素的水平改变而变化(或多个因素的水平组合所处的具体的特定条件的改变而改变)。举一个通俗的例子:假定用不同种类的药物(设为 A 药与 B 药)治疗某病患者时,发现对男性患者而言, A 药的疗效优于 B 药,而对女性患者而言, A 药的疗效劣于 B 药。在这种情况下,就说“药物种类”与“患者性别”之间存在不可忽视的交互作用。

1.3 无法考察交互作用的多因素设计类型种类

在一个试验研究项目中,同时考察的因素数目 ≥ 2 时,就属于多因素试验研究。如何进行多因素设计,关键在于是否需要考察交互作用。若需要考察,考察哪些交互作用。考察的交互作用项越多、交互作用的阶数越高,对应的多因素设计类型就越复杂。一般来说,试验因素与区组因素之间的交互作用可以忽略不计,但在样本含量允许的条件下,最好先采取可以估计交互作用效应的统计模型对试验数据进行统计分析,依据初步统计分析结果决定是否需要在最终的统计分析模型中考虑交互作用项。

无法考察交互作用的多因素设计类型有如下几种:无重复试验的随机区组设计、无重复试验的平衡不完全随机区组设计、无重复试验的拉丁方设计、无重复试验的正交拉丁方设计、交叉设计、嵌套(或系统分组)设计,本文只介绍随机区组设计与拉丁方设计。

2 无法考察交互作用的多因素设计类型之要领

2.1 无重复试验的随机区组设计

2.1.1 问题与设计架构

【例 1】某研究者以小白鼠为受试对象,研究 3 种抗癌药物(A 药、B 药和 C 药)对其肉瘤的抑瘤疗效,所做的试验安排和得到的试验结果见表 1。问:这是什么试验设计类型?具体的操作是什么?

表 1 不同药物作用后小白鼠肉瘤重量

体重分组	肉瘤重量(g)			
	药物种类:	A	B	C
1		X	X	X
2		X	X	X
3		X	X	X
4		X	X	X
5		X	X	X

注:表中“X”代表待测定的肉瘤重量

2.1.2 无重复试验的随机区组设计及有关重要内容

表 1 的设计被称为无重复试验随机区组设计或无重复试验配伍组设计。该设计的特点是:可以考察一个试验因素和一个单纯型(只涉及一个重要非试验因素)或复合型(由多个重要非试验因素水平全面组合而成)区组因素(即重要非试验因素)对评价指标的影响。该设计的要求是:试验因素与区组因素之间的交互作用可以忽略不计。该设计与单因素 K (≥ 2) 水平设计的区别是:它比单因素 K (≥ 2) 水平设计多考察了一个区组因素,而且,在对受试对象进行分组时,它以分层随机化取代了单因素 K (≥ 2) 水平设计时采用的“完全随机化”。

2.1.3 无重复试验的随机区组设计的具体操作方法

第 1 步:由研究者确定试验因素的名称和具体水平(如本例,试验因素的名称为抗癌药物种类,共有 3 个水平,分别为 A 药、B 药和 C 药)。

第 2 步:根据药物性质和研究目的,确定受试对象的种类,制订纳入和排除标准,此处从略。在本例中,假定有根据地选定了小白鼠作为受试对象,并制订出了严格的纳入和排除标准。

第 3 步:根据文献资料或预试验结果估算出所需要的最少样本含量,在本例中,假定经过计算总共需要 15 只小白鼠。

第 4 步:假定由基本常识和专业知识的得知,小白鼠的原始体重可能对抗癌药物的疗效有较大的影响,于是,拟将“小白鼠体重分组”作为一个单纯型区组因素。先将 15 只染有肉瘤的小白鼠按体重大小形成 5 个区组,使每个区组内的 3 只小白鼠体重最接近。

第 5 步:随机决定每个区组中的 3 只小白鼠接受 3 种药物中的一种药物治疗,即在每个区组内对小白鼠进行随机分组(必须确保每个药物组都可以被分配到相同数目的小白鼠)。

第 6 步:以肉瘤的重量为指标,在标准操作规程指导下并实时精准地进行严格的质量控制,确保给药方法规范、观察和测定的方法与时间统一,从而获得原始、精准的试验结果。

第 7 步:将测自每个区组、每种药物条件下的那只小白鼠的“肉瘤重量数值”准确地填写入表 1 中相应位置上去,最终的试验结果见表 2。

表 2 不同药物作用后小白鼠肉瘤重量

体重分组	肉瘤重量(g)			
	药物种类:	A	B	C
1		0.82	0.65	0.51
2		0.73	0.54	0.23
3		0.43	0.34	0.28
4		0.41	0.21	0.31
5		0.68	0.43	0.24

2.1.4 随机区组设计的变形

第 1 种情形: 区组因素为“窝别”, 这是以动物为受试对象开展试验研究的优势, 因为以“窝别”为区组因素, 可以有效地控制“遗传因素”对评价指标的影响, 更好地显露试验因素不同水平之间的差别。

第 2 种情形: 可以构造出“具有重复试验的随机区组设计”。在例 1 中, 若各体重组中有 6 只小白鼠, 于是, 可将每个区组中的 6 只小白鼠随机均分入 3 个药物组中去, 这样就构造出“具有 2 次重复试验的随机区组设计”。显然, 对试验资料进行方差分析时, 误差项的自由度增大了, 结论的可信度就更高了。

表 3 用血压计多次测量家兔血压值

兔号	血压计种类及血压值(kPa)					
	测定次序:	1	2	3	4	5
1		C(X)	A(X)	E(X)	D(X)	B(X)
2		B(X)	E(X)	D(X)	C(X)	A(X)
3		A(X)	D(X)	C(X)	B(X)	E(X)
4		D(X)	B(X)	A(X)	E(X)	C(X)
5		E(X)	C(X)	B(X)	A(X)	D(X)

注: 表中“X”代表待测定的血压值

【例 3】某试验研究的资料见表 4。研究者关心的是药液(A~G)、离体肠管标本(a~g)、用药次序

第 3 种情形: 可以构造出含“复合型区组因素的随机区组设计”。在例 1 中, 可同时考虑小白鼠的体重、性别、年龄三个重要的非试验因素, 由它们复合成一个区组因素(即在同一个人区组内的小白鼠性别、年龄相同且体重最为接近)。这种“复合型区组因素”可以有效地控制多个重要非试验因素对评价指标的影响。

第 4 种情形: 在某些实际应用场合下, 由于每个区组内的受试对象的数目少于试验因素的水平数(例如有 4 种治疗脚气的药物, 而每位患有脚气病的患者只有两只脚), 此时, 可依据动态平衡的原则, 构造出被称为“平衡不完全随机区组设计”, 此设计比较繁琐, 需要时请参阅文献[11]。

2.2 无重复试验的拉丁方设计

2.2.1 问题与设计架构

【例 2】某研究者以家兔为受试对象, 研究 A、B、C、D、E 五种血压计测定的血压结果是否存在差异。所做的试验安排和得到的设计架构见表 3。问: 这是什么试验设计类型? 具体的操作方法是什么?

(1~7) 对作用强度的影响是否具有统计学意义。

表 4 不同药液、不同标本和不同用药次序对作用强度的影响

标本编号	药液种类与作用强度							
	用药次序:	1	2	3	4	5	6	7
a		A(21)	B(19)	C(0)	D(0)	E(5)	F(5)	G(2)
b		B(25)	E(4)	A(3)	G(0)	F(1)	D(2)	C(0)
c		C(0)	F(7)	G(10)	B(11)	D(7)	A(6)	E(4)
d		D(10)	G(4)	E(7)	F(7)	C(0)	B(17)	A(7)
e		E(6)	D(0)	B(9)	C(0)	A(1)	G(4)	F(5)
f		F(7)	C(0)	D(10)	A(11)	G(3)	E(6)	B(15)
g		G(3)	A(6)	F(3)	E(12)	B(26)	C(0)	D(6)

注: 各行上的标本被分成 7 份分别接受不同的药液处理

【例 4】为了研究 5 个不同剂量的甲状腺提取液对豚鼠甲状腺肿的影响,考虑到鼠的种系和体重对观测指标可能有一定的影响,设计试验时,最好将

这两个重要的非试验因素一并安排。根据专业知识得知,这三个因素之间的交互作用可忽略不计,设计格式和收集的资料见表 5。

表 5 5 个不同剂量的甲状腺提取液对豚鼠甲状腺重的影响

种 系	甲状腺提取液的剂量(字母) [(甲状腺重(g)/200g 体重)]					
	体重:	1	2	3	4	5
1		C(65)	E(85)	A(57)	B(49)	D(79)
2		E(82)	B(63)	D(77)	C(70)	A(46)
3		A(73)	D(68)	C(51)	E(76)	B(52)
4		D(92)	C(67)	B(63)	A(41)	E(68)
5		B(81)	A(56)	E(99)	D(75)	C(66)

注: A、B、C、D、E 所代表的具体剂量从略;每行上的 5 个数据测自种系相同的 5 只豚鼠

2.2.2 无重复试验的拉丁方设计及有关重要内容

表 3、表 4、表 5 的设计被称为无重复试验拉丁方设计。该设计的特点是:可以考察一个试验因素和一个单纯型(只涉及一个重要非试验因素)或复合型的(由多个重要非试验因素水平全面组合而成)区组因素(即重要非试验因素)以及一个名叫“测定次序”的区组因素对评价指标的影响。该设计的要求是:试验因素与两个区组因素之间的交互作用可以忽略不计。该设计与单因素 $K(\geq 2)$ 水平设计的区别是:它比单因素 $K(\geq 2)$ 水平设计多考察了两个区组因素。

2.2.3 无重复试验的拉丁方设计的具体操作方法

以表 3 的设计为例,介绍其具体操作方法:

第 1 步:由研究者确定试验因素的名称和具体取哪几个水平(如本例中,试验因素的名称为血压计种类,其有 5 个水平,分别为 A、B、C、D、E)。

第 2 步:根据研究目的,确定受试对象的种类,制订纳入和排除标准,此处从略。在本例中,假定有根据地选定了家兔作为受试对象,并制订出了严格的纳入和排除标准。

第 3 步:根据文献资料或预试验结果估算出所需要的最少样本含量(注意:在传统的拉丁方设计中,样本含量就是试验因素的水平数),在本例中,样本含量为 5 只家兔。

第 4 步:假定由基本常识和专业知识得知:家兔个体差异比较大,于是,拟将“家兔编号”作为一个单纯型区组因素,而将测定次序作为另一个区组因素。

第 5 步:将五种血压计的代号 A、B、C、D、E 随机排列成 5 行 5 列的方阵,简称为 5×5 拉丁方阵(注意:最早人们使用的是拉丁字母,而不是英文字母,由此而得名“拉丁方阵”),在本例中,方阵的具体排列形式见表 3 中表身部分。再在表 3 的左边放置“家兔编号”及其水平 1~5,在表 3 的表头短横线下方放置“测定次序”及其水平 1~5。

第 6 步:以血压值为指标,在标准操作规程指导下并实时精准地进行严格的质量控制,确保测定方法规范,统一观察和测定的方法与时间,从而获得原始、精准的试验数据。

第 7 步:将测自每只家兔在每个测定时间点上的“血压值”准确地填写入表 3 中相应位置上去,最终的试验结果见表 6。

表 6 每只家兔在 5 个不同时间点上被测得的血压值

兔 号	血压计代号及血压值(kPa)					
	测定次序:	1	2	3	4	5
1		C(X ₁₁)	A(X ₁₂)	E(X ₁₃)	D(X ₁₄)	B(X ₁₅)
2		B(X ₂₁)	E(X ₂₂)	D(X ₂₃)	C(X ₂₄)	A(X ₂₅)
3		A(X ₃₁)	D(X ₃₂)	C(X ₃₃)	B(X ₃₄)	E(X ₃₅)
4		D(X ₄₁)	B(X ₄₂)	A(X ₄₃)	E(X ₄₄)	C(X ₄₅)
5		E(X ₅₁)	C(X ₅₂)	B(X ₅₃)	A(X ₅₄)	D(X ₅₅)

注:表中“X_{ij}”代表第 i 行第 j 列上测定出来的具体血压值(因为没有真实的原始数据)

2.2.4 拉丁方设计的变形

根据拉丁方设计中“横向区组因素”各水平的具体情况,可将拉丁方设计计划分成以下五种类型:

第一种类型:为“不可拆分的单个体型区组因素”(即每行上的多个数据重复测自同一个个体),见表 3。

第二种类型:为“可拆分的单个体型区组因素”(即每行上的多个数据测自同一个标本分解后的多个子标本),见表 4。

第三种类型:为“多个体型区组因素”(即每行上的 K 个数据测自 K 个条件接近的个体),见表 5。

第四种类型:安排一个试验因素和三个区组因素的拉丁方设计,被称为正交拉丁方设计,见表 7。在表 7 中,“血压计代号”为试验因素,而“护士代号”“受试者号”和“测定号”属于三个区组因素。

表 7 四台血压计测定结果

受试者号	血压计代号与护士代号(血压值(kPa))				
	测定号:	1	2	3	4
1		A α (X)	B β (X)	C γ (X)	D δ (X)
2		B γ (X)	A δ (X)	D α (X)	C β (X)
3		C δ (X)	D γ (X)	A β (X)	B α (X)
4		D β (X)	C α (X)	B δ (X)	A γ (X)

注: A - D 分别代表 4 台血压计、 α - δ 分别代表 4 名护士;圆括号中的“X”代表相应位置上的血压值

第五种类型:在上述第二种和第三种拉丁方设计中,若各行与各列交叉处有两个或两个以上独立的受试对象接受试验,就称为“有重复试验的拉丁方设计”。

2.2.5 正确运用拉丁方设计的要领

拉丁方设计正确与否,取决于两个方面:试验因素的性质和横向区组因素的具体情况。

情形一:当试验因素为“药物种类”或“剂量大小”时,应尽可能采取“多个体型区组因素”,也可采取“可拆分的单个体型区组因素”。此时应尽可能

避免使用“不可拆分的单个体型区组因素”。

情形二:当试验因素为“重量计量器种类(如秤或天平)”或“体温计种类”时,可以选择“不可拆分的单个体型区组因素”(可以减少样本含量),因为试验因素不会改变评价指标的数值。

情形三:当试验因素为“血压计种类”时,最好选择“多个体型区组因素”。此时若选择“不可拆分的单个体型区组因素”,若受试对象为动物,基本可行;若受试对象为人,结果的准确度会有所下降,这是由于人的“心理因素”可能会暂时影响评价指标的数值。

参考文献

- [1] 宋辉,方灵芝,陶用富,等.酒精戒断所致震颤谵妄与非震颤谵妄患者 DNA 氧化损伤的比较[J].四川精神卫生,2016,29(3):238-241.
- [2] 蔡颖莲,陈晓东,刘国雄,等.广东省怀集县中学生自杀未遂状况调查及相关因素分析[J].四川精神卫生,2016,29(3):249-253.
- [3] Young D, Lamb SE, Shah S, et al. High-frequency oscillation for acute respiratory distress syndrome[J]. N Engl J Med, 2013, 368(9): 806-813.
- [4] Vain NE, Satragno DS, Gorenstein AN, et al. Effect of gravity on volume of placental transfusion: a multicentre, randomised, non-inferiority trial[J]. Lancet, 2014, 384(9939): 235-240.
- [5] 任露泉. 试验优化设计与分析[M]. 2 版. 北京: 高等教育出版社, 2001: 246-278.
- [6] 田口玄一. 实验设计法(上)[M]. 北京: 机械工业出版社, 1987: 170-325.
- [7] 方开泰,马长兴. 正交与均匀试验设计[M]. 北京: 科学出版社, 2001: 83-211.
- [8] 王万中. 试验的设计与分析[M]. 北京: 高等教育出版社, 2004: 333-357.
- [9] 胡良平. 统计学三型理论在实验设计中的应用[M]. 北京: 人民军医出版社, 2006: 139-165.
- [10] Douglas CM. Design and analysis of experiments[M]. 北京: 人民邮电出版社, 2007: 119-558.
- [11] 薛仲三. 医学统计方法和原理[M]. 北京: 人民卫生出版社, 1978: 335-345.

(收稿日期:2017-02-11)

(本文编辑:吴俊林)