

试验设计类型之无法考察交互作用的多因素设计:嵌套设计

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】 本文目的是介绍一种无法考察交互作用的多因素设计类型, 即嵌套设计。通过介绍四个具体实例, 总结出两种不同的嵌套设计: ①试验因素之间具有自然属性上的嵌套关系; ②试验因素对评价指标的影响存在主次关系。通过对实例的解读, 揭示嵌套设计具有一个其他多因素设计所缺乏的明显特点, 即因素的水平是“纵向连接”而不是“横向组合”的; 并从正确合理应用的角度考量, 指出了使用嵌套设计时应注意的事项。

【关键词】 试验因素; 设计类型; 嵌套设计; 正交设计; 均匀设计; 组合设计

中图分类号: R195.1

文献标识码: A

doi:10.11886/j.issn.1007-3256.2017.02.001

Types of the multifactor experimental designs with no interaction: nested 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】 This paper aimed to introduce a type of the multifactor experimental designs with no interaction: nested design. Two different kinds of nested designs were summarized by introducing four concrete actual cases: ① There was a nested relationship between the experimental factors in natural; ② According to the influence of the experimental factors on the evaluating indicator, the experimental factors had a primary and secondary ones. Through the analysis of the example, an obvious characteristic of the nested design was revealed that the levels of the factors were "vertical linkage" rather than "horizontal combination". From the point of view of the proper and reasonable application, the matters need attention when using nested design are also pointed out in the paper.

【Keywords】 Experimental factors; Type of experimental designs; Nested design; Orthogonal design; Uniform design; Combination design

1 嵌套设计概述

“试验设计类型”就是在做试验之前, 采取某种方式或结构来安排“试验因素”。事实上, 在运用统计学解决一系列实际问题时, 不难发现, 人们在以下两种场合会涉及到与“试验设计类型”有关的概念和知识。

第一种场合, 就是在开展“试验研究”之前, 无论是在以动物或样品为受试对象的试验研究中, 还是在以人(健康志愿者或某病患者)为受试对象的临床试验研究中, 都涉及到“试验因素(因为它是试验研究中三要素之一)”。既然涉及到“试验因素”, 就必然涉及到“试验因素的个数”“每个试验因素的水平数以及具体的水平是什么”“如何将这些试验

因素及其水平进行组合”“是否需要考察试验因素之间的交互作用以及考察多少项”等诸多问题。为正确合理地解答前述问题的方案或架构, 就是要确定试验研究中的“试验设计类型”。

第二种场合, 就是对所获得的科研资料进行各种统计分析(特别是进行广义差异性分析)时。此时, 可能会有人提出疑问: 在做试验之前, 就已经确定了试验设计类型, 为何到对资料进行统计分析时, 又一次面临选定“试验设计类型”呢? 为了清楚地回答这个问题, 必需按下面两种情形分别阐述:

情形一, 拟分析产生于某种特定试验设计类型的科研资料。

事实上, 即使科研资料是来自事先确定的某种试验设计类型, 但在对资料进行统计分析时, 分析者需要结合课题的大小、不同的研究目的和资料的具

体情况,不仅需要对整个数据集做分析,还需要从不同角度对数据集进行剖分,即形成所谓的“亚组”,并对剖分后的各子数据集进行分析。有时,还需要针对分析结果,对无统计学意义的因素不予考虑,即在初步统计分析模型中,删除明显无统计学意义的因素,例如 $P > 0.05$ 所对应的项,只分析最有可能具有统计学意义的那些因素。在前面这些情况下,都可能导致原先的“试验设计类型”发生改变。因此,对来自特定试验设计类型的试验资料进行一系列统计分析的过程中,需要不断判定每次分析是基于什么样的试验设计类型。但在实际分析资料时,很多人可能根本就没有意识到这一点。

情形二,拟分析产生于某种非试验研究的科研资料。

很多人可能认为上面所写的“情形二”根本不存在。例如,在分析 2×2 列联表资料或高维列联表资料时,人们还需要事先判定其来自某种试验设计类型吗?事实上,大多数列联表资料来自“调查研究”,根本不涉及“试验”。然而,只要冷静地思考一下:即使是产生于病例对照研究的 2×2 列联表资料,它仍然涉及一个两水平的因素,一个二值的结果变量。我们仍然可以称呼它为“单因素两水平设计一元定性(确切地说,是二值的)资料”。再举个例子,进行直线相关分析与回归分析时,很少有人关注此时的资料来自什么“试验设计类型”,得到数据后直接录入计算机,调用统计软件进行分析。此时,很可能就出错了。因为分析者正在进行直线相关与回归分析的资料可能根本不符合特定的“试验设计类型”要求。那么,适合进行直线相关与回归分析的

资料必须是来自什么“试验设计类型”呢?答案是“单组设计”。因为来自真正“单组设计”的受试对象必须是“同质的”,即全部受试对象在所观测的两个定量变量上的取值仅受个体差异(应排除遗传因素)的影响,不受任何未考察或测量的其他因素的影响。例如,当你试图研究某地正常成年人身高与体重之间的关系时,若受试对象中夹杂着一定比例的“肥胖者”或“婴幼儿”或“贫血患者”,得出的统计分析结论肯定是错误的。

综上所述,可以概括地得出如下的经验:①应将习惯上所讲的“试验设计类型”改为“设计类型”;②在试验研究之前,就应认真考虑拟选用何种“设计类型”来具体安排因素及其水平;③在对来自任何研究类型的科研资料进行统计分析的全过程中,始终要关注“设计类型”。

本文将要介绍的“嵌套设计”,更多的是应该基于上述“第二种场合下的情形二”的角度,更有实际意义。

2 两种不同的嵌套设计^[1-2]

2.1 因素之间存在自然属性上的嵌套关系

【例 1】为研究萝卜叶子中 M 物质的含量,随机采集三棵萝卜(设为 A_1 、 A_2 、 A_3),各棵萝卜中随机取几片叶子(设为 B_1 、 B_2 、 B_3)作测定,结果见表 1。显然,M 物质含量的多少,不仅取决于所取自的叶子,更主要是取决于所取自的萝卜,即不同萝卜之间的变异大于同一棵萝卜叶子之间的变异,试判断其所对应的试验设计类型。

表 1 萝卜叶子中 M 物质的测定结果

编 号	M 物质含量(单位)							
	A_1 (B_1)	B_2)	A_2 (B_1)	B_2	B_3)	A_3 (B_1)	B_2)	
1	3.28	3.52	2.46	1.87	2.19	2.77	3.74	
2	3.09	3.48	2.44	1.92	2.19	2.66	3.44	
3	3.31	4.07	-	2.10	-	-	-	

【例 2】某研究者用某种动物进行了如下的试验研究:将 120 只某种小动物随机均分为 12 个小组,这 12 个小组由两个试验因素的水平组合而形成。第一个试验因素为“药物种类”,其两个水平分

别为“PDS 药”与“PTS 药”;第二个试验因素为“药物浓度”。将各组中动物评价指标(即骨髓红系祖细胞的含量)的算术平均值与标准差整理在表 2 中,试判断其所对应的试验设计类型。

表 2 PDS 与 PTS 对骨髓红系祖细胞的增殖作用($\bar{x} \pm s, n = 10$)

药物种类	骨髓红系祖细胞的含量(个/ 0.5×10^5)						
	药物浓度(ng/mL):	0.0	2.5	12.5	25.0	50.0	100.0
PDS		115.2 \pm 20.6	143.8 \pm 20.8	144.3 \pm 29.6	154.3 \pm 28.2	178.5 \pm 30.2	165.9 \pm 22.4
PTS		115.2 \pm 20.6	95.9 \pm 12.9	89.1 \pm 14.0	85.8 \pm 14.3	49.5 \pm 12.9	6.5 \pm 2.7

注:PDS,人参二醇皂苷;PTS,人参三醇皂苷

2.2 因素对主要评价指标的影响存在主次关系

【例 3】在某项化合物的试验研究中,涉及到催化剂的种类(因素 A)和温度(因素 B),由专业知识可知,对该化合物转化率的影响,催化剂的作用大于温度,而且,各催化剂条件下所用的温度不完全相同,设计格式和资料见表 3,试判断其所对应的试验设计类型。

表 3 不同催化剂在不同温度下对某化合物转化率影响的观测结果

试验批次	催化剂与温度 (°C):	转化率(%)		
		甲 (70 80 90)	乙 (55 65 75)	丙 (90 95 100)
1		82 91 85	65 62 56	71 75 85
2		84 88 83	61 59 60	67 78 89

【例 4】某中医临床医生收治了三种癌症患者,根据中医理论,将他们划分为血瘀证与非血瘀证两种类型,测定两项定量指标(即 MaxSUV 和 MeanSUV)的数值。将各组中患者各指标的算术平均值与标准差整理在表 4 中,试判断其所对应的试验设计类型。

表 4 两组不同组织类型的 NSCLC 肺部 ROI 的 MaxSUV 和 MeanSUV 比较($\bar{x} \pm s$)

组别		MaxSUV	MeanSUV
非血瘀证	腺癌(n=25)	4.73 ± 1.10	3.03 ± 0.05
	鳞癌(n=21)	4.38 ± 1.06	3.31 ± 0.93
	大细胞癌(n=8)	4.45 ± 1.47	3.57 ± 1.09
血瘀证	腺癌(n=25)	8.71 ± 1.38	5.43 ± 1.06
	鳞癌(n=24)	7.11 ± 2.11	4.91 ± 1.22
	大细胞癌(n=8)	8.24 ± 2.81	6.07 ± 1.84

3 对上述两种嵌套设计的解读

3.1 对表 1 和表 2 两个实际问题的解读

表 1 很容易理解,不同棵萝卜的结果变量取值之间的差别会比较大,而同一棵萝卜不同叶中结果变量取值之间的差别相对来说会小一些。可将一棵萝卜比作一个“家庭”,同一个家庭各“代次”在某主要评价指标上取值之间的差别会大于某一个特定“代次”的若干个“同辈”之间的差别。

表 2 稍微难理解一点。“药物种类”与“药物剂量”之间的关系似乎不同于“不同棵萝卜”与“同一棵萝卜不同叶”之间的关系。关键在于具体地给药方案是什么,若每个受试对象同时接受两种药物治疗,其列表格式与表 2 是不同的,此时,一个因素叫万方数据

做“PDS 浓度”,它有 6 个水平;另一个因素叫做“PTS 浓度”,它也有 6 个水平。它们的水平全面组合总共有 36 种,也就是说,此时共有 36 个试验组。可以考察两个因素之间存在的交互作用项是否具有统计学意义。而像表 2 这样,每种药物的每个浓度构成一个试验组,总共有 $2 \times 6 = 12$ 个试验组。不同药物对评价指标影响的差别明显会大于同一种药物不同浓度之间的差别。

在上述两种情形中,不同水平之间差别大的那个因素可被称为“父辈因素”(例如表 1 中不同棵萝卜或表 2 中药物种类)、不同水平之间差别小的那个因素可被称为“子辈因素”(例如表 1 中同一棵萝卜不同叶片或表 2 中同一种药物不同浓度)。这两种情形都属于“因素之间存在自然属性上的嵌套关系”。

3.2 对表 3 和表 4 两个实际问题的解读

在表 3 中,催化剂和温度是做此项试验研究所必需的两个条件,但它们之间不存在“自然属性”上的嵌套关系。结合专业知识和实际操作的经验,使用的催化剂种类不同,需要加热的温度也是不同的,甚至还可以取不同数目(即水平个数)的具体温度。由表 3 中的试验结果可清楚地看出:不同的催化剂所导致的“转化率”之间的差别较大,而同一种催化剂下不同温度之间的差别较小。

在表 4 中,也存在与表 3 相同的现象,即“非血瘀证”与“血瘀证”结果之间的差别较大,而同一种症候下三种癌症之间的结果差别较小。

在上述两种情形中,不同水平之间差别大的那个因素可被称为“大组因素”(例如表 3 中不同催化剂或表 4 中是否血瘀证)、不同水平之间差别小的那个因素可被称为“小组因素”(例如表 3 中同一种催化剂下的不同温度或表 4 中同一种中医症候下的癌症种类)。这两种情形都属于“因素对主要评价指标的影响之间存在主次关系”。

3.3 嵌套设计具有一个其他多因素设计所缺乏的明显特点

从前述的 4 个例子中不难看出,在嵌套设计中,每个受试对象所接受的处理不是全部试验因素的各水平在“横向”上的全部或部分组合,而是各层因素(例如父辈因素与子辈因素、大组因素与小组因素)的水平在“纵向”上的“连接”,这是不同于其他多因素设计的一个明显特点,借助表 3 中的实际问题看得格外清楚:其第 1 列上的“82%”和“84%”两个数

据,它们都是“在甲催化剂前提条件下选取 70℃ 时”的两批重复试验的结果。在全部试验中,不允许出现“在甲催化剂前提条件下”“选取 55℃ 或 100℃ 的试验条件(因为温度太低或太高,甲催化剂都不能发挥其应有的功能)”,类似地,也不允许出现“在丙催化剂前提条件下”“选取 55℃ 或 70℃ 的试验条件(因为温度太低,丙催化剂会失去其应有的功能)”。

4 运用嵌套设计时应注意的两个事项

4.1 在分析嵌套设计定量资料时无法考察试验因素之间交互作用项的效应大小

由上面所交代的“嵌套设计的明显特点”可知,全部试验数据中没有包含反映因素之间交互作用的信息。换言之,在嵌套设计中,不适合将一个试验因素控制在特定的水平上,来考察另一个试验因素各水平对评价指标影响的差异是否有统计学意义。例如,在表 3 中,若将催化剂控制在“甲催化剂”条件下,只能得出在 70℃、80℃、90℃ 这三种特定温度条件下转化率之间差别如何的结论,而无法得出 8 种不同温度条件下的转化率之间差别如何的结论。同理,分别控制在“乙”或“丙”催化剂条件下,也无法得出 8 种不同温度条件下的转化率之间差别如何的结论。这就好比不同家庭之间情况相差很大,不便于比较;而在同一个家庭内部的同一辈人中,才具有一定的可比性。用“家庭”来举例,实际上就是对前述的“因素水平之间是纵向传递而不是横向组合”

的一种直观而又形象的诠释。

4.2 判定一个定量资料是否来自嵌套设计的意义

其意义不在于“试验安排”,而在于呈现资料 and 选择合适的统计分析方法对资料进行差异性分析。因为此时的“试验安排”完全可由基本常识和专业知识来决定(例如不同的催化剂所需要的温度变化范围是不同的),不涉及复杂的“多因素试验设计知识和技能(例如正交设计、均匀设计、组合设计等^[3-5])”;但在用列表方式呈现试验资料,特别是在试图选择合适的统计分析方法对定量资料进行差异性分析时,若不能准确地判定待分析的资料属于嵌套设计的定量资料,就不能精准地选择“嵌套设计定量资料方差分析模型”去拟合资料。

参考文献

- [1] 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: 85-88.
- [2] 胡良平. 科研设计与统计分析[M]. 北京:军事医学科学出版社, 2012: 237-238.
- [3] 姬振豫. 正交设计的方法与理论[M]. 香港:世界科技出版社, 2001: 98-132.
- [4] 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.
- [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, 2003: 246-278.

(收稿日期:2017-04-05)

(本文编辑:陈霞)



科研方法专题策划人——胡良平教授简介

胡良平,男,1955年8月出生,教授,博士生导师,曾任军事医学科学院研究生部医学统计学教研室主任和生物医学统计学咨询中心主任、国际一般系统论研究会中国分会概率统计系统专业理事会常务理事和北京大学

口腔医学院客座教授;现任世界中医药学会联合会临床科研统计学专业委员会会长、中国生物医学统计学会副会长,《中华医学杂志》等10余种杂志编委和国家食品药品监督管理局评审专家。主编统计学专著45部,参编统计学专著10部;发表第一作者学术论文220余篇,发表合作论文

130余篇,获军队科技成果和省部级科技成果多项;参加并完成三项国家标准的撰写工作;参加三项国家科技重大专项课题研究工作。在从事统计学工作的30年中,为几千名研究生、医学科研人员、临床医生和杂志编辑讲授生物医学统计学,在全国各地作统计学学术报告100余场,举办数十期全国统计学培训班,培养多名统计学专业硕士和博士研究生。近几年来,参加国家级新药和医疗器械项目评审数十项、参加100多项全军重大重点课题的统计学检查工作。归纳并提炼出有利于透过现象看本质的“八性”和“八思维”的统计学思想,独创了逆向统计学教学法和三型理论。擅长于科研课题的研究设计、复杂科研资料的统计分析与SAS实现、各种层次的统计学教学培训和咨询工作。