

科研方法专题

试验设计类型之概述

沈 宁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】 本文目的是全面介绍与试验设计类型有关的概念, 以及学习和掌握试验设计类型知识的意义。通过介绍“研究类型、设计类型、比较类型和资料类型”的概念、临床试验研究中常见试验设计类型的实例并从不同角度来划分设计类型的思路, 读者能够比较全面和深刻地把握设计类型的要领。本文首次将多因素设计类型归纳为“伴随式多因素试验设计”与“统筹式多因素试验设计”两大类, 极大地方便了广大科研工作者结合自己的科研实际, 合理选定并规划出所需要的多因素试验设计类型。

【关键词】 研究类型; 设计类型; 比较类型; 重复测量设计; 析因设计; 回归设计

中图分类号: R195.1

文献标识码: A

doi: 10.11886/j.issn.1007-3256.2017.01.001

Overview of the experimental design type

Shen Ning¹, 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 aim of this article is to comprehensively introduce the concepts involved with the experimental design type, and illustrate the meaning of studying and apprehension of the knowledge relating to the design of experiments and clinical trials. By introducing the four concepts "research type, design type, comparative type, data type", giving the examples of experimental design commonly used in clinical research, and illustrating the categorization of experimental designs from different aspects, the readers may thoroughly and deeply understand the principles of experimental designs. This article, for the first time, establishes the concepts of "concomitant multifactor design" and "overall planning multifactor design" to categorize an experiment involving multiple factors. Such categorization facilitates the process of choosing design type for researchers who deal with multiple factors in an experiment or a clinical trial, based on their individual research specification.

【Keywords】 Research type; Design type; Comparative type; Repeated measurement design; Factorial design; Regression design

1 统计学中经常提及的四种“类型”

1.1 问题的提出

在学习和使用统计学的过程中, 人们常会提及四种“类型”, 即“研究类型、设计类型、比较类型和资料类型”。而且, 人们常对前两种“类型”的概念感到模糊; 对后两种“类型”的概念有时也吃不太准。因此, 常在平时的表达、论文写作, 甚至在讲课中, 出现一些误解或差错。下面, 就用通俗的语言, 将前述提及的四种“类型”作一诠释。

1.2 四种“类型”的概念

1.2.1 研究类型

人们观察或研究事物或现象的一种手法, 被称

为一种“研究类型”。例如某研究者希望在有限的人力、物力、财力和时间的前提下, 获得比较准确的“现阶段全国成年人口中某精神疾病患病率”的基础数据。该研究者会采取调查方法(通常会采取抽样调查方法, 而非普查方法), 而不是试验方法, 当然, 此时也很难采用“文献研究方法”。这实际上说明研究类型可以粗分为“调查研究、试验研究、文献研究”三类。当然, 在实践中, 还会涉及到一种研究类型, 即“理论研究类型”。例如陈景润研究“哥德巴赫猜想”时, 采用的就是“纯数学推导”。

事实上, 在调查研究中, 根据研究目的、要求和具体情况的不同, 人们又可将调查研究划分成以下四种类型, 即“横断面研究、前瞻性队列研究、回顾性病例对照研究和混合研究”。其中, 横断面研究最为常用, 它能以相对少的人力、物力、财力和时间, 获得关于被研究对象的大量信息; 其缺点在于原因

项目基金: 国家高技术研究发展计划课题资助(2015AA020102)

与结果之间的关系并不明确,基于横断面研究所得的信息,很难对过去以及未来作出任何有价值的估计或预测。前瞻性队列研究的优点是所获得的结果比较精准可靠,缺点是质量控制十分困难且费时、费力、费钱。回顾性病例对照研究所获得的结论具有一定的参考价值,但所提取信息的精准程度通常不够高。在某些特殊的场合下,采用队列与病例对照联合起来的混合研究^[1],可以充分发挥队列研究与病例对照研究各自的优点,有效克服各自的缺点。

1.2.2 设计类型^[2]

设计类型通常是指“试验设计类型”。学习和掌握试验设计类型的意义在于:其一,有利于科学、高效地安排试验因素并合理处置区组因素;其二,有利于在试验过程中及时精准地收集试验数据;其三,有利于对所获得的试验数据进行正确的统计分析。

所谓试验设计类型,它是由“外部轮廓”与“内部情况”共同决定的安排因素及其水平组合的一种“架构”。所谓“外部轮廓”,就是在观察或试验中,合理安排试验因素、区组因素及其水平组合的一种框架(常以表格或流程图形式呈现);所谓“内部情况”就是要结合实际问题,明确交代具备以下哪些具体情形:是否具有“区组因素”;是否具有“重复测量因素”;是否对“时段”或“区域”进行了分割;因素在“自然属性”或“对结果变量影响的重要性”方面是否存在“主次关系”;是否可考察因素之间的全部交互效应;全部试验点在空间上是否具有正交性或均匀性;设计矩阵是否具有某些优良性。

将一个具体试验安排的“外部轮廓”与“内部情况”结合起来,使用者就能比较准确地判定其所对应的试验设计类型。

值得注意的是:在文献中,人们习惯于用“组别”代替“因素”“水平”或“因素的水平组合”,常给读者识别资料究竟来自什么样的试验设计类型增添了不小的麻烦。现举例如下:

【例 1】在文献[3]的表 3 中,将 12 个因素以先后顺序放置在一张统计表的最左边,冠以“组别”作为其总称。

剖析:这样表达试验研究情况与资料,给人的错觉是:这是一个“单因素 12 水平设计”问题。将“组别”换成“因素名称”,就不容易造成误解了。

【例 2】在文献[4]的表 3 中,第 1 列呈现出 6 行,其中前 4 行分别代表“是否同卵”与“是否高收缩压”两个 2 水平因素的水平全面组合,而后 2 行则是前 4 行分别取 2 行合并的结果。

剖析:这样表达试验研究情况与资料,不难看出,在此张统计表中,“组别”的含义是令人费解的,因为它既即包含了 2×2 析因设计的结构,还包含了合并同卵和异卵者“有高收缩压”与“无高收缩压”两种情形。这在正确判定设计类型的前提下合理选择统计分析方法设置了多重障碍。

【例 3】在文献[5]的表 1 中,研究者安排了 5 组试验,其“组别”的含义分别代表“正常组”、“氯丙嗪 10 mg/kg(ig)”、“氯丙嗪 20 mg/kg(ig)”、“氯丙嗪 40 mg/kg(ig)”和“氯丙嗪 40 mg/kg(ig) + 新斯的明 0.5 mg/kg(ip)”。

剖析:这样表达试验研究情况与资料,使读者感到十分困惑——“组别”下面的 5 个组究竟代表一个因素的 5 个水平,还是代表两个或多个因素水平的非全面组合?这种安排应该被判定为什么试验设计类型呢?

由基本常识可知,前述提及的 5 组试验,既不代表一个因素有 5 个水平,又不是某两个因素水平的全面组合,也不是有根据地从多个因素水平全面组合中抽取的有代表性的试验点,实际上,那 5 个组应该被视为“多因素非平衡组合试验”的产物,没有资格被称为某种多因素设计类型。

1.2.3 比较类型

所谓试验比较类型,实际上是指如何对试验结果进行差异性分析的四类方法的总称。统计学上最开始提出的“假设检验”方法实际上属于“一般差异性检验”,大约在 20 世纪末,人们又提出了另外三类“假设检验”方法,即所谓的“非劣效性检验”“等效性检验”和“优效性检验”,这三类检验方法在本质上仍属于“差异性检验”的范畴,它们与“一般差异性检验”的区别在于:前者中涉及“非零的界值”,而后者中的“界值”被设定为 0。现在常把这四类假设检验统一称为“试验比较类型”。

“试验比较类型”与前述的试验设计类型有什么关系呢?一般差异性检验似乎与开始的试验设计没有太大关联性,其实不然。事实上,试验设计三要素是否把握得合适、试验设计四原则是否遵守得严格、选择什么样的试验设计类型,这些(例如如何估计出合适的样本含量)都与将采取的一般差异性检验有直接关系;尤其是在“非劣效性检验”“等效性检验”和“优效性检验”中,涉及到“在临床上或实际工作中有意义的界值”。这些(非劣效性、优效性)或这个(等效性)界值需要由多位同行专家依据专业知识和已有的经验在试验设计阶段共同商量确定,而不是等到试

验结束后根据实际试验结果再来确定。

1.2.4 资料类型

所谓资料类型,是指在所获得的资料中,无论是原因还是结果变量及其取值具有何种“性质”。例如人们常收集的原因变量有“性别、血型、有无某种疾病的家族史等”与“年龄、血压等”,前者为“定性原因变量”,后者为“定量原因变量”。人们常收集的结果变量有“是否患高血压、是否患糖尿病、治疗结果为‘痊愈、好转、无效、死亡’中的哪一档”与“经过治疗后的血糖值、血小板数值等”,前者为“定性结果变量”,后者为“定量结果变量”。在“定性变量”中,若某个定性变量的若干个具体取值之间表现为一定的等级关系或存在虽不精确但可区分大小的数量关系,则应将此类定性变量再命名为“有序或等级变量”,连同变量名及其具体的取值被称为“有序或等级资料”。

值得注意的是:迄今为止,仍有一些论文作者称“定性变量或资料”为“计数变量或资料”,这是不妥的!所谓“计数变量或资料”是指变量在每个个体身上的取值是数值(多数情况下,为0或正整数)而不是“属性”(男、女性别;ABO血型系统)或“等级”(优、良、中、差等评价等级)或“代号”(如用“0”代表未患某病、用“1”代表患了该病)。在医学研究中,真正的“计数变量或资料”通常有:某地某年癫痫患者经过治疗后,在未来一年内每位患者“复发的次数”;抽查某地正常成年人,测得每位受试者每分钟的脉搏次数;某地区在某年内各次出现食物中毒的人数。

1.3 四种“类型”中的难点

在前述提及的四种类型中,最难掌握的是“设计类型”。这是由于不仅设计类型的种类繁多,而且有些设计类型十分相似,在实际问题中难以明确区分。最难之处在于一些实际问题中的设计类型是“可变的”。也就是说,根据设计者和数据分析者对实际问题把握的全面程度和要求高低、可能涉及的因素多少、因素之间的相互关系以及对相关专业知识的掌握程度,会采用动态的眼光去探索每个具体问题中最合适的设计类型。正因如此,在后面的文章中,笔者将由浅入深,逐一展开,目的是使读者学以致用。

2 临床试验研究中常见试验设计类型举例

2.1 单因素设计举例

在文献[6]中,研究者最关注的试验因素是受

试者的年龄,分为“青中年”与“老年”两个水平。若只关注两个“年龄”组在各项评价指标上的差异情况,则此研究可被视为“单因素两水平设计”问题。

2.2 多因素设计举例

事实上,在文献[6]中,研究者除了关注“年龄”这个试验因素外,还考虑了“性别”“民族”“有无配偶”“受教育程度”和“职业”等“重要非试验因素”。若将“年龄”和其他重要非试验因素一并纳入考虑,则此研究也应被视为某种多因素试验设计类型。类似的设计类型在本刊的学术论文中经常被使用,如文献[7-8]。

在文献[9]中,研究者最关注的试验因素有两个,其一,在患者使用了艾司西酞普兰基础上,是否对他们进行干预,即“干预与否”是第1个试验因素;其二,观测疗效的时间(分为“治疗前”“治疗1月末”“治疗2月末”三个时间点)是第2个试验因素。这项研究所对应的设计类型为“带有一个协变量(即治疗前)且具有一个重复测量因素的两因素设计”。类似的设计类型在本刊的学术论文中也经常被使用,如文献[10]。

3 试验设计类型种类的划分

3.1 基于不同角度来划分设计类型的种类

设计类型的名称很多,例如单因素与多因素设计,显然,这是从一个试验项目所涉及的因素个数角度来划分的;还有“非回归设计与回归设计”的说法,这是从分析试验数据采用的是“回归分析”还是“非回归分析”的角度来划分的。类似地,还可罗列出许多种划分设计类型种类的方法。

值得一提的是,“单因素与多因素设计”是没有多大实用价值的称谓,因为在“单因素设计”中还包含着多个具体的“单因素设计类型”;同样,在“多因素设计”中,也包含着几十个具体的“多因素设计类型”。诸如“ $N(\geq 2)$ 因素设计(即两因素设计、四因素设计、七因素设计等)”的说法或名称,对于具体的试验安排或对已获得的试验资料进行差异性分析都是没有任何帮助的。因为在各种统计分析中,通常只是在进行差异性分析时,人们才会特别关注资料来自什么试验设计类型。事实上,在进行直线相关与回归分析时,也应先辨别其所对应的设计类型是否为单组设计。若不是,很可能会出现误用现象。当然,“非回归设计与回归设计”也是没有多大实用价值的划分结果,因为“非回归设计”与“回归设计”的每一大类中,都包含了很多具体的“设计类型”。

3.2 基于构造原理角度来划分设计类型的种类

对于某些简单的试验研究场合,人们只需要基于“功能”角度来划分设计类型的种类即可,例如单组设计、单因素 $K(\geq 2)$ 水平设计、配对设计;而对于某些较复杂或很复杂的试验研究场合,人们则需要基于“构造原理”角度来划分设计类型的种类,这样做的结果不仅言简意赅,而且可以形象生动。例如随机区组设计、平衡不完全随机区组设计、交叉设计、拉丁方设计、嵌套设计、裂区设计、重复测量设计、析因设计、分式析因设计、正交设计、均匀设计、稳健设计、A 最优设计、D 最优设计和 G 最优设计等。

3.3 两类模式的多因素试验设计

3.3.1 携带式多因素试验设计

在临床试验研究中,实际工作者经常采用的多因素试验设计类型可以被概括为“携带式多因素试验设计”。在这类多因素试验设计中,研究者通常以一个单因素两水平(如试验组与对照组)设计或单因素多水平(如单相抑郁、双向 I 型、双向 II 型^[11])设计为“骨架”,顺便记录下伴随的其他影响因素的取值^[6-8],故称其为“伴随式多因素试验设计”。

3.3.2 统筹式多因素试验设计

在以动物或样品为受试对象的试验研究中,实际工作者经常采用的多因素试验设计类型可以被概括为“统筹式多因素试验设计”。在这类多因素试验设计中,研究者通常以全部拟考察的试验因素全部或部分水平组合为“骨架”,必要时可纳入少量来自受试对象的属性因素,如窝别、性别、体重分组、年龄分组等,它们常被简称为“区组因素”。这类多因素试验设计的具体设计名称常常为前面提及的“基于构造原理角度来划分”的那些设计类型。后续系列文章中将要详细介绍的多因素试验设计类型就属于此类。

4 如何正确选定一个试验研究问题中的试验设计类型

4.1 何时需要选定试验设计类型

由基本常识可知,在整个试验研究过程中,在两个时间段上需要选定试验设计类型,其一,在正式开展试验研究之前,就应根据试验研究目的和研究者

所具备的前提条件,选定合适的试验设计类型;其二,在完成了试验且收集到全部试验数据之后,拟对试验数据进行各种统计分析之前,首要任务是需要多次选定试验设计类型。因为拟采取的统计分析方法不同,将需要从全分析集中提取不同的数据子集,因而它们也就对应着不同的设计类型。

4.2 正式开展试验研究之前确定试验设计类型的一般方法

人们习惯于将复杂的多因素试验研究问题尽可能简单化,例如人们最热衷于将全部受试对象分入试验组与对照组中去。然后记录每位受试对象的各种“基线资料”,通常包括性别、血型、职业、受教育程度、生活地点与环境、生活习惯与方式、年龄、血压、各种血脂指标、生理指标、是否患有某些疾病、与某些疾病有关的家族史等。人们将前述的资料视为来自“单因素两水平(即试验与对照)设计类型”,而把前述提及的“基线资料”视为“协变量”,读者从文献[6-8]中可以非常明显地看出前述提及的那些“特征”。由此可知,很多实际工作者就是通过所谓的“协变量”这个“杀手锏”来达到“化繁为简”的目的。换句话说,实际工作者基本上都习惯于采用“携带式多因素设计”取代“统筹式多因素设计”。

事实上,基于上述“化繁为简”做法得出的结论,通常是存在问题的,因为它隐含着下面两个很强但几乎不太可能满足的假定:其一,试验组与对照组(简称为试验因素)中的受试对象在所有“基线资料(包括未检测的其他各种因素或指标)”上是均衡一致的;其二,试验因素与所有“基线资料”之间的交互作用是可以忽略不计的。

正因如此,在试验之前,就应结合基本常识和专业常识,从“基线资料”中提取出不可忽视的重要非试验因素,将其与试验因素,即试验组与对照组一视同仁,由它们联合起来构成一个特定的多因素试验设计类型,并将未被列入“重要非试验因素”的其他“基线资料”视为“协变量”。这样做,才是明智之举。

4.3 在对试验数据进行统计分析之前确定试验设计类型的一般方法

在前面“3.2”中所讨论的两种情形(即单因素两水平设计与某种特定的多因素设计)下且在对所获得的试验数据进行分析之前,又一次面临着确定试验设计类型的问题,因为在那两种情形下,都存在较多的“协变量”。于是,如何正确处置“协变量”,便成为合理选择统计分析方法的“关键问题”。

事实上,应该将“协变量”分解成两种类型的

“因素或变量”。其一,为定性的;其二,为定量的。在差异性分析中,“协方差分析”中的“协”实际上指的是“定量协变量(因素)”,此类“协变量”是不会改变原先确定的设计类型的;问题是:在差异性分析中如何引入“定性协变量(因素)”呢?它或它们的加入,会使原先的设计类型变得更加复杂。在实践中,人们几乎一律采取“回归分析”的手段,将“定量与定性协变量(因素)”考虑在统计分析之中。然而,这实际上是一种基于“黑箱原理”理念指导下的“馄饨”做法。因为此时由全部自变量组成的“子集”可能是来自多个未知“总体”的一个样本,而且,自变量中既有服从各种分布的定量变量,也有三种类型(即二值的、多值名义的和多值有序的)的定性变量。能够很好驾驭前述两种情形资料的回归分析模型和求解方法,迄今为止,几乎还未被发现。当然也就不太可能在现今的国际通用统计软件包中找到相应的、合理的解决方案。

综上所述,在多因素差异性分析中,正确判定试验设计类型是合理选择统计分析方法的关键环节;而借助“回归分析”来替代“多因素差异性分析”应该十分谨慎,其重点应放在“合理处置自变量”上。

参考文献

- [1] 刘忠厚. 骨质疏松学[M]. 北京: 科学出版社, 1998: 515-526.
[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: 129-276.
[3] 钟丽, 宋丽娟, 张涛. 泸州市在校大学生饮酒现状及危险因素[J]. 四川精神卫生, 2016, 29(4): 360-366.
[4] 李吉祝, 王金东, 徐春生, 等. 青岛市中老年双生子轻度认知障碍患病率及其与血压的关系[J]. 四川精神卫生, 2016, 29(5): 453-456.
[5] 张家瑞, 任丽娜, 赵均铭, 等. 氯丙嗪致大鼠便秘与炎性细胞因子表达的对照研究[J]. 四川精神卫生, 2016, 28(2): 116-118.
[6] 佟钙玉, 戴立磊, 邹绍红. 新疆地区有自杀意念的青年与老年抑郁症患者临床症状比较[J]. 四川精神卫生, 2016, 29(5): 414-417.
[7] 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.
[8] 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.
[9] 公维谦, 陈玉社. 艾司西酞普兰联合康复训练治疗脑卒中后抑郁的效果[J]. 四川精神卫生, 2016, 29(5): 410-413.
[10] 刘成文, 程从彬, 王慢利. 帕罗西汀联合认知行为疗法治疗强迫症的临床对照研究[J]. 四川精神卫生, 2016, 29(3): 234-237.
[11] 林康广, 卢睿, 陈玲玉, 等. 单双相抑郁患者的情感气质特征及其与抗抑郁治疗反应的关系[J]. 四川精神卫生, 2016, 29(3): 211-215.

(收稿日期: 2017-02-11)

(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。