中国卫生统计1990年第7卷第5期

具有重复测量的多因素设计类型及方差分析

军事医学科学院科技部统计教研室　胡良平

提　要　本文以两试验因素的三种不同安排为例，介绍了具有重复测量的多因素设计类型及其资料的方差分析方法。笔者认为：在析因设计中有意识地加入区组因素，可以提高对试验误差的估计精度。

关键词　重复试验　重复测量　区组因素　析因设计

两个名词概念的约定

“重复”和“因素”是方差分析中两个重要概念，“重复”有三层含意，其一，试验条件不变，但K次试验是在K个受试对象（或样品）上实施的，收集到K个数据；其二，试验条件不变，但K次试验是在同一个受试对象（或样品）上重复进行的，收集到K个数据；其三，试验条件部分或全部作有规律地变动，但受试对象（或样品）不变，即从同一个受试对象（或样品）上收集到K个数据。这是不同意义下的“重复”，为了区别起见，不妨把前两种称为“重复试验”，把第三种称为“重复测量”。“因素”常有两层含意，其一，是指试验者要重点考查的试验因素（简称因素）；其二，是指试验者为更精确地估计试验误差，而特意安排的重要非试验因素，称为区组因素。

具有重复测量的双因素设计类型

（一）不同的设计类型及其数据的方差分析结果

1．设计类型

设有三组试验数据，如表1、2、3所示，首先让我们来看看它们在设计上有何异同点。

表1的设计：A、B两因素构成四个不同的试验条件，16个受试对象被完全随机地均分到四个处理组中去。并且，任何一个受试对象只接受四种处理中的一种，这是具有4次重复

表1　因素A，B对大鼠肉瘤生长的影响

重复试验号	测量指标				
	*A_1B_1	A_1B_2	A_2B_1	A_2B_2	合计
1	3.0	8.0	5.0	10.0	26.0
2	1.1	6.5	7.2	12.0	26.8
3	4.2	3.2	8.0	9.5	24.9
4	3.2	5.5	9.0	7.0	24.7
合计	11.5	23.2	29.2	38.5	102.4

*代表A，B水平组合，A代表药物，B代表部位

表2　因素A，B对大鼠肉瘤生长的影响

试验因素A与区组因素S	测量指标		
因素B：	B_1	B_2	合计
A_1：			
S_1	3.0	8.0	11.0
S_2	1.1	6.5	7.6
S_3	4.2	3.2	7.4
S_4	3.2	5.5	8.7
A_2：			
S_5	5.0	10.0	15.0
S_6	7.2	12.0	19.2
S_7	8.0	9.5	17.5
S_8	9.0	7.0	16.0
合计	40.7	61.7	102.4

*A代表药物，B代表部位

表3　因素A，B对大鼠肉瘤生长的影响

大鼠编号	测量指标				
	*A_1B_1	A_1B_2	A_2B_1	A_2B_2	合计
S_1	3.0	8.0	5.0	10.0	26.0
S_2	1.1	6.5	7.2	12.0	26.8
S_3	4.2	3.2	8.0	9.5	24.9
S_4	3.2	5.5	9.0	7.0	24.7
合计	11.5	23.2	29.2	38.5	102.4

*代表A，B水平组合，A代表药物，B代表部位。

试验的2×2析因设计。

表 2 的设计：A、B 两因素也构成四个不同的试验条件，但 8 个受试对象被完全随机地均分到按 A_1、A_2 划分的两组中去，并表现为“区组”的形式，任何一个受试对象先后接受两种处理，这两种处理中，要么只含 A_1 或者要么只含 A_2，但分别含有 B_1 和 B_2。此设计称为具有一个重复测量的两因素设计[1,2]。这里“具有一个重复测量”是就受试对象与两因素的关系而言的，如前四个受试对象，均在 A 因素取定 A_1 水平的条件下，B 因素依次从 B_1 变到 B_2，分别从同一个受试对象身上测得了两个数据，后四个受试对象也有类似的情况。

表 3 的设计：A、B 两因素还是构成四个不同的试验条件，但只有 4 个受试对象参加试验，并表现为“区组”的形式。任何一个受试对象都依次接受四种不同的处理。此设计称为具有两个重复测量的两因素设计[1,2]，即在两因素各水平同时变化的条件下，从同一个受试对象身上通过“重复测量”获得了 4 个数据。

表 1 至表 3 在设计上的异同，试验性质、

1）　表 1 设计资料方差分析的结果

变异来源	离均差平方和	自由度	均方	F值	P值
因素A	68.0625	1	68.0625	21.17303	<0.01
因素B	27.5625	1	27.5625	8.574204	0.013
$A\times B$	0.3599854	1	0.3599854	0.111985	0.744
误差	38.57501	12	3.214584		

2）　表 2 设计资料方差分析的结果

变异来源	离均差平方和	自由度	均方	F值	P值
因素A	68.0626	1	68.0625	44.74116	<0.01
误差1	9.127502	6	1.52125		
因素B	27.5625	1	27.5625	5.615926	0.056
$A\times B$	0.3600464	1	0.3600464	0.0733603	0.796
误差2	29.44751	6	4.907919		

3）　表 3 设计资料方差分析的结果

变异来源	离均差平方和	自由度	均方	F值	P值
因素A	68.06256	1	68.06256	24.3011	0.016
误差1A	8.402405	3	2.800802		
因素B	27.56256	1	27.56256	3.513779	0.158
误差1B	23.53241	3	7.844137		
$A\times B$	0.3599243	1	0.3599243	0.1825471	0.698
误差2	5.915039	3	1.97168		

目的、受试对象、试验因素及其个数、数据个数诸方面相同，受试对象的个数、重复的含义等方面不同。

2. 方差分析的结果

（二）设计类型不同，误差的分解原则有别

从表 1 的设计可看出，因素 A 与 B 的地位是平等的。因此，衡量因素 A、B 和交互作用 $A\times B$ 重要与否的尺度应该是统一的，即由

$$SS_e = SS_{总} - SS_A - SS_B - SS_{A\times B}$$

$$\nu_e = \nu_{总} - \nu_A - \nu_B - \nu_{A\times B}$$

可算得误差的均方 $MS_e = SS_e/\nu_e$，用它作为度量 A、B 和 $A\times B$ 重要与否的公共尺度。

从表 2 的设计可看出，因素 A 与 B 的地位是不完全平等的，受试者先按 A 因素的两个水平分为两个大组，而 B 因素的两个水平先后作用于两个大组中每一个受试者，因此，A 相当于是一级因素，B 相当于是二级因素。由于试验因素 A 的作用包含在区组因素 S 的作用之内（以下为避免符号混淆，用 G 代替 S），于是，由

$$SS_{e1} = SS_G - SS_A,\quad \nu_{e1} = \nu_G - \nu_A$$

可算得一级误差的均方 $MS_{e1} = SS_{e1}/\nu_{e1}$，用它作为度量因素 A 重要与否的尺度。而由

$$SS_{e2} = SS_{总} - SS_A - SS_B - SS_{A\times B} - SS_{e1}$$

$$\nu_{e2} = \nu_{总} - \nu_A - \nu_B - \nu_{A\times B} - \nu_{e1}$$

可算得二级误差的均方 $MS_{e2} = SS_{e2}/\nu_{e2}$，用它作为度量因素 B 和交互作用 $A\times B$ 重要与否的尺度。

从表 3 的设计来看，仅比表 1 的设计多了一个区组因素，因素 A、B 的地位基本上是平等的，但正是由于多了一个“区组因素”，使得用来度量因素 A、B 重要与否的尺度不同了，即可以用试验因素 A 与区组因素 S 的交互作用的均方来作出 A 对试验结果影响大小的判断，同样，可以用试验因素 B 与区组因素 S 的交互作用的均方来作出 B 对试验结果影响大小的判断，这就相当于有了两个地位平等的一级误差。然后，从总的离差平方和及自由度中减去 A、B、$A\times B$ 以及两个一级误差的离差平方和及自由度，可得到一个二级误差的离差平方和

及自由度，用其均方作为度量 $A\times B$ 重要与否的尺度。与表 1 相比，表 3 的设计使得对误差的估计更加精细、更加精确。

（三）方差分析中的计算公式

1．重复测量与一个因素的各水平有关

这种设计类型如表 2 所示。由表 2 可直接算得 $SS_{总}$、SS_G、SS_A、SS_B、$SS_{A\times B}$，进而可以求出一级和二级误差的离差平方和 SS_{e1} 及 SS_{e2}。

离差平方和的计算：

校正数 $C=(102.4)^2/16=655.36$

$SS_{总}=(3.0^2+1.1^2+\cdots+7.0^2)-C==134.56$

$SS_G=(11.0^2+7.6^2+\cdots+16.0^2)/2--C=77.19$

$SS_A=[(11.0+7.6+\cdots+8.7)^2++(15.0+19.2+\cdots+16.0)^2]/8-C=[34.7^2+67.7^2]/8-C=68.0625$

$SS_B=(40.7^2+61.7^2)/8-C=27.5625$

$SS_{A\times B}=[(3.0+1.1+\cdots+3.2)^2++(8.0+6.5+\cdots+5.5)^2+\cdots++(10.0+12.0+\cdots+7.0)^2]/4-C-SS_A-SS_B=0.36$

$SS_{e1}=SS_G-SS_A=9.1275$

$SS_{e2}=SS_{总}-SS_A-SS_B-SS_{A\times B}-SS_{e1}==29.4475$

自由度的计算：

$\nu_{总}=16-1=15$，$\nu_A=2-1=1$，$\nu_B=2--1=1$，$\nu_G=8-1=7$，$\nu_{A\times B}=\nu_A\times\nu_B=1$，$\nu_{e1}=\nu_G-\nu_A=6$，$\nu_{e2}=\nu_{总}-\nu_A-\nu_B-\nu_{A\times B}-\nu_{e1}=6$

F值的计算：

$MS_A=SS_A/\nu_A$，$MS_B=SS_B/\nu_B$，$MS_{A\times B}=SS_{A\times B}/\nu_{A\times B}$，$MS_{e1}=SS_{e1}/\nu_{e1}$，$MS_{e2}=SS_{e2}/\nu_{e2}$

$F_A=MS_A/MS_{e1}=44.741$，$F_B=MS_B/MS_{e2}=5.616$，$F_{A\times B}=MS_{A\times B}/MS_{e2}=0.073$

2．重复测量与两个因素的各种水平组合有关

这种设计类型如表 3 所示。由表 3 可直接算得 $SS_{总}$、SS_G、SS_A、SS_B、$SS_{A\times B}$，进而可以求出两个一级误差、一个二级误差的离差平方和 SS_{e1A}、SS_{e1B} 和 SS_{e2}。

离差平方和的计算：

校正数 $C=(102.4)^2/16=655.36$

$SS_{总}=(3.0^2+8.0^2+\cdots+7.0^2)-C=134.56$

$SS_G=(26.0^2+26.8^2+24.9^2+24.7^2)/4-C=0.725$

$SS_A=[(11.5+23.2)^2+(29.2+38.5)^2]/8-C=68.0625$

$SS_B=[(11.5+29.2)^2+(23.2+38.5)^2]/8-C=27.5625$

$SS_{A\times B}=(11.5^2+23.2^2+29.2^2+38.5^2)/4-C-SS_A-SS_B=0.36$

$SS_{e1A}=[(3.0+8.0)^2+(1.1+6.5)^2+\cdots+(9.0+7.0)^2]/2-C-SS_G-SS_A=8.4025$

$SS_{e1B}=[(3.0+5.0)^2+(1.1+7.2)^2+\cdots+(5.5+7.0)^2]/2-C-SS_G-SS_B=23.5325$

$SS_{e2}=SS_{总}-SS_A-SS_B-SS_{A\times B}-SS_G-SS_{e1A}-SS_{e1B}=5.915$

自由度的计算：

$\nu_{总}=16-1=15$，$\nu_A=2-1=1$，$\nu_B=2-1=1$，$\nu_{A\times B}=\nu_A\times\nu_B=1$，$\nu_G=4-1=3$，$\nu_{e1A}=(4-1)\times(2-1)=3$，$\nu_{e1B}=(4-1)\times(2-1)=3$，$\nu_{e2}=\nu_{总}-\nu_A-\nu_B-\nu_{A\times B}-\nu_G-\nu_{e1A}-\nu_{e1B}=3$

F值的计算：

$MS_A=SS_A/\nu_A$，$MS_B=SS_B/\nu_B$，$MS_{A\times B}=SS_{A\times B}/\nu_{A\times B}$，$MS_{e1A}=SS_{e1A}/\nu_{e1A}$，$MS_{e1B}=SS_{e1B}/\nu_{e1B}$，$MS_{e2}=SS_{e2}/\nu_{e2}$

$F_A=MS_A/MS_{e1A}=24.301$，$F_B=MS_B/MS_{e1B}=3.514$，$F_{A\times B}=MS_{A\times B}/$

* 这里是用计算器计算的结果，与前面计算机计算的结果略有出入，所用程序为[2]。

$MS_{e2} = 0.183$

小　结

本文分析了相似但又有区别的三种设计类型的异同点，通过例子介绍了“具有重复测量的两因素设计类型”的方差分析方法。形如表2、3的设计可以推广到多因素时的情形，至少可以在含有三个试验因素时实现[1,2]。笔者认为：就“重复测量”而言，本质上是配对设计的扩展，在很多具体的试验中是行不通的。若将这些设计类型中的区组因素的各水平由现在的一个一个受试者换成条件相近的一组一组受试者，可以使其实用性大大增加。由本文例子的分析结果可知：在析因设计中有意识地加入“区组因素”后，可以对总误差进行更精细地分解，从而可以达到更精确地估计试验误差之目的。

The Design Types of Multifactor with Repeated Measurements and Analysis of Variance of the Data *Hu Liangping, Military Medical Academy of Sciences, Beijing*

This paper takes the three different designs contained two experimental factors as an example in order to introduce the design types of multifactor with repeated measurements and the analysis of variance of the data. The author believes that putting the block factor into the factorial design intentionally will benefit the estimated precision of experimental error.

Key words Repeated experiments Repeated measurements Block factor Factorial design

参考文献

1. General Editor P. R. Krishnaith. Handbook of statistics, Analysis of Variance. V.1. First Edition. Amsterdam. New York. Oxford. North-Holland Publishing Company. 1980:41—63.
2. Northwest Analytical, Inc. The NWASTATPAK. Version 2.1. Portland, Oregon. 1982: 12.1.1—12.8.3.

·简讯·

全国卫生统计工作会议在京召开

首次全国卫生统计工作会议于1990年5月25日～29日在北京召开。各省自治区卫生厅计财处长及主管统计工作的同志参加了会议。卫生部部长陈敏章、副部长顾英奇、国家统计局副局长孙兢新、国家中医药管理局副局长诸国本及卫生部各司（局）长到会，并作了重要讲话。

会议认真总结了全国近几年卫生统计工作成绩和经验，讨论制定了1990～2000年《全国卫生统计工作发展规划设想》，听取了先进单位的典型经验介绍，表彰了一批卫生统计工作和计算机应用工作的先进集体和个人，成立了专家咨询小组。

陈敏章在讲话中充分肯定了统计信息在卫生事业管理和决策中的地位和作用。他说，各级卫生部门要切实加强领导，把统计信息工作列入议事日程，并自觉运用统计信息进行管理和决策。同时要改革和完善卫生统计体制，加快全国卫生统计信息网络系统的建设。今后卫生部所有的调查和统计资料的公布都归口于卫生统计信息中心管理。

顾英奇在讲话中指出，卫生统计工作的主要任务是搞好统计服务和统计监督。要积极探索对统计数据实行全面质量管理的科学方法，建立统计数据质量控制、审核、考核等项制度，加强同有关部门的联系和合作，及时反馈各方面的信息，提高统计服务的水平。

（孙高）