论 著 文章编号:0253-3626(2008)12-1509-04

析因设计资料单独效应分析的SAS实现

张旭东，彭 斌，王润华
（重庆医科大学公共卫生学院卫生统计教研室，重庆 400016）

【摘 要】目的：探讨析因设计资料单独效应分析的SAS实现方法。方法：将析因设计的总的误差均方作为标准误，通过构造参数的对比系数矩阵，利用SAS的GLM过程中Contrast语句编程实现单独效应分析，并结合实例给出SAS程序来实现三因素析因设计资料的统计分析。结果：给出了设计参数对比系数矩阵的方法。结论：通过构造设计矩阵，在SAS中编写Contrast语句来实现单独效应分析，是一种方便可行的方法。

【关键词】析因设计；单独效应；对比系数矩阵
【中图法分类号】O212.6 【文献标识码】A 【收稿日期】2008-03-06

Analysis of simple effect on factorial design with SAS

ZHANG Xu-dong, et al
(*Department of Health Statistics, School of Public Health, Chongqing Medical University*)

【Abstract】*Objective*: To explore the approach to analyzing simple effect on factorial design with SAS. *Methods*: The mean square of error coming from factorial design was considered as standard error(SE), and simple effect analysis under Contrast statement in GLM was conducted by designing the contrast coefficient matrix of the parameter, and SAS program was realized based on an example of 3-way factorial design. *Results*: he procedures of designing the contrast coefficient matrix of the parameter were achieved. *Conclusion*: It is convenient and feasible to conduct simple effect analysis under Contrast statement by constructing design matrix.

【Key words】Factorial design; Simple effect; Contrast coefficient matrix

在对析因设计资料进行分析时，我们通常先对各因素的主效应和交互效应进行分析，如果因素间存在交互作用时，只对各因素主效应进行整体检验的意义不大，此时需要对各因素水平间的单独效应做进一步分析[1]。在一般的参考书和文献中以两因素析因设计资料的讨论居多，而有关三因素的讨论较少，并且涉及对三因素析因设计资料的单独效应分析的讨论和实例更是鲜见，而在日常中此实验设计类型也常用到，因此本文将讨论通过构造参数的对比系数矩阵，利用SAS编程来实现其单独效应分析，并对三因素的析因设计资料作实例分析。

单独效应分析一般是在其他因素的水平固定时，比较某一因素不同水平的差别。分析单独效应时，应当利用析因分析得到的总的误差均方作为标准误，这样不仅使各次比较具有统一的尺度，而且使计算统计量的分母有较大的自由度，并充分利用了整体数据信息，使结论更加可靠[2]。析因分析及其基于总的误差均方的单独效应分析可以通过SAS中GLM过程实现。下面着重探讨对比系数矩阵的设计，并在SAS中编写Contrast语句来实现析因设计资料的单独效应分析。

作者介绍：张旭东（1983- ），男，硕士，
研究方向：多元统计方法在医学中的应用。
通讯作者：王润华，男，教授，E-mail:wrhzyp@163.com。

1 原理和方法

1.1 三因素方差分析模型

假设处理因素A、B、C分别有r、s、t个水平，总共有$r\times s\times t$种组合，每种组合下重复例数为，则$r\times s\times t$以析因设计资料的方差分析模型[3]可构造为：$X_{abck}=\mu+\alpha_a+\beta_b+\gamma_c+(\alpha\beta)_{ab}+(\alpha\gamma)_{ac}+(\beta\gamma)_{bc}+(\alpha\beta\gamma)_{abc}+\varepsilon_{abck}$。模型中$\alpha=1,2,\cdots,r;b=1,2,\cdots,t;k=1,2,\cdots,n;$。$X_{abck}$表示第$abc$组中第$k$个观察值；$\mu$为常数项，表示总体平均值；$\alpha_a$、$\beta_b$和$\gamma_c$表示处理因素$A$、$B$、$C$分别取$a$、$b$、$c$水平的主效应，$(\alpha\beta)_{ab}$、$(\alpha\gamma)_{ac}$、$(\beta\gamma)_{bc}$$(\alpha\beta\gamma)_{abc}$分别代表因素A、B、C间一级与二级交互作用的效应，并满足以下条件：

$$\sum_{a=1}^{r}\alpha_a=0,\ \sum_{b=1}^{s}\beta_b=0,\ \sum_{c=1}^{t}\gamma_c=0,\ \sum_{a=1}^{r}(\alpha\beta)_{ab}=0,\ \sum_{b=1}^{s}(\alpha\beta)_{ab}=0,\ \sum_{a=1}^{r}$$

$(\alpha\gamma)_{ac}=0$, $\sum_{c=1}^{t}(\alpha\gamma)_{ac}=0$, $\sum_{b=1}^{s}(\beta\gamma)_{bc}=0$, $\sum_{c=1}^{t}(\beta\gamma)_{ac}=0$, $\sum_{a=1}^{r}(\alpha\beta\gamma)_{abc}=0$, $\sum_{b=1}^{s}(\alpha\beta\gamma)_{abc}=0$, $\sum_{c=1}^{t}(\alpha\beta\gamma)_{abc}=0$; ε_{abck}代表实验误差，是随机效应，$\varepsilon_{abck}\sim N(0,\sigma^2)$，且各$\varepsilon_{abck}$独立。

两因素析因设计的方差分析模型可以从三因素的模型简化而来，以$r\times s$析因设计为例，其方差分析模型可构造为：$X_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ijk}$；并满足：$\sum_{i=1}^{r}\alpha_i=0$, $\sum_{j=1}^{s}\beta_j=0$, $\sum_{i=1}^{t}\gamma_{ij}=0$, $\sum_{j=1}^{s}\gamma_{ij}=0$；$\varepsilon_{ijk}$的意义同三因素模型。

1.2 对比系数矩阵(矩阵)

在SAS的一般线性模型中，Contrast语句提供了一种获得一般假设检验的技巧，即通过规定矩阵来构造线性假设检验$L\beta=0$[4]，其中L为L矩阵，也称为对比系数矩阵，β为一般线性模型中的参数向量。进行单独效应分析时，可以对因素各水平的参数设计对比系数矩阵，并利用SAS中Contrast语句来实现。以最简单的两因素(A和B)析因设计为例，其水平数分别为3和4，现要比较A因素第一水平和第三水平间效应的差异，则可以构造关于A因素的矩阵和参数向量：$L=(1,0,-1)$，$\beta=(A_1,A_2,A_3)'$，根据线性模型假设$L\beta=0$，可以得到表达式：$1\times A_1+0\times A_2-1\times A_3=0$，化简后即等价于$A_1=A_3$，这正是我们希望进行比较的假设检验。在SAS中Contrast语句的一般形式为：

contrast 'label' effect value

其中引号内的label用以标注识别对比，effect是识别在Model语句中出现的效应，value表示和效应(effect)相关的矩阵向量的元素。上例中A因素的第一和第三水平的比较在SAS中可写为：

class *A B*; model x=*A B*; contrast 'A_1 *vs* A_3' A 1 0 -1;

当A与B存在交互作用时，要比较A因素第一水平和第三水平间效应的差异，就需要固定B因素的某一个水平来进行分析。以固定B因素的第二个水平为例，可以通过构造交互效应表来设计矩阵向量。因素在设计表中的排列位置与它们在class语句中的顺序有关，在此class语句中A因素在B因素之前，所以在设置表格时将A的水平放在行，B的水平放在列；同时将A因素中需要对比的两个水平(A_1和A_3)的系数分别设为1和-1，B因素中被固定的水平(B_2)的系数设为1，其余水平的系数均设置为0(见图1的横纵标目)。

然后将行和列的系数进行交叉相乘，将所得的值填进相应的交叉单元格，这样得到中间大方格的矩阵即为A与B的交互作用矩阵(见图1的中间大方格)，Contrast语句中就可以根据这个矩阵从上到下依次写出每行的向量，从而得到交互效应项的L矩阵，即：A*B 0 1 0 0 0 0 0 0 0 -1 0 0 。

再将交互作用矩阵的各行和各列分别求和，把所得的值填进表格相应行和列的单元格(见图1的右侧行合计和下侧列合计)，得到的向量在Contrast语句中就分别代表因素A的主效应(最右边一列)和因素B的主效应(最下边一行)，因此就可以根据主效应向量构造Contrast语句，即：A 1 0 -1；由于因素B的系数全部为0，所以在Contrast语句中可以不用写出来。

余下的右下角最后一个单元格为相应的总行合计或总列合计(见图1的右下单元格)，它代表截距项的系数，由于系数为0，所以在Contrast语句中也不需要进行指定。

通过以上的设计矩阵，就可以在SAS的GLM过程中构造

			B				
			B_1	B_2	B_3	B_4	
			0	1	0	0	
	A_1	1	0	1	0	0	1
A	A_2	0	0	0	0	0	0
	A_3	-1	0	-1	0	0	-1
			0	0	0	0	0

图 1 对比系数矩阵

Contrast语句来实现B_2水平下A_1和A_3水平间效应的差异比较，即：

contrast 'A_1 *vs* A_3 *in* B_2' A 1 0 -1 A*B 0 1 0 0 0 0 0 0 0 -1 0 0;

三因素析因设计资料的单独效应分析，其构造设计矩阵的方法与上述一致，以A、B、C三因素($2\times2\times3$)为例，当各因素间均存在交互效应时，以固定A因素的第二个水平和B因素的第一个水平，来比较C因素第一和第三水平效应的差异。由于三因素析因设计方差分析模型中包含主效应A、B、C及交互效应A*B、A*C、B*C、A*B*C，并且A*C、B*C及A*B*C效应中均包含有C因素，则当Contrast语句中指定为比较C因素水平间的效应差异时，在设计矩阵的时候就需要构造关于C、A*C、B*C及A*B*C效应的对比系数，可构造设计矩阵(见图2、图3、图4)。一般当模型中有交互项不显著时，则将此交互项的平方和与自由度合并到误差中去，进行重新计算[5]。

这样就可以根据图2、图3、图4的系数矩阵，编写Contrast语句，实现在A_2B_1组合下，C因素第一水平和第三水平效应的差异检验，即：

contrast '*C1 vs C3 in A2 and B1*' *C* 1 0 -1
*A*C* 0 0 0 1 0 -1 *B*C* 1 0 -1 0 0 0
*A*B*C* 0 0 0 0 0 0 1 0 -1 0 0 0;

2 三因素析因设计资料实例分析

【实 例】在治疗肝癌的药物研究中，设置器官、药物、时间 3 个因素，器官因素(A)有 2 个水平：肝脏(A_1)和心脏(A_2)；药物因素(B)有 2 个水平：分别为“丝裂霉素＋高分子物质＋磁性物质”(实验组B1)和“丝裂霉素”(对照组 B_2)；时间因素(C)有 3 个

			C			
			C_1	C_2	C_3	
			1	0	-1	
A	A_1	0	0	0	0	0
	A_2	1	1	0	-1	0
			1	0	-1	0

图 2 对比系数矩阵

			C			
			C_1	C_2	C_3	
			1	0	-1	
B	B_1	1	1	0	-1	0
	B_2	0	0	0	0	0
			1	0	-1	0

图3 对比系数矩阵

				C			
				C_1	C_2	C_3	
				1	0	-1	
A	A_1	B	B_1 0	0	0	0	0
			B_2 0	0	0	0	0
	A_2	B	B_1 1	1	0	-1	0
			B_2 0	0	0	0	0
				1	0	-1	0

图 4 对比系数矩阵

水平:分别为给药后 15 min(C_1)、30 min(C_2)和 60 min(C_3)。将 60 只小鼠随机分为 12 组(2×2×3 种组合),每组 5 只。观察指标(反应变量)为组织中丝裂霉素的浓度(μg/g),见表 1[6],行三因素析因设计资料的方差分析。

2.1 析因设计的方差分析

主要考察器官、药物、时间3个因素的主效应和交互效应。使用SAS/STAT中的GLM过程,编写程序为:

```
data example1 ;
  do A= 1 to 2;
do id= 1 to 5;
    do B= 1 to 2;
 do C=1 to 3;
input x@@;output;
 end;
    end;
end;
  end;
datalines;
0.118 9 0.349 8 0.240 4 0.348 2 0.620 4 0.396 8
……(省略的数据行)
0.799 9 3.361 7 0.562 8 0.163 7 0.394 2 0.119 6
;
proc glm;
class A B C;
model x=A B C A*B A*C B*C A*B*C;
  run;
```

结果见表2。

由表2可见,对各因素主效应及其交互效应的检验中,$P<0.000\ 1$,按水平,均有统计学意义,可以认

表1 小鼠不同器官组织的丝裂霉素浓度(μg/g)检测结果

	药物实验组(B_1)			药物对照组(B_2)		
	15 min(C_1)	30 min(C_2)	60 min(C_3)	15 min(C_1)	30 min(C_2)	60 min(C_3)
心脏(A_1)	0.118 9	0.349 8	0.240 4	0.348 2	0.620 4	0.396 8
	0.123 6	0.322 7	0.267 6	0.364 6	0.654 4	0.393 5
	0.133 3	0.348 8	0.250 5	0.378 0	0.677 9	0.394 2
	0.103 1	0.311 9	0.264 2	0.356 2	0.631 2	0.377 0
	0.092 0	0.327 0	0.243 4	0.359 6	0.622 1	0.391 8
肝脏(A_2)	0.778 7	3.615 3	0.564 3	0.161 3	0.377 4	0.119 4
	0.779 8	3.465 4	0.569 1	0.166 3	0.356 6	0.092 9
	0.756 0	3.498 0	0.579 9	0.150 2	0.374 8	0.105 0
	0.774 5	3.317 4	0.585 9	0.112 4	0.382 9	0.098 5
	0.799 9	3.361 7	0.562 8	0.163 7	0.394 2	0.119 6

表2 处理因素及交互作用的方差分析表

变异来源	自由度(df)	离均差平方和(SS)	均方(MS)	*F*	*P*
A	1	4.659 7525 4	4.659 752 54	3 393.71	<0.000 1
B	1	5.025 604 65	5.025 604 65	3 660.16	<0.000 1
C	2	9.855 487 47	4.927 743 73	3 588.89	<0.000 1
A*B	1	9.843 282 08	9.843 282 08	7 168.88	<0.000 1
A*C	2	5.790 761 17	2.895 380 59	2 108.71	<0.000 1
B*C	2	4.847 196 05	2.423 598 03	1 765.11	<0.000 1
A*B*C	2	5.876 447 03	2.938 223 52	2 139.91	<0.000 1
误差	48	0.065 906 73	0.001 373 06		

表3 C因素单独效应分析表(A_2B_1组合下)

对比水平	自由度(df)	离均差平方和(SS)	均方(MS)	*F*	*P*
C1VSC3	1	0.105 452 36	0.105 452 36	76.80	<0.000 1
C1VSC2	1	17.872 748 72	17.872 748 72	13 016.8	<0.000 1
C2VSC3	1	20.723 905 76	20.723 905 76	15 093.3	<0.000 1

为3个因素间均存在一级和二级交互效应,此时需要进一步分析各因素的单独效应。

2.2 单独效应分析

以固定A因素的第二个水平(肝脏)和B因素的第一个水平(实验组)为例(即控制在A_2B_1组合下),来分析C因素第一水平和第三水平效应的差异比较,构造设计矩阵(参见图2、图3、图4),利用SAS中Contrast语句编程实现。

```
data example2;
set example1;
proc glm data=example2;
class A B C;
model x=A B C A*B A*C B*C A*B*C;
contrast 'C1 vs C3 in A2 and B1' C 1 0 -1
A*C 0 0 0 1 0 -1 B*C 1 0 -1 0 0 0 0
A*B*C 0 0 0 0 0 0 1 0 -1 0 0 0 0;
run;
```

同样,我们也可以对C因素其他水平之间进行比较,如:

```
contrast 'C1 vs C1 in A2 and B1' C 1 -1 0
A*C 0 0 0 1 -1 0 B*C 1 -1 0 0 0 0
A*B*C 0 0 0 0 0 0 1 0 -1 0 0 0 0;
contrast 'C2 vs C3 in A2 and B1' C 0 1 -1
A*C 0 0 0 0 1 -1 B*C 0 1 -1 0 0 0
A*B*C 0 0 0 0 0 0 0 1 0 -1 0 0 0 0;
```

在A_2B_1组合下对C因素的单独效应分析结果见表3。

从表3可知,在固定A因素的第二个水平(肝脏)和B因素的第一个水平(实验组)之下,对时间因素各水平间的效应进行两两比较的检验中,$P<0.000\ 1$,按水平,差异均具有统计学意义。其它任意两个因素均固定在某水平情况下的第三个因素各水平间的效应分析可参照上述方法实现。

通过以上设计参数对比系数矩阵,利用SAS中Contrast语句来实现析因设计资料的单独效应分析是一种方便合理的方法,并且利用析因分析得到的总的误差均方作为标准误,使结论更加可靠,同时Contrast语句不仅适用于一般线性模型,还可以用于混合效应模型等。

参 考 文 献

[1] 金丕焕.医用统计方法[M]. 第2版.上海:复旦大学出版社,2003. 79-84.

[2] 胡良平.现代统计学与SAS应用[M]. 北京:军事医学科学出版社,2002.98-103.

[3] 曹素华.实用医学多因素统计方法[M]. 上海:上海医科大学出版,1998.10-12.

[4] 高惠璇.SAS系统SAS/STAT软件使用手册[M]. 北京:中国统计出版社,1997.318-319.

[5] 盛 骤,谢式千.概率论与数理统计[M]. 北京:第3版.高等教育出版社,2001.290-291.

[6] 颜 虹.医学统计学[M]. 第1版.北京:人民卫生出版社,2005. 272-273.

(责任编辑:关蕴良)