返回总目录

目　　录

第六部分
一般线性模型与四型离差平方和

第 31 章　一般线性模型：统计程序 PROC GLM

31.1　PROC GLM 程序概述

GLM 是一般线性模型 (General Linear Model) 的简称，其推算参数的理论基础是最小误差平方法 (The Least Squares Method)，最适用于不平衡的实验设计，亦即各组 (或各细格) 的观察体个数不等。若你的实验是一个平衡的实验设计，则你应该用 ANOVA 程序 (见第 26 章) 来执行变异数分析，以节省计算的时间与电脑的记忆空间。

PROC GLM 可以被应用在多种不同的统计分析上，如：

a. 变异数分析 (特别是不平衡的实验设计)
b. 共变量分析 (Analysis of Covariance)
c. 极多变量变异数分析 (MANOVA)
d. 重复观察的变异数分析 (又称 Split-Plot Factorial Analysis)
e. 一元回归分析
f. 复回归分析
g. 极多项式回归分析 (Polynomial Regression)
h. 加权回归分析 (Weighted Regression)
i. 二项式反应面分析 (Response Surface Modeling)
j. 净相关的计算等。

本章的重点在于介绍 PROC GLM 内有关变异数分析的指令 (亦即上述 a 到 d 的用法)。至于 GLM 程序在回归分析上的应用（上述 e 到 j 的用途），则与 PROC REG 的基本语法类似（请参阅第 18 章的内容或本章第 31.4 节）。

31.2　统 计 模 型

欲详细了解有关变异数分析之统计模型的专有名词，请读者参阅第 26 章 (ANOVA) 内第 26.2 及第 26.3 节。以下简单介绍 PROC GLM 所能处理的几种模型。

若以 A，B，C，等字母代表实验设计的自变量，Y 代表因变量 ，X1，X2 与 X3 代表连续变量，则 PROC GLM 可处理的几种变异数分析之模型及其 SAS 指令如下：

实验设计	SAS 指令
一因子的变异数分析	MODEL Y = A;
主效果模型	MODEL Y = A B C;
交互效果模型	MODEL Y = A B A*B;
镶嵌效果模型	MODEL Y = B(A) C(B A);
极多变量变异数分析	MODEL Y1 Y2 = A B;
共变量分析	MODEL Y = A X1;

组内斜线模型	MODEL Y = A X1(A);
共斜线模型	NODEL Y = A X1 X1*A;

31.3　如何撰写 PROC GLM 程序

PROC GLM 含十七道指令。其中只有 PROC GLM 和 MODEL 是必要的，不可省略。其他十五道指令则可有可无。但当实验设计内含一个以上的自变量时，读者必须用 CLASS 指令标明。下面请读者特别注意各指令出现的次序。

不可省略

```
PROC GLM    选项串;
```

必须放在 MODEL 与 MEANS 指令前

```
  CLASS     自变量名称串;
```

不可省略

```
  MODEL     因变量串= 实验效果 /选项串;
```

必须放在 MODEL 指令之后，而且可在交谈式环境下执行

```
  MEANS     效果名称串 /选项串;
  CONTRAST  '比较式的名字' 各组效果的系数据
            / 选项;
  ESTIMATE  '估计值的名字' 各组效果的系数据
            /选项串;
  LSMEANS   效果名称串 /选项串;
  MANOVA    H= 效果名称
            E= 效果名称
            M= 变量的转换式
            PREFIX = 新变量的名称代号
            MNAMES=新变量的名称串/选项串;
```

必须放在 MODEL 指令之后，而且可在交谈式环境下执行

```
  OUTPUT    OUT=输出资料文件名称
            关键字=变量名称串;
  RANDOM    效果名称串 /选项串;
  REPEATED  重复变量的名称串组数 (组名)
            变量的转换 /选项串;
  TEST      H= 效果名称串
            E= 效果名称 /选项串;
```

这些指令可以放在此程序的任何一处，RUN；指令之前

```
  ABSORB    变量名称串;
  BY        变量名称串;
            FREQ 变量名称;
```

```
  ID        变量名称串;
  WEIGHT    变量名称;
```

下面是除 PROC GLM 以外其余十六道指令的功能简介：

PROC GLM	指令功能简介
CLASS	标明自变量
MODEL	界定统计模型
MEANS	计算各组的平均数
CONTRAST	以线性方程序重新组合参数据执行检定
ESTIMATE	检验参数的线性组合
LSMEANS	计算根据最小误差法求得的平均数
MANOVA	执行多变量变异数分析
OUTPUT	界定一个输出资料文件，使其包含预测值与预测误差
RANDOM	宣告某些效果是随机效果，然后计算它的变异数的均方
REPEATED	执行重复观察之实验设计的变异数分析
TEST	检定某些实验效果时，界定检定的分子与分母
ABSORB	简化模型中的主效果
BY	将资料文件分成几个部分，分别对其执行统计分析
FREQ	表明观察体重复出现的次数
ID	观察体的识别编号
WEIGHT	与 FREQ 作用类似，旨在标明数据的加权值

下面将对这些指令作详尽的介绍：

指令 #1　PROC GLM 选项串;

在此指令后有下列六选项：

(1) DATA= 输入资料文件名称

指明对那一个 SAS 资料文件执行分析。若省略此选项，则 SAS 会自动找出在此程序前最后形成的 SAS 资料文件，并对它执行分析。

(2) ORDER= FREQ 或

ORDER= DATA 或

ORDER= INTERNAL 或

ORDER= FORMATTED (内设值)

界定自变量内组别的次序，这个选项和 CONTRAST 及 ESTIMATE 指令是息息相关的。

当 ORDER=FREQ 时，观察体个数最多的那一组就是第一组，以下类推。

当 ORDER=DATA 时，组别是按照输入资料文件中各组第一次出现的次序而决定的。

当 ORDER=INTERNAL 时，组别按其代号由小到大 (如 1，2，3，等) 排列，或按各组名称的英文字母顺序排列 (如：FEMALE 在 MALE 之前)。

当 ORDER=FORMATTED 时，则组别的顺序以外部的格式 (External Format) 而决定。这也是本选项的内设值。

(3) MANOVA

要求 PROC GLM 将含一个或一个以上遗漏数据的观察体剔除。当读者以交互式 (Interactive Mode) 方式进行多变量的变异数分析时，最好界定此选项。

(4) MULTIPASS

要求 PROC GLM 在必要情况下重读输入资料文件内的数据。由于这个选项会占用极多的记忆空间，同时耗时很多，除非必要，读者可以省略此选项。

(5) OUTSTAT=(含分析结果的) 输出资料文件名称

这个选项会界定一个含分析结果的输出资料文件。此输出资料文件将含离差平方和 (SS)、F 检定值以及各实验效果的显著程度。若读者同时界定 MANOVA 指令中的 CANONICAL 选项但未界定 M= 的选项，则典型相关分析的结果也会纳入此输出资料文件内。

(6) NOPRINT

要求 PROC GLM 抑止分析结果在报表上的打印。除非读者只想制造某些输出资料文件而不太想看到分析的结果，否则这个选项不太有用。

指令 #2　CLASS 自变量名称串；

这道指令也可以写成 CLASSES 自变量名称串；此指令标明资料文件中到底哪些是统计模型的自变量。这些自变量可以是数值的或文字的。若是文字变量，则其长度不可超过十六个字母。

指令 #3　MODEL 因变量串＝实验效果 / 选项串；

删除号 (/) 之前的部分：(因变量串＝实验效果) 要求你必须先决定何种统计模型适用于你现在要分析的数据，然后根据 26 章 (ANOVA) 第 26.3 节的原则将它写出。

删除号 (/) 后的选项可分五大类来讨论。

第一类选项　与截距的界定有关，有两个选项：

(1) NOINT

要求 GLM 程序将截距 (常数) 的参数排除在模型之外。

(2) INT (或 INTERCEPT)

要求 GLM 程序印出截距的统计检定。

第二类选项　与报表的打印有关，有三个选项：

(1) NOUNI

此选项抑止有关单变量变异数分析之结果的打印，最适用于多变量或重复观察的变异数分析。

(2) SOLUTION

要求 GLM 程序印出常态公式的解 (亦即一般线性模型中参数的估计)。当省略 CLASS 指令时，GLM 程序会自动印出此解。

(3) TOLERANCE

印出容忍量。其定义是 $1-R^2$，在此 R^2= 复相关系数的平方。有关容忍量的详细解释，请见第 17 章第 17.4 节。

第三类选项　与虚无假设的检定有关，有九个选项：

(1) E

要求 GLM 程序印出所有可估计函数 (Estimable Functions) 的值。

(2) E1

要求 GLM 程序只印出每一效果的第一型可估计函数值 (Type Ⅰ Estimable Function)。

(3) E2

要求 GLM 程序只印出每一效果的第二型可估计函数值 (Type Ⅱ Estimable Function)。

(4) E3

要求 GLM 程序只印出每一效果的第三型可估计函数值 (Type Ⅲ Estimable Function)。

(5) E4

要求 GLM 程序只印出每一效果的第四型可估计函数值 (Type Ⅳ Estimable Function)。

(6) SS1

要求 GLM 程序只印出每一效果的第一型离差平方的总和 (Type Ⅰ Sum of Squares)。

(7) SS2

要求 GLM 程序只印出每一效果的第二型离差平方的总和 (Type Ⅱ Sum of Squares)。

(8) SS3

要求 GLM 程序只印出每一效果的第三型离差平方的总和 (Type Ⅲ Sum of Squares)。

(9) SS4

要求 GLM 程序只印出每一效果的第四型离差平方的总和 (Type Ⅳ Sum of Squares)。若读者已选用 E1，E2，E3 或 E4，则 GLM 会自动印出与其相对应的 SS1，SS2，SS3，SS4。这一类选项的内设值是 E1，E3 或 SS1，SS3。

第四类选项　下列两个选项控制计算过程的打印：

(1) XPX

要求印出 (X'X) 的向量积矩阵。

(2) INVERSE (或 I)

要求印出 (X'X) 的反矩阵或 (X'X) 之通用式反矩阵 (Generalized Inverse Matrix)。

第五类选项　可用来调整统计的精确性，有一个选项：

(1) ZETA= 极小的正实数

ZETA 选项控制第三型与第四型可估计函数值之可估计性检定的敏感度。此选项的内设值是 10 的 -8 次方。这个内设值足以应付大多数的模型检定。

指令 #4　MEANS 效果名称串 / 选项串;

此指令的前半部 (删除号之前) 可用来要求 GLM 程序算出某些自变量 (和其交互作用或镶嵌作用) 中各组 (各细格) 的平均数。比方说：SEX 表示性别 (下分男、女)，RACE 表示种族 (下分黑、白)，则我们可用下列的 SAS 指令算出资料文件中男人、女人、黑人、白人、男黑人、男白人、女黑人及女白人在因变量年薪 (SALARY) 上的平均数：

```
PROC ANOVA;
    CLASS SEX RACE;
    MODEL SALARY=SEX RACE;
    MEANS SEX RACE SEX*RACE;
```

删除号 (/) 之后可用的选项有二十七个。前十七个选项是用来对 MEANS 指令中所列的主效果执行不同的显著性考验。以前例而言，MEANS 指令会比较男与女，及黑人与白人之间的年薪差异；后十个选项则与统计检定的各项事宜有关。

(1) BON

执行显著性 t 检定，其理论基础是班弗尼氏的不等律 (Bonferroni Inequality)。

(2) DUNCAN

执行唐肯氏多范围检定 (Duncan's Multiple-Range Test)。

(3) DUNNETT (控制组组别)

这个选项界定唐那氏的两组平均数之双尾检定。唐那氏 (Dunnett) 的检定依据 t 分配而且必须是实验组与控制组平均数的比较。因此，括号内必须指明控制组的组别，请看下面的程序：

```
MEANS A/DUNNETT ('CONTROL');
```

根据这个指令的语法 A 效果的 CONTROL 组就是控制组。若控制组的组别是以数字来表示的 (如 2)，则不必再加单引号，如

```
MEANS A/DUNNETT (2);
```

这个选项的控制组一般是设定在一组 (内设值)。若控制组不只一组时，读者可同时在括号内提及，如：

```
MEANS A B C/DUNNETT ('FIRST' 'SECOND' 'THIRD');
```

根据上述指令的语法，A 效果的控制组是第 FIRST，B 效果的控制组是第 SECOND 组，C 效果则是第 THIRD 组。

(4) DUNNETTL (控制组组名)

这个选项界定唐那氏的两组平均数之单尾检定，而且预期的差异必须是负值 (亦即实验组的平均数小于控制组的平均数)，因此临界值定在 t 分配的下端。

有关控制组的内设值以及撰写语法，请参见上面 (3) DUNNETT 的说明。

(5) DUNNETTU (控制组组名)

这个选项界定唐那氏的两组平均数之单尾检定，而且预期的差异必须是正值 (亦

即实验组的平均数大于控制组的平均数)，因此临界值定在 t 分配的上端。

有关控制组的内设值以及撰写语法，请参见上面 (3) DUNNETT 的说明。

(6) GABRIEL

执行贵博氏的多重比较 (Gabriel's Multiple Comparison Procedure)。

(7) REGWF

执行 Ryan-Einot-Gabriel-Welsch 的 F 检定。

(8) REGWQ

执行 Ryan-Einot-Gabriel-Welsch 的 t 检定。

(9) SCHEFFE

执行沙菲氏的多重比较检定。

(10) SIDAK

执行 Sidak 两组平均数比较的 t 检定。

(11) SMM [或 (12) GT2]

执行 Sidak 的独立样本 t 检定。当两组人数不等时，此法也就是哈氏 (Hochberg) 的 GT2 法。

(13) SNK

执行纽曼-库尔 (Newman-Keuls) 的两组样本平均数差的 t 检定。

(14) T [或 (15) LSD]

执行配对组 t 检定。因为 GLM 程序所处理的可能是不平衡的设计，故其结果与费契尔的最小显著差 (LSD) 的检定结果不一定完全相同。

(16) TUKEY

执行土其氏的 HSD 检定。

(17) WALLER

执行 Waller-Duncan 的 K-ratio 之 t 检定。

(18) ALPHA=P

界定统计检定的显著程度，内设值是 .05。当此选项与前述选项 (2) DUNCAN 并用时，ALPHA 的值必须是 .10，.05，及 .01 三者之一。

(19) LINES

将读者选用的显著性检定的分析结果 (亦即各平均数) 做由大到小的排列。若某一对平均数之间无显著的差异，则 SAS 将它们印在同一行上，并以虚线将它们与其他有显著差异的平均数分开。当读者选用 DUNCAN，REGWF，REGWQ，SNK 或 WALLER 等显著性检定 (或当实验设计是平衡，或当实验设计只含两细格时)，此选项会自动包括在分析过程内；否则读者必须另外附加。再者，此选项最适用于平衡的实验设计。若细格内的人数不等，GLM 程序会先计算出各细格人数的调和平均数 (Harmonic Mean)，并用此调和平均数来比较主效果的平均数差异。然而若各细格内的人数差异太大时，某些比较的显著结果会过于乐观。

(20) CLDIFF

将 BON，GABRIEL，SCHEFFE，SIDHK，SMM，GT2，T，LSD 或 TUKEY 等显著性检定的结果用信赖区间的方式表示。当实验设计是一个不平衡的设计时，

CLDIFF 选项会自动包括在分析过程内。当读者选用 DUNCAN，REGWF，REGWQ，SNK 或 WALLER 时，则必须另外附加此选项。

(21) CLM

将 MEANS 指令中所提到的效果之各组平均数以信赖区间的方式表示。此选项必须与 BON，GABRIEL，SCHEFFE，SIDAK，SMM，T 以及 LSD 等联用。

(22) NOSORT

与上述 CLDIFF 或 CLM 选项合用，抑止平均数按大小重新作排列。

(23) E=效果名称

此选项界定上述各显著性检定的分母。若省略此选项，则实验设计的余差均方值 (MSResidual) 就自动成为分母。

(24) DEPONLY

要求 GLM 程序只印出因变量的平均数。若省略此选项，则 GLM 程序会印出资料文件中所有连续变量的平均数。

(25) ETYPE=1 (或 2，或 3，或 4)

界定 F 检定中分母矩阵的均方 (Mean Square) 类型。内设值是分析检定中最高的一型。

(26) HTYPE=1 (或 2，或 3，或 4)

与前述 WALLER 选项并用。此选规界定 F 检定中分子矩阵的平均方类型。内设值是分析检定中最高的一型。

(27) KRATIO=正整数

与 WALLER 选项联用。这个比例 (第一类型错误 / 第二类型错误) 的值若订为 50，100 与 500，则大约与 ALPHA 值 .10，.05 与 .01 相对应。这个选项的内设值是 100。

指令 #5　CONTRAST “比较式的名字”各组效果的系数 / 选项串;

请读者仔细阅读下页几个示范的例子，以便了解这个指令的格式。首先，我们假设有一个二因子的主效果实验设计：A 分为五组，B 分为两组。

```
MODEL Y= A B;
CONTRAST 'A LINEAR & QUADRATIC'
          A -2 -1  0  1 2,
          A  2 -1 -2 -1 2;
CONTRAST 'CONTROL VS OTHERS'
        A -1 .25 .25 .25 .25;
CONTRAST 'ONE VS TWO' B -1 1;
```

由上例我们可以归纳出几点原则：

- “比较式的名字”必须放在单引号内，名字的长度以二十个字母为限；命名的方式不拘，但不可夹带分号 (;)。
- 各组效果系数前必须先注明所要比较的效果。这些效果必须是 MODEL 指令中出

现过的。如上例中我们不能比较 A*B 的交互效果，因为 MODEL 指令中无此效果。这些系数的总和必须是 0，而且只能是整数或小数 (SAS 不接受任何分数作为系数)，各系数之间要以空格隔开。

- 若同一个 CONTRAST 指令内含一个以上的比较式，则以逗号（ , ）将系数串隔开。

删除号 (/) 后的选项有四：

(1) E

印出线性函数的向量，L。

(2) E= 效果名称

界定以 E 的效果为 CONTRAST 指令中 F 检定的分母。内设值是误差的平均方 (MS Error)。

(3) ETYPE= 1(或 2，或 3，或 4)

计算选项 E= 效果名称中，效果的离差均方之类型。

(4) SINGULAR=极小的正实数 (如 0.007)

这个选项用来检定 CONTRAST 指令所导出的线性函数，是否为可估计的 (Estimable)。其检定的标准如下：

以 i 代表 L (线性函数之矩阵) 的某一横列，

$$H=(X'X)^{-1}X'X,$$

如果下式成立，则 L_i 的值被 SAS 认为不能估计出来，

若 $L_i = 0$，而且 $ABS[L_i-(LH)_i]>$ 极小的正实数 (如 0.007)

或适当 $L_i \neq 0$，而且 $ABS[L_i-(LH)_i] > ABS(L_i)*$极小的正实数 (如 0.007)

这个选项的内设值等于 10 的 -4 次方。

指令 #6　ESTIMATE “估计值的名字”各组效果的系数 / 选项串;

这个指令与上述的 CONTRAST 指令类似，它们遵循同样的原则。但除此之外，ESTIMATE 指令还可印出 t 检定的值，t 检定的分母 (即平均误差的值)，以及其统计显著程度。

请看下面的例子：

```
MODEL Y = A:
ESTIMATE 'A1 VS A2' A 1 -1;
ESTIMATE '1/3(A1+A2)-2/3 A3' A 1 1 -2/DIVISOR= 3;
ESTIMATE '1/3(A1+A2)-2/3 A3' A .33333 .33333 -.66667;
```

上面第二和第三式的意义完全相同，现在让我们来讨论删除号 (/) 后的三个选项：

(1) DIVISOR= 整数

GLM 用此整数当做分母来除删除号前的效果系数。

(2) E

印出线性函数的向量，L。

(3) SINGULAR= 极小的正实数

以此数为标准，检定 ESTIMATE 指令所导出的线性函数是否为可估计的 (Estimable)。其检定标准与前述 CONTRAST 指令中同一选项完全一致，故不再赘述。内设值也等于 10 的 -4 次方。

指令 #7　LSMEANS 效果名称串 / 选项串；

LSMEANS 是以最小误差平方法所估计之平均数的代称 (英文称 Least Squares Means)。下页示范 LSMEANS 的语法：

```
PROC GLM;
    CLASS A B;
    MODEL A B A*B;
    LSMEANS A A*B;
```

上面的程序指示 SAS 以最小误差平方法估计 A 及 A*B 两效果内各组 (或各细格) 的矫正平均数，好似整个实验设计是一个平衡的设计。请注意：LSMEANS 指令里所提的效果，必须是 MODEL 指令里已经提过的效果。

删除号 (/) 后的选项有十个，分述如下：

(1) E

印出最小误差平方平均数计算过程中所用到的可估计函数值。有关 E 的定义，在下一章 (第 32 章) 内有详细的说明。

(2) STDERR

印出 t 检定 (H0：LSM= 0) 的分母与其显著程度。

(3) TDIFF

印出各平均数对比较的 t 值以及其统计显著程度。

(4) PDIFF

印出各平均数对比较后的统计显著程度，与上述 (2) STDERR 不同。

(5) E= 效果名称

须与上述 STDERR，TDIFF，及 PDIFF 等选项合用，作用在于指定某一个效果的平均方做为 t 检定的分母。若读者选用 STDERR，TDIFF 及 PDIFF 选项，但省略此选项，则 GLM 自动以误差的平均方 (MS Error) 为 t 检定的分母。

(6) ETYPE=1 (或 2，或 3，或 4)

计算选项 E=效果名称 中，效果的离差平均方之类型。

(7) SINGULAR=极小的正实数

以此数为标准，检定 LSMEANS 指令所导出的线性函数是否为可估计的 (Estimable)。其检定标准与前述 CONTRAST 指令中同一选项完全一致，故不再赘述。内设值也等于 10 的 -4 次方。

(8) OUT= 输出资料文件名称

界定一个输出资料文件，内含 LSMEANS 指令所导出的矫正平均数，平均数的标准误差，以及平均数间的共变异数 (如果读者同时界定下一个选项 COV)。

(9) COV

要求将矫正平均数之间的共变异数，纳入上述 OUT= 输出资料文件内。此选项必须与 OUT=选项联用而且 LSMEANS 的效果必须只有一个。

(10) NOPRINT

要求 GLM 程序不将分析的结果打印在报表上。如果读者撰写 LSMEANS 指令的目的只是为了产生一个 OUT= 的输出资料文件，则此选项会十分有用。

指令 #8　MANOVA　H＝效果名称　E＝效果名称

M＝变量的转换式

PREFIX＝新变量的名称代号

MNAMES＝新变量的名称串 / 选项串;

此指令要求多变量变异数分析 (MANOVA)，同时也导致一种对遗漏数据的特殊处理法。下面分别介绍此指令的各部分：

(1) H= 效果名称 (或 _ALL_ 或 INTERCEPT)

界定多变量变异数分析所检验的假设矩阵。H=的效果必须已被包含在 MODEL 指令里。当读者有意通盘地对 MODEL 指令中所提到的所有效果执行多变量变异数分析时，则可用 H=_ALL_ 表示。在 GLM 程序中，这些效果将经由四种方法进行多变量变异数分析，亦即：Hotelling-Lawley Trace，Pillai's Trace，Wilks' Criterion 和 Roy's MaximumRoot Criterion。这四种分析的结果仍依据 F 分配来判断其显著程度。当 H=INTERCEPT 时，表示读者有意对模型中的截距或总平均数 (Grand Mean) 作统计的检定。

(2) E= 效果名称

界定 F 检定的分母。若省略此选项，则余差的平均方 (MS Residual) 就自动成为分母。

(3) M=变量的转换式

界定因变量的转换式，下例示范这个转换式的写法：

```
MODEL Y1-Y5=A B(A);
MANOVA H=A E=B(A)
      M=Y1-Y2, Y2-Y3, Y3-Y4, Y4-Y5
      PREFIX=DIFF;
```

上列的指令将原有的因变量转换成相邻两平均数的差。转换式的格式是

M = 转换变量 {±转换变量....};

此处的**转换变量**可以是原始变量或常数乘以原因变量。{}中的部分可有可无。若含一个以上的转换式，则以逗点 (,) 相隔。

(4) PREFIX=新变量名称代号

在上面的例子中，由于有 PREFIX=DIEF 这个选项，因此新变量将被命名为 DIFF1，DIFF2，DIFF3 及 DIFF4。请注意：这个名称代号必须是八个字母以内的名字，数字 1，2 等分别与转换式 1，2 等对应。

(5) MNAMES=新变量名称串

这个选项的用途与选项 PREFIX 类似，但不同之处是这个选项为选项 M= 中转换过的每一个新变量一个不同的名称。这些名称之间不以数字 1，2 等相连，名称之间仍以空格相隔。

删除号 (/) 后的选项，有下列七个：

(1) PRINTH

印出 F 检定的分子矩阵 (即假设矩阵)。

(2) PRINTE

印出 F 检定分母矩阵与其净相关矩阵。

(3) HTYPE=1 (或 2，或 3，或 4)

界定假设矩阵的变异数平方值的型态 (可等于 1，2，3，或 4)。内设值是分析过程中所用过最高型的值。

(4) ETYPE=1 (或 2，或 3，或 4)

界定 F 检定中分母矩阵的变异数平方值的型态 (可等于 1，2，3，或 4)。内设值是分析过程中所用过最高型的值。

(5) ORTH

要求转换式(在选项 M= 中所形成的新变量) 先经过标准化正交 (Orthonormalization) 的处理。

(6) CANONICAL

对 F 检定中的分子与分母矩阵进行典型分析。此分析的结果与另一统计程序 PROCCANDISC 的分析结果应该完全一致。

(7) SUMMARY

印出每一因变量的变异数分析摘要表。若曾选用选项 M= ，则变异数分析摘要表是根据转换后的因变量所形成的。

我们现在举两个例子示范 MANOVA 指令的写法：

例 1

```
PROC GLM;
     CLASS A B;
     MODEL Y1-Y5=A B(A);
     MANOVA H=A E=B(A)/PRINTH PRINTE
            HTYPE=1 ETYPE=1;
     MANOVA H=B(A)/PRINTE;
     MANOVA H=A E=B(A)
           M=2*Y2-2*Y3, 4*Y4-6*Y5+2*Y1;
```

在 MODEL 指令中，我们看到有五个因变量(Y1 到 Y5)，故可采用多变量变异数分析。

第一个 MANOVA 指令中，F 检定的分子是 A 效果矩阵，分母则是 B(A) 效果矩阵，这两矩阵将分别被印出。程序还要求计算出这两个矩阵的第一型离差平方总和。

在第二及第三个 MANOVA 指令中，由于没有规定哪一型的离差平方和，因此 GLM

程序会自动计算第一型与第三型的离差平方和。

第二个 MANOVA 指令中，没有规定分母的矩阵，故 GLM 程序会采用余差的均方(MS Residual，亦即选项 E= 的内设值) 为 F 检定的分母。

第三个 MANOVA 指令，请读者注意 M= 选项。原因变量经过转换后，由于未使用 PREFIX=或 MNAMES=选项，因此 GLM 程序自动称这两个转换过的新变量为 MVAR1 及 MVAR2。这些新变量与 A 效果之间的关系是由这个 MANOVA 指令所检验的。

例 2

```
PROC GLM;
    CLASS GROUP;
    MODEL DOSE1-DOSE4=GROUP;
    MANOVA H=GROUP
           M=3*DOSE1-DOSE2+DOSE3+3*DOSE4,
           DOSE1-DOSE2-DOSE3+DOSE4,
           -DOSE1+3*DOSE2-3*DOSE3+DOSE4
           MNAMES=LINEAR QUADRATIC CUBIC/PRINTE;
```

此例中，原因变量经过选项 M= 做线性转换、抛物线转换及三次方曲线转换。然后新因变量经 NNAMES= 选项命名为线性，抛物线性及三次曲线值 。选项 PRINTE 指示 SAS 印出 F 检定的分母矩阵 (在此例中，由于无 E= 选项，故分母矩阵即是余差的矩阵) 以及多项式值之间的净相关。

指令 #9　OUTPUT　OUT＝输出资料文件名称　关键字＝变量名称串;

本指令包括两个部分：OUT= 与关键字= 。

OUT=输出资料文件名称

这个资料文件含原输入资料文件的所有变量，以及本指令中所提到的关键字 (如：PREDICTED，RESIDUAL 等)。

关键字=变量名称串

下页列举十六种关键字及其定义：

(1) PREDICTED (或 P)= 预测值。

(2) RESIDUAL (或 R)= 预测误差。

(3) L95M= 因变量平均数的 95% 信赖区间之下限。

(4) U95M= 因变量平均数的 95% 信赖区间之上限。

(5) L95= 因变量预测值的 95% 信赖区间之下限，这个值考虑了抽样误差及参数估计值的变异数。

(6) U95= 因变量预测值的 95% 信赖区间之上限，这个值考虑了抽样误差及参数估计值的变异数。

(7) STDP= 预测值平均数的标准误差。

(8) KSTDR= 误差的标准误差。

(9) STDI= 个别预测值的标准误差。

(10) STUDENT= 经过标准化的误差。

(11) COOKD= 库格氏影响力的统计值。

(12) H= 影响力，定义是 $X_i(X'X)^{-1}X_i$

(13) PRESS= 除去一个观察体后所求得的该观察体之预测误差。

(14) RSTUDENT= 除去一个观察体后所求得的该观察体的标准化误差。

(15) DFFITS= 将观察体对预测值的影响力加以标准化。

(16) COVRATIO= 将观察体对回归系数的共变异数之影响力加以标准化。

指令 #10　RANDOM 效果名称串 / 选项串；

这个指令可用来指出 MODEL 指令所含的各项效果中，哪个 (些) 是随机效果。从这个界定中，GLM 程序会自动印出第三型、第四型的变异数平方，或平均数比较的平方值。读者可在 MODEL 指令之后，多次界定 RANDOM 指令。若省略 RANDOM 指令，则 GLM 程序视 MODEL 指令中所有的效果为固定效果 (Fixed Effects)。

删除号 (/) 后的选项串：

(1) Q

印出所有固定效果的二次式函数值 (Quadratic Forms)。

(2) TEST

要求 GLM 程序对 RANDOM 指令中所提的各式随机效果执行适当的 F 检定，并且 F 检定的分母完全根据各效果变异数均方的期望值 (Expected Mean Squares) 而来。唯一值得注意的是：若 A，B 两主效果被宣告成随机效果，这并不代表 A*B 一定被 SAS 视为随机效果。因此下页两个 RANDOM 指令所得的 F 检定是不一样的：

```
RANDOM A B/TEST;
RANDOM A B A*B/TEST;
```

指令 #11　REPEATED 重复变量的名称串组数 (组名) 变量的转换 / 选项串；

假设有三种实验各种控制在四个不同的时间进行，则每一个被试有十二个分数。假如这十二个分数分别以 Y1－Y12 表示，则下面的指令可代表这十二个分数的统计分析：

```
REPEATED TRIAL 3(A B C), TIME 4 (T1 T2 T3 T4);
```

这个指令言简意赅地说明了下列的数据结构：

因变量	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12
TRIAL 的值	1	1	1	1	2	2	2	2	3	3	3	3
TIME 的值	1	2	3	4	1	2	3	4	1	2	3	4

现在让我们用这个例子来解释 REPEATED 指令的写法：

重复变量的名称

即上例中的 TRIAL 及 TIME。重复变量必须与因变量有关。重复变量的名称不可与

输入资料文件内任何其他的变量名称相同，它的长度不可超过八个字母。

组数

界定上述重复变量的组数。若该变量的组数为 1，则可以省略此选项。从上面的例子中，我们可看出重复变量 TRIAL 有三组，而 TIME 有四组，所以其排列组合共产生了十二个分数 (以 Y1－Y12 表示)。

(组名)

这个选项的值必须包含在括号内。括号内的值用来标明各分组，其个数须与组数吻合。如 TRIAL 这个重复变量有三组，即 A，B 与 C。组名与组名之间以空格分隔。TIME 变量则有四组，分别以 T1，T2，T3，T4 等表示。

变量的转换

下面的 (1) 与 (2) 变量转换均属于正交的转换；其余则属非正交的转换。每一转换数的平均数比较 (Contrast) 有 1 度的自由度：

(1) POLYNOMIAL

产生多项式的正交比较，如：直线式、抛物线式，及三次曲线式的比较。

(2) HELMERT

比较同一变量内一组平均数与其后各组平均数的平均。如 TRIAL 中，比较 A 组平均数及 B，C 两组平均数的平均。

(3) PROFILE

比较同一变量内相邻两组的平均数。

(4) CONTRAST (参考组之组别)

读者先选定变量中的某一组为参考组，然后其他各组依序与此参考组做比较。如 CONTRAST (A) 表示 A 是参考组，所以 A 与 B，A 与 C 的平均数作比较，参考组组别的内设值是最后一组。这是内设的转换方法。

(5) MEAN (参考组之组别)

比较同一变量内某一组平均数与其他各组的平均，但不比较参考组平均数与其他各组平均数的平均。 参考组组别的内设值是最后一组。 如：MEAN 表示 TRIAL 变量中的 C 是参考组 (因为 C 组是最后一组)。因此比较 A 组平均数与 B，C 两组平均数的平均；但不比较 C 组平均数与 A，B 两组平均数的平均。

请读者注意：指令中若含一个以上的重复变量，则以逗点分隔这些变量。每一变量内的资料，如：名称、组数据(组别)、变量的转换，应当以空格分隔。若读者同时界定 CONTRAST 与 TEST 指令，则 REPEATED 指令必须在这两个指令之后。

删除号 (/) 后的选项名称串

下列九个选项可置于 REPEATED 指令的删除号之后：

(1) NOM

不印出多变量变异数分析的结果，只印出单变量变异数分析的结果。

(2) NOU

与上述选项相反，不印出单变量变异数分析的结果，只印出多变量变异数分析的结果。

(3) PRINTM

印出变量转换之 M 矩阵的转置矩阵，亦即 M' 矩阵。

(4) PRINTH

印出多变量变异数分析的分子矩阵。

(5) PRINTRE

印出多变量变异数分析的分母矩阵。当统计假设之间彼此不独立时，此选项同时执行球形假设 (Sphericity) 的检定。

(6) PRINTV

印出每一个多变量检定的特性根与特性向量。

(7) SUMMARY

印出每一个变量转换式的变异数分析摘要表。

(8) CANONICAL

针对因变量所导出的 H 与 E 矩阵进行典型分析，其分析结果应与 PROC CANDISC 程序的分析结果相似。

(9) HTYPE=1 (或 2　或 3　或 4)

界定 F 检定中分子矩阵的变异数平方值的型态 (可等于 1，2，3 或 4)。内设值是分析过程中所用过最高型的值。

指令 #12　TEST H＝效果名称串　E＝效果名称 / 选项串；

一般而言，GLM 程序以 MODEL 指令中的各项效果为分子，以余差的平均方 (MS Residual) 为分母，执行 F 检定。但读者可利用 TEST 指令自行指定其他效果的平均方 (Mean Squares 或 MS) 为分母进行额外的 F 检定。这种情形在重复观察的实验设计里最需要。

但是读者必须注意：在不平衡的实验设计里各项效果的平均方不一定彼此独立。所以即使虚无假设成立，各效果平均方的期望值 (Expected Value) 不一定是各组 (或各细格) 在母群中的变异数 (σ^2)。如此所形成的 F 检定则非正统的 F 值。SAS 对读者所自选的 H 与 E 效果名称 (即 F 检定的分子，分母) 不负任何责任。因此读者事前应多参阅统计书籍，或利用 RANDOM 指令来预测效果的期望值。

现将此指令的各部分介绍如下：

H=效果名称串

界定 F 检定的分子。这个效果名称必须是在 MODEL 指令中出现过的。

E=效果名称

界定 F 检定的分母，个数只限一个。同样的，此效果必须是在 MODEL 指令中出现过的。

删除号 (/) 后的选项串有两个：

(1) HTYPE=1 (或 2　或 3　或 4)

界定被测效果 (即 F 检定的分子) 的 MS 型态。这四种型态的定义在第 32 章有详尽的介绍。

(2) ETYPE=1 (或 2　或 3　或 4)

界定 F 检定中分母的 MS 型态。同样地，请读者参阅第 32 章的说明。

下面以一个重复观察的例子示范 TEST 指令的语法：

```
PROC GLM;
    CLASS A B C ;
    MODEL Y=A B(A) C A*C B*C(A);
    TEST H=A E=B(A)/HTYPE=1 ETYPE=1;
    TEST H=C A*C E=B*C(A)/HTYPE=1 ETYPE=1;
```

指令 #13　ABSORB 变量名称串;

此指令的主要用途在于节省 SAS 计算时所使用的内存与时间。

假若一个自变量下有许多组，而且这个自变量与其他自变量之间没有任何交互作用，则读者可在此指令中提出这个自变量的名称。如此一来，所有其他自变量效果的计算都会因此而简化。

若读者在此指令中提出一个以上的自变量，则 GLM 程序自动假设右边的变量是镶嵌在左边变量的效果内。

另外有两点，请读者在选用 ABSORB 指令时注意：

一、输入资料文件内的数据，必须依照 ABSORB 指令中列举的变量做由小到大的重新排列。这个步骤可藉 PROC SORT 达成。

二、若选用 ABSORB 指令，则 OUTPUT 指令无效，GLM 程序无法产生输出资料文件。

有关 ABSORB 指令及其使用方法的详细说明，请读者参阅本章第 31.8 节。

指令 #14　BY 变量名称串;

SAS 依据此指令所列举的变量，将资料文件分成几个小的资料文件；然后对每一个小的资料文件分别执行 GLM 分析。当读者选用此指令时，资料文件内的数据必须先依照 BY 变量串的值做由小到大的重新排列，这个步骤可藉 PROC SORT 达成。

指令 #15　FREQ 变量名称;

此变量的值代表资料文件中各观察体重复出现的次数。当这个值小于 1 时，这些观察体的数据便被排除在分析之外。若这个值不是一个整数 (如：5.8)，则 SAS 自动取其整数的部分 (即 5)。

指令 #16　ID 变量名称串;

适用于回归分析，作用是识别各观察体。

指令 #17　WEIGHT 变量名称;

这个指令的作用是将因变量做不等的加权调整。调整的幅度视 WEIGHT 变量的值而定。当读者选用此指令时，实验设计仍然照旧；自由度与样本数不变，但平均数的计算与平均数间的比较会受影响。对参数的估计则由下式导出：

$$\beta = (X'WX)^{-1}X'WY$$

在此式中，W 代表加权值的大小，亦即 WEIGHT 变量的值。

有一种加权值会导致最佳线性不偏估计值 (英文简称 B.L.U.E，即 Best Linear Unbiased Estimates)，这种加权值即等于各组内余差变异数的倒数。

31.4　用 PROC GLM 执行回归分析

若读者想利用 GLM 程序执行 (单/复) 回归分析，则你只需考虑 PROC GLM，MODEL，OUTPUT，ABSORB，BY，FREQ，ID 以及 WEIGHT 等指令。有关这些指令的撰写，读者可参阅第 18 章 PROCREG 的内容以及范例。

31.5　用 PROC GLM 执行单变量变异数分析

若读者只想利用 PROC GLM 执行单变量的变异数分析，则可省略 MANOVA 指令的撰写。其余的指令仍然有效。关于 GLM 程序在单变量变异数分析上的应用，读者可参考本章范例的例二与例四。

31.6　用 PROC GLM 程序执行多变量变异数分析

多变量变异数分析会用到 GLM 程序中所有的十七道指令，所以是最复杂的分析方法。关于 GLM 程序在多变量变异数上的应用，读者可参考本章范例的例五。

31.7　范　　例

例一：平衡的块试验设计与平均数的比较

本资料文件 (PLANTS) 的数据来自 Stenstrom (1940) 的实验。该实验的目的在比较金鱼草在七种土壤 (TYPE) 里生长的情形，每一种土壤中种三盆金鱼草 (BLOCK)。这个实验是一个平衡的设计。现在我们利用此资料文件来示范 GLM 程序。

在 SAS 程序里，由于并没有在 MODEL 指令中指定变异数平方的类型，故 GLM 程序按内设值，自动印出第一和第三型的变异数平方。因为这是一个平衡的设计，故这两型变异数平方的值相等。

在第二个 GLM 程序中，选用了 ORDER=INPUT 选项。因此，在 CONTRAST 指令串中，平均数比较的顺序是根据输入资料文件内各组数据第一次出现的顺序而定的。另外，MODEL 指令中的 SOLUTION 选项要求 GLM 程序列出所有参数的估计值，MEANS 指令则将七种土壤的平均数作两两比较。

程　序

```
DATA PLANTS;
     INPUT TYPE $ @;
```

```
        DO BLOCK=1 TO 3;
            INPUT STEMLENG @;  OUTPUT;   END;   CARDS;
    CLARION  32.7  32.3  31.5
    CLINTON  32.1  29.7  29.1
    KNOX     35.7  35.9  33.1
    O'NEILL  36.0  34.2  31.2
    COMPOST  31.8  28.0  29.2
    WABASH   38.2  37.8  31.9
    WEBSTER  32.5  31.1  29.7
    ;
    PROC GLM ORDER=DATA;
        CLASS TYPE BLOCK;
        MODEL STEMLENG= TYPE BLOCK;
        MEANS TYPE / SNK;
        /*-TYPE-ORDER ------------CLRN-CLTN-KNOX-ONEL-CPST-WBSH-WSTR*/
        CONTRAST  'COMPOST VS OTHERS'   TYPE  -1  -1  -1  -1   6  -1  -1 ;
        CONTRAST  'RIVER SOILS VS.NON'  TYPE  -1  -1  -1  -1   0   5  -1 ,
                                        TYPE  -1   4  -1  -1   0   0  -1 ;
        CONTRAST  'GLACIAL VS DRIFT'    TYPE  -1   0   1   1   0   0  -1 ;
        CONTRAST  'CLARION VS WEBSTER'  TYPE  -1   0   0   0   0   0   1 ;
        CONTRAST  'KNOX VS ONEILL'      TYPE   0   0   1  -1   0   0   0 ;
    RUN;
```

结　果

分析的结果显示：金鱼草的生长情形随七种土壤以及三种盆栽之不同而改变(F=10.80，P=0.0002)。

利用 SNK 的事后检定以及 CONTRAST 指令，我们可以下结论说：WABASH 的土壤最优，COMPOST 的土壤最劣。KNOX 与 O'NEILL 两种不分轩轾；CLARION 与 WEBSTER 两种土壤间亦无显著的差异。

报表 31.1　平衡的块试验设计与平均数的比较

```
                General Linear Models Procedure

                   Class Level Information

Class    Levels    Values

TYPE          7    CLARION CLINTON KNOX O'NEILL COMPOST WABASH WEBSTER
BLOCK         3    1 2 3

                Number of observations in data set = 21
```

Dependent Variable: STEMLENG

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	C.V.
Model	8	142.1885714	17.7735714	10.80	0.0002	0.878079	3.939745
Error	12	19.7428571	1.6452381			Root MSE	STEMLENG Mean
Corrected Total	20	161.9314286				1.282668	32.5571429

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TYPE	6	103.1514286	17.1919048	10.45	0.0004
BLOCK	2	39.0371429	19.5185714	11.86	0.0014

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TYPE	6	103.1514286	17.1919048	10.45	0.0004
BLOCK	2	39.0371429	19.5185714	11.86	0.0014

Student-Newman-Keuls test for variable: STEMLENG

NOTE: This test controls the type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha= 0.05 df= 12 MSE= 1.645238

Number of Means	2	3	4	5	6	7
Critical Range	2.2818421	2.7939265	3.1092353	3.3381252	3.5177134	3.6652997

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	TYPE
A	35.967	3	WABASH
A			
B A	34.900	3	KNOX
B A			
B A C	33.800	3	O'NEILL
B C			
B D C	32.167	3	CLARION
D C			
D C	31.100	3	WEBSTER
D			
D	30.300	3	CLINTON
D			
D	29.667	3	COMPOST

Dependent Variable: STEMLENG

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
COMPOST VS OTHERS	1	29.24198413	29.24198413	17.77	0.0012
RIVER SOILS VS.NON	2	48.24694444	24.12347222	14.66	0.0006
GLACIAL VS DRIFT	1	22.14083333	22.14083333	13.46	0.0032
CLARION VS WEBSTER	1	1.70666667	1.70666667	1.04	0.3285

```
KNOX VS ONEILL          1      1.81500000      1.81500000        1.10     0.3143
```

例二：非平衡型的实验设计：二因子的单变量变异数分析

本资料文件 (A) 的数据由 Kutner (1974) 所提供。其出处是 Afifi 与 Azen (1972) 合著的书：以电脑为工具的统计分析。两个自变量分别是 DRUG 和 DISEASE 。请读者注意在 MODEL 指令中，程序要求四型变异数平方值的打印。

程　序

```
DATA A;

    INPUT DRUG DISEASE @;

    DO I= 1 TO 6;

      INPUT Y @;

      OUTPUT;
    END;
    CARDS;
1 1 42 44 36 13 19 22
1 2 33  . 26  . 33 21
1 3 31 -3  . 25 25 24
2 1 28  . 23 34 42 13
2 2  . 34 33 31  . 36
2 3  3 26 28 32  4 16
3 1  .  .  1 29  . 19
3 2  . 11  9  7  1 -9
3 3 21  1  .  9  3  .
4 1 24  .  9 22 -2 15
4 2 27 12 12 -5 16 15
4 3 22  7 25  5 12  .
;
PROC GLM;
    CLASS DRUG DISEASE;
    MODEL Y=DRUG DISEASE DRUG*DISEASE / SS1 SS2 SS3 SS4;
RUN;
```

结　果

由于本实验属非平衡型的实验设计，因此第一与第二型离差平方和的结果不尽相同，然而第三与第四型离差平方和的结果则相同。虽然它们的数据不尽相等，我们仍然可以下

结论说：DRUG 的效果达显著水准 (p<.0001)，然而 DISEASE (p=.1709) 或两者间的交互作用 (p=.3764) 均未达显著水准。

报表 31.2　非平衡型的实验设计：二因子的单变量变异数分析

General Linear Models Procedure

Class Level Information

Class	Levels	Values
DRUG	4	1 2 3 4
DISEASE	3	1 2 3

Number of observations in data set = 72

NOTE: Due to missing values, only 58 observations can be used in this analysis.

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	C.V.
Model	11	4347.859195	395.259927	3.53	0.0012	0.457752	56.20146
Error	46	5150.416667	111.965580			Root MSE	Y Mean
Corrected Total	57	9498.275862				10.58138	18.8275862

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DRUG	3	3194.609195	1064.869732	9.51	0.0001
DISEASE	2	413.697900	206.848950	1.85	0.1691
DRUG*DISEASE	6	739.552100	123.258683	1.10	0.3764

Source	DF	Type II SS	Mean Square	F Value	Pr > F
DRUG	3	3115.336496	1038.445499	9.27	0.0001
DISEASE	2	413.697900	206.848950	1.85	0.1691
DRUG*DISEASE	6	739.552100	123.258683	1.10	0.3764

Dependent Variable: Y

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DRUG	3	3041.947623	1013.982541	9.06	0.0001
DISEASE	2	411.194966	205.597483	1.84	0.1709
DRUG*DISEASE	6	739.552100	123.258683	1.10	0.3764

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
DRUG	3	3041.947623	1013.982541	9.06	0.0001
DISEASE	2	411.194966	205.597483	1.84	0.1709
DRUG*DISEASE	6	739.552100	123.258683	1.10	0.3764

例三：共变量分析

本资料文件 (DRUG TEST) 的数据由 Snedecor 与 Cochran (1967，P. 422) 所提供。

其目的在探讨两种药物对癫疯病菌的效果。实验效果的代号如下：

DRUG　两种抗生素 (A 与 D) 及一组控制组 (F)

X　　治疗前癫疯病菌的数量

Y　　治疗后癫疯病菌的数量

十位病人被分配到 DRUG 自变量下的各组。癫疯病菌的数量是由每一病人身体上六个部位病菌感染的程度而定的。治疗后的病菌数量 (Y) 是这个实验的因变量，治疗前癫疯病菌的数量 (X)则是 Y 的共变量。

程　序

```
DATA DRUGTEST;
     INPUT DRUG $ X Y @@;
     CARDS;
A  11   6  A   8   0  A  19  11
A   6   4  A  10  13  A   3   0
D   6   0  D   6   2  D  18  18
D   8   4  D  19  14  D  15   9
F  16  13  F  13  10  F  21  23
F  16  12  F  12   5  F  12  20
;
PROC GLM;
     CLASS DRUG;
     MODEL Y=DRUG X;
     LSMEANS DRUG / STDERR PDIFF;
RUN;
```

结　果

利用共变量分析法分析这组资料，结果显示治疗后的癫疯病菌数量明显地受治疗前癫疯病菌的数量所影响 (F=20.21，P=0.0005)。然而，这两种抗生素的效果与控制组 (未接受治疗) 不分上下 (F=0.84，P=0.4516)。请注意，在共变量分析法中，实验效果的检定是根据第三型的离差平方和而非第一型的离差平方和。

报表 31.3　共变量分析

```
                    General Linear Models Procedure

                       Class Level Information

                      Class    Levels    Values

                      DRUG          3    A D F

                    Number of observations in data set = 18

  Dependent Variable: Y
Source          DF   Sum of Squares    Mean Square   F Value   Pr > F   R-Square          C.V.

Model            3      581.6768838     193.8922946     10.68   0.0006   0.695971      46.75927

Error           14      254.1008940      18.1500639                      Root MSE        Y Mean
```

```
Corrected      17      835.7777778                                    4.260289    9.11111111
Total


Source          DF        Type I SS    Mean Square   F Value    Pr > F

DRUG             2      214.7777778    107.3888889      5.92    0.0137
X                1      366.8991060    366.8991060     20.21    0.0005

Source          DF      Type III SS    Mean Square   F Value    Pr > F

DRUG             2       30.5637001     15.2818501      0.84    0.4516
X                1      366.8991060    366.8991060     20.21    0.0005


          DRUG                Y        Std Err      Pr > |T|    LSMEAN
                         LSMEAN         LSMEAN    H0:LSMEAN=0    Number

               A       8.2481907      1.8315796      0.0005         1
               D       7.9946786      1.7396259      0.0004         2
               F      11.0904640      1.8431443      0.0001         3

                  Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

                    i/j          1          2          3
                      1          .     0.9212     0.3154
                      2     0.9212          .     0.2436
                      3     0.3154     0.2436          .
  NOTE: To ensure overall protection level, only probabilities associated with

        pre-planned comparisons should be used.
```

例四：三因子的单变量变异数分析与平均数的比较

本资料文件 (ONE) 的数据由 Cochran 与 Cox (1957，P.176) 所提供。这个实验探讨电击对萎缩肌肉的影响，各项效果的代号如下：

REP	实验重复的次数 (一或二次)
TIME	电流通过的时间 (一到四级时间单位)
CURRENT	电流的强度 (一到四级)
NUMBER	每天受电击的次数 (一到三次)
Y	肌肉经电击后的重量 (因变量)

每一位受试者接受两次电击，但电流通过肌肉的时间，电流的强度与每天接受的次数则因人而异。这个例子的程序主在示范 CONTRAST 指令的写法。

第一个 CONTRAST 指令是针对 TIME 的主效果 (三度自由度，因此三个平均数比较) 以及 CURRENT 在每一 TIME 组内的简单主效果 (Simple Main Effect) 而撰写的。

第二个 CONTRAST 指令 O 则直接比较 CURRENT 的第一与第二组的平均数。

程　序

```
DATA ONE;
     DO REP=1 TO 2;
       DO TIME=1 TO 4;
          DO CURRENT=1 TO 4;
             DO NUMBER=1 TO 3;
```

```
                    INPUT Y @@;
                      OUTPUT;
                    END;
                  END;
                END;
              END;
              CARDS;
        72 74 69 61 61 65 62 65 70 85 76 61
        67 52 62 60 55 59 64 65 64 67 72 60
        57 66 72 72 43 43 63 66 72 56 75 92
        57 56 78 60 63 58 61 79 68 73 86 71
        46 74 58 60 64 52 71 64 71 53 65 66
        44 58 54 57 55 51 62 61 79 60 78 82
        53 50 61 56 57 56 56 56 71 56 58 69
        46 55 64 56 55 57 64 66 62 59 58 88
        ;
        PROC GLM;
              CLASS REP CURRENT TIME NUMBER;
              MODEL Y=REP CURRENT| TIME| NUMBER;
              CONTRAST 'TIME IN CURRENT 3'
              TIME  1 0 0 -1  CURRENT*TIME 0 0 0 0 0 0 0 0 1 0 0 -1,
              TIME  0 1 0 -1  CURRENT*TIME 0 0 0 0 0 0 0 0 0 1 0 -1,
              TIME  0 0 1 -1  CURRENT*TIME 0 0 0 0 0 0 0 0 0 0 1 -1;
              CONTRAST 'CURR 1 VS. CURR 2' CURRENT 1  -1;
        RUN;
```

(上述程序，经我们在不同的 PC 上测试后，发现十六位元的 PC 可能无法顺利执行此程序，这是因为十六位元 PC 的内存空间不足所致。)

结　果

从变异数分析表看来，整个实验设计只有一个效果达显著水准，亦即电流的主效果(F=10.51，P=0.0001)，其余的效果均未达显著水准。

因此，利用 CONTRAST 指令对 TIME 的主效果，以及 CURRENT 在每一 TIME 组内的简单主效果做平均数比较时，其检定的结果均未达显著水准。

最后，CONTRAST 指令比较电流第一组与第二组的平均数差异，结果显示：这两组间的差异未达统计上显著的程度。因此，我们可以推测：电流效果可能存在于一、三、四组间或二、三、四组间或三、四组间。

报表 31.4　三因子的单变量变异数分析与平均数的比较

General Linear Models Procedure

Class Level Information

Class	Levels	Values
REP	2	1 2
CURRENT	4	1 2 3 4
TIME	4	1 2 3 4
NUMBER	3	1 2 3

Number of observations in data set = 96

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	C.V.
						0.643805	13.05105
Model	48	5782.916667	120.477431	1.77	0.0261	Root MSE	Y Mean
Error	47	3199.489583	68.074246			8.250712	63.2187500
Corrected Total	95	8982.406250					

Source	DF	Type I SS	Mean Square	F Value	Pr > F
REP	1	605.010417	605.010417	8.89	0.0045
CURRENT	3	2145.447917	715.149306	10.51	0.0001
TIME	3	223.114583	74.371528	1.09	0.3616
CURRENT*TIME	9	298.677083	33.186343	0.49	0.8756
NUMBER	2	447.437500	223.718750	3.29	0.0461
CURRENT*NUMBER	6	644.395833	107.399306	1.58	0.1747
TIME*NUMBER	6	367.979167	61.329861	0.90	0.5023
CURRENT*TIME*NUMBER	18	1050.854167	58.380787	0.86	0.6276

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	1	605.010417	605.010417	8.89	0.0045
CURRENT	3	2145.447917	715.149306	10.51	0.0001
TIME	3	223.114583	74.371528	1.09	0.3616
CURRENT*TIME	9	298.677083	33.186343	0.49	0.8756
NUMBER	2	447.437500	223.718750	3.29	0.0461
CURRENT*NUMBER	6	644.395833	107.399306	1.58	0.1747
TIME*NUMBER	6	367.979167	61.329861	0.90	0.5023
CURRENT*TIME*NUMBER	18	1050.854167	58.380787	0.86	0.6276

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
TIME IN CURRENT 3	3	34.83333333	11.61111111	0.17	0.9157
CURR 1 VS. CURR 2	1	99.18750000	99.18750000	1.46	0.2334

例五：多变量的变异数分析

本资料文件 (SKULL) 由奥勒冈大学的教授 A.Anderson 提供，旨在决定性别 (SEX) 对四种反应 (LENGTH，BASILAR，ZYGOMAT 与 POSTORB) 的效果。

虚无假设是性别不会影响受试者在这四种反应上的成绩。

程　序

```
DATA SKULL;
     INPUT SEX $ LENGTH BASILAR ZYGOMAT POSTORB @@;
     CARDS;
M  6460  4962  3286  1100  M  6252  4773  3239  1061
M  6264  4806  3179  1054  M  6622  5113  3365  1071
M  6441  4918  3153  1061  M  6281  4821  3133  1071
M  6573  4977  3392  1110  M  6563  5025  3234  1090
M  6535  4939  3261  1065  M  6573  4962  3320  1091
M  6302  4761  3204  1135  M  6449  4921  3256  1068
M  6368  4824  3258  1130  M  6372  4844  3306  1137
M  6229  4746  3257  1153  M  6391  4834  3244  1169
M  6787  5181  3334  1104  M  6384  4834  3195  1064
M  6340  4791  3300  1110  M  6394  4879  3272  1241
M  6348  4886  3160   991  M  6534  4990  3310  1028
F  6287  4845  3218   996  F  6583  4992  3300  1107
F  6432  4790  3249  1117  F  6450  4888  3259  1060
F  6424  4855  3322  1065  F  6615  5088  3280  1179
F  6521  5011  3208   989  F  6416  4889  3200  1001
F  6540  4997  3320  1078  F  6780  5259  3358  1174
F  6472  4954  3125  1178  F  6476  4896  3148  1066
F  6693  5177  3236  1131  F  6328  4792  3214  1018
F  6266  4721  3257  1031  F  6660  5146  3374  1069
F  6331  4819  3278  1008  F  6298  4683  3270  1150
;
PROC GLM;
    CLASS SEX;
    MODEL LENGTH BASILAR ZYGOMAT POSTORB=SEX;
    MANOVA H=SEX / PRINTE PRINTH;
    TITLE 'MULTIVARIATE ANALYSIS OF VARIANCE';
RUN;
```

结　果

当我们检视四个多变量变异数分析的检定时，会发现它们的值经转换成 F 值后，大小均完全一致 (亦即 F=0.8018，P=0.5323)。这个结果与单变量变异数分析的结果是完全相同的 (F 值介于 0.00 与 1.02 间) 。所以，我们可下结论说：男女在这四种反应上的结果是不相上下的。

报表 31.5　极多变量变异数分析

MULTIVARIATE ANALYSIS OF VARIANCE

General Linear Models Procedure

Class Level Information

Class	Levels	Values
SEX	2	F M

Number of observations in data set = 40

Dependent Variable: LENGTH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	C.V.
Model	1	21068.17071	21068.17071	1.02	0.3189	0.026139	2.227958
Error	38	784930.92929	20656.07709			Root MSE	LENGTH Mean
Corrected Total	39	805999.10000				143.7222	6450.85000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEX	1	21068.17071	21068.17071	1.02	0.3189

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEX	1	21068.17071	21068.17071	1.02	0.3189

Dependent Variable: BASILAR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	C.V.
Model	1	11468.21237	11468.21237	0.62	0.4342	0.016174	2.756780
Error	38	697567.76263	18357.04638			Root MSE	BASILAR Mean
Corrected Tota l	39	709035.97500				135.4882	4914.72500

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEX	1	11468.21237	11468.21237	0.62	0.4342

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEX	1	11468.21237	11468.21237	0.62	0.4342

Dependent Variable: ZYGOMAT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	C.V.
Model	1	5.38282827	5.38282827	0.00	0.9726	0.000031	2.060864
Error	38	171189.717172	4504.99255715			Root MSE	ZYGOMAT Mean
Corrected Total	39	171195.100000				67.11924	3256.85000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEX	1	5.38282828	5.38282828	0.00	0.9726

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEX	1	5.38282828	5.38282828	0.00	0.9726

Dependent Variable: POSTORB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square	C.V.
Model	1	2832.272980	2832.272980	0.83	0.3689	0.021300	5.378600
Error	38	130136.702020	3424.650053			Root MSE	POSTORB Mean
Corrected Total	39	132968.975000				58.52051	1088.02500

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SEX	1	2832.272980	2832.272980	0.83	0.3689

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEX	1	2832.272980	2832.272980	0.83	0.3689

E = Error SS&CP Matrix

	LENGTH	BASILAR	ZYGOMAT	POSTORB
LENGTH	784930.92929	701469.40404	191075.85859	71561.838384
BASILAR	701469.40404	697567.76263	148810.80808	44647.494949
ZYGOMAT	191075.85859	148810.80808	171189.71717	32681.676768
POSTORB	71561.838384	44647.494949	32681.676768	130136.70202

Partial Correlation Coefficients from the Error SS&CP Matrix / Prob > |r|

DF = 37	LENGTH	BASILAR	ZYGOMAT	POSTORB
LENGTH	1.000000 0.0	0.947981 0.0001	0.521256 0.0007	0.223906 0.1706
BASILAR	0.947981 0.0001	1.000000 0.0	0.430628 0.0062	0.148185 0.3680
ZYGOMAT	0.521256 0.0007	0.430628 0.0062	1.000000 0.0	0.218960 0.1805
POSTORB	0.223906 0.1706	0.148185 0.3680	0.218960 0.1805	1.000000 0.0

H = Type III SS&CP Matrix for SEX

	LENGTH	BASILAR	ZYGOMAT	POSTORB
LENGTH	21068.170707	15543.94596	-336.7585859	-7724.688384
BASILAR	15543.94596	11468.212374	-248.4580808	-5699.219949
ZYGOMAT	-336.7585859	-248.4580808	5.3828282828	123.47323232
POSTORB	-7724.688384	-5699.219949	123.47323232	2832.2729798

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SS&CP Matrix for SEX E = Error SS&CP Matrix

Characteristic Root	Percent	Characteristic Vector V'EV=1			
		LENGTH	BASILAR	ZYGOMAT	POSTORB
0.0916399077	100.00	-0.00277942	0.00192900	0.00109960	0.00194135
0.0000000000	0.00	-0.00268751	0.00336127	-0.00031297	-0.00055254
0.0000000000	0.00	0.00072144	0.00055117	0.00268193	0.00097546
0.0000000000	0.00	0.00037806	-0.00041005	-0.00019310	0.00186465

Manova Test Criteria and Exact F Statistics for
the Hypothesis of no Overall SEX Effect
H = Type III SS&CP Matrix for SEX E = Error SS&CP Matrix

S=1　M=1　N=16.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.91605299	0.8018	4	35	0.5323
Pillai's Trace	0.08394701	0.8018	4	35	0.5323
Hotelling-Lawley Trace	0.09163991	0.8018	4	35	0.5323
Roy's Greatest Root	0.09163991	0.8018	4	35	0.5323

例六：重复观察的变异数分析

本资料文件 (DOGS) 的数据由 Cole 与 Grizzle (1966) 所提供。实验的目的在研究两个自变量 (药物与狗身体内胺基酸的分泌作用) 与一个因变量 (狗血液中胺基酸的浓度) 的关系。第一个自变量药物 (DRUG) 下分两组：吗啡或 C6H9NO3 麻醉药。第二个自变量胺基酸的分泌作用 (DEPL)下也分两组：健康的与退化的。十六对狗 (有一对资料不全，故分析时将其剔除)，经历四次测量：注射药物后 0 分钟，1 分钟，3 分钟 及 5 分钟(分别以 HIST0，HIST1，HIST3，HIST5 代表)。这些数据经过对数的转换后，成为此分析的因变量数据。

GLM 程序中，选项 NOUNI 抑止了单变量变异数的分析，这是因为重复观察值代表整体的数据。若将它们分开，则会失去彼此在时间上的关系。另外，指令 REPEATED 中，POLYNOMIAL 选项规定将重复观察值作多项式的正交转换。

程　序

```
DATA DOGS;
    INPUT DRUG $ DEPL $ HIST0 HIST1 HIST3 HIST5;
    LHIST0=LOG(HIST0); LHIST1=LOG(HIST1);
    LHIST3=LOG(HIST3); LHIST5=LOG(HIST5);
    CARDS;
MORPHINE  N  .04   .20   .10   .08
MORPHINE  N  .02   .06   .02   .02
MORPHINE  N  .07  1.40   .48   .24
MORPHINE  N  .17   .57   .35   .24
MORPHINE  Y  .10   .09   .13   .14
```

```
MORPHINE  Y  .12   .11   .10    .
MORPHINE  Y  .07   .07   .06   .07
```

```
MORPHINE  Y  .05   .07   .06   .07
TRIMETH   N  .03   .62   .31   .22
TRIMETH   N  .03  1.05   .73   .60
TRIMETH   N  .07   .83  1.07   .80
TRIMETH   N  .09  3.13  2.06  1.23
TRIMETH   Y  .10   .09   .09   .08
TRIMETH   Y  .08   .09   .09   .10
TRIMETH   Y  .13   .10   .12   .12
TRIMETH   Y  .06   .05   .05   .05
;
PROC GLM;
    CLASS DRUG DEPL;
    MODEL LHIST0 LHIST1 LHIST3 LHIST5=DRUG DEPL DRUG*DEPL/NOUNI;
    REPEATED TIME 4(0 1 3 5) POLYNOMIAL / SHORT SUMMARY;
RUN;
```

结　果

经过对数转换后的时间效果达显著水准 (F=24.0326，P=0.0001)；转换后的时间与药物间的交互效果亦达显著水准 (F=5.7832，P=0.0175)；转换后的时间与分泌作用间的交互效果达显著水准 (F=21.3112，P=0.0002)。最后，时间、药物与分泌作用的三因子交互效果也达到统计上的显著水准 (F=12.4775，P=0.0015)。

利用 REPEATED 指令检定未经对数转换的时间之函数 (包括线性的、抛物线的以及三次曲线的)，结果显示：药物、分泌作用以及两者间的交互效果在时间的线性以及抛物线值上均达显著水准 (P<0.05)。然而，时间经过三次曲线转换后，只有分泌作用的效果达统计的显著水准 (P<0.0001)。

报表 31.6　重复观察的变异数分析

```
                General Linear Models Procedure
                   Class Level Information

              Class    Levels    Values

              DRUG          2    MORPHINE TRIMETH
              DEPL          2    N Y

             Number of observations in data set = 12
```

NOTE: Observations with missing values will not be included in this analysis.

Thus, only 11 observations can be used in this analysis.

Repeated Measures Analysis of Variance

Repeated Measures Level Information

Dependent Variable	LHIST0	LHIST1	LHIST3	LHIST5
Level of TIME	0	1	3	5

Manova Test Criteria and Exact F Statistics for

the Hypothesis of no TIME Effect

H = Type III SS&CP Matrix for TIME E = Error SS&CP Matrix

S=1 M=0.5 N=1.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.11053046	13.4121	3	5	0.0079
Pillai's Trace	0.88946954	13.4121	3	5	0.0079
Hotelling-Lawley Trace	8.04727987	13.4121	3	5	0.0079
Roy's Greatest Root	8.04727987	13.4121	3	5	0.0079

Manova Test Criteria and Exact F Statistics for

the Hypothesis of no TIME*DRUG Effect

H = Type III SS&CP Matrix for TIME*DRUG E = Error SS&CP Matrix

S=1 M=0.5 N=1.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.39402509	2.5632	3	5	0.1680
Pillai's Trace	0.60597491	2.5632	3	5	0.1680
Hotelling-Lawley Trace	1.53790945	2.5632	3	5	0.1680
Roy's Greatest Root	1.53790945	2.5632	3	5	0.1680

Manova Test Criteria and Exact F Statistics for

the Hypothesis of no TIME*DEPL Effect

H = Type III SS&CP Matrix for TIME*DEPL E = Error SS&CP Matrix

S=1 M=0.5 N=1.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.11385354	12.9720	3	5	0.0085
Pillai's Trace	0.88614646	12.9720	3	5	0.0085
Hotelling-Lawley Trace	7.78321382	12.9720	3	5	0.0085
Roy's Greatest Root	7.78321382	12.9720	3	5	0.0085

Manova Test Criteria and Exact F Statistics for

the Hypothesis of no TIME*DRUG*DEPL Effect

H = Type III SS&CP Matrix for TIME*DRUG*DEPL E = Error SS&CP Matrix

S=1 M=0.5 N=1.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.25059156	4.9843	3	5	0.0580
Pillai's Trace	0.74940844	4.9843	3	5	0.0580
Hotelling-Lawley Trace	2.99055732	4.9843	3	5	0.0580
Roy's Greatest Root	2.99055732	4.9843	3	5	0.0580

Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DRUG	1	4.81572376	4.81572376	2.23	0.1791
DEPL	1	3.88791503	3.88791503	1.80	0.2217
DRUG*DEPL	1	3.29812165	3.29812165	1.53	0.2565
Error	7	15.12660428	2.16094347		

Univariate Tests of Hypotheses for Within Subject Effects

Source: TIME

DF	Type III SS	Mean Square	F Value	Pr > F	Adj G - GH - F	Pr > F
3	8.51342151	2.83780717	28.32	0.0001	0.0001	0.0001

Source: TIME*DRUG

DF	Type III SS	Mean Square	F Value	Pr > F	Adj G - G	Pr > F H - F
3	1.26187310	0.42062437	4.20	0.0178	0.0467	0.0178

Source: TIME*DEPL

DF	Type III SS	Mean Square	F Value	Pr > F	Adj G - G	Pr > F H - F
3	9.10044033	3.03348011	30.28	0.0001	0.0001	0.0001

Source:TIME*DRUG*DEPL

DF	Type III SS	Mean Square	F Value	Pr > F	Adj G - G	Pr > F H - F
3	1.71675141	0.57225047	5.71	0.0051	0.0215	0.0051

Source: Error(TIME)

DF	Type III SS	Mean Square	
21	2.10407489	0.10019404	Greenhouse-Geisser Epsilon = 0.5672
			Huynh-Feldt Epsilon = 1.0518

TIME.N represents the nth degree polynomial contrast for TIME

Contrast Variable: TIME.1(线性)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	1.50845432	1.50845432	16.99	0.0045
DRUG	1	0.89164165	0.89164165	10.04	0.0157
DEPL	1	0.96103952	0.96103952	10.82	0.0133
DRUG*DEPL	1	1.39320838	1.39320838	15.69	0.0055
Error	7	0.62158198	0.08879743		

Contrast Variable: TIME.2(抛物线)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	3.72723077	3.72723077	28.71	0.0011
DRUG	1	0.34417030	0.34417030	2.65	0.1475
DEPL	1	4.26438450	4.26438450	32.85	0.0007
DRUG*DEPL	1	0.28962340	0.28962340	2.23	0.1789
Error	7	0.90865653	0.12980808		

Contrast Variable: TIME.3(三次曲线)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	3.27773642	3.27773642	39.98	0.0004
DRUG	1	0.02606115	0.02606115	0.32	0.5905

```
DEPL                          1   3.87501631   3.87501631     47.27       0.0002
DRUG*DEPL                     1   0.03391963   0.03391963      0.41       0.5406
Error                         7   0.57383637   0.08197662
```

31.8 注 意 事 项

■ ABSORB 指令及其使用方法

ABSORPTION 是电脑计算法之一，旨在节省计算的时间与内存空间。最适于用在块试验设计的变异数分析 (Block Design)。

请看下面的两个范例：它们的统计模型事实上完全相同，变量 HERD 是一个块试变量 (BlockVariable)。

例 1 (使用 ABSORB 指令)

```
PROC GLM;
      ABSORB HERD;
      CLASS A B;
      MODEL Y=A B A*B;
```

例 2 （不用 ABSORB 指令)

```
    PROC GLM;
         CLASS HERD A B;
         MODEL Y=HERD A B A*B;
```

在例 1 中，当 ABSORB 指令被用在块试变量 HERD 上时，GLM 程序只计算第一型的离差平方和，因此节省了第二、三及四型离差平方和的计算。所以例 1. 的分析会比例 2.更有效率。

另外，读者也可在 ABSORB 指令中同时包含好几个效果。请看下面的三个例子：(例 3. 与例 4. 的统计模型事实上完全相同)

例 3 （不用 ABSORB 指令)

```
PROC GLM;
     CLASS HERD COW A B ;
     MODEL Y=HERD COW(HERD) A B A*B;
```

例 4 （将 ABSORB 指令用在 HERD，COW 变量串上)

```
PROC GLM;
     ABSORB HERD COW;
     CLASS A B;
     MODEL Y=A B A*B;
```

例 5 (将 ABSORB 指令用在 A，B 变量串上)

```
PROC GLM
    ABSORB A B;
    CLASS HERD COW;
    MODEL Y=HERD COW(HERD);
```

例 5 的程序所导出的效果有四，即：A，B(A)，HERD 与 COW(HERD)。其中最后两项，亦即 HERD 与 COW(HERD)，和 ABSORB 的变量 A，B 或 A*B 完全无关。

为什么 ABSORB 指令会节省计算的时间与空间？这是因为经过 ABSORB 指令宣告的效果必从自变量的矩阵 (亦即 X'X) 中剔除。如此，X'X 的行列数减少，其反矩阵的计算时间也相对地减少。

下面的实例显示使用 ABSORB 指令后的益处：这个实验总共有六千八百七十五个自由度。第一种处理法不用 ABSORB 指令，第二种处理法则使用了 ABSORB 指令。

第一处理法 (不用 ABSORB 指令)

```
DATA A;
    DO HERD=1 TO 40;
      N=1+RANUNI(1234567)*60;
      DO COW=1 TO N;
        DO TRTMENT=1 TO 3;
          DO REP=1 TO 2;
            Y=HERD/5+COW/10+TRTMENT+RANNOR(1234567);
            OUTPUT;
          END;
        END;
      END;
    END;
    DROP N;
PROC GLM;
    CLASS HERD COW TRTMENT;
    MODEL Y=HERD COW(HERD) TRTMENT;
RUN;
```

这个分析将会占用 6 Megabytes 的电脑空间，而第二种处理法则会大大节省分析的时间：

第二处理法 (使用 ABSORB 指令)

```
PROC GLM;
    ABSORB HERD COW;
    CLASS TRTMENT;
```

```
        MODEL Y=TRTMENT;
    RUN;
```

因为使用 ABSORB 指令，因此可将自变量矩阵减至 4*4 的正方矩阵，其所占用的电脑空间将大大减少，分析的结果见报表 31.7：

报表 31.7　ABSORB 指令的示范

General Linear Models Procedure

Class Level Information

Class	Levels	Values
TRTMENT	3	1 2 3

Number of observations in data set = 6876

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1147	52049.89684	45.37916	44.17	0.0001
Error	5728	5884.55788	1.02733		
Corrected Total	6875	57934.45472			

R-Square	C.V.	Root MSE	Y Mean
0.898427	12.48196	1.013574	8.12031348

Source	DF	Type I SS	Mean Square	F Value	Pr > F
HERD	39	36230.70920	928.99254	904.28	0.0001
COW(HERD)	1106	11375.42905	10.28520	10.01	0.0001
TRTMENT	2	4443.758590	2221.879295	2162.77	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRTMENT	2	4443.758590	2221.879295	2162.77	0.0

■ 有关重复观察的正确统计分析

重复观察的实验设计所收集的数据可用单变量或多变量的变异数分析法处理。此处，我们将讨论这两种方法在 SAS 程序里的异同。

(1) 单变量变异数分析

首先让我们假设有一资料文件，称作 "OLD"，利用单变量变异数分析的 SAS 程序如下：

```
    PROC GLM DATA=OLD;
        CLASS GROUP SUBJ TIME;
        MODEL Y=GROUP SUBJ(GROUP) TIME
        GROUP*TIME;
        TEST H=GROUP E=SUBJ(GROUP);
```

(2) 极多变量变异数分析

然而，另有一种处理观察体的方法将更节省打字时间。首先我们将同样的数据以另一种排列方式呈现，称此资料文件为 NEW：

GROUP	Y1	Y2	Y3
1	15	19	25
1	21	18	17
2	14	12	16
.	.	.	.
.	.	.	.
3	14	18	16

然后用下列的指令处理以便获得多变量变异数的结果：

```
PROC GLM DATA=NEW;
    CLASS GROUP;
    MODEL Y1-Y3=GROUP/NOUNI;
    REPEATED TIME;
```

若读者有意改用多变量变异数分析法，则可利用下列的指令将 "OLD" 资料文件转换成适合多变量分析的资料文件：

```
PROC SORT DATA=OLD;
    BY GROUP SUBJ;
DATA NEW(KEEP=Y1-Y3 GROUP);
    ARRAY YY {3} Y1-Y3;
    DO I=1 TO 3;
      SET OLD;
      BY GROUP SUBJ;
      YY{TIME}=Y;
      IF LAST.SUBJ THEN RETURN;
    END;
```

同样地，读者可利用另一个统计程序 PROC TRANSPOSE 来调整 "OLD" 资料文件，使它适合多变量的变异数分析：

```
PROC SORT DATA=OLD;
    BY GROUP SUBJ;
PROC TRANSPOSE OUT=NEW(RENAME=(_1=Y1 _2=Y2 _3=Y3));
    BY GROUP SUBJ;
    ID TIME;
```

(3) 重复观察值的五种线性转换

以下，我们将介绍指令 REPEATED 里所包含的五种线性转换方式。这些线性转换都是针对重复观察的变量。转换后的新变量会自动成为多变量异数分析中的因变量。下面介绍五种线性转换的选项：

CONTRAST 选项

是这五种选项的内设值。最适用于控制组与一个或一个以上实验组的比较。比方说：五组动物接受不同的药物治疗。其中第一组是控制组，接受糖水。其他四组是实验组，接受试验中的 "新医疗" 药物。CONTRAST 选项可有效地用第一组与其他四组一一比较。

```
PROC GLM;
    MODEL D1-D5=/NOUNI;
    REPEATED DRUG 5 CONTRAST (1);
```

其转换矩阵如下：

$$M = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

HELMERT 选项

这个线性转换适用于一组与其它组平均数的比较。经此线性转换后，读者就容易找出重复观察值中的临界点 (亦即平均数开始稳定的点)。下面的例子比较男女受试对四个实验的不同反应：

```
PROC GLM;
    CLASS SEX;
    MODEL RESP1-RESP4=SEX/NOUNI;
    REPEATED TRTMENT 4 HELMERT/CANON;
```

这一串指令后面所代表的转换矩阵如下：

$$M = \begin{pmatrix} 1 & -0.3333 & -0.3333 & -0.3333 \\ 0 & 1 & -0.5000 & -0.5000 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

MEAN 选项

MEAN 选项的功用与 CONTRAST 选项十分类似。唯一不同的是所谓的控制组在此不只限于固定的一组而已，而是各组轮流当控制组。所以下面的指令所导出的转换矩阵 (M) 略有不同。本例的数据与 CONTRAST 选项的例子相同：五组动物接受不同的药物治疗。

```
PROC GLM;
    MODEL D1-D5=/NOUNI;
    REPEATED DRUG 5 MEAN;
```

其转换矩阵如下：

$$M = \begin{pmatrix} 1 & -.25 & -.25 & -.25 & -.25 \\ -.25 & 1 & -.25 & -.25 & -.25 \\ -.25 & -.25 & 1 & -.25 & -.25 \end{pmatrix}$$

-.25　-.25　-.25　1　-.25

若读者有意从比较中省略一组，则可将其组别放在括号内，置于 MEAN 选项之后，如：MEAN (5)。如此，第五组将从平均数的比较中省略。

PROFILE 选项

这个线性转换最适用于不同性质的重复观察值。假若有四种不同的教学法分别试用于几所公立学校。由于教学法之间有质与量的差别，PROFILE 选项可提供两两教学法的比较。请看下面的示范：

```
PROC GLM;
    CLASS SCHOOL;
    MODEL T1-T4=SCHOOL/NOUNI;
    REPEATED METHOD 4 PROFILE;
```

其转换矩阵为：

$$M = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

如此，相邻两组 (前后两种教学法) 可做比较。新的变量 (即相邻两组的平均数差) 将以 METHOD.1，METHOD.2，与 METHOD.3 来代表。

POLYNOMIAL 选项

若欲找出平均数的临界点，则读者需检验标准系数。这些标准系数可由选项 CANON 得出。当这些标准系数在某一个重复观察点之后突然变得极小，则这个点就是所谓的临界点。这类型的转换最适用于等量(或等值)的重复观察实验。假使实验设计不符合此条件，则读者必须将不等的值包含在括号里。在一些使用药物的实验或时间序列的连续观察体，这个选项十分有用。本章第 31.7 节的例六与下面的示范就是很好的例子。假如重复变量 DOSE 下分五组：1，2，5，10 和 20 公克；这些不等的药量分别给不同的 GROUP。则选项 POLYNOMIAL 的使用自动导出直线性、抛物线性、三次以及四次式曲线的趋势分析 (Trend Analysis)。

```
PROC GLM;
    CLASS GROUP;
    MODEL R1-R5=GROUP/NOUNI;
    REPEATED DOSE 5(1 2 5 10 20)
            POLYNOMIAL/SUMMARY;
```

前述的指令导出下面的转换矩阵来转换数据：

$$M = \begin{bmatrix} -.4250 & -.3606 & -.1674 & .1545 & .7984 \\ .4349 & .2073 & -.3252 & -.7116 & .3946 \\ -.4331 & .1366 & .7252 & -.5108 & .0821 \\ .4926 & -.7800 & .3743 & -.0936 & .0066 \end{bmatrix}$$

经过转换后的新变量将以 DOSE.1，DOSE.2，DOSE.3，与 DOSE.4 代表。如上所述，其几何的意义分别是直线性、抛物线性、三次式或四次式曲线的分析。

■ 遗漏数据的处理

在单变量变异数分析或者含 MANOVA，REPEATED 指令的多变量变异数分析中，若观察体在任何一个自变量或因变量上有遗漏数据，则 GLM 程序会将此观察体排除在分析之外。

不含 MANOVA 或 REPEATED 指令的多变量变异数分析中，即使观察体在其中之一的因变量上有遗漏数据，PROC GLM 仍会将它包括在分析内。

第 32 章　离差平方和(SS)的四种类型及其函数

32.1　四类型的 SS 是什么

本章介绍 SAS 变异数分析程序中最核心的概念，即四种类型的离差平方和(Sum of Squares)，其定义与统计的检定。

在 SAS 的变异数分析程序里 (如：GLM，VARCOMP，ANOVA)，每一种实验效果的离差平方和都被归纳成第一型、第二型、第三型、第四型。这种分类是便于统计分析的检定，而非统计学上公认的分类方式。若读者对这四型的离差平方和有兴趣，可参考 Freund，Littell 及 Spector（1986）合著的 “SAS System for Linear Models”。

32.2　在变异数分析里，哪些线性函数是可估计的

一般的线性模型均可简化成下列的方程式：

$$Y = X\beta + \in$$

由此导出 $E(Y)=X\beta$，因为残差 (或作余差，即上式中的$\in$) 之平均数等于 0。变异数分析的最终目的就是估计 β 矩阵的元素；或是这些元素的线性组合，如：$L\beta$。到底那一种线型组合 (或线性函数) 才是可估计的 (Estimable)? 这个问题的答案有其充份且必要的条件。现简述如下：

若一组线性组合的系数可使下式成立，则我们说此线性组合是可以估计的：

$$L\beta = E(KY)$$

K 是一组 Y 的线性组合系数。

由于 $E(KY)=E(KX\beta+K\in)=KX\beta=L\beta$，因此 L 的导出是根据 X 矩阵中横列的线性组合而来的。只要这个线性组合存在，则 L 就是可估计的参数线性函数。

进一步说，由于 $X=[X(X'X)^{-1}(X'X)]$，因此 L 也可以由 (X'X) 或 $(X'X)^{-1}(X'X)$ 等矩阵中横列的组合导出。

当 L 的可估计性被建立后，则 β 的估计就十分简单了。根据最小平方误差的理论，β 可用下式估计：

$$\beta = b = (X'X)^{-1}X'Y$$

所以 Lb 也就是 $L\beta$ 的最小平方误差的估计值。若将 Lb 以统计检定的观念来处理，则下列的虚无假设可用 F 检定来考验：

$$H_0: L\beta = 0$$

这个 F 检定的分子是由 Lb 的离差平方和决定的，亦即：

$$SS(Lb)=(Lb)'[L(X'X)^{-1}L']^{-1}Lb$$

F 检定的分母则依不同的变异数分析模型而定。下面三段分别探讨上述理论在一因子及三因子变异数分析与复回归分析等例子里的应用。

32.3　一因子的变异数分析

假如有一个一因子的变异数分析 (自变量下分为三组)，其线性模型如下：

$Y=\mu+A_i+E$，i=1，2，3

另外，假定此实验中有六名被试。则 X 与 β 的矩阵可定义如下：

$$X=\begin{pmatrix}1&1&0&0\\1&1&0&0\\1&0&1&0\\1&0&1&0\\1&0&0&1\\1&0&0&1\end{pmatrix}\qquad \beta=\begin{pmatrix}\mu\\A1\\A2\\A3\end{pmatrix}$$

根据这些定义，X 矩阵的每一横列都可演绎成一个可估计的线性函数 L 。由于 X 的横列间有重复的现象，我们可使之简化，然后定义一个 X* 矩阵：

$$X^*=\begin{pmatrix}1&1&0&0\\1&0&1&0\\1&0&0&1\end{pmatrix}$$

这个 X* 矩阵和 X 矩阵一样，其每一直行均可导出一个 L 函数。

再进一步的推演，可将 X* 转化成 X** 矩阵如下：

$$X^{**}=\begin{pmatrix}1&0&0&1\\0&1&0&-1\\0&0&1&-1\end{pmatrix}$$

到底这三个矩阵之间的关系如何？这个问题的答案可从它们各自对应的 L 函数看出。首先，让我们定义 L1，L2，与 L3 为三个非零的实数 (也就是说，这三个值不可同时等于 0)，则它们所导出的可估计之 L 函数如下：

对 X* 矩阵而言：

L=L1*(1 1 0 0)+L2*(1 0 1 0)+L3*(1 0 0 1)

上式也可以写成：

L=(L1+L2+L3，L1，L2，L3)

对 X** 矩阵而言：

L=(L1，L2，L3，L1-L2-L3)

从上面的例子中，我们或许对所谓的 L 函数有一个粗浅的认识。从 X* 和 X** 矩阵所导出的 L 函数，其第一个元素是另外其他元素的总和。这是因为在变异数分析里，各实验效果 (Treatment Effects) 加起来等于 0；所以，主效果的自由度是组数减 1。任何一个函数若符合上述的条件，则此函数就是一个 L 函数。

然而，并非所有的 L 函数都是最简洁的。那么，到底那一种函数才是最好的呢？在 SAS 的系统里，最简洁的 L 函数是由 $(X'X)^{-1}(X'X)$ 矩阵组合所导出的。

下面让我们继续探讨这个 L 函数与参数估计值的关系。假设有一个一因子的实验设计，其数学模型是

$Y=\mu+A_i+\in$，i=1，2，3

这个模型的可估计函数 Lβ 等于 L1*μ+L2*A1+L3*A2+(L1-L2-L3)*A3。

于是 L=(L1，L2，L3，L1-L2-L3)，而参数 (见本节初的 β 矩阵) 中只有 A1，A2，A3 对我们最有意义。由于 L1=0 (即 μ 对应的系数必须是 0)，其余 L2 与 L3 则可轮流以非零的实数代入。鉴于 L 函数只有两个自由度，因此首先可用 L2=1，L3=0 带入，然后再以 L2=0，L3=1 带入，如此 L 成了一个 2*4 的矩阵如下：

$$L=\begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

这个 L 矩阵与 β 矩阵的相乘积可用来检验：

H_0：A1=A2=A3=0 (这个虚无假设是变异数分析中最基本的)

从 Lβ 所计算出来的平均方值 (MS) 则成为 F 检定的分子。其自由度是 2 (与 L 的两列相对应)。

下面的 L 矩阵也可导出同样的离差平方和：

$$L^*=\begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 1 & -2 & 1 \end{bmatrix}$$

L* 与上述的 L 有两点雷同之处；第一：两个矩阵的横列都是彼此线性独立的。第二：第一直行的元素必须都是 0 (因我们不需要估计 μ 的值)。如此，L* 与 L 的功能是完全一样的，但 L 的结构似乎更简单一点。

32.4　三因子变异数分析与其主效果的参数估计

假如有一个三因子的实验设计内含 A，B，C 三个主效果。五个数据点分属于这三个变量的不同组，整个实验设计如下图所示：

数据点	变　项		
识别号	A	B	C
1	1	2	1
2	1	1	2
3	2	1	3
4	2	2	2
5	2	2	2

从上述的设计中，我们所能导出的 L 函数如下：

参数 (β)	L 的系数
μ	L1
A1	L2
A2	L1-L2
B1	L4

B2	L1-L4
C1	L6
C2	L1+L2-L4-2*L6
C3	-L2+L4+L6

上述的系数只有四个，即 L1，L2，L4 与 L6，因此 L 矩阵的自由度是 4。根据上节的讨论，我们可知任何一个 4*8 的矩阵，只要其四个横列间是线性独立的，则此矩阵的横列就代表 L 函数，所导出的平均方值 (MS) 可用来进行下列的检定：

H_0：$L\beta=0$，或

H_0：μ=A1=A2=B1=B2=C1=C2=C3=0

由于一般研究者对 μ 的检定没有兴趣，我们进一步讨论其他只适合测 A 或 B 或 C 主效果的 L 函数。鉴于这个三因子的实验是一个不平衡的设计 (即各细格人数不等)，下面所形成的函数只能用来检验最高级的假设 [Maximum Rank Hypothesis (MRH)]：

- 如果我们只对 A 效果有兴趣，则 L1=L4=L6=0，同时 L2=0 (因为 C2 与 C3 必须与 0 的系数相对应)。如此 A1-A2 不能被估计，因为 L 矩阵的 Rank 等于 0，而且相对应的平均方值也是 0。
- 如果我们只想估计 B 的主效果，则 L1=L2=L6=0 。由于 C2 与 C3 必须与 0 的系数相对应，所以 L4=0。如此，B1-B2 不能被估计，其理由与上述 A1-A2 不能被估计是相同的。
- 最后 C 变量的参数是唯一可被估计的，因为我们先使 L1=L2=L4=0，其结果则是：

  C1-2*C2+C3=K (不等于 0 的任何实数)。

  这个函数所导出的平均方值可适用于下列假设的检验：

  H_0：C1=2*C2-C3=0

  从上述的讨论看来，在这个不平衡设计的实验中，我们无法估计 A 或 B 的主效果，但可以估计 C 的主效果。

32.5　复回归分析与其统计模型

假若一个复回归分析有三个自变量与一个因变量，其函数关系如下：

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\in$$

让我们进一步假设 (X'X) 是满秩的矩阵，所以参数的最小误差估计值存在。在这些条件下，L 函数的元素如下表所示：

参数 (β)	L 系数
β_0	L1
β_1	L2
β_2	L3
β_3	L4

若读者只想检定某一个主效果，如 β_2 (亦即 H_0：β_2=0)，则你可定 L1=L2=L4=0，

然后定 L3=1。则此函数所导出来的平均方值可用来检验上述的虚无假设，此检定的自由度为 1 (因为只有一个参数被估计)。一般而言，只要回归分析中的 (X'X) 矩阵是满秩的，则每一个参数，以及它们之间的线性组合，都是可估计的。

反之，若回归分析中的自变量之间不互相独立；比方说：X3=2*X1+3*X2，则其 L 函数的结构与上述的会截然不同，请看下表：

参数 (β)	L 系数
β_0	L1
β_1	L2
β_2	L3
β_3	2*L2+3*L3

这个例子中的参数 β_0 是可估计的。然而 β_1，β_2 与 β_3 等参数则不可估计，因为它们之间不独立，有函数的关系存在，由此所导出的条件离差平方和 (Conditional Sum of Squares) 等于 0。

32.6　第一型离差平方和与其函数

利用上述回归分析的统计模型 (亦即 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\in$)，让我们对第一型离差平方和作如下的定义：

参数 (β)	第一型离差平方和 (SSI)
β_1	SS (β_1)
β_2	SS (β_2|β_1)
β_3	SS (β_3|β_1 , β_2)

从上表中，我们可知第一型离差平方和与模型效果的前后顺序有关。因为此型的平方和是所谓条件的平方和。所以 β_2 效果是依 β_1 效果的大小作调整的。同理，β_3 是对 β_1 与 β_2 两个效果作调整的。

与上述第一型平方和相对应的 L 函数的形成如下：

L1=(X1'X1|X1'X2|X1'X3)

L2=(0|X2'M1X2|X2'M1X3)

L3=(0|0|X3'M2X3)

其中，

$M1=1 - X1(X1'X1)^{-1}X1'$

$M2=M1 - M1X2(X2'M1X2)^{-1}X2'M1$

总而言之，第一型平方和有下列几个特点：

●对参数 β_3 的估计排除掉其他两个参数 β_1 与 β_2。

●对参数 β_2 的估计往往 (特别在不平衡的实验设计下) 牵涉到 β_3，但排除 β_1。

●对参数 β_1 的统计检定牵涉另外两个参数 β_2 与 β_3 的估计值。

●适用于以下三种实验设计：

平衡的变异数分析，镶嵌式的变异数分析，与多项式复回归分析。

32.7　第二型离差平方和与其函数

根据第 32.6 节所述的回归分析统计模型 (亦即 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+E$)，第二型离差平方和的定义如下：

参数 (β)	第二型离差平方和 (SSⅡ)
β_1	SS (β_1|β_2，β_3)
β_2	SS (β_2|β_1，β_3)
β_3	SS (β_3|β_1，β_2)

如果用三因子 (A，B，C) 变异数分析的符号来表示，则第二型平方和的定义如下：

效　果	第二型离差平方和 (SSⅡ)
A	SS (A|B，C)
B	SS (B|A，C)
C	SS (C|A，B)

由上述定义看来，第二型平方和的导出将排除其他参数的估计。所以不适用于有交互作用的变异数分析，镶嵌式变异数分析，或多项式的复回归分析。请看下面所举的几个例子：

- 若有一个二因子变异数分析，其模型如下：
  MODEL Y= A　B　A*B;
  则第二型平方和的计算分别是：
  SS(A|μ，B)，SS(B|μ，A)，及 SS(A*B|μ，A，B)
- 若是镶嵌式变异数分析，其模型如下：
  MODEL Y=A　B(A)　C(A B);
  则第二型平方和的计算分别是：
  SS(A|μ)，SS(B(A)|μ，A)，SS(C(A B)|μ，A，B(A))
- 若有一个一元二次的复回归分析，其模型如下：
  MODEL Y=X　X*X;
  则第二型平方和的计算分别是：
  SS(X|μ，X*X)，SS(X*X|μ，X)

除了前面所述一般性的应用外，让我们进一步看第二型平方和在二因子 (2*2) 变异数分析上的定义。一个标准的二因子变异数分析通常有三个效果，即：A ，B (主效果) 和 A*B (交互效果)。所导出的参数与其 L 函数的系数分别列举在表一：

表一

参数 (β)	L 函数的系数
μ	L1
A1	L2
A2	L1-L2
B1	L4
B2	L1-L4

AB11	L6
AB12	L2-L6
AB21	L4-L6
AB22	L1-L2-L4+L6

如果这个实验是平衡的设计，则上表可展开成表二如下：

表二

对应的系数			
效果	A	B	A*B
μ	0	0	0
A1	L2	0	0
A2	-L2	0	0
B1	0	L4	0
B2	0	-L4	0
AB11	.5L2	.5L4	L6
AB12	.5L2	-.5L4	-L6
AB21	-.5L2	.5L4	-L6
AB22	-.5L2	-.5L4	L6

如果这个实验是不平衡的设计；假设最后一个细格 (即 AB22) 只有一人，其他三细格各有两人，则表一可展开成表三如下：

表三

对应的系数			
效果	A	B	A*B
μ	0	0	0
A1	L2	0	0
A2	-L2	0	0
B1	0	L4	0
B2	0	-L4	0
AB11	.6L2	.6L4	L6
AB12	.4L2	-.6L4	-L6
AB21	-.6L2	.4L4	-L6
AB22	-.4L2	-.4L4	L6

比较表二与表三之后，读者可发现第二型平方和的计算视实验设计的平衡与否而定。特别在检定 A 与 B 的主效果时，不平衡的设计使第二型平方和的计算受各组 (细格) 人数多少或比例的影响。

综合以上的定义和例子，我们建议你将第二型平方和用在下列几种实验设计里：

- 完全平衡的设计
- 只牵涉主效果的设计
- 纯粹的回归分析，以及
- 任何具独立性，不牵涉其他效果的实验效果

32.8　第三型离差平方和与其函数

欲了解第三型的离差平方和，让我们首先示范到底其 L 函数系数是如何决定的? 假设有一个二因子的 (2*2) 实验设计，其 A*B 的第三型系数如下：

效果	一般表示法	A*B 的第三型系数
μ	L1	0
A1	L2	0
A2	L1-L2	0
B1	L4	0
B2	L1-L4	0
AB11	L6	L6
AB12	L2-L6	-L6
AB21	L4-L6	-L6
AB22	L1-L2-L4+L6	L6

上表的第三直行 (即 A*B 的第三型系数) 是依据下列的步骤建立的：

(1) 将所有与 μ，A，B 主效果有关的系数定为 0，所以 L1=L2=L4=0。

(2) 将 L1=L2=L4=0 带入四个 A*B 效果的一般表示，即可得到 A*B 的第三型系数。

现将上面的原则用在主效果 A 的第三型系数上：

(1) 将与 A 效果完全无关的函数系数订为 0，因此 L1(μ 的系数) 与 L4 (B 效果系数) 等于 0。

(2) 慎重选择 L6 与 L2 的值，使 A 与 A*B 的系数是正交的。在此，设 L6=.5L2。

这两个步骤产生如下的系数表：

效果	一般表示法	A 效果的第三型系数
μ	L1	0
A1	L2	L2
A2	L1-L2	-L2
B1	L4	0
B2	L1-L4	0
AB11	L6	.5L2
AB12	L2-L6	.5L2
AB21	L4-L6	-.5L2
AB22	L1-L2-L4+L6	-.5L2

下面让我们看第三型的平方和如何应用在一个不平衡的实验设计：N1-N6 代表六组的人数，它们的值不尽相同但均非零。

让我们进一步假设一个二因子 (3*3) 的实验设计的人数分配如下表所示：

	B 变项		
A	1	2	3
变		N1	N2
量	N3		N4
	N5	N6	

由于上列的实验设计中，有三细格的人数均为 0，而且它们正在这个 3*3 矩阵的主轴上。所以，这个实验设计的第三型函数系数如下：

对应的系数

效果	A	B	A*B
μ	0	0	0
A1	L2	0	0
A2	L3	0	0
A3	-L2-L3	0	0
B1	0	L5	0
B2	0	L6	0
B3	0	-L5-L6	0
AB12	.667L2+.333L3	.333L5+.667L6	L8
AB13	.333L2-.333L3	-.333L5-.667L6	-L8
AB21	.333L2+.667L3	.667L5+.333L6	-L8
AB23	-.333L2+.333L3	-.667L5-.333L6	L8
AB31	-.333L2-.667L3	.333L5-.333L6	L8
AB32	-.667L2-.333L3	-.333L5+.333L6	-L8

32.9　第四型离差平方和与其函数

试看下列的不平衡 3*3 实验设计：N1-N5 代表各细格的人数；它们的值不尽相同但均不等于 0。

		B 变项		
		1	2	3
A 变量	1	N1	N1	
	2	N3	N4	
	3			N5

与上述实验设计相对应的第四型函数系数如下：

对应的系数

效果	A	B	A*B
μ	0	0	0
A1	-L3	0	0
A2	L3	0	0
A3	0	0	0
B1	0	L5	0
B2	0	-L5	0
B3	0	0	0
AB11	-.5L3	.5L5	L8
AB12	-.5L3	-.5L5	-L8
AB21	.5L3	.5L5	-L8
AB22	.5L3	-.5L5	L8
AB33	0	0	0

32.10　四型离差平方和的比较

■二因子单变量变异数分析的四型 SS

效果	第一型	第二型	第三型 (＝第四型)
A	SS(α |μ)	SS(α |μ,β)	SS(α |μ,β,αβ)
B	SS(β |μ,α)	SS(β |μ,α)	SS(β |μ,α,αβ)
A*B	SS(αβ |μ,α,β)	SS(αβ |μ,α,β)	SS(αβ |μ,α,β)

■不同变异数分析的四型 SS

下表中，罗马数字 Ⅰ，Ⅱ，Ⅲ 及 Ⅳ 分别代表第一、二、三及四型离差平方和：

效果	平衡设计	不平衡但等比的设计	不平衡设计(无空组)	不平衡且有空组的设计
A	Ⅰ＝Ⅱ＝Ⅲ＝Ⅳ	Ⅰ＝Ⅱ,Ⅲ＝Ⅳ	Ⅰ≠Ⅱ,Ⅲ＝Ⅳ	Ⅰ≠Ⅱ≠Ⅲ≠Ⅳ
B	Ⅰ＝Ⅱ＝Ⅲ＝Ⅳ	Ⅰ＝Ⅱ,Ⅲ＝Ⅳ	Ⅰ＝Ⅱ,Ⅲ＝Ⅳ	Ⅰ＝Ⅱ,Ⅲ≠Ⅳ
A*B	Ⅰ＝Ⅱ＝Ⅲ＝Ⅳ	Ⅰ＝Ⅱ＝Ⅲ＝Ⅳ	Ⅰ＝Ⅱ＝Ⅲ＝Ⅳ	Ⅰ＝Ⅱ＝Ⅲ＝Ⅳ