第六章 方差分析与协方差分析

在实际中经常要对在不同条件下进行试验或观察得到的数据进行分析，以判断不同条件对结果有无影响。此时协变量往往表示某种效应(条件)存在(成立)与否，因而往往取 0，1 两个值。在统计学上称这类分析为方差分析(analysis of variance，ANOVA)。

在方差分析中：

- 称试验结果或观察指标为响应；
- 在试验或观察中改变其状态时对响应有影响的因素称为因子(factor)，常用字母 $A,B,C\cdots$ 表示；因子在试验或观察中所取的状态称为因子的水平(level)，因子 A 的水平常记为 $A_1,A_2\cdots$；
- 假定在同一条件下的试验结果来自正态分布的一个样本；不同条件下正态总体是相互独立的，期望可能不同但方差相同。
- 要判断不同条件对响应有无影响就是要检验各个正态总体的期望是否相等。

6.1 单向分类模型

设在试验或观察中只有一个因子 A，假定因子 A 有 a 个水平，在 $A_i(i=1,\cdots,a.)$ 水平下获得 n_i 个响应值 $y_{i1},\cdots,y_{i,n_i}$，它们来自正态总体 $N(\mu_i,\sigma^2)$ 的样本且 a 个正态总体相互独立。其模型有两种表达：

1. 均值模型

$y_{ij}=\mu_i+e_{ij}$，$i=1,\cdots,a.$，$j=1,\cdots,n_i$，$\{e_{ij}\}$ 为 $i.i.d\sim N(0,\sigma^2)$，要检验的假设为 H_0：$\mu_1=\mu_2=\cdots=\mu_a$。

2. 主效应模型

记 $N=\sum_{i=1}^{a}n_i$，$\mu=\dfrac{\sum_{i=1}^{a}n_i\mu_i}{N}$ 称为总平均，$\alpha_i=\mu_i-\mu$ 称为因子 A 的第 i 水平的效应。

注意有约束条件 $\sum_{i=1}^{a}n_i\alpha_i=0$，此时模型可以表为

$y_{ij}=\mu+\alpha_i+e_{ij}$，$i=1,\cdots,a.$，$j=1,\cdots,n_i$，$\sum_{i=1}^{a}n_i\alpha_i=0$，$\{e_{ij}\}$ 为 $i.i.d\sim N(0,\sigma^2)$，要检验的假设为 H_0：$\alpha_1=\alpha_2=\cdots=\alpha_a$。

将主效应模型写成矩阵形式，设计矩阵为

$$X_{N\times(a+1)}=\begin{pmatrix} E_{n_1} & E_{n_1} & 0 & \cdots & 0\\ E_{n_2} & 0 & E_{n_2} & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ E_{n_a} & 0 & 0 & \cdots & E_{n_a}\end{pmatrix},$$

$Y=(y_{11},\cdots,y_{1n_1},y_{21},\cdots,y_{2n_2}\cdots y_{a1},\cdots,y_{an_a})'$，相应的有误差向量 e，$\beta=(\mu,\alpha_1,\cdots\alpha_a)'$，$L=(0,n_1,\cdots n_a)$，主效应模型写成矩阵形式：

$$\begin{cases}Y=X\beta+e\\ L\beta=0\end{cases},\quad e\sim N(0,\sigma^2I_N)\ 。$$

由于 $rank(X)=a$，$\mu(X')\cap\mu(L')=\{0\}$，$rank\begin{pmatrix}X\\L\end{pmatrix}=a+1$，故约束 $L\beta=0$ 为 side condition。由正规方程 $X'X\beta=X'Y$ 得到

$$N\mu+\sum_{i=1}^{a}n_i\alpha_i=y_{..}$$

$$n_i\mu+n_i\alpha_i=y_{i.},\quad i=1,\cdots,a.$$

加上约束条件：$\sum_{i=1}^{a}n_i\alpha_i=0$，得到参数估计：

$\hat{\mu}=\bar{y}_{..}$，$\hat{\alpha}_i=\bar{y}_{i.}-\bar{y}_{..}$，$i=1,\cdots,a$。

注：对一般单向分类模型 $y_{ij}=\mu+\alpha_i+e_{ij}$，$i=1,\cdots,a.$，$j=1,\cdots,n_i$，上述估计并不是 μ,α_i 的无偏估计，因为 μ,α_i 是不可估的。由于 $rank(X)=a$ 不是列满秩的，此时至多有 a 个线性无关的可估函数，$\mu+\alpha_i$，$i=1,\cdots,a.$ 已为 a 个线性无关的可估函数，故任可估函数都可表为它们的线性组合。

此外，由于 $L\beta=0$ 为 side condition，加上此条件后，使得模型参数 μ,α_i 条件可估，上述估计是条件无偏估计。

定义 6.1.1：称线性函数 $\sum_{i=1}^{a}c_i\alpha_i$ 为一个对照 (contrast)，如果 $\sum_{i=1}^{a}c_i=0$。

注：对照不含 μ。

定理 6.1.1：对于一般单向分类模型 $\sum_{i=1}^{a}c_i\alpha_i$ 可估 $\Leftrightarrow\sum_{i=1}^{a}c_i\alpha_i$ 为一个对照，此时其 BLUE 为 $\sum_{i=1}^{a}c_i\hat{\alpha}_i$，即 $\sum_{i=1}^{a}c_i\bar{y}_{i.}$。

对单向分类模型，感兴趣的是考察因 A 的 a 个水平效应是否有显著差异，即检验 H_0：$\alpha_1=\alpha_2=\cdots=\alpha_a$ 或等价检验假设 H_0：$\alpha_1-\alpha_a=\cdots=\alpha_{a-1}-\alpha_a=0$。由于 $\alpha_i-\alpha_a$

$1\le i\le a-1$ 为 $a-1$ 个对照，即 $a-1$ 个可估函数，构造 F 统计量来进行检验。残差平方和

$$ESS=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\hat{\mu}-\hat{\alpha}_i)^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2,$$

在假设下，$ESS_H=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{..})^2$，检验 F 统计量为 $F=\dfrac{(ESS_H-ESS)/(a-1)}{ESS/(N-a)}$，当 H_0 为真时 $F\sim F_{a-1,N-a}$ 分布，故给定水平 α，当

$F>F_{a-1,N-a}(\alpha)$ 时拒绝 H_0，认为因子 A 的水平效应有显著差异。记总和 $TSS=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{..})^2$，则 $TSS=ESS+SS_A$，

这里 $SS_A=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\bar{y}_{..})^2=\sum_{i=1}^{a}n_i(\bar{y}_{i.}-\bar{y}_{..})^2$，是由因子 A 的水平变化所引起的观测数据的变差平方和，常称为因子 A 的平方和。若把 A 的每个水平 A_i 下的数据看成一组，则

SS_A也称为组间平方和，ESS也称为组内平方和。令$MS_A = SS_A/(a-1)$，$MS_E = ESS/(N-a)$，称为平均平方和，简称为均方和，此时$F = \frac{MS_A}{MS_E}$。注意到$\hat{\sigma}^2 = MS_E$为σ^2的无偏估计。通常把上述要计算的结果列成表格，称为方差分析表。

单因素方差分析表

来源	自由度	平方和	均方和	F值
因子A	$a-1$	SS_A	MS_A	$F = \frac{MS_A}{MS_E}$
误差E	$N-a$	ESS	MS_E	
总和T	$N-1$	TSS		

当因子A的各水平效应有显著差异时，常常需要进一步对一切$i \neq j$同时检验如下$\frac{a(a-1)}{2}$个假设H_0^{ij}：$\mu_i = \mu_j$，H_1^{ij}：$\mu_i \neq \mu_j$。该检验称为多重比较(multiple comparison)。也等价于对所有效应之差$\alpha_i - \alpha_j (i \neq j)$作出置信区间，若某个$\alpha_i - \alpha_j$的置信区间不含有0点，则表明$A_i, A_j$水平效应有显著差异。

一般地，对m个对照$\sum_{i=1}^{a} c_i^{(k)} \alpha_i (k = 1,2,\cdots m)$，需要作置信水平为$1-\alpha$的同时置信区间。

由 Bonferroni 方法得到同时置信区间为：

$$\sum_{i=1}^{a} c_i^{(k)} \bar{y}_{i.} \pm t_{N-a}\left(\frac{\alpha}{2m}\right)\hat{\sigma}\sqrt{\sum_{i=1}^{a}\frac{\left[c_i^{(k)}\right]^2}{n_i}}, k = 1,2,\cdots m。$$

特别对$m = \frac{a(a-1)}{2}$个形如$\alpha_i - \alpha_j$的对照的置信水平为$1-\alpha$的 Bonferroni 区间为：

$$(\bar{y}_{i.} - \bar{y}_{j.}) \pm t_{N-a}\left(\frac{\alpha}{2m}\right)\hat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}, 1 \le i \neq j \le a。$$

通常对A的水平较少时$(a = 2,3,4)$有效。

所有对照$\sum_{i=1}^{a} c_i \alpha_i$的置信系数为$1-\alpha$的 Scheffe 区间为：

$$\sum_{i=1}^{a} c_i \bar{y}_{i.} \pm \hat{\sigma}\sqrt{(a-1)F_{a-1,N-a}(\alpha)\sum_{i=1}^{a}\frac{c_i^2}{n_i}}。$$

特别对$m = \frac{a(a-1)}{2}$个形如$\alpha_i - \alpha_j$的对照的置信水平为$1-\alpha$的 Scheffe 区间为：

$$(\bar{y}_{i.} - \bar{y}_{j.}) \pm \hat{\sigma}\sqrt{(a-1)F_{a-1,N-a}(\alpha)\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}，$$

$1 \le i \neq j \le a$。

Tukey 的方法

定义 6.1.1：设$z_1,\cdots,z_n \sim N(0,1)$，$mW^2 \sim \chi_m^2$且所以随机变量相互独立，称随机变量

$$Q_{n,m}=\frac{\max\limits_{1\le i\le n} z_i-\min\limits_{1\le i\le n} z_i}{W}$$

的分布为参数n,m的学生化级差分布(studentized range distribution)。(其 α 分位点的表格可以参见 Christensen. R(1996) Analysis of Variance, Design and Regression: Applied Statistical Methods. Chapman & Hall, London.)

构造同时置信区间的 Tukey 方法：

设$y_i \sim N(\mu_i,\sigma^2)$，$i=1,\cdots,n$，$m\frac{\hat\sigma^2}{\sigma^2}\sim\chi_m^2$且所有随机变量相互独立，则所有$\mu_i-\mu_j(i\ne j)$的置信系数为$1-\alpha$的同时置信区间为

$$y_i-y_j\pm\hat\sigma Q_{n,m}(\alpha)。$$

注：Tukey 方法要求正态分布的方差相同。

对于单向分类模型，Tukey 方法只能用于平衡数据，即$n_1=\cdots=n_a=n$，此时$\bar y_{i.}\sim N(\mu+\alpha_i,\frac{\sigma^2}{n})$，$i=1,\cdots,a$，$\frac{(N-a)\hat\sigma^2}{\sigma^2}\sim\chi_{N-a}^2$，有所有随机变量相互独立，故对一切$\alpha_i-\alpha_j(i\ne j)$的置信系数为$1-\alpha$的同时置信区间为

$$\bar y_{i.}-\bar y_{j.}\pm\frac{\hat\sigma}{\sqrt n}Q_{a,N-a}(\alpha)。$$

进一步可以证明，对单向分类平衡模型，所有对照$\sum_{i=1}^a c_i\alpha_i$的置信系数为$1-\alpha$的Tukey 区间为

$$\sum_{i=1}^a c_i\bar y_{i.}\pm Q_{a,N-a}(\alpha)\frac{\hat\sigma}{2\sqrt n}\sum_{i=1}^a|c_i|。$$

6.2 两向分类模型

设在试验或观察中有两个因子A与 B，假定因子A有a个水平，因子B有 b个水平，在因子A取第i个水平，因子B取第j个水平时(简记为条件A_iB_j)响应服从正态分布$N(\mu_{ij},\sigma^2)$，记总平均为$\mu=\sum_{i=1}^a\sum_{j=1}^b\mu_{ij}/(ab)$，

又记$\bar\mu_{i.}=\sum_{j=1}^b\mu_{ij}/b$，$\bar\mu_{.j}=\sum_{i=1}^a\mu_{ij}/a$。

称$\alpha_i=\bar\mu_{i.}-\mu$为因子$A$的第 i个水平的效应(有约束$\sum_{i=1}^a\alpha_i=0$)，称$\beta_j=\bar\mu_{.j}-\mu$为因子B的第 j个水平的效应(有约束$\sum_{j=1}^b\beta_j=0$)。

若对一切$i=1,\cdots,a$，$j=1,\cdots,b$有

$$\mu_{ij}=\mu+\alpha_i+\beta_j$$

则此时模型称为效应可加模型，否则记

$$\gamma_{ij}=\mu_{ij}-\mu-\alpha_i-\beta_j$$

称为因子 A的第 i个水平与因子 B的第 j个水平的交互效应(interaction)(有约束 $\sum_{i=1}^{a}\gamma_{ij}=0,\ j=1,\cdots,b,\ \sum_{j=1}^{b}\gamma_{ij}=0,\ i=1,\cdots,a$)。此时模型称为有交互作用的模型。

6.2.1 效应可加模型的方差分析

此时在每种水平组合下只要进行一次试验即可，若记在A_iB_j下试验结果为 y_{ij}，则模型为：

$y_{ij}=\mu+\alpha_i+\beta_j+e_{ij},\ i=1,\cdots,a,\ j=1,\cdots,b,$ e_{ij} 相互独立且$e_{ij}\sim N(0,\sigma^2)$。写成矩阵形式

$$Y=X\beta+e,\ e\sim N(0,\sigma^2 I_{ab}),$$

其中$\beta=(\mu,\alpha_1,\cdots,\alpha_a,\beta_1,\cdots,\beta_b)'$，$rank(X)=a+b-1$，

$$X_{ab\times(a+b+1)}=\begin{pmatrix} E_b & E_b & & & I_b \\ E_b & & E_b & & I_b \\ \vdots & & & \ddots & \vdots \\ E_b & & & E_b & I_b \end{pmatrix}$$

正规方程$X'X\beta=X'Y$为：

$$ab\mu+b\sum_{i=1}^{a}\alpha_i+ab\sum_{j=1}^{b}\beta_j=y_{..}$$

$$b\mu+b\alpha_i+\sum_{j=1}^{b}\beta_j=y_{i.},\ i=1,\cdots,a.$$

$$a+\sum_{i=1}^{a}\alpha_i+a\beta_j=y_{.j},\ j=1,\cdots,b.$$

$\sum_{i=1}^{a}\alpha_i=0,\ \sum_{j=1}^{b}\beta_j=0$为 side condition.

参数估计：

$$\hat{\mu}=\bar{y}_{..},$$

$$\hat{\alpha}_i=\bar{y}_{i.}-\bar{y}_{..},\ i=1,\cdots,a.$$

$$\hat{\beta}_j=\bar{y}_{.j}-\bar{y}_{..},\ j=1,\cdots,b.$$

因此

$$ESS=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2,$$

$\hat{\sigma}^2=\dfrac{ESS}{(a-1)(b-1)}$为$\sigma^2$的无偏估计。

定理 6.2.1：对无交互效应的两向分类模型：

1.$\sum_{i=1}^{a}c_i\alpha_i$可估$\Leftrightarrow\sum_{i=1}^{a}c_i\alpha_i$是一个对照，此时其 BLUE 为$\sum_{i=1}^{a}c_i\bar{y}_{i.}$；

2.$\sum_{j=1}^{b}d_j\beta_j$可估$\Leftrightarrow\sum_{j=1}^{b}d_j\beta_j$是一个对照，此时其 BLUE 为$\sum_{j=1}^{b}d_j\bar{y}_{.j}$。

感兴趣的检验主要有下面两个：

$$H_{0A}:\ \alpha_1=\cdots=\alpha_a$$

$$H_{0B}:\ \beta_1=\cdots=\beta_b$$

或等价

$$H_{0A}:\ \alpha_1-\alpha_a=\cdots=\alpha_{a-1}-\alpha_a=0$$

$$H_{0B}:\ \beta_1-\beta_b=\cdots=\beta_{b-1}-\beta_b=0$$

在假设H_{0A}下，

$$ESS_{H_{0A}}=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j})^2。$$

当H_{0A}成立时，

$$F_A=\frac{(ESS_{H_{0A}}-ESS)/(a-1)}{ESS/\left[(a-1)(b-1)\right]}\sim F_{a-1,(a-1)(b-1)},$$

给定水平α，当$F_A>F_{a-1,(a-1)(b-1)}(\alpha)$时拒绝$H_{0A}$，否则接受。

同样的，在假设H_{0B}下，

$$ESS_{H_{0B}}=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.})^2。$$

当H_{0B}成立时，

$$F_B=\frac{(ESS_{H_{0B}}-ESS)/(b-1)}{ESS/\left[(a-1)(b-1)\right]}\sim F_{b-1,(a-1)(b-1)},$$

给定水平α，当$F_B>F_{b-1,(a-1)(b-1)}(\alpha)$时拒绝$H_{0B}$，否则接受。

令

$$SS_A=SS_{H_{0A}}-ESS=\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{i.}-\bar{y}_{..})^2,$$

$$SS_B=SS_{H_{0B}}-ESS=\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2,$$

$$TSS=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{..})^2,$$

则有$TSS=SS_A+SS_B+ESS$。

无重复试验无交互作用两因素方差分析表

来源	平方和	自由度	均方和	F值
因子 A	SS_A	$a-1$	MS_A	$F_A=\frac{MS_A}{MS_E}$
因子 B	SS_B	$b-1$	MS_B	$F_B=\frac{MS_B}{MS_E}$
误差 E	ESS	$(a-1)(b-1)$	MS_E	
总和 T	TSS	$ab-1$		

若经F检验，假设H_{0A}被拒绝，表明因子A的a个水平的效应有显著差异，与单向分类模型一样，希望构造对照$\alpha_i-\alpha_{i'}$的同时置信区间，类似的若假设H_{0B}被拒绝时，构造对照$\beta_j-\beta_{j'}$的同时置信区间。

Bonferroni 区间：

任m个$\alpha_i-\alpha_{i'}(i\neq i')$置信系数为$1-\alpha$的同时置信区间为

$$(\bar{y}_{i.}-\bar{y}_{i'.})\pm t_{(a-1)(b-1)}\left(\frac{\alpha}{2m}\right)\hat{\sigma}\sqrt{\frac{2}{b}},$$

任m个$\beta_j-\beta_{j'}(j\neq j')$置信系数为$1-\alpha$的同时置信区间为

$$(\bar{y}_{.j}-\bar{y}_{.j'})\pm t_{(a-1)(b-1)}\left(\frac{\alpha}{2m}\right)\hat{\sigma}\sqrt{\frac{2}{a}}。$$

Scheffe 区间：

所有形如$\alpha_i-\alpha_{i'}(i\neq i')$的对照置信系数为$1-\alpha$的同时置信区间为

$$(\bar{y}_{i.}-\bar{y}_{i'.})\pm\hat{\sigma}\sqrt{(a-1)\frac{2}{b}F_{(a-1)(b-1)}(\alpha)},$$

所有形如$\beta_j-\beta_{j'}(j\neq j')$的对照置信系数为$1-\alpha$的同时置信区间为

$$(\bar{y}_{.j}-\bar{y}_{.j'})\pm\hat{\sigma}\sqrt{(b-1)\frac{2}{a}F_{(a-1)(b-1)}(\alpha)}。$$

Tukey 区间：

所有对照$\alpha_i-\alpha_{i'}(i\neq i')$的置信系数为$1-\alpha$的同时置信区间为

$$(\bar{y}_{i.}-\bar{y}_{i'.})\pm Q_{a,(a-1)(b-1)}(\alpha)\frac{\hat{\sigma}}{\sqrt{b}},$$

所有对照$\beta_j-\beta_{j'}(j\neq j')$的置信系数为$1-\alpha$的同时置信区间为

$$(\bar{y}_{.j}-\bar{y}_{.j'})\pm Q_{b,(a-1)(b-1)}(\alpha)\frac{\hat{\sigma}}{\sqrt{a}}。$$

6.2.2 有交互作用模型的方差分析

此时在每种水平的组合下必须要进行多次试验才可以进行数据分析。通常在每种水平的组合下重复进行的试验次数相同，

设均为 c次(称为平衡数据)，记在A_iB_j条件下试验第 k次结果为y_{ijk}，则模型可以表为

$$y_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+e_{ijk},$$

其中$i=1,\cdots,a$， $j=1,\cdots,b$，$k=1,\cdots,c$，e_{ijk}独立同分布，共同分布为$N(0,\sigma^2)$。若将模型写成矩阵形式$Y=X\beta+e$，则X 为$abc\times(ab+a+b+1)$矩阵，$rank(X)=ab$。

正规方程为

$$abc\mu+bc\sum_{i=1}^{a}\alpha_i+ac\sum_{j=1}^{b}\beta_j+c\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij}=y_{...},$$

$$bc\mu+bc\alpha_i+c\sum_{j=1}^{b}\beta_j+c\sum_{j=1}^{b}\gamma_{ij}=y_{i..},1\le i\le a,$$

$$ac\mu+c\sum_{i=1}^{a}\alpha_i+ac\beta_j+c\sum_{i=1}^{a}\gamma_{ij}=y_{.j.},1\le j\le b,$$

$$c\mu+c\alpha_i+c\beta_j+c\gamma_{ij}=y_{ij.},1\le i\le a,1\le j\le b。$$

(独立个数只有最后的ab个方程)

要附加$a+b+ab+1-ab=a+b+1$个 side condition，可选为

$$\begin{cases}\sum_{i=1}^{a}\alpha_i=0\\ \sum_{j=1}^{b}\beta_j=0\\ \sum_{j=1}^{b}\gamma_{ij}=0,1\le i\le a\\ \sum_{i=1}^{a}\gamma_{ij}=0,1\le j\le b\end{cases}$$

此时得到参数估计

$$\hat{\mu}=\bar{y}_{...}，$$

$$\hat{\alpha}_i=\bar{y}_{i..}-\bar{y}_{...}，\ 1\le i\le a，$$

$$\hat{\beta}_j=\bar{y}_{.j.}-\bar{y}_{...}，\ 1\le j\le b，$$

$\hat{\gamma}_{ij}=\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}$，$1\le i\le a$，$1\le j\le b$。

$ESS=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(y_{ijk}-\bar{y}_{ij.})^2$，自由度为$abc-rank(X)=ab(c-1)$，（$c>1$）故$\hat{\sigma}^2=ESS/\left[ab(c-1)\right]$为$\sigma^2$的无偏估计。

定理 6.2.2：对有交互效应平衡数据两向分类模型，下列ab个函数构成了极大线性无关的可估函数组：

$$\alpha_i-\alpha_{i+1}+\bar{\gamma}_{i.}-\bar{\gamma}_{i+1.}，\ 1\le i\le a-1，$$

$$\beta_j-\beta_{j+1}+\bar{\gamma}_{.j}-\bar{\gamma}_{.j+1}，\ 1\le j\le b-1，$$

$$\gamma_{ij}-\bar{\gamma}_{i.}-\bar{\gamma}_{.j}+\bar{\gamma}_{..}，\ 1\le i\le a-1，\ 1\le j\le b-1，$$

$$\mu+\sum_{i=1}^{a}\alpha_i/a+\sum_{j=1}^{b}\beta_j/b+\bar{\gamma}_{..}。$$

对有交互效应的两向分类模型，α_i并不能反映因子水平A_i的优劣，因为可能与因子B的水平有关，不同的B_j水平下A_i的优劣不一样，因此对模型首先要检验交互效应是否存在，即检验

$$H_{0A\times B}：\ \gamma_{ij}=0，\ 1\le i\le a，\ 1\le j\le b。$$

对于模型$y_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+e_{ijk}$，$r_{ij}$不是可估的，但加上前述的 side condition 即考虑有约束的线性模型时，r_{ij}是条件可估的。

应用第 4 章有初始约束的假设检验的结果

$$F=\frac{(ESS_{LH_{0A\times B}}-ESS_L)/(m_1-m_2)}{ESS_L/(n-m_1)}$$

其中$m_1=rank\begin{pmatrix}X\\L\end{pmatrix}-rank(L)=ab$，$n=abc$，

$$m_2=rank\begin{pmatrix}X\\L\end{pmatrix}-rank\begin{pmatrix}L\\H\end{pmatrix}=a+b-1\quad，$$

$$ESS_L=ESS=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(y_{ijk}-\bar{y}_{ij.})^2$$

在$H_{0A\times B}$下得残差平方和$ESS_{LH_{0A\times B}}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(y_{ijk}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$，令$SS_{A\times B}=ESS_{LH_{0A\times B}}-ESS_L$，则

$$SS_{A\times B}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2。$$

在$H_{0A\times B}$下，

$$F_{A\times B}=\frac{SS_{A\times B}/[(a-1)(b-1)]}{ESS/[ab(c-1)]}\sim F_{(a-1)(b-1),ab(c-1)}$$分布。

给定水平 α，当$F_{A\times B}>F_{(a-1)(b-1),ab(c-1)}(\alpha)$时拒绝$H_{0A\times B}$。

关于因子效应得检验，考虑有约束条件下的模型，要检验如下两个假设

$$H_{0A}：\ \alpha_1=\cdots=\alpha_a=0；$$

$$H_{0B}：\ \beta_1=\cdots=\beta_b=0；$$

都是条件可估函数(注：H_{0A}只有$a-1$个独立的检验，H_{0B}只有$b-1$个独立的检验)。

假设H_{0A}的F检验统计量为

$$F_A=\frac{(ESS_{LH_{0A}}-ESS_L)/(m_1-m_2)}{ESS_L/(n-m_1)}$$

此处$n=abc$，$m_1=ab$，$m_2=ab-a+1$，在H_{0A}下，模型为

$$\begin{cases}y_{ijk}=\mu+\beta_j+\gamma_{ij}+e_{ijk},\\ \sum_{j=1}^{b}\beta_j=0,\sum_{j=1}^{b}\gamma_{ij},1\le i\le a-1,\sum_{i=1}^{a}\gamma_{ij}=0,1\le j\le b.\end{cases}$$

此时最小二乘解为 $\hat{\mu}=\bar{y}_{...}$，$\hat{\beta}_j=\bar{y}_{.j.}-\bar{y}_{...}$，$\hat{\gamma}_{ij}=\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}$，因此得到 $ESS_{LH_{0A}}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(y_{ijk}-\bar{y}_{ij.}+\bar{y}_{i..}-\bar{y}_{...})^2$，令 $SS_A=ESS_{LH_{0A}E}-ESS_L=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{y}_{i..}-\bar{y}_{...})^2$。在 H_{0A}下，$F_A=\dfrac{SS_A/(a-1)}{ESS/[ab(c-1)]}\sim F_{a-1,ab(c-1)}$分布，给定水平 α，当$F_A>F_{a-1,ab(c-1)}(\alpha)$时拒绝$H_{0A}$。

同理检验假设H_{0B}的F统计量为

$$F_B=\frac{SS_B/(b-1)}{ESS/[ab(c-1)]},$$

其中 $SS_B=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{y}_{.j.}-\bar{y}_{...})^2$。在 H_{0B} 下，$F_B=\dfrac{SS_B/(b-1)}{ESS/[ab(c-1)]}\sim F_{b-1,ab(c-1)}$分布，给定水平 α，当$F_B>F_{b-1,ab(c-1)}(\alpha)$时拒绝$H_{0B}$。

令$TSS=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(y_{ijk}-\bar{y}_{...})^2$，则由于

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(y_{ijk}-\bar{y}_{...})^2$$
$$=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(y_{ijk}-\bar{y}_{ij.})^2+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$$
$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{y}_{i..}-\bar{y}_{...})^2+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{y}_{.j.}-\bar{y}_{...})^2$$
$$TSS=ESS+SS_{A\times B}+SS_A+SS_B$$

有交互效应平衡数据两因素方差分析表

来源	自由度	平方和	均方和	F值
因子 A	$a-1$	SS_A	MS_A	$F_A=\dfrac{MS_A}{MS_E}$
因子 B	$b-1$	SS_B	MS_B	$F_B=\dfrac{MS_B}{MS_E}$
交互效应 $A\times B$	$(a-1)(b-1)$	$SS_{A\times B}$	$MS_{A\times B}$	$F_{A\times B}=\dfrac{MS_{A\times B}}{MS_E}$
误差 E	$ab(c-1)$	ESS	MS_E	
总和 T	$abc-1$	TSS		

在各种条件A_iB_j下，响应变量的均值(指标均值)μ_{ij}的无偏估计如下：

当交互作用显著时：

$\hat{\mu}_{ij}=\hat{\mu}+\hat{\alpha}_i+\hat{\beta}_j+\hat{\gamma}_{ij}=\bar{y}_{ij.}$，方差为$\dfrac{\sigma^2}{c}$；

当交互作用不显著而两因子显著时：

$\hat{\mu}_{ij}=\hat{\mu}+\hat{\alpha}_i+\hat{\beta}_j=\bar{y}_{i..}+\bar{y}_{.j.}-\bar{y}_{...}$，方差为$\dfrac{[1+(a-1)+(b-1)]\sigma^2}{abc}$；

当交互作用不显著且因子B也不显著时：

$\hat{\mu}_{ij}=\hat{\mu}+\hat{\alpha}_i=\bar{y}_{i..}$，方差为$\dfrac{\sigma^2}{bc}$；

当交互作用不显著且因子A也不显著时：

$\hat{\mu}_{ij}=\hat{\mu}+\hat{\beta}_j=\bar{y}_{.j.}$，方差为$\dfrac{\sigma^2}{ac}$；

当交互作用不显著且两因子都不显著时：

$\hat{\mu}_{ij}=\hat{\mu}=\bar{y}_{...}$，方差为$\dfrac{\sigma^2}{abc}$；

由于$\hat{\sigma}^2 = MS_E$与$\hat{\mu}_{ij}$相互独立，因此μ_{ij}的置信水平为$1-\alpha$的置信区间为：

$$\hat{\mu}_{ij} \pm t_{ab(c-1)}\left(\frac{\alpha}{2}\right)\hat{\sigma}\sqrt{\frac{1+\text{显著因子自由度之和}}{abc}}。$$

6.4 单向分类模型的正态性与方差齐性检验

在前面所有模型假设中，都假设观测误差e满足：诸分量相互独立；正态分布；方差齐性(homogeneity of variance，每次测量值的方差相等)。若某一条假设不满足，检验假设的统计量一般不服从F分布，此时方差分析的结果就不可靠，甚至会导致错误的结论。一般说来对某个具体问题，只要在试验过程中随机化实现的好，试验结果相互独立性一般是容易满足的。

正态性检验

对单向分类模型$y_{ij} = \mu + \alpha_i + e_{ij}$，$i = 1,\cdots,a$，$j = 1,\cdots,n_i$。记残差$\hat{e}_{ij} = y_{ij} - \bar{y}_{i.}$，则$E\hat{e}_{ij} = 0$，$Var(\hat{e}_{ij}) = \frac{n_i - 1}{n_i}\sigma^2$，

$$E\hat{e}_{ij}\hat{e}_{i'j'} = \begin{cases} 0, i \neq i' \\ -\frac{\sigma^2}{n_i}, i = i', j \neq j' \end{cases}。$$

在同一水平下，残差方差相同但不独立，在不同水平下残差方差不等但相互独立。若作如下线性变换：

$$z_{il} = \sqrt{\frac{l}{l+1}}\left(\frac{1}{l}\sum_{j=1}^{l}\hat{e}_{ij} - \hat{e}_{i,j+1}\right) = \sqrt{\frac{l}{l+1}}\left(\frac{1}{l}\sum_{j=1}^{l}y_{ij} - y_{i,j+1}\right),$$

$l = 1,\cdots,n_i - 1$；$i = 1,\cdots,a$。

将$N = \sum_{i=1}^{a} n_i$个残差变为$N - a$个z_{il}。这$N - a$个z_{il}满足$Ez_{il} = 0$，$Var(z_{il}) = \sigma^2$，$Cov(z_{il}, z_{i'l'}) = 0$(除非$i = i', l = l'$)。把$\{z_{il}\}$看作从$N(0,\sigma^2)$总体中抽出的一组独立样本，用通常检验误差正态分布的方法作检验。

方差齐性的检验

理论研究表明，当正态性假定不满足时对F检验影响较小(除非偏离正态程度很严重)，但F检验对方差齐性的偏离较为敏感。把单因素a个水平当作a个正态总体，$y_{ij} = \mu + \alpha_i + e_{ij}$，$j = 1,\cdots,n_i$，$i = 1,\cdots,a$，$e_{ij}$相互独立，$e_{ij} \sim N(0,\sigma_i^2)$。要检验的假设为：

H_0：$\sigma_1^2 = \cdots = \sigma_a^2$；$H_1$：$\sigma_i^2$不全相等。

1. Levene 检验

只用于平衡数据 $n_1 = n_2 = \cdots = n$，记 $\hat{e}_{ij} = y_{ij} - \bar{y}_{i.}$，$l_{ij} = \hat{e}_{ij}^2$，$SS_{组内} = \sum_{i=1}^{a}\sum_{j=1}^{n}(l_{ij} - \bar{l}_{i.})^2$，$SS_{组间} = \sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{l}_{i.} - \bar{l}_{..})^2$，在 H_0 下近似的有 $L = \frac{a(n-1)}{a-1}\frac{SS_{组间}}{SS_{组内}} \sim F_{a-1,a(n-1)}$ 分布，给定水平 α，当 $L > F_{a-1,a(n-1)}(\alpha)$ 拒绝 H_0。

2.Hartley 检验(最大 F 比检验)

设第 i 个水平误差平方和为 $SS_{E_i} = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$，在正态分布下 $SS_{E_i} \sim \sigma^2\chi^2_{n_i-1}$ 分布，令 $MS_{E_i} = SS_{E_i}/(n_i - 1)$，$F_{\max} = \frac{\max\limits_{1\le i\le a} MS_{E_i}}{\min\limits_{1\le i\le a} MS_{E_i}}$，当 $n_1 = n_2 = \cdots = n$ 时，在 H_0 下 $F_{\max} \sim F_{\max_{a,n-1}}$ 分布，若 n_i 不全等近似的用 $F_{M,m}$ 分布代替(其中 M, m 为达到极值时 MS_{E_i} 的自由度)。

3.Cochran 检验(最大方差检验)

只用于平衡数据 $n_1 = n_2 = \cdots = n$，令 $G_{\max} = \frac{\max MS_{E_i}}{\sum_{i=1}^{a} MS_{E_i}}$，在 H_0 下 $G_{\max} \sim G_{\max_{a,n-1}}$，给定水平 α，查表得临界值 $G_{\max_{a,n-1}}(\alpha)$，当 $G_{\max} > G_{\max_{a,n-1}}(\alpha)$ 拒绝 H_0。

4.Bartlett 检验

令 $MS_E = \frac{1}{N-a}\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$，

$MS_{E_i} = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2/(n_i - 1)$，

$C = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a}\frac{1}{n_i - 1} - \frac{1}{N-a}\right)$，

$B = \frac{1}{C}\left[(N-a)\ln MS_E - \sum_{i=1}^{a}(n_i - 1)\ln MS_{E_i}\right]$

Bartlett 证明在大样本下，$B \sim \chi^2_{a-1}$，给定水平 α，当 $B > \chi^2_{a-1}(\alpha)$，拒绝 H_0。一般在 n_i 都 ≥ 5 时，使用该检验是适当得。

5.修正的 Bartlett 检验

针对样本量< 5时，不能使用 Bartlett 检验的缺点，Box 提出修正的 Bartlett 检验统计量，令 $\tilde{B} = \frac{f_2 BC}{f_1(A - BC)}$，这里 B, C 如上，$f_1 = a - 1$，$f_2 = \frac{a+1}{(C-1)^2}$，$A = \frac{f_2}{2 - C + 2/f_2}$。

Box 证明了在 H_0 下统计量 $\tilde{B}$ 的分布近似为 F_{f_1,f_2} 分布，对给定水平 α，当 $\tilde{B} > F_{f_1,f_2}(\alpha)$ 时拒绝 H_0。通常 f_2 的值不是整数，此时可通过 F 分布的分位数表进行内插法得到。

6.5 协方差分析

在方差分析中，假定除可控变量因素外其他因素对响应变量没有影响。当知道有些协变量也会影响响应变量，却不能够控制或不感兴趣时，可以在实验处理前予以观测，然后在排除这些协变量对观测变量影响的条件下，分析可控变量因素对观测变量的作用，从而更加准确地对控制因素进行评价。

这样影响试验结果(响应变量) y的因素既有像方差分析中所讨论的定性因素$A,B\cdots$，又有像回归分析中所涉及的定量变量$z_1,z_2\cdots$时，可采用协方差分析的方法(analysis of covariance，ANCOVA)。定性的因素称为因子，定量的变量称为协变量。

一般协方差分析模型为：

$Y=X\beta+Z\gamma+e$，$e\sim N(0,\sigma^2 I_n)$。

$X=(x_{ij})_{n\times p}$，其元素非 0 即 1，$rank(X)=r$，β为效应向量，$Z=(z_{ij})_{n\times q}$为协变量，一般假定$rank(Z)=q$，$\mu(X)\cap\mu(Z)=\{0\}$，γ为回归系数。记$P_X=X(X'X)^{-}X'$，为$\mu(X)$上的投影矩阵。

性质 1：$Z'(I_n-P_X)Z$可逆，$rank[(I_n-P_X)Z]=q$。

性质 2：$\mu(X\vdots Z)=\mu(X\vdots(I_n-P_X)Z)$。

性质 3：设P为$\mu(X\vdots Z)$上的投影矩阵，则$P=P_X+(I_n-P_X)Z[Z'(I_n-P_X)Z]^{-1}Z'(I_n-P_X)$。

最小二乘解β^*,γ^*应使得$X\beta^*+Z\gamma^*=PY$。令

$$\gamma^*=[Z'(I_n-P_X)Z]^{-1}Z'(I_n-P_X)Y,$$

则$E\gamma^*=\gamma$，$E(Y-Z\gamma^*)=X\beta$，故回归系数γ是可估的。设相对应的纯方差分析模型(称为子模型)为$Y=X\beta+e$，对纯方差分析模型来说若$c'\beta$可估，则$\exists b$，$c=X'b$，此时对协方差模型来说$c'\beta$也是可估的，因为$b'(Y-Z\gamma^*)$为其无偏估计。令$\hat\beta=(X'X)^{-}X'Y$为纯方差分析模型最小二乘解，即$X\hat\beta=P_XY$，令

$$\beta^*=\hat\beta-(X'X)^{-}X'Z\gamma^*,$$

则$X\beta^*+Z\gamma^*=PY$，故由上定义的β^*,γ^*为协方差分析模型的最小二乘解。且易计算得$Var(\gamma^*)=\sigma^2[Z'(I_n-P_X)Z]^{-1}$，$Cov(\hat\beta,\gamma^*)=0$，$Cov(\beta^*,\gamma^*)=-\sigma^2(X'X)^{-}X'Z[Z'(I_n-P_X)Z]^{-1}$，记$X_Z=(X'X)^{-}X'Z$，则$\beta^*=\hat\beta-X_Z\gamma^*$，$Var(\beta^*)=\sigma^2\left[(X'X)^{-}(X'X)(X'X)^{-}+X_Z[Z'(I_n-P_X)Z]^{-1}X_Z'\right]$。

残差平方和：

$$\begin{aligned}ESS^*&=Y'(I_n-P)Y\\&=Y'(I_n-P_X)Y-Y'(I_n-P_X)Z[Z'(I_n-P_X)Z]^{-1}Z'(I_n-P_X)Y\\&=ESS-\gamma^{*}Z'(I_n-P_X)Y\end{aligned}$$

其中，ESS为纯方差分析模型的残差平方和。

由于$rank(P)=r+q$，故$\hat\sigma^2=\dfrac{ESS^*}{n-r-q}$为协方差分析模型误差方差$\sigma^2$的无偏估计。

下面研究假设检验问题。对协方差分析模型主要感兴趣的检验有两个，一是对方差分析部分的假设H_{01}：$H\beta=0$，另一个是检验协变量是否有影响，即检验H_{02}：$\gamma=0$。首先看检验H_{01}：$H\beta=0$。在约束$H\beta=0$下，设纯方差分析模型的解为$\hat{\beta}_H$，残差平方和为$ESS_H=Y'Y-\hat{\beta}_H'X'Y=Y'QY$，这里$Q=I_n-P_M$，其中$P_M$为空间$\{X\beta|H\beta=0,\beta\in R^p\}$上的投影矩阵。

设ESS_H^*为协方差分析模型在约束$H\beta=0$下的残差平方和，由于ESS_H^*与ESS_H的关系与ESS^*与ESS的关系一样，因此

$$ESS_H^*=Y'QY-(Y'QZ)(Z'QZ)^{-1}(Z'QY),$$

这里与前面性质 1 一样，$Z'QZ$可逆。因此检验H_{01}的F统计量为

$$F_1=\frac{(ESS_H^*-ESS^*)/m}{ESS^*/(n-r-q)},$$

这里$m=rank(H)$。

当H_{01}成立时，$F_1\sim F_{m,n-r-q}$，因此给定水平α，当$F_1>F_{m,n-r-q}(\alpha)$时拒绝H_{01}。

对于假设H_{02}，检验的F统计量为

$$F_2=\frac{(ESS-ESS^*)/q}{ESS^*/(n-r-q)},$$

当H_{02}成立时，$F_2\sim F_{q,n-r-q}$，因此给定水平α，当$F_2>F_{m,n-r-q}(\alpha)$时拒绝H_{02}。

6.6 含一个协变量的两向分类模型

此时协方差分析模型为：

$y_{ij}=\mu+\alpha_i+\beta_j+\gamma z_{ij}+e_{ij}$，$i=1,\cdots,a$，$j=1,\cdots,b$，$e_{ij}\sim N(0,\sigma^2)$且所有误差$e_{ij}$相互独立，$\sum_{i=1}^{a}\alpha_i=\sum_{j=1}^{b}\beta_j=0$(side condition)。相对应的纯方差分析模型为$y_{ij}=\mu+\alpha_i+\beta_j+e_{ij}$，$i=1,\cdots,a$，$j=1,\cdots,b$。

纯方差分析模型，残差平方和

$$ESS=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2=Y'N_XY,$$

其中$N_X=I_n-P_X$，因此

$$Z'N_XY=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})(z_{ij}-\bar{z}_{i.}-\bar{z}_{.j}+\bar{z}_{..}),$$

$$Z'N_XZ=\sum_{i=1}^{a}\sum_{j=1}^{b}(z_{ij}-\bar{z}_{i.}-\bar{z}_{.j}+\bar{z}_{..})^2。$$

回归系数γ的估计为

$$\gamma^*=[Z'N_XZ]^{-1}Z'N_XY$$

$$=\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})(z_{ij}-\bar{z}_{i.}-\bar{z}_{.j}+\bar{z}_{..})}{\sum_{i=1}^{a}\sum_{j=1}^{b}(z_{ij}-\bar{z}_{i.}-\bar{z}_{.j}+\bar{z}_{..})^2}$$

对于纯方差模型此时估计为$\hat{\beta}=(X'X)^{-}X'Y$，这里

$$\hat{\beta}=(\bar{y}_{..},\bar{y}_{1.}-\bar{y}_{..},\cdots\bar{y}_{a.}-\bar{y}_{..},\bar{y}_{.1}-\bar{y}_{..},\cdots\bar{y}_{.b}-\bar{y}_{..})'$$

因此，
$(X'X)^{-}X'Z=(\bar{z}_{..},\bar{z}_{1.}-\bar{z}_{..},\cdots\bar{z}_{a.}-\bar{z}_{..},\bar{z}_{.1}-\bar{z}_{..},\cdots\bar{z}_{.b}-\bar{z}_{..})'$。
从而对协方差模型，设估计为$\beta^{*}=(\mu^{*},\alpha_1^{*},\cdots\alpha_a^{*},\beta_1^{*},\cdots\beta_b^{*})$，则

$$\mu^{*}=\bar{y}_{..}-\gamma^{*}\bar{z}_{..},$$

$$\alpha_i^{*}=\bar{y}_{i.}-\bar{y}_{..}-\gamma^{*}(\bar{z}_{i.}-\bar{z}_{..})，\ 1\le i\le a.,$$

$$\beta_j^{*}=\bar{y}_{.j}-\bar{y}_{..}-\gamma^{*}(\bar{z}_{.j}-\bar{z}_{..})，\ 1\le j\le b.。$$

此时，协方差分析模型残差平方和

$$ESS^{*}=ESS-\gamma^{*}(Z'N_XY)$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$$

$$-\frac{\left[\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})(z_{ij}-\bar{z}_{i.}-\bar{z}_{.j}+\bar{z}_{..})\right]^2}{\sum_{i=1}^{a}\sum_{j=1}^{b}(z_{ij}-\bar{z}_{i.}-\bar{z}_{.j}+\bar{z}_{..})^2}$$

。

对任意对照$\sum_{i=1}^{a}c_i\alpha_i,\sum_{j=1}^{b}d_j\beta_j$都是可估的，其BLUE分别为$\sum_{i=1}^{a}c_i(\bar{y}_{i.}-\gamma^{*}\bar{z}_{i.}),\sum_{j=1}^{b}d_j(\bar{y}_{.j}-\gamma^{*}\bar{z}_{.j})$。
特别，$i\ne i',j\ne j'$，$\alpha_i-\alpha_{i'},\beta_j-\beta_{j'}$的BLUE分别为$\bar{y}_{i.}-\bar{y}_{i'.}-\gamma^{*}(\bar{z}_{i.}-\bar{z}_{i'.})$，$\bar{y}_{.j}-\bar{y}_{.j'}-\gamma^{*}(\bar{z}_{.j}-\bar{z}_{.j'})$，方差分别为$\sigma^2\left[\frac{2}{b}+\frac{(\bar{z}_{i.}-\bar{z}_{i'.})^2}{Z'N_XZ}\right]$，$\sigma^2\left[\frac{2}{a}+\frac{(\bar{z}_{.j}-\bar{z}_{.j'})^2}{Z'N_XZ}\right]$。

利用上面结果可以给出$i\ne i',j\ne j'$，$\alpha_i-\alpha_{i'},\beta_j-\beta_{j'}$的各种同时置信区间。

对此模型，关心三个检验问题：

H_{01}：$\alpha_1=\cdots=\alpha_a$；

H_{02}：$\beta_1=\cdots=\beta_b$；

H_{03}：$\gamma=0$。

在假设H_{01}下，纯方差模型残差平方和为

$$ESS_{H_1}=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j})^2=Y'QY，$$

因此$Y'QZ=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j})(z_{ij}-\bar{z}_{.j})$，

$Z'QZ=\sum_{i=1}^{a}\sum_{j=1}^{b}(z_{ij}-\bar{z}_{.j})^2$，从而在假设$H_{01}$下，协方差分析模型的残差平方和为

$$ESS_{H_1}^{*}=Y'QY-Y'QZ(Z'QZ)^{-1}(Z'QY)$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j})^2-\frac{\left[\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{.j})(z_{ij}-\bar{z}_{.j})\right]^2}{\sum_{i=1}^{a}\sum_{j=1}^{b}(z_{ij}-\bar{z}_{.j})^2}$$

此检验的F统计量为

$$F_1=\frac{(ESS_{H_1}^{*}-ESS^{*})/(a-1)}{ESS^{*}/(ab-a-b)},$$

给定水平α，当$F_1>F_{a-1,ab-a-b}(\alpha)$时拒绝$H_{01}$。同理，可以得到其他两个检验的$F$统计量。