重复测量设计的分析

严 琦

(华东师范大学统计系,上海,200062)

[摘要] 给出重复测量设计中等相关系数情况的模型,并根据其特殊的协差阵,通过正交变换,将多元线性模型的检验问题化为 G-M 模型的检验问题,从而节省了自由度,提高了效率.

[关键词] 重复测量设计;正交变换;G-M 模型

[中图分类号]O212.4; [文献标识码]A; [文章编号]1001-4616(2001)04-0043-04

0 引言

重复测量设计(Repeated Measures)是对同一因变量进行重复测度.可以是同一条件下的进行的重复测度,目的在于研究各种处理间是否存在显著性差异的同时,研究受试者之间的差异、受试者几次测量之间的差异以及测试者与各种处理之间的交互效应.也可以是不同条件下的重复测度,目的在于研究各种处理之间是否存在显著性差异的同时,研究形成重复测量条件间的差异以及这些条件与处理间的交互效应.

例如在教育心理研究中,为比较受试者对三种视觉刺激(处理)的反应时(因变量),用一个重复测量设计去获得受试者之间与受试者之内(一个受试者的几次测量间)的变异的比较.若受试者之间变异大于受试者之内的变异,视觉刺激反应时的实验是可行的,否则是不可行的.再例如研究社会某些条件对人类特定方面的特性的影响,往往在某一地区采样 100,另一地区采样 100….两个地区的社会条件对研究对象来说可能是独立的,但同一地区的 100 个个体同处于相同社会条件下,彼此并不独立,此时需重复测量的处理方法.

在重复测量设计的方差分析中总偏差平方和被分解为处理间的偏差平方和、受试者之间的偏差平方和、受试者之内的偏差平方和.这些偏差平方和除以各自的自由度得到效应的均方.它们与误差均方之商即为 F 检验的 F 值.本文将从线性模型的角度,导出一些结论.

1 模型

设有 n 个个体,来自 G 个组,每个个体重复测量 p 次 $n = n_1 + n_2 + \cdots + n_G$,其中 n_i 为来自第 i 组的个体的个数.

数据结构:

$$y_{ij} = \mu_{ij} + \alpha_i^I + \alpha_{ij}^{IO} + \varepsilon_{ij}, i = 1, \cdots, n, j = 1, \cdots, p.$$

y_{ij}:第 i 个个体第 j 次测量值.

收稿日期:2001-03-17

作者简介:严琦,1975—,女,华东师范大学统计系硕士研究生,主要从事概率统计的学习与研究.

μ_{ij} :固定效应.

若 $i, i' \in (g)$,则 $\mu_{ij} = \mu_{i'j} = \mu_j^{(g)}$,$\alpha_i^I$,$\alpha_{ij}^{IO}$,$\varepsilon_{ij}$为随机误差.

α_i^I 独立同分布,服从 $N(0, \sigma_I^2)$.

ε_{ij}独立同分布,服从 $N(0, \sigma^2)$.

α_{ij}^{IO} 同分布,服从 $N(0, \sigma_{IO}{}^2)$ 满足约束条件 :$\sum_{j=1}^{p} \alpha_{ij}^{IO} = 0$.

写成矩阵形式,$Y = A\mu + \varepsilon$ 即为:

$$\begin{pmatrix} y_{11} & \cdots & y_{1p} \\ & \vdots & \\ y_{n1} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} 1_{n1} & & \\ & & \\ & & 1_{nG} \end{pmatrix} \begin{pmatrix} \mu_1^{(1)} & \cdots & \mu_1^{(G)} \\ & \vdots & \\ \mu_p^{(1)} & \cdots & \mu_p^{(G)} \end{pmatrix} + \varepsilon .$$

将参数 $\mu_j^{(g)}$ 重整:

$$\begin{aligned} \mu_j^{(g)} &= \mu_+^{(+)}/pG + (\mu_j^{(+)}/G - \mu_+^{(+)}/pG) + (\mu_+^{(g)}/p - \mu_+^{(+)}/pG) \\ &\quad + (\mu_j^{(g)} - \mu_j^{(+)}/G - \mu_+^{(g)}/p - \mu_+^{(+)}/pG) \\ &= \theta + \alpha_j + \beta^{(g)} + \gamma_j^{(g)} . \end{aligned}$$

约束条件:$\sum_{j=1}^{p} \alpha_j = \sum_{g=1}^{G} \beta^{(g)} = \sum_{j=1}^{p} \gamma_j^{(g)} = \sum_{g=1}^{G} \gamma_j^{(g)} = 0$.

记 $a = (\alpha_1, \cdots, \alpha_j)^T$,$b = (\beta^{(1)}, \cdots, \beta^{(G)})^T$,$\gamma = (\gamma_j^{(g)})$,

则 $\mu = \theta 1_G 1_p^T + 1_G a^T + b 1_p^T + \gamma$,

即 $Y = A(\theta 1_p 1_G^T + a 1_G^T + 1_p b^T + \gamma) + \varepsilon$,$\varepsilon \sim N(0, I_n \otimes \sum)$,

$\sum = [\sigma^2 + (1-\rho)\sigma_{IO}{}^2] I_p + (\sigma_I{}^2 + \rho\sigma_{IO}{}^2) J_p \triangleq \sigma_1 I_p + \sigma_2 J_p$.

2　检验问题

2.1　检验组间有无差异

(a) 此问题可表达为 :H_0 :$(\mu_1^{(1)} \cdots \mu_p^{(1)}) = \cdots = (\mu_1^{(G)} \cdots \mu_p^{(G)})$,

H_1 :至少有两组均值向量不等.

对于该假设,可根据多元分析中有关检验多组均值的理论,构造 Λ 统计量:

$$\Lambda = \frac{\sum_{g=1}^{G} Y^{(g)} (I_p - \frac{1}{n_g} J_p) Y^{(g)}}{\sum_{g=1}^{G} \sum_{i\in(g)} (Y_i - \bar{Y})(Y_i^{(g)} - \bar{Y})^T} \sim \Lambda(p, n-G, G-1).$$

但此法要求有较大的 n,需 $n_i > p$ 方可行. 而实验次数的增多,要花费更多的人力物力财力. 注意到该模型的特殊结构,可考虑如下检验:

(b) H_0 :$\mu_1^{(1)} + \cdots + \mu_p^{(1)} = \cdots = \mu_1^{(G)} + \cdots + \mu_p^{(G)} \Leftrightarrow \beta^{(1)} = \cdots = \beta^{(G)} = 0$,

H_1 :至少有一个 $\beta^{(g)} \neq 0$,

令 $Z_i = y_{i1} + \cdots + y_{ip}$,

则 $z_i \sim N[(\mu_1^{(g)} + \cdots + \mu_p^{(g)}), \rho^2(\sigma_I{}^2 + \sigma^2) I_p \quad i \in (g)]$.

即 Z_i 成为独立的变量,此时模型成为 G-M 模型. 根据线性模型的一般理论可得检验统计量:

$$F=\frac{\left[\sum_{g=1}^{G}\left(\sum_{i\in(g)}Z_i\right)^2/n_g-\frac{Z_+^2}{n}\right]/(G-1)}{\left[\sum_{i=1}^{n}Z_i^2-\sum_{g=1}^{G}\left(\sum_{i\in(g)}Z_i\right)^2/n_g\right]/(n-G)}$$

$$=\frac{\left[\sum_{g=1}^{G}\left(\sum_{i\in(g)}y_{i+}\right)^2/n_gp-\frac{Y_{++}^2}{np}\right]/(G-1)}{\left[\sum_{i=1}^{n}y_{i+}^2/p-\sum_{g=1}^{G}\left(\sum_{i\in(g)}y_{i+}\right)^2/n_gp\right]/(n-G)}.$$

2.2 检验测量间有无差异

(a) 问题表达为

$$H_0:\mu_1^{(1)}+\cdots+\mu_1^{(G)}=\cdots=\mu_p^{(g)}+\cdots+\mu_p^{(G)}\Leftrightarrow\alpha_1=\cdots=\alpha_j=0.$$

令 $\tilde{\Gamma}=\begin{bmatrix}1/\sqrt{p} & \cdots & 1/\sqrt{p}\\ & \Gamma & \end{bmatrix}$,其中 $\Gamma_{(p-1)\times p}$ 为 $p-1$ 个 p 维正交对比构成的矩阵,

$$Y\Gamma^T=\begin{bmatrix}Z_{11} & \cdots & Z_{1,P-1}\\ \cdots & \cdots & \cdots\\ Z_{n1} & \cdots & Z_{n,P-1}\end{bmatrix}=Z^*\sim N[A\mu\Gamma^T,\ln\otimes(\sigma^2+(1-\rho)\sigma_{IO^2}I_{p-1}),$$

即通过正交变换将 y_{ij} 变成行、列都互相独立的 z_{ij},

$\mu^*=\mu\Gamma^T$,问题变为:

$$\begin{cases}Z^*=A\mu^*+\varepsilon^*\\ EZ^*=A\mu^*\\ DZ^*=\ln\otimes(\sigma^2+(1-\rho)\sigma_{IO^2}I_{p-1})\end{cases}$$

$H_0:(1_G\cdots1_G)^T\mu^*=0$,

$H_1\neq H_0$,

将之拉直后即为 G-M 线性模型.故可根据线性模型理论构造检验统计量(过程略)并用正交变换的性质将 Z 变回为 Y,可得检验统计量为:

$$F=\frac{\left[\sum_{g=1}^{G}\left(\frac{1}{n_g}\right)\right]^{-1}\left\{\left[\sum_{j=1}^{p}\sum_{g=1}^{G}\left(\sum_{i\in(g)}y_{i+}/n_g\right)\right]^2-\frac{1}{p}\left[\sum_{g=1}^{G}\sum_{j=1}^{p}\sum_{i\in(g)}\frac{y_{ij}^2}{n_g}\right]\right\}/(p-1)}{\left[\sum_{i=1}^{n}\sum_{j=1}^{p}y_{ij}^2-\sum_{i=1}^{n}\frac{y_{i+}^2}{p}-\sum_{g=1}^{G}\left(\sum_{j=1}^{p}\sum_{i\in(g)}\frac{y_{ij}^2}{n_g}\right)^2+\sum_{g=1}^{G}\left(\sum_{j=1}^{p}\sum_{j\in(g)}y_{ij}\right)^2/n_gp\right]/(n-G)(p-1)}.$$

(b) 问题表述为

$$H_0:(\mu_1^{(1)}\cdots\mu_1^{(G)})=\cdots=(\mu_p^{(g)}+\cdots+\mu_p^{(G)}),$$

$H_1\neq H_0$,

此时

$$\bar{Y}=I_P\otimes A\bar{\mu}+\bar{\varepsilon}.$$

注意到方差为$\sum\otimes I_n$是非对角阵,故一般不考虑此检验,而仅就(a)中的原假设作检验.

2.3 检验组与测量间有无交互效应

$H_0:\gamma=0=(\gamma_j^{(g)})$,

H_1:至少有一个 $\gamma_j^{(g)}\neq0\cdots$,

此时可仿照 2 中(a) 情形,作一个正交变换,使其成为行列都不相关的独立变量,再应用 G-M 线性模型理论可得 F 检验统计量

$$F=\frac{\left[\sum_{g=1}^{G}\sum_{j=1}^{p}\frac{1}{n_g}\left(\sum_{i\in(g)}y_{ij}\right)^2-\frac{1}{p}\sum_{j=1}^{p}\sum_{g=1}^{G}\sum_{i\in(g)}\frac{y_{ij}^2}{n_g}-\frac{1}{n}\sum_{j=1}^{p}y_{+j}^2+\frac{1}{np}y_{++}^{\;2}\right]/(p-1)(G-1)}{\left[\sum_{i=1}^{n}\sum_{j=1}^{p}y_{ij}^2-\sum_{i=1}^{n}\frac{y_{i+}^2}{p}-\sum_{g=1}^{G}\left(\sum_{j=1}^{p}\sum_{i\in(g)}\frac{y_{ij}^2}{n_g}\right)^2+\sum_{g=1}^{G}\left(\sum_{j=1}^{p}\sum_{j\in(g)}y_{ij}\right)^2/n_gp\right]/(n-G)(p-1)}.$$

3　几点说明

1. 本文所述的方法要求方差有特殊的结构,故检验均值以前应先检验方差是否具有该特殊结构,为此可用 Mauchly 提出的 $w=(\det R)/q^{-1}\mathrm{tr}(R)$,其中 $q=p-1$,$R=CSC^T$,C 为 $q\times p$ 正交阵.

2. 本文所述的方法可用 SPSS、SAS 等统计软件包实现.

致谢:感谢王静龙导师的启发和指导.

[参考文献]

[1]　CROWDER M J,HAND D J. Analysis of Repeated Measures[M]. Chapman and hall,1991.

[2]　陈尧庭,方开泰. 多元统计分析引论[M]. 北京:科学出版社,1990.

[3]　陈希孺,王松桂. 近代回归分析——原理方法及应用[M]. 合肥:安徽教育出版社,1987.

[4]　茆师松,王静龙. 数理统计[M]. 上海:华东师范大学出版社,1998.

The Analysis of Repeated Measurements

Yan Qi

(Department of Statistics,Hua Dong Normal University,Shanghai,200062,PRC)

Abstract:This paper presents a model of repeated measurements whose correlation ratios are equal,with the assumption of the special situation we can transfer the multiple linear model to G-M model by orthogonalization. Then F-tests can be still feasible and optimal in testing the contrast of means,and the efficiency is increased by economizing on freedom degree.

Key words:repeated measurements;orthogonalization;G-M model

[责任编辑:陆炳新]