论著

文章编号:1000-5404(2006)07-0732-02

单样本重复测量资料的混合线性模型及其应用

许汝福[1],沈渝菊[2],肖瑞卿[3] (第三军医大学:[1]军事预防医学院军队流行病学教研室,[2]西南医院呼吸内科,[3]西南医院输血科,重庆 400038)

提 要:目的 探索单样本重复测量资料的统计分析方法。**方法** 给出了混合线性模型拟合方法和实例分析。**结果** 通过混合线性模型拟合,使单样本重复测量资料得以合理分析。**结论** 混合线性模型可以有效地、全面地分析单样本重复测量资料。

关键词: 单样本;重复测量;混合线性模型

中图法分类号: R181.2; R311 **文献标识码:** A

Mixed linear model and its application to repeated measurements in single-sample

XU Ru-fu[1], SHEN Yu-ju[2], XIAO Rui-qing[3] ([1]Department of Military Epidemiology, College of Preventive Medicine, [2]Department of Respiratory Diseases, [3]Department of Blood Transfusion, Southwest Hospital, Third Military Medical University, Chongqing 400038, China)

Abstract: Objective To explore analysis method of repeated measurements in single-sample. Methods Mixed linear model was presented and an example of repeated measurements in single-sample was analyzed. Results The reasonable results were obtained for repeated measurements in single-sample by the methods of mixed linear model. Conclusion Mixed linear model can be used to analyze repeated measurement data in single-sample.

Key words: single-sample; repeated measurements; mixed linear model

在医学研究中常常遇到同一受试对象或样品接受多种处理因素后,在不同时间上作多次重复观测,一般称为单样本重复测量资料。统计分析一般采用单变量统计分析方法,常常损失资料所蕴含的信息量,有时可能得出不切实际的解释[1,2]。本研究应用析因设计方差分析思想,拟合混合线性模型进行分析,可同时分析各处理因素和时间因素的作用,并给出应用实例分析。

1 混合线性模型拟合方法

析因设计是多因素设计,其处理组数是各因素水平数的乘积,要求每个处理重复r次。析因设计方差分析可分析各因素的主效应及各因素间的交互作用。如两因素析因设计,假定A因素有i个水平,B因素有j个水平,共有$i\times j$种处理,每个处理重复r次,析因分析模型为:

$$y_{ijk}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+e_{ijk} \quad (1)$$

其中y_{ijk}为实验观测结果,μ为总均数,α_i为A因素第i个水平的效应,β_j为B因素第j个水平的效应,$(\alpha\beta)_{ij}$为A因素第i个水平与B因素第j个水平的交互效应(交互作用),e_{ijk}为随机误差项。

对于三因素析因分析,假定A因素有i个水平,B因素有j个水平,C因素有m个水平,共有$i\times j\times m$种处理,每个处理重复r次,析因分析模型为:

$$y_{ijmk}=\mu+\alpha_i+\beta_j++\gamma_m+(\alpha\beta)_{ij}+(\alpha\gamma)_{im}+(\beta\gamma)_{jm}+e_{ijmk} \quad (2)$$

其中y_{ijmk}为实验观测结果α_i,γ_m的意义同α_i,$(\alpha\gamma)_{im}$和$(\beta\gamma)_{jm}$的意义同$(\alpha\beta)_{ij}$,$(\alpha\beta\gamma)_{ijm}$为A因素第$i$个水平、B因素第$j$个水平与C因素第$m$个水平的二阶交互效应,$e_{ijmk}$为随机误差项。

由于单样本进行多种处理、分别在不同时间上作多次重复观测,各种处理的例数为1,拟合模型时不考虑最高阶交互作用。对于单一处理因素、分别在不同时间上作重复观测,只考虑处理因素和时间因素的主效应,不考虑处理因素和时间因素的交互作用,拟合下列混合线性模型:

$$y_{ijk}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+e_{ijk} \quad (3)$$

其中y_{ijk}为实验观测结果,μ为总均数,α_i为处理因素第i个水平的效应,β_j为时间因素第j个水平的效应,e_{ijk}为随机误差项。

对于两种处理因素、分别在不同时间上作重复观测,不考虑处理因素和时间因素的二阶交互作用,拟合下列混合线性模型:

$$y_{ijmk}=\mu+\alpha_i+\beta_j+\gamma_m+(\alpha\beta)_{ij}+(\alpha\gamma)_{im}+(\beta\gamma)_{jm}+e_{ijmk} \quad (4)$$

其中y_{ijmk}为实验观测结果,α_i为处理因素A第i个水平的效应,β_j为处理因素B第j个水平的效应,γ_m为时间因素第m个水平

作者简介: 许汝福(1963-),男,四川省成都市人,副教授,主要从事卫生统计学方法与应用、流行病学调查等方面的研究,发表论文20余篇。电话:(023)68752276

收稿日期: 2005-06-16;**修回日期:** 2005-12-07

的效应，$(\alpha\beta)_{ij}$、$(\alpha\gamma)_{im}$、$(\beta\gamma)_{jm}$为一阶交互效应，e_{ijmk}为随机误差项。

重复测量资料的分析与一般混合线性模型一致，可利用 SPSS 11.0 的“*General Linear Model*（一般线性模型）”—“*Univariate Analysis of Variance*（单因素方差分析）”进行拟合。

2　实例应用

表 1 是 1 例 B 型血全血样品分成 12 份，每份 60 ml，分别在不同温度（20、37 ℃）、加入不同剂量 O 型全血或 O 型红细胞（不加、加入 9 ml O 型全血、加入 12 ml O 型全血、加入 15 ml O 型全血、加入 18 ml O 型全血、加入 18 ml O 型红细胞）后在不同时间（1、2、4、8、12 h）检测的血浆游离血红蛋白（FHb，mg/L）资料。

表 1　B 型全血在不同温度、加入不同剂量 O 型全血或红细胞后在不同时间的血浆游离血红蛋白

处理编号	温度（℃）	剂量（ml）	检测时间（h）				
			1	2	4	8	12
1	20	60 ml B 型全血	56.28	67.46	86.44	148.72	165.01
2	20	60 ml B 型全血 +9 ml O 型全血	66.44	59.64	84.99	158.59	183.28
3	20	60 ml B 型全血 +12 ml O 型全血	58.80	69.72	77.03	119.61	213.42
4	20	60 ml B 型全血 +15 ml O 型全血	50.56	85.26	72.91	166.98	193.54
5	20	60 ml B 型全血 +18 ml O 型全血	65.94	76.03	90.93	171.78	222.60
6	20	60 ml B 型全血 +18 ml O 型红细胞	67.70	83.51	97.56	146.49	172.20
7	37	60 ml B 型全血	67.63	67.46	78.55	145.86	239.90
8	37	60 ml B 型全血 +9 ml O 型全血	64.52	71.24	73.42	183.32	261.32
9	37	60 ml B 型全血 +12 ml O 型全血	68.55	66.46	71.91	153.80	261.95
10	37	60 ml B 型全血 +15 ml O 型全血	64.44	94.17	107.61	134.15	237.56
11	37	60 ml B 型全血 +18 ml O 型全血	71.66	87.96	107.36	181.55	251.55
12	37	60 ml B 型全血 +18 ml O 型红细胞	71.32	76.78	86.95	141.92	225.98

本资料是单样本（1 名 B 型血全血样品）分别在不同温度（*temp*）、加入不同剂量 O 型全血或 O 型红细胞（*dose*）12 种处理后在不同时间（*time*）重复观测血浆游离血红蛋白（FHb），属于单样本重复测量资料。经直线相关与回归分析，12 种处理的游离血红蛋白量与检测时间的相关性见表 2，最小的相关系数 r 为 0.961，最大的为 0.997，均具有显著的正相关关系，随着检测时间的延长，游离血红蛋白量均明显增高；但回归系数 b 最小的为 9.603，最大的为 18.900，相差近 1 倍，说明变化趋势不完全相同；球对称条件检验，近似 $\chi^2 = 23.260$，$\nu = 9$，$P = 0.009$，数据不满足球对称条件；经方差齐性检验，$F = 9.679$，$P < 0.001$，方差不齐。因此，该资料不满足方差齐性条件和球对称条件，不能按常规作单变量方差分析。

表 2　12 种处理的游离血红蛋白量与检测时间的相关系数与回归关系

处理编号	1	2	3	4	5	6	7	8	9	10	11	12
r	0.959	0.977	0.979	0.964	0.966	0.961	0.974	0.992	0.997	0.993	0.997	0.980
b	13.970	12.114	18.900	13.413	18.035	13.186	15.934	14.902	16.437	9.603	13.989	10.550
P	0.001	0.004	0.004	0.008	0.009	0.010	0.005	0.001	0.001	0.001	0.001	0.003

应用 SPSS 11.0 软件，按公式（4）拟合混合线性模型。将数据整理成“*FHb*、*dose*、*temp*、*time*”4 列，FHb 为应变量，*temp*、*dose*、*time* 为确定的因素变量，考虑温度、剂量和时间的主效应以及一阶交互作用，拟合 FHb 关于“*dose + temp + time + dose * temp + dose * time + temp * time*”的混合线性模型（方差分析模型），剂量间比较有显著意义（$F = 3.066$，$P = 0.033$）；温度间比较有显著意义（$F = 19.896$，$P < 0.001$）；时间之间比较有显著意义（$F = 321.721$，$P < 0.001$）。温度与时间之间存在交互作用（$F = 9.465$，$P < 0.001$），剂量与时间之间交互作用不显著（$F = 1.498$，$P = 0.187$），温度与剂量之间交互作用不显著（$F = 0.291$，$P = 0.912$）。

如果表 1 资料按单因素方差分析方法计算，分析一个因素时不考虑其他两个因素，则时间之间比较差异显著（$F = 142.134$，$P < 0.001$），剂量间比较差异不显著（$F = 0.119$，$P = 0.988$），温度间比较差异不显著（$F = 0.829$，$P = 0.366$）。

3　讨论

对同一受试对象或样品（单样本）接受多种处理因素后、在不同时间上作多次重复观测的资料，进行单变量方差分析时，要求资料必须满足各组的对比协方差阵相等的齐性条件和球对称条件[1]。由于重复测量数据误差的多层次性和数据间存在的自相关性等原因，协方差阵常常不满足单变量方差分析的球对称条件[1,2]，需要用球对称系数对单变量方差分析的统计量进行校正。各组的对比协方差阵相等的齐性检验和球对称系数的计算较繁琐，使单变量方差分析在重复测量数据的分析中受到限制。混合线性模型考虑了数据的非独立性和分布于不同水平上的误差，可以同时考虑各处理因素和时间因素的作用及其交互作用（不考虑最高级交互作用），且计算简便。

实例资料不满足各组的对比协方差阵相等的齐性条件和球对称条件，单变量方差分析受到时间因素的影响，结果忽略了剂量和温度的作用。而拟合混合线性模型进行分析，剂量、温度和时间 3 个主效应均有显著意义，游离血红蛋白与血液保存温度的变化、保存时间的长短以及加入 O 型全血或 O 型红细胞的剂量有关，与文献[3,4]的结论一致。通过实例分析，说明混合线性模型是一种较好的单样本重复测量数据分析方法，可以有效地、全面地分析单样本重复测量资料。

参考文献：

[1] 陈长生，徐勇勇. 重复观测数据单变量方差分析的前提条件的检验[J]. 中国卫生统计，2000，17(2)：74－76.

[2] 陈长生，徐勇勇，吴　冰. 医学重复观测资料序列相关假设检验[J]. 第四军医大学学报，1998，19(5)：581－582.

[3] 杜静平，邢天荣，马奎力，等. 不同温度及时间对保存血液有效期和质量的影响[J]. 中国应用生理学杂志，2002，18(4)：362－364.

[4] 田兆嵩. 临床输血学[M]. 第 2 版. 北京：人民卫生出版社，2002：124.

（编辑　龙　亮）