

用 SPSS11.0 实现对重复测量资料的方差分析

邱 宏 金如锋 赵 玲* 魏建子* 沈学勇*

(上海中医药大学预防医学与卫生统计教研室 上海201203)

摘 要: 应用 SPSS11.0 中 GLM(一般线性模型)的 Repeated Measures 和 Multivariate 过程对重复测量设计资料进行重复测量方差分析和多元方差分析,并进行不同时间点和不同组间的两两比较。

关键词: 重复测量设计; 一般线性模型; 重复测量方差分析; 多元方差分析

近年来,临床和基础研究中采用重复测量设计方案的研究日益增多,常见如治疗前和治疗后不同时间测量的某指标的数值资料,不少已发的文章对此类资料的分析都采用单因素方差分析甚至 t 检验等不妥当的方法,而没有考虑到某指标在不同时间点上的关联性或该指标随时间变化的趋势。许多统计学教科书对重复测量资料的方差分析方法有论述但不全面,特别是方差分析有统计学意义时对其后的进一步在不同时点或不同组间两两比较的方法较少介绍。考虑到 SPSS 统计软件的普及实用性,在此以一实例介绍用 SPSS11.0 实现对重复测量资料的方差分析,以期对采用重复测量设计方案的临床和基础科研工作者提供可借鉴的方法。

1 资料和方法

1.1 资料来源

本研究资料来源于国家中医药管理局和上海市重点学科建设项目、上海市科技发展基金科技攻关项目(03DZ19554-2)。雄性家兔48只,体重1.8kg~2.8kg,随机分为4组:A 空白对照组、B 模型对照组、C 激光内关组(即左侧内关穴激光照射复合针灸治疗组)和 D 激光对照组(即内关穴旁开对照点激光照射复合针灸治疗组)。B、C、D 三组注射垂体后叶素制作家兔心肌缺血和心动过缓模型,C、D 组造模后即开启复合激光针灸治疗仪,均照射20min。各组均记录40分钟心电图变化。效应指标为心率,以心电图直接记录。

1.2 统计方法

用一般线性模型(General Linear Model)的 Repeated Measures 过程实现重复测量资料的方差分析,并用 Multivariate 过程实现组间的两两比较。

数据文件格式为模型数据格式。以 group 变量表示资料

的分组(1~4分别代表模型对照、激光对照、激光内关和空白对照4组),以 t_0 - t_7 变量分别代表造模前和造模后5min、10min、15min、20min、25min、30min、40min 等不同时间点的心率。

2 分析步骤及结果

2.1 通过球形检验(Mauchly's Test of Sphericity)的结果判断重复测量数据之间是否存在相关性

如果该检验 $P > 0.05$,说明重复测量数据之间实际上不存在相关性,数据符合 Huynh-Feldt 条件,可按单因素方差分析方法来处理;如果 $P < 0.05$,说明重复测量数据之间存在相关性,不可按单因素方差分析方法处理。实际应用中的重复测量设计资料以后者多见,应使用重复测量的方差分析模型。

Analyze→General Linear Model→Repeated Measures

Within-subject factor name 框:改为 t “定义重复测量的变量名为 t ”

Number of levels 框:键入8;add “重复测量的次数为8次”

Define

Within-subject variables 框: t_0 - t_7 “ t_0 - t_7 代表8次测量结果”

Between subject factor 框:group

Model:

选中 Custom “自定义模型”

Within-subject Model 框: t “分析8次重复测量间有无趋势”

Between subject Model 框:group “只分析主效应”

Continue

OK

主要结果如下:

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig	Epsilon		
					Greenhaus e-Geisser	Huynh-Feldt	Lower-bound
T	0.002	251.620	27	0.000	0.412	0.474	0.143

收稿日期:2005-07-13

* 上海中医药大学针灸推拿学院

球形检验的结果 $P < 0.01$, 说明8次重复测量的数据间存在高度的相关性, 宜用多元方差分析进行检验或按表右侧给出的三种校正方法对一元方差分析进行校正(校正系数为

Epsilon)。一般推荐使用 Greenhouse-Geisser 的校正结果。

2.2 分析时间、分组因素的作用以及时间和分组之间的交互作用。

Teste of Within-Subjects Effects

Measure: MEASURE.1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
T	Sphericity Assumed	257833.602	7	36833.372	162.240	0.000
	Greenhouse-Geisser	257833.602	2.882	89460.488	162.240	0.000
	Huynh-Feldt	257833.602	3.316	77758.069	162.240	0.000
	Lower-bound	257833.602	1.000	257833.602	162.240	0.000
T * GROUP	Sphericity Assumed	90098.159	21	4290.389	18.898	0.000
	Greenhouse-Geisser	90098.159	8.646	10420.448	18.898	0.000
	Huynh-Feldt	90098.159	9.948	9057.339	18.898	0.000
	Lower-bound	90098.159	3.000	30032.720	18.898	0.000
Error(T)	Sphericity Assumed	69925.365	308	227.030		
	Greenhouse-Geisser	69925.365	126.812	551.409		
	Huynh-Feldt	69925.365	145.897	479.279		
	Lower-bound	69925.365	44.000	1589.213		

此处 t 和 $t * \text{group}$ 的 P 值均 < 0.05 , 时间因素以及时间因素和分组的交互作用有统计学意义, 说明测量指标有随时间变化的趋势并且时间因素的作用随着分组的不同而不同。

Teste of Between-Subjects Effects

Measure: MEASURE.1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	17166610.628	1	17166610.63	2843.363	0.000
GROUP	164584.195	3	54861.398	9.087	0.000
Error	265647.052	44	6037.433		

group 的 P 值 < 0.05 , 说明分组因素起作用。

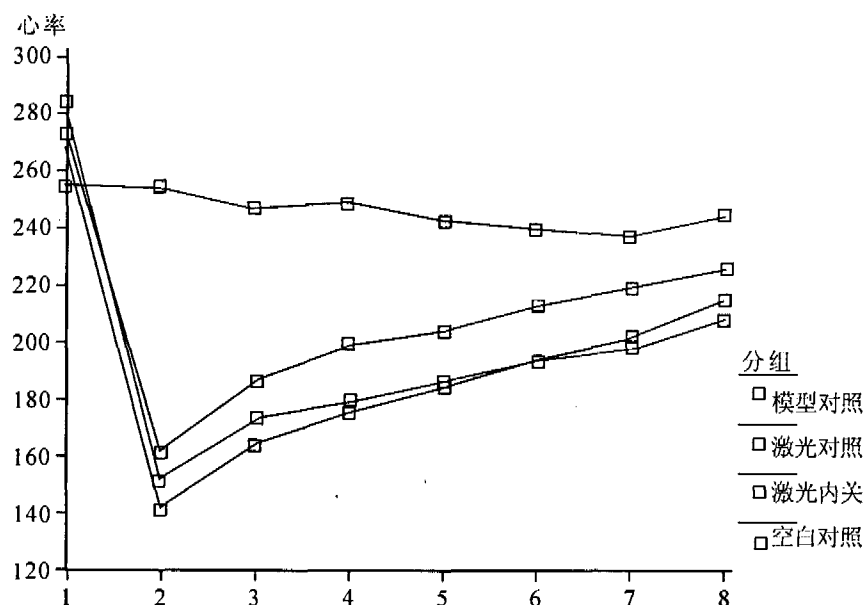
这里可以借助于 Plots 子菜单作出8次重复测量资料的均数变化趋势图。

Horizontal Axis: t “横轴表示时间”

Separate Lines: group “用不同的折线表示分组”

Add

结果如下:



由上图可以直观地看出测量指标随时间的变化趋势以及4组资料随时间变化趋势的不同。

我们发现绝大多数研究者都希望继续分析每组资料在不同时间点上的差别和每一个时间点上4组均数的差别, 这就需

要进一步做不同时间点或不同组间的两两比较。

2.3 每个分组在8个时间点上的两两比较

Data→Split Files

选中 Organize Output by Groups: group “按分组对数据文件进行拆分”

Analyze→General Linear Model→Repeated Measures

Within-subject factor name 框: 改为 t “定义重复测量的变量名为 t”

Number of levels 框: 键入 8: “重复测量的次数为 8 次”

Within-subject variables 框: t0-t7 “t0-t7 代表 8 次测量结果”

选中 Custom “自定义模型”

Within-subject Model 框: t “分析 8 次重复测量间的趋势”

Display Means for: t

选中 Compare main effects

Confidence interval adjustment: Bonferroni

“当校正系数 Epsilon < 0.7 时推荐选择 Bonferroni 法进行两两比较”

2.4 每个时间点上 4 个分组之间的两两比较

Analyze→General Linear Model→Multivariate

“调用多元方差分析过程”

Dependent Variables: 选入 t0-t7

Fixed Factors: group

选中 Custom

Model 框: group

Post Hoc Tests for: group

选中 LSD “选择 LSD 为两两比较的方法，

两两比较方法的选择原则同单因素方差分析”

Display Means for: group

Display: Descriptive statistics “输出对资料的描述”

2.5 两两比较结果

2.5.1 每个时间点上 4 个分组之间的两两比较

表1 4组家兔在不同时间点上的心率比较($\bar{x} \pm s$)

时间	n	模型对照	激光对照	激光内关	空白对照
造模前	12	284.08±45.498	272.08±42.694	273.42±42.786	254.50±29.312
造模后 5 min	12	151.92±26.548*	141.92±24.608*	162.08±25.759*	254.42±31.598
造模后 10 min	12	173.17±18.265*	164.58±19.341*	186.17±27.587*	246.08±36.56
造模后 15 min	12	179.50±23.020*	175.25±17.848*	198.67±33.792*#	248.33±33.027
造模后 20 min	12	186.25±24.595*	183.92±19.256*	203.67±34.634*	241.83±36.336
造模后 25 min	12	193.50±26.756*	193.33±17.981*	212.25±37.979*	238.92±35.184
造模后 30 min	12	198.17±27.725*	200.92±18.118*	218.67±40.529	236.92±33.209
造模后 40 min	12	207.67±31.477*	214.25±18.361*	225.58±39.160	243.92±33.854

注: * 与同一时间点上的空白对照组比较, $P < 0.05$; # 与同一时间点上的激光对照组比较, $P < 0.05$ 。

2.5.2 每个分组在 8 个时间点上的两两比较 因每个分组在 8 个时间点上的两两比较结果输出内容较多, 本文省略, 研究者可根据自己的分析目的和需要选用。本例资料结合均数图即可直观看出各组心率随时间的变化趋势, 可不必做 8 个时间点上的两两比较。

3 讨论

重复测量是指对同一观察对象的同一观察指标在不同时间或环境下进行的多次测量, 用于分析观察指标的变化趋势及有关的影响因素。临床和基础研究中常见这样的设计: 中西医结合治疗哮喘, 为了动态观察治疗方案对病人生命质量的改善情况, 测量了病人治疗前后不同时间(治疗前、治疗后 1 个

月、3 个月、半年、1 年、2 年) 的生命质量量表的评分; 中医药防治老年痴呆的作用机理研究中, 为了动态观察药物对痴呆大鼠空间记忆能力的改善情况, 测量了大鼠在治疗前后不同时间(治疗前、治疗后 1、2、3、4、5、6、7 天) 进行水迷宫实验的潜伏期资料。此类设计均属重复测量设计。

含有多水平的重复测量因子的方差分析中, 有特别的假设: 复合对称假设与球形假设。由于这两个假设在方差分析中很少能够满足, 分析重复测量设计资料的多元方差分析方法近年来越来越普及。复合对称的假设要求重复测量结果的方差与协方差是一致的, 这是重复测量单变量方差分析有效的充分条件(即得到的 F 值服从 F 分布), 但不是必要条件。球形假设是单变量方差分析有效的充分必要条件, 说明每一个对

正交投影鉴别决明子色谱指纹图谱峰纯度

方洪壮 沈德凤 任淑清
(佳木斯大学药学院 佳木斯154007)

摘要: 应用正交投影法对色谱指纹图谱的峰进行鉴别。用对照组分光谱与目标色谱峰前后的零浓度区光谱构建正交投影阵,对目标色谱峰的光谱进行投影,以投影前后光谱的夹角余弦作为判据来鉴别目标色谱峰的峰纯度。对决明子药材色谱指纹图谱的峰纯度鉴别,识别出大黄素、大黄素甲醚及大黄酚三个纯色峰。正交投影鉴别色谱指纹图谱的峰纯度方法结果可靠,背景的干扰可消除。

关键词: 正交投影; 色谱指纹图谱; 色谱峰纯度; 决明子

准确识别色谱峰的纯度是色谱的定性和定量分析的前提。由于中药的化学成分多为次生代谢产物,彼此间的化学结构差别很小、理化性质近似,其色谱峰在色谱指纹图谱上不可避免有重叠,准确判别色谱峰纯度更重要。鉴别色谱指纹图谱色谱峰的纯度、确认色谱峰不仅增加指纹图谱的信息量,而且有助于对中药的深入研究。色谱峰纯度的定性可归结为仪器方法和化学计量学方法^[1],其中后一方法又是前一方法的基础。本研究根据正交投影的原理和多组分色谱相关光谱的方法^[2,3],应用正交投影法对中药决明子药材的色谱指纹图谱中的色谱峰进行了纯度鉴别,识别出三种蒽醌类成分的纯色峰。实验结果显示,所提出的方法结果可靠,有一定的实用价值。

1 方法原理

设 X 为采用联用色谱测得的中药色谱指纹图谱的 $N \times M$ 阶二维数据矩阵,它们的行方向表征了色谱的流出信息(M 个流出时间点),列方向表征了光谱信息(N 个吸收波长);用 X_a 表示 X 中的待鉴别的目标组分色谱峰的 $N \times m$ 阶二维数据

矩阵,考虑到仪器的背景,可把矩阵 X_a 分成两个部分,即:

$$X_a = X_b + X_c \quad (1)$$

X_c 表示来自目标组分的主要信息,而 X_b 则表示来自仪器的背景。在目标色谱峰前后的零浓度区内分别各选取一点 s (目标色谱峰前)与 e (目标色谱峰后),以此两点的光谱矢量 B_s 和 B_e 及对照组分的光谱矢量 B_c ,组成光谱矩阵 A :

$$A = [B_s, B_e, B_c] \quad (2)$$

并构造正交投影阵:

$$M = I - AA^+ \quad (3)$$

其中, I 为单位矩阵, A^+ 为 A 的广义逆阵。用矩阵 M 对 X_c 沿光谱方向的每一光谱矢量 x_{cj} 进行投影运算:

$$x_{cj}^* = M x_{cj} \quad (4)$$

x_{cj} 表示 X_c 的第 j 条光谱, x_{cj}^* 表示 x_{cj} 投影后的残余光谱。为消除不等性噪声^[4]的影响,令投影后所得的残余光谱 x_{cj}^* 与投影前的原始光谱 x_{cj} 的夹角余弦为 r_j ,即:

$$r_j = x_{cj}^* x_{cj}^T / (\|x_{cj}^*\| \|x_{cj}\|) \quad (5)$$

并以 r_j 作为鉴别色谱峰纯度的判据^[3]。

由于 M 扣除了 x_{cj} 与 A 相关的光谱信息, x_{cj}^* 保留了 x_{cj}

象内重复测量数据间相互独立。这个假设的本质以及是否违反该假设在方差分析教科书中没有详细描述。当违反了复合对称假设或球形假设,单变量方差分析得到的结果就是错误的。但如果我们使用多变量准则进行两组或多组重复测量资料比较的检验,他们之间就不需要相互独立。这个优点也是为什么多元方差分析方法越来越广泛应用于含有多个水平的重复测量因素假设检验的原因。采用重复测量设计的研究常可见到这种情形:单变量方差分析与多元方差分析的结果存在明显差异,这意味着重复测量因素各水平间的差异与对象中的差异相关,在球形假设不满足的前提下应以多元方差分析的结果为准。

本研究介绍的分析重复测量资料的方法属于一般线性模型(General Linear Model),国内有不少学者(如赵耐青教授)

提出混合效应模型(Mixed Model)的分析方法,由于篇幅所限和笔者水平的限制,在此不进行进一步的介绍和两模型的比较。

参 考 文 献

- 1 张文彤,主编. SPSS11.0统计分析教程(高级篇). 第1版. 北京:希望电子出版社,2002,6:32~63.
- 2 余松林,向惠云,编著. 重复测量资料分析与 SAS 程序. 第1版. 科学出版社,2004,3:1~48.
- 3 赵耐青. 重复测量资料的统计分析方法. 复旦大学公共卫生学院卫生统计和社会医学教研室内部资料.