152-155号

维普资讯 http://www.cqvip.com

第22卷 第2期 重庆医科大学学报 Vol.22 No.2
1997 ACTA UNIVERSITATIS SCIENTIAE MEDICINAE CHONGQING 1997

第一主成份划值区间分类方法的探讨

卫生统计学教研室 王润华* 曾 军

提 要 本文对第一主成份分类方法及适用条件作了清晰的阐述。将此方法用于小学生情感状态资料，明确显示获得了好的分类效果，实际符合率达到了93%。R181.2

关键词 主成份分析;分类方法

中图分类号 R181.2 O212.2 R195.1

生物医学研究经常遇到依据研究对象的多指标数据将观察对象分类及分型问题。一般采用专业经验法，即从经验出发，制定多指标分类标准后，再依据观察对象的各指标数据加以分类。但面对有较多新指标的情况，经验不足便成了突出的问题。如何正确地制定新指标分类标准，特别是当新指标较多时，无疑难度较大。怎样对研究对象分类的正确性进行评价，也会十分棘手。鉴于此，医学多元统计方法提供了指标或样品聚类分析法[1]，在相当程度和范围内促进了分类学研究的发展。对众多指标或样品进行聚类分析的前提条件是:指标间具有较好的相关性，或样品间距离的相近性。否则必然导致分类的类别过多而与专业不符，同时也不易获得较好的分类效果。医学研究实践中时而出现的多指标数据变差较大，相关性较弱的情况，促使我们进一步探索更合理、更有效的分类学方法。主成份分析是一种按照多种指标数据提供的多种相关信息，经过一定形式的线性组合，成为含有某种专业意义的综合指标的方法。不同形式的线性组合方式即构成了不同的综合变量，表达不同的生物医学意义。因此常用来指示相应的生理或病理类型。作为分类学方法，它亦在分类学研究中占有一定的重要地位。但由于人们对其在分类学研究中的意义认识不足，加之实践甚少，故至今尚未能得到推广应用。在主成份分析中，第一主成份即第一综合指标所概括的原始各指标的信息量最大，故亦是最重要的主成份。根据划定的第一主成份取值区间对观察对象进行分类，尤其适用于区分为两种类型，或伴有中间过渡类型的三种类别的医学研究资料。本文作者正是在这一方法的应用研究方面作了有益的探讨。

1 资料来源与方法

资料来源于1996年3月对重庆市4所公立小学四年级学生的情感状态调查。问卷采用改进的成人NCTB(神经行为功能核心测验组合)中的测验－1[2]。共设置50个情感状态变量，每一情感状态均按5级记分。各分值含意为:0(一点也无)，1(略有一点)，2(有一些)，3(相当多)，4(非常多)。填写方法是先由调查者当面解释词意，待学生理解后即刻由学生自行独立选择分值，互不干扰。50个情感状态变量表达如下:

1 对人友好	18 心神不定	35 脾气倔强
2 紧张感	19 不专心	36 孤立无援
3 乏力感	20 疲劳感	37 萎靡不振
4 头脑清醒	21 助人为乐	38 困惑感
5 轻快感	22 生气	39 警觉性很高
6 为某事而难过	23 气馁	40 受骗感
7 无精打彩	24 忿恨不满	41 办事有效率
8 有点气恼	25 寂寞感	42 信任他人
9 悲感	26 伤心	43 劲头十足
10 活跃	27 糊涂	44 卑微感
11 易发怒	28 心情愉快	45 健忘
12 忧郁感	29 痛苦感	46 无忧无虑
13 精神饱满	30 焦虑	47 惊恐感
14 恐慌感	31 好斗	48 内疚感
15 安逸感	32 脾气好	49 精力充沛
16 自卑感	33 情绪低落	50 对事情缺乏判断力
17 富有同情心	34 懒散	

将收集的445份资料全部输入IBM－486微机，建立数据库，应用SAS[3]统计软件进行如下分析:①全部指标的聚

* 男，1944—03出生，教授

维普资讯 http://www.cqvip.com

类分析，用以确定划类基本依据。在确定划分类别的基础上，制定划类标准和将观察对象先验归类。在观察对象已有先验类别的前提下，应用逐步判别分析，并对判别分类效果进行考查。②全部指标的主成份分析，按标准化第一主成份取值制定划类标准和将观察对象先验归类。在观察对象已有先验类别的前提下，进行逐步判别分析，并对判别分类效果进行考查。逐步判别分析均采用统一程式，根据445例观察对象编号用系统随机抽样法(即等距抽样)。以不能被5整除者共356例作为模型样本，用以建立判别函数式组；以能够被5整除者共89例作为检验样本，用以检验判别效果。两种样本均实施判别归类结果与先验类别比较的符合率分析。

2 分析结果

2.1 采用相似系数系统聚类法对50个指标进行指标系统聚类分析，在相关系数为0.5时开始有指标聚集成类，中间聚类结果专业意义难于解释，当相关系数变动为0.23时，指标聚为正向情感类(恰包括16个正向情感)和负向情感类(恰包括34个负向情感)。这提示小学生情感类型可以在正向和负向两类基础上划类。由于正常小学生正向与负向情感类型均占少数，而波动型居多，其分析一般呈正态分布，故决定将观察对象情感状态分为3种类型：①正向情感型，记T_1；②波动情感型，记T_2；③负向情感型，记T_3。划类标准规定为：①16个积极正向情感总积分＞48和34个消极负向情感总积分＜50者，定为T_1；②16个积极正向情感总积分＜40和34个消极负向情感总积分＞60者定为T_3；③其它定为T_2。445例两种样本观察对象划定的先验类别结果见表1中的(A)。

表1 两种方法对445例小学生两个样本情感状态分型结果的分布

方法	样本	例数	正向型T_1		波动型T_2		负向型T_3	
			频数	%	频数	%	频数	%
经验法(A)	模型样本	356	60	16.9	246	69.1	50	14.0
	检验样本	89	12	13.5	56	62.9	21	23.6
合计		445	72	16.2	302	67.9	71	16.0
第一主成份法(B)	模型样本	356	55	15.5	244	68.5	57	16.0
	检验样本	89	20	22.5	53	59.6	16	18.0
合计		445	75	16.9	297	66.7	73	16.4

根据指标聚类分析结果提示，采用经验标准划类后，用逐步判别分析法对模型样本资料建立判别函数式组($P<0.05$)内含17个情感状态变量X_i ($i=3、4、5、9、15、17、26、30、31、32、34、38、41、42、46、47、49$)。判别函数组略。经模型样本356例内回代，与其先验类别T_i比较，判别符合率为79.21%。检验样本判别符合率仅为73.03%。经反复测试，划类标准无论怎样调整，内含如何变化，后续分析的判别符合率均低于80%。

2.2 情感状态的主成份划值区间分类—逐步判别分析

由于在聚类分析结果提示下采用经验制定划类标准未取得满意效果，作者对资料采用了主成份分析。获得第一主成份表达负向情感，第二主成份表达正向情感，其余主成份内的变量组合含义复杂。故首先考虑以标准化第一主成份和第二主成份联用，按正态分布法制定划类标准，按标准进行先验归类，并执行逐步判别分析($P<0.05$)，获得13个情感变量X_i($i=1、5、10、18、20、25、29、33、37、39、40、41、42$)组成判别函数式组。但回代及检验

表2 第一主成份划值区间分类法模型样本356例小学生情感型别判别式

	正向情感型T_1	波动情感型T_2	负向情感型T_3
截距	−32.9696	−12.5189	−7.5209
X_3	1.8302	1.0050	0.5578
X_5	2.1454	1.8620	2.1499
X_6	0.8126	0.7857	0.2340
X_8	1.8954	0.9506	0.3151
X_9	1.0293	0.4889	−0.0512
X_{11}	0.4617	0.4116	0.2020
X_{19}	0.9014	0.8843	0.4042
X_{25}	1.1202	0.5171	0.1968
X_{29}	1.2582	0.6094	0.1862
X_{37}	0.9709	0.3401	0.0022
X_{39}	1.5181	0.6317	0.0606
X_{40}	2.0571	1.2497	0.5604
X_{41}	1.9047	1.1117	0.3483
X_{42}	1.3595	0.5411	−0.1133
X_{48}	1.6390	1.3472	1.2482
X_{50}	1.4067	0.6177	0.4471
X_{50}	1.3717	0.8784	0.2787

维普资讯 http://www.cqvip.com

样本判别符合率分别为76.69%和65.17%。继之采用标准化第一主成份划值区间法，规定标准化第一主成份取值$Z_1<-1$为正向情感类，记T_1；$-1<Z_1<1$为波动情感类，记T_2；$Z_1>1$为负向情感类，记T_3。3种类型情感观察对象在模型样本及检验样本中的分布见表1中的(B)。同样在划类、归类基础上进行逐步判别分析($P<0.05$)，模型样本356例资料所获得的有判别意义的指标变量X_i和建立的判别式组见表2。回代及检验判别效果见表3。

表3 第一主成份划值区间分类法两种样本的判别效果考查

"先验类别"	模型样本判属类型			检验样本判属类型		
	T_1	T_2	T_3	T_1	T_2	T_3
T_1	52	3	0	20	0	0
T_2	0	230	3	1	49	3
T_3	11	8	49	0	2	14
判别符合率(%)	92.98			93.26		

3 讨论

对多指标系统的研究对象进行分类或分型问题是生物医学常见的研究课题。虽然分类方法较多，但分类效果应以客观评价依据来考查。这些评价依据包括：①划类符合专业知识，②划类应避免采用各单一指标的简单叠加法，最好选用综合指标法，③制定划类标准的指标数据应具有易进行统计的分布，④应尽可能利用多元逐步判别分析方法对分类效果进行考查，以取得模型样本和实际应用中的较高的符合率。纯经验分类法在依据单一指标或较少指标对研究对象进行分类或分型工作中曾有过重要的贡献，在适用条件下也将会继续发挥作用。但随着多指标系统分类学研究课题的涌现，就表现出该方法的能力不足。对全部指标数据的分布缺乏判断力、对全部指标中哪些指标对分类有重要影响难于鉴别、对分类效果缺乏考查依据是其重要缺陷。因此研究者不约而同地把目光投向指标聚类分析方法。聚类分析属多元统计分析，它一般分为指标聚类分析(R型聚类分析和样品聚类分析(Q型聚类分析)。指标聚类分析结果对相同指标系统的分类有总体性指导作用。它要求各指标数据具有较好的相关性，以便在经验干预下获得理想的聚类，并找出每类内的典型代表性指标变量，但具体分类标准尚需人为地根据典型变量分布特征加以确定。样品聚类分析一般只对某样本中的样品分类起划类作用。就指标聚类分析而言，遇有指标间相关性较弱的情况，也难于获得符合专业知识的理想聚类，更难于选择类内典型变量。因而使确定划类标准缺乏依据和概括性，导致不良的分类效果。本文所采用的资料就是这样一种情况，聚类后计算两类内各情感变量的类内平均相关指数$R^2=\Sigma r_i^2/(m-1)<0.14$，无法选择类内典型变量。既使按正态分布人为确定划类标准，最终判别符合率也较低(<80%)。

作者所采用第一标准化主成份划值区间分类方法用于小学生情感指标分类满足了分类效果的客观评价依据。由于是综合指标法，它概括了50个指标的主要信息。由于是标准化主成份，它具有标准正态分布，易于统计且有分布预见性。由于划类符合专业知识，故能获得较高的判别符合率(93%)。根据统计学原理可以推论：指标变量间的相关程度愈高，第一主成份所概括的信息量愈大。因而只要划类标准符合专业知识，获得良好判别符合率的可能性亦愈大。这些均提示：①第一主成份划值区间分类法适用于指标分类，②它更适合于两种类别或兼有中间过渡型的三种类别的分类情况，③它对指标变量数、指标变量分布和指标变量间的相关程度无严格要求。总之，是一种值得进一步探讨和倡导使用的分类学方法。

参考文献

1.黄正南.医用多因素分析，湖南科学技术出版社，1995年第3版 156
2.梁友信.工业卫生与职业病 1987;13(6):331
3.SAS/STAT User's Guide Release 6.03 Edition,SAS Institute Inc,1988

(收稿日期 1996－09－13)

维普资讯 http://www.cqvip.com

Deliberate classified method of first principal component

Wang Runhua, et al
Department of health statistics

This paper gives a clear explanation of classified method and applicable condition of first principal component.Writers make use this method in analysis of emotional data of schoolchildren shows that it acquires a good classificatory result and rate of conforinable to reality attains to 93%.

KEY WORDS Principal component analysis;Classified method

正常大白鼠24小时动态心电图分析(HOLTER)与方法探讨

临床学院内科 覃 数* 陈运贞 王志华 杨自荣

提 要 本文报道了大白鼠24小时动态心电图的分析结果;同时,探讨HOLTER记录仪的改装与电极安放等方法学问题.

中图分类号 R-332;R540.41
关键词 大白鼠; 动态心电图(HOLTER)

以大白鼠为模型研究心律失常,因其价廉易得,故应用日益广泛。心电图是研究心律失常的重要指标。大白鼠的常规心电图记录,因其时间短,动物又处于麻醉状态,导致信息大量丢失,使实验结果复杂化。24小时动态心电图可克服上述缺点,其结果可能更可靠,更有价值。目前尚未见这方面的报道。本文分析40只大白鼠24小时动态心电图,并对其方法学进行探讨。

1 材料与方法

1.1 健康成年Wistar大白鼠40只(本校实验动物中心提供),体重250-300克,雌雄不拘。

1.2 3%水合氯醛,腹腔注射麻醉。胸前区剪毛,常规消毒。于其皮下埋置五根自制电极。上胸部并列按放2根负极,下胸部左右各安放1根正极,无关电极则埋于右侧下胸部,全部电极导线通过皮下隧道,从大白鼠后背穿出,缝合切口,将大白鼠置于铁丝笼内(25×6×7cm),相对限制其旋转。三日后记录心电。

1.3 HOLTER磁带记录仪(Diagnostic Medical Instrument Inc U.S.A).

1.4 回放扫描分析系统(Compaq Prolinea MT4/66 Grand pacific Company HOLTER 3.0 1995).

2 结果与讨论

2.1 大白鼠HOLTER分析结果

2.1.1 心率 最慢心率242±58.57/min,最快心率439±73.43/min,平均心率356±47.21/min。其心率范围141-678/min。

2.1.2 心率的昼夜变化 将24小时分为四个时段,即12-18pm、18-0pm、0-6am、6-12am。比较这四个时段的心率变化,无昼夜波动特征。t检验,无显著差异($P>0.5$)。

2.1.3 心律 40只大白鼠均为窦性心律。共记录到房性早搏总数85次,平均2.125/24h,最高发生率42/24h。室性早搏总数31次,平均为0.775/24h,最高发生率14/24h。未发现房性或室性心动过速。

国内外均有关于大白鼠心电图的报道,其心率为443±74.27/min[1],347±27/min[2]。前者与本实验的最快心率接近,后者则类似于本实验的平均心率,可见常规心电图记录受当时动物的麻醉状态和挣扎情况影响较大。

本实验提示:大白鼠心率缺乏昼夜波动,可能与小动物持续高代谢状态和该动物的生活规律有关;其次,不易发生心律失常,有人分析动物电生理特点,发现大鼠心肌易损期为2.9mS,而狗则为36mS[3],加之鼠心脏体积小,故电生理较稳定,心律失常发生率低。这也是大鼠能很好耐受实验心脏手术的原因之一。

2.2 大白鼠HOLTER记录方法的研究

HOLTER磁带记录仪,磁带走速1mm/秒,回

* 男,1957—10出生,主治医师