

社会预警研究中判别分析技术的应用*

冯江平¹, 黄三宝¹, 王欣¹, 陈虹¹, 王二平²

(1. 云南师范大学西南边疆稳定与发展研究中心, 云南昆明 650092; 2. 中国科学院心理研究所, 北京 100101)

[摘要] 近年来, 各种社会突发事件和群体性冲突事件时有发生, 极大地影响了社会稳定, 探寻具有实用价值的社会预警技术, 有效开展预测预报, 无疑具有重要的政治意义和社会意义。通过引入社会心理学中的判别分析技术, 建立起多元分析的社会预警判别函数, 运用回代检测和实证检验的方法, 对取自云南省 11 个边疆少数民族县的 382 份调查问卷进行验证, 判别分析判正率达 66.7%, 说明这一技术可用于预测社会群体性事件的发生概率, 并可发现影响群体性事件发生的主要因素。

[关键词] 社会预警; 判别分析; 判别模型; 边疆民族地区

[中图分类号] B84-05 **[文献标识码]** A **[文章编号]** 1000-5110(2010)04-0048-05

一、问题的提出

我国正处在经济转型时期, 社会各阶层的利益在重新分配, 许多体制性、制度性的问题逐渐暴露出来; 加入 WTO 后, 我国经济系统更易受到世界社会和经济动荡的影响; 腐败和社会丑恶现象孳生, 部分国家公职人员的不良工作作风等等, 都会引发群众的不满并构成种种社会不稳定因素。近年来, 国内各种社会突发事件和群体性冲突时有发生, 并呈上升趋势。因此, 开展社会预警研究, 对于构建和谐社会具有十分重要的政治意义和社会意义。由于社会预警研究在国内刚刚起步, 还缺乏能够帮助政府部门对群体性事件进行有效预测预报的实用技术和方法, 因而其不仅具有重要的学术价值, 也更有重要的应用价值。

二、研究思路与目的

对预警地区的居民进行社会态度调查, 并建立预警模型; 然后, 运用判别分析技术, 建立预警的判别函数, 并检验其判别效果; 建立置信区间进行判别, 以有效预知预警地区的状况, 从而及时采取防范措施。

三、研究对象

研究对象的选择, 根据《云南年鉴·2006》提

供的社会、经济数据, 遵循分层随机抽样原则, 于 2008 年 1—3 月, 选取云南省 11 个边疆少数民族聚居县: 宾川县、澜沧县、元江县、玉龙县、屏边县、泸水县、勐海县、盈江县、沧源县、河口县和麻栗坡县的居民作为本研究对象, 共发放问卷 440 份, 回收有效问卷 382 份, 回收有效率 86.8%。

四、研究工具

采用“居民社会态度调查问卷”, 主要包括六方面内容: 一是生活满意度, 包括个人对自身和社会现状的总体判断, 对收入状况、社会地位的满意度等; 二是压力感受状况, 包括子女教育、住房及就业等的压力; 三是对政府工作的满意度, 包括对中央、地方及所在乡镇政府工作的满意度; 四是对社会问题判断, 涉及社会治安、社会风气、环境污染、腐败、贫富差距等热点社会问题; 五是社会风险应对方式(行为意向), 即个人在风险压力下倾向于采取的应对措施, 根据是否有利于解决问题和社会稳定, 将社会公众的行为意向分为 4 种恰当应对行为和 5 种不恰当应对行为——忍受或发牢骚、联合亲友对抗、罢工罢课、集体示威、集体离开当地; 六是个人的背景情况等人口学特

* [收稿日期] 2010-03-15

[基金项目] 国家社会科学基金西部项目“边疆少数民族地区社会预警指数的构建——以云南省为例”(项目号: 07XSH007)。

[作者简介] 冯江平(1955—), 男, 山西陵川人, 云南师范大学教授, 主要研究领域为社会心理学, 人力资源管理。

征。另外,由于研究对象为边疆民族地区,所以,还增加了国家和民族认同感、边境地区安全感等问题。

五、研究方法

运用判别分析技术进行研究,并主要采用Fisher判别法。判别分析(Discriminant Analysis)是多元统计分析中用于判别样品所属类型的一种统计方法,是根据已知类别的事物的性质(自变量),建立函数式(自变量的线性组合,即判别函数),然后对未知类别的新事物进行判断以将之归入已知的类别中。^[1]

Fisher判别法是1936年提出来的,其最大的优点在于对总体的分布并未提出特定的要求。其基本思想是:^{[1][2]}从两个总体中抽取具有p个指标的样品观测数据,借助方差分析的思想构造一个判别函数,亦称作判别式: $y = c_1x_1 + c_2x_2 + \dots + c_px_p$,其中 c_1, c_2, \dots, c_p 确定的原则是使两个总体组间的区别最大,同时每个组内部的离差最小。确定了判别式后,对于一个新的样品,将它的p个指标值代入判别式中求出y值,然后与判别临界值(或称分界点)进行比较,可以判别它应归于哪一个总体。^[3]判别分析结果中报告以下几项内容:标准化典型判别函数,在标准化典型判别函数中,各个预测变量的系数代表了该变量在区分被试时作用的大小;相应的还需要报告各个判别函数的几个指标:特征值(Eigenvalue)、贡献率、典型相关系数以及Wilks' Lambda检验的结果;结构矩阵,代表各个预测变量与几个判别函数之间的相关;

表1 典型判别函数估计表

判别函数	特征值	贡献率	累积贡献率	典型相关系数
1	.223	100.0%	100.0%	.427

表2为判别函数显著性检验表,Wilks' Lambda值为0.817,卡方值为40.441,自由度为5,达到了显著性水平,这表明此判别函数达到了统计学上的显著效果,是有意义的。

表2 典型判别函数显著性检验

判别函数	Wilks' Lambda	Chi-square	df	Sig.
1	.817	40.441	5	.000

从标准化典型判别函数和合并组内相关矩阵

判对率,依据判别函数对各个样本点重新进行分类后与原有分数之间的比例关系;Fisher线性判别函数,可直接用于新被试的分类。^{[4][5]}

六、研究过程

(一)判别函数的建立

根据逐步回归分析的结果,选取了对不恰当行为意向有重要影响的五个自变量最终作为判别分析的自变量,即解决就业问题和推行计划生育两项政府工作的满意状况以及贫富差距、环境污染和农产品销售三个社会问题的严重性判断状况。将五种不恰当行为进行因素分析,得到两个因子,分别命名为:集体不恰当行为意向和个体不恰当行为意向,其中前者包含了联合亲友对抗、罢工罢课和集体示威三种行为意向,后者包含了忍受或发牢骚行为意向。由于群体性事件大多是由群体性不恰当行为意向所导致,故本研究采用集体不恰当行为作为因变量,依据其总分高低各取27%作为高不恰当行为意向组(即群体性事件发生可能性高组)和低不恰当行为意向组(群体性事件发生可能性低组)。

在研究的判别分析中,组别数为两组,预测变量有五个,所以判别函数只有一个。^[4]正如表1结果显示的估计值,判别函数的特征值为0.223,贡献率为100%,典型相关系数为0.427。社会科学的问题特征值为0.223,这说明此函数具有较好的判别效果,^{[6][7]}即自变量可以有效区分不恰当行为意向高、低两个组别,典型相关系数则表示判别函数与组别间关联的程度。

来看,与判别函数相关较密切程度依次是贫富差距问题严重性判断、推行计划生育工作的满意状况、解决就业问题的满意状况、环境污染问题严重性判断和农产品销售问题严重性判断。可以看出,贫富差距的判别效果最明显。具体见表3。

所以得到的标准化典型判别函数为:

$$D1 = 0.512X1 + 0.592X2 + 0.399X3 - 0.546X4 - 0.478X5 \quad (1)$$

表 3 典型判别函数系数

自变量	标准化典型判别系数	合并组内相关矩阵
x1 解决就业问题	.512	-.436
x2 贫富差距	.592	-.422
x3 农产品销售	.399	.409
x4 推行计划生育	-.546	.365
x5 环境污染	-.478	.340

表 4 为未标准化的典型判别函数系数表,未标准化典型判别系数及常数项已在表中列出。

表 4 未标准化的典型判别函数

自变量	未标准化典型判别系数
x1 解决就业问题	.552
x2 贫富差距	.659
x3 农产品销售	.398
x4 推行计划生育	-.596
x5 环境污染	-.457
常数项 (Constant)	-1.407

如表 5 所示,我们得到了两组 Fisher 线性判别函数如下:

第一组即高不恰当行为意向组的 Fisher 线性判别函数为:

$$F1 = 5.233X1 + 4.248X2 + 3.273X3 + 3.008X4 + 3.237X5 - 29.575$$

$$447X4 + 2.808X5 - 30.901 \tag{2}$$

第二组即低不恰当行为意向组的 Fisher 线性判别函数为:

$$F2 = 4.714X1 + 3.628X2 + 2.899X3 + 4.008X4 + 3.237X5 - 29.575 \tag{3}$$

表 5 Fisher 判别函数系数

自变量	不恰当行为意向分组	
	高不恰当	低不恰当
x1 解决就业问题	5.233	4.714
x2 贫富差距	4.248	3.628
x3 农产品销售	3.273	2.899
x4 推行计划生育	3.447	4.008
x5 环境污染	2.808	3.237
常数项 (Constant)	-30.901	-29.575

(二)判别分析的检验

判别分析的特点是根据已掌握的若干样本的数据信息总结出客观事物分类的规律性,建立判别公式和判别准则然后,当遇到新的样本点时,只要根据总结出来的判别公式和判别准则就能判别该样本点所属的类别。而对未知类别的样本进行判类,就很可能发生误判,所以要对判别分析的结果进行检验,计算样本错判比率和正确判定的比率则是一种检验判别效果最实用的方法。所谓错判就是把原来为第一类的样本判给了第二类。对于正确判定的比率应该达到多少才能接受,并没有严格的规则。不过一些研究者建议,正确判

定的比率至少应该超过随机分类比率的 25%。例如:如果分成两类,正确判定的比率应该超过 75% (100%/2+25%);如果分成三类,正确判定的比率应该超过 58% (100%/3+25%) 才能考虑接受判别分析的结果^[7]。本文对判别分析结果采用回代判定,原有的分类当中,第一类的正确率为 66.7%,第二类的正确率为 67.0%,总的正确率为 66.8%。因为本研究是基于居民社会态度的调查,态度的影响因素很复杂,这就导致了社会心理问题比一般的经济类、自然科学类的研究正确率要低一些,故我们认为此结果是可接受的。

表6 分类结果

Group	观察值	Predicted Group Membership		Total
		高不恰当	低不恰当	
Original	高不恰当	34	103	372
	低不恰当	69	103	131
	分类正确百分比	33.3%	100.0%	100.0%
		67.0%	100.0%	100.0%

本研究对预警地区的年数据进行实证检验,随机选取了A地和B地作为判别分析检验的实证效标,将A地和B地的样本数据代入:

$$F1 = 5.233X1 + 4.248X2 + 3.273X3 + 3.447X4 + 2.808X5 - 30.901 \quad (2)$$

$$F2 = 4.714X1 + 3.628X2 + 2.899X3 + 4.008X4 + 3.237X5 - 29.575 \quad (3)$$

即得到两地的行为意向组 Fisher 线性判别函数,结果显示,A地有66.7%的个体可以归为高不恰当行为意向组;B地有65.8%可以归为低不恰当行为意向组。

根据调查员反应的情况和反馈的社情观察报告,以及对调查问卷数据进行的描述统计分析的结果,A地居民的社会态度和各方面的满意度较低,边境安全和民族问题反映较严重,居民选择消极或过激行为意向可能性的百分比较大;而B地各方面的情况反应较好,这恰好佐证了上述实证检验的效果。由此可见,所建立的判别函数判正率比较高。

(三)高不恰当行为意向预警判别的尝试

通过以上判别函数的建立及判别效果的分析 and 验证,根据以上高不恰当行为意向组的 Fisher 线性判别函数来判断各地的行为意向类别,采用95%和99%的置信区间进行估计。随机选取C地建立预警判别区间。将C地样本数据代入高不恰当行为意向判别函数,该数据95%和99%的置信区间分别为[20.7%,76.3%]和[11.9%,85.1%]。这样要判定以后C地每年行为的类别,就将C地的数据代入判别函数统计预测高不恰当行为的个体所占的比例,当落入[20.7%,76.3%]和[11.9%,85.1%]区间或超过区间的上限时,可以预测95%或99%的可能性属于高不恰当行为。运用此函数判断C地数据中高不恰当行为意向的比例为53.7%,落在了置信区间内,即此地高不恰当行为意向比较突出。同样,其他地

区的判断也依据此研究逻辑和方法进行,不再一一阐述。

七、讨论分析

(一)判别分析技术可以较有效地预测受调查地区群体性事件发生的概率

在判别函数的建立及检验的基础上,运用 Fisher 判别法可以对未知样本的分类进行判别,即将待判样本的数据分别代入 Fisher 判别函数中,根据数值大小判断是否将其行为意向归为该类别。本研究所建立的判别函数的特征值为0.223,这在对社会心理学问题的研究中是可以接受的,所建立的判别函数达到了统计学上的显著效果,是有意义的。本研究所建立的判别函数通过回代检验和实证校标检验的结果都比较理想;同时,运用置信区间进行预警判别,随机选取一个样本地区估计其区间的上下限,判别效果较好。因此,运用此技术和研究思路,可对其他样本地区的群体性事件发生的可能性进行预警判别,帮助政府部门进行社会预警,从而能够未雨绸缪,及早采取防范措施。

(二)通过判别分析可发现影响群体性事件发生的主要因素,其也是有效的预测变量

从判别分析的结果看,就业问题、贫富差距、农产品销售、计划生育和环境污染等五个方面的问题,能够显著区分边疆少数民族地区的高、低两类不恰当行为意向。也就是说,政府对这五方面问题解决得好坏,将在很大程度上决定着该地区居民的行为取向——即影响群体性事件发生概率的主要因素。因而它们可作为判别当地居民行为意向的预测变量,并以此来判定待判样本的类别。

研究结果表明,受调查地区居民的不恰当的行为意向,应当成为关注的重点。对群体性事件发生状况的判别,可主要针对不恰当的行为意向进行划分和判别,将不恰当行为划分为高、低两组,有利于找出更具危险性的样本群体,并提高社

会预警的准确性。

(三)判别分析技术的应用还需不断完善

回代判定发现,原有分类中第一类的正确率为 66.7%,第二类的正确率为 67.0%,判正率虽可接受,却不是很高。随机选取 A 地和 B 地作为判别分析检验的实证校标,将样本数据代入 Fisher 线性判别函数进行判定的结果也基本类似。而随机选取 C 地建立预警判别区间,将 C 地样本数据代入判别函数,其结果虽落在置信区间内,而判别比例为 53.7%,显示判别结果似乎不够稳定。其原因,一是社会心理问题的影响因素极其复杂,实证计算有一定困难;二是本研究中可能还有一些影响因素尚未探明,今后还将对这一方法的实际应用进行更多的后续研究以及检验和修正。此外,也需寻找其他更为有效的社会预警的研究技术和方法。

[参 考 文 献]

[1] 于秀林,任雪松.多元统计分析[M].北京:中国统计出版社,2003.

[2] 李映,焦李成.基于核 Fisher 判别分析的目标识别[J].西安电子科技大学学报(自然科学版),2003,(2).

[3] 周宏力,董景荣.上市公司投资价值的 Fisher 判别分析[J].统计与决策,2006,(7).

[4] 吴明隆.SPSS 统计应用务实[M].北京:中国铁道出版社,2000.

[5] 赵良渊,何大卫,王彤.多元线性回归方程中共线影响点的诊断[J].中国卫生统计,2004,(2).

[6] 林清山.多变项分析统计法—社会及行为科学研究适用[M].台北:台湾华东书局股份有限公司,1995.

[7] 王宏伟.我国上市公司财务困境预测的实证研究[D].成都:西南财经大学,2002.

The Application of Discriminatory Analysis to the Research of Early-warning System for Social Emergency Events

FENG Jiang-ping¹, HUANG San-bao¹, WANG Xin¹, CHEN Hong¹, WANG Er-ping²

(1. Research Center of Borderland Stability and Development, Yunnan Normal University, Kunming 650092, China; 2. Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)

Abstract: In recent years in China, there have been some social emergency events and group conflicts which have much disturbed our social stability. The study of practical and effective early-warning technology for such events has much political and social significance. Adopting the technology of discriminatory analysis in social psychology and the multi-analysis of the discriminant function for the early-warning system plus some tests, this research has analyzed 382 questionnaires from eleven border and minority counties in Yunnan Province, and concluded that the accuracy of this technology is 66.7%, which has proved to be effective in predicting the probability of emergency events and group conflicts as well as the possible causes.

Key words: early-warning system for emergency events; discriminatory analysis; discriminative model; border and minority areas

[责任编辑: 邹建达]