

农业经济管理分类的研究

达瓦¹, 卓玛², 晓红¹ (1. 西藏农牧学院公共教学部, 西藏林芝 860000; 2. 西藏农牧学院教务科, 西藏林芝 860000)

摘要 运用多元统计分析中的分层聚类分析和判别分析方法, 借助 SPSS 统计软件进行统计分析, 以西藏自治区的 35 个农业县为研究对象, 根据 2007 年西藏农业县农林牧渔业产值、粮食产量、工业总产值等主要统计指标为分类变量, 对农业县进行分类, 根据所得出的结论, 客观、合理地提出合理性的建议。

关键词 聚类分析; 判别分析; 农业经济

中图分类号 F302.5 **文献标识码** A **文章编号** 0517 - 6611 (2009)08 - 03859 - 02

Classification Study on the Agriculture Economy Management

DA Wa et al (Public Teaching Department, Tibet Agricultural and Animal Husbandry College, Linzhi, Tibet 860000)

Abstract Making use of the method of the cluster analysis and discriminate analysis, with the help of SPSS, taking 35 Tibetan agricultural counties as the research objects, and the main covariance index of Tibetan agricultural counties, such as agricultural, forestry, animal husbandry, and fishery production value, grain yield and the total production industrial value as variables, and based on the conclusion, objective and reasonable suggestions were put forward.

Key words Cluster Analysis; Discriminate Analysis; Agricultural economy

西藏自治区有 1 个市辖区、1 个县级市和 71 个县, 根据产业结构、自然地理环境和开发特点, 将其分成 21 个边境县、35 个农业县、14 个牧业县、24 个半农半牧县、18 个“一江两河”开发县和 11 个粮食基地县。由于西藏地处青藏高原的西南部, 受特殊地理位置的影响, 各县的产业结构有所不同, 所以科学、合理地分析农业经济问题具有重要的意义。笔者以 35 个农业县和 24 个半农半牧县中的 2 个县为研究对象, 以 6 个主要统计指标为因子, 应用多元统计分析中分层聚类分析和判别分析原理, 借助 SPSS 统计软件研究 35 个农业县的经济状况的分类。

1 数据来源

数据来源于《西藏统计年鉴 2008》的 2007 年各县(区、市)主要统计指标。把主要统计指标乡村从业人数 (X_1)、农林牧渔业产值 (X_2)、工业总产值 (X_3)、农林牧渔业产值比前一年增加的指数 (X_4)、农作物播种面积 (X_5)、粮食产量 (X_6) 定义为变量, 得到原始数据, 见表 1 (略去其他的只显示部分数据)^[1]。

表 1 原始数据
Table 1 Original data

县(市) Counties (Cities)	X_1	X_2 万元	X_3 万元	X_4 万元	X_5 hm ²	X_6 万 t
城关区	7 964	7 132	117 944	89.7	1 524	0.39
尼木县	15 135	7 221	612	95.7	2 443	1.27
曲水县	17 971	11 576	8 033	105.4	4 222	2.45
堆龙德庆县	22 029	15 317	30 279	106.6	5 545	2.48
达孜县	15 734	12 496	12 169	115.0	4 568	2.48
墨竹工卡县	14 585	22 674	28 903	134.6	5 215	2.32
左贡县	19 842	17 952	806	125.5	3 719	1.62
...

2 数学原理

聚类分析 (Cluster Analysis) 和判别分析 (Discriminate Analysis) 是用于解决分类问题的多元统计分析方法。聚类分

析是根据事物本身的特性对被研究对象进行分类, 使得同一类中个体有较大的相似性, 不同类中的个体有较大的差异。采用分层聚类, 是根据样本的亲疏程度将最相似的对象结合在一起, 以逐次聚合的方式把所有个案分类, 逐步合并直到最后合并成为一类。判别分析是根据已知类别的事物性质, 利用建立的判别函数, 对未知类别的新事物进行判断使之归入已知的类别中^[2]。

3 统计量结果

3.1 分层聚类分析 由于各指标的观测值量纲不同或者在相同量纲上数量级不同, 对原始数据进行标准化处理, 得到标准化数据 (表 2)^[3]。

表 2 原始数据的标准化结果
Table 2 Standardized results of original data

县(市) Counties (Cities)	X_1	X_2	X_3	X_4	X_5	X_6
城关区	-0.87	-0.63	4.61	-1.58	-1.05	-1.08
尼木县	-0.14	-0.62	-0.41	-1.17	-0.71	-0.46
曲水县	0.14	-0.08	-0.09	-0.49	-0.06	0.37
堆龙德庆县	0.55	0.39	0.86	-0.41	0.42	0.40
达孜县	-0.08	0.04	0.08	0.18	0.06	0.40
墨竹工卡县	-0.20	1.31	0.80	1.54	0.30	0.28
左贡县	0.33	0.72	-0.40	0.91	-0.25	-0.21
...

在 SPSS 统计软件下运行 Analyze→Classify→Hierarchical Cluster, 选定相关设置, 运行后输出, 得到样本处理的概要情况, 35 个个案都进入分析过程, 均为有效样本。分层聚类的第 1 步使观察值 14 与 25 合并, 合并后的差异系数为 0.121; 第 2 步使观察值 16 与 24 合并, 合并后的差异系数为 0.140; 依次合并, 直到将所有观察值全部合并为止^[4]。

由图 1 可知, 聚类为 4 的类别是: 第 1 类 (拉萨城关区)、第 2 类 (其余 32 个县)、第 3 类 (日喀则市)、第 4 类 (南木林县)。从分类的结果看, 拉萨城关区和日喀则市分别作为市辖区和县级市, 在交通、通信、科技和政策等方面有着得天独厚的有利条件, 其经济实力和效益都要比其他县优越很多, 它们各自归为一类正好说明了这一点; 而“南木林县”作为日

作者简介 达瓦 (1968 -), 男, 藏族, 西藏扎囊人, 副教授, 从事数学模型、应用数学方面的研究。

收稿日期 2008-12-10

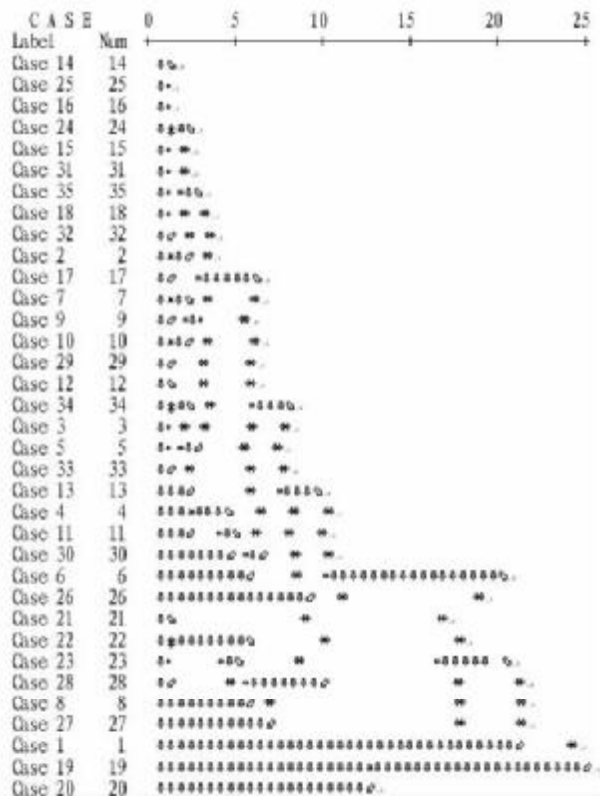


图1 分层聚类的冰状示意

Fig.1 Dendrogram of hierarchical clustering

喀则地区最大的县,是全区乡村从业人员人数最多的一个县,地理上与地区首府很近,也决定了其单独为一类的地位。分类的结果与实际情况比较吻合,说明了聚类分析的结果是有效的。第2类的32个农业县是西藏农业的主力军,基本分布在偏远的基层,各方面条件较差,在农业经济管理和政策方面应该是西藏政府和农业部门重点需要考虑的地区。其他3类中3个县(或区)在科技推广、产业结构调整、农村现代化等方面起率先典范作用,当好全区的农业经济跨越式发展和可持续发展的领头羊。

3.2 判别分析 在进行分类研究中,常常把聚类分析和判别分析结合起来用,一般是先进行聚类分析,再进行判别分析,就可以进行样品的识别。判别分析是在事先有聚类分析结果的情况下,根据一定的准则把被研究对象归类。此处从24个半农半牧县中挑出林周县和昌都县为被研究对象,在前面的聚类分析基础上,分析这2个县在经济状况上应该与哪些县基本接近。为此,重新列出数据表格,在最后一列添加一个“类别序列”列,在2个县的两个统计指标已知情况下,其“类别序号”填写为“待定”后得数据(表3)。

表3 判别分析数据(data)

Table 3 Data of discriminate analysis

县(市) Counties (Cities)	X_1	X_2 万元	X_3 万元	X_4 万元	X_5 hm ²	X_6 万 t	类别序号 Sort sequence
城关区	7 964	7 132	117 944	89.7	1 524	0.39	1
尼木县	15 135	7 221	612	95.7	2 443	1.27	2
曲水县	17 971	11 576	8 033	105.4	4 222	2.45	2
...
林周县	12 435	15 761	491	127.8	5 977	2.05	待定
昌都县	9 615	13 907	131	123.2	3 533	1.06	待定

在 SPSS 统计软件下运行 Analyze→Classify→Discriminant, 选定相关设置, 运行后输出, 得到有效样本 35 个, 占 94.6%; 分类变量范围之外的个案有 2 个, 占 5.4%。3 个典型判别函数的方差百分比分别是 64.1%、20.4%、15.4%, 典型相关系数分别是 0.861、0.690、0.638。经典判别函数的卡方检验可知, 3 个典型判别函数卡方检验的差异(0.05 水平)都不显著, 表明了各组变量的均数不存在显著的差异, 表示适合作判别分析^[5]。从标准化的典型判别系数可得出判别函数为:

$$y_1 = -0.391x_1 - 0.364x_2 + 0.959x_3 - 0.561x_4 + 1.349x_5 - 0.927x_6; y_2 = 0.618x_1 + 0.335x_2 + 0.438x_3 + 0.525x_4 - 0.038x_5 - 0.434x_6; y_3 = -0.607x_1 + 0.382x_2 + 0.082x_3 - 0.122x_4 - 1.427x_5 + 2.199x_6。$$

根据判别函数, 林周县和昌都县归为第 2 类, 由此可以认为这 2 个县在经济状况上与处于第 2 类的各农业县相当。研究其他一些县与已经分好类别的县进行比对或者排比, 可以重复以上的判别分析步骤来实现比较。

4 结语

聚类分析和判别分析已经广泛应用到作物品种分类、土壤分类、经济分析、地质勘探、天气预报等各个领域。通过分析可以知道, 对西藏农业县进行分类, 有利于农业经济状况和效益的比较、排序, 为农业县经济管理、决策部署、产业结构调整和生产经营活动提供咨询服务。运用多元统计分析原理, 借助 SPSS 统计软件进行分析可以使分类更科学合理、符合实际。

参考文献

[1] 中国统计局. 西藏统计年鉴 2008 [M]. 北京: 中国统计出版社, 2008: 278-310.
 [2] 袁志发. 多元统计分析 [M]. 北京: 科学出版社, 2002.
 [3] 周爽. 社会统计分析—SPSS 应用教程 [M]. 北京: 清华大学出版社, 2006.
 [4] 高祥宝. 数据分析与 SPSS 应用 [M]. 北京: 清华大学出版社, 2007.
 [5] 李清明. 统计学原理 [M]. 上海: 复旦大学出版社, 2004.

(上接第 3835 页)

[14] 董伦, 解缙, 胡广, 等. 《明太祖实录》卷 243 [M]. 洪武二十八年十一月戊寅.
 [15] 易时中. 《夏津县志》卷 2《食货志》[M]. 天一阁藏明代方志选刊本.
 [16] 邓鞅. 《濮州志》卷 2《食货志》[M]. 天一阁藏明代方志选刊本.
 [17] 周家齐. 《高唐州志》卷 3《仓储》[M]. 光绪 33 年刻本.
 [18] 陈嗣良 [M]. 曹县志 (上部) [M]. 山东省曹县档案馆, 1981: 131-132.
 [19] 吕鹏云. 《巨野县志》卷 8《杂志·物产》[M]. 崇祯间增刻本.
 [20] 于俊. 《临清州志》卷 11《市廛志》[M]. 乾隆元年刻本.
 [21] 裴景煦. 《曹州府曹县乡土志·商》[M]. 光绪 33 年抄本.

[22] 王象晋纂辑. 伊钦恒诠释. 群芳谱诠释 [M]. 北京: 农业出版社, 1985: 155.
 [23] 谢国楨. 明代社会经济史料选编 (上) [M]. 福州: 福建人民出版社, 1980: 50.
 [24] 陈振汉. 清实录经济史料·农业编 (第 1 分册) [M]. 北京: 北京大学出版社, 1989: 10.
 [25] 许慎. 明清时期山东商品经济的发展 [M]. 北京: 中国社会科学出版社, 1998: 30-31.
 [26] 陆武. 《山东通志》卷 8《物产》[M]. 天一阁藏明代方志选刊本.