维普资讯 http://www.cqvip.com

59-64

第25卷 第4期 西北农业大学学报 Vol. 25 No. 4
1997年8月 Acta Univ. Agric. Boreali-occidentalis Aug. 1997

农业产业区划中的主元分析

刘静义[1] 王明俊[2] 温天舜[1]
(1 郑州粮食学院基础部.2 河南电力调度通信局.郑州 450052)

摘 要 以黄淮海平原(河南省部分)的农业产业区划为例,用主元分析法讨论了分区划片过程,并建立了相应的规则和模式.

关键词 农业产业区划,主元分析,黄淮海平原
中图分类号 F304.5, S11

农业产业化被称为中国农村改革的"第二次飞跃",它虽立足于农业资源条件,但涉及自然和社会、经济、技术等方方面面的情况。黄淮海平原地区是河南省农业的主体部分(占全省总人口的66.6%,约占总耕地的70%)。河南省黄淮海平原综合开发方案[1]的制定始于1988年,已通过有关部门鉴定,取得了阶段性成果并在继续实施中,为区划模型的讨论提供了良好的基础。但因起步较早,覆盖面宽,区划上的数学处理有可改进之处,尚难作为新形势下农业产业区划的范例。本文将对区划的数学方法及相应法则作新的探索。

1 变量主元分析

文献[1]在区划中选择了21个变量,即①地貌;②无霜期;③年降水量(mm);④沙丘面积/沙土面积(%);⑤沙土面积/土地面积(%);⑥盐碱面积/土地面积(%);⑦砂姜黑土/土地面积(%);⑧林地面积/土地面积(%);⑨草地面积/土地面积(%);⑩水面面积/土地面积(5);⑪棉花面积/土地面积(%);⑫油料面积/土地面积(%);⑬果园面积/土地面积(%);⑭可灌面积/耕地面积(%);⑮耕地面积/土地面积(%);⑯万公顷排灌机械(台);⑰人口密度(人/hm²);⑱人均耕地(hm²);⑲粮食单产(kg/hm²);⑳人均肉类产量(kg/人);㉑人均纯收入(元/人)。然后把各地的实测数据用标准差标准化方法[2]处理后,用系统聚类(类平均)法[3]得到初步结果,再根据自然条件和社会经济技术条件相对一致性、开发治理方向和关键措施的一致性和行政区划的完整性3项原则,分成8个综合开发治理区。从数学角度看,有以下问题值得商榷:

1)变量选择缺乏依据。如果说所选变量仅限于农业生产开发内容,是因为"起步早,覆盖面宽"、受当时形势和认识水平局限的话,那么,规划农业生产开发而不考虑光照、肥力等因素的影响,则不应该。

2)聚类方法应改进。其一,不应忽视各变量间存在的相似、交叉、相悖关系,让它们以同等地位参与聚类,势必导致彼此的重迭、抵消,一定程度上要抹杀或歪曲区片自身特征和彼此差异。应该说,文献[1]中聚类结果不理想,是有这方面原因的。其二,没有聚类图、

收稿日期 1996-12-20
课题来源 河南省"九五"社科规划项目
作者简介 刘静义,男,1943年生,副教授

维普资讯 http://www.cqvip.com

分区图，不便参照。

3)调整划片的3项原则有主观随意性。

要抓住主要矛盾，使决策更具科学性，使区划方法更规范、更简明，进行变量主元分析[4~6]是必要的。为便于比较，这里仍用文献[1]中数据。

记样本资料阵为 $X_{72\times21}$(72行，每行对应一个县级地方；21列，每列对应前文中的一个变量)，用下式

$$R=\frac{1}{n-1}X^TX \qquad (n=72) \tag{1}$$

求出变量间的相关矩阵 $R_{21\times21}$。(1)式中，X^T 表示数据阵 X 的转置阵。由

$$|R-\lambda E|=0 \tag{2}$$

找出 R 的各特征根 $\lambda_1,\cdots,\lambda_m$(从大到小排列)。式中 E 是与 R 同阶的单位阵。解

$$\begin{cases}(R-\lambda E)=0\\ L_\theta^TL_\theta=1 \qquad (\theta=1,\cdots,m)\end{cases} \tag{3}$$

得与 $\lambda_\theta(\theta=1,\cdots,m)$ 对应的特征向量 L_θ，以各特征值(根)累积贡献率≥0.85为标准，选 m 个特征向量构造新变量 y_θ

$$y_\theta=x_{i1}l_\theta+\cdots+x_{i21}l_{\theta21} \qquad (i=1,\cdots,72;\theta=1,\cdots,m) \tag{4}$$

取代原数据阵 X 的各列变量，得主数据阵 $Y_{72\times10}$

$$Y=XL \tag{5}$$

本例中 $m=10$，$L_\theta=(l_{\theta1},\cdots,l_{\theta21})^T$，其具体值见附表。

附表 选出的10个特征向量值

对应变量编号	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}
1	−0.0126	−0.2538	0.1046	−0.1862	−0.2589	−0.3817	−0.3141	−0.3037	−0.4239	0.1 27
2	0.3544	0.0916	0.2417	−0.1326	−0.0320	0.1674	−0.0254	−0.0527	0.1886	−0.3083
3	−0.3853	0.0083	0.1293	−0.0698	0.1559	0.1558	0.1026	0.0215	0.1592	−0.0894
4	−0.1653	−0.2054	0.0454	−0.1112	0.4518	0.1594	0.0757	−0.1560	0.3745	0.5055
5	−0.2132	−0.3116	−0.1350	−0.1641	−0.0236	0.0107	−0.0467	−0.1105	−0.0182	−0.5817
6	−0.1340	−0.1597	−0.3445	0.3094	0.1749	−0.1593	0.2000	−0.1478	0.1682	−0.1696
7	0.3180	0.1295	0.0048	−0.1594	0.2991	0.1727	−0.1137	−0.2065	−0.2035	0.0339
8	−0.1357	−0.2615	0.3065	0.0069	0.2587	0.2095	−0.3993	0.0950	0.1407	−0.0424
9	−0.0574	−0.1054	0.2701	−0.0615	0.3564	−0.1418	0.3598	0.4464	−0.4636	−0.1908
10	0.1670	−0.2308	0.2312	0.2098	−0.1498	0.1656	−0.2144	0.2696	−0.1373	0.3345
11	−0.2314	−0.0387	−0.2348	−0.2897	−0.1318	0.3516	0.0864	0.4267	−0.0456	0.1386
12	0.2559	−0.2717	−0.0662	−0.2021	0.0527	−0.1261	0.2135	−0.2629	−0.1303	0.0475
13	0.2020	−0.2922	0.0949	−0.3700	0.1280	−0.0049	−0.2460	0.0964	0.2096	−0.1575
14	−0.3666	0.1345	0.1365	0.1076	−0.0292	0.0520	0.0376	−0.0752	−0.1095	0.0104
15	−0.0368	0.2512	−0.2942	−0.4727	0.1596	0.0224	−0.0034	−0.0112	−0.1059	0.1633
16	0.2414	0.0364	0.2527	0.1517	0.2790	0.2199	0.2236	−0.2829	−0.2979	0.0362
17	−0.0529	0.4141	0.0604	−0.2168	0.2011	−0.2547	−0.2428	0.0843	−0.1094	−0.0548
18	0.0998	−0.3332	0.1888	−0.2209	−0.1625	0.3332	0.3427	−0.1520	−0.0908	0.1183
19	−0.1719	0.2748	0.3560	−0.1551	−0.0848	0.1524	0.1437	−0.2622	0.1657	−0.0890
20	0.0236	0.0015	0.3191	−0.2084	−0.2470	−0.4504	0.3605	0.2134	0.2700	0.0514
21	−0.3034	0.1071	0.2216	−0.1715	−0.3012	0.1819	−0.0492	−0.1781	−0.0013	0.1386

维普资讯 http://www.cqvip.com

2　主元涵义分析

2.1　主元涵义

本例求出了10个主元,这说明描述自然和社会经济技术条件的变量众多、角度各异,且数据在不同方向上的分布是相对均匀的。相关阵 R 各特征根 $\lambda_1 \sim \lambda_{10}$ 依次为0.2234,0.1589,0.1060,0.0773,0.0661,0.0571,0.0479,0.0445,0.0377,0.0318,其大小差别不太显著也说明了这一情况。

因后几个特征根数值较小,对应贡献率有限,下面把分析重点放在前边几个主元上。

(4)式表明,对主元 y_θ 影响最大的,是特征向量 L_θ 中绝对值最大的分量,若干个绝对值较大分量对应条件变量的共同作用,是决定该主元性质的主导因素。在附表中,用数据下边加横线标出了绝对值超过0.3000的各 L_θ 分量位置,从左边第一列可找出它所对应变量的编号。

在第一主元 y_1 中,照上述办法可查出起主要作用的依次是:0.3853(3.年降水量)、−0.3666(14.可灌面积/土地面积)、0.3544(2.无霜期)、0.3108(6.砂姜黑土/土地面积)、−0.3034(21.人均收入)。以上的数字是 L_1 绝对值较大的分量,后面括号中的数码为对应变量的编号。因年降水量、无霜期、砂姜黑土所占比例等生态因素对应的各系数和 y_1 同号,它们之中的一个或多个增大(或减小),将导致第一主元 y_1 的增大(或减小)。这说明,目前农业科技含量不足、机械化程度不高、人为影响气候和改造环境(如改土治水)能力有限,一定程度上还停留在"靠天吃饭"的自然状态。可灌面积所占比例和人均收入对应系数与 y_1 变化方向相反(异号),意味着当人为努力的影响不够大,经济条件较差时,对"天"的依赖程度会变得更大。这就是说,第一主元综合表达的是生态因素对农业经济的影响,它是农业产业区划中须首先考虑的问题。

在第二主元 y_2 中,占据主导位置的是:0.4141(17.人口密度)、−0.3332(18.人均耕地)、−0.3116(5.沙地面积/土地面积)。说明人口总量过于庞大、城市化水平太低,是制约农业产业化进程的头等障碍。决不能忽视我国人均耕地低于世界平均水平,而人口密度却大于几乎所有国家这一严峻现实。一方面,必须坚定不移地执行计划生育国策,千方百计解决人口素质低、贫困人口多、就业压力大(特别是要开辟1.5亿农业剩余劳力的就业门路)等紧迫问题。另一方面,要以切实有效的措施防止耕地退化(沙化),改造中低产田,提高单位面积产量和复种指数,严格控制非农占地,大力开发宜农荒地。因此,称这一主元为生产条件因素是合适的。

第三主元 y_3 中以下变量位居前列:0.3560(19.粮食单产)、−0.3445(6.盐碱面积/土地面积)、0.3191(20.人均肉类产量)、0.3065(8.林地面积/土地面积)。其中,粮食单产和人均肉类产量反映生产发展水平,盐碱地面积表达中低产田的影响,林地比例是生态环境优劣的标志(后二者是对生产条件和生态环境重要性的强调)。所以,应称这一主元为生产水平因素。

对其它主元的解释,可仿照上面的分析进行。为突出主要矛盾,不妨从第4～10主元($y_4 \sim y_{10}$)的第一主条件分别为耕地/土地、沙丘/沙土、人均肉类产量、林地/土地、草地/土地、草地/土地、沙地/土地着眼进行分析。

维普资讯 http://www.cqvip.com

显而易见，后7个主元各条件变量仍属于生态、生产条件、生产水平范畴，可简单地将其理解为是对前3个主元的加强和补充。

以上分析表明，生态、生产条件和生产水平三因素，是农业产业区划要考虑的主要方面(主变量)。

2.2 主元外延分析

以下将三主元按其内涵向外延伸，运用到区划中去。

从农业角度看，生态因素包含：气候、土壤、生物、地理、人为垦殖活动等影响农作物生长和分布的环境条件。既然如此，将农业产业区划和三主元的结合点，选在流域上是比较合适的。这是因为①生物群落和地理环境都沿流域作梯度变化，一个流域构成一个自然系统，是一个独立的存在；②农业产业化要实现农工贸一体化，或者说要以市场经济下的工业方式经营农业，这就使得它的建设成了一项系统工程，须从现有经济基础(其中少不了生产条件和生产水平)入手，整体布局，多向开发，加长生产链条，实行规模经营，加快流通过程，提高产出效益；③目前农业开发的中心是改土治水，综合考虑全流域的兴利除害规划，从全流域的自然(生态因素)和社会经济技术条件(生产条件、生产水平因素)着眼，统筹兼顾各经济部门的不同要求，相应制定的分期方案、工程措施、指标体系和经济效益，都具有较高水准的大局观念和合理性。

综上所述，“以流域划片，按政区调整”可作为农业产业区划的实施规则。之所以要“按政区调整”，是因为政区划分本身就立足于地域特征，这样作也便于调控指导和组织管理。应该说，这一规则比文献[1]中的“三项原则”减少了主观随意性，更好地体现出大局观念和实事求是的精神。

3 主元聚类与区划

称从主数据阵 Y 入手的聚类为主元聚类，以区别文献[1]中直接从数据阵 X 出发的聚类。只要能恰当地筛选出主元，它将比普通聚类更好地体现事物自身因个性分离和共性聚集的倾向。

在本例中，先用下式计算各行(地方)间的距离 D_{ik}

$$D_{ik}=\sum_{j=1}^{m}(y_{ij}-y_{kj})^2 \qquad (i,k=1,\cdots,72;i\neq k,m=10) \tag{6}$$

再用递推公式(系统聚类法)

$$D_{ri}=\alpha_p D_{ip}+\alpha_q D_{iq}+\beta D_{pq}+\gamma|D_{ip}-D_{iq}| \tag{7}$$

迭代，(按类平均方式)在(7)式中

令
$$\alpha_p=\frac{n_p}{n_r},\alpha_q=\frac{n_q}{n_r},n_r=n_p+n_q,\beta=\gamma=0 \tag{8}$$

这样，得出了优于文献[1]中的结果。

在主元聚类基础上，依据“以流域划片，按政区调整”规则，即可得出聚类图(图1)和区划结果(图2)。

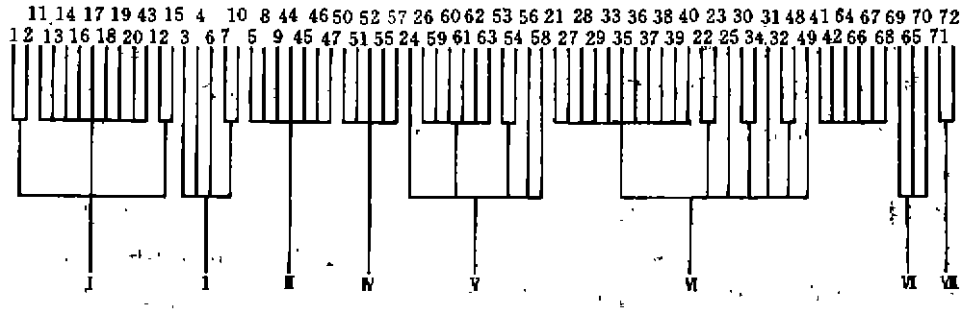

图 1　主元聚类图

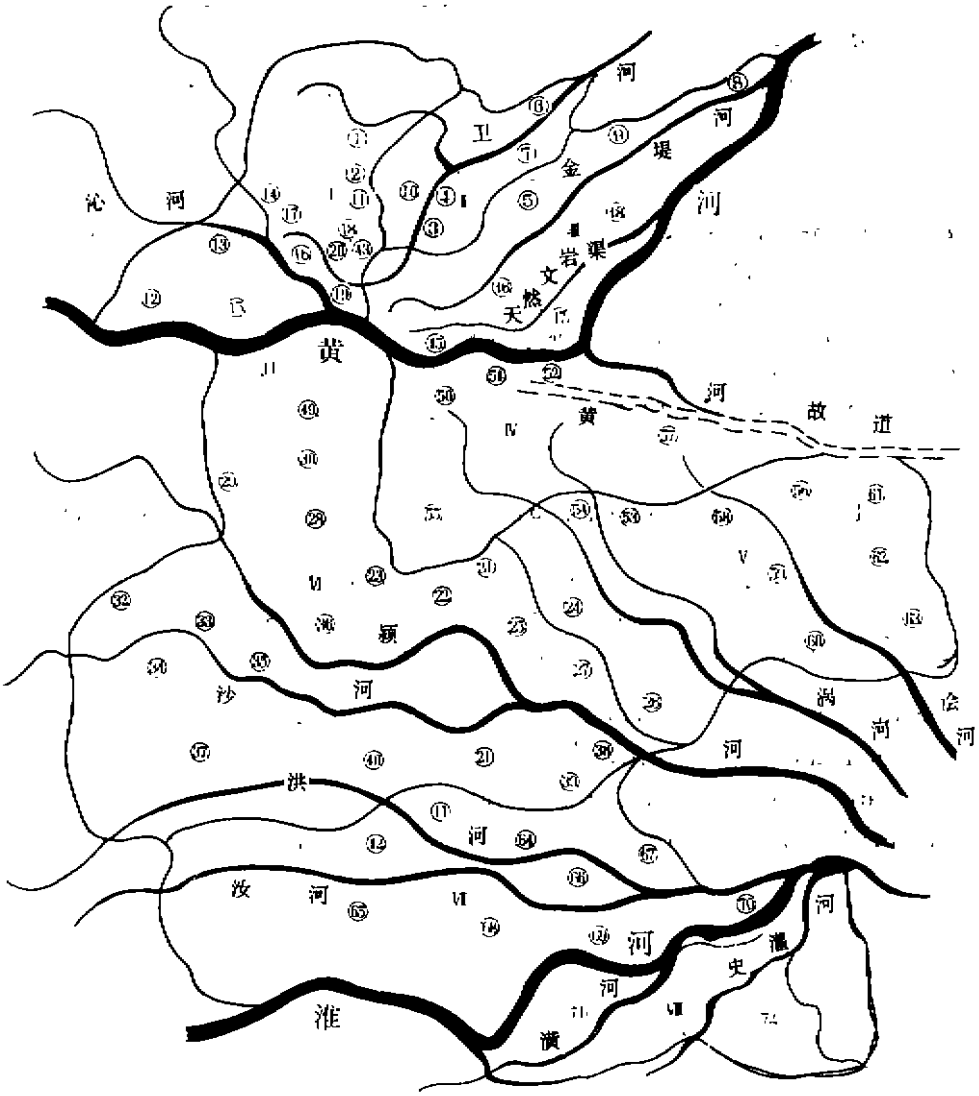

图 2　区划示意图

维普资讯 http://www.cqvip.com

图2中，编号Ⅰ～Ⅷ代表下列各产业区：

Ⅰ 沁游河倾斜平原褐土区；Ⅱ 卫河-马颊河泛滥平原潮土区；Ⅲ 金堤河-天然文岩渠洼涝平原沙碱土区；Ⅳ 黄河故道沙丘洼地沙碱区；Ⅴ 沱涂-涡惠河泛滥平原潮土区；Ⅵ 沙颍河倾斜泛滥平原褐土潮土区；Ⅶ 洪汝河淮北低平原砂姜黑土区；Ⅷ 淮河-史灌河波状平原水稻土区。

高新技术本质上是数学技术，一项研究水平的高低，其数学含量和质量是重要标志。这一论点，已被越来越多的科技工作者认同。加上考虑到农业产业化在中国尚处于起步阶段，而区划是其前期工作中的关键环节，有必要规范其方法和模式，这是本文的出发点。因资料所限，示例侧重于农业基础条件分析。一般来说，农业产业区划还要涉及到市场因素、生产能力、主导产业、中介组织等方面的变量，问题要复杂得多。但是，这些变量的加入，只会增大主元分析的必要性，而不是减弱它。虽然，那时得出的主元和区划规则具体内容可能有变化，但本文数学模型的意义不受影响，其合理性和有效性经得起检验。

参考文献

1 河南省黄淮海平原农业开发办公室. 河南省黄淮海平原中低产田综合治理与农业技术开发. 1990年11月

2 陈希孺. 数理统计引论. 北京：科学出版社，1981

3 王学仁、王松桂. 实用多元统计分析. 上海：上海科学技术出版社、1990

4 天津大学概率教研室. 应用概率统计. 天津：天津大学出版社，1990

5 刘静义，王有安、温天舜等. 高等数学新讲(下册). 北京：北京科学技术出版社，1995

6 温天舜，周大强，刘一勋等. 随机数据分析与处理. 开封：河南大学出版社，1993

The Pivot Analysis in Agro-Industrial Zone Distribution

Liu Jingyi[1] **Wang Mingjun**[2] **Wen Tianshun**[1]

(1 *Department of Basic Courses, Zhengzhou Cereal College, Zhengzhou*, 450052)

(2 *Henan Provincial Department of Electricity Distribution, Zhengzhou*, 450052)

Abstract Based on the accurate instances of the agro-industrial zone distribution in North China Plain, the zone distribution is discussed as well as its corresponding scheme and model are established by means of pivot analysis.

Key words agro-industrial zone distribution, pivot analysis, North China Plain