

# 对内蒙古旗县社会经济发展状况的因子分析

吴 静

(内蒙古财经学院统计学系 内蒙古 呼和浩特 010051)

**摘要:**用因子分析法对内蒙古自治区 1999 年 89 个旗县的 11 项经济发展指标进行分析,以期发现旗县之间社会经济发展状况的差异,从而为内蒙古旗县经济协调发展提供决策依据。

**关键词:**社会经济发展状况;因子分析;差异

**中图分类号:**O212.4 **文献标识码:**A

## 一、引言

社会经济发展状况是一个地区竞争力的体现。笔者利用多元统计方法中的因子分析法对内蒙古各旗县的社会经济发展状况进行因子分析,以期发现旗县之间经济发展的差异,从而为内蒙古旗县经济协调发展提供决策依据。

## 二、因子分析的基本思想

在科学研究中往往需要对反映事物的多个变量进行大量的观测,收集大量数据以便进行分析和寻找规律。多变量大样本无疑会为科学研究提供丰富的信息,但也在一定程度上增加了数据采集的工作量,更重要的是增加了问题分析的复杂性。由于各变量间存在一定的相关关系,因此有可能用较少的综合指标分别综合存在于各变量中的各类信息,这样就可以对综合指标根据专业知识和指标所反映的独特含义给予命名。因子分析正是用少数几个因子去描述多个变量之间关系的多元统计方法。被描述的变量( $x_1, x_2, \dots, x_p$ )是可以观测的随机变量,即

显在变量,而不可观测的潜在变量是描述多个变量的因子( $F_1, F_2, \dots, F_m$ )。因子分析利用这些潜在变量或本质因子去解释可观测的变量,使复杂的现象简单化。

1904 年查尔斯·斯皮尔曼(Charles Spearman)在美国心理学杂志上发表了第一篇有关因子分析的文章,在以后的几十年中,因子分析的理论 and 数学基础得到了不断的发展和完善。50 年代来,随着计算机的普及和各种统计软件的出现,使得因子分析方法在各个领域里得到了成功的应用。

### 1. 因子分析的数学模型

因子分析的数学模型是:

$$x_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \epsilon_1$$

$$x_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \epsilon_2$$

⋮

$$x_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \epsilon_p$$

写成矩阵形式为:  $X = AF + \epsilon$ 。其中  $X$  为原始变量向量,  $A$  为公因子负荷系数矩阵,  $F$  为因子向量,  $\epsilon$  为残差向量。公因子  $F_1, F_2, F_3, \dots, F_m$  之间彼此不相关,称为正交模型。因子分析的目的就是通过模型  $X = AF + \epsilon$  以  $F$  代替  $X$ , 由于  $m < p$ , 从而达到简化变量维数的愿望。

### 2. 因子模型中因子载荷、公因子方差和方差贡

作者:吴静(1958)女,(蒙古族),内蒙古赤峰市人,内蒙古财经学院统计学系副教授,东北财经大学博士研究生。

献的统计意义

因子载荷  $a_{ij}$  指第  $i$  个变量与第  $j$  个公共因子的相关系数即  $x_i$  依赖  $F_j$  的份量, 它反映了第  $i$  个变量在第  $j$  个公共因子上的相对重要性。

公因子方差也称变量共同度, 指观测变量中由公因子决定的比例, 也即因子载荷阵  $A$  中第  $i$  行元素的平方和, 即

$$h_i^2 = \sum_{j=1}^m a_{ij}^2 \quad i=1, \dots, p$$

$h_i^2$  刻划了全部公共因子对变量  $x_i$  的总方差所做的贡献,  $h^2$  越接近于 1, 说明该变量的原始信息被所选取的公共因子说明的越充分。

公因子  $F_j$  的方差贡献指每个公因子对数据的解释能力, 可以用该因子所解释的总方差来衡量, 通常称为该因子的贡献, 它等于因子载荷阵  $A$  中第  $j$  列元素的平方和, 即

$$S_j = \sum_{i=1}^p a_{ij}^2 \quad j=1, \dots, m$$

$S_j$  表示同一公共因子  $F_j$  对诸变量所提供的方差贡献的总和, 它是衡量公共因子相对重要性的指标。

表 1 相关系数矩阵

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$
$x_1$	1.000	0.205	0.816	0.102	-0.200	-0.228	-0.030	-0.015	-0.011	0.782	0.806
$x_2$	0.205	1.000	-0.057	0.651	0.217	0.243	0.802	0.582	0.580	0.195	0.370
$x_3$	0.816	-0.057	1.000	0.001	-0.354	-0.379	-0.062	-0.057	-0.097	0.723	0.638
$x_4$	0.102	0.651	0.001	1.000	0.006	0.036	0.555	0.453	0.346	0.116	0.268
$x_5$	-0.200	0.217	-0.354	0.006	1.000	0.783	0.037	0.123	0.151	-0.307	-0.208
$x_6$	-0.228	0.243	-0.379	0.036	0.783	1.000	0.101	0.126	0.199	-0.241	-0.134
$x_7$	-0.030	0.802	-0.062	0.555	0.037	0.101	1.000	0.464	0.495	0.086	0.162
$x_8$	-0.015	0.582	-0.057	0.453	0.123	0.126	0.464	1.000	0.769	0.179	0.132
$x_9$	-0.011	0.580	-0.097	0.346	0.151	0.199	0.495	0.769	1.000	0.226	0.152
$x_{10}$	0.782	0.195	0.723	0.116	-0.307	-0.241	0.086	0.179	0.226	1.000	0.695
$x_{11}$	0.806	0.370	0.638	0.268	-0.208	-0.134	0.162	0.132	0.152	0.695	1.000

通过计算相关系数矩阵, 得出 KMO 为 0.666, Bartlett's 检验值为 734.795, P 值小于 0.05, 这说明变量之间存在着较强的相关关系, 适合进行因子分析。

3. 计算公因子方差、R 的特征值、贡献率、累计贡献率及方差最大正交旋转矩阵。计算结果见表 2、表 3。

表 2 R 的特征值、贡献率和累计贡献率

因子	公因子方差	特征值	初始因子( $F_1-F_{11}$ )		选取因子( $F_1-F_4$ )	
			方差贡献率(%)	累计方差贡献率(%)	方差贡献率(%)	累计方差贡献率(%)
1	0.927	3.746	34.053	34.053	34.053	34.053
2	0.902	3.322	30.196	64.248	30.196	64.248
3	0.823	1.402	12.743	76.992	12.743	76.992
4	0.735	0.883	8.025	85.016	8.025	85.016
5	0.887	0.480	4.363	89.379		
6	0.884	0.328	2.985	92.364		
7	0.780	0.271	2.465	94.829		
8	0.856	0.211	1.918	96.747		
9	0.897	0.186	1.695	98.442		
10	0.841	0.125	1.133	99.575		
11	0.819	0.047	0.425	100.000		

三、对内蒙古旗县社会经济发展状况的因子分析

资料取自 2000 年内蒙古统计年鉴, 共包括内蒙古 1999 年 89 个旗县的 11 项经济发展指标。分别以  $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}$  代表国内生产总值、人均国内生产总值、财政支出、在岗职工平均货币工资、人均农业总产值、人均猪牛羊肉总产量、人均地方财政收入、每万人拥有医院卫生院床位数、每万人拥有医生数、城乡居民储蓄存款余额、工业总产值。现对内蒙古 89 个旗县的 11 项经济发展指标进行因子分析, 具体分析步骤如下:

1. 对原始数据进行标准化处理, 以消除量纲的影响。

2. 求标准化数据的相关阵 R,  $R = (r_{ij})_{p \times p}$

$$r_{ij} = \frac{\sum_{\alpha=1}^n (x_{\alpha i} - \bar{x}_i)(x_{\alpha j} - \bar{x}_j)}{\sqrt{\sum_{\alpha=1}^n (x_{\alpha i} - \bar{x}_i)^2} \sqrt{\sum_{\alpha=1}^n (x_{\alpha j} - \bar{x}_j)^2}}$$

相关阵计算结果见表 1。

表 3

方差最大正交旋转矩阵

变 量		因 子			
		F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>
国内生产总值(万元)	x <sub>1</sub>	0.957	0.053	-0.046	-0.070
人均国内生产总值(元)	x <sub>2</sub>	0.171	0.833	0.232	0.353
财政支出(万元)	x <sub>3</sub>	0.862	-0.094	-0.258	-0.072
在岗职工平均货币工资(元)	x <sub>4</sub>	0.072	0.845	-0.035	0.124
人均农业总产值(元)	x <sub>5</sub>	-0.175	0.021	0.923	0.062
人均猪牛羊肉总产量(吨)	x <sub>6</sub>	-0.157	0.068	0.920	0.086
人均地方财政收入(元)	x <sub>7</sub>	-0.030	0.835	-0.001	0.285
每万人拥有医院卫生院床位数(床)	x <sub>8</sub>	0.019	0.339	0.046	0.859
每万人拥有医生数(人)	x <sub>9</sub>	0.041	0.281	0.119	0.896
城乡居民储蓄存款余额(万元)	x <sub>10</sub>	0.865	0.006	-0.172	0.251
工业总产值(万元)	x <sub>11</sub>	0.863	0.268	-0.026	0.026

4. 确定和解释综合因子。根据表 2 可知,按累计方差贡献率大于 85% 的原则,变量相关系数矩阵有四大特征根: 3.746、3.322、1.402、0.883,它们一起解释了变量 X 标准化方差的 85.016%。同时,公因子方差在 0.736—0.927 之间,证明全体变量能较好地被综合因子 F<sub>1</sub>、F<sub>2</sub>、F<sub>3</sub>、F<sub>4</sub> 解释。前四个综合因子包含的原始数据信息总量已达 85.016%,说明用 F<sub>1</sub>、F<sub>2</sub>、F<sub>3</sub>、F<sub>4</sub> 代表原来的 11 个变量已有 85.016% 的把握。四个综合因子的线性组合如下:

$$Z_1 = 0.957x_1 + 0.171x_2 + 0.862x_3 + 0.072x_4 - 0.175x_5 - 0.157x_6 - 0.030x_7 + 0.019x_8 - 0.041x_9 + 0.865x_{10} + 0.863x_{11}$$

$$Z_2 = 0.053x_1 + 0.833x_2 - 0.094x_3 + 0.845x_4 + 0.021x_5 + 0.068x_6 + 0.835x_7 + 0.339x_8 + 0.281x_9 + 0.006x_{10} + 0.268x_{11}$$

$$Z_3 = -0.046x_1 + 0.232x_2 - 0.258x_3 - 0.035x_4 + 0.923x_5 + 0.920x_6 - 0.001x_7 + 0.046x_8 + 0.119x_9 - 0.172x_{10} - 0.026x_{11}$$

$$Z_4 = -0.070x_1 + 0.353x_2 - 0.072x_3 + 0.124x_4 + 0.062x_5 + 0.086x_6 + 0.285x_7 + 0.859x_8 + 0.896x_9 + 0.251x_{10} + 0.026x_{11}$$

综合因子的经济意义由各线性组合中权数较大的几个指标的综合意义来确定。综合因子在 F<sub>1</sub> 在 x<sub>1</sub>、x<sub>3</sub>、x<sub>10</sub>、x<sub>11</sub> 上有较大载荷,即它们的系数大于其它变量的系数,所以, F<sub>1</sub> 主要是国内生产总值、财政支出、城乡居民储蓄存款余额、工业总产值这四个指标的综合反映,是反映经济发展综合实力总量的综合因子。F<sub>2</sub> 在 x<sub>2</sub>、x<sub>4</sub>、x<sub>7</sub> 上有较大载荷,即 F<sub>2</sub> 是人均国内生产总值、在岗职工平均货币工资、人均地方财政收入这三个指标的综合反映,是反映社会福利的综合因子。F<sub>3</sub> 在 x<sub>5</sub>、x<sub>6</sub> 上有较大载荷,即 F<sub>3</sub> 是人均农业总产值、人均猪牛羊肉总产量这两个指标的综合反映,是反映农牧业生产状况的综合因子。F<sub>4</sub> 在 x<sub>8</sub>、x<sub>9</sub> 上有较大载荷,即 F<sub>4</sub> 是每万人拥有医院卫生院床位数、每万人拥有医生数这两个指标的综合反映,是反映医疗保健状况的综合因子。我们将 11 个指标按高载荷分成四类,见表 4。

表 4

高载荷指标分类表

类别	高载荷指标	因子命名
1	x <sub>1</sub> : 国内生产总值 x <sub>3</sub> : 财政支出 x <sub>10</sub> : 城乡居民储蓄存款余额 x <sub>11</sub> : 工业总产值	总量因子
2	x <sub>2</sub> : 人均国内生产总值 x <sub>4</sub> : 在岗职工平均货币工资 x <sub>7</sub> : 人均地方财政收入	福利因子
3	x <sub>5</sub> : 人均农业总产值 x <sub>6</sub> : 人均猪牛羊肉总产量	农牧业因子
4	x <sub>8</sub> : 每万人拥有医院卫生院床位数 x <sub>9</sub> : 每万人拥有医生数	医疗保健因子

#### 5. 社会经济发展状况的评价和排序

我们记 Z<sub>1</sub>、Z<sub>2</sub>、Z<sub>3</sub>、Z<sub>4</sub> 分别为各旗县在四个综合因子方面的得分,为各旗县的社会经济发展状况总得分。将标准化的数据代入 Z<sub>1</sub>、Z<sub>2</sub>、Z<sub>3</sub>、Z<sub>4</sub> 式中就可计算出各旗县的社会经济发展状况在四个综合因

子方面的得分(表略)。再分别计算 Z<sub>1</sub>、Z<sub>2</sub>、Z<sub>3</sub>、Z<sub>4</sub> 与各因子方差贡献率的乘积,加总之后即为各旗县的社会经济发展状况总得分。依得分大小排序就可得到 89 个旗县在四个影响各旗县的社会经济发展状况的主要方面和总得分的名次。见表 5。

表 5

社会经济发展状况的评价和排序

地 区	社会经济发展状况总得分		地 区	社会经济发展状况总得分		地 区	社会经济发展状况总得分	
	Z	排序		Z	排序		Z	排序
科尔沁区	1.944	1	额尔古纳市	0.092	31	林西县	-0.319	61
东胜市	1.699	2	土默特右旗	0.075	32	巴林左旗	-0.338	62
临河市	1.005	3	宁城县	0.073	33	科尔沁右翼前旗	-0.341	63
锡林浩特市	0.947	4	阿拉善左旗	0.050	34	四子王旗	-0.367	64
达拉特旗	0.811	5	土默特左旗	0.042	35	察哈尔右翼前旗	-0.370	65
二连浩特市	0.808	6	赤峰市红山区	0.024	36	武川县	-0.373	66
东乌珠穆沁旗	0.746	7	根河市	-0.003	37	正蓝旗	-0.374	67
伊金霍洛旗	0.721	8	五原县	-0.040	38	巴林右旗	-0.383	68
鄂温克自治旗	0.625	9	苏尼特右旗	-0.042	39	突泉县	-0.385	69
包头九元区	0.560	10	扎鲁特旗	-0.057	40	商都县	-0.391	70
准格尔旗	0.557	11	托克托县	-0.059	41	凉城县	-0.417	71
海拉尔市	0.524	12	科尔沁左翼后旗	-0.062	42	科尔沁左翼中旗	-0.421	72
鄂托克旗	0.476	13	莫力达瓦旗	-0.064	43	克什克腾旗	-0.431	73
赤峰市元宝山区	0.438	14	敖汉旗	-0.078	44	多伦县	-0.440	74
苏尼特左旗	0.405	15	额济纳旗	-0.083	45	和林格尔县	-0.444	75
陈巴尔虎旗	0.368	16	赤峰市松山区	-0.110	46	库伦旗	-0.444	76
乌拉特前旗	0.349	17	开鲁县	-0.135	47	科尔沁右翼中旗	-0.447	77
霍林郭勒	0.325	18	达尔罕茂明安联合旗	-0.138	48	清水河县	-0.448	78
乌兰浩特市	0.306	19	固阳县	-0.150	49	阿鲁科尔沁旗	-0.450	79
西乌珠穆沁旗	0.301	20	鄂托克前旗	-0.153	50	镶黄旗	-0.451	80
满洲里市	0.288	21	翁牛特旗	-0.167	51	察哈尔右翼中旗	-0.454	81
乌审旗	0.272	22	磴口县	-0.171	52	太仆寺旗	-0.457	82
阿巴嘎旗	0.256	23	乌拉特中旗	-0.189	53	正镶白旗	-0.463	83
杭锦旗	0.238	24	阿荣旗	-0.231	54	察哈尔右翼后旗	-0.482	84
丰镇市	0.204	25	奈曼旗	-0.238	55	阿尔山市	-0.517	85
新巴尔虎左旗	0.196	26	鄂伦春自治旗	-0.238	56	卓资县	-0.536	86
集宁市	0.137	27	扎兰屯市	-0.241	57	喀喇沁旗	-0.565	87
杭锦后旗	0.129	28	扎赉特旗	-0.255	58	化德县	-0.580	88
牙克石市	0.111	29	阿拉善右旗	-0.263	59	兴和县	-0.635	89
新巴尔虎右旗	0.095	30	乌拉特后旗	-0.303	60			

在社会经济发展状况总得分中,有许多地区得分为负数,但这并不表明社会经济发展水平为负,它表明的是该地区与平均水平的位置关系,社会经济发展状况的平均水平为零点,这是因为我们在整个分析过程中将数据标准化的结果。

下面采用聚类分析中的类平均法,用欧式距离对各地区社会经济发展状况总得分进行聚类分析,将各个地区划分到不同的类别之内。

聚类分析也是一种多元统计方法。在聚类分析中,基本的思想是认为我们所研究的观测量或变量之间存在着程度不同的亲疏关系。我们可根据多个观测量的多个观测指标,具体找出一些能够度量观测量或变量之间相似程度的统计量,以这些统计量作为划分类型的依据,把一些相似程度较大的观测

量或变量聚到一个类别之内,从而表现观测量或变量之间的亲疏关系。

通过聚类分析,得到如下分析结果:

第一,从发展水平看,所选旗县可分为四个集团,排序第一、第二的科尔沁区和东胜市为第一集团,它们的总得分远远高于其它旗县,为 1.944 和 1.699;从排序第三的临河市到排序第十五的苏尼特左旗为第二个集团,总得分在 0.405 至 1.005 之间;从排序第十六的陈巴尔虎旗到排序第三十七的根河市为第三个集团,总得分在 -0.003 至 0.368 之间;从排序第三十八的五原县到排序第八十九的兴和县为第四个集团,总得分在 -0.635 至 -0.040 之间。

第二,从各集团所占结构而言,明显呈上小下大的塔形结构。即第一集团有 2 个 (下转第 31 页)

易体制的变迁与发展都将对世界纺织品贸易格局产生巨大的影响。因此,目前世界纺织品和服装贸易的格局是:发达国家主要从事技术及资本密集型纺织服装生产,产品侧重于个性化、流行化、环保化及高档化;我国及其他发展中国家则主要生产、加工中低档纺织品及服装。注视世界纺织品贸易的发展,有针对性地采取相应的发展策略至关重要。例如近年来美国通过 PSP 方式增加从中美洲和部分拉美国家的进口。PSP 是美国政府实行的一种单向的贸易优惠安排,针对美国进口原材料进行加工组装,然后再将产品返销美国,这些国家这一类产品在返销美国时,美国海关仅对产品的增值部分征收关税。在纺织品贸易上,美国对来自中美洲和部分南美国家的以 PSP 方式生产的纺织品不仅有关税优惠,而且在纺织品配额上给予金额的满足,从而极大鼓励了这一地区 PSP 纺织品贸易的发展,使其在美国市场的份额逐年攀升。1994 年到 1998 年,美国从与其有 PSP 安排的国家进口纺织品以平均每年 25.41% 的速度增长,而同期从其他无 PSP 安排的国家

进口纺织品年平均增长率仅为 7.56%。因此,中国应积极拓宽出口产品的渠道,加大对出口一些我国具有资源禀赋的产品,例如羊绒制品等。

在当前国际纺织品贸易逐步自由化,市场竞争日趋激烈的形式下,应以长远发展眼光,增强紧迫感,不断提高国际竞争力,为纺织品出口的长期发展,特别是纺织品配额限制取消后占据更大的市场份额奠定良好基础。

## 参考文献:

- [1] 李善同等. WTO: 中国与世界[M]. 北京: 中国发展出版社.
- [2] 徐诚. 向低位发展——探析国际纺织品市场价格走势[J]. 国际贸易, 2000, (1).
- [3] 贾荣昌. 鄂尔多斯期百家企业[M]. 呼和浩特: 远方出版社.
- [4] 国务院发展中心. WTO 大众读本[M]. 北京: 中国发展出版社.

(上接第 15 页)稳定,没有民族地区的小康就没有全国的小康,没有民族地区的现代化就不能说实现了全国的现代化。”

综上所述,在民族地区工作中,通过加强民族团结、维护祖国统一和社会稳定代表和体现中国最广大人民的根本利益,与满足民族地区各族人民群众的切

身利益是统一的。二者统一于民族地区不断发展先进生产力、发展先进文化的过程中。民族地区发展先进生产力,发展先进文化,要以维护和实现好这两个利益作为出发点和归宿。在民族地区工作中全面贯彻“三个代表”重要思想,必须坚持“三个代表”之间的统一性,实现三者与实践中的真正统一。

(上接第 27 页)旗县,第二集团有 13 个旗县,第三个集团有 22 个旗县,第四个集团有 52 个旗县。理想的结构应为枣核形,而目前状况显示我区各旗县发展水平非常不均衡,地区协调还有很长的路要走。

第三,在发展最差的第四个集团中,乌兰察布盟所属旗县所占比重最大,这说明乌盟本身的发展与其它盟市的协调发展应当引起有关方面的高度重视。

需要说明,各旗县社会经济发展状况总得分排序的高低与笔者所选取的指标有关,如果全部选取人均指标,不选择总量指标,可能得出另外一种排序结果,但总量指标也是一个地区综合实力的体现,故仍然选

取了几个总量指标作为分析变量。

总而言之,用多元统计中的因子分析方法对社会经济发展状况进行综合评价是比较贴近客观实际的,具有一定的理论意义和实践意义。

## 参考文献:

- [1] 何晓群. 现代统计分析方法与应用[M]. 北京: 中国人民大学出版社, 1998, 316—343.
- [2] 2000 年内蒙古统计年鉴[M]. 北京: 中国统计出版社, 2000, 393—570.