

# 江西统计资料

( 53 )

江西省统计局编印

二〇〇九年六月二十四日

## 多元统计分析在 EXCEL 中的实现

多元统计分析作为研究多维随机变量之间相互依赖关系以及内在规律的一门科学，它在经济研究的各个领域都有着广泛的运用，并取得了许多卓有成效的应用成果。多元统计分析已经越来越成为多类学科进行科学研究、数据分析、数据处理必不可少的重要方法。多元统计分析软件为我们提供了便利，如：TSP、SPSS、SAS 和 R 软件，本文介绍在 EXCEL 中如何实现多元统计分析。

### 一、基本原理

多元统计分析在 EXCEL 中的实现原理基于多元统计分析 BASIC 程序，以 EXCEL 工作簿进行数据维护和管理，利用 EXCEL 的 VBA 程序开发多元统计分析宏，包括：多元回归分析宏、主成分分析宏、

因子分析宏、聚类分析宏、判别分析宏、对应分析宏、典型相关分析宏，在将结果以 EXCEL 数表形式进行分析宏打包，最终形成应用方便的“多元统计分析程序”，只要进行安装，在 EXCEL 的菜单上就出现了“多元统计分析程序”，应用方便直观。

## 二、应用描述

### （一）简单回归分析在 EXCEL 中的实现

#### 1、问题及背景

通过经济分析可知：国内生产总值（GDP）与固定资产投资有密切关系，研究发现两变量之间存在线性关系。根据江西 1990 – 2007 年的 GDP 与全社会固定资产投资数据，研究它们的数量规律性，探讨江西固定资产投资与 GDP 的数量关系。

#### 2、分析过程

第一步：录入原始数据，数据格式如下图 1。



	A	B	C	D
1	年份	GDP (亿元)	全社会固定资产投资 (亿元)	
2	1990	428.62	70.65	
3	1991	479.37	97.08	
4	1992	572.55	125.36	
5	1993	723.04	185.5	
6	1994	948.16	237.45	
7	1995	1169.73	284.18	
8	1996	1409.74	355.85	
9	1997	1605.77	384.3	
10	1998	1719.87	454.77	
11	1999	1853.65	491.48	
12	2000	2003.07	548.2	

图 1 原始数据

第二步：启动多元统计分析系统，选择“回归分析”。单击确定，并选择其中的“多元线性回归”。

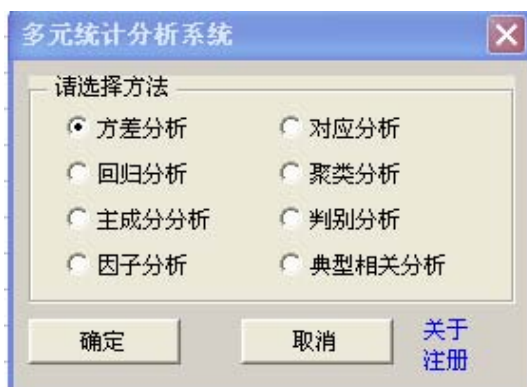


图2 选择项对话框

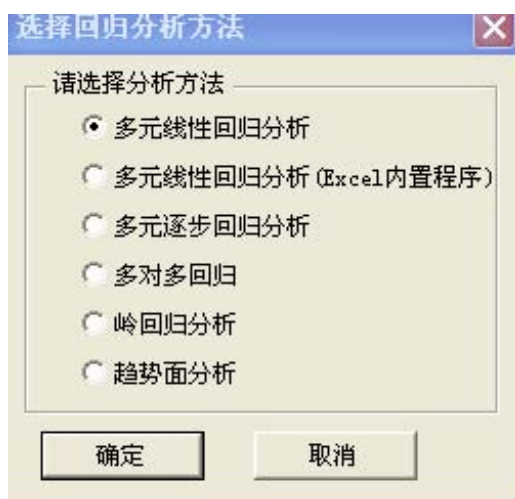


图3 选择项对话框

第三步：填写系统需要信息。样品数为年份数 18，自变量个数为 1，待预报样品为 0，如图 4。

多元线性回归分析

输入

样品数：18

自变量数：1

待预报样品数：

请选择自变量数据区域：heet1!\$C\$2:\$C\$19

请选择因变量数据区域：heet1!\$B\$2:\$B\$19

输出选项

☐ 输出区域

☒ 新工作表组

☐ 新工作簿

确定 取消

图 4 回归分析对话框

第四部：单击确定，得到运行结果，如图 5。

A2					
	A	B	C	D	E
1	*线性回归计算结果*				
3	样品数	N =18			
4	自变量数	P =1			
5	待预报样品数	NN =0			
31	均值表				
32		X1	Y		
33	均值	2112.8544	898.60667		
34					
35	正规方程的系数矩阵，L（）：				
36	37241404.9	23939611			
40	正规方程系数矩阵L（）的逆矩阵，C（）：				
41	2.68518E-08				
44	回归系数：				
45	B0	B1			
46	-459.5835715	0.6428224			
47	回归方程：				
48	Y=-459.5836+0.6428*X1				

图 5 一元回归分析结果

### 3、输出结果和分析结论

回归方程为：

$$y = 459.5836 + 0.6428x_1$$

其中，复相关系数为  $R = 0.9804$ ，说明回归方程拟合优度较高。而回归系数  $t = 19.9257$ ，查  $t$  分布表  $t_{0.025}(16) = 2.1199$ ，小于  $t$  值，因此回归系数显著。

查  $F$  分布表， $F_{0.05}(1, 16) = 4.49$ ，由回归结果知， $F = 313.3765 > 4.49$ ，因此回归方程也显著。

我们可以根据回归方程进行预测。若已知 2009 年江西全社会固定资产投资可以预测 2009 年江西 GDP 数据。程序中如果输入待预报样品信息，可以直接获得待预报样品的预测结果。

## (二) 多元线性回归模型在 EXCEL 中的实现

### 1、问题及背景

在我国，居民消费是在国内生产总值经过初次分配和再分配后形成的，因此考虑人均国内生产总值作为影响人均消费的一个因素；另外，在进行收入分配时，必须考虑到消费者已经实现了的消费，保持消费的连续性，因此，分析当年的消费必须考虑上年已经实现的消费。

为探寻我国的人均消费模型，我们搜集整理了我国 1990 - 2007 年人均消费及人均 GDP 数据。

### 2、分析过程

第一步：录入数据，数据格式如图 6。

Microsoft Excel - 多元表2

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工具(T) 数据(D)

宋体 12 B I U

	A	B	C	D	E
	年份	人均居民消费(元) y	人均GDP(元) x1	前期人均居民消费(元) x2	
1					
2	1990	833	1644	788	
3	1991	932	1893	833	
4	1992	1116	2311	932	
5	1993	1393	2998	1116	
6	1994	1833	4044	1393	
7	1995	2355	5046	1833	
8	1996	2789	5846	2355	
9	1997	3002	6420	2789	
10	1998	3159	6796	3002	
11	1999	3346	7159	3159	
12	2000	3632	7858	3346	

图6 原始数据

第二步，启动多元统计宏，选择回归分析，单击确定，并选择其中的“多元线性回归分析”。

第三步：根据系统要求信息填写，样品数为18，自变量数为2，带预报样品数1个。

多元线性回归分析

输入

样品数：18

自变量数：2

待预报样品数：1

请选择自变量数据区域：heet1!\$C\$2:\$D\$19

请选择因变量数据区域：heet1!\$B\$2:\$B\$19

输出选项

☐ 输出区域

☒ 新工作表组

☐ 新工作簿

确定 取消

图7 多元回归对话框

第四步：单击确定，得到多元线性回归运算结果如图 8。

	A	B	C	D
1	*线性回归计算结果*			
3	样品数 N = 18			
4	自变量数 P = 2			
5	待预报样品数 NN = 1			
1	均值表			
2		X1	X2	Y
3	均值	7892.5	3005	3354.6111
5	正规方程的系数矩阵, L( ) :			
6	419022464.5	136444523	151418572	
7	136444523	45952128	50267716	
1	正规方程系数矩阵 L( ) 的逆矩阵, C( ) :			
2	7.20404E-08	-2.13908E-07		
3	-2.13908E-07	6.56913E-07		
6	回归系数 :			
7	B0	B1	B2	
8	227.6903114	0.155598569	0.63189986	
9	回归方程 :			
0	Y=227.6903+0.1556*X1+0.6319*X2			

图 8 多元回归结果

### 3、输出结果和分析结论

具体输出结果有原始数据表，均值表，正规方程系数矩阵，逆矩阵，回归系数，回归方程，方差分析表，复相关系数，自变量的 t 检验值，各个自变量的偏回归平方和，偏回归系数，观测值，回归值，残差值，以及预测结果。由输出结果，得出以下结论：

$$y = 227.6903 + 0.1556x_1 + 0.6319x_2$$

t 值 (6.3243) (8.5054)

从实际意义来看，各个回归系数均大于零，没有明显错误。由回归方程可知，如果其他变量保持不变，则人均 GDP 每增加 1 元，人均居民消费将增加 0.16 元；同理，如果其他变量保持不变，前期居民消费每增加 1 元，人均居民消费将增加 0.63 元。

回归系数显著性检验：当显著性水平分别为 0.01，0.05 时，t 分布临界值分别为  $t_{0.005} (18 - 2 - 1) = 2.9467$ ， $t_{0.025} (18 - 2 - 1) = 2.1315$ 。由上述方程各自变量的 t 检验值可知，变量  $x_1$ 、 $x_2$  均显著。

回归方程的显著性检验：查 F 分布表可知  $F_{0.05} (2, 15) = 3.68$ ，由多元回归结果，F 统计量的值为 3292.19，显然大于 3.55，因此回归方程显著。复相关系数 R 为 0.9989，从统计意义来看其方程的拟合度很高，总体显著性好。

预测：若已知 2009 年的前期（2008 年）人均居民消费和 2009 年的人均 GDP 值，则可以预测 2009 年的人均居民消费。

$$y_{2008} = 227.6903 + 0.1556x_1 + 0.6319x_2$$

此步也可在程序中得以实现，即待预报样品数为 1，输入相应的 2008 年的前期（2007 年）人均居民消费和 2008 年的人均 GDP 值，即可得到运算结果。

### （三）岭回归分析在 EXCEL 中的实现

#### 1、问题及背景

城镇居民的住房需求受多种因素影响，有宏观上的地区经济总体水平，固定资产投资状况的影响，也有微观上居民收入、储蓄的影响。从搜集数据方便的角度出发，我们以北京市为例，研究北京市城镇居民人均居住面积的影响因素，选择如下指标进行分析：城市人均住房面积（y，平方米），城镇居民人均可支配收入（ $x_1$ ，元），人均城镇储蓄存款余额（ $x_2$ ，元），人均 GDP（ $x_3$ ，元），房地产开发投

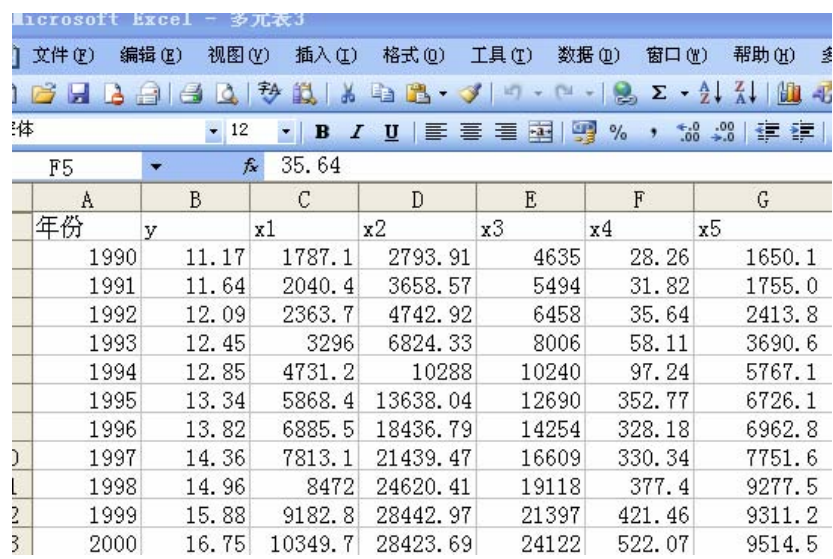


资额（x4，亿元），人均固定资产投资（x5，元）。

由于各影响因素间存在多重共线性问题，如果仍使用最小二乘法，参数估计值的方差会出现偏差，因此，我们使用岭回归来修正。

2、分析过程

第一步：录入原始数据，数据格式如下图 9。



	A	B	C	D	E	F	G
	年份	y	x1	x2	x3	x4	x5
	1990	11.17	1787.1	2793.91	4635	28.26	1650.1
	1991	11.64	2040.4	3658.57	5494	31.82	1755.0
	1992	12.09	2363.7	4742.92	6458	35.64	2413.8
	1993	12.45	3296	6824.33	8006	58.11	3690.6
	1994	12.85	4731.2	10288	10240	97.24	5767.1
	1995	13.34	5868.4	13638.04	12690	352.77	6726.1
	1996	13.82	6885.5	18436.79	14254	328.18	6962.8
	1997	14.36	7813.1	21439.47	16609	330.34	7751.6
	1998	14.96	8472	24620.41	19118	377.4	9277.5
	1999	15.88	9182.8	28442.97	21397	421.46	9311.2
	2000	16.75	10349.7	28423.69	24122	522.07	9514.5

图 9 原始数据

第二步：启动多元统计分析宏，选择“回归分析”，单击确定，选定其中的“岭回归分析”。

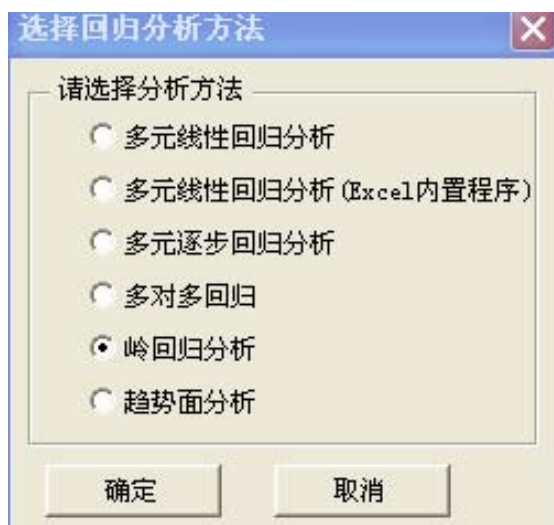


图 10 选择对话框

第三步：同前所述，根据系统要求输入样品数位 18，自变量为 5，带预报样品 1 个。

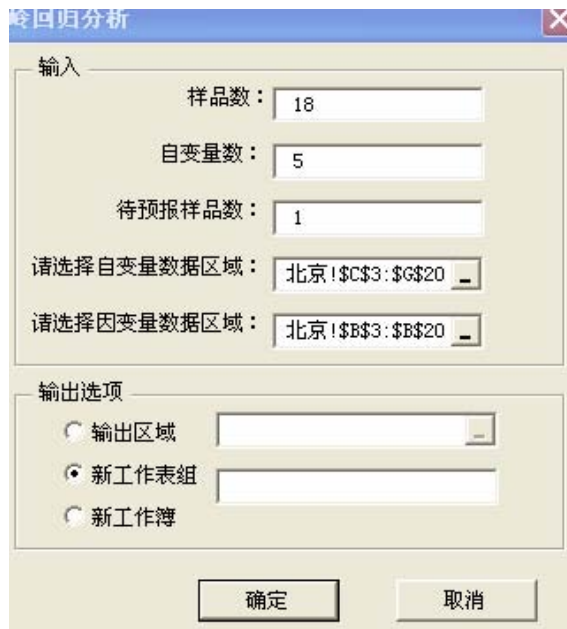


图 11 岭回归对话框

第四步：单击确定，得到岭回归分析结果。

Microsoft Excel - 修改后的多元表3								
文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工具(T) 数据(D) 窗口(W) 帮助(H) 多元分析								
宋体 12 B I U % 100%								
C12	fx							
	A	B	C	D	E	F	G	H
1	<< 岭 回 归 分 析 结 果 >>							
2	样 品 数 N = 18							
3	自 变 量 数 M = 5							
4	待预报的样品数 NN = 1							
5								
6	最小二乘回归计算:							
7								
8	回归方程的系数向量							
9	B1	0.0017976						
10	B2	-0.00034						
11	B3	0.0011326						
12	B4	-0.028616						
13	B5	-9.58E-05						
14								
15								
16	回归拟合结果							
17	顺序号	观测值	回归值	剩余值				
18	1	11.17	10.167538	-1.002462				
19	2	11.64	11.089427	-0.550573				
20	3	12.09	12.649621	0.5596211				
21	4	12.45	12.793845	0.3438448				
22	5	12.85	12.920621	0.0706213				
23	6	13.34	13.430768	0.0907684				
24	7	13.82	13.897403	0.0774029				
25	8	14.36	14.078522	-0.281478				
26	9	14.96	15.012718	0.0527184				
27	10	15.88	15.749687	-0.130313				
28	11	16.75	16.353938	-0.396062				

图 12 岭回归分析结果

### 3、输出结果和分析结论

(1) 分别以  $k = \frac{k_t}{4}$ ,  $k = \frac{k_t}{2}$ ,  $k = \frac{3k_t}{4}$ ,  $k_t$ , 建立岭回归估计方程。

(2) 根据各岭回归的剩余值计算误差平方和可知, 当  $k = \frac{k_t}{4}$  时, 误差平方和最小, 因此选择  $k = \frac{k_t}{4}$  时的预报结果, 相应的岭回归方程为:

$$y = 0.0017979x_1 - 0.00034x_2 + 0.0011326x_3 + 0.028616x_4 - 0.0000958x_5$$

由上述方程, 我们可以进行因素分析, 其中, 对居民人均居住面积影响较大的因素主要有: 人均可支配收入 ( $x_1$ )、人均 GDP ( $x_3$ ) 和房地产开发投资 ( $x_4$ ), 与人均储蓄存款 ( $x_2$ ) 和人均固定资产投资 ( $x_5$ ) 影响不大。

#### (四) 主成分分析在 EXCEL 中的实现

##### 1、问题及背景

随着生产力水平的不断提高, 我国居民生活水平不断提高, 生活质量也在不断改善。但是受个地区生产力发展水平不平衡的影响, 我国各地区居民生活质量也表现为不平衡, 为了分析各地区居民生活质量的状况并进行综合评价, 我们选取如下指标体系对 2007 年全国 31 个省市、自治区的居民生活质量状况进行评价分析, 即职工人均工资 ( $x_1$ , 元), 人均住宅建筑面积 ( $x_2$ , 平方米), 城市用水普及率 ( $x_3$ , %), 城市燃气普及率 ( $x_4$ , %), 人均城市道路面积 ( $x_5$ , 平方米), 人均公共绿地面积 ( $x_6$ , 平方米), 批发零售贸易额 ( $x_7$ , 亿元), 旅游外汇收入 ( $x_8$ , 百万美元)。

##### 2、分析过程

第一步: 录入数据, 数据格式如图 13。

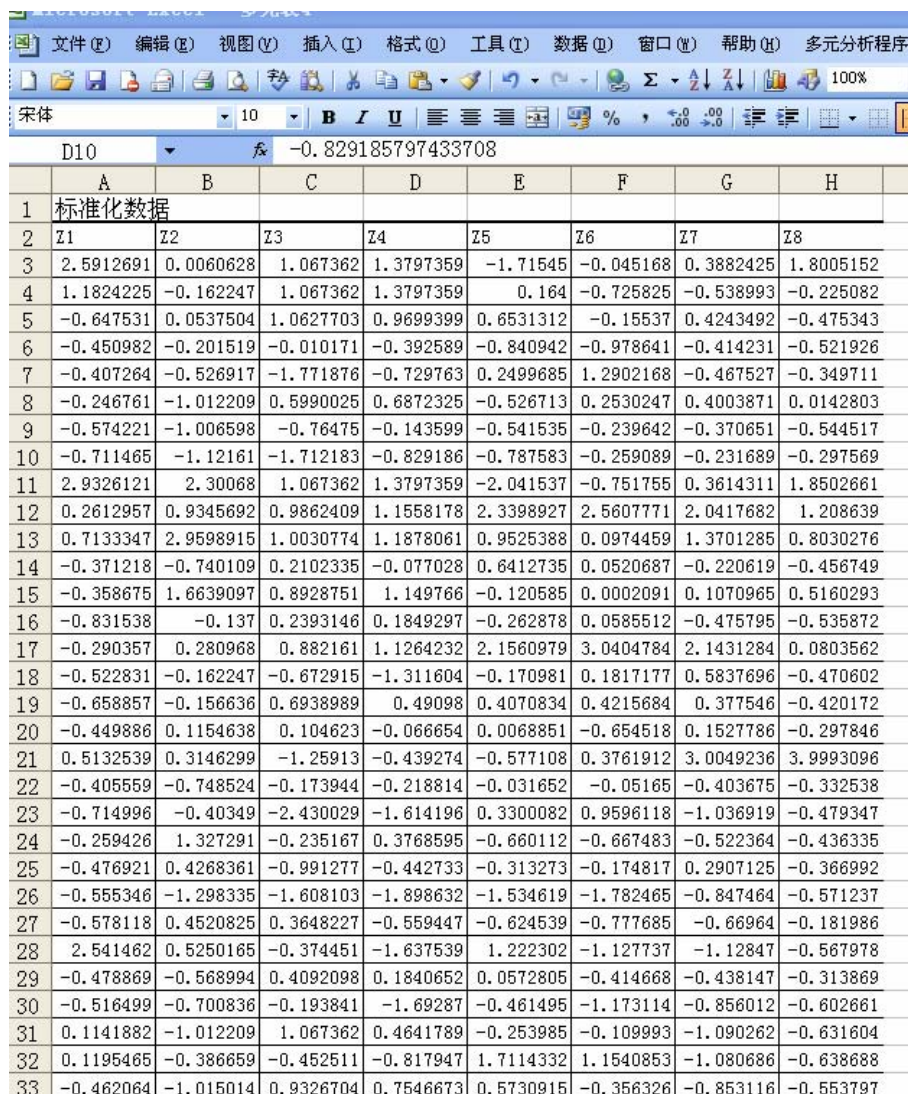
	A	B	C	D	E	F	G	H	I
		职工平均 工资 (元)	人均住宅 建筑面积 (平方 米)	城市用水 普及率 (%)	城市燃气 普及率 (%)	人均城市道 路面积 (平 方米)	人均公共绿 地面积 (平 方米)	批发零售 贸易商品 销售额 (亿元)	旅游外汇 收入 (百 万美元)
1									
2		x1	x2	x3	x4	x5	x6	x7	x8
3	北 京	46507	26.65	100	100	5.6	8.57	3335.7	4579.62
4	天 津	34938	26.05	100	100	11.94	7.52	1350.6	778.7132
5	河 北	19911	26.82	99.97	95.26	13.59	8.4	3413	309.1147
6	山 西	21525	25.91	92.96	79.5	8.55	7.13	1617.7	221.7051
7	内 蒙 古	21884	24.75	81.45	75.6	12.23	10.63	1503.6	544.8549
8	辽 宁	23202	23.02	96.94	91.99	9.61	9.03	3361.7	1227.862
9	吉 林	20513	23.04	88.03	82.38	9.56	8.27	1711	179.3142
10	黑 龙 江	19386	22.63	81.84	74.45	8.73	8.24	2008.5	642.6955
11	上 海	49310	34.83	100	100	4.5	7.48	3278.3	4672.974
12	江 苏	27374	29.96	99.47	97.41	19.28	12.59	6875.7	3469.001
13	浙 江	31086	37.18	99.58	97.78	14.6	8.79	5437.8	2707.897
14	安 徽	22180	23.99	94.4	83.15	13.55	8.72	2032.2	344.0049
15	福 建	22283	32.56	98.86	97.34	10.98	8.64	2733.8	2169.363
16	江 西	18400	26.14	94.59	86.18	10.5	8.73	1485.9	195.5364
17	山 东	22844	27.63	98.79	97.07	18.66	13.33	7092.7	1351.849
18	河 南	20935	26.05	88.63	68.87	10.81	8.92	3754.3	318.01
19	湖 北	19818	26.07	97.56	89.72	12.76	9.29	3312.8	412.64
20	湖 南	21534	27.04	93.71	83.27	11.41	7.63	2831.6	642.176
21	广 东	29443	27.75	84.8	78.96	9.44	9.22	8937.7	8705.52
22	广 西	21898	23.96	91.89	81.51	11.28	8.56	1640.3	577.0794
23	海 南	19357	25.19	77.15	65.37	12.5	10.12	284.6	301.6004
24	重 庆	23098	31.36	91.49	88.4	9.16	7.61	1386.2	382.31
25	四 川	21312	28.15	86.55	78.92	10.33	8.37	3126.9	512.4284

图 13 原始数据

第二步：启动多元统计分析程序，选择“主成分分析”。

第三步：确定并输入系统要求信息，样品数位  $n = 31$ ， $p = 8$ 。特征值贡献率一般选择为 85%，如下图 14。

### 3、输出结果



	A	B	C	D	E	F	G	H
1	标准化数据							
2	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8
3	2.5912691	0.0060628	1.067362	1.3797359	-1.71545	-0.045168	0.3882425	1.8005152
4	1.1824225	-0.162247	1.067362	1.3797359	0.164	-0.725825	-0.538993	-0.225082
5	-0.647531	0.0537504	1.0627703	0.9699399	0.6531312	-0.15537	0.4243492	-0.475343
6	-0.450982	-0.201519	-0.010171	-0.392589	-0.840942	-0.978641	-0.414231	-0.521926
7	-0.407264	-0.526917	-1.771876	-0.729763	0.2499685	1.2902168	-0.467527	-0.349711
8	-0.246761	-1.012209	0.5990025	0.6872325	-0.526713	0.2530247	0.4003871	0.0142803
9	-0.574221	-1.006598	-0.76475	-0.143599	-0.541535	-0.239642	-0.370651	-0.544517
10	-0.711465	-1.12161	-1.712183	-0.829186	-0.787583	-0.259089	-0.231689	-0.297569
11	2.9326121	2.30068	1.067362	1.3797359	-2.041537	-0.751755	0.3614311	1.8502661
12	0.2612957	0.9345692	0.9862409	1.1558178	2.3398927	2.5607771	2.0417682	1.208639
13	0.7133347	2.9598915	1.0030774	1.1878061	0.9525388	0.0974459	1.3701285	0.8030276
14	-0.371218	-0.740109	0.2102335	-0.077028	0.6412735	0.0520687	-0.220619	-0.456749
15	-0.358675	1.6639097	0.8928751	1.149766	-0.120585	0.0002091	0.1070965	0.5160293
16	-0.831538	-0.137	0.2393146	0.1849297	-0.262878	0.0585512	-0.475795	-0.535872
17	-0.290357	0.280968	0.882161	1.1264232	2.1560979	3.0404784	2.1431284	0.0803562
18	-0.522831	-0.162247	-0.672915	-1.311604	-0.170981	0.1817177	0.5837696	-0.470602
19	-0.658857	-0.156636	0.6938989	0.49098	0.4070834	0.4215684	0.377546	-0.420172
20	-0.449886	0.1154638	0.104623	-0.066654	0.0068851	-0.654518	0.1527786	-0.297846
21	0.5132539	0.3146299	-1.25913	-0.439274	-0.577108	0.3761912	3.0049236	3.9993096
22	-0.405559	-0.748524	-0.173944	-0.218814	-0.031652	-0.05165	-0.403675	-0.332538
23	-0.714996	-0.40349	-2.430029	-1.614196	0.3300082	0.9596118	-1.036919	-0.479347
24	-0.259426	1.327291	-0.235167	0.3768595	-0.660112	-0.667483	-0.522364	-0.436335
25	-0.476921	0.4268361	-0.991277	-0.442733	-0.313273	-0.174817	0.2907125	-0.366992
26	-0.555346	-1.298335	-1.608103	-1.898632	-1.534619	-1.782465	-0.847464	-0.571237
27	-0.578118	0.4520825	0.3648227	-0.559447	-0.624539	-0.777685	-0.66964	-0.181986
28	2.541462	0.5250165	-0.374451	-1.637539	1.222302	-1.127737	-1.12847	-0.567978
29	-0.478869	-0.568994	0.4092098	0.1840652	0.0572805	-0.414668	-0.438147	-0.313869
30	-0.516499	-0.700836	-0.193841	-1.69287	-0.461495	-1.173114	-0.856012	-0.602661
31	0.1141882	-1.012209	1.067362	0.4641789	-0.253985	-0.109993	-1.090262	-0.631604
32	0.1195465	-0.386659	-0.452511	-0.817947	1.7114332	1.1540853	-1.080686	-0.638688
33	-0.462064	-1.015014	0.9326704	0.7546673	0.5730915	-0.356326	-0.853116	-0.553797

图 14 标准化数据

表 1 相关系数矩阵

1	0.45368462	0.3213314	0.30061604	-0.1544334	-0.124737	0.12471718	0.5116906
0.4536846	1	0.3633485	0.4447214	8.64E-02	9.19E-02	0.42920958	0.4572615
0.3213314	0.36334848	1	0.81940855	0.17265732	5.90E-02	0.20753184	0.127153
0.300616	0.4447214	0.8194085	1	0.10766705	0.2856137	0.42134374	0.3479959
-0.154433	8.64E-02	0.1726573	0.10766705	1	0.676262	0.22518991	-0.18758
-0.124737	9.19E-02	5.90E-02	0.28561374	0.67626199	1	0.54043642	0.1971603
0.1247172	0.42920958	0.2075318	0.42134374	0.22518991	0.5404364	1	0.7379227
0.5116906	0.45726152	0.127153	0.34799594	-0.1875805	0.1971603	0.73792265	1

表 2 特征向量（列向量）

0.2930428	-0.4209675	1.64E-02	-0.5540994	-0.5687322	0.0359355	-0.2927063	0.1385388
0.3978385	-0.1636193	1.76E-02	-0.4466635	0.74528607	0.2153528	3.92E-03	-0.115284
0.3597059	-7.62E-02	-0.61395	0.18117673	-6.77E-02	-0.360343	-8.53E-02	-0.559768
0.4451167	-2.60E-02	-0.399894	0.35769053	-4.38E-02	0.3868322	0.18672127	0.571765
0.1342118	0.60736792	-0.17416	-0.4728305	-5.75E-02	-0.417346	0.3327679	0.2664061
0.2532684	0.58780346	0.1478781	-3.05E-02	-0.252408	0.5658678	-0.1550175	-0.399498
0.4294408	0.18367487	0.3992999	0.29032502	0.13034757	-0.400454	-0.5534572	0.2333378
0.4021891	-0.2034268	0.50093	0.15463396	-0.1749013	-0.134531	0.65657152	-0.205533

表 3 累计贡献率表

NO	特征值 H ( )	百分率 LH	累计率
1	3.17348689	0.3966859	0.39668586
2	1.84289522	0.2303619	0.62704776
3	1.31075387	0.1638442	0.790892
4	0.71978434	0.089973	0.88086504
5	0.52464857	0.0655811	0.94644611
6	0.25665007	0.0320813	0.97852737
7	0.09336483	0.0116706	0.99019797
8	0.0784162	0.009802	1

由特征值得累计百分率确定方差贡献的百分数 85%，由此选择主成分 M=4。



表 4 主成分得分值

序号	主成分 1	主成分 2	主成分 3	主成分 4
1	2.40904642	-2.572493	0.18454726	0.45200189
2	0.79622331	-0.9687	-1.6544103	-0.1425199
3	0.68501346	0.6375637	-1.2554179	0.61988636
4	-1.1392592	-0.822117	-0.2728166	0.42411504
5	-1.2722629	1.3071919	1.14910045	-0.4683632
6	0.21743427	0.1054209	-0.3682968	1.30297285
7	-1.4192839	0.0413774	0.13789039	0.64941594
8	-2.0301742	0.022436	1.20861838	0.55541517
9	3.20773603	-3.720066	0.19713001	-0.5864278
10	3.64313416	2.6874171	0.34494632	-0.3748649
...				

4、分析结论

(1) 相关系数矩阵显示，8 个原始指标间存在一定的相关性，说明这些指标间存在信息重叠。所以可以利用主成分分析把这些变量重新综合为互不相关的新的综合变量。

(2) 由特征值表和累计贡献率，我们选择了 4 个重要主成分，其中由特征向量表的各个主成分表达式为：

$$F_1 = 0.2930x_1 + 0.3978x_2 + 0.3597x_3 + 0.4451x_4 + 0.1342x_5 + 0.2533x_6 + 0.4294x_7 + 0.4022x_8$$

$$F_2 = -0.4210x_1 - 0.1636x_2 - 0.0762x_3 + 0.026x_4 + 0.607x_5 + 0.5878x_6 + 0.1837x_7 - 0.2034x_8$$

$$F_3 = 0.0164x_1 + 0.0176x_2 - 0.6140x_3 + 0.03999x_4 - 0.1742x_5 + 0.1479x_6 + 0.3993x_7 + 0.5009x_8$$

$$F_4 = -0.5541x_1 - 0.4467x_2 + 0.1812x_3 + 0.3577x_4 - 0.4728x_5 - 0.0305x_6 + 0.2903x_7 + 0.1546x_8$$

(3) 各地区居民生活质量水平的综合排名和地区分布  
首先，根据四个主成分的贡献确定各自的权数。

$$W1 = 3.17 / (3.17 + 1.84 + 1.31 + 0.72) = 0.45$$



$W2 = 1.84 / (3.17 + 1.84 + 1.31 + 0.72) = 0.26$

$W3 = 1.31 / (3.17 + 1.84 + 1.31 + 0.72) = 0.19$

$W3 = 0.72 / (3.17 + 1.84 + 1.31 + 0.72) = 0.10$

其次，计算各地区居民生活质量综合得分和排名。综合得分计算公式为：

$F = 0.45F1 + 0.26F2 + 0.19F3 + 0.10F4$

最后得主成分综合得分如表 5。

表 5 主成分分值及综合评价表

地 区	主成分 1	主成分 2	主成分 3	主成分 4	综合值	排名
江 苏	3.643	2.687	0.345	-0.375	2.366	1
山 东	2.857	3.454	-0.022	0.121	2.192	2
广 东	2.544	-0.551	4.322	0.942	1.917	3
浙 江	3.340	-0.168	-0.229	-1.042	1.312	4
福 建	1.627	-0.378	-0.662	0.196	0.528	5
北 京	2.409	-2.572	0.185	0.452	0.495	6
上 海	3.208	-3.720	0.197	-0.586	0.455	7
湖 北	0.367	0.887	-0.704	0.576	0.320	8
河 北	0.685	0.638	-1.255	0.620	0.297	9
辽 宁	0.217	0.105	-0.368	1.303	0.186	10
内 蒙 古	-1.272	1.307	1.149	-0.468	-0.061	11
河 南	-0.959	0.538	0.980	-0.057	-0.111	12
四 川	-0.633	0.053	0.746	-0.083	-0.137	13
宁 夏	-0.845	1.718	-0.279	-1.525	-0.139	14
安 徽	-0.541	0.736	-0.538	0.107	-0.144	15
湖 南	-0.297	-0.128	-0.229	0.208	-0.190	16
天 津	0.796	-0.969	-1.654	-0.143	-0.222	17
江 西	-0.570	0.246	-0.641	0.533	-0.261	18
陕 西	-0.549	0.033	-0.746	0.469	-0.333	19
广 西	-0.901	0.256	-0.155	0.297	-0.339	20
新 疆	-0.470	0.364	-1.670	0.555	-0.379	21
重 庆	-0.122	-0.900	-0.398	-0.244	-0.389	22
海 南	-2.313	1.266	1.549	-1.002	-0.518	23
吉 林	-1.419	0.041	0.138	0.649	-0.537	24
青 海	-0.563	-0.267	-1.581	0.458	-0.577	25
云 南	-0.749	-0.766	-0.367	0.081	-0.598	26
黑 龙 江	-2.030	0.022	1.209	0.555	-0.623	27
山 西	-1.139	-0.822	-0.273	0.424	-0.736	28
西 藏	-0.745	-1.097	-0.179	-3.255	-0.980	29
甘 肃	-2.222	-0.614	0.038	-0.129	-1.165	30
贵 州	-3.354	-1.401	1.094	0.363	-1.629	31

由上述主成分综合评价得分，2007 年我国 31 个省市自治区的居民生活质量状况发展不平衡。总体上，东部沿海省市居民生活质量水平较高，中部和西部地区相对较低。特别是前十名的省市中有 9 名来自东部沿海地区。具体分布情况见表 6。

表 6 居民生活质量水平分布表

	东部地区	中部地区	西部地区
第 1 到 10 名	江苏、山东、广东、浙江、福建、北京、上海、河北、辽宁	湖北	
第 11 到 20 名	天津	河南、安徽、湖南、江西	内蒙古、四川、宁夏、陕西、广西
第 21 到 31 名	海南	吉林、黑龙江、山西	新疆、重庆、青海、云南、西藏、甘肃、贵州

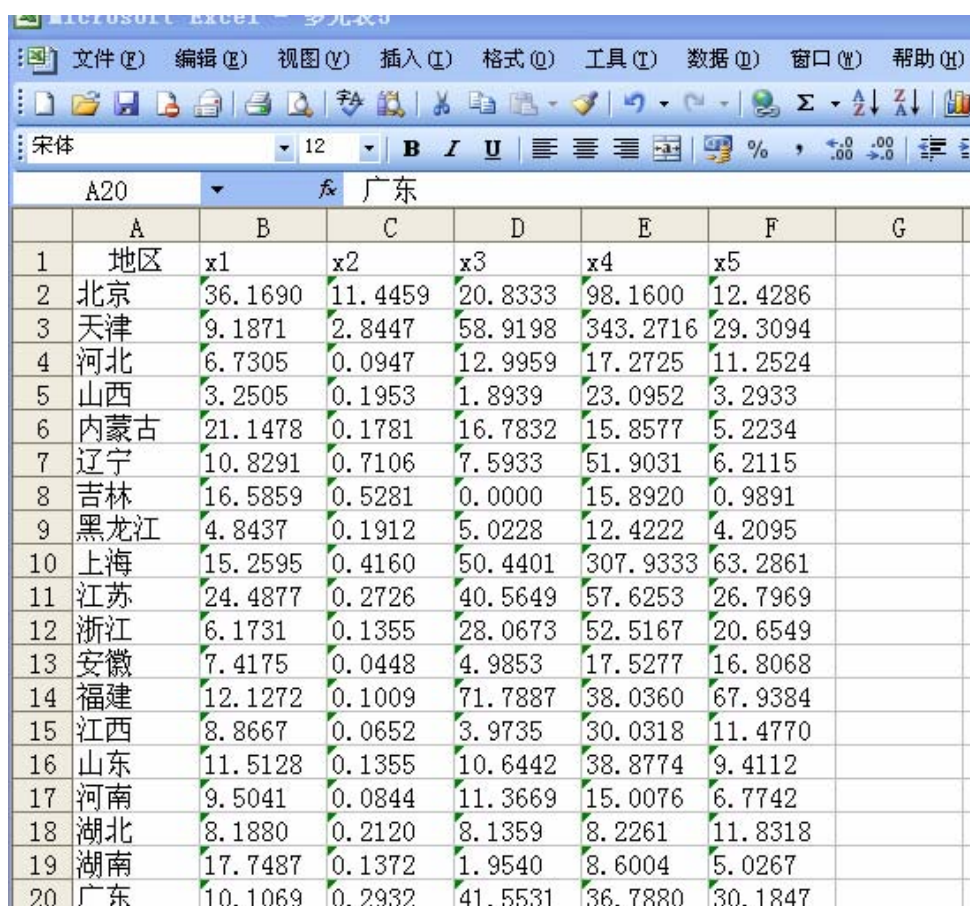
(五) 因子分析在 EXCEL 中的实现

1、问题及背景

实施科技大开放战略是实现地区经济发展的迫切要求。为了分析比较各地区科技大开放的发展状况，我们选择因子分析进行综合评价。具体选择 2006 年全国 30 个省市、自治区（西藏除外）如下指标数据来分析：高新企业出口额占全部出口额比重（ $x_1, \%$ ），每万人高新企业个数（ $X_2$ ，个），三资企业拥有发明专利占大中型工业企业发明专利的比重（ $x_3, \%$ ），人均拥有大中型工业企业技术引进经费和购买国内技术经费（ $x_4$ ，元），三资企业科学家和工程师人数占整个大中型工业企业科学家和工程师人数比重（ $x_5, \%$ ）

2、分析过程

第一步：录入数据，数据格式如图 15。



	A	B	C	D	E	F	G
1	地区	x1	x2	x3	x4	x5	
2	北京	36.1690	11.4459	20.8333	98.1600	12.4286	
3	天津	9.1871	2.8447	58.9198	343.2716	29.3094	
4	河北	6.7305	0.0947	12.9959	17.2725	11.2524	
5	山西	3.2505	0.1953	1.8939	23.0952	3.2933	
6	内蒙古	21.1478	0.1781	16.7832	15.8577	5.2234	
7	辽宁	10.8291	0.7106	7.5933	51.9031	6.2115	
8	吉林	16.5859	0.5281	0.0000	15.8920	0.9891	
9	黑龙江	4.8437	0.1912	5.0228	12.4222	4.2095	
10	上海	15.2595	0.4160	50.4401	307.9333	63.2861	
11	江苏	24.4877	0.2726	40.5649	57.6253	26.7969	
12	浙江	6.1731	0.1355	28.0673	52.5167	20.6549	
13	安徽	7.4175	0.0448	4.9853	17.5277	16.8068	
14	福建	12.1272	0.1009	71.7887	38.0360	67.9384	
15	江西	8.8667	0.0652	3.9735	30.0318	11.4770	
16	山东	11.5128	0.1355	10.6442	38.8774	9.4112	
17	河南	9.5041	0.0844	11.3669	15.0076	6.7742	
18	湖北	8.1880	0.2120	8.1359	8.2261	11.8318	
19	湖南	17.7487	0.1372	1.9540	8.6004	5.0267	
20	广东	10.1069	0.2932	41.5531	36.7880	30.1847	

图 15 原始数据

第二步：启动多元统计分析程序，选择因子分析

第三步：同主成分分析。方差贡献率为 0.85。

第四步：单击确定，得到因子分析运行结果。

图 16 选择项对话框

### 3、输出结果和分析结论

(1) 由累计贡献率可得，取前三个因子其累计贡献率 89%，大于 85%，说明前三个因子已经代表原始数据的绝大多数信息。

表 7 特征根和累计贡献率

因子	特征根	方差贡献率%	累计贡献率%
1	2.3690	47.3790	47.3790
2	1.4214	28.4283	75.8074
3	0.6376	12.7522	88.5596
4	0.3764	7.5280	96.0875
5	0.1956	3.9125	100.0000

(2) 由因子载荷矩阵得因子模型为：

表 8

指                    标	因子 1	因子 2	因子 3
高新企业出口额占全部出口额比重 (x1,%)	0.0057	0.2102	0.1602
每万人高新企业个数 (x2, 个)	0.0137	0.8989	0.3106
三资企业拥有发明专利占大中型工业企业发明专利的比重(x3,%)	0.9302	0.0983	0.0696
人均拥有大中型工业企业技术引进经费和购买国内技术经费(x4,元)	0.7033	0.5345	0.6656
三资企业科学家和工程师人数占整个大中型工业企业科学家和工程师人数比重 (x5,%)	0.9294	-0.0765	-0.0022

(3) 由因子载荷矩阵可以看出, 因子 1 与 x3、x5 的相关系数大, 定义为引进人才和专利因子; 因子 2 与 x2 相关系数大, 定义为高新企业发展因子; 因子 3 与 x4 相关系数大, 定义为引进技术因子。

(4) 计算综合因子得分, 并以各因子和综合因子得分进行排名。

表 9                      2006 年旋转后各因子的方差贡献表

	因子 1	因子 2	因子 3
方差贡献	2.2238	1.1535	1.0506

表 10                      2006 年各省市、区科技大开放差异因子分析表

地区	引进人才和专利		高新企业发展因子		引进技术因子		因子总得分	
	得分	排序	得分	排序	得分	排序	得分	排序
北 京	-0.433879	19	4.286188	1	1.731464	2	5.798335	3
天 津	2.050796	3	2.343233	2	-1.326309	30	5.870058	2
河 北	-0.210366	10	-0.219194	16	-0.552758	10	-1.301381	18
山 西	-0.695711	26	0.113982	6	-1.035115	26	-2.503136	28
内蒙古	-0.227096	11	-0.361701	23	0.775518	6	-0.107478	11
辽 宁	-0.402680	18	0.294138	4	-0.401755	18	-0.978274	16
吉 林	-0.777560	27	0.026218	7	0.202707	27	-1.485932	24
黑龙江	-0.619190	25	-0.016213	10	-0.813422	25	-2.250237	25
上 海	2.913491	1	0.537215	3	-0.382770	1	6.696561	1
江 苏	1.043561	4	-0.610864	28	1.221295	3	2.899131	5
浙 江	0.487165	6	-0.214025	15	-0.587553	6	0.219198	9
安 徽	-0.230659	12	-0.291708	21	-0.502254	12	-1.377092	20
福 建	2.800588	2	-1.457638	30	0.664805	7	5.245007	4

续表

地区	引进人才和专利		高新企业发展因子		引进技术因子		因子总得分	
	得分	排序	得分	排序	得分	排序	得分	排序
江 西	-0.362364	16	-0.135109	13	-0.466809	16	-1.452103	23
山 东	-0.230659	13	-0.101669	12	-0.230009	13	-0.880695	14
河 南	-0.363691	17	-0.201313	14	-0.328134	17	-1.385730	21
湖 北	-0.330825	15	-0.234371	17	-0.408253	15	-1.434946	22
湖 南	-0.606189	22	-0.274389	19	0.390462	9	-1.254331	17
广 东	1.027752	5	-0.554314	27	0.003024	13	1.649290	7
广 西	-0.009124	9	-0.527870	25	-0.001664	14	-0.630935	13
海 南	0.114052	7	-0.666920	29	0.028821	12	-0.485385	12
重 庆	-0.293106	14	-0.284328	20	1.125393	4	0.202555	10
四 川	-0.609089	24	-0.267439	18	0.336849	10	-1.309090	19
贵 州	0.001581	8	-0.532432	26	0.888367	5	0.322673	8
云 南	-0.470020	20	-0.312008	22	0.484373	8	-0.896249	15
陕 西	-0.606314	23	-0.399257	24	3.325352	1	1.684752	6
甘 肃	-0.845022	30	0.123928	5	-1.003222	26	-2.790195	29
青 海	-0.783917	28	-0.055848	11	-1.243799	28	-3.114431	30
宁 夏	-0.514449	21	-0.005661	9	-1.274688	29	-2.489750	27
新 疆	-0.813105	29	-0.000631	8	-0.619915	24	-2.460194	26

结论：2006 年科技大开放排在前五位的省市为：上海、天津、北京、福建和江苏。

课 题 主 持：孙菊生

课题组成员：刘晓红 龚 晨