

《应用多元统计分析》（第五版）

JMP13.1 的应用

王学民

说明：本文侧重于 JMP13.1 的菜单操作，其输出结果有许多与 SAS9.3 的输出结果相同或类似，对于这些部分本文未作输出说明，可参见《应用多元统计分析》（原《应用多元分析》，第五版，王学民编著）各章附录 1（SAS 的应用）或书中的有关例题。

读者可从

<https://anyshare.sufe.edu.cn/#/link/B9F2F217DF9A179950462AF6B590145F?path=>下载《应用多元统计分析》（第五版）配书资料，下载的资料中有一个“《应用多元统计分析》（第五版）Excel 数据”文件夹，本文均从该文件夹中打开数据表。

在主窗口（见图 1）中，选择文件⇒打开...，即出现如图 2 所示的“打开数据文件”窗口，选择窗口左边列表框中 Excel 数据所在的文件夹，双击窗口中的数据表名即出现“Excel 导入向导”窗口（见图 3），在该窗口中点击导入，即可打开数据表。

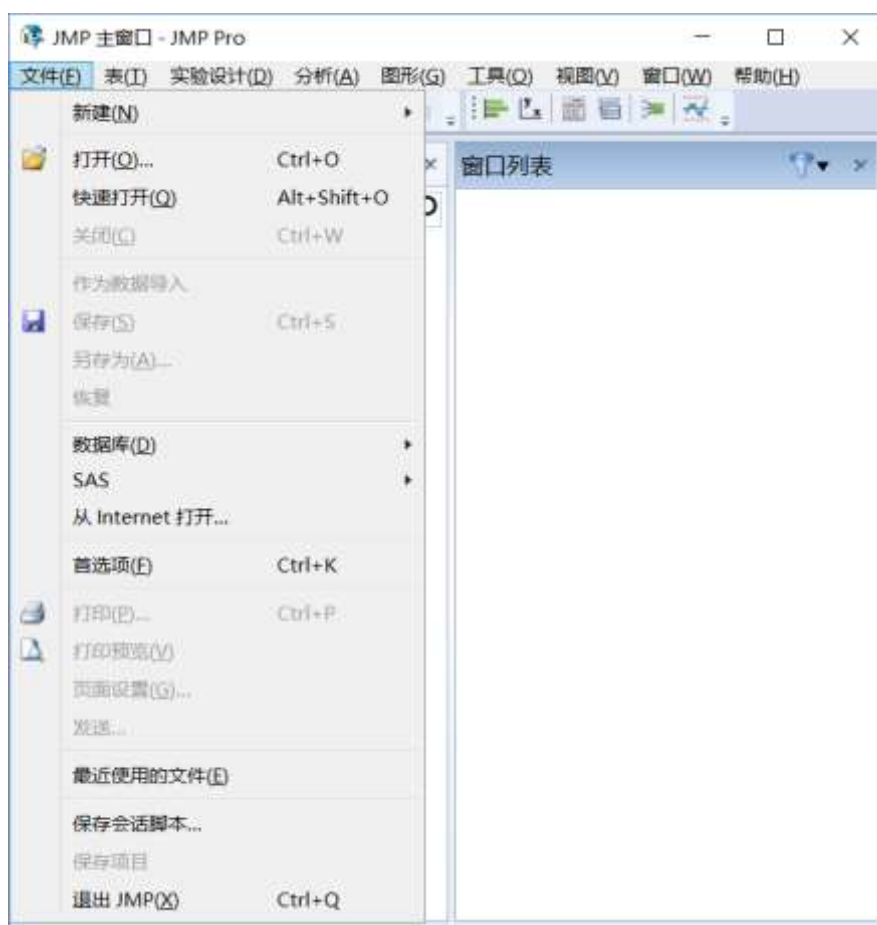


图 1

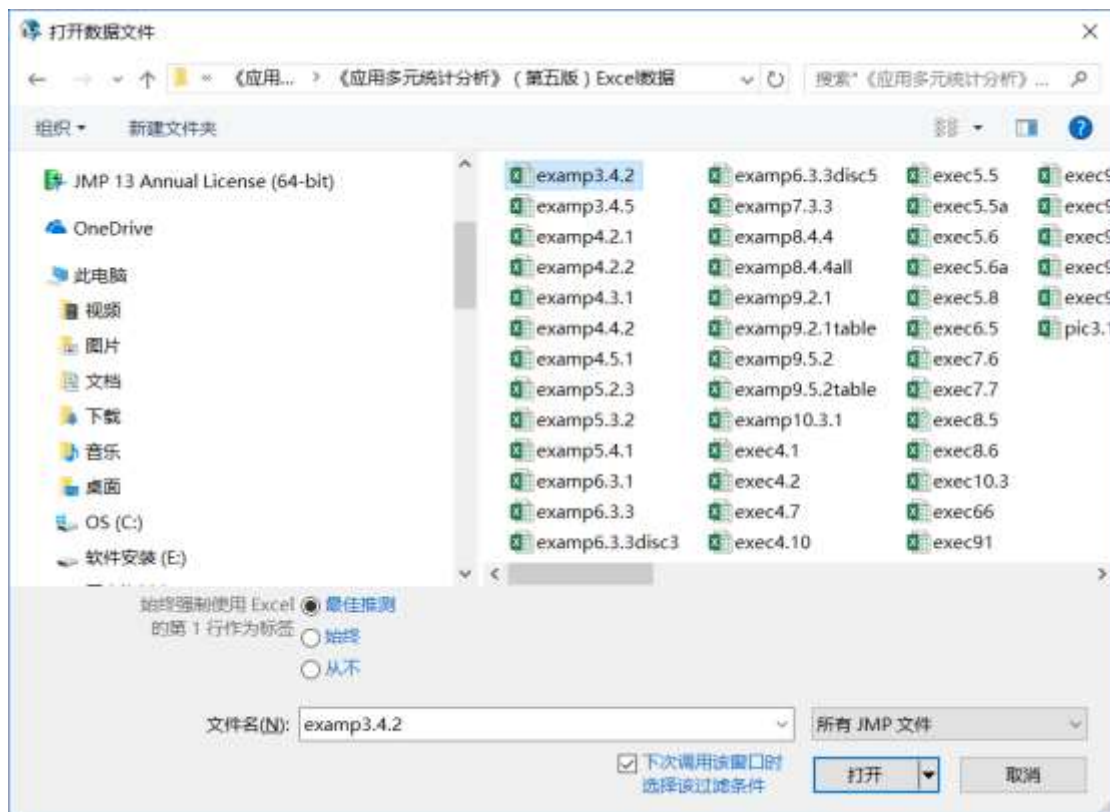


图 2

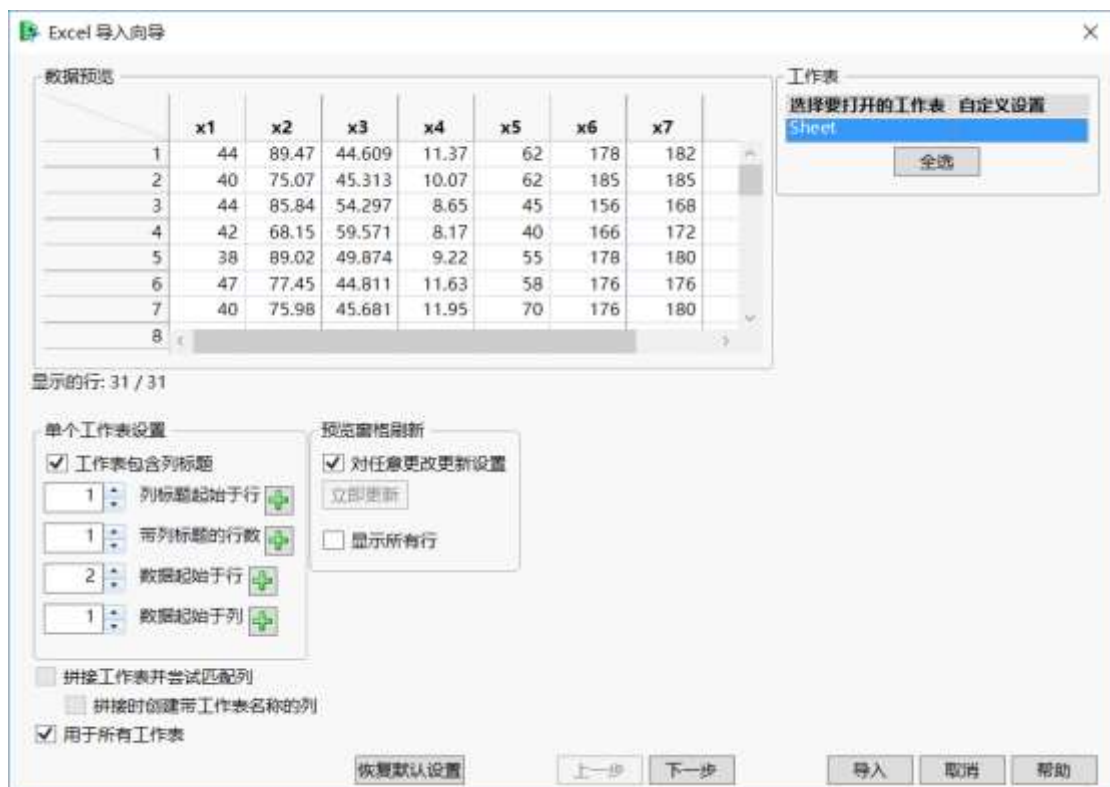


图 3

第三章 多元正态分布

一、对例 3.4.2 进行相关分析等

打开 examp3.4.2.xlsx 数据表（见图 3.1）⇒ 选择 分析 ⇒ 多元方法 ⇒ 多元，随即出现“多元与相关性”对话框（见图 3.2）⇒ 选择 x1,x2,x3,x4,x5,x6,x7，并点击 Y, 列（或将其拖入“Y, 列”列表框内）⇒ 确定 ⇒ 在出现的“多元”窗口中（见图 3.3），点击“多元”旁的红色小三角或在右边空白处点击右键，出现图中菜单，做该菜单中的选择可得相应的结果。



图 3.1

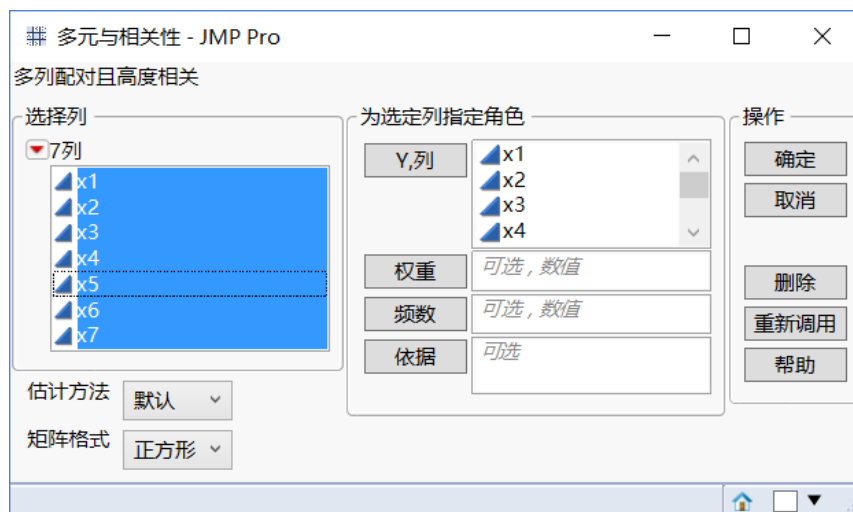


图 3.2

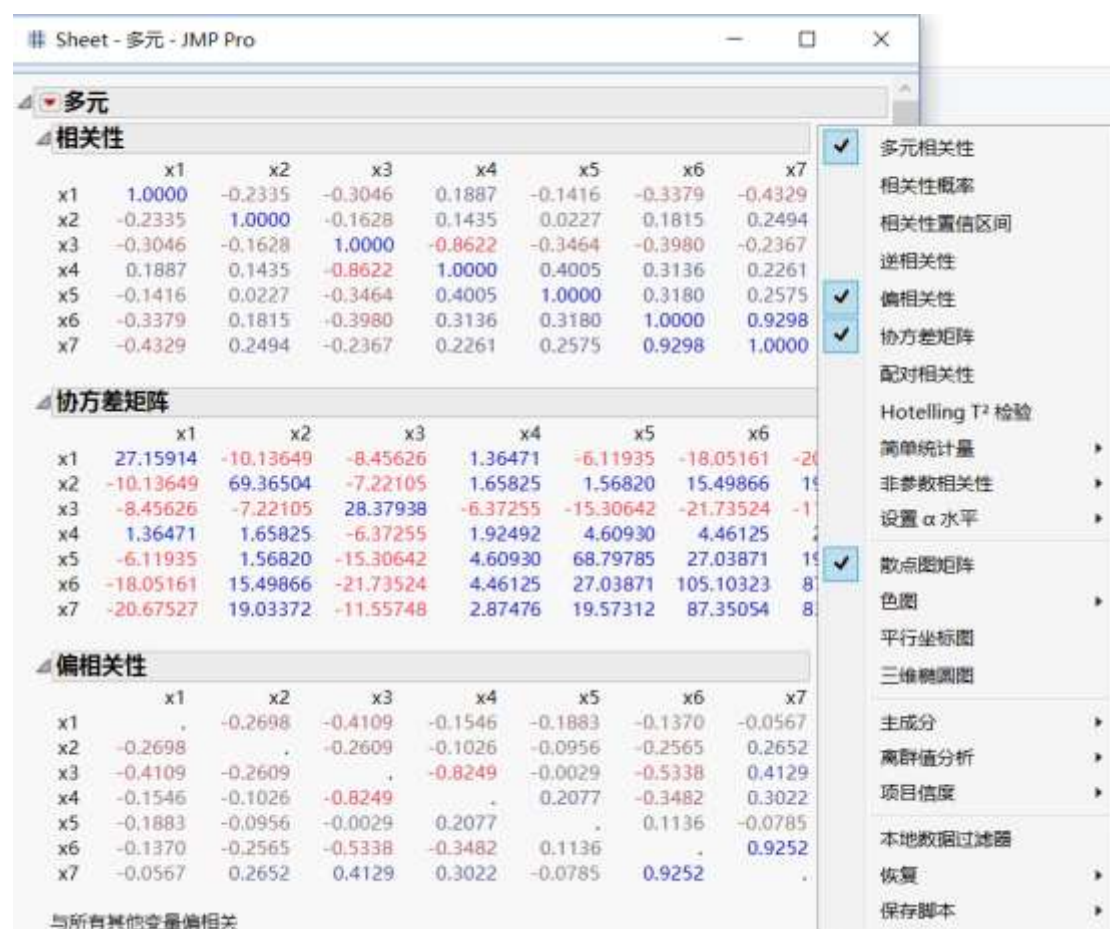


图 3.3

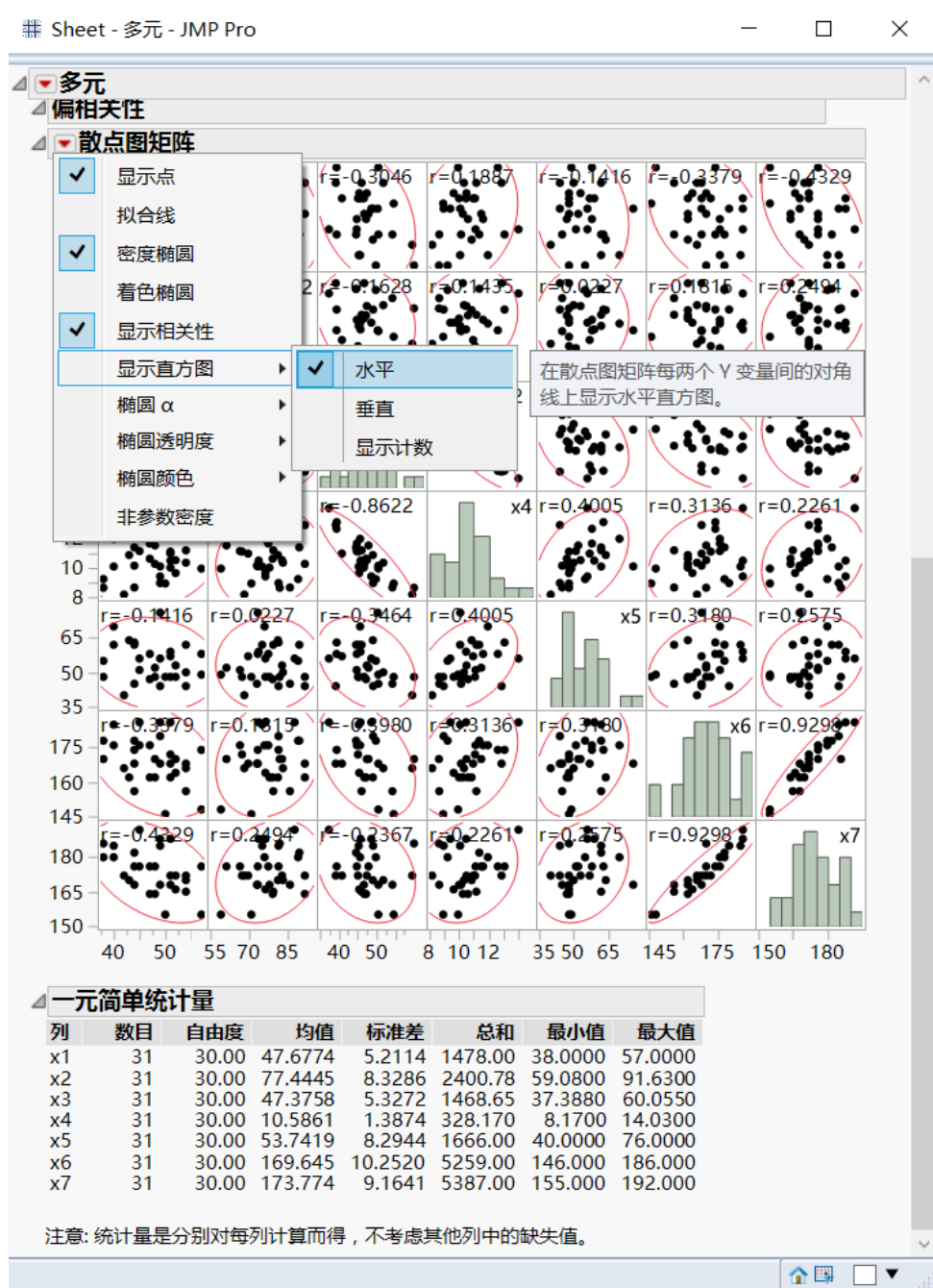


图 3.3 (续)

注: “偏相关性”中的值是给定其他 5 个变量后求得的两个变量间的偏相关系数。

在图 3.3 (续) 中, 点击“散点图矩阵”旁的红色小三角菜单, 做图中选择。在图 3.3 的菜单中, 选择三维椭圆图, 出现如图 3.4 所示的对话框, 点击**确定**, 得到可旋转的三维椭圆图, 如图 3.5 所示。

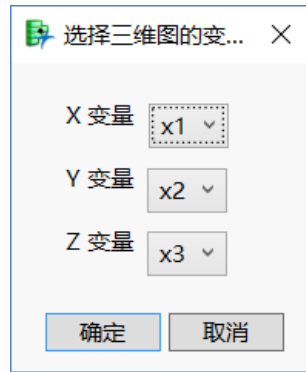


图 3.4

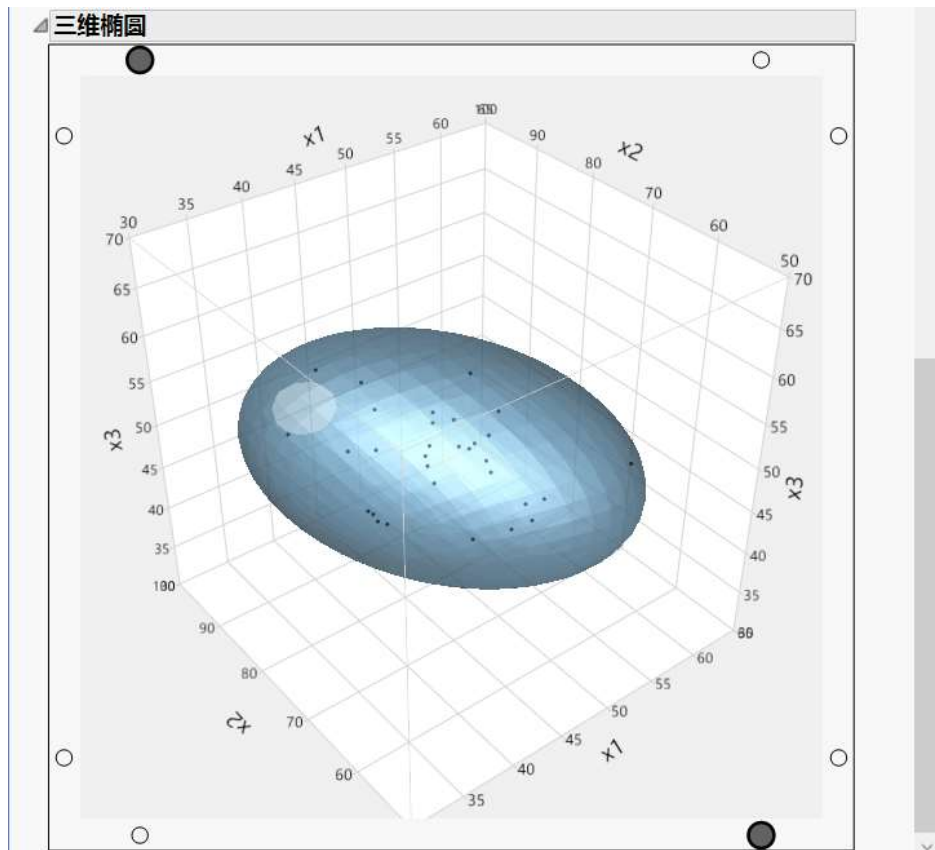


图 3.5

二、对 pic3.1.2.xlsx 数据表作二元正态密度椭圆线

打开 pic3.1.2.xlsx 数据表 \Rightarrow 分析 \Rightarrow 以 X 拟合 Y \Rightarrow 在图 3.6 中选择 Y \rightarrow Y, 响应: 选择 X \rightarrow X, 因子 \Rightarrow 确定 \Rightarrow 在随即出现的窗口中 (见图 3.7), 点击“二元拟合”旁的红色小三角, 做该图中菜单的选择 \Rightarrow 在弹出的对话框中 (见图 3.8), 填入概率值 \rightarrow 确定, 重复操作, 在图 3.8 中分别填入概率值: 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,0.95,0.99, 生成图 3.9。



图 3.6

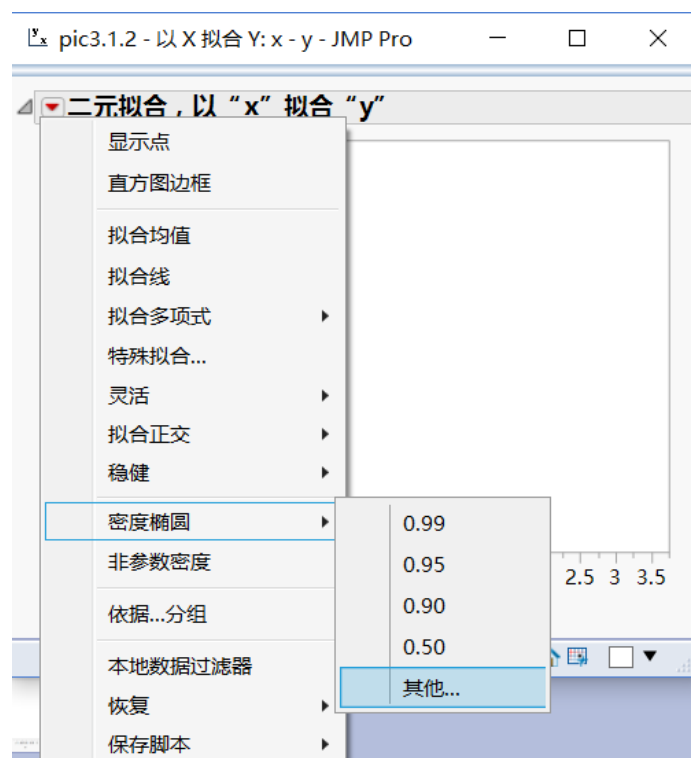


图 3.7

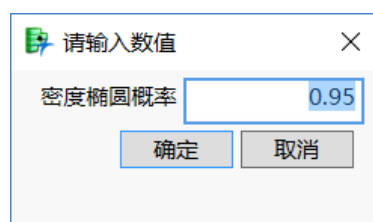


图 3.8

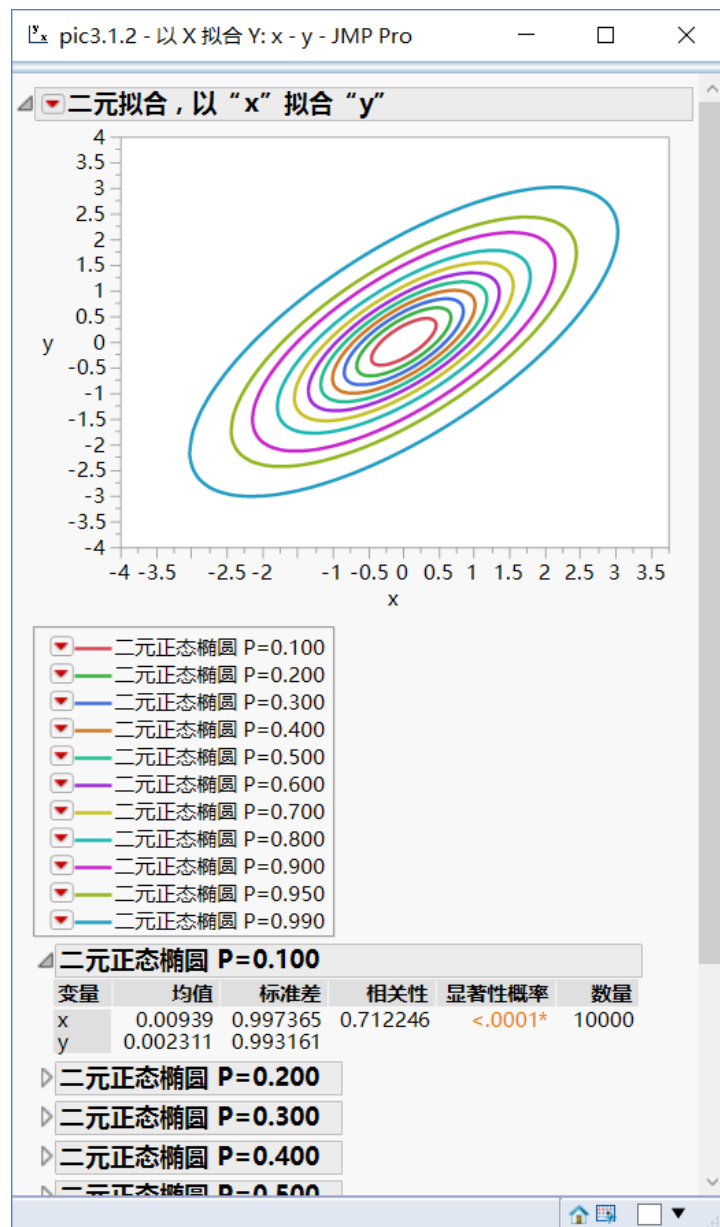


图 3.9

三、对 pic3.1.2.xlsx 数据表作二元密度等高线

打开 pic3.1.2.xlsx 数据表 \Rightarrow 图形 \Rightarrow 等高线图 (见图 3.10) \Rightarrow 在图 3.11 中, 选择 x, y \rightarrow X; 选择 f \rightarrow Y \Rightarrow 确定。随即出现图 3.12。

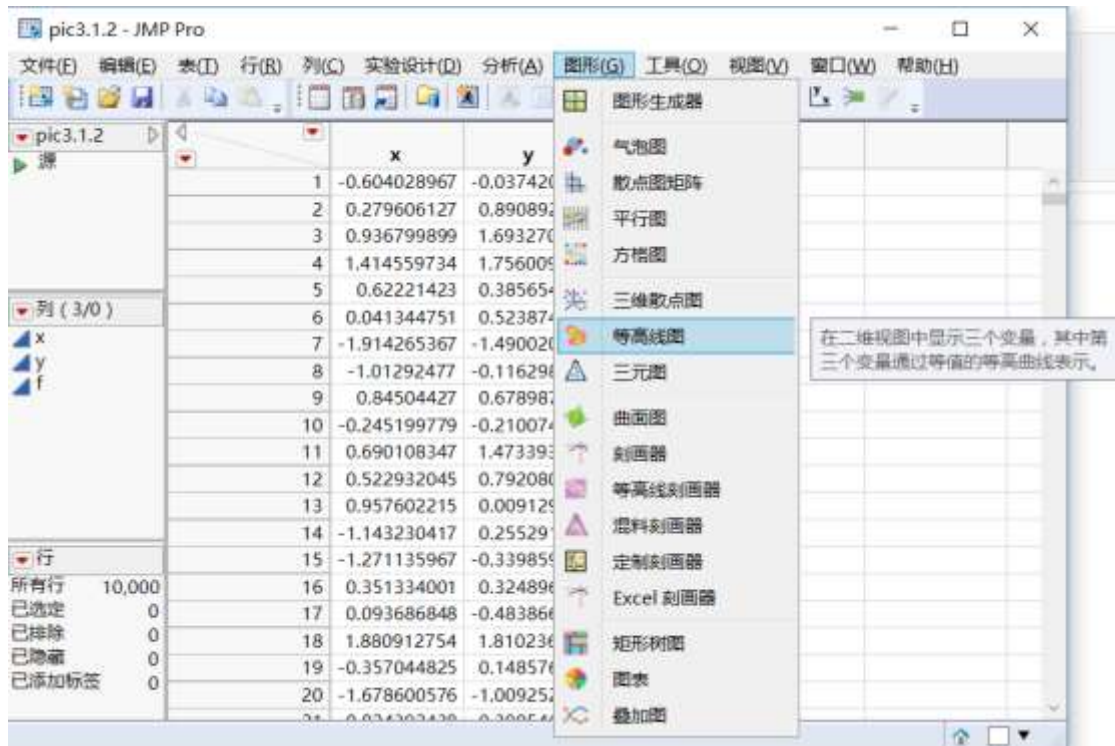


图 3.10

注：该数据表中(x, y)为 10000 对独立的二元正态分布随机数，f 为其密度值。

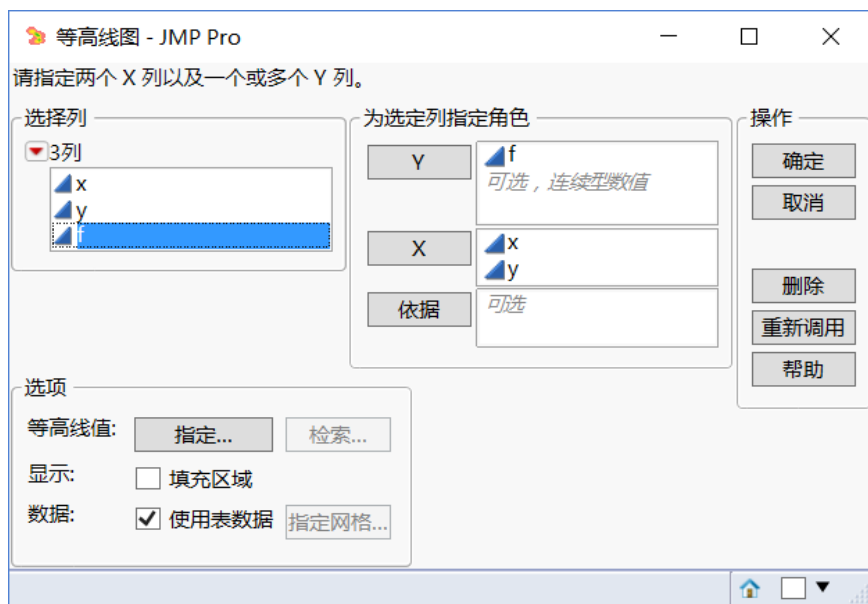


图 3.11

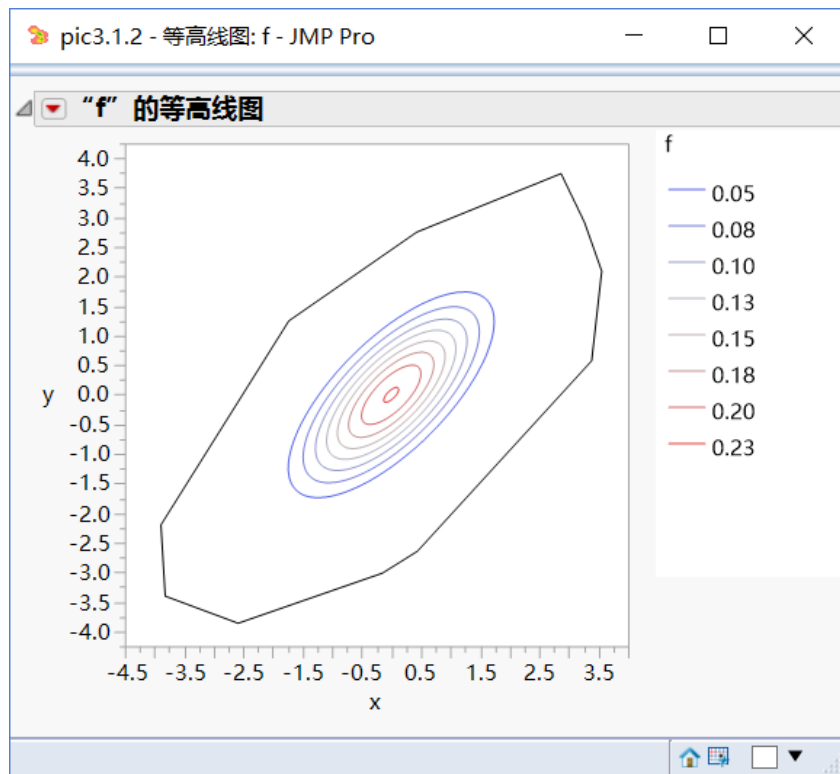


图 3.12

在图 3.12 中右侧的内方框与外方框之间点击右键，在弹出菜单中选择显示数据点（见图 3.13）⇒ 在右侧的内方框内点击右键，作图 3.14 中的选择。可见，散点呈椭圆状，并且沿着椭圆线点的密集程度相同。

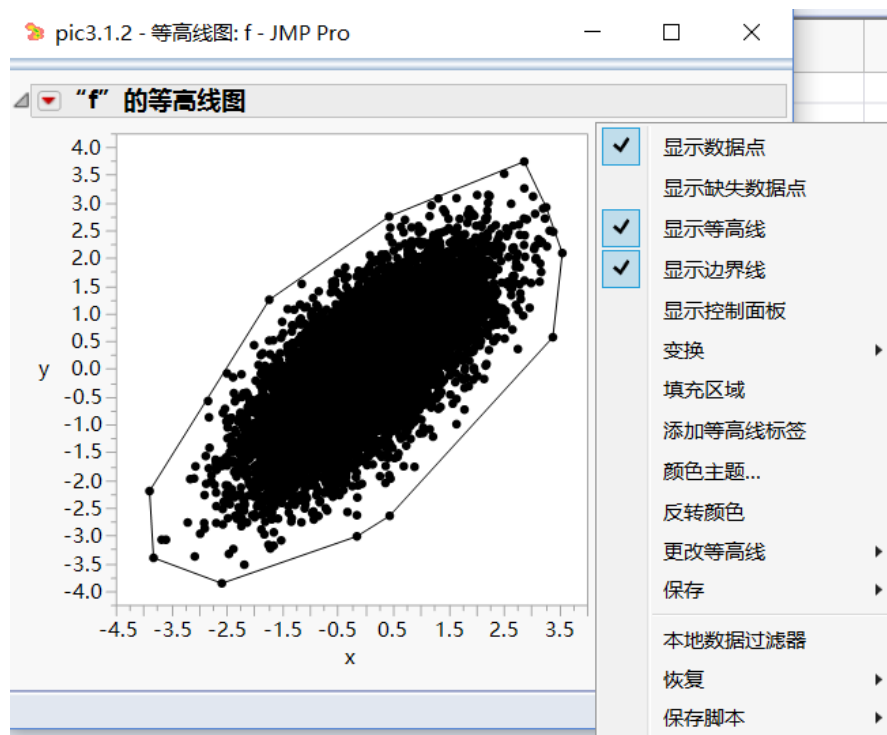


图 3.13

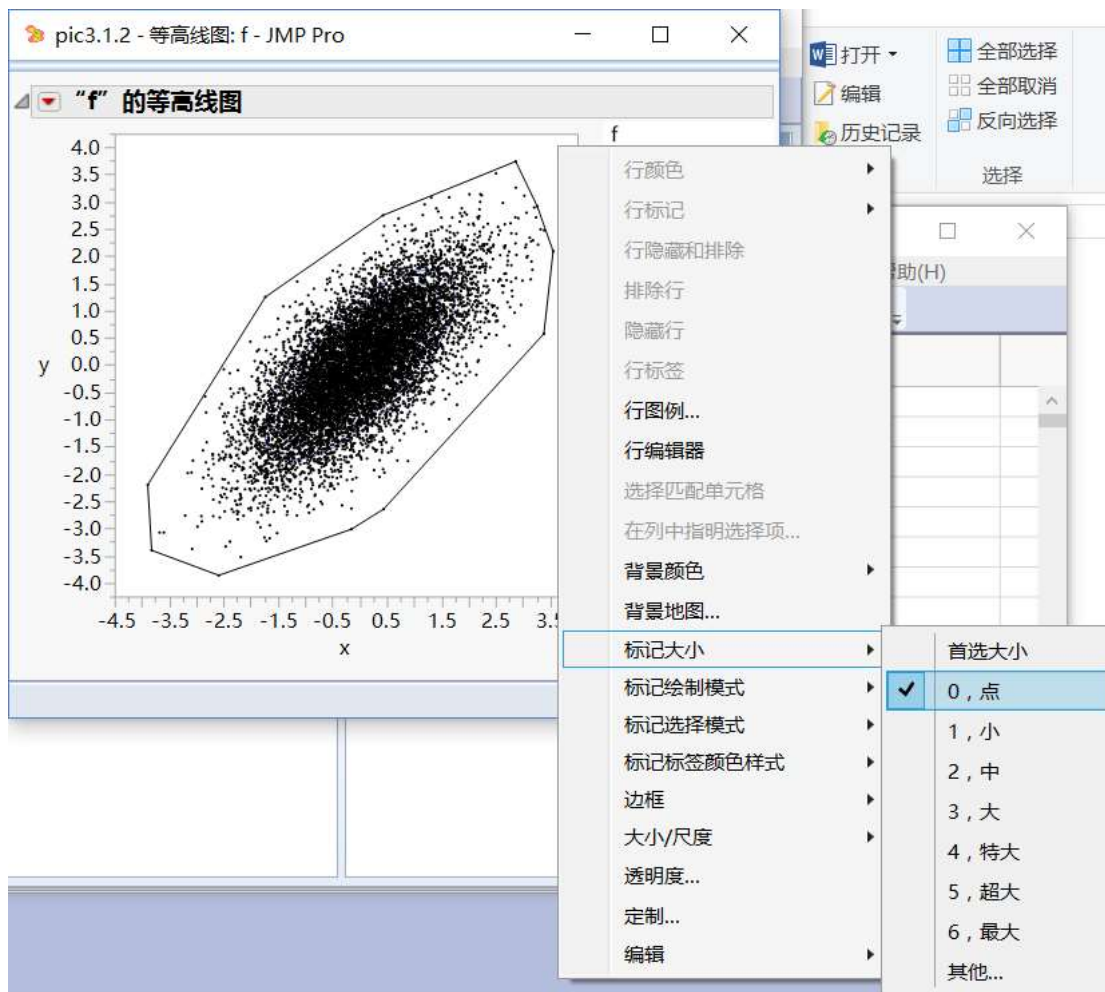


图 3.14

第四章 多元正态总体的统计推断

一、对例 4.2.1 进行检验

打开 examp4.2.1.xlsx 数据表 \Rightarrow 选择 分析 \Rightarrow 多元方法 \Rightarrow 多元，随即出现“多元与相关性”对话框（类似于图 3.2） \Rightarrow 选择 x1,x2,x3，并点击 Y, 列 \Rightarrow 确定 \Rightarrow 在出现的“多元”窗口中（类似于图 3.3），点击“多元”旁的红色小三角（见图 4.1） \Rightarrow 选择 Hotelling T² 检验 \Rightarrow 在出现的“为原假设指定均...”窗口中填入图 4.2 中的数据 \Rightarrow 确定，即生成图 4.3。

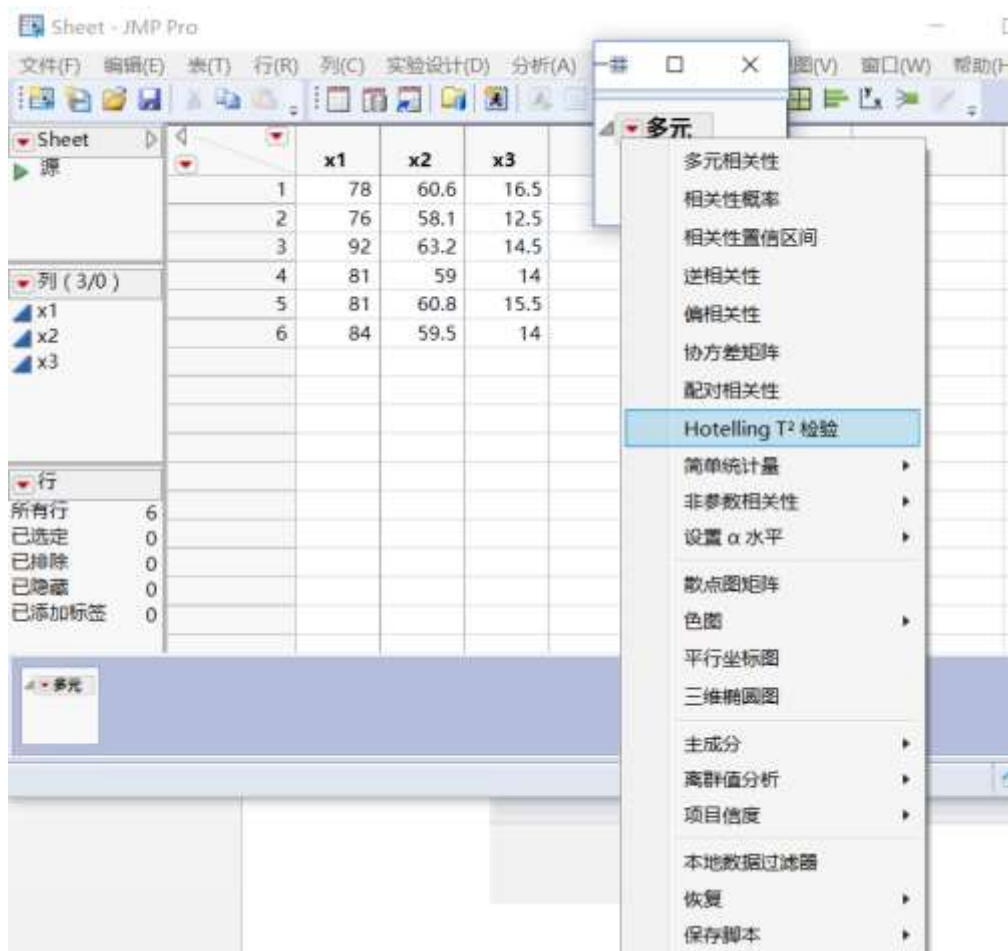


图 4.1

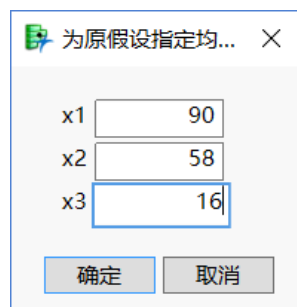


图 4.2

变量	均值	假设均值	检验统计量	F 比	概率>F
x1	82.0000	90.0000	420.445	84.0889	0.0022
x2	60.2000	58.0000			
x3	14.5000	16.0000			

图 4.3

二、对例 4.5.1 进行一元方差分析

打开 examp4.5.1.xlsx 数据表 ⇒ 在变量 g 处点击右键，弹出一菜单（见图 4.4）⇒ 列信息...，出现如图 4.5 所示的对话框 ⇒ 在“建模类型”列表框中，选择名义型（或有序型）⇒

确定，变量 g 即设置为名义型（或有序型）。

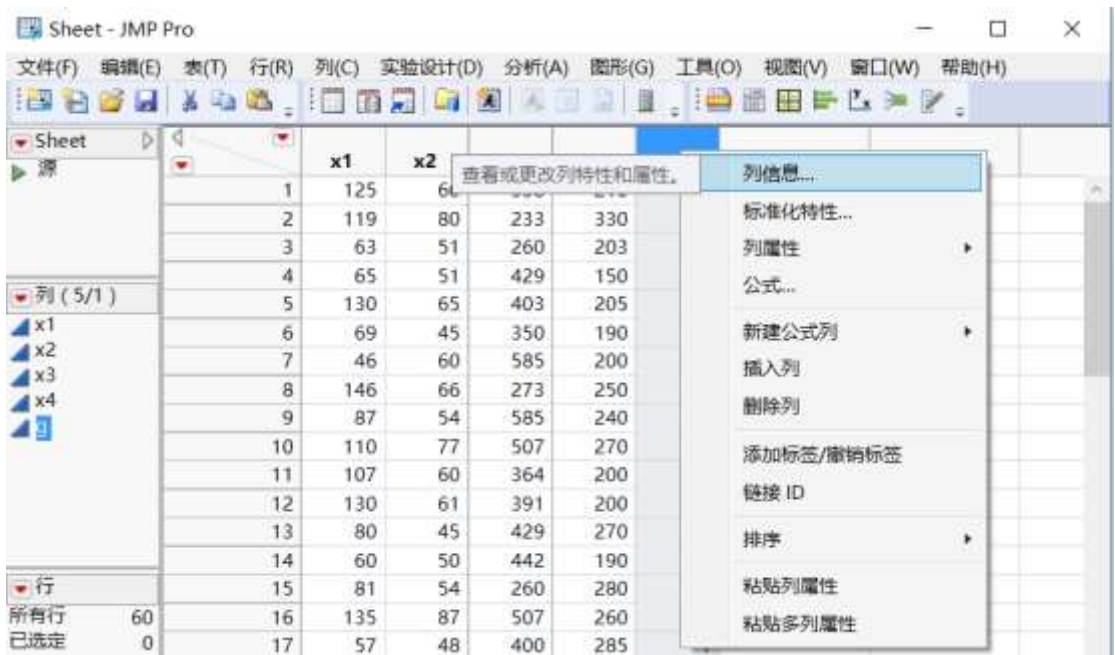


图 4.4

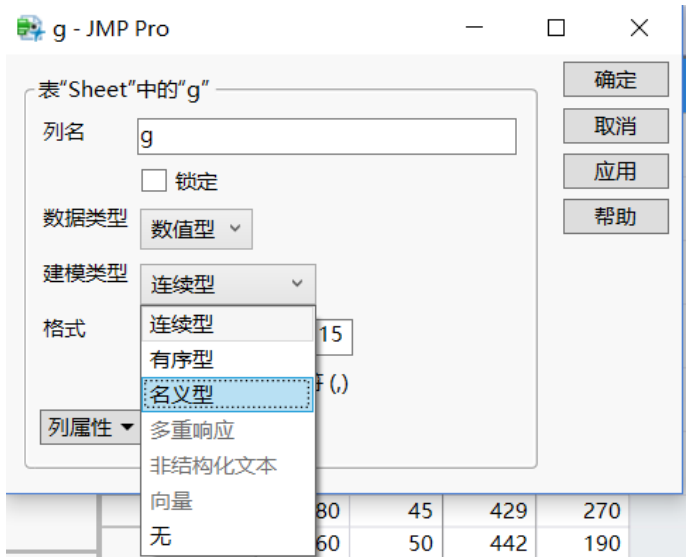


图 4.5

在图 4.4 中，选择分析⇒以 X 拟合 Y，出现“以 X 拟合 Y”对话框（见图 4.6）⇒选择 x1,x2,x3,x4→Y，响应；选择 g→X，因子⇒确定，即出现图 4.7⇒按住 Ctrl 键（该操作可使对这四个变量所作的分析同时进行）的同时点击“单因子分析”旁的红色小三角或右边空白处点击右键，做图中的选择。



图 4.6

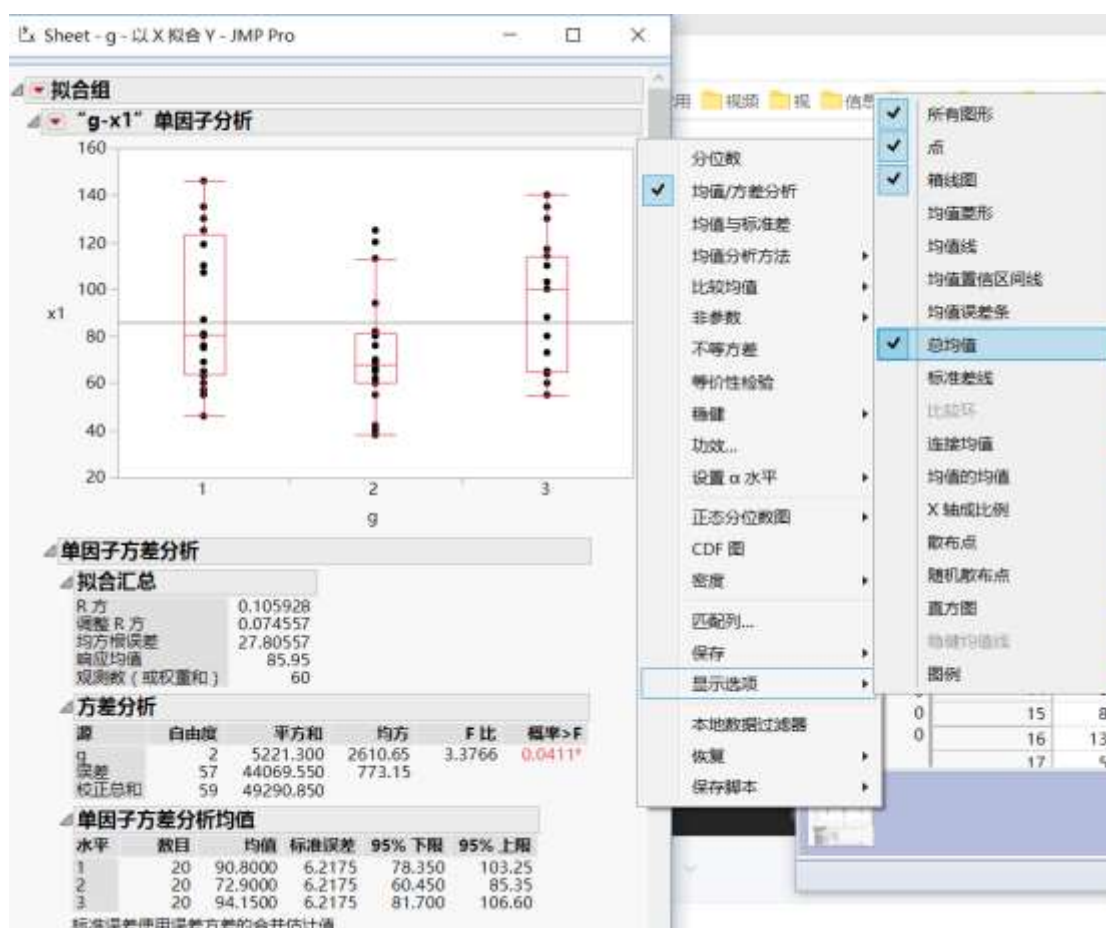


图 4.7

注：这里的输出省略了关于 x2, x3, x4 的一元方差分析结果。

三、对例 4.5.1 进行多元方差分析

类似于图 5.10 的操作，选中“显示典型详细信息”即可，生成图 4.8。

检验	值	近似的 F 值	分子自由度	分母自由度	概率>F
Wilks Lambda	0.6663595	3.0379	8	108	0.0040*
Pillai 迹	0.3612359	3.0309	8	110	0.0041*
Hotelling-Lawley	0.4592792	3.0673	8	74.856	0.0048*
Roy 最大根	0.3360454	4.6206	4	55	0.0027*

图 4.8

第五章 判别分析

一、对书中例 5.2.3 中的数据作 Bayes 判别

打开 examp5.2.3.xlsx 数据表 \Rightarrow 将变量 g 设置为名义型(或有序型) \Rightarrow 在图 5.1 中, 选择 分析 \Rightarrow 多元方法 \Rightarrow 判别, 即出现“判别”对话框(见图 5.2) \Rightarrow 选择 $x_1, x_2, x_3, x_4 \rightarrow Y$, 协变量; 选择 $g \rightarrow X$, 类别; 在“判别方法”列表框中, 使用缺省的线性, 共同协方差 \Rightarrow 确定 \Rightarrow 在出现的窗口中, 点击“判别分析”旁的红色小三角(见图 5.3), 作菜单中的选择, 即得到图 5.4。

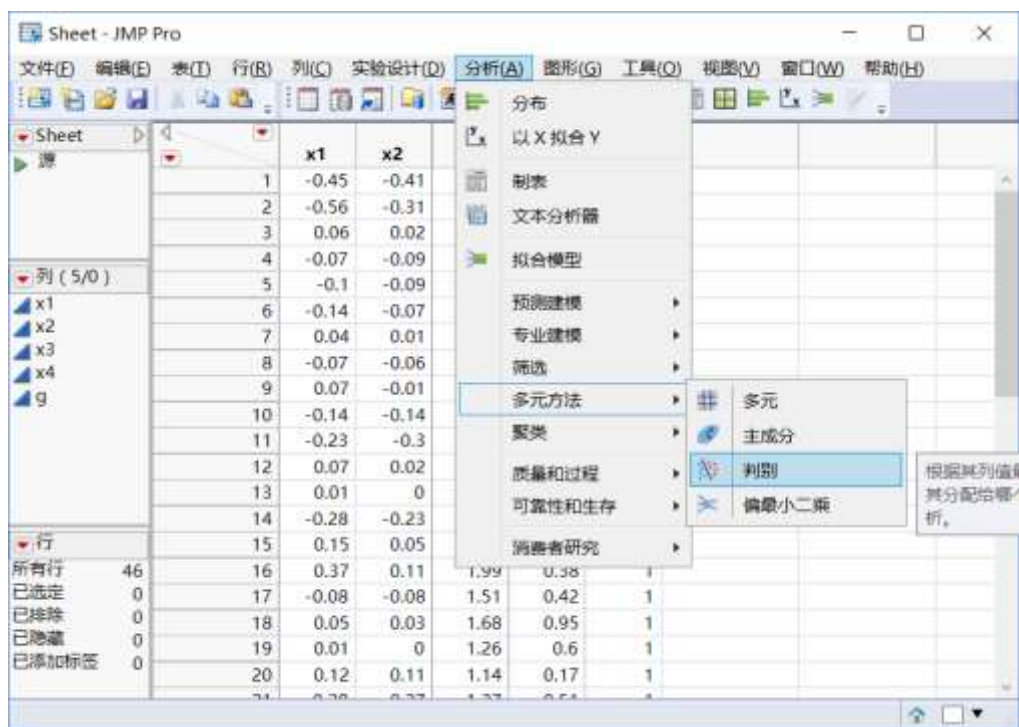


图 5.1

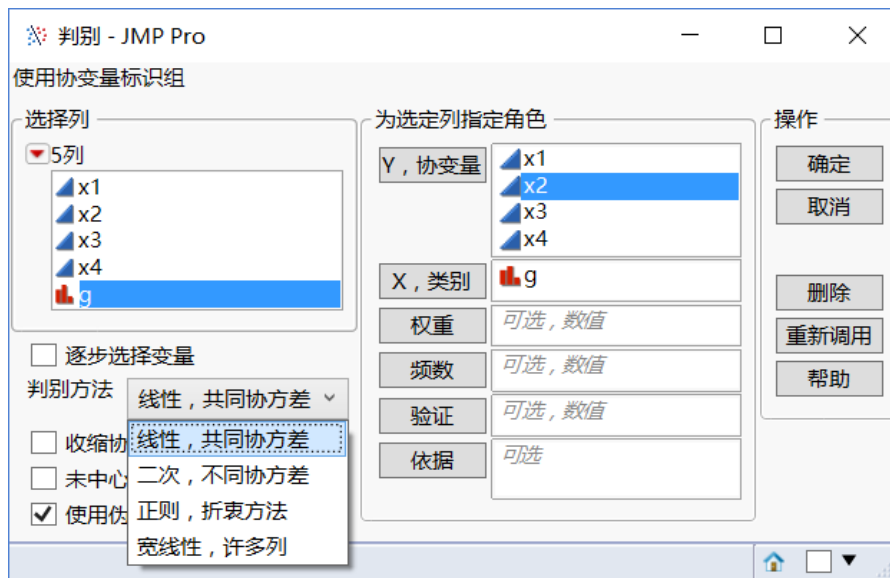


图 5.2

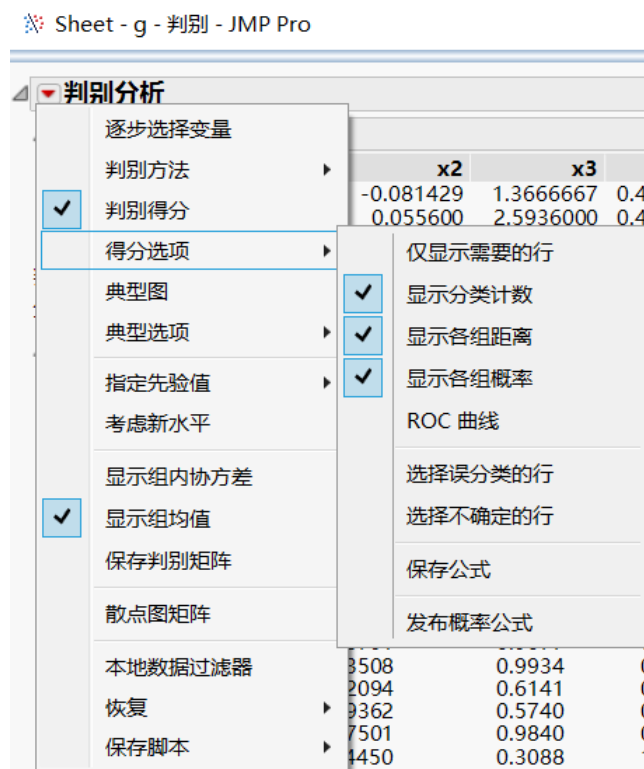


图 5.3

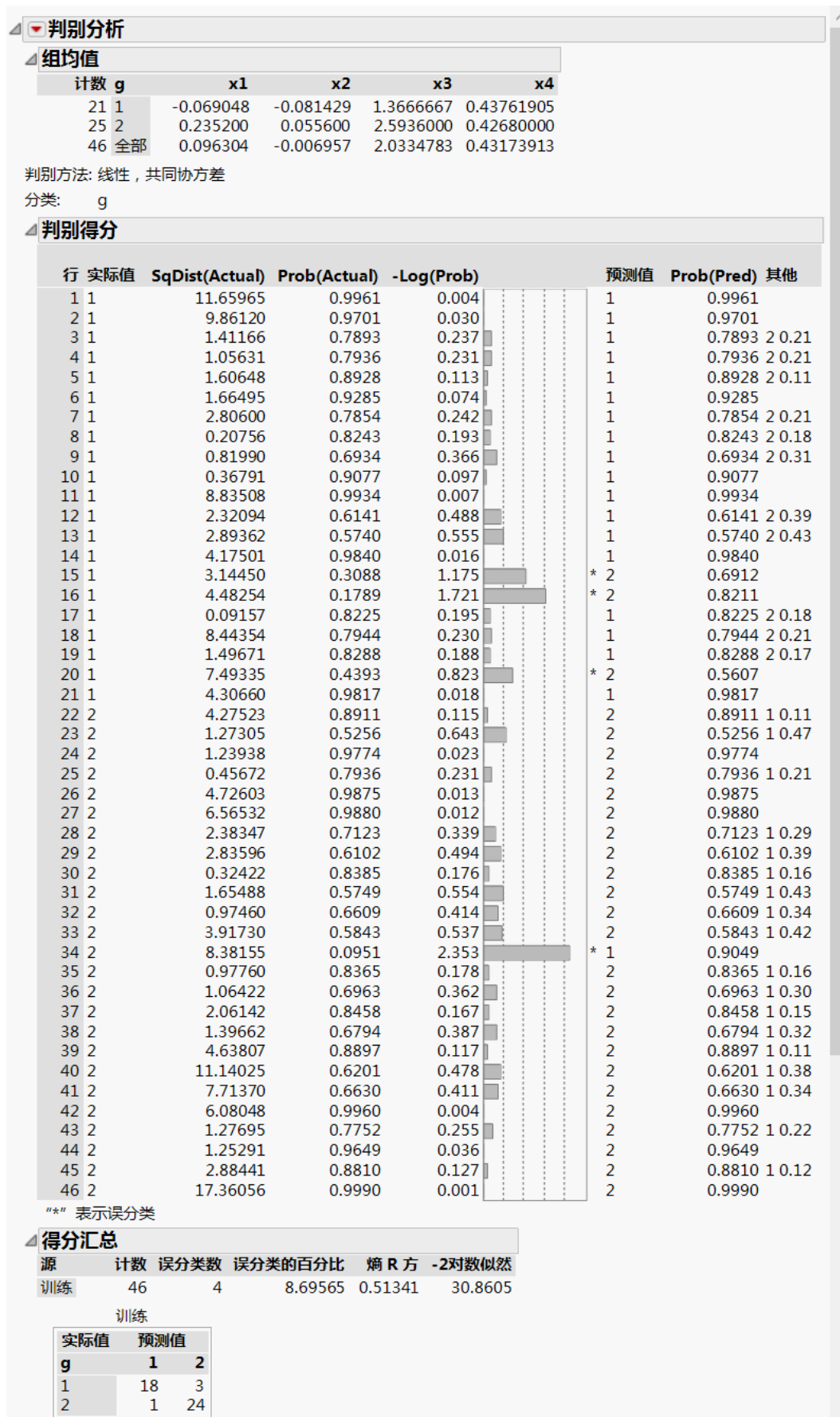


图 5.4

各组概率			
行	实际值	1	2
1	1	0.99608	0.00392
2	1	0.97009	0.02991
3	1	0.78928	0.21072
4	1	0.79365	0.20635
5	1	0.89276	0.10724
6	1	0.92845	0.07155
7	1	0.78538	0.21462
8	1	0.82425	0.17575
9	1	0.69340	0.30660
10	1	0.90770	0.09230
11	1	0.99343	0.00657
12	1	0.61406	0.38594
13	1	0.57395	0.42605
14	1	0.98402	0.01598
15	1	0.30883	0.69117
16	1	0.17892	0.82108
17	1	0.82255	0.17745
18	1	0.79444	0.20556
19	1	0.82884	0.17116
20	1	0.43932	0.56068
21	1	0.98171	0.01829
22	2	0.10890	0.89110
23	2	0.47439	0.52561
24	2	0.02261	0.97739
25	2	0.20644	0.79356
26	2	0.01253	0.98747
27	2	0.01204	0.98796
28	2	0.28771	0.71229
29	2	0.38979	0.61021
30	2	0.16149	0.83851
31	2	0.42511	0.57489
32	2	0.33913	0.66087
33	2	0.41568	0.58432
34	2	0.90494	0.09506
35	2	0.16346	0.83654
36	2	0.30369	0.69631
37	2	0.15422	0.84578
38	2	0.32058	0.67942
39	2	0.11028	0.88972
40	2	0.37986	0.62014
41	2	0.33705	0.66295
42	2	0.00401	0.99599
43	2	0.22476	0.77524
44	2	0.03506	0.96494
45	2	0.11899	0.88101
46	2	0.00096	0.99904

到各组的平方距离			
行	实际值	1	2
1	1	11.65965	22.735637
2	1	9.8612002	16.819554
3	1	1.4116586	4.0528832
4	1	1.0563054	3.750414
5	1	1.6064816	5.8449034
6	1	1.6649545	6.7912436
7	1	2.8060045	5.4006158
8	1	0.2075586	3.2984289
9	1	0.8198986	2.4520487
10	1	0.3679075	4.9395984
11	1	8.8350798	18.872429
12	1	2.3209388	3.2497913
13	1	2.8936168	3.4896099
14	1	4.1750061	12.415226
15	1	3.1444996	1.5332933
16	1	4.4825428	1.4351569
17	1	0.0915689	3.1589867
18	1	8.4435422	11.147323
19	1	1.4967141	4.6515165
20	1	7.493352	7.0054908
21	1	4.3065969	12.272517
22	2	8.4793711	4.2752297
23	2	1.4780814	1.2730528
24	2	8.7723169	1.2393794
25	2	3.1497296	0.456718
26	2	13.459587	4.7260344
27	2	15.37933	6.5653234
28	2	4.196577	2.383466
29	2	3.7323178	2.8359645
30	2	3.6185608	0.324218
31	2	2.258559	1.6548812
32	2	2.308924	0.9745979
33	2	4.5983816	3.9172998
34	2	3.8748332	8.381553
35	2	4.243073	0.9776039
36	2	2.7237755	1.064221
37	2	5.4652033	2.0614166
38	2	2.8988345	1.3966219
39	2	8.813758	4.6380748
40	2	12.120523	11.140253
41	2	9.0666554	7.7136979
42	2	17.109821	6.0804836
43	2	3.753217	1.2769503
44	2	7.8828783	1.2529074
45	2	6.8884962	2.8844108
46	2	31.26258	17.360563

图 5.4 (续)

在图 5.2 的“判别方法”列表框中，选择二次，不同协方差⇒确定⇒仍作图 5.3 中的选择，得图 5.5。

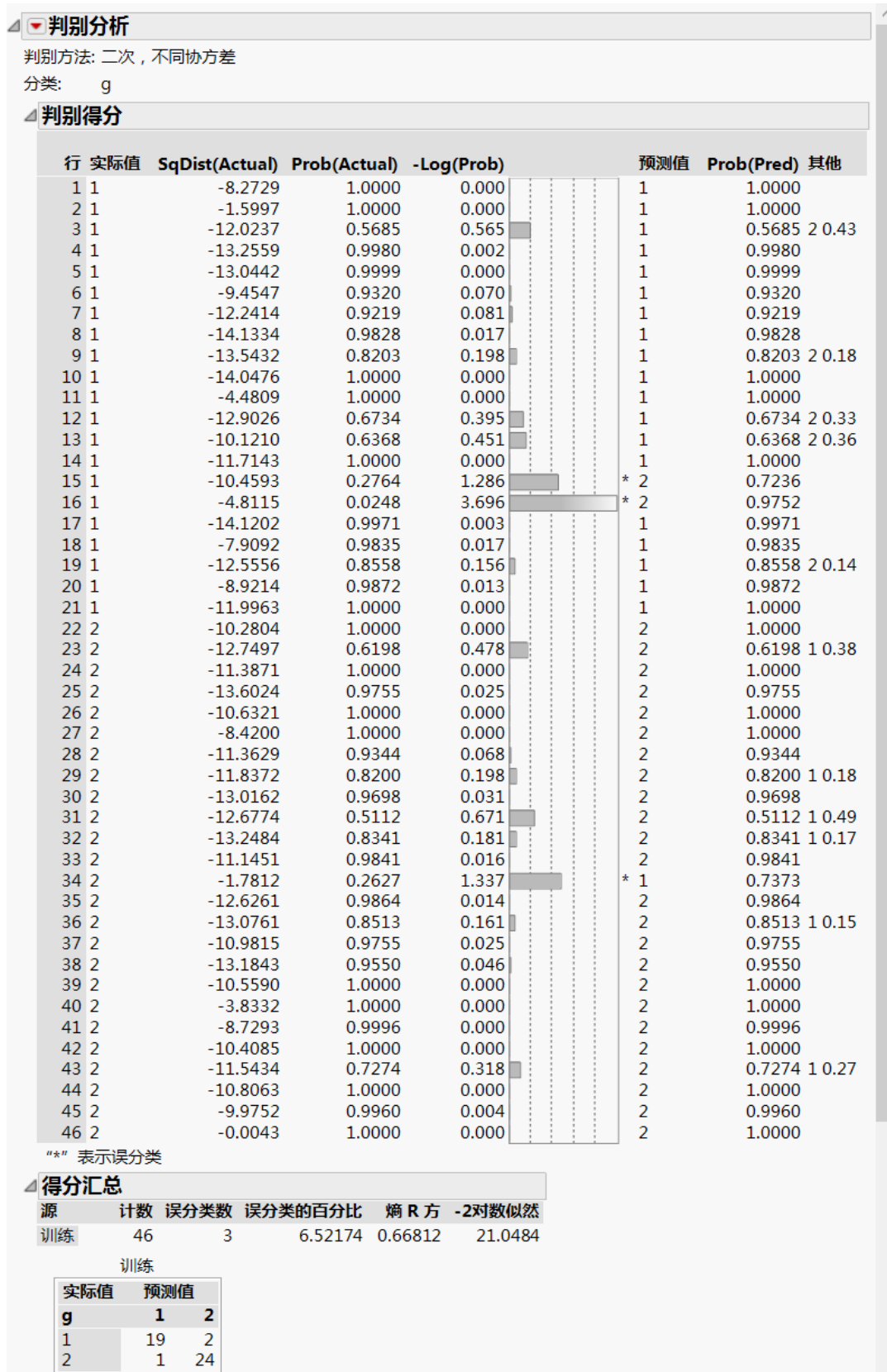


图 5.5

各组概率			
行	实际值	1	2
1	1	1.00000	0.00000
2	1	1.00000	0.00000
3	1	0.56846	0.43154
4	1	0.99801	0.00199
5	1	0.99992	0.00008
6	1	0.93195	0.06805
7	1	0.92191	0.07809
8	1	0.98277	0.01723
9	1	0.82032	0.17968
10	1	1.00000	0.00000
11	1	1.00000	0.00000
12	1	0.67342	0.32658
13	1	0.63683	0.36317
14	1	1.00000	0.00000
15	1	0.27635	0.72365
16	1	0.02483	0.97517
17	1	0.99705	0.00295
18	1	0.98349	0.01651
19	1	0.85584	0.14416
20	1	0.98723	0.01277
21	1	1.00000	0.00000
22	2	0.00000	1.00000
23	2	0.38017	0.61983
24	2	0.00000	1.00000
25	2	0.02448	0.97552
26	2	0.00000	1.00000
27	2	0.00000	1.00000
28	2	0.06564	0.93436
29	2	0.17996	0.82004
30	2	0.03023	0.96977
31	2	0.48877	0.51123
32	2	0.16595	0.83405
33	2	0.01590	0.98410
34	2	0.73734	0.26266
35	2	0.01355	0.98645
36	2	0.14871	0.85129
37	2	0.02450	0.97550
38	2	0.04504	0.95496
39	2	0.00000	1.00000
40	2	0.00001	0.99999
41	2	0.00042	0.99958
42	2	0.00000	1.00000
43	2	0.27261	0.72739
44	2	0.00001	0.99999
45	2	0.00398	0.99602
46	2	0.00000	1.00000

到各组的平方距离			
行	实际值	1	2
1	1	-8.27286	161.94474
2	1	-1.599705	59.61451
3	1	-12.02373	-11.47259
4	1	-13.25595	-0.822955
5	1	-13.04424	5.7277688
6	1	-9.454736	-4.220612
7	1	-12.24141	-7.304367
8	1	-14.13342	-6.045498
9	1	-13.54322	-10.50612
10	1	-14.04758	13.157095
11	1	-4.480918	94.457519
12	1	-12.90259	-11.45521
13	1	-10.12097	-8.997709
14	1	-11.71429	55.725857
15	1	-10.45934	-12.38459
16	1	-4.811489	-12.15229
17	1	-14.12025	-2.473915
18	1	-7.909163	0.2657259
19	1	-12.55565	-8.993261
20	1	-8.921366	-0.225332
21	1	-11.99628	72.646502
22	2	14.347431	-10.28044
23	2	-11.7721	-12.74974
24	2	21.356718	-11.38713
25	2	-6.232238	-13.60235
26	2	52.801802	-10.63206
27	2	63.616057	-8.419959
28	2	-6.051662	-11.36292
29	2	-8.803991	-11.83722
30	2	-6.079975	-13.01622
31	2	-12.58751	-12.6774
32	2	-10.01919	-13.24844
33	2	-2.89441	-11.14512
34	2	-3.845594	-1.781219
35	2	-4.051231	-12.62606
36	2	-9.586638	-13.07612
37	2	-3.613082	-10.98149
38	2	-7.07625	-13.1843
39	2	14.477546	-10.55903
40	2	19.783446	-3.833224
41	2	6.8339663	-8.729308
42	2	77.995011	-10.40853
43	2	-9.580632	-11.54342
44	2	13.509588	-10.80627
45	2	1.0689614	-9.975175
46	2	157.4424	-0.004343

图 5.5（续）

注：“平方距离”就是书上的“广义平方距离”，可以是负值。

二、对书中例 5.3.2 进行 Bayes 判别

打开 [examp5.2.3.xlsx](#) 和 [examp5.3.2.xlsx](#) 数据表，将 [examp5.3.2.xlsx](#) 中未判企业的四个变量值复制至原样本数据后面一行（见图 5.6）⇒ 将变量 g 设置为名义型（或有序型）⇒ 在

类似图 5.1 中，选择分析⇒多元方法>⇒判别，即出现“判别”对话框（类似图 5.2）⇒选择 x1,x2,x3,x4→Y，协变量；选择 g→X，类别⇒确定⇒在出现的窗口中，点击“判别分析”旁的红色小三角（见图 5.7）⇒指定先验值>⇒其他...⇒在弹出的图 5.8 所示的对话框中作图中的输入⇒确定，随即得到图 5.9 的结果。



图 5.6

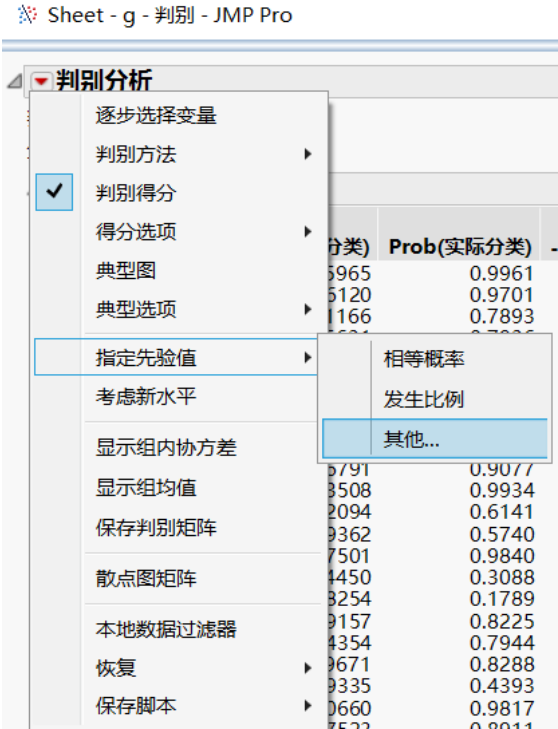


图 5.7

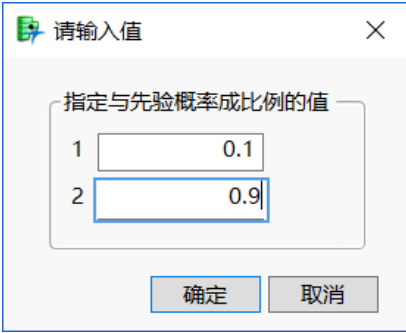


图 5.8

Sheet - g - 判别 - JMP Pro

判别分析

判别得分

行	实际分类	SqDist(实际分类)	Prob(实际分类)	-Log(Prob)	预测值	Prob(预测分类)	其他
37	2	2.27214	0.9801	0.020	2	0.9801	
38	2	1.60734	0.9502	0.051	2	0.9502	
39	2	4.84880	0.9864	0.014	2	0.9864	
40	2	11.35097	0.9363	0.066	2	0.9363	
41	2	7.92442	0.9465	0.055	2	0.9465	
42	2	6.29120	0.9996	0.000	2	0.9996	
43	2	1.48767	0.9688	0.032	2	0.9688	
44	2	1.46363	0.9960	0.004	2	0.9960	
45	2	3.09513	0.9852	0.015	2	0.9852	
46	2	17.57128	0.9999	0.000	2	0.9999	
47					2	0.5230	1 0.48

“*” 表示误分类

图 5.9

三、对书中例 5.4.1 中的数据作 Fisher 判别

打开 examp5.4.1.xlsx 数据表 \Rightarrow 将变量 g 设置为名义型（或有序型） \Rightarrow 在与图 5.2 类似的对话框中，点击确定 \Rightarrow 作图 5.10 中菜单的选择，并点击点着色，即得到图 5.11。

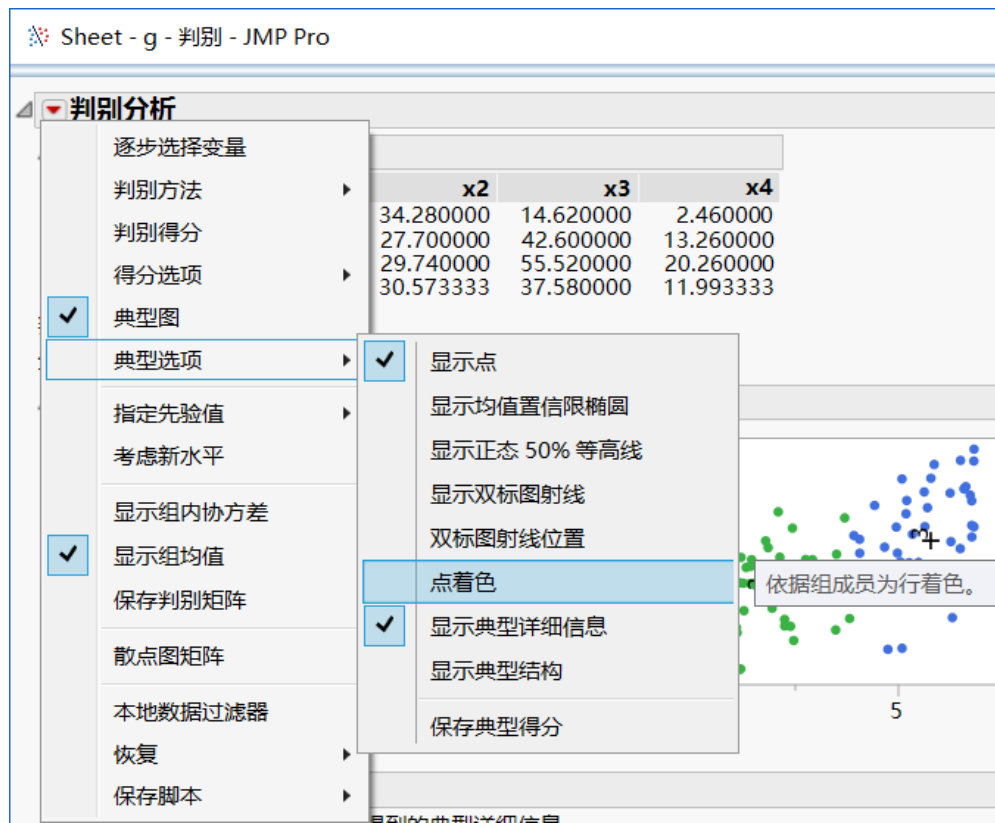


图 5.10

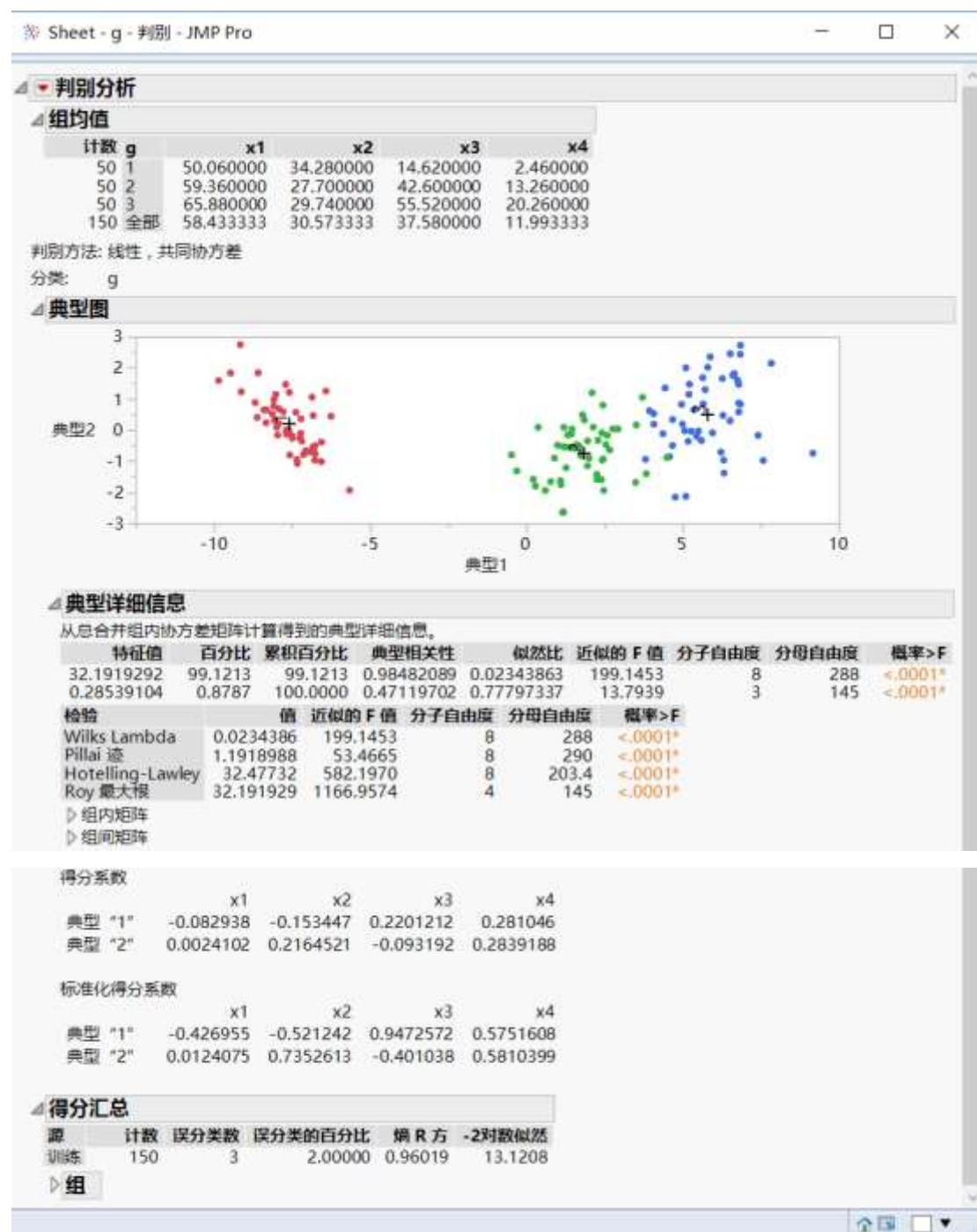


图 5.11

注: 这里的“得分系数”和“标准化得分系数”分别是书上输出 5-1.9 中的“原始典型系数”和“合并类内标准化典型系数”。

第六章 聚类分析

一、对书中例 6.3.3 用 Ward 方法聚类

打开 examp6.3.3.xlsx 数据表 \Rightarrow 分析 \Rightarrow 聚类 \Rightarrow 层次聚类, 即出现“聚类”对话框 (见图 6.1) \Rightarrow 选择 x1,x2,x3,x4,x5,x6,x7,x8 \rightarrow Y, 列; 选择 region \rightarrow 标签; 对“选项”列表框做图

中的选择⇒确定⇒从“层次聚类”相应的红色小三角菜单中（见图 6.2），作图中选择，即生成图 6.3。

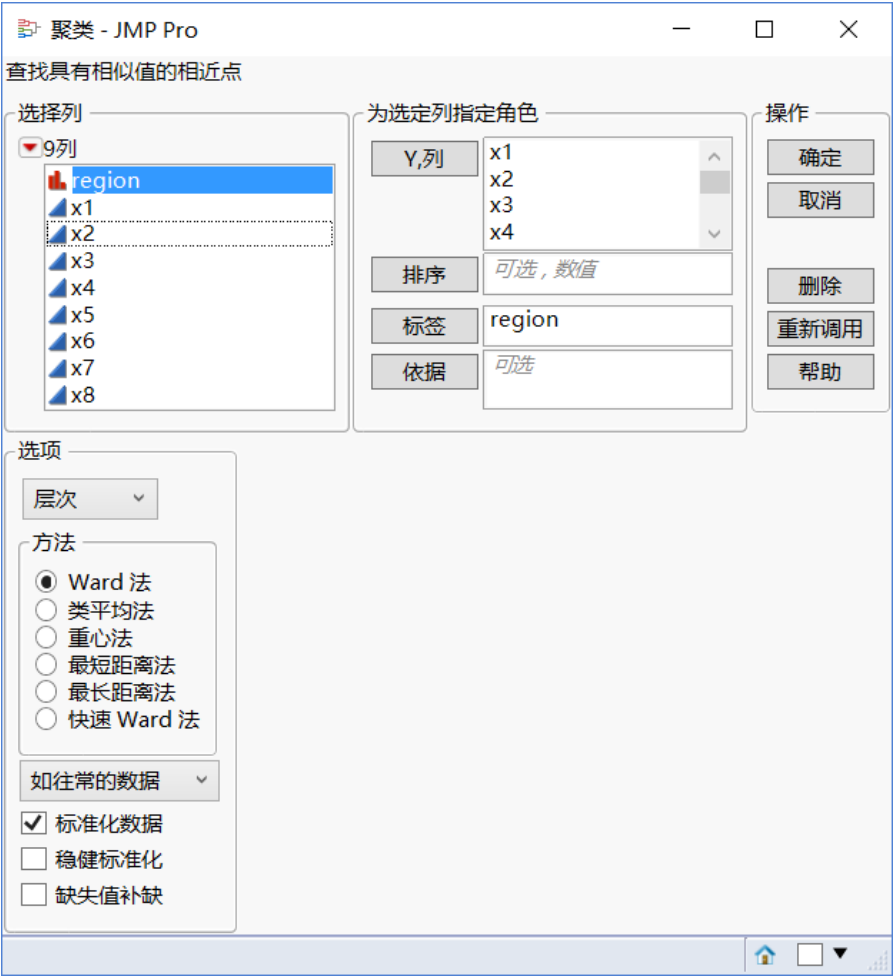


图 6.1



图 6.2

在图 6.3 中，点击“聚类历史记录”旁的展开按钮，得图 6.4。

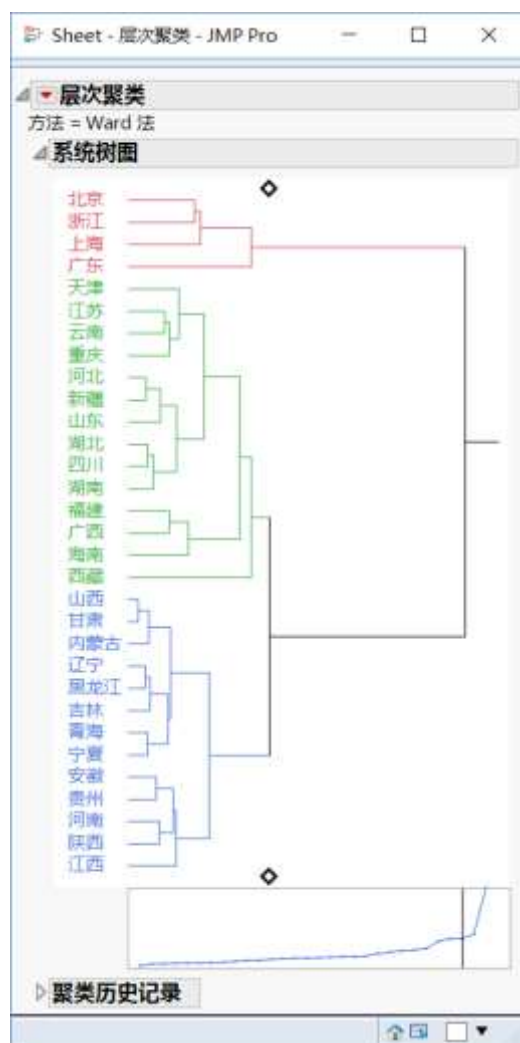


图 6.3

聚类数	距离	前导对象	连接对象
30	0.32559044	山西	甘肃
29	0.56143933	辽宁	黑龙江
28	0.58168392	河北	新疆
27	0.63257216	湖北	四川
26	0.63791323	青海	宁夏
25	0.66595326	山西	内蒙古
24	0.70958145	辽宁	吉林
23	0.82741173	湖北	湖南
22	0.91578198	安徽	贵州
21	1.02282615	河南	陕西
20	1.04729155	河北	山东
19	1.18947071	江苏	云南
18	1.30490467	辽宁	青海
17	1.34270944	江苏	重庆
16	1.34726395	福建	广西
15	1.38551457	山西	辽宁
14	1.46329058	安徽	河南
13	1.56136592	安徽	江西
12	1.58292721	河北	湖北
11	1.65222111	天津	江苏
10	1.96590760	福建	海南
9	2.17199323	北京	浙江
8	2.35849392	北京	上海
7	2.48282693	天津	河北
6	2.65378510	山西	安徽
5	3.64466442	天津	福建
4	4.02111846	北京	广东
3	4.02468232	天津	西藏
2	4.61702849	天津	山西
1	11.00138219	北京	天津

图 6.4

在图 6.2 中，点击聚类数→在出现的对话框中，填入 3→确定，即聚 3 类；点击平行坐标图，产生图 6.5；再点击保存聚类，聚类结果将出现在原数据表中，见图 6.6。

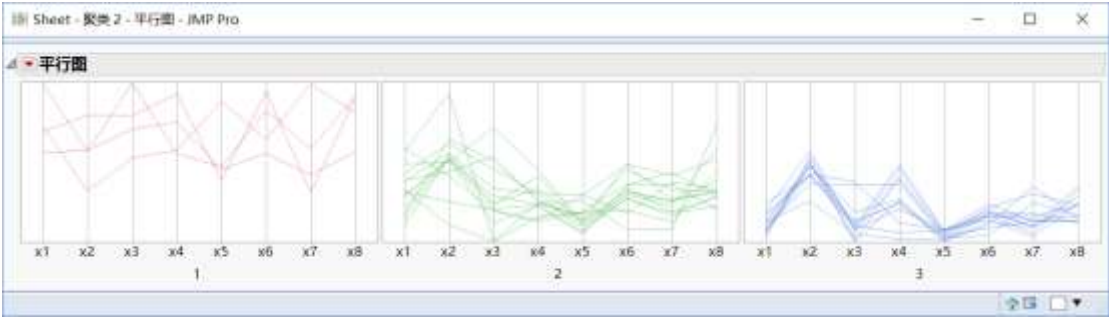
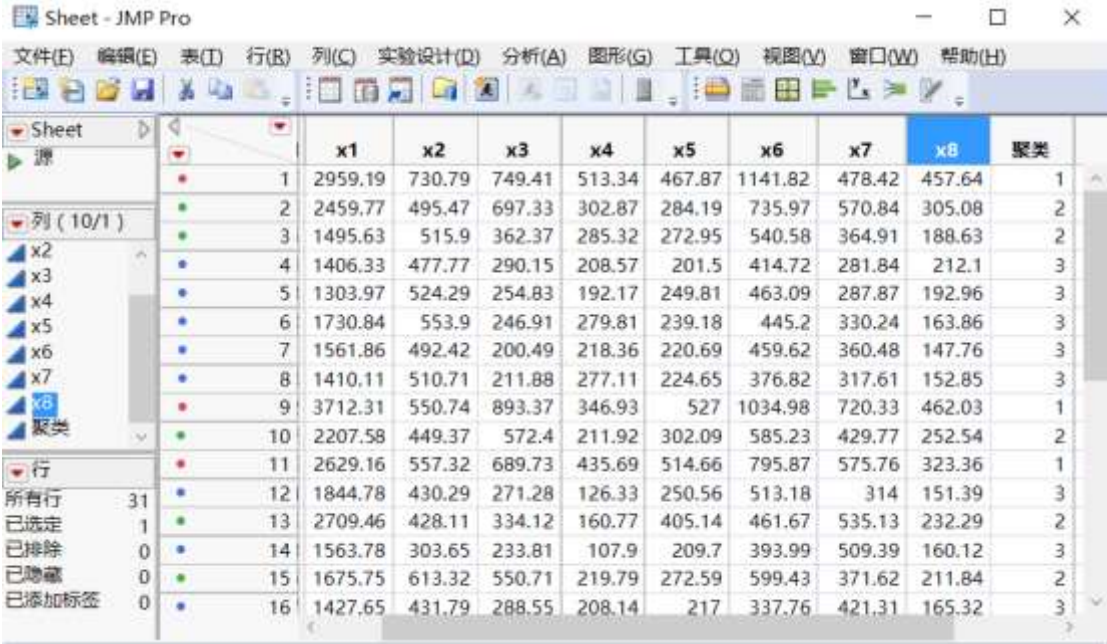


图 6.5

Sheet - JMP Pro

文件(F) 编辑(E) 表(T) 行(R) 列(C) 实验设计(D) 分析(A) 图形(G) 工具(O) 视图(V) 窗口(W) 帮助(H)



		x1	x2	x3	x4	x5	x6	x7	x8	聚类
1	2959.19	730.79	749.41	513.34	467.87	1141.82	478.42	457.64	1	
2	2459.77	495.47	697.33	302.87	284.19	735.97	570.84	305.08	2	
3	1495.63	515.9	362.37	285.32	272.95	540.58	364.91	188.63	2	
4	1406.33	477.77	290.15	208.57	201.5	414.72	281.84	212.1	3	
5	1303.97	524.29	254.83	192.17	249.81	463.09	287.87	192.96	3	
6	1730.84	553.9	246.91	279.81	239.18	445.2	330.24	163.86	3	
7	1561.86	492.42	200.49	218.36	220.69	459.62	360.48	147.76	3	
8	1410.11	510.71	211.88	277.11	224.65	376.82	317.61	152.85	3	
9	3712.31	550.74	893.37	346.93	527	1034.98	720.33	462.03	1	
10	2207.58	449.37	572.4	211.92	302.09	585.23	429.77	252.54	2	
11	2629.16	557.32	689.73	435.69	514.66	795.87	575.76	323.36	1	
12	1844.78	430.29	271.28	126.33	250.56	513.18	314	151.39	3	
13	2709.46	428.11	334.12	160.77	405.14	461.67	535.13	232.29	2	
14	1563.78	303.65	233.81	107.9	209.7	393.99	509.39	160.12	3	
15	1675.75	613.32	550.71	219.79	272.59	599.43	371.62	211.84	2	
16	1427.65	431.79	288.55	208.14	217	337.76	421.31	165.32	3	

图 6.6

二、对书中例 6.3.3 用 k 均值法聚类

在图 6.1 的“选项”列表框中选择 k 均值，如图 6.7 所示⇒确定，出现图 6.8⇒在“聚类数…”框内填入 5⇒执行，出现图 6.9（缺省时“聚类标准差”未展开）⇒从“K 个均值聚类数=5”相应的红色小三角菜单中（见图 6.10），选保存聚类，聚类结果即保存在原数据表中（见图 6.11）。



图 6.7

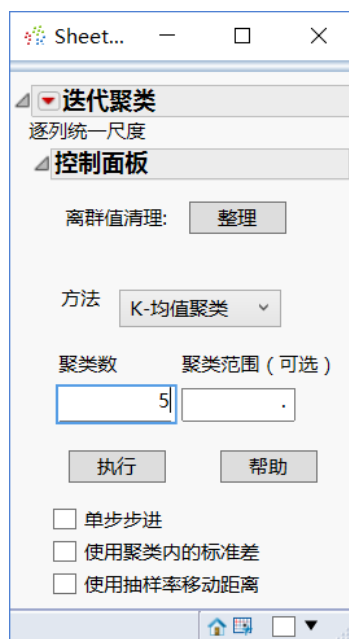


图 6.8

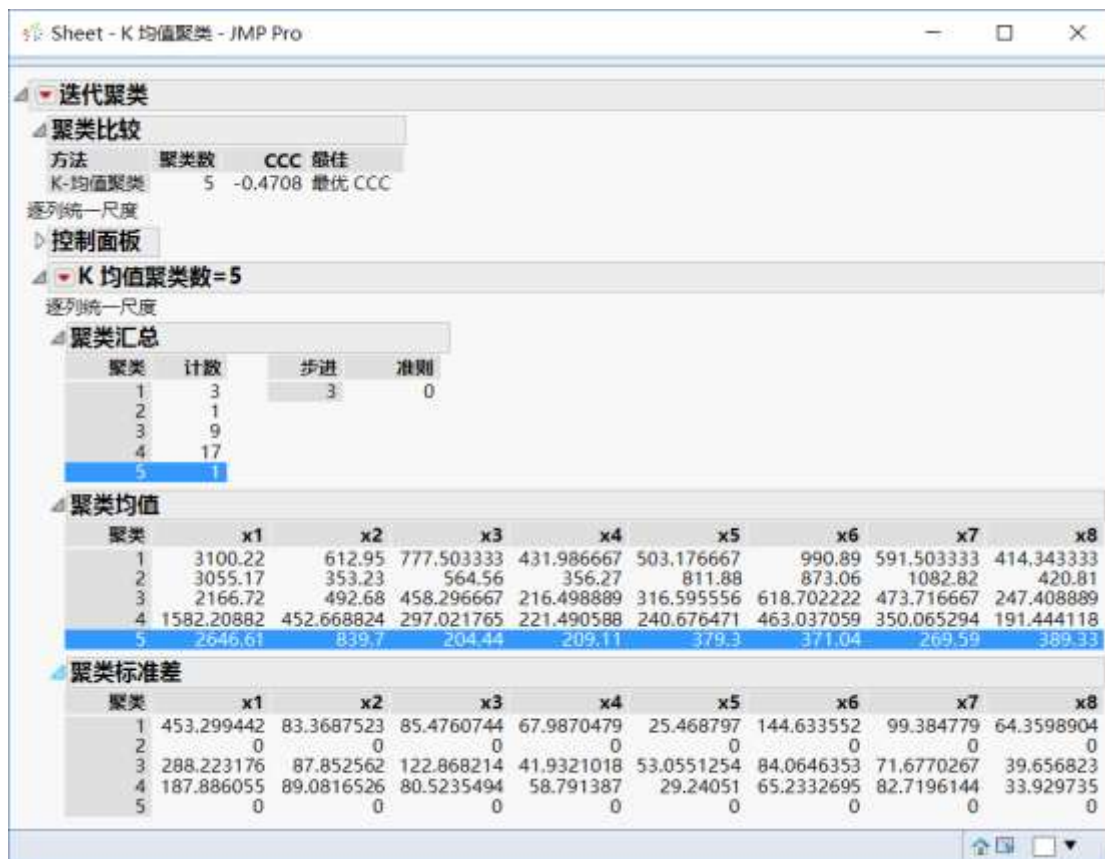


图 6.9

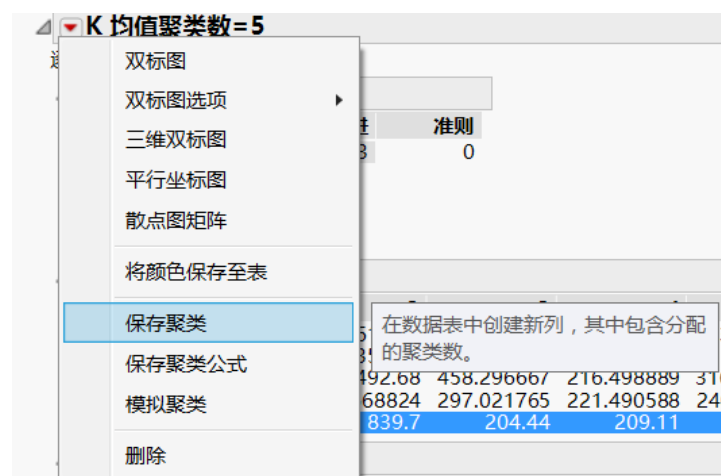


图 6.10

	region	x1	x2	x3	x4	x5	x6	x7	x8	聚类	距离
1	北京	2959.19	730.79	749.41	513.34	467.87	1141.82	478.42	457.64	1	3.3310016684
2	天津	2459.77	495.47	697.33	302.87	284.19	735.97	570.84	305.08	3	4.224628795
3	河北	1495.63	515.9	362.37	285.32	272.95	540.58	364.91	188.63	4	1.2152758591
4	山西	1406.33	477.77	290.15	208.57	201.5	414.72	281.84	212.1	4	0.5352351525
5	内蒙古	1303.97	524.29	254.83	192.17	249.81	463.09	287.87	192.96	4	0.8907660902
6	辽宁	1730.84	553.9	246.91	279.81	239.18	445.2	330.24	163.86	4	1.4251069683
7	吉林	1561.86	492.42	200.49	218.36	220.69	459.62	360.48	147.76	4	0.6738539832
8	黑龙江	1410.11	510.71	211.88	277.11	224.65	376.82	317.61	152.85	4	1.4089123782
9	上海	3712.31	550.74	893.37	346.93	527	1034.98	720.33	462.03	1	3.7083290486
10	江苏	2207.58	449.37	572.4	211.92	302.09	585.23	429.77	252.54	3	0.6548110649
11	浙江	2629.16	557.32	689.73	435.69	514.66	795.87	575.76	323.36	1	3.2407174603
12	安徽	1844.78	430.29	271.28	126.33	250.56	513.18	314	151.39	4	1.7984109689
13	福建	2709.46	428.11	334.12	160.77	405.14	461.67	535.13	232.29	3	3.3624103538
14	江西	1563.78	303.65	233.81	107.9	209.7	393.99	509.39	160.12	4	4.6007891608
15	山东	1675.75	613.32	550.71	219.79	272.59	599.43	371.62	211.84	3	2.6282940475
16	河南	1427.65	431.79	288.55	208.14	217	337.76	421.31	165.32	4	0.8642138478
17	湖北	1783.43	511.88	282.84	201.01	237.6	617.74	523.52	182.52	4	2.1630655031
18	湖南	1942.23	512.27	401.39	206.06	321.29	697.22	492.6	226.45	3	0.529674276
19	广东	3055.17	353.23	564.56	356.27	811.88	873.06	1082.82	420.81	2	0
20	广西	2033.87	300.82	338.65	157.78	329.06	621.74	587.02	218.27	3	4.0660247096
21	海南	2057.86	186.44	202.72	171.79	329.65	477.17	312.93	279.19	4	7.677130198
22	重庆	2303.29	589.99	516.21	236.55	403.92	730.05	438.41	225.8	3	1.7783045239
23	四川	1974.28	507.76	344.79	203.21	240.24	575.1	430.36	223.46	3	1.0933607546
24	贵州	1673.82	437.75	461.61	153.32	254.66	445.59	346.11	191.48	4	1.5206129761
25	云南	2194.25	537.01	369.07	249.54	290.84	561.91	407.7	330.95	3	1.7144017801
26	西藏	2646.61	839.7	204.44	209.11	379.3	371.04	269.59	389.33	5	0
27	陕西	1472.95	390.89	447.95	259.51	230.61	490.9	469.1	191.34	4	1.7241270089
28	甘肃	1525.57	472.98	328.9	219.86	206.65	449.69	249.66	228.19	4	0.6756931422
29	青海	1654.69	437.77	258.78	303	244.93	479.53	288.56	236.51	4	1.3710894906
30	宁夏	1375.46	480.89	273.84	317.32	251.08	424.75	228.73	195.93	4	1.997848419
31	新疆	1608.82	536.05	432.46	235.82	250.28	541.3	344.85	214.4	4	1.3204441044

图 6.11

第七章 主成分分析

一、对书中例 7.3.3 作主成分分析

打开 examp6.3.3.xlsx 数据表 \Rightarrow 分析 \Rightarrow 多元方法 \Rightarrow 主成分，即出现“主成分”对话框（见图 7.1） \Rightarrow 选择 x1,x2,x3,x4,x5,x6,x7,x8 并点击 Y, 列 \Rightarrow 确定 \Rightarrow 从“主成分：基于相关性”相应的红色小三角菜单中（见图 7.2），作图中选择；在选择得分图时，弹出如图 7.3 所示的框，输入 2 \Rightarrow 确定，即生成图 7.4。

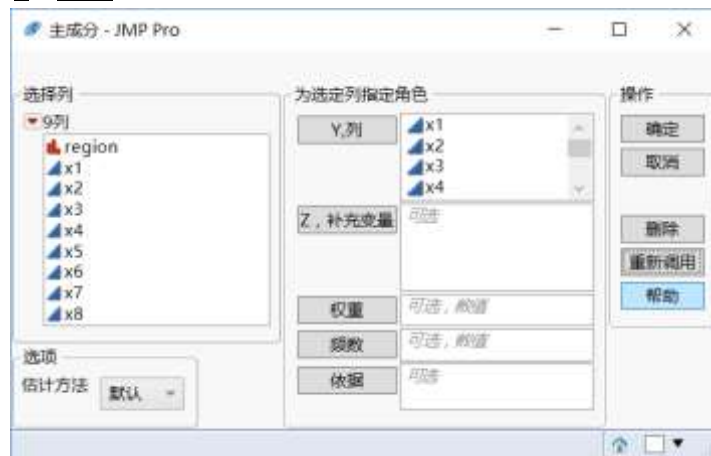


图 7.1



图 7.4

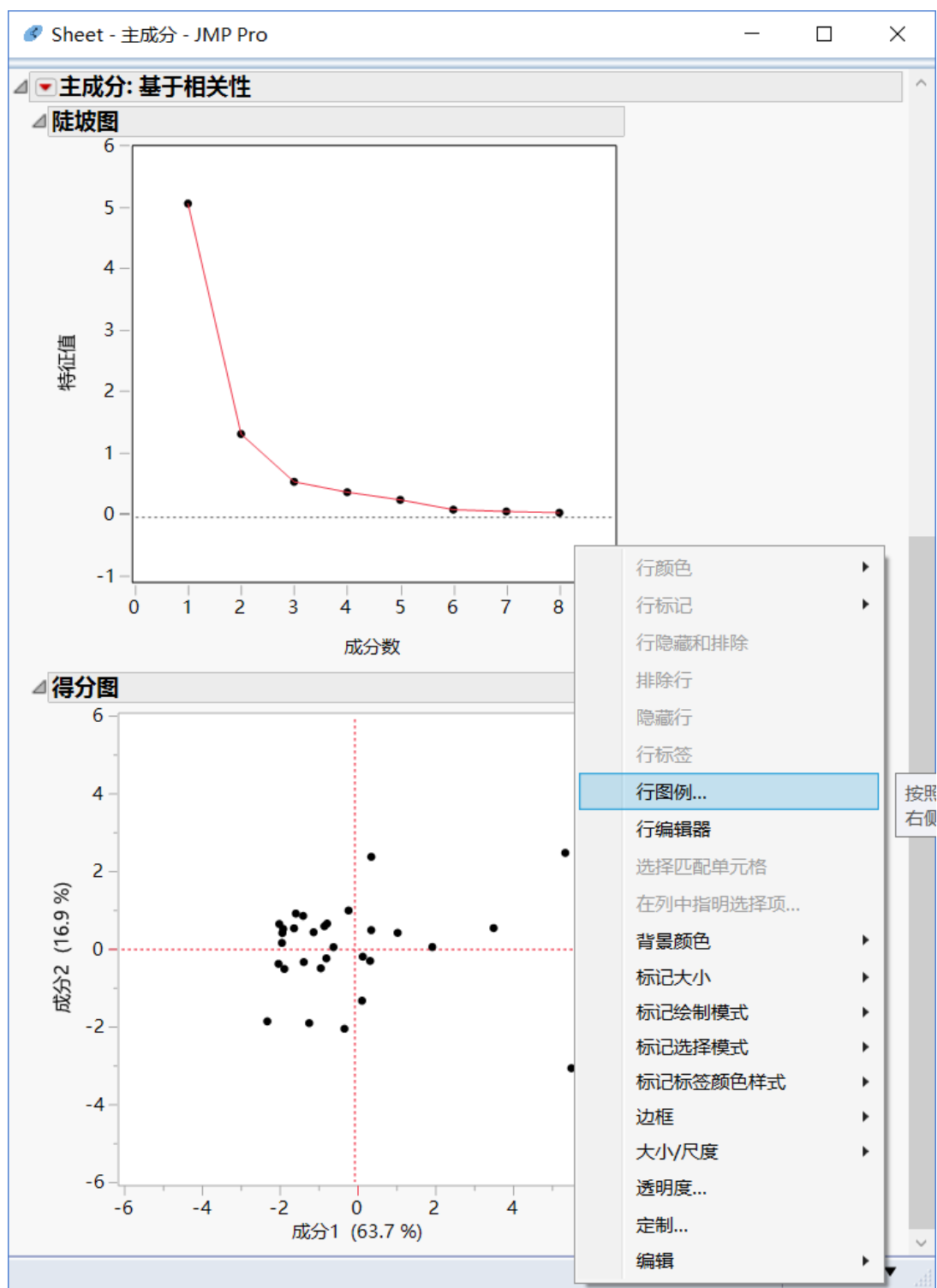


图 7.4 (续)

在得分图框内点击右键, 选择弹出菜单中的行图例..., 出现如图 7.5 所示的“按列标记”对话框 \Rightarrow 在变量框中选择 region, 在“标记”列表框中选择字母数字 \Rightarrow 确定, 即出现图 7.6 中标记了的得分图。

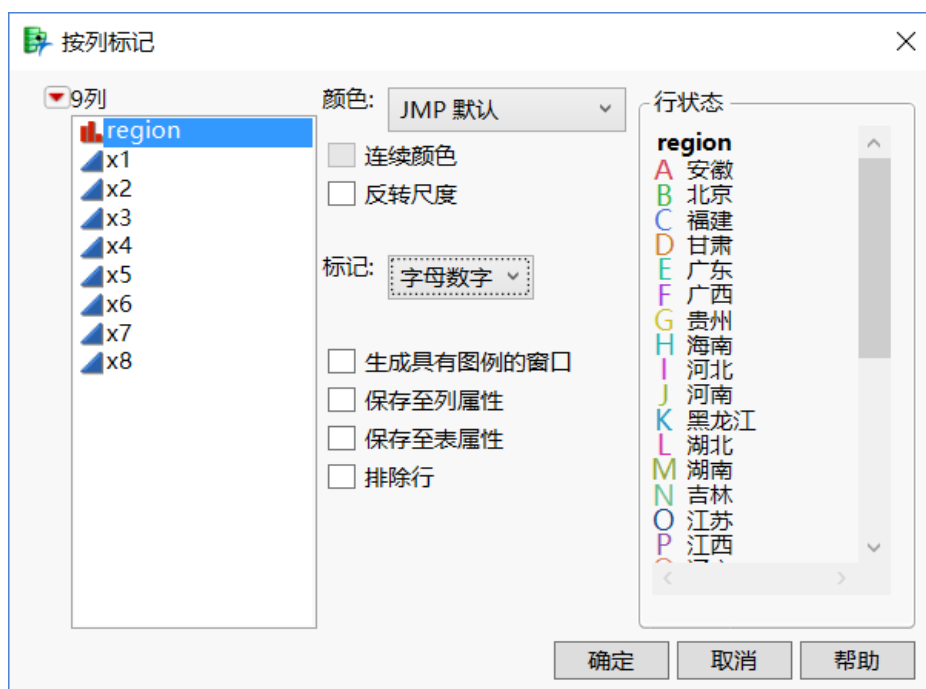


图 7.5

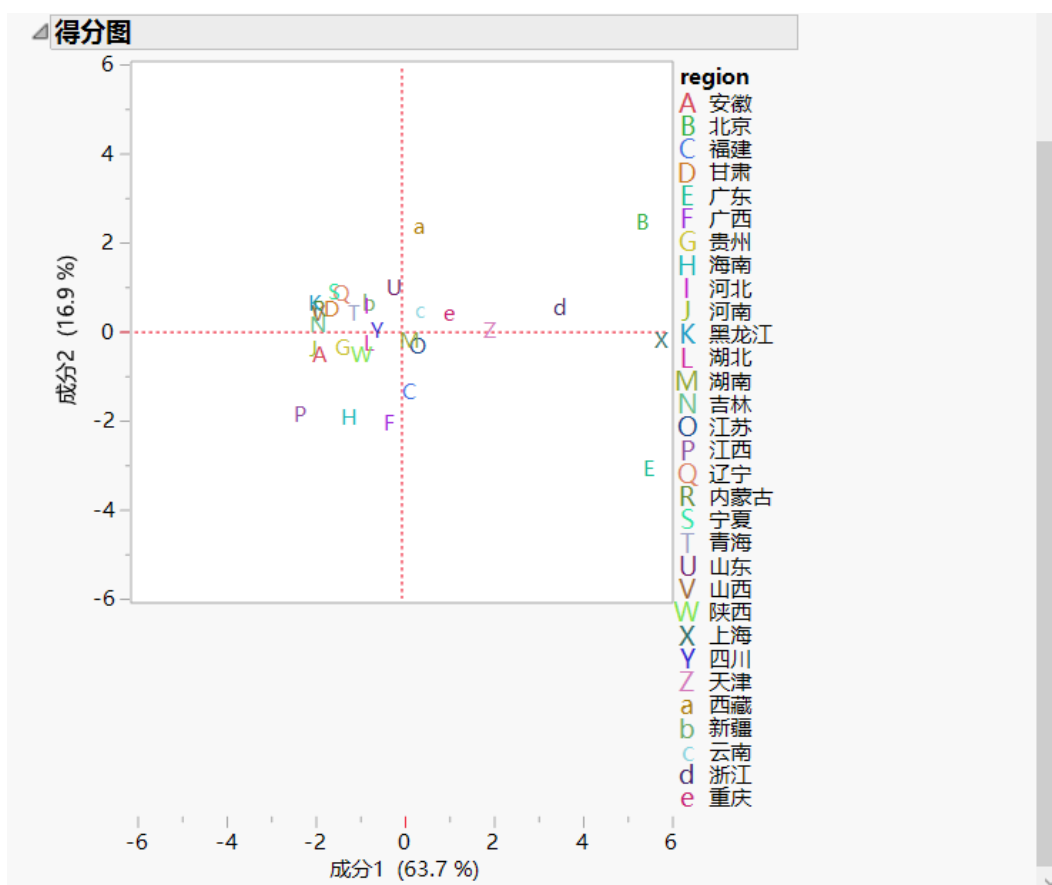



图 7.6

注：如希望调整得分图中字母的大小，则可在图 7.4（续）的菜单中点击标记大小，从中选择所需要的大小。此外，通过拉得分图的边框还可以将图形拉大，以使图形显示得更清楚。

在图 7.2 中点击保存主成分，弹出如图 7.7 所示的框⇒输入 2⇒确定，前两个主成分得分随即出现在原数据表中，见图 7.8。


请输入数值
×

指定要保存的成分数:

确定

取消

图 7.7

		region	x1	x2	x3	x4	x5	x6	x7	x8	主成分1	主成分2
B	1	北京	2959.19	730.79	749.41	513.34	467.87	1141.82	478.42	457.64	5.426442824	2.4666100706
Z	2	天津	2459.77	495.47	697.33	302.87	284.19	735.97	570.84	305.08	2.0064021721	0.044897784
I	3	河北	1495.63	515.9	362.37	285.32	272.95	540.58	364.91	188.63	-0.769577208	0.5804860187
V	4	山西	1406.33	477.77	290.15	208.57	201.5	414.72	281.84	212.1	-1.848721831	0.404449961
R	5	内蒙古	1303.97	524.29	254.83	192.17	249.81	463.09	287.87	192.96	-1.826726062	0.5099104145
Q	6	辽宁	1730.84	553.9	246.91	279.81	239.18	445.2	330.24	163.86	-1.313555872	0.8448344764
N	7	吉林	1561.86	492.42	200.49	218.36	220.69	459.62	360.48	147.76	-1.859765766	0.1513591305
K	8	黑龙江	1410.11	510.71	211.88	277.11	224.65	376.82	317.61	152.85	-1.927574639	0.6366917161
X	9	上海	3712.31	550.74	893.37	346.93	527	1034.98	720.33	462.03	5.8665800816	-0.195590149
O	10	江苏	2207.58	449.37	572.4	211.92	302.09	585.23	429.77	252.54	0.4072064428	-0.311959647
d	11	浙江	2629.16	557.32	689.73	435.69	514.66	795.87	575.76	323.36	3.5839360229	0.5318411174
A	12	安徽	1844.78	430.29	271.28	126.33	250.56	513.18	314	151.39	-1.796734483	-0.519304406
C	13	福建	2709.46	428.11	334.12	160.77	405.14	461.67	535.13	232.29	0.2011572053	-1.337538882
P	14	江西	1563.78	303.65	233.81	107.9	209.7	393.99	509.39	160.12	-2.234432278	-1.867172919
U	15	山东	1675.75	613.32	550.71	219.79	272.59	599.43	371.62	211.84	-0.147463977	0.9838427967
J	16	河南	1427.65	431.79	288.55	208.14	217	337.76	421.31	165.32	-1.947265673	-0.388124368
L	17	湖北	1783.43	511.88	282.84	201.01	237.6	617.74	523.52	182.52	-0.717010904	-0.247235911
M	18	湖南	1942.23	512.27	401.39	206.06	321.29	697.22	492.6	226.45	0.2190214152	-0.203553132
E	19	广东	3055.17	353.23	564.56	356.27	811.88	873.06	1082.82	420.81	5.5835416167	-3.071999293
F	20	广西	2033.87	300.82	338.65	157.78	329.06	621.74	587.02	218.27	-0.251547629	-2.058477339
H	21	海南	2057.86	186.44	202.72	171.79	329.65	477.17	312.93	279.19	-1.15751891	-1.913072019
e	22	重庆	2303.29	589.99	516.21	236.55	403.92	730.05	438.41	225.8	1.1156184438	0.409970592
Y	23	四川	1974.28	507.76	344.79	203.21	240.24	575.1	430.36	223.46	-0.533650385	0.0417889163
G	24	贵州	1673.82	437.75	461.61	153.32	254.66	445.59	346.11	191.48	-1.298147689	-0.341976439
c	25	云南	2194.25	537.01	369.07	249.54	290.84	561.91	407.7	330.95	0.435733524	0.4790944646
a	26	西藏	2646.61	839.7	204.44	209.11	379.3	371.04	269.59	389.33	0.4373180482	2.3649994504
W	27	陕西	1472.95	390.89	447.95	259.51	230.61	490.9	469.1	191.34	-0.859474363	-0.501061268
D	28	甘肃	1525.57	472.98	328.9	219.86	206.65	449.69	249.66	228.19	-1.549423484	0.526219683
T	29	青海	1654.69	437.77	258.78	303	244.93	479.53	288.56	236.51	-1.045191936	0.4260940415
S	30	宁夏	1375.46	480.89	273.84	317.32	251.08	424.75	228.73	195.93	-1.501711016	0.906911419
b	31	新疆	1608.82	536.05	432.46	235.82	250.28	541.3	344.85	214.4	-0.697463691	0.6470637195

图 7.8

在图 7.2 中，如选择三维得分图，则出现图 7.9，可作三维旋转。当然，该例并不适合取第三主成分，这里只是作一展示。

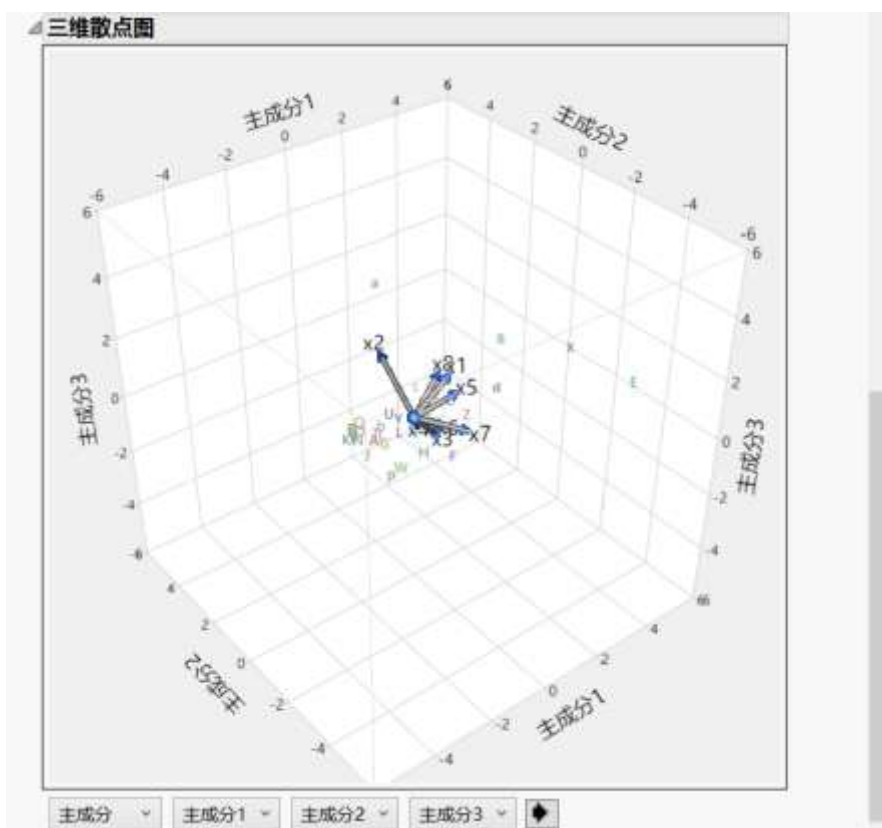


图 7.9

第八章 因子分析

一、对书中习题 6.5 中的数据作因子分析

打开 exec6.5.xlsx 数据表 ⇒ 分析 ⇒ 多元方法 ⇒ 主成分，即出现“主成分”对话框（见图 8.1）⇒ 选择 x1,x2,x3,x4,x5,x6,x7,x8 并点击 Y, 列 ⇒ 确定 ⇒ 从“主成分：基于相关性”相应的红色小三角菜单中（见图 8.2），选择特征值和因子分析，随即出现图 8.3。

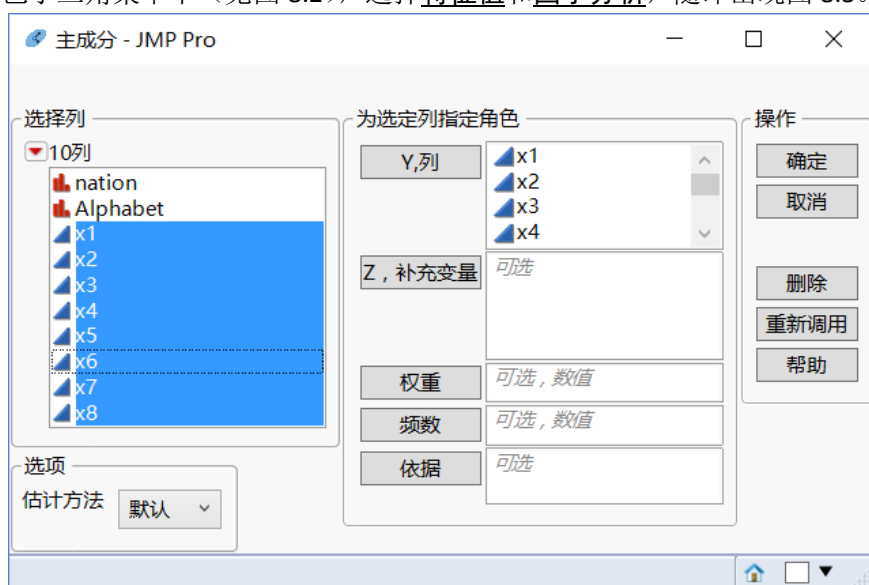


图 8.1



图 8.2

1.主成分法

选择因子分解及旋转方法

因子分解方法

☒ 主成分
☐ 最大似然

先验公因子方差

☒ 主成分 (对角线=1)
☐ 公因子分析 (对角线=SMC)

因子数

旋转方法

最大方差法

注意: 建议对直角旋转采用 Varimax 旋转, 对斜角旋转采用 Quartimin 旋转。

确定

取消

图 8.3

作图 8.3 中的选择 \Rightarrow 确定，在生成的窗口中展开所有按钮，如图 8.4 所示。



图 8.4



图 8.4 (续 1)

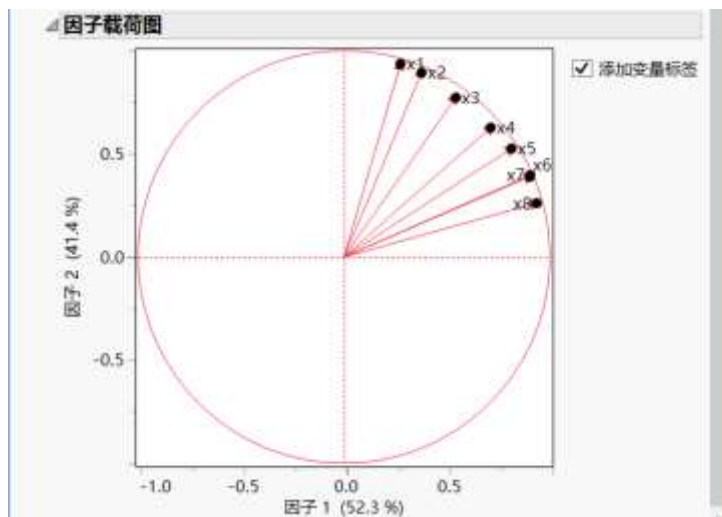


图 8.4 (续 2)

在图 8.5 中点击保存旋转成分，两个旋转因子得分即出现在原数据表中，见图 8.6。



图 8.5

	nation	Alphabet	x1	x2	x3	x4	x5	x6	x7	x8	因子1	因子2
1	阿根廷	a	10.39	20.81	46.84	1.81	3.7	14.04	29.36	137.72	0.3169772437	-0.213093462
2	澳大利亚	a	10.31	20.06	44.84	1.74	3.57	13.28	27.66	128.3	-0.571156834	-0.79421059
3	奥地利	a	10.44	20.81	46.82	1.79	3.6	13.26	27.72	135.9	-0.576639622	0.1900216845
4	比利时	b	10.34	20.68	45.04	1.73	3.6	13.22	27.45	129.95	-0.78405685	-0.304385999
5	百慕大	b	10.28	20.58	45.91	1.8	3.75	14.68	30.55	146.62	1.4463107841	-1.244244271
6	巴西	b	10.22	20.43	45.21	1.73	3.66	13.62	28.62	133.13	-0.013822046	-0.913721148
7	缅甸	b	10.64	21.52	48.3	1.8	3.85	14.45	30.28	139.95	0.4027017637	0.7073693832
8	加拿大	c	10.17	20.22	45.68	1.76	3.63	13.55	28.09	130.15	-0.167174377	-0.847257113
9	智利	c	10.34	20.8	46.2	1.79	3.71	13.61	29.3	134.03	0.0276323912	-0.259595695
10	中国	c	10.51	21.04	47.3	1.81	3.73	13.9	29.13	133.53	-0.128108347	0.3933185979
11	哥伦比亚	c	10.43	21.05	46.1	1.82	3.74	13.49	27.88	131.35	-0.462596428	0.3063004836
12	库克群岛	c	12.18	23.2	52.94	2.02	4.24	16.7	35.38	164.7	2.063991826	3.8934948444
13	哥斯达黎加	c	10.94	21.9	48.66	1.87	3.84	14.03	28.81	136.58	-0.484088861	1.9345311701
14	捷克斯洛伐克	c	10.35	20.65	45.64	1.76	3.58	13.42	28.19	134.32	-0.384754699	-0.370667326
15	丹麦	d	10.56	20.52	45.89	1.78	3.61	13.5	28.11	130.78	-0.597530171	0.0320141749
16	多米尼加	d	10.14	20.65	46.8	1.82	3.82	14.91	31.45	154.12	2.2081659239	-1.549856477
17	芬兰	f	10.43	20.69	45.49	1.74	3.61	13.27	27.52	130.87	-0.783172311	-0.097026388
18	法国	f	10.11	20.38	45.28	1.73	3.57	13.34	27.97	132.3	-0.29089558	-0.957014626
19	德意志民主共和国	g	10.12	20.33	44.87	1.73	3.56	13.17	27.42	129.92	-0.54784803	-0.907131069
20	德意志联邦共和国	g	10.16	20.37	44.5	1.73	3.53	13.21	27.61	132.23	-0.466826307	-0.979258927

图 8.6

2.主因子法

将图 8.3 “先验公因子方差” 中的选择改为公因子分析（对角线=SMC） ⇒ 确定 ⇒ 在随即出现的窗口中点击“主成分：基于相关性”旁的红色小三角，在菜单中撤销特征值的选项，将窗口中所有的展开按钮处于展开状态，产生图 8.7。

Sheet - 主成分 - JMP Pro

主成分：基于相关性

因子分析：主成分/ 最大方差法

先验公因子方差估计值: SMC

x1	0.87726
x2	0.88802
x3	0.84494
x4	0.88368
x5	0.92716
x6	0.95500
x7	0.96676
x8	0.90472

简化相关性矩阵的特征值

数目	特征值	百分比	累积百分比
1	6.5302	90.102	90.102
2	0.7778	10.732	100.834
3	0.0506	0.698	101.532
4	0.0061	0.084	101.617
5	-0.0135	-0.187	101.430
6	-0.0149	-0.205	101.225
7	-0.0356	-0.491	100.734
8	-0.0532	-0.734	100.000

根据“因子数准则”会保留 2 个因子。

未旋转的因子载荷

	因子 1	因子 2
x1	0.806881	0.496025
x2	0.857850	0.412118
x3	0.899705	0.216008
x4	0.938577	0.023869
x5	0.955676	-0.114320
x6	0.938318	-0.282122
x7	0.946168	-0.281197
x8	0.873969	-0.378138

图 8.7

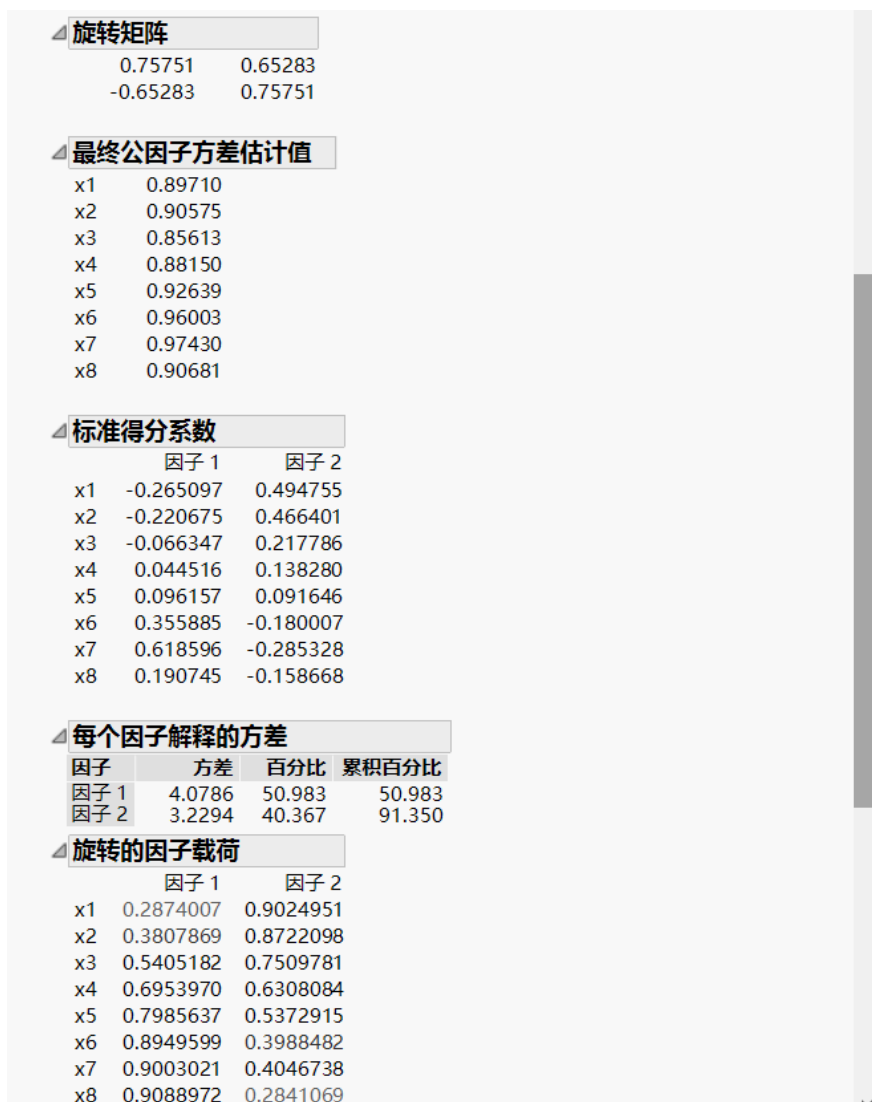


图 8.7 (续 1)

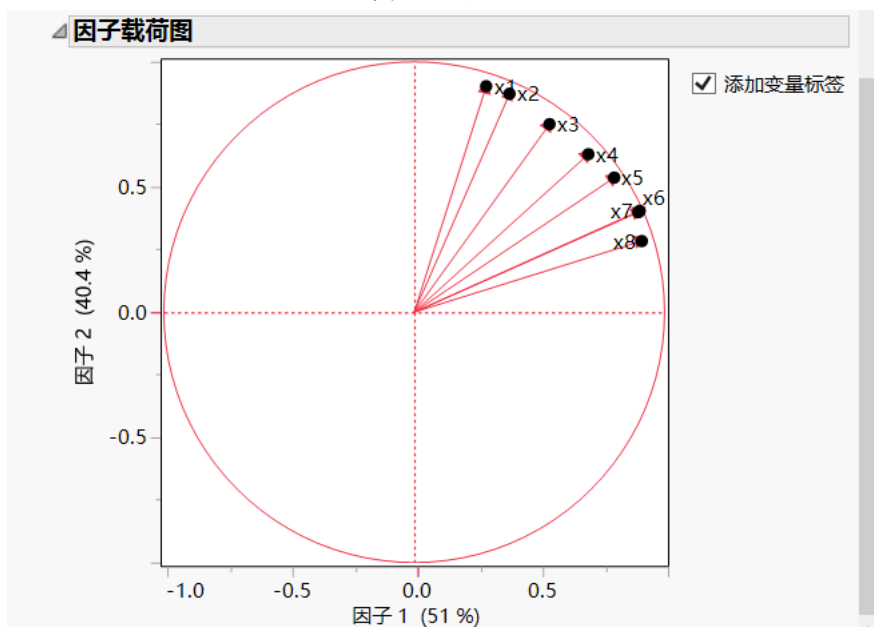


图 8.7 (续 2)

3. 极大似然法

将图 8.3 “因子分解方法” 中的选择改为最大似然，“先验公因子方差” 中的选择仍为主成分（对角线=1） [也可选择为公因子分析（对角线=SMC）] ⇒ 确定，在生成的窗口中将所有的展开按钮处于展开状态，生成图 8.8。

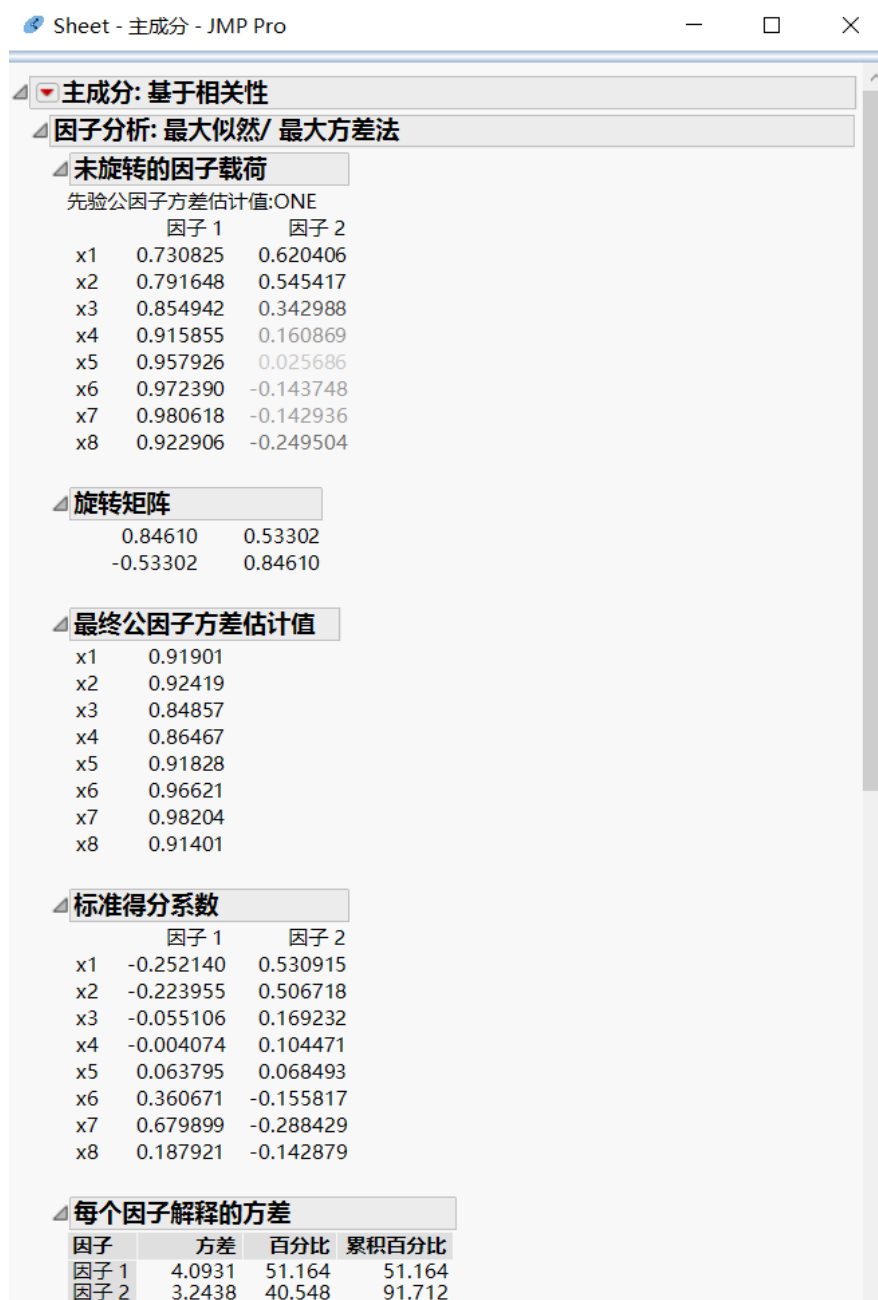


图 8.8

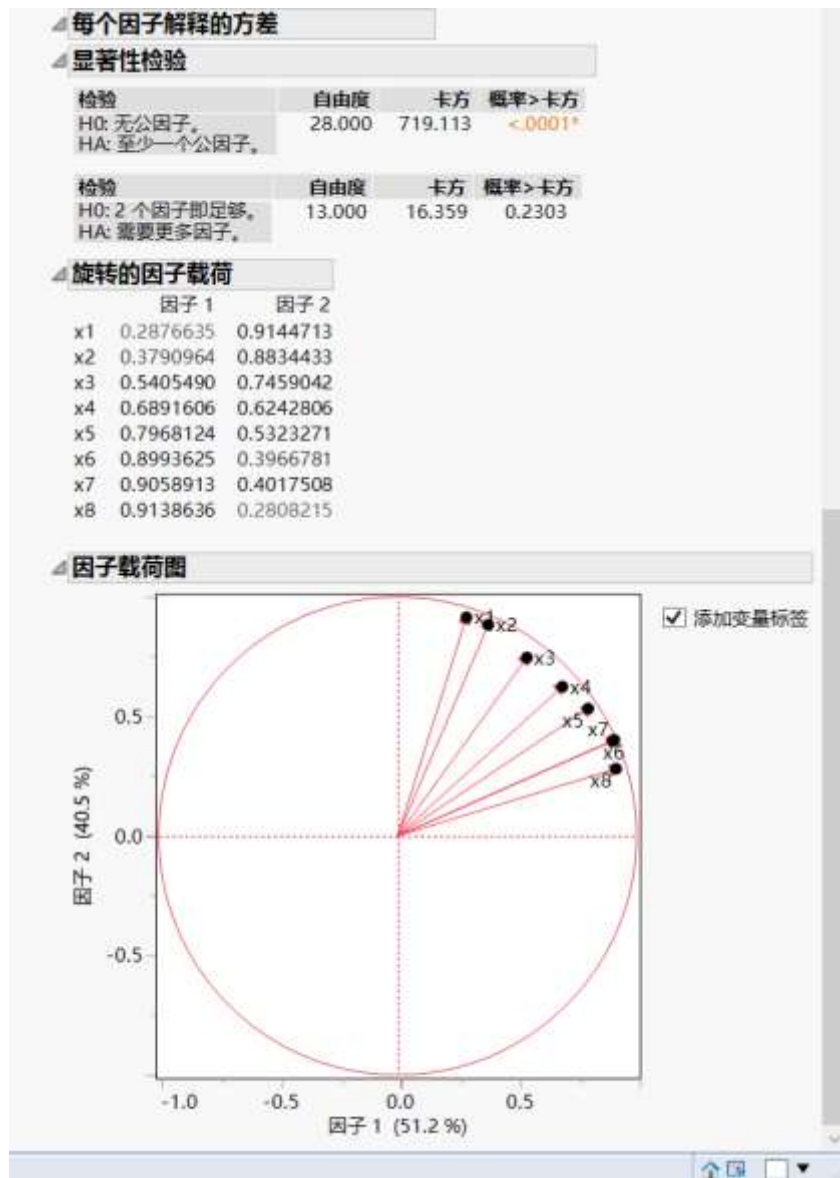


图 8.8 (续)

第九章 对应分析

一、对书中例 9.2.1 中的数据作对应分析

打开 examp9.2.1.xlsx 数据表 \Rightarrow 类似于图 4.4 的操作, 将变量“心理健康状况”设置为有序型(或名义型) \Rightarrow 分析 \Rightarrow 以 X 拟合 Y, 即出现“以 X 拟合 Y”对话框(见图 9.1) \Rightarrow 选择父母社会经济地位 \rightarrow Y, 响应: 选择心理健康状况 \rightarrow X, 因子 \Rightarrow 确定 \Rightarrow 在随即出现的图 9.2 中, 点击“列联分析”旁的红色小三角, 作图中的选择, 即出现图 9.3。



图 9.1

“心理健康状况-父母社会经济地位”列联分析

	E	合计
马森克图	36	21
列联表	2.17	1.27
检验	3.58	9.68
设置 α 水平	1.73	6.84
对应分析	97	71
Cochran Mantel Haenszel 检验	8.94	4.78
关联度量	36.27	26.37
精确检验	0.11	11.19
显示选项	54	54
制成数据表	3.25	3.25
本地数据过滤器	0.38	24.88
恢复	8.92	14.92
保存脚本	78	71
	8.70	4.28
	23.43	

图 9.2

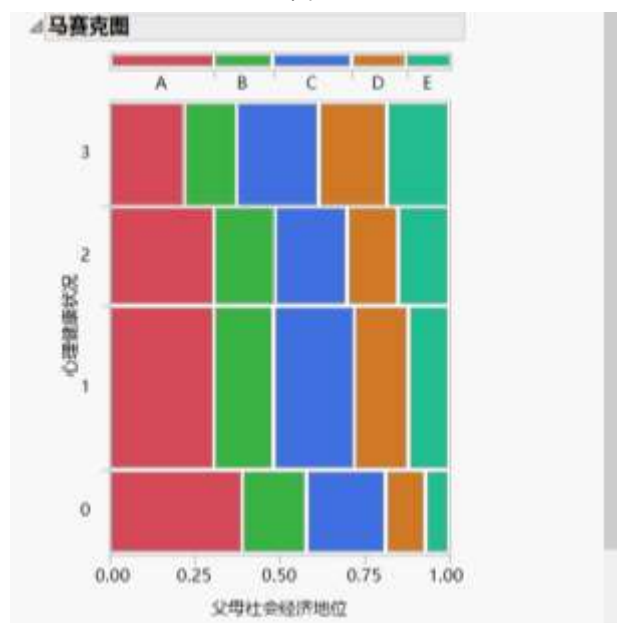


图 9.3

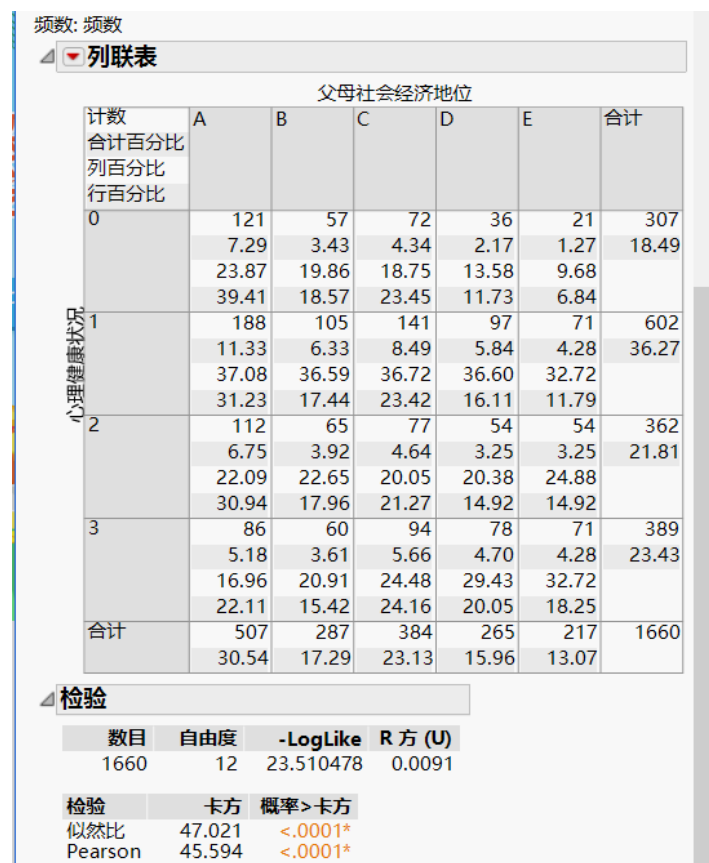


图 9.3 (续 1)

注：列联表格子中的第 2、3、4 行的单位为%，其第 2 行元素（除最后一行格子和最后一列格子外）构成对应矩阵，其最后一行格子的第 2 行元素为列边缘频率，最后一列格子的第 2 行元素为行边缘频率。格子中第 3 行元素构成列轮廓矩阵，第 4 行元素构成行轮廓矩阵。

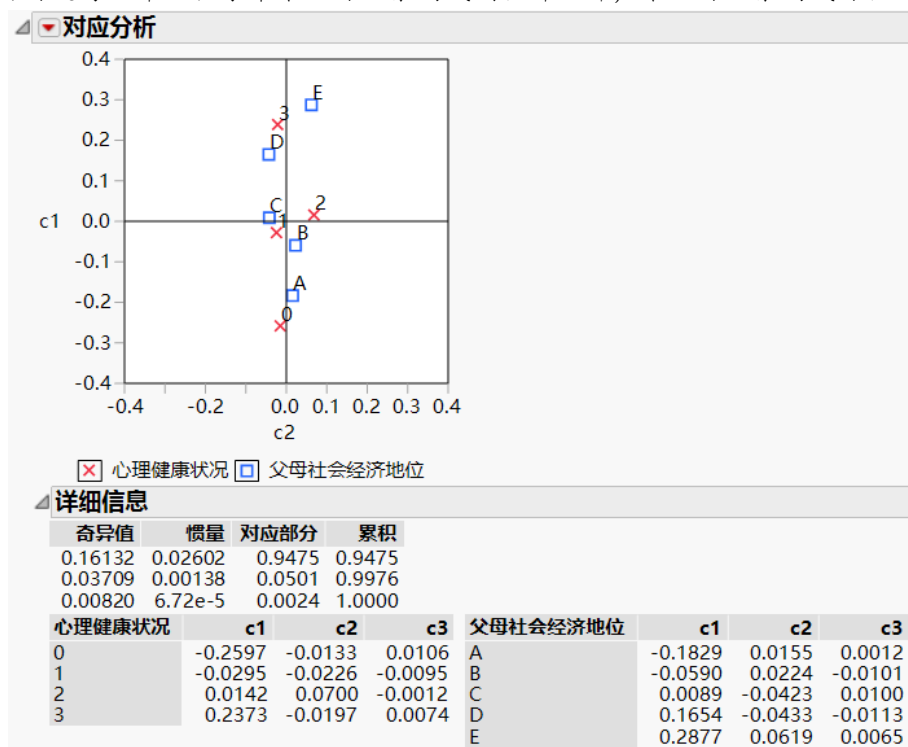


图 9.3 (续 2)

注：图中的二维对应分析图可以拉大，以便看得更清楚。

若希望得到三维对应分析图（本例无此必要），则从“对应分析”相应的红色小三角菜单，选择三维对应分析，出现图 9.4，该图可进行三维旋转。

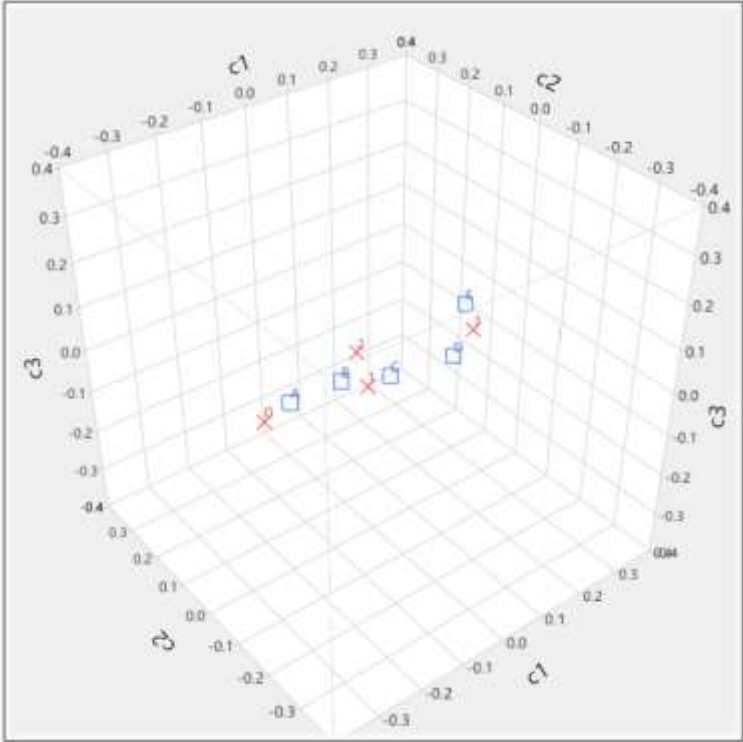


图 9.4

在图 9.3（续 1）中，点击“列联表”旁的红色小三角，作图 9.5 中的选择，生成图 9.6。

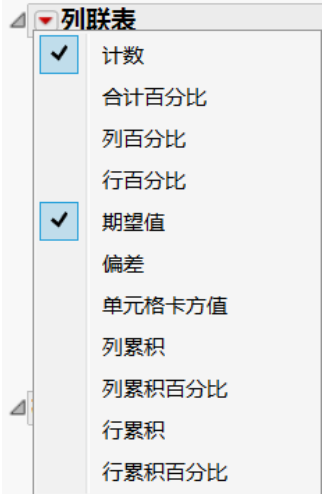


图 9.5

		父母社会经济地位					
心理健康状况	计数	A	B	C	D	E	合计
	期望值						
	0	121	57	72	36	21	307
		93.7645	53.0777	71.0169	49.009	40.1319	
	1	188	105	141	97	71	602
		183.864	104.081	139.258	96.1024	78.6952	
	2	112	65	77	54	54	362
		110.563	62.5867	83.7398	57.7892	47.3217	
3		86	60	94	78	71	389
		118.809	67.2548	89.9855	62.0994	50.8512	
合计		507	287	384	265	217	1660

图 9.6

作图 9.7 中的操作 ⇒ 确定，产生另一马赛克图，如图 9.8 所示。



图 9.7

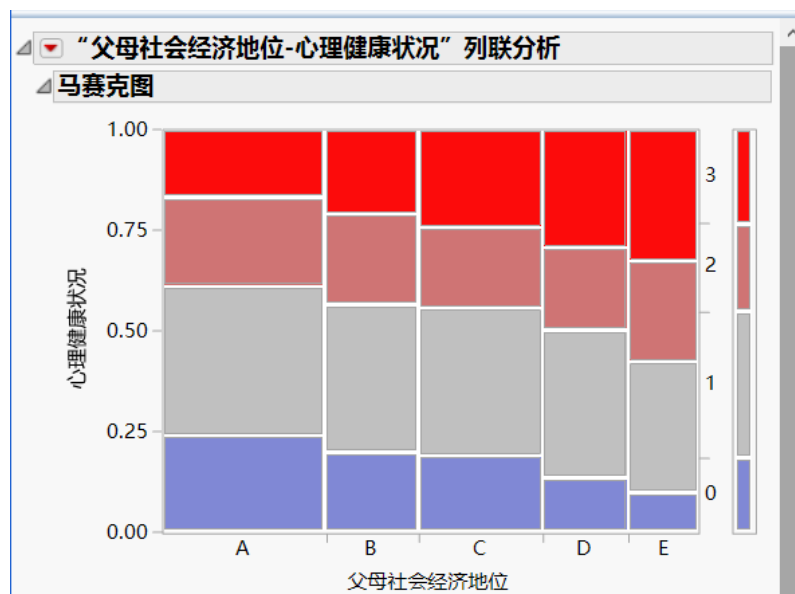


图 9.8