

《应用多元统计分析》（第五版）

SPSS24 的应用

王学民

说明：本文主要侧重 SPSS24 的菜单操作，对其许多输出结果，读者可参见《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，并可通过对照书中的 SAS 输出结果理解本文的 SPSS 输出结果。

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》SPSS 数据”文件夹，本文均从该文件夹中打开数据表。

在 SPSS 的数据编辑器窗口中，选择文件⇒打开⇒数据...，即出现如图 1 所示的“打开数据”窗口⇒在“查找范围”框中选择“《应用多元统计分析》SPSS 数据”所在的文件夹。

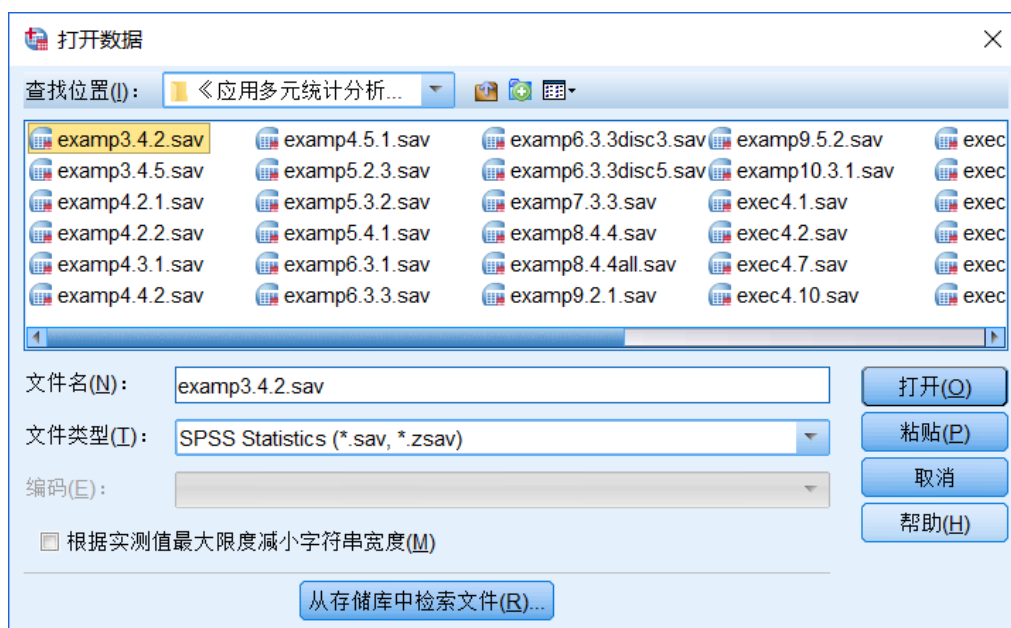


图 1

第三章 多元正态分布

一、对例 3.4.2 进行相关分析等

在图 1 的列表框中双击 examp3.4.2.sav 数据表，随即出现“SPSS 数据编辑器”窗口（见图 3.1）⇒分析⇒相关⇒双变量...⇒在出现的“双变量相关性”对话框中（见图 3.2），将年龄[x1]、体重[x2]、肺活量[x3]、1.5 英里跑的时间[x4]、休息时的脉搏[x5]、跑步时的脉搏[x6]和跑步时的最大脉搏[x7]选入“变量”列表框⇒选项...⇒作图 3.3 中的选项⇒

继续⇒确定，输出结果如图 3.4 所示。

	x1	x2	x3	x4	x5	x6	x7	变量
1	44.0	89.47	44.609	11.37	62.0	178.0	182.0	
2	40.0	75.07	45.313	10.07	62.0	185.0	185.0	
3	44.0	85.84	54.297	8.65	45.0	156.0	168.0	
4	42.0	68.15	59.571	8.17	40.0	166.0	172.0	
5	38.0	89.02	49.874	9.22	55.0	178.0	180.0	
6	47.0	77.45	44.811	11.63	58.0	176.0	176.0	
7	40.0	75.98	45.681	11.95	70.0	176.0	180.0	
8	43.0	81.19	49.091	10.85	64.0	162.0	170.0	
9	44.0	81.42	39.442	13.08	63.0	174.0	176.0	
10	38.0	81.87	60.055	8.63	48.0	170.0	186.0	
11	44.0	73.03	50.541	10.13	45.0	168.0	168.0	
12	45.0	87.66	37.388	14.03	56.0	186.0	192.0	
13	45.0	66.45	44.754	11.12	51.0	176.0	176.0	
14	47.0	79.15	47.273	10.60	47.0	162.0	164.0	
15	54.0	83.12	51.855	10.33	50.0	166.0	170.0	
16	49.0	81.42	49.156	8.95	44.0	180.0	185.0	
17	51.0	69.63	40.836	10.95	57.0	168.0	172.0	
18	51.0	77.04	46.872	10.00	48.0	162.0	168.0	

图 3.1

双变量相关性

变量(V):

- 年龄 [x1]
- 体重 [x2]
- 肺活量 [x3]
- 1.5英里跑的时...
- 休息时的脉搏 [x5]
- 跑步时的脉搏 [x6]
- 跑步时的最大脉...

相关系数

☒ 皮尔逊(N) ☐ 肯德尔 tau-b(K) ☐ 斯皮尔曼(S)

显著性检验

☒ 双尾(I) ☐ 单尾(L)

☒ 标记显著性相关性(E)

确定 粘贴(P) 重置(R) 取消 帮助

图 3.2

双变量相关性：选项

统计

☒ 平均值和标准差(M)

☒ 叉积偏差和协方差(C)

缺失值

☒ 成对排除个案(P)

☐ 成列排除个案(L)

继续(C) 取消 帮助

图 3.3

描述统计			
	平均值	标准差	个案数
年龄	47.677	5.2114	31
体重	77.4445	8.32857	31
肺活量	47.37581	5.327231	31
1.5英里跑的时间	10.5861	1.38741	31
休息时的脉搏	53.742	8.2944	31
跑步时的脉搏	169.645	10.2520	31
跑步时的最大脉搏	173.774	9.1641	31

相关性								
		年龄	体重	肺活量	1.5英里跑的时间	休息时的脉搏	跑步时的脉搏	跑步时的最大脉搏
年龄	皮尔逊相关性	1	-.234	-.305	.189	-.142	-.338	-.433 [*]
	显著性 (双尾)		.206	.096	.309	.447	.063	.015
	平方和与叉积	814.774	-304.095	-253.688	40.941	-183.581	-541.548	-620.258
	协方差	27.159	-10.136	-8.456	1.365	-6.119	-18.052	-20.675
	个案数	31	31	31	31	31	31	31
体重	皮尔逊相关性	-.234	1	-.163	.144	.023	.182	.249
	显著性 (双尾)	.206		.382	.441	.904	.328	.176
	平方和与叉积	-304.095	2080.951	-216.632	49.748	47.046	464.960	571.012
	协方差	-10.136	69.365	-7.221	1.658	1.568	15.499	19.034
	个案数	31	31	31	31	31	31	31
肺活量	皮尔逊相关性	-.305	-.163	1	-.862 ^{**}	-.346	-.398 [*]	-.237
	显著性 (双尾)	.096	.382		.000	.056	.027	.200
	平方和与叉积	-253.688	-216.632	851.382	-191.176	-459.193	-652.057	-346.724
	协方差	-8.456	-7.221	28.379	-6.373	-15.306	-21.735	-11.557
	个案数	31	31	31	31	31	31	31
1.5英里跑的时间	皮尔逊相关性	.189	.144	-.862 ^{**}	1	.401 [*]	.314	.226
	显著性 (双尾)	.309	.441	.000		.026	.086	.221
	平方和与叉积	40.941	49.748	-191.176	57.748	138.279	133.837	86.243
	协方差	1.365	1.658	-6.373	1.925	4.609	4.461	2.875
	个案数	31	31	31	31	31	31	31
休息时的脉搏	皮尔逊相关性	-.142	.023	-.346	.401 [*]	1	.318	.258
	显著性 (双尾)	.447	.904	.056	.026		.081	.162
	平方和与叉积	-183.581	47.046	-459.193	138.279	2063.935	811.161	587.194
	协方差	-6.119	1.568	-15.306	4.609	68.798	27.039	19.573
	个案数	31	31	31	31	31	31	31
跑步时的脉搏	皮尔逊相关性	-.338	.182	-.398 [*]	.314	.318	1	.930 ^{**}
	显著性 (双尾)	.063	.328	.027	.086	.081		.000
	平方和与叉积	-541.548	464.960	-652.057	133.837	811.161	3153.097	2620.516
	协方差	-18.052	15.499	-21.735	4.461	27.039	105.103	87.351
	个案数	31	31	31	31	31	31	31
跑步时的最大脉搏	皮尔逊相关性	-.433 [*]	.249	-.237	.226	.258	.930 ^{**}	1
	显著性 (双尾)	.015	.176	.200	.221	.162	.000	
	平方和与叉积	-620.258	571.012	-346.724	86.243	587.194	2620.516	2519.419
	协方差	-20.675	19.034	-11.557	2.875	19.573	87.351	83.981
	个案数	31	31	31	31	31	31	31

图 3.4

选择分析 ⇒ 相关 ⇒ 偏相关... ⇒ 在“偏相关”对话框 (见图 3.5) 中, 将肺活量[x3]、1.5 英里跑的时间[x4]、休息时的脉搏[x5]、跑步时的脉搏[x6]和跑步时的最大脉搏[x7]选入“变量”列表框, 将年龄[x1]、体重[x2]选入“控制”列表框 ⇒ 确定, 得到图 3.6。

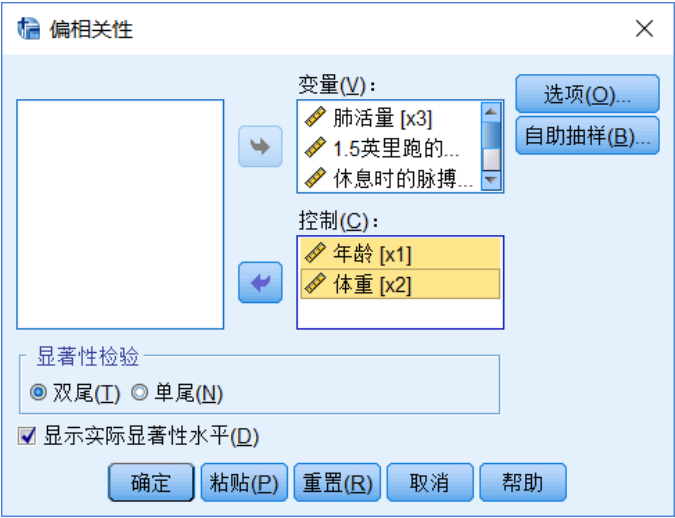


图 3.5

相关性			肺活量	1.5英里跑的时间	休息时的脉搏	跑步时的脉搏	跑步时的最大脉搏
控制变量			肺活量	1.5英里跑的时间	休息时的脉搏	跑步时的脉搏	跑步时的最大脉搏
年龄 & 体重	肺活量	相关性	1.000	-.855	-.430	-.552	-.405
		显著性 (双尾)	.	.000	.020	.002	.029
		自由度	0	27	27	27	27
	1.5英里跑的时间	相关性	-.855	1.000	.450	.396	.325
		显著性 (双尾)	.000	.	.014	.033	.085
		自由度	27	0	27	27	27
	休息时的脉搏	相关性	-.430	.450	1.000	.293	.225
		显著性 (双尾)	.020	.014	.	.123	.241
		自由度	27	27	0	27	27
	跑步时的脉搏	相关性	-.552	.396	.293	1.000	.924
		显著性 (双尾)	.002	.033	.123	.	.000
		自由度	27	27	27	0	27
	跑步时的最大脉搏	相关性	-.405	.325	.225	.924	1.000
		显著性 (双尾)	.029	.085	.241	.000	.
		自由度	27	27	27	27	0

图 3.6

二、对例 3.4.2 数据表作散点图矩阵

选择图形⇒旧对话框⇒散点图/点图...⇒在出现的如图 3.7 所示的“散点图/点图”对话框中，选择矩阵散点图⇒定义，随即出现“散点图矩阵”对话框（见图 3.8）⇒将年龄[x1]、体重[x2]、肺活量[x3]、1.5 英里跑的时间[x4]、休息时的脉搏[x5]、跑步时的脉搏[x6]和跑步时的最大脉搏[x7] 选入到“矩阵变量”列表框中⇒确定，生成如图 3.9 所示的散点图矩阵。



图 3.7

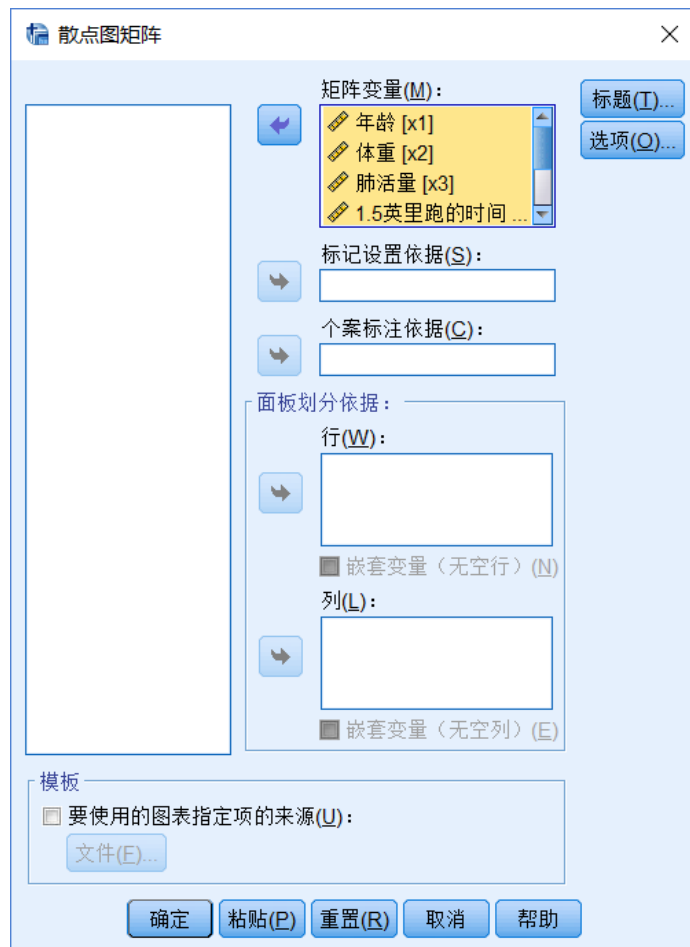


图 3.8

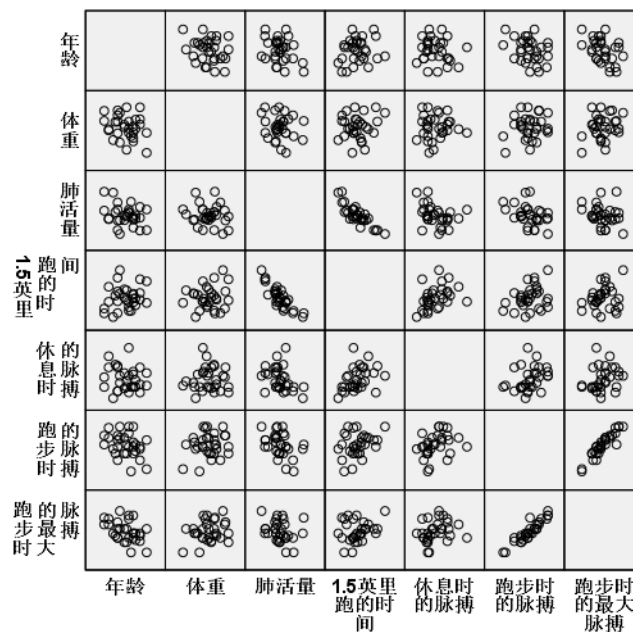


图 3.9

在图 3.9 中双击以激活该图，出现“图表编辑器”窗口，如图 3.10 所示 ⇒ 选项 ⇒ 显示沿对角线绘制的图表，即出现该散点图矩阵对角线上的直方图。

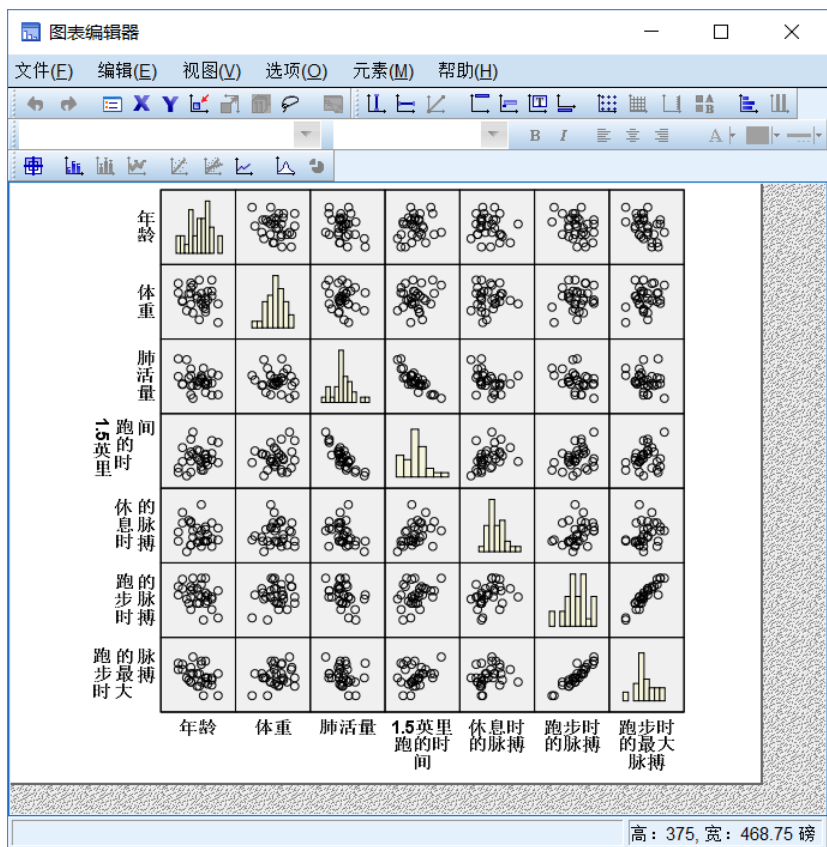


图 3.10

三、对例 3.4.2 数据表作三维散点图

在图 3.7 的对话框中，选择**三维散点图**⇒**定义**⇒在出现“三维散点图”对话框中（见图 3.11），将**年龄[x1]**、**体重[x2]**和**肺活量[x3]**分别选入到“**Y轴**”、“**X轴**”和“**Z轴**”框中⇒**确定**，随即生成三维散点图（见图 3.12）⇒在图 3.12 中双击以激活该图，出现“图表编辑器”窗口，如图 3.13 所示⇒**编辑**⇒**三维旋转**，即弹出一对话框（见图 3.14），此时“图表编辑器”窗口中的鼠标图案变为一只手⇒按住鼠标左键拖动，三维散点图即开始依拖动的方向进行旋转。

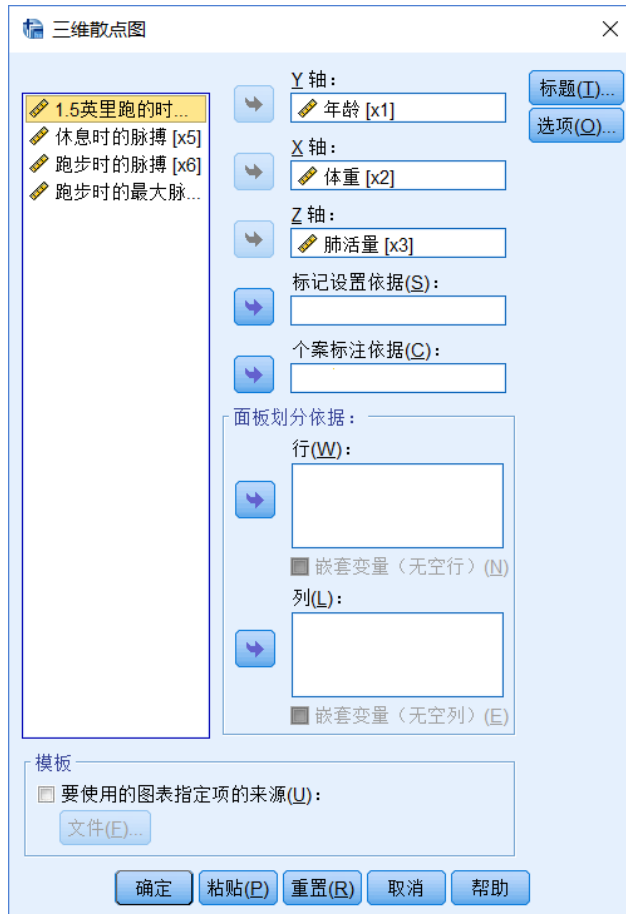


图 3.11

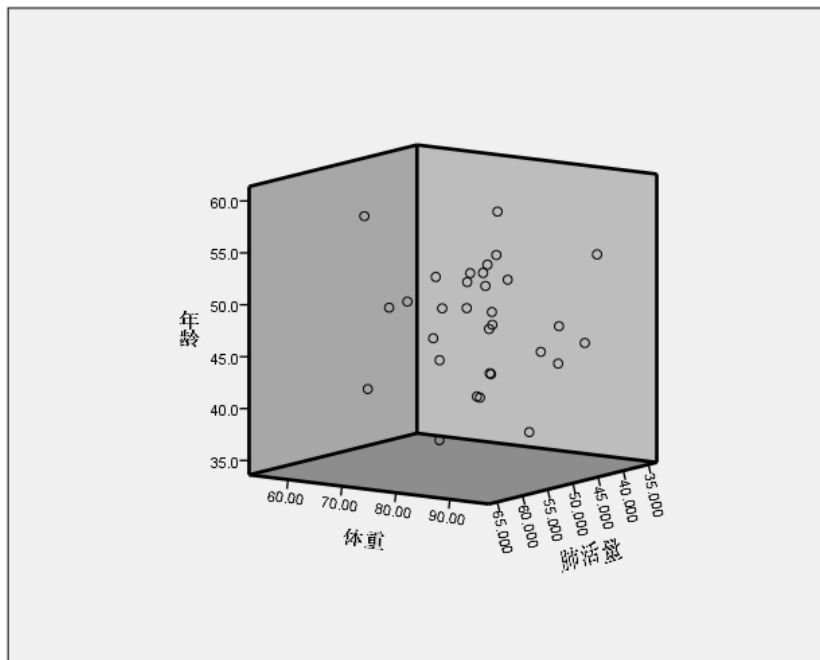


图 3.12

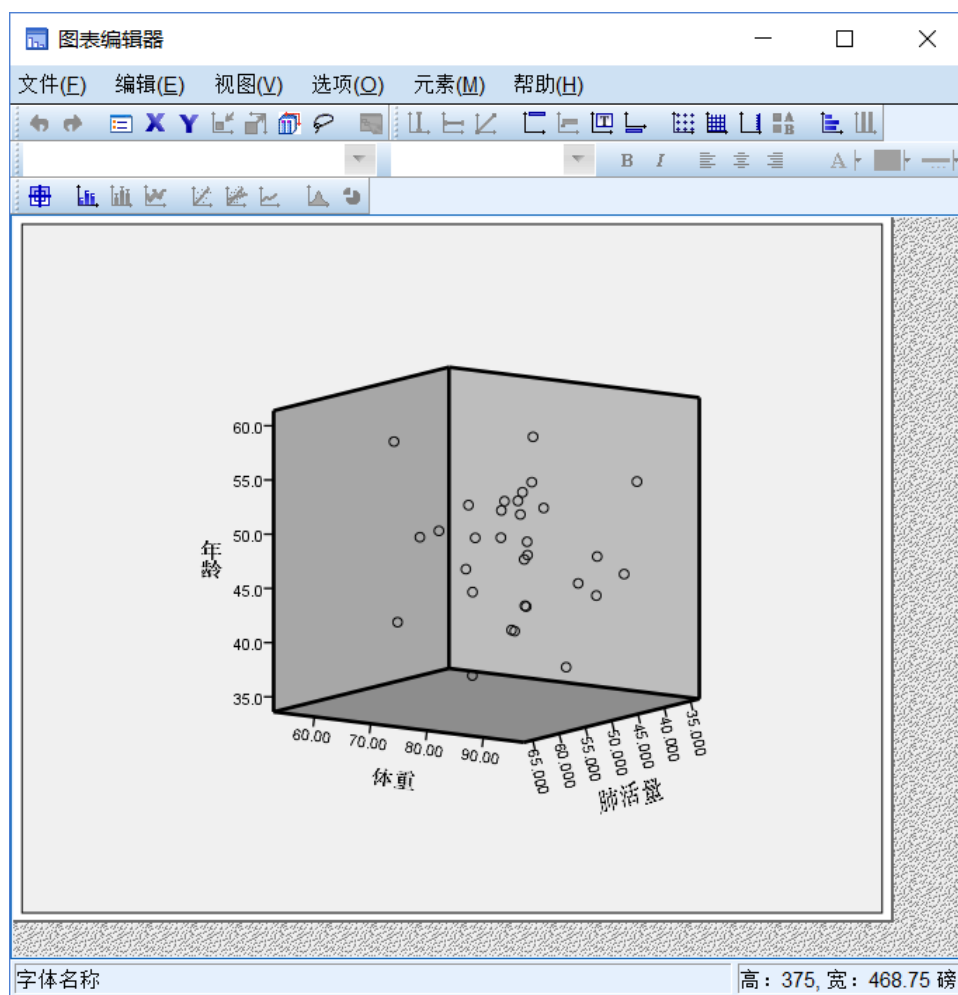


图 3.13

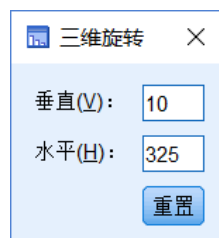


图 3.14

第四章 多元正态总体的统计推断

一、对例 4.5.1 进行多元方差分析和协差阵的齐次性检验

打开 examp4.5.1.sav 数据表 ⇒ 分析 ⇒ 一般线性模型 > ⇒ 多变量... ⇒ 在“多变量”对话框中（见图 4.1），将商品甲销售额[x1]、商品乙销售额[x2]、商品丙销售额[x3]和商品丁销售额[x4]选入“因变量”列表框中，将销售方式[g]选入“固定因子”列表框中 ⇒ 选项... ⇒ 在弹出的“多变量：选项”对话框中（见图 4.2），作图中的选择 ⇒ 继续 ⇒ 确定，生成图 4.3。

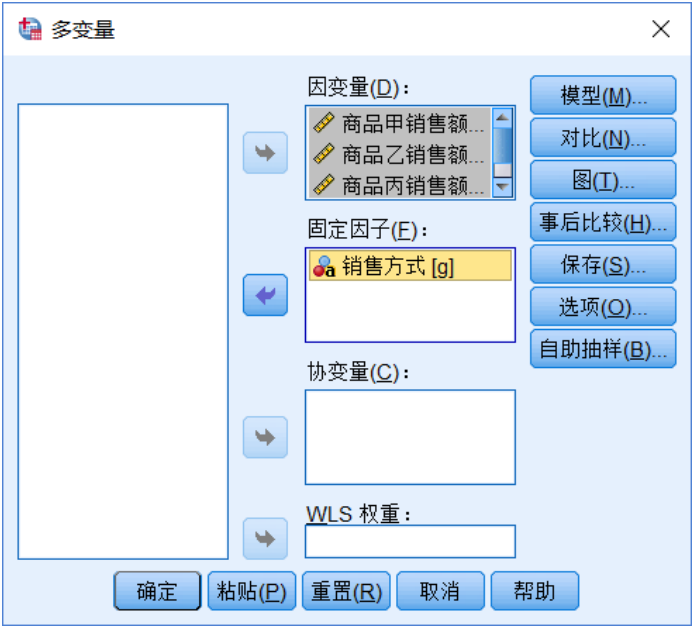


图 4.1

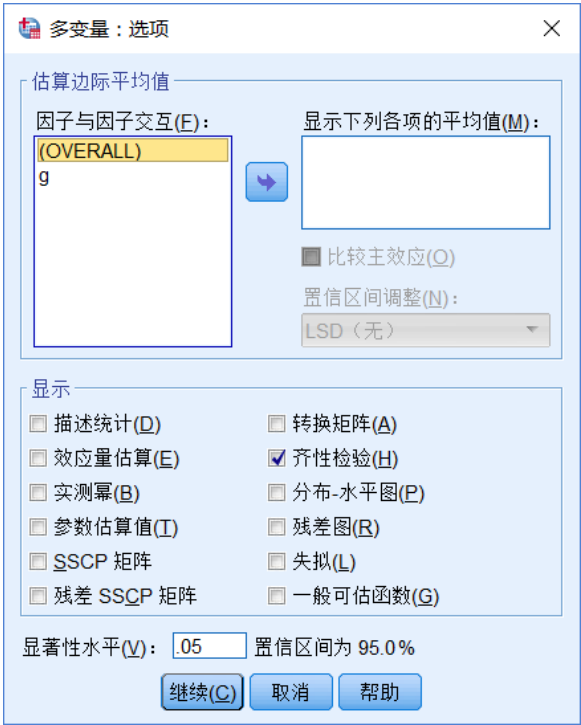


图 4.2

主体间因子

个案数

销售方式	1	20
	2	20
	3	20

协方差矩阵的博克斯等同性检验^a

博克斯 M	25.587
F	1.148
自由度 1	20
自由度 2	11662.473
显著性	.291

检验“各个组的因变量实测协方差矩阵相等”这一原假设。

a. 设计: 截距 + g

多变量检验^a

效应		值	F	假设自由度	误差自由度	显著性
截距	比莱轨迹	.988	1137.064 ^b	4.000	54.000	.000
	威尔克 Lambda	.012	1137.064 ^b	4.000	54.000	.000
	霍特林轨迹	84.227	1137.064 ^b	4.000	54.000	.000
	罗伊最大根	84.227	1137.064 ^b	4.000	54.000	.000
g	比莱轨迹	.361	3.031	8.000	110.000	.004
	威尔克 Lambda	.666	3.038 ^b	8.000	108.000	.004
	霍特林轨迹	.459	3.043	8.000	106.000	.004
	罗伊最大根	.336	4.621 ^c	4.000	55.000	.003

a. 设计: 截距 + g

b. 精确统计

c. 此统计是生成显著性水平下限的 F 的上限。

误差方差的莱文等同性检验^a

	F	自由度 1	自由度 2	显著性
商品甲销售额	2.382	2	57	.101
商品乙销售额	2.458	2	57	.095
商品丙销售额	1.014	2	57	.369
商品丁销售额	.580	2	57	.563

检验“各个组中的因变量误差方差相等”这一原假设。

a. 设计: 截距 + g

图 4.3

主体间效应检验

源	因变量	III 类平方和	自由度	均方	F	显著性
修正模型	商品甲销售额	5221.300 ^a	2	2610.650	3.377	.041
	商品乙销售额	518.533 ^b	2	259.267	1.615	.208
	商品丙销售额	2480.833 ^c	2	1240.417	.166	.848
	商品丁销售额	38529.300 ^d	2	19264.650	8.008	.001
截距	商品甲销售额	443244.150	1	443244.150	573.296	.000
	商品乙销售额	182050.417	1	182050.417	1134.326	.000
	商品丙销售额	10020506.67	1	10020506.67	1337.543	.000
	商品丁销售额	4012437.600	1	4012437.600	1668.007	.000
g	商品甲销售额	5221.300	2	2610.650	3.377	.041
	商品乙销售额	518.533	2	259.267	1.615	.208
	商品丙销售额	2480.833	2	1240.417	.166	.848
	商品丁销售额	38529.300	2	19264.650	8.008	.001
误差	商品甲销售额	44069.550	57	773.150		
	商品乙销售额	9148.050	57	160.492		
	商品丙销售额	427028.500	57	7491.728		
	商品丁销售额	137115.100	57	2405.528		
总计	商品甲销售额	492535.000	60			
	商品乙销售额	191717.000	60			
	商品丙销售额	10450016.00	60			
	商品丁销售额	4188082.000	60			
修正后总计	商品甲销售额	49290.850	59			
	商品乙销售额	9666.583	59			
	商品丙销售额	429509.333	59			
	商品丁销售额	175644.400	59			

a. R 方 = .106 (调整后 R 方 = .075)

b. R 方 = .054 (调整后 R 方 = .020)

c. R 方 = .006 (调整后 R 方 = -.029)

d. R 方 = .219 (调整后 R 方 = .192)

图 4.3 (续)

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.sav` 数据表 ⇒ 分析 ⇒ 分类 ⇒ 判别式... ⇒ 在“判别分析”对话框中，将现金流量/总债务[x1]、净收入/总资产[x2]、流动资产/流动债务[x3]和流动资产/净销售额[x4]选入“自变量”列表框中；将组别[g]选入“分组变量”列表框中（见图 5.1）⇒ 定义范围...⇒在弹出的“判别分析：定义范围”对话框中（见图 5.2），作图中的输入⇒继续⇒统计...⇒在弹出的“判别分析：统计”对话框中（见图 5.3），作图中的选择⇒继续；选择分类...⇒在弹出的“判别分析：分类”对话框中（见图 5.4），作图中的选择⇒继续；选择保存...⇒在弹出的“判别分析：保存”对话框中（见图 5.5），作图中的选择⇒继续⇒确定，生成图 5.6 和图 5.7。



图 5.1

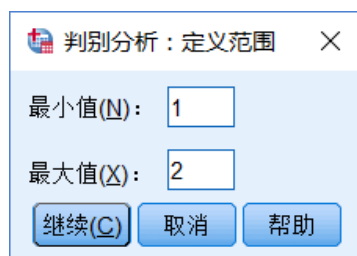


图 5.2

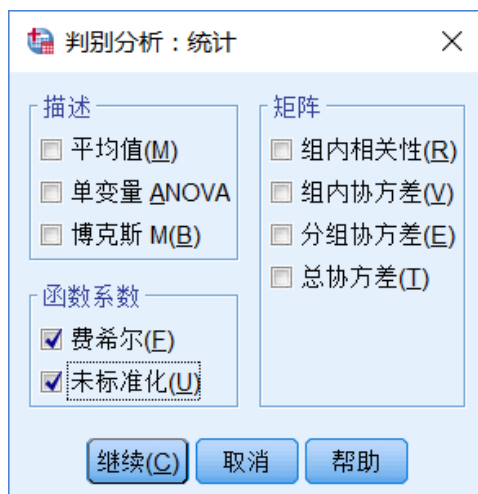


图 5.3

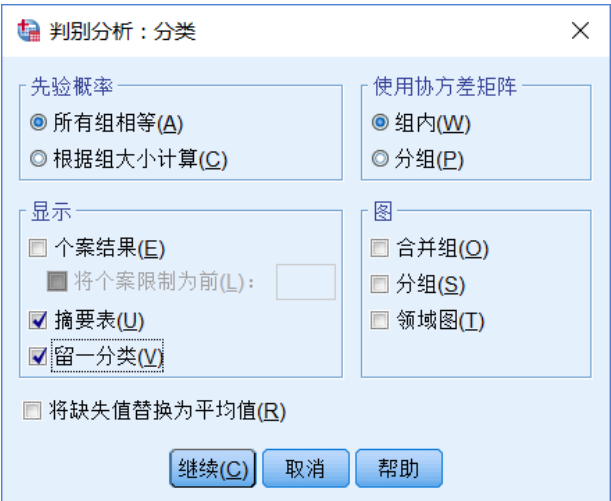


图 5.4

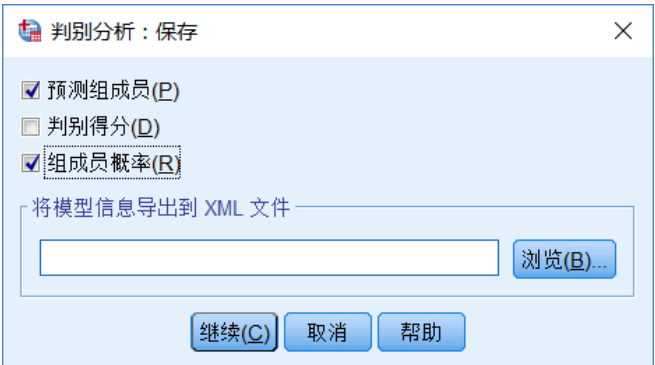


图 5.5

分类统计

分类处理摘要

已处理	47
排除	0
缺失或超出范围组代码	0
至少一个缺失判别变量	0
已在输出中使用	47

组的先验概率

组别	先验	在分析中使用的个案	
		未加权	加权
1.0	.500	21	21.000
2.0	.500	25	25.000
总计	1.000	46	46.000

分类函数系数

	组别	
	1.0	2.0
现金流量/总债务	4.035	5.295
净收入/总资产	-18.387	-10.020
流动资产/流动债务	1.616	3.306
流动资产/净销售额	12.194	9.949
(常量)	-5.075	-7.447

费希尔线性判别函数

图 5.6

分类结果^{a,c}

		预测组成员信息			
		组别	1.0	2.0	总计
原始	计数	1.0	18	3	21
		2.0	1	24	25
		未分组个案	1	0	1
	%	1.0	85.7	14.3	100.0
		2.0	4.0	96.0	100.0
		未分组个案	100.0	.0	100.0
交叉验证 ^b	计数	1.0	18	3	21
		2.0	2	23	25
		未分组个案	1	0	1
	%	1.0	85.7	14.3	100.0
		2.0	8.0	92.0	100.0
		未分组个案	100.0	.0	100.0

- a. 正确地对 91.3% 个原始已分组个案进行了分类。
b. 仅针对分析中的个案进行交叉验证。在交叉验证中，每个个案都由那些从该个案以外的所有个案派生的函数进行分类。
c. 正确地对 89.1% 个进行了交叉验证的已分组个案进行了分类。

图 5.6 (续)

	x1	x2	x3	x4	g	Dis_1	Dis1_1	Dis2_1
31	.17	.07	1.80	.52	2.0	2.0	.42511	.57489
32	.15	.05	2.17	.55	2.0	2.0	.33913	.66087
33	-.10	-.01	2.50	.58	2.0	2.0	.41568	.58432
34	.14	-.03	.46	.26	2.0	1.0	.90494	.09506
35	.14	.07	2.61	.52	2.0	2.0	.16346	.83654
36	.15	.06	2.23	.56	2.0	2.0	.30369	.69631
37	.16	.05	2.31	.20	2.0	2.0	.15422	.84578
38	.29	.06	1.84	.38	2.0	2.0	.32058	.67942
39	.54	.11	2.33	.48	2.0	2.0	.11028	.88972
40	-.33	-.09	3.01	.47	2.0	2.0	.37986	.62014
41	.48	.09	1.24	.18	2.0	2.0	.33705	.66295
42	.56	.11	4.29	.44	2.0	2.0	.00401	.99599
43	.20	.08	1.99	.30	2.0	2.0	.22476	.77524
44	.47	.14	2.92	.45	2.0	2.0	.03506	.96494
45	.17	.04	2.45	.14	2.0	2.0	.11899	.88101
46	.58	.04	5.06	.13	2.0	2.0	.00096	.99904
47	-.16	-.10	1.45	.51	1.0	1.0	.89141	.10859

图 5.7

注：Dis_1 表示经判别归属的组，Dis1_1 和 Dis2_1 分别表示归属第 1 组和第 2 组的后验概率。

二、对书中例 5.4.1 中的数据作 Fisher 判别

打开 examp5.4.1.sav 数据表 → 分析 → 分类 → 判别式... → 在“判别分析”对话框中，将花萼长[x1]、花萼宽[x2]、花瓣长[x3]和花瓣宽[x4]选入“自变量”列表框中；将组别[g]选入“分组变量”列表框中（见图 5.8）→ 定义范围...，即弹出“判别分析...”对话框 → 在“最小”框中填入 1，在“最大”框中填入 3 → 继续 → 统计... → 在弹出的“判别分析：统计”对话框中，选择“函数系数”一栏里的未标准化 → 继续；分类... → 在弹出的“判别分析：分类”对话框中，选择“图”一栏里的合并组 → 继续 → 确定，生成图 5.9。



图 5.8

分析个案处理摘要

未加权个案数		个案数	百分比
有效		150	100.0
排除	缺失或超出范围组代码	0	.0
	至少一个缺失判别变量	0	.0
	既包括缺失或超出范围组代码，也包括至少一个缺失判别变量	0	.0
	总计	0	.0
总计		150	100.0

组统计

组别		有效个案数 (成列)	
		未加权	加权
1	花萼长	50	50.000
	花萼宽	50	50.000
	花瓣长	50	50.000
	花瓣宽	50	50.000
2	花萼长	50	50.000
	花萼宽	50	50.000
	花瓣长	50	50.000
	花瓣宽	50	50.000
3	花萼长	50	50.000
	花萼宽	50	50.000
	花瓣长	50	50.000
	花瓣宽	50	50.000
总计	花萼长	150	150.000
	花萼宽	150	150.000
	花瓣长	150	150.000
	花瓣宽	150	150.000

图 5.9

典则判别函数摘要

特征值				
函数	特征值	方差百分比	累计百分比	典型相关性
1	32.192 ^a	99.1	99.1	.985
2	.285 ^a	.9	100.0	.471

a. 在分析中使用了前 2 个典则判别函数。

威尔克 Lambda				
函数检验	威尔克 Lambda	卡方	自由度	显著性
1 直至 2	.023	546.115	8	.000
2	.778	36.530	3	.000

标准化典则判别函数系数		
函数		
	1	2
花萼长	-.427	.012
花萼宽	-.521	.735
花瓣长	.947	-.401
花瓣宽	.575	.581

结构矩阵		
函数		
	1	2
花瓣长	.706 [*]	.168
花萼宽	-.119	.864 [*]
花瓣宽	.633	.737 [*]
花萼长	.223	.311 [*]

判别变量与标准化典则判别函数之间的汇聚组内相关性
变量按函数内相关性的绝对大小排序。

*. 每个变量与任何判别函数之间的最大绝对相关性

典则判别函数系数		
函数		
	1	2
花萼长	-.083	.002
花萼宽	-.153	.216
花瓣长	.220	-.093
花瓣宽	.281	.284
(常量)	-2.105	-6.661

未标准化系数

组质心处的函数		
函数		
组别	1	2
1	-7.608	.215
2	1.825	-.728
3	5.783	.513

按组平均值进行求值的未标准化典则判别函数

图 5.9（续 1）

分类处理摘要

已处理		150
排除	缺失或超出范围组代码	0
	至少一个缺失判别变量	0
已在输出中使用		150

组的先验概率

组别	先验	在分析中使用的个案	
		未加权	加权
1	.333	50	50.000
2	.333	50	50.000
3	.333	50	50.000
总计	1.000	150	150.000

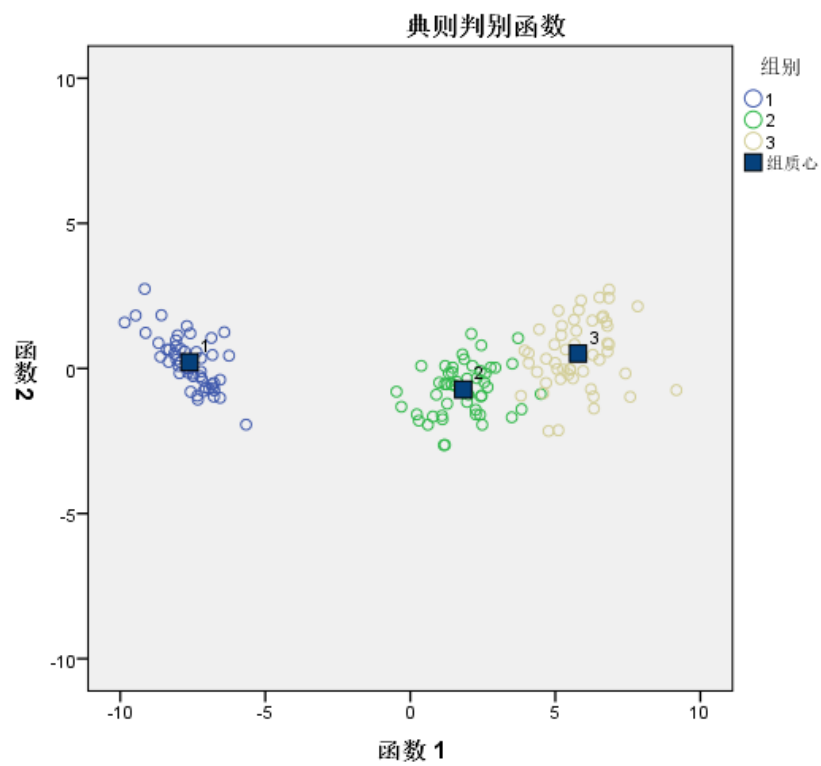


图 5.9 (续 2)

三、书中例 5.5.1 的逐步判别

在图 5.8 中，选择使用步进法⇒确定，产生图 5.10。

步进统计

输入/除去的变量 ^{a,b,c,d}									
步骤	输入	威尔克 Lambda			精确 F				显著性
		统计	自由度 1	自由度 2	自由度 3	统计	自由度 1	自由度 2	
1	花瓣长	.059	1	2	147.000	1180.161	2	147.000	.000
2	花萼宽	.037	2	2	147.000	307.105	4	292.000	.000
3	花瓣宽	.025	3	2	147.000	257.503	6	290.000	.000
4	花萼长	.023	4	2	147.000	199.145	8	288.000	.000

在每个步骤中，将输入可以使总体威尔克 Lambda 最小化的变量。

a. 最大步骤数为 8。
b. 要输入的最小偏 F 为 3.84。
c. 要除去的最大偏 F 为 2.71。
d. F 级别、容差或 VIN 不足，无法进行进一步计算。

包括在分析中的变量				
步骤		容差	要除去的 F	威尔克 Lambda
1	花瓣长	1.000	1180.161	
2	花瓣长	.857	1112.954	.599
	花萼宽	.857	43.035	.059
3	花瓣长	.736	38.724	.038
	花萼宽	.749	54.577	.044
	花瓣宽	.669	34.569	.037
4	花瓣长	.365	35.590	.035
	花萼宽	.609	21.936	.031
	花瓣宽	.649	24.904	.032
	花萼长	.348	4.721	.025

未包括在分析中的变量					
步骤		容差	最小容差	要输入的 F	威尔克 Lambda
0	花萼长	1.000	1.000	119.265	.381
	花萼宽	1.000	1.000	49.160	.599
	花瓣长	1.000	1.000	1180.161	.059
	花瓣宽	1.000	1.000	960.007	.071
1	花萼长	.428	.428	34.323	.040
	花萼宽	.857	.857	43.035	.037
	花瓣宽	.765	.765	24.766	.044
2	花萼长	.358	.358	12.268	.032
	花瓣宽	.669	.669	34.569	.025
3	花萼长	.348	.348	4.721	.023

威尔克 Lambda									
步骤	变量数	Lambda	自由度			精确 F			
			1	2	3	统计	1	2	显著性
1	1	.059	1	2	147	1180.161	2	147.000	.000
2	2	.037	2	2	147	307.105	4	292.000	.000
3	3	.025	3	2	147	257.503	6	290.000	.000
4	4	.023	4	2	147	199.145	8	288.000	.000

图 5.10

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.sav 数据表 ⇒ 分析 ⇒ 分类 ⇒ 系统聚类... ⇒ 在“系统聚类分析”对话框中,将食品[x1]、衣着[x2]、家庭设备用品及服务[x3]、医疗保健[x4]、交通和通讯[x5]、娱乐教育文化服务[x6]、居住[x7]和杂项商品和服务[x8]选入“变量”列表框中;将地区[region]选入“个案标注依据”列表框中(见图 6.1) ⇒ 统计... ⇒ 在弹出的“系统聚类分析:统计”对话框中(见图 6.2),作图中的选择→继续; 图... ⇒ 在弹出的“系统聚类分析:图”对话框中(见图 6.3),作图中的选择→继续; 方法... ⇒ 在弹出的“系统聚类分析:方法”对话框中(见图 6.4),作图中的选择→继续 ⇒ 确定,生成图 6.5。



图 6.1

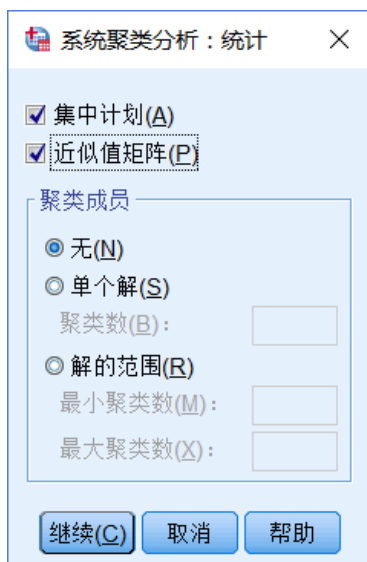


图 6.2

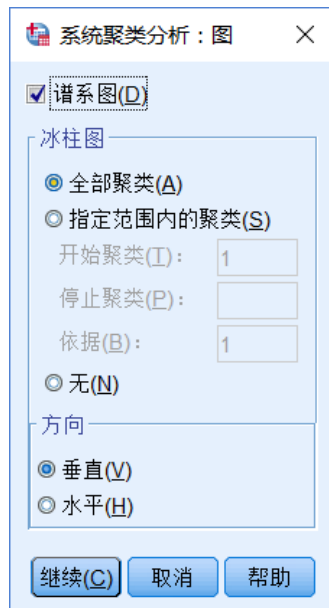


图 6.3

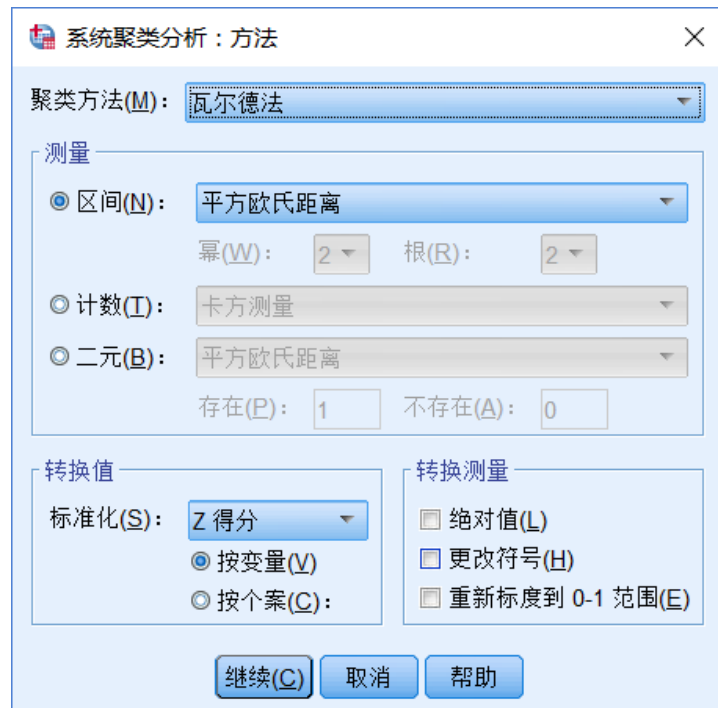


图 6.4

近似值矩阵															
个案	平方欧氏距离														
	1:北京	2:天津	3:河北	4:山西	5:内蒙古	6:辽宁	7:吉林	8:黑龙江	9:上海	10:江苏	11:浙江	12:安徽	13:福建	14:江西	15:山东
1:北京	.000	20.510	43.169	58.569	58.325	50.256	60.117	59.508	10.838	36.059	9.435	63.908	48.276	79.565	37.219
2:天津	20.510	.000	10.554	16.916	18.226	15.213	18.053	19.647	16.261	3.635	6.094	18.008	10.991	23.039	7.384
3:河北	43.169	10.554	.000	2.142	2.044	1.120	2.026	1.865	48.413	4.645	19.478	4.827	9.422	9.535	2.587
4:山西	58.569	16.916	2.142	.000	.510	1.939	1.177	1.465	61.635	6.793	31.298	2.560	10.375	5.815	5.134
5:内蒙古	58.325	18.226	2.044	.510	.000	1.810	.948	1.554	62.433	7.677	31.449	2.365	10.531	6.623	4.597
6:辽宁	50.256	15.213	1.120	1.939	1.810	.000	1.009	.630	56.492	7.455	25.119	4.434	9.780	9.629	4.674
7:吉林	60.117	18.053	2.026	1.177	.948	1.009	.000	.817	62.957	8.045	31.353	2.001	9.178	5.058	6.076
8:黑龙江	59.508	19.647	1.865	1.465	1.554	.630	.817	.000	66.192	9.992	31.009	4.711	12.321	8.207	6.989
9:上海	10.838	16.261	48.413	61.635	62.433	56.492	62.957	66.192	.000	30.268	10.567	60.127	36.341	69.611	39.705
10:江苏	36.059	3.635	4.645	6.793	7.677	7.455	8.045	9.992	30.268	.000	13.848	6.290	4.385	10.523	3.088
11:浙江	9.435	6.094	19.478	31.298	31.449	25.119	31.353	31.009	10.567	13.848	.000	33.853	20.180	42.314	17.425
12:安徽	63.908	18.008	4.827	2.560	2.365	4.434	2.001	4.711	60.127	6.290	33.853	.000	6.555	3.281	6.800
13:福建	48.276	10.991	9.422	10.375	10.531	9.780	9.178	12.321	36.341	4.385	20.180	6.555	.000	8.713	9.981
14:江西	79.565	23.039	9.535	5.815	6.623	9.629	5.058	8.207	69.611	10.523	42.314	3.281	8.713	.000	13.859
15:山东	37.219	7.384	2.587	5.134	4.597	4.674	6.076	6.989	39.705	3.088	17.425	6.800	9.981	13.859	.000
16:河南	64.995	17.738	2.987	1.297	1.935	2.682	1.144	1.696	63.918	7.408	32.498	2.770	8.756	2.995	7.052
17:湖北	47.833	10.583	2.527	3.908	3.366	3.198	2.228	4.576	46.670	4.623	22.933	3.214	5.984	5.878	4.017
18:湖南	36.410	5.899	3.056	5.993	5.347	5.333	5.319	7.818	34.082	1.984	15.154	4.969	4.654	9.471	2.580
19:广东	36.531	31.834	56.775	71.067	70.466	65.391	68.574	72.082	17.834	42.392	20.593	67.477	37.482	67.153	56.013
20:广西	53.905	11.402	8.461	9.203	9.695	11.012	8.162	12.025	42.220	4.797	24.099	5.449	3.630	4.277	10.862
21:海南	66.196	21.341	12.375	9.223	11.031	13.495	10.497	12.966	57.159	10.213	34.701	7.176	8.191	6.496	18.354
22:重庆	26.528	4.595	5.736	11.043	9.943	8.353	10.457	12.465	25.418	2.879	9.392	9.854	6.421	18.098	2.964
23:四川	43.453	8.261	1.990	2.699	2.783	2.684	2.490	4.405	42.175	2.348	20.317	2.684	4.664	6.668	2.593
24:贵州	56.677	12.590	3.439	2.035	2.567	4.631	3.245	4.958	52.313	3.084	27.578	1.677	6.495	4.421	3.717
25:云南	31.607	5.907	4.330	5.933	6.495	5.582	7.304	8.297	31.768	2.830	13.983	8.104	5.588	13.478	4.255
26:西藏	41.852	25.335	19.627	19.756	18.449	16.826	21.309	21.272	47.730	20.204	29.674	23.419	18.464	35.185	16.565
27:陕西	50.198	9.636	1.957	3.159	4.276	4.170	3.547	4.131	48.749	3.681	21.913	4.946	8.492	5.519	4.893
28:甘肃	53.696	14.745	1.908	.212	.926	2.092	1.823	2.062	56.984	5.519	28.340	2.834	9.966	7.011	4.392
29:青海	47.049	13.402	1.521	1.826	2.774	1.800	2.542	1.949	52.460	6.296	22.948	5.290	9.850	9.144	6.456
30:宁夏	51.943	17.436	1.571	1.884	2.422	1.463	2.593	1.025	60.850	9.040	26.608	6.439	13.732	11.281	6.747
31:新疆	42.526	8.978	.677	1.810	1.854	1.990	2.661	3.199	45.424	2.900	19.935	3.910	8.333	9.184	1.041

图 6.5

注：“近似值矩阵”中的近似值是指各对象之间的距离或相似系数，图 6.5 中是指样品之间的距离。

集中计划

阶段	组合聚类		系数	首次出现聚类的阶段		下一个阶段
	聚类 1	聚类 2		聚类 1	聚类 2	
1	4	28	.106	0	0	6
2	6	8	.421	0	0	7
3	3	31	.760	0	0	11
4	17	23	1.160	0	0	8
5	29	30	1.567	0	0	13
6	4	5	2.010	1	0	16
7	6	7	2.514	2	0	13
8	17	18	3.198	4	0	19
9	12	24	4.037	0	0	17
10	16	27	5.083	0	0	17
11	3	15	6.180	3	0	19
12	10	25	7.595	0	0	14
13	6	29	9.298	7	5	16
14	10	22	11.100	12	0	20
15	13	20	12.916	0	0	21
16	4	6	14.835	6	13	25
17	12	16	16.976	9	10	18
18	12	14	19.414	17	0	25
19	3	17	21.920	11	8	24
20	2	10	24.650	0	14	24
21	13	21	28.515	15	0	26
22	1	11	33.232	0	0	23
23	1	9	38.795	22	0	27
24	2	3	44.959	20	19	26
25	4	12	52.002	16	18	29
26	2	13	65.285	24	21	28
27	1	19	81.455	23	0	30
28	2	26	97.653	26	0	29
29	2	4	118.970	28	25	30
30	1	2	240.000	27	29	0

图 6.5（续 1）



图 6.5（续 2）

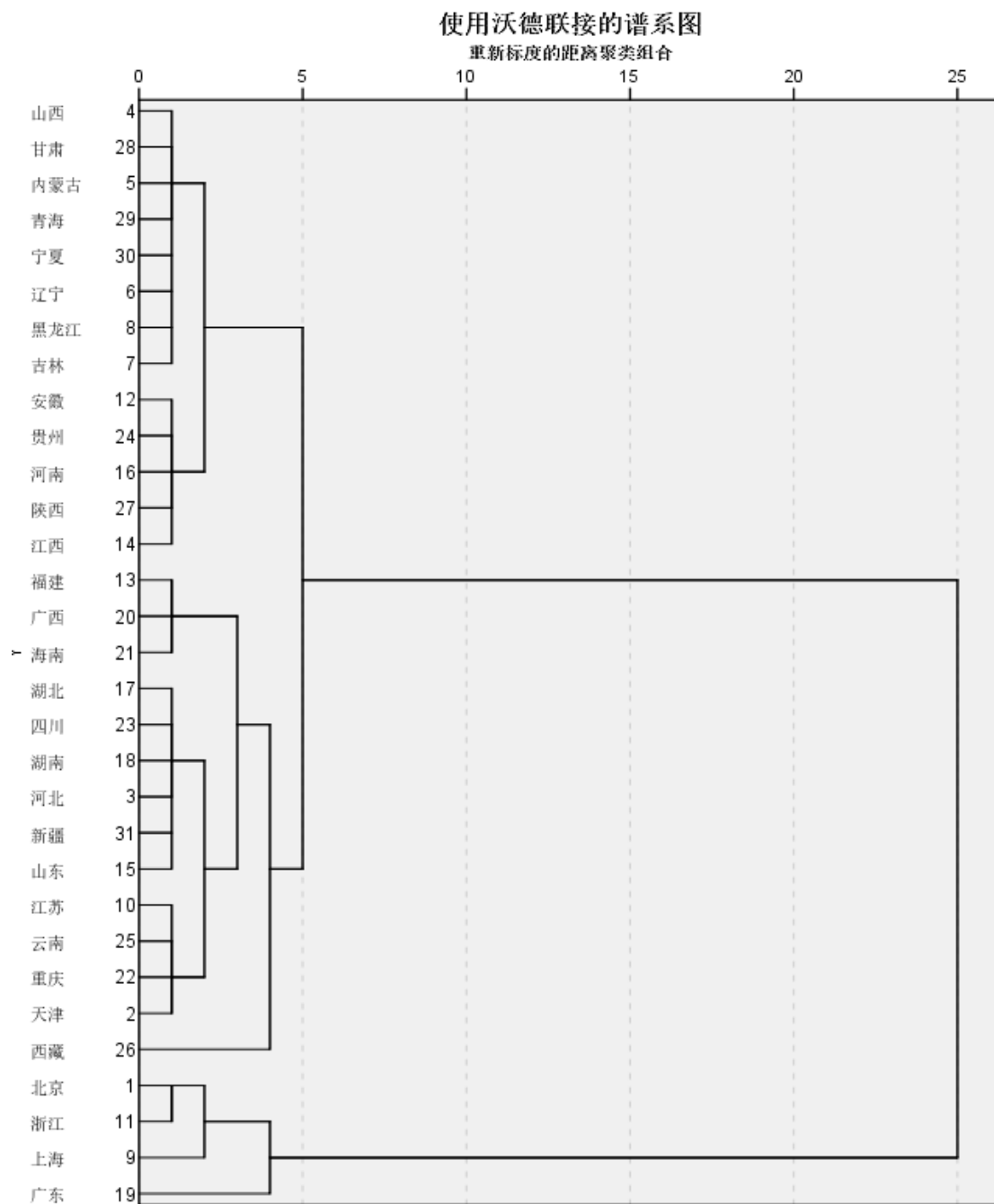


图 6.5 (续 3)

注：

1. 在“凝聚计划”表中，“首次出现阶段集群”列里的“0”表示该样品是第一次出现在聚类过程中。“下一个阶段”列表示在该步骤中合并的类下一次将在第几步中与其他类再进行合并。

2. “冰柱图”也是一种查看聚类结果的图形，它虽不如谱系图（即树形图）容易看，但图形的分辨率更高。指定聚类个数，然后在此个数上水平切一刀，即可得到分类结果。在 SAS 和 JMP 中，因为树形图的分辨率高，故不需要画冰柱图。

二、对书中例 6.3.3 用 k 均值法聚类

打开 examp6.3.3.sav 数据表 \Rightarrow 分析 \Rightarrow 描述统计 \Rightarrow 描述...，出现“描述性”对话框 \Rightarrow 将食品[x1]、衣着[x2]、家庭设备用品及服务[x3]、医疗保健[x4]、交通和通讯[x5]、娱乐

教育文化服务[x6]、居住[x7]和杂项商品和服务[x8]选入“变量”列表框中（见图 6.6）；选择将标准化值另存为变量⇒确定，生成图 6.7，标准化后的变量分别表示为 Zx1 至 Zx8。

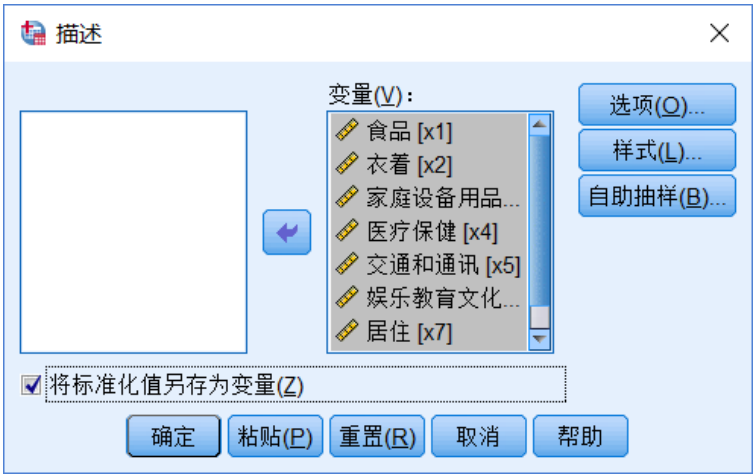


图 6.6

	x8	Zx1	Zx2	Zx3	Zx4	Zx5	Zx6
1	457.64	1.66732	2.00667	1.96425	3.11488	1.22417	3.01963
2	305.08	.81636	.05310	1.67480	.67756	-.20939	.87805
3	188.63	-.82645	.22271	-.18683	.47433	-.29711	-.15298
4	212.10	-.97861	-.09384	-.58821	-.41446	-.85475	-.81712
5	192.96	-1.15302	.29236	-.78451	-.60438	-.47771	-.56188
6	163.86	-.42567	.53818	-.82853	.41052	-.56067	-.65628
7	147.76	-.71360	.02778	-1.08652	-.30109	-.70498	-.58019
8	152.85	-.97217	.17962	-1.02321	.37925	-.67407	-1.01711
9	462.03	2.95057	.51194	2.76434	1.18779	1.68566	2.45586
10	252.54	.38665	-.32961	.98047	-.37567	-.06968	.08263
11	323.36	1.10498	.56657	1.63256	2.21567	1.58935	1.19413
12	151.39	-.23153	-.48800	-.69308	-1.36683	-.47186	-.29757
13	232.29	1.24181	-.50610	-.34383	-.96801	.73459	-.56937
14	160.12	-.71033	-1.53934	-.90133	-1.58026	-.79075	-.92651
15	211.84	-.51954	1.03147	.85992	-.28453	-.29992	.15756
16	165.32	-.94228	-.47555	-.59710	-.41944	-.73378	-1.22322
17	182.52	-.33606	.18934	-.62884	-.50201	-.57300	.25417
18	226.45	.08549	.10257	.02004	.44252	.08017	.67257

图 6.7

在图 6.7 中，选择分析⇒分类⇒K-均值聚类...⇒在“K 平均值聚类分析”对话框中，将 Zscore：食品[x1]、Zscore：衣着[x2]、Zscore：家庭设备用品及服务[x3]、Zscore：医疗保健[x4]、Zscore：交通和通讯[x5]、Zscore：娱乐教育文化服务[x6]、Zscore：居住[x7]和 Zscore：杂项商品和服务[x8]选入“变量”列表框中（见图 6.8）；将地区[region]选入“个案标注依据”列表框中；在“聚类数”列表框中，填入 5⇒选项...⇒在弹出的“K 均值聚类分析”对话框中（见图 6.9），作图中的选择⇒继续⇒确定，生成图 6.10。

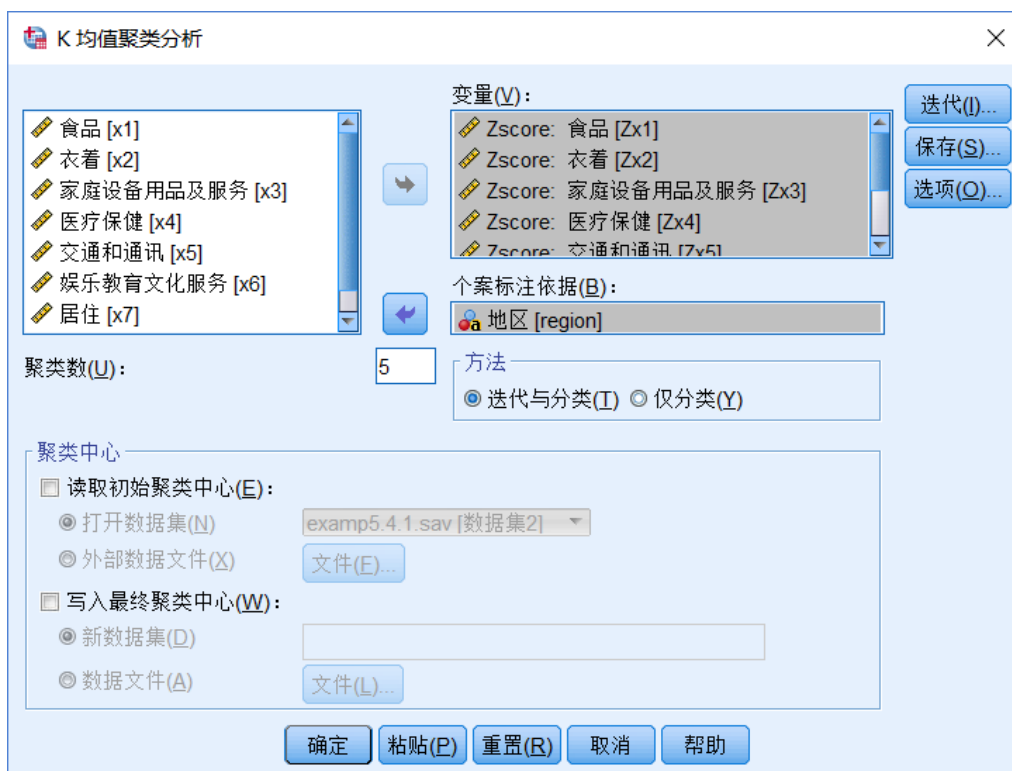


图 6.8

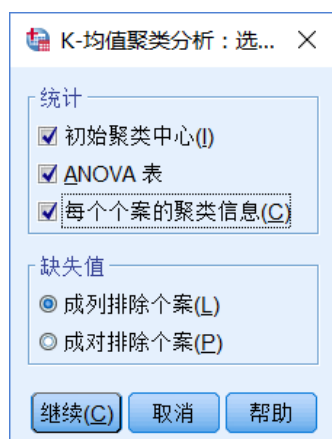


图 6.9

初始聚类中心					
	聚类				
	1	2	3	4	5
Zscore: 食品	1.66732	1.83087	.13154	1.13472	-.97217
Zscore: 衣着	2.00667	-1.12774	-2.51238	2.91081	.17962
Zscore: 家庭设备用品及服务	1.96425	.93690	-1.07412	-1.06456	-1.02321
Zscore: 医疗保健	3.11488	1.29595	-.84039	-.40821	.37925
Zscore: 交通和通讯	1.22417	3.90905	.14541	.53291	-.67407
Zscore: 娱乐教育文化服务	3.01963	1.60144	-.48758	-1.04761	-1.01711
Zscore: 居住	.28577	3.88031	-.69845	-.95621	-.67062
Zscore: 杂项商品和服务	2.43702	2.01877	.41047	1.66127	-1.02429

迭代历史记录 ^a					
	聚类中心中的变动				
迭代	1	2	3	4	5
1	1.825	.000	1.641	.000	1.579
2	.000	.000	.000	.000	.000

a. 由于聚类中心中不存在变动或者仅有小幅变动，因此实现了收敛。任何中心的最大绝对坐标变动为 .000。当前迭代为 2。初始中心之间的最小距离为 3.601。

聚类成员			
个案号	地区	聚类	距离
1	北京	1	1.825
2	天津	5	3.075
3	河北	5	.800
4	山西	5	1.187
5	内蒙古	5	1.261
6	辽宁	5	1.221
7	吉林	5	1.280
8	黑龙江	5	1.579
9	上海	1	1.926
10	江苏	5	1.662
11	浙江	1	1.800
12	安徽	5	1.641
13	福建	3	1.688
14	江西	3	1.609
15	山东	5	1.486
16	河南	5	1.475
17	湖北	5	1.231
18	湖南	5	1.380
19	广东	2	.000
20	广西	3	.999
21	海南	3	1.641
22	重庆	5	2.224
23	四川	5	.738
24	贵州	5	1.263
25	云南	5	1.691
26	西藏	4	.000
27	陕西	5	1.292
28	甘肃	5	1.067
29	青海	5	1.305
30	宁夏	5	1.631
31	新疆	5	.583

图 6.10

最终聚类中心						
	聚类					
	1	2	3	4	5	
Zscore: 食品	1.90763	1.83087	.18842	1.13472	-.42919	
Zscore: 衣着	1.02839	-1.12774	-1.53016	2.91081	.05693	
Zscore: 家庭设备用品及服务	2.12038	.93690	-.65949	-1.06456	-.16343	
Zscore: 医疗保健	2.17278	1.29595	-1.09782	-.40821	-.13704	
Zscore: 交通和通讯	1.49973	3.90905	.05751	.53291	-.41687	
Zscore: 娱乐教育文化服务	2.22321	1.60144	-.42705	-1.04761	-.25069	
Zscore: 居住	.95831	3.88031	.33155	-.95621	-.32387	
Zscore: 杂项商品和服务	1.94533	2.01877	-.23369	1.66127	-.39006	

最终聚类中心之间的距离					
聚类	1	2	3	4	5
1		4.643	6.494	6.027	5.818
2	4.643		6.910	8.102	7.443
3	6.494	6.910		5.207	2.186
4	6.027	8.102	5.207		4.198
5	5.818	7.443	2.186	4.198	

ANOVA						
	聚类		误差		F	显著性
	均方	自由度	均方	自由度		
Zscore: 食品	4.938	4	.394	26	12.527	.000
Zscore: 衣着	5.589	4	.294	26	19.005	.000
Zscore: 家庭设备用品及服务	4.457	4	.468	26	9.518	.000
Zscore: 医疗保健	5.311	4	.337	26	15.768	.000
Zscore: 交通和通讯	6.537	4	.148	26	44.132	.000
Zscore: 娱乐教育文化服务	5.151	4	.361	26	14.249	.000
Zscore: 居住	5.368	4	.328	26	16.370	.000
Zscore: 杂项商品和服务	5.438	4	.317	26	17.147	.000

由于已选择聚类以使不同聚类中个案之间的差异最大化，因此 F 检验只应该用于描述目的。实测显著性水平并未因此进行修正，所以无法解释为针对“聚类平均值相等”这一假设的检验。

每个聚类中的个案数目		
聚类	1	3.000
	2	1.000
	3	4.000
	4	1.000
	5	22.000
有效		31.000
缺失		.000

图 6.10（续）

第七章 主成分分析

一、对书中例 6.3.3 作主成分分析

打开 examp6.3.3.sav 数据表 \Rightarrow 分析 \Rightarrow 降维 \Rightarrow 因子... \Rightarrow 在“因子分析”对话框中，将食品[x1]、衣着[x2]、家庭设备用品及服务[x3]、医疗保健[x4]、交通和通讯[x5]、娱乐教育文化服务[x6]、居住[x7]和杂项商品和服务[x8]选入“变量”列表框中（见图 7.1） \Rightarrow 描述... \Rightarrow 在弹出的“因子分析：描述”对话框中（见图 7.2），作图中的选择 \rightarrow 继续；提取... \Rightarrow 在弹出的“因子分析：提取”对话框中（见图 7.3），作图中的选择 \rightarrow 继续 \Rightarrow 确定，生成图 7.4。



图 7.1

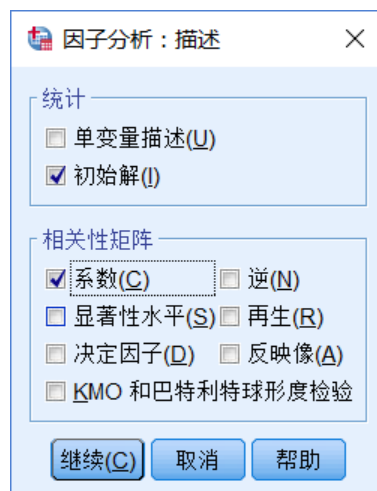
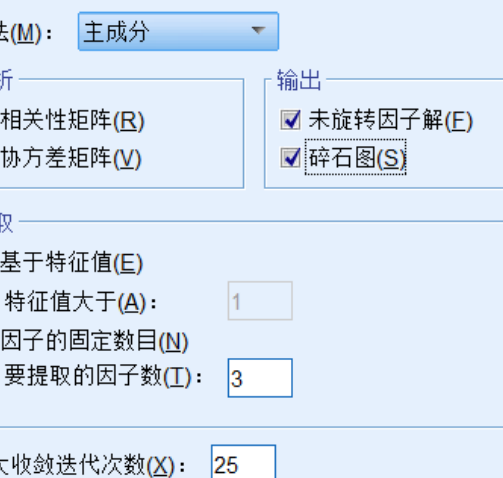


图 7.2



因子分析: 提取

方法(M): 主成分

分析

- ☒ 相关性矩阵(R)
- ☐ 协方差矩阵(V)

输出

- ☒ 未旋转因子解(E)
- ☒ 碎石图(S)

提取

- ☐ 基于特征值(E)
特征值大于(A): 1
- ☒ 因子的固定数目(N)
要提取的因子数(I): 3

最大收敛迭代次数(X): 25

继续(C) 取消 帮助

相关性矩阵									
		食品	衣着	家庭设备用品及服务	医疗保健	交通和通讯	娱乐教育文化服务	居住	杂项商品和服务
相关性	食品	1.000	.247	.698	.468	.828	.769	.670	.877
	衣着	.247	1.000	.258	.423	.086	.255	-.201	.349
	家庭设备用品及服务	.698	.258	1.000	.621	.585	.856	.569	.667
	医疗保健	.468	.423	.621	1.000	.531	.684	.314	.628
	交通和通讯	.828	.086	.585	.531	1.000	.708	.800	.776
	娱乐教育文化服务	.769	.255	.856	.684	.708	1.000	.647	.745
	居住	.670	-.201	.569	.314	.800	.647	1.000	.525
	杂项商品和服务	.877	.349	.667	.628	.776	.745	.525	1.000

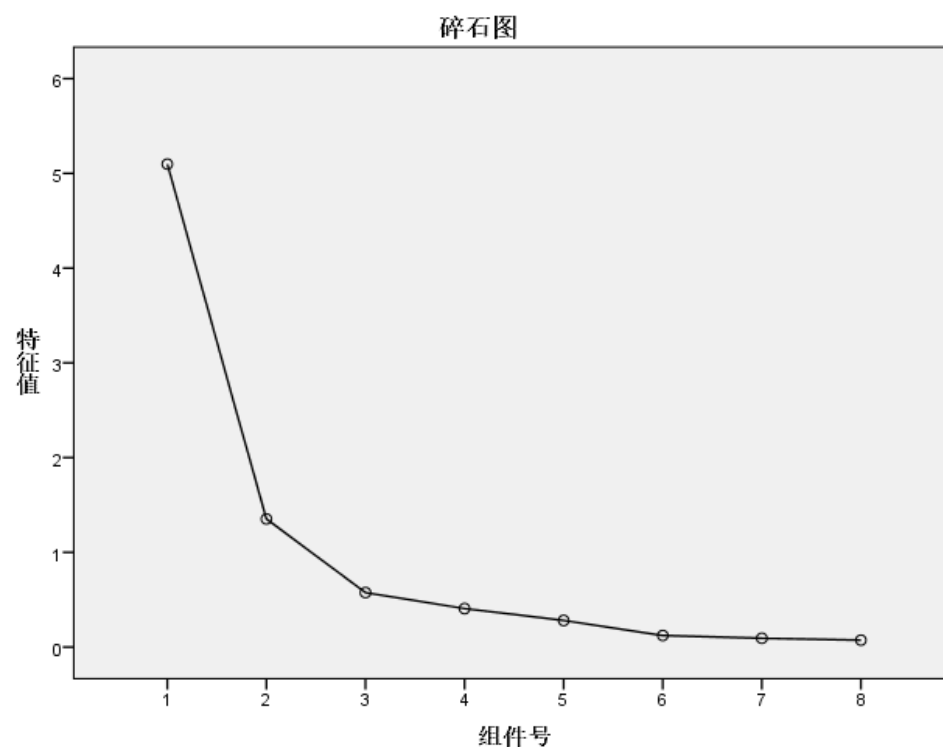
	初始	提取
食品	1.000	.927
衣着	1.000	.911
家庭设备用品及服务	1.000	.835
医疗保健	1.000	.814
交通和通讯	1.000	.884
娱乐教育文化服务	1.000	.896
居住	1.000	.876
杂项商品和服务	1.000	.882

提取方法：主成分分析法。

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	5.098	63.721	63.721	5.098	63.721	63.721
2	1.352	16.903	80.625	1.352	16.903	80.625
3	.575	7.184	87.809	.575	7.184	87.809
4	.406	5.079	92.887			
5	.281	3.516	96.403			
6	.122	1.528	97.932			
7	.093	1.158	99.090			
8	.073	.910	100.000			

提取方法：主成分分析法。

图 7.4



成分矩阵^a

	成分		
	1	2	3
食品	.905	-.090	.315
衣着	.298	.871	.252
家庭设备用品及服务	.847	.076	-.335
医疗保健	.722	.401	-.362
交通和通讯	.876	-.270	.212
娱乐教育文化服务	.916	.032	-.235
居住	.737	-.577	-.026
杂项商品和服务	.895	.112	.262

提取方法：主成分分析法。

a. 提取了 3 个成分。

图 7.4 (续)

注：

(1) “成分矩阵”是因子分析中主成分法的载荷矩阵，它同时也是原始变量与主成分之间的相关矩阵。

(2) SPSS 中可以得到因子得分，但无法得到主成分得分（除非自己排公式计算），也就无法画出主成分的散点图。

第八章 因子分析

一、对书中习题 6.5 中的数据作因子分析

1.主成分法

打开 exec6.5.sav 数据表 ⇒ 分析 ⇒ 降维 ⇒ 因子... ⇒ 在“因子分析”对话框中，将

100 米（秒）[x1]、200 米（秒）[x2]、400 米（秒）[x3]、800 米（分）[x4]、1500 米（分）[x5]、5000 米（分）[x6]、10000 米（分）[x7]和马拉松（分）[x8]选入“变量”列表框中（参考图 7.1）⇒**描述...**→在弹出的“因子分析：描述”对话框中（参见图 7.2），作图中的选择→**继续**；**提取...**→在弹出的“因子分析：提取”对话框中（见图 8.1），作图中的选择→**继续**；**旋转...**→在弹出的“因子分析：旋转”对话框中（见图 8.2），作图中的选择→**继续**；**得分...**→在弹出的“因子分析：得分”对话框中（见图 8.3），作图中的选择→**继续**⇒**确定**，生成图 8.4 和图 8.5。

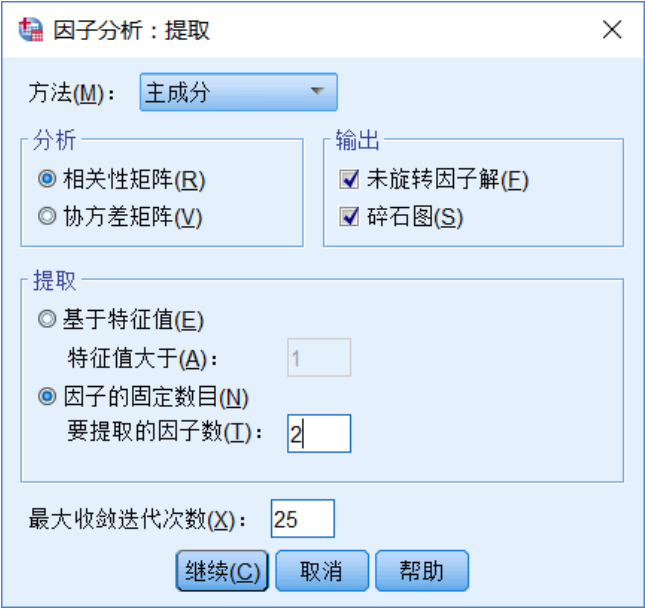


图 8.1

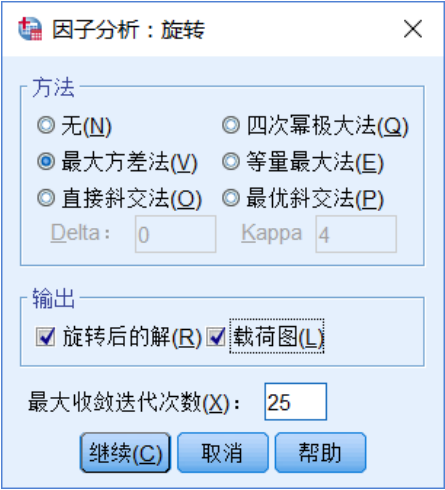


图 8.2

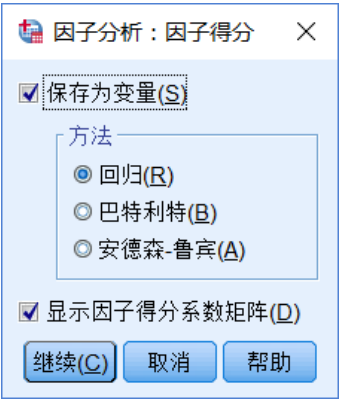


图 8.3

相关性矩阵

		100米（秒）	200米（秒）	400米（秒）	800米（分）	1500米（分）	5000米（分）	10000米（分）	马拉松（分）
相关性	100米（秒）	1.000	.923	.841	.756	.700	.619	.633	.520
	200米（秒）		1.000	.851	.807	.775	.695	.697	.596
	400米（秒）			1.000	.870	.835	.779	.787	.705
	800米（分）				1.000	.918	.864	.869	.806
	1500米（分）					1.000	.928	.935	.866
	5000米（分）						1.000	.975	.932
	10000米（分）							1.000	.943
	马拉松（分）								1.000

公因子方差

	初始	提取
100米（秒）	1.000	.950
200米（秒）	1.000	.939
400米（秒）	1.000	.892
800米（分）	1.000	.900
1500米（分）	1.000	.938
5000米（分）	1.000	.965
10000米（分）	1.000	.973
马拉松（分）	1.000	.943

提取方法：主成分分析法。

总方差解释

成分	总计	初始特征值		提取载荷平方和			旋转载荷平方和		
		方差百分比	累积 %	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	6.622	82.777	82.777	6.622	82.777	82.777	4.186	52.323	52.323
2	.878	10.970	93.747	.878	10.970	93.747	3.314	41.424	93.747
3	.159	1.992	95.739						
4	.124	1.551	97.289						
5	.080	.999	98.288						
6	.068	.850	99.137						
7	.046	.580	99.717						
8	.023	.283	100.000						

提取方法：主成分分析法。

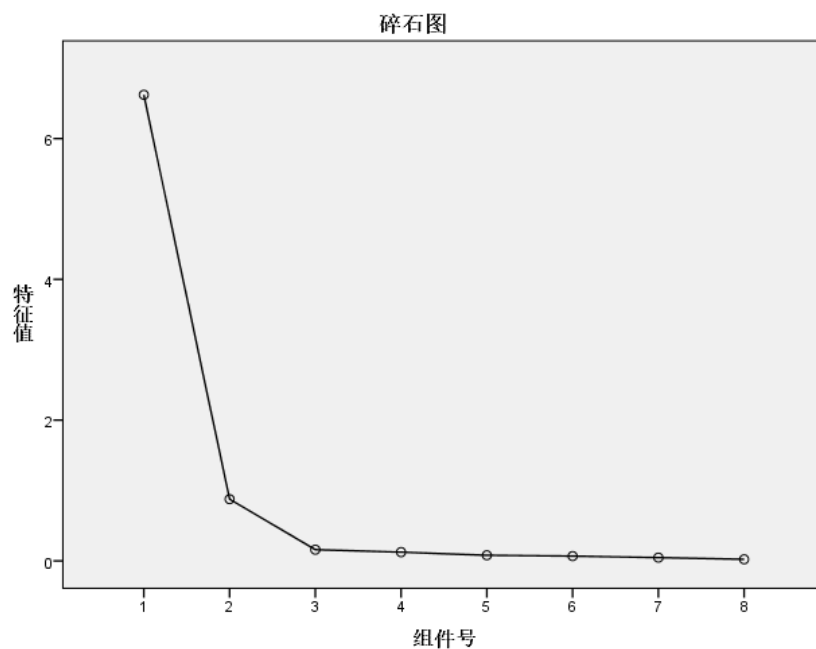


图 8.4

成分矩阵^a

	成分	
	1	2
100米 (秒)	.817	.531
200米 (秒)	.867	.432
400米 (秒)	.915	.233
800米 (分)	.949	.012
1500米 (分)	.959	-.131
5000米 (分)	.938	-.292
10000米 (分)	.944	-.287
马拉松 (分)	.880	-.411

提取方法：主成分分析法。

a. 提取了 2 个成分。

旋转后的成分矩阵^a

	成分	
	1	2
100米 (秒)	.274	.935
200米 (秒)	.376	.893
400米 (秒)	.543	.773
800米 (分)	.712	.627
1500米 (分)	.813	.525
5000米 (分)	.902	.389
10000米 (分)	.903	.397
马拉松 (分)	.936	.261

提取方法：主成分分析法。

旋转方法：凯撒正态化最大方差法。

a. 旋转在 3 次迭代后已收敛。

成分转换矩阵

成分	1	2
1	.759	.651
2	-.651	.759

提取方法：主成分分析法。

旋转方法：凯撒正态化最大方差法。

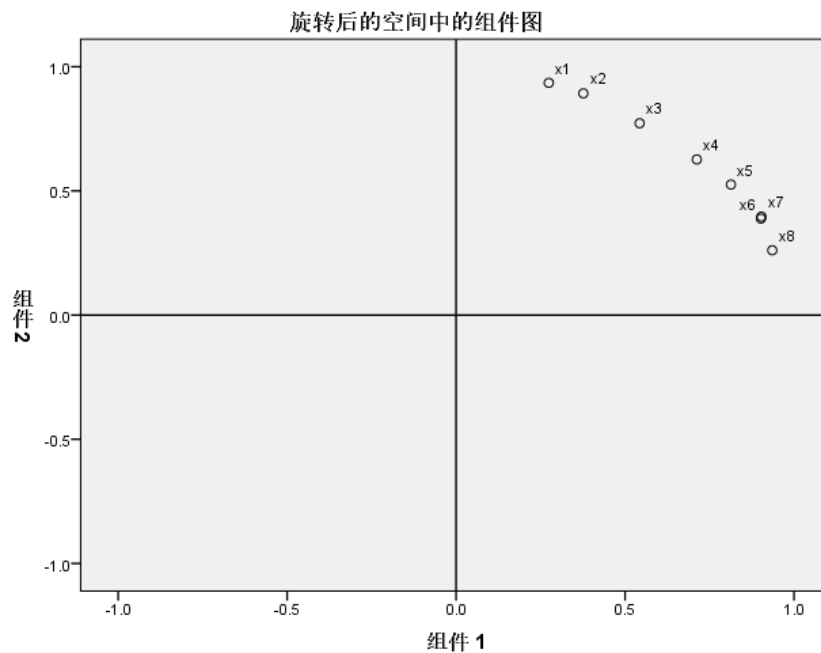


图 8.4 (续 1)

成分得分系数矩阵

	成分	
	1	2
100米 (秒)	-.300	.540
200米 (秒)	-.222	.459
400米 (秒)	-.068	.291
800米 (分)	.100	.103
1500米 (分)	.207	-.019
5000米 (分)	.324	-.161
10000米 (分)	.321	-.156
马拉松 (分)	.406	-.269

提取方法：主成分分析法。
 旋转方法：凯撒正态化最大方差法。
 组件得分。

成分得分协方差矩阵

成分	1	2
1	1.000	.000
2	.000	1.000

提取方法：主成分分析法。
 旋转方法：凯撒正态化最大方差法。
 组件得分。

图 8.4 (续 2)

	x5	x6	x7	x8	FAC1_1	FAC2_1
1	3.70	14.04	29.36	137.72	.31698	-.21309
2	3.57	13.28	27.66	128.30	-.57116	-.79421
3	3.60	13.26	27.72	135.90	-.57664	.19002
4	3.60	13.22	27.45	129.95	-.78406	-.30439
5	3.75	14.68	30.55	146.62	1.44631	-1.24424
6	3.66	13.62	28.62	133.13	-.01382	-.91372
7	3.85	14.45	30.28	139.95	.40270	.70737
8	3.63	13.55	28.09	130.15	-.16717	-.84726
9	3.71	13.61	29.30	134.03	.02763	-.25960
10	3.73	13.90	29.13	133.53	-.12811	.39332
11	3.74	13.49	27.88	131.35	-.46260	.30630
12	4.24	16.70	35.38	164.70	2.06399	3.89349
13	3.84	14.03	28.81	136.58	-.48409	1.93453
14	3.58	13.42	28.19	134.32	-.38475	-.37067
15	3.61	13.50	28.11	130.78	-.59753	.03201
16	3.82	14.91	31.45	154.12	2.20817	-1.54986

图 8.5

2.主因子法

在图 8.1 的“因子分析：提取”对话框中，作图 8.6 中的选择→继续⇒确定，生成图 8.7。

因子分析：提取

方法(M): 主轴因式分解

分析

- ☒ 相关性矩阵
- ☐ 协方差矩阵

提取

- ☐ 基于特征值
- ☒ 因子的固定数目(N)

特征值大于(A): 1

要提取的因子数(I): 2

最大收敛迭代次数(X): 1

输出

- ☒ 未旋转因子解(E)
- ☐ 碎石图(S)

继续(C) 取消 帮助

图 8.6

公因子方差

	初始	提取
100米（秒）	.877	.897
200米（秒）	.888	.906
400米（秒）	.845	.856
800米（分）	.884	.881
1500米（分）	.927	.926
5000米（分）	.955	.960
10000米（分）	.967	.974
马拉松（分）	.905	.907

提取方法：主轴因式分解法。

总方差解释

因子	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	6.622	82.777	82.777	6.530	81.627	81.627	4.079	50.983	50.983
2	.878	10.970	93.747	.778	9.723	91.350	3.229	40.367	91.350
3	.159	1.992	95.739						
4	.124	1.551	97.289						
5	.080	.999	98.288						
6	.068	.850	99.137						
7	.046	.580	99.717						
8	.023	.283	100.000						

提取方法：主轴因式分解法。

图 8.7

因子矩阵^a

	因子	
	1	2
100米（秒）	.807	.496
200米（秒）	.858	.412
400米（秒）	.900	.216
800米（分）	.939	.024
1500米（分）	.956	-.114
5000米（分）	.938	-.282
10000米（分）	.946	-.281
马拉松（分）	.874	-.378

提取方法：主轴因式分解法。

a. 已尝试提取 2 个因子。需要进行 1 次以上的迭代。（收敛性 = .020）。已终止提取。

旋转后的因子矩阵^a

	因子	
	1	2
100米（秒）	.287	.902
200米（秒）	.381	.872
400米（秒）	.541	.751
800米（分）	.695	.631
1500米（分）	.799	.537
5000米（分）	.895	.399
10000米（分）	.900	.405
马拉松（分）	.909	.284

提取方法：主轴因式分解法。

旋转方法：凯撒正态化最大方差法。

a. 旋转在 3 次迭代后已收敛。

因子转换矩阵

因子	1	2
1	.758	.653
2	-.653	.758

提取方法：主轴因式分解法。

旋转方法：凯撒正态化最大方差法。

图 8.7（续）

3.极大似然法

在图 8.1 的“方法”列表框中，选择“最大似然”，其他操作与主成分法类似，其主要结果如图 8.8 所示。

公因子方差		
	初始	提取
100米（秒）	.877	.919
200米（秒）	.888	.924
400米（秒）	.845	.849
800米（分）	.884	.865
1500米（分）	.927	.918
5000米（分）	.955	.966
10000米（分）	.967	.982
马拉松（分）	.905	.914

提取方法：最大似然法。

总方差解释									
因子	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	6.622	82.777	82.777	6.407	80.088	80.088	4.093	51.164	51.164
2	.878	10.970	93.747	.930	11.624	91.712	3.244	40.548	91.712
3	.159	1.992	95.739						
4	.124	1.551	97.289						
5	.080	.999	98.288						
6	.068	.850	99.137						
7	.046	.580	99.717						
8	.023	.283	100.000						

提取方法：最大似然法。

因子矩阵 ^a		
	因子	
	1	2
100米（秒）	.731	.620
200米（秒）	.792	.545
400米（秒）	.855	.343
800米（分）	.916	.161
1500米（分）	.958	.026
5000米（分）	.972	-.144
10000米（分）	.981	-.143
马拉松（分）	.923	-.250

提取方法：最大似然法。

a. 提取了 2 个因子。需要进行 5 次迭代。

图 8.8

拟合优度检验

卡方	自由度	显著性
16.359	13	.230

旋转后的因子矩阵^a

	因子	
	1	2
100米（秒）	.288	.914
200米（秒）	.379	.883
400米（秒）	.541	.746
800米（分）	.689	.624
1500米（分）	.797	.532
5000米（分）	.899	.397
10000米（分）	.906	.402
马拉松（分）	.914	.281

提取方法：最大似然法。

旋转方法：凯撒正态化最大方差法。

a. 旋转在 3 次迭代后已收敛。

因子转换矩阵

因子	1	2
1	.846	.533
2	-.533	.846

提取方法：最大似然法。

旋转方法：凯撒正态化最大方差法。

因子得分系数矩阵

	因子	
	1	2
100米（秒）	-.252	.531
200米（秒）	-.224	.507
400米（秒）	-.055	.169
800米（分）	-.004	.104
1500米（分）	.064	.068
5000米（分）	.361	-.156
10000米（分）	.680	-.288
马拉松（分）	.188	-.143

提取方法：最大似然法。

旋转方法：凯撒正态化最大方差法。

因子评分方法：回归法。

因子得分协方差矩阵

因子	1	2
1	.973	.031
2	.031	.943

提取方法：最大似然法。

旋转方法：凯撒正态化最大方差法。

因子评分方法：回归法。

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.sav 数据表（已通过数据⇒个案加权...将频数[f]作为个案加权系数）
⇒分析⇒降维>⇒对应分析...⇒在“对应分析”对话框中，将心理健康状况[row] 选入“行”列表框中（见图 9.1）⇒定义范围(D)...，出现“对应分析：定义行范围”对话框（见图 9.2）⇒在“最小值”和“最大值”列表框中分别填入 0 和 3⇒更新⇒继续；将父母社会经济地位[column] 选入“列”列表框中⇒定义范围(F)...，出现“对应分析：定义列范围”对话框（见图 9.3）⇒在“最小值”和“最大值”列表框中分别填入 1 和 5⇒更新⇒继续⇒统计...⇒在弹出的“对应分析：统计”对话框中（见图 9.4），作图中的选择⇒继续⇒确定，生成图 9.5。



图 9.1

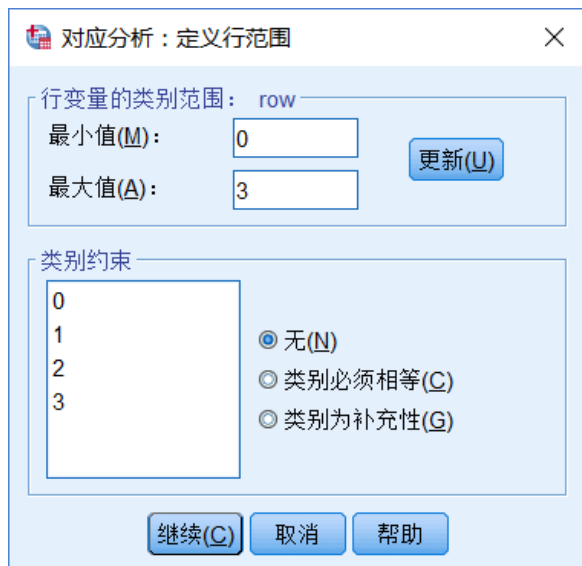


图 9.2

对应分析：定义列范围

列变量的类别范围： column

最小值(M): 1

最大值(A): 5

更新(U)

类别约束

1
2
3
4
5

☒ 无(N)
☐ 类别必须相等(C)
☐ 类别为补充性(G)

继续(C) 取消 帮助

图 9.3

对应分析：统计

☒ 对应表(C)
☐ 行点概述(R)
☐ 列点概述(L)
☐ 对应表的排列(P)
 最大排列维数(M): 1

☒ 行概要(O)
☒ 列概要(U)

以下对象的置信度统计：
☐ 行点(W) ☐ 列点(N)

继续(C) 取消 帮助

图 9.4

对应表

心理健康状况	父母社会经济地位					活动边际
	1	2	3	4	5	
0	121	57	72	36	21	307
1	188	105	141	97	71	602
2	112	65	77	54	54	362
3	86	60	94	78	71	389
活动边际	507	287	384	265	217	1660

行概要

心理健康状况	父母社会经济地位					活动边际
	1	2	3	4	5	
0	.394	.186	.235	.117	.068	1.000
1	.312	.174	.234	.161	.118	1.000
2	.309	.180	.213	.149	.149	1.000
3	.221	.154	.242	.201	.183	1.000
数量	.305	.173	.231	.160	.131	

列概要

心理健康状况	父母社会经济地位					数量
	1	2	3	4	5	
0	.239	.199	.188	.136	.097	.185
1	.371	.366	.367	.366	.327	.363
2	.221	.226	.201	.204	.249	.218
3	.170	.209	.245	.294	.327	.234
活动边际	1.000	1.000	1.000	1.000	1.000	

摘要

维	奇异值	惯量	卡方	显著性	惯量比例		置信度奇异值	
					占	累积	标准差	相关性
1	.161	.026			.947	.947	.023	-.052
2	.037	.001			.050	.998	.025	
3	.008	.000			.002	1.000		
总计		.027	45.594	.000 ^a	1.000	1.000		

a. 12 自由度

行点和列点

对称 正态化

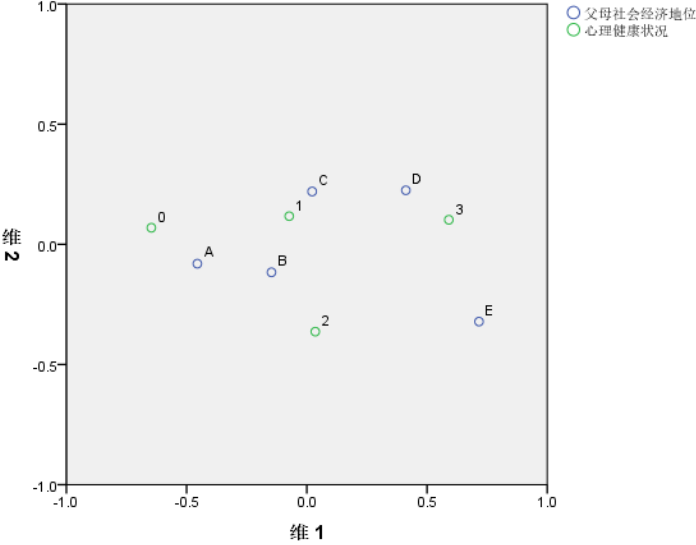


图 9.5

第十章 典型相关分析

一、对书中例 10.3.1 中的数据作典型相关分析

打开 `examp10.3.1.sav` 数据表 ⇒ **分析** ⇒ **相关** ⇒ **典型相关性** ⇒ 在“典型相关性”对话框中，将**体重[x1]**、**腰围[x2]**和**脉搏[x3]**选入“集合 1”列表框，将**引体向上[y1]**、**起坐次数[y2]** 和**跳跃次数[y3]**选入“集合 2”列表框 ⇒ **确定**，生成图 10.2。



图 10.1

典型相关性设置

值	
集合 1 变量	x1 x2 x3
集合 2 变量	y1 y2 y3
集中的数据集	无
评分语法	无
用于评分的相关性	3

典型相关性

	相关性	特征值	威尔克统计	F	分子自由度	分母自由度	显著性
1	.796	1.725	.350	2.048	9.000	34.223	.064
2	.201	.042	.955	.176	4.000	30.000	.949
3	.073	.005	.995	.085	1.000	16.000	.775

H0 for Wilks 检验是指当前行和后续行中的相关性均为零

集合 1 标准化典型相关系数

变量	1	2	3
x1	.775	-1.884	-.191
x2	-1.579	1.181	.506
x3	.059	-.231	1.051

集合 2 标准化典型相关系数

变量	1	2	3
y1	.349	-.376	-1.297
y2	1.054	.123	1.237
y3	-.716	1.062	-.419

图 10.2

集合 1 非标准化典型相关系数

变量	1	2	3
x1	.031	-.076	-.008
x2	-.493	.369	.158
x3	.008	-.032	.146

集合 2 非标准化典型相关系数

变量	1	2	3
y1	.066	-.071	-.245
y2	.017	.002	.020
y3	-.014	.021	-.008

集合 1 典型载荷

变量	1	2	3
x1	-.621	-.772	-.135
x2	-.925	-.378	-.031
x3	.333	.041	.942

集合 2 典型载荷

变量	1	2	3
y1	.728	.237	-.644
y2	.818	.573	.054
y3	.162	.959	-.234

集合 1 交叉载荷

变量	1	2	3
x1	-.494	-.155	-.010
x2	-.736	-.076	-.002
x3	.265	.008	.068

集合 2 交叉载荷

变量	1	2	3
y1	.579	.048	-.047
y2	.651	.115	.004
y3	.129	.192	-.017

图 10.2 (续)