

第 3 章 SPSS 基础知识

SPSS 是世界上最早的统计分析软件,由美国斯坦福大学的三位研究生 Norman H. Nie、C. Hadlai Hull 和 Dale H. Bent 于 1968 年研究开发成功,同时成立了 SPSS 公司,并于 1975 年成立法人组织,在芝加哥组建了 SPSS 总部。1984 年 SPSS 总部推出了世界上第一个统计分析软件微机版本 SPSS/PC+,开创了 SPSS 微机系列产品的开发方向,极大地扩充了它的应用范围,并使其能很快地应用于自然科学、技术科学和社会科学的各个领域。世界上许多有影响的报刊杂志纷纷就 SPSS 的自动统计绘图、数据的深入分析、使用方便和功能齐全等方面给予了高度的评价。

学习目标:

- (1) 零起步学习 SPSS 19.0 软件,了解和掌握 SPSS 基本功能;
- (2) 学习 SPSS 图形绘制,并能结合实例进行应用求解;
- (3) 学习 SPSS 进行回归分析,并能结合实例进行应用求解;
- (4) 学习 SPSS 进行因子分析,并能结合实例进行应用求解;
- (5) 学习 SPSS 进行系统聚类分析,并能结合实例进行应用求解。

3.1 数据的输入和保存

当打开 SPSS 19.0 后,界面如图 3-1 所示。

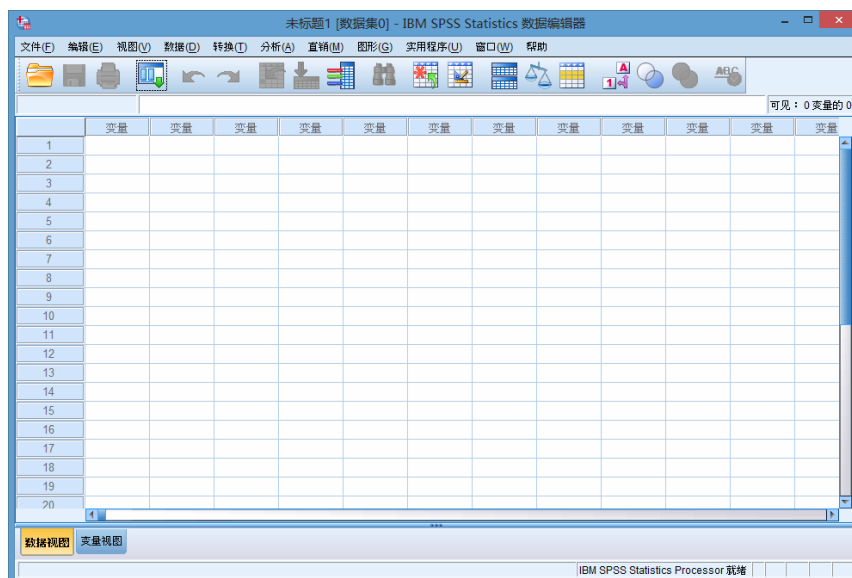


图 3-1 SPSS 19.0 界面

界面包括变量视图和数据视图。变量视图，常用语对变量进行设置，可以写名称、设置数据类型和数据精度等，数据视图则为用户输入的数据。

单击“变量视图”按钮，进行变量的命名、数据精度设置和数据类型设置等等，具体如图 3-2 所示。

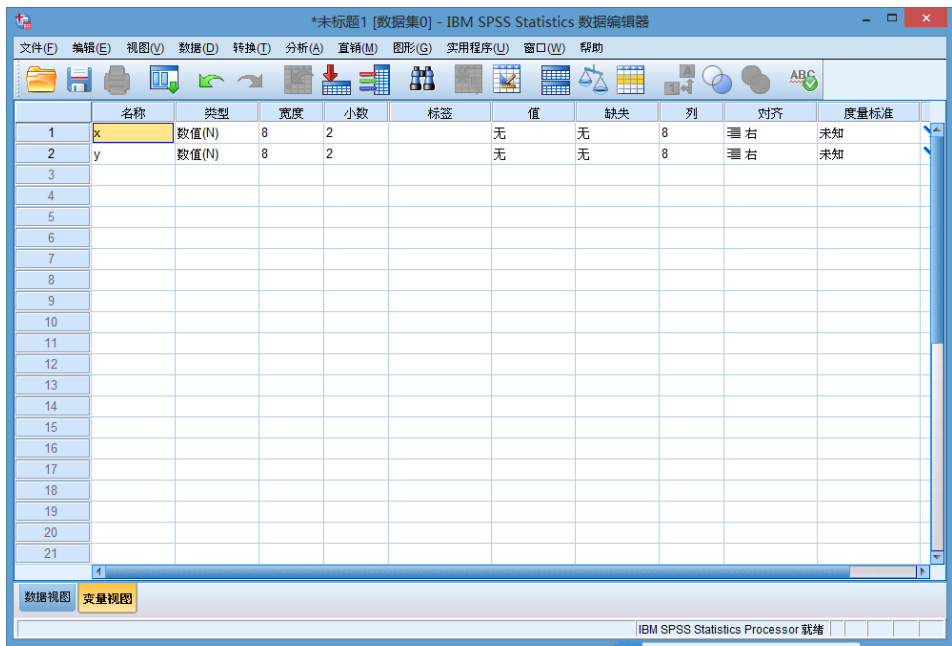


图 3-2 变量命名

单击“保存”按钮，实现该数据分析文件的保存，弹出相应的数据输出窗口，如图 3-3 所示。

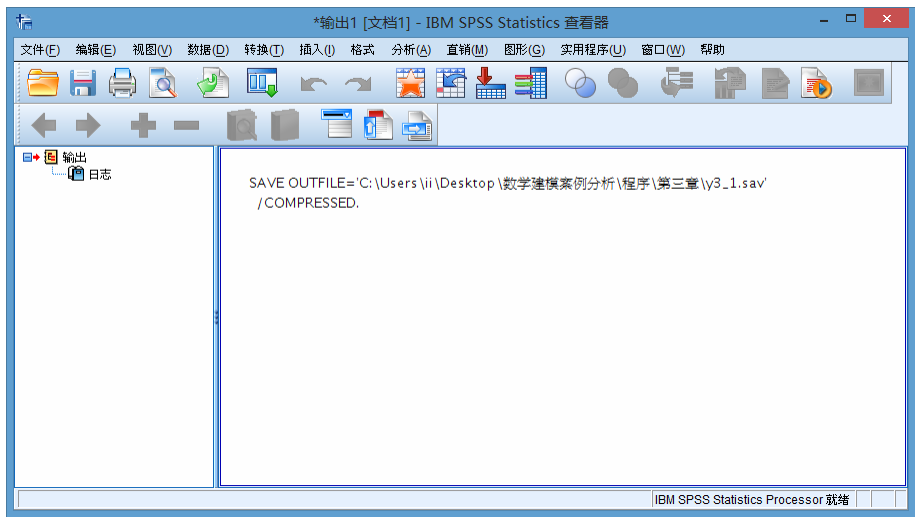


图 3-3 数据输出窗口

在数据视图中，输入相应的 x 和 y 对应的数值，如图 3-4 所示。

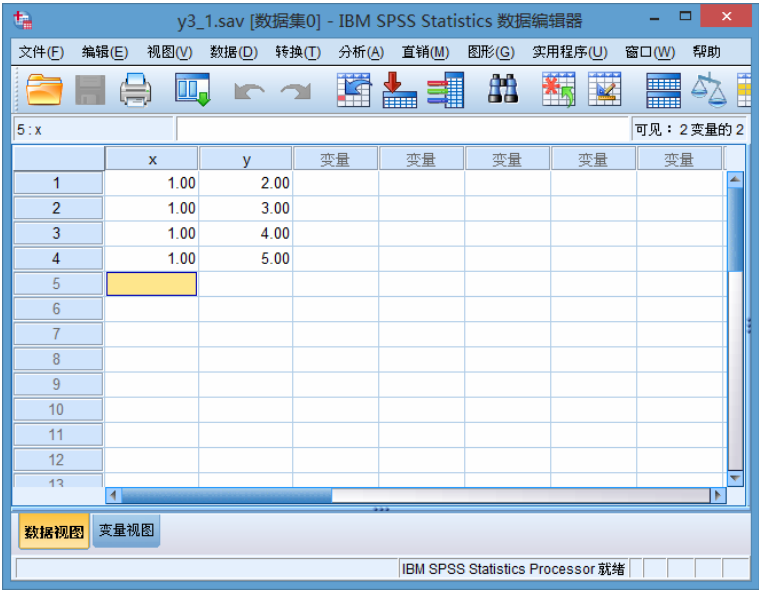


图 3-4 数据输入

SPSS 软件集成了大量的函数调用功能，用户只需要进行简单的单击加载，即可完成数据的分析。SPSS 数据分析下拉菜单如图 3-5 所示。

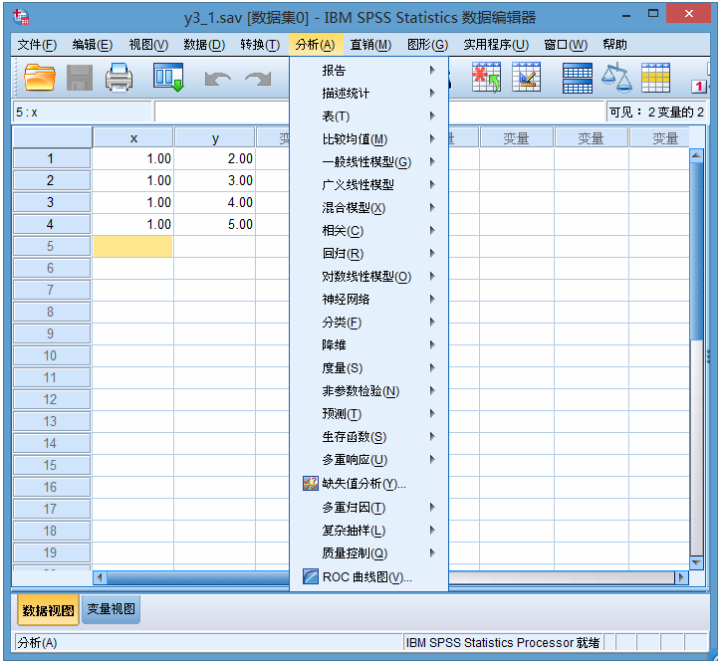


图 3-5 SPSS 数据分析下拉菜单

常用数据分析包括比较均值、一般线性模型、广义线性模型、回归分析、相关分析、神经网络、分类和降维等等。对于常见的数据分析问题，该菜单功能基本能够进行一定的初步分析，具体的回归分析包括自动线性模型、线性、曲线估计、部分最小平方和权重估计等，具体如图 3-6 所示。



图 3-6 回归分析

同样的对于相关分析，主要包括双变量、偏相关和距离相关等，如图 3-7 所示。



图 3-7 相关分析

对于常见的预测分析，SPSS 同样给出序列图、季节性分解和自相关等分析方法，具体如图 3-8 所示。

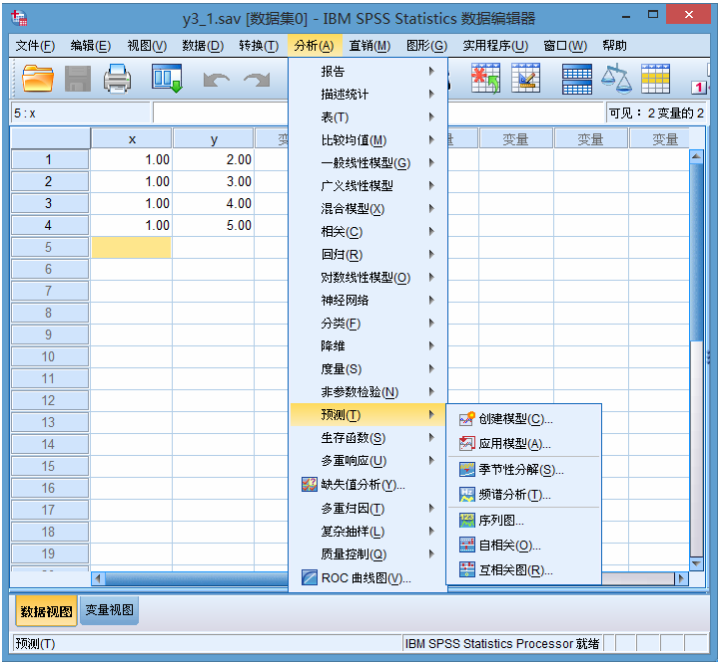


图 3-8 预测分析

对于经济学、评价因子值和得分计算等问题，常涉及到相应的降维算法，即因子分析、对应分析和最优尺度等计算，SPSS 在降维菜单下同样给出了对应的分析方法，如图 3-9 所示。

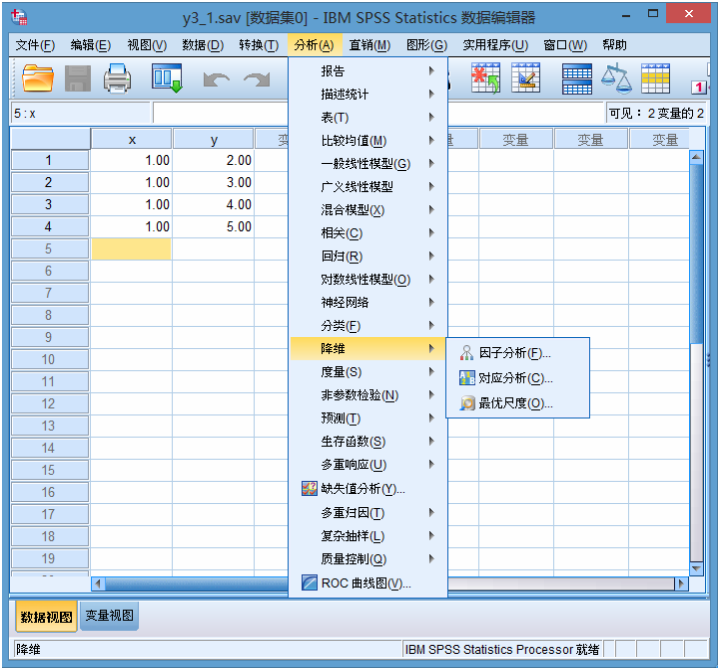


图 3-9 降维分析

SPSS 应用分析方法很多，软件集成功能相当强大，用户可根据自己的需求，进行相应的设置，进行分析求解。

3.2 SPSS 图形绘制

3.2.1 散点图绘制

实际问题常常伴有大量的数据产生，数据本身使得问题变得更加明确和清晰，然而数据的不规则性导致实际分析数据常常遇到一些困难。从表象去分析数据内在的固有属性显得很重要，一方面用户可以根据自身经验知识，从散点图判断数据的属性，另一方面散点图也能提供给用户方法的选择，一次回归、二次曲线或者是指数回归等等，因此散点图绘制尤为重要。SPSS 自带强大的绘图工具箱，用户能够很简便地操作工具箱查看数据属性。

分析数据如表 3-1 所示。

表 3-1 数据表

阅读能力	2480	2810	2910	2750	2530
识字量	71	87	89	86	75
阅读能力	3140	2830	2890	2820	3230
识字量	98	83	90	93	95
阅读能力	2330	2920	2970	2800	2770
识字量	73	99	86	96	80
阅读能力	2870	2540	2930	2080	2990
识字量	82	80	97	71	90
阅读能力	2550	2470	2690	2420	2550
识字量	87	69	82	73	75
阅读能力	2650	2790	2450	2950	2400
识字量	75	82	83	98	64

单击“分析”下拉菜单下的回归选项，选择回归下的曲线估计，即可选择数据散点图选项，具体如图 3-10 所示。

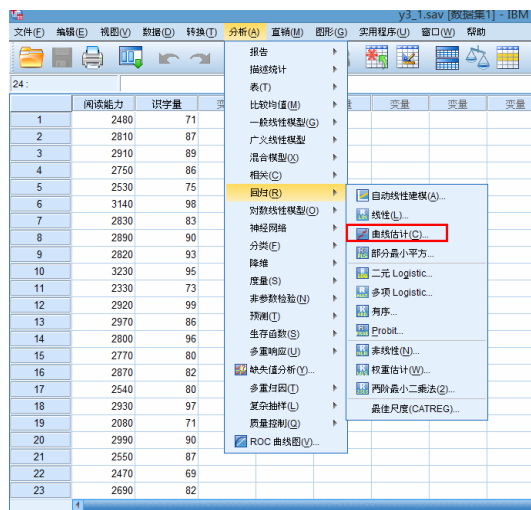


图 3-10 曲线估计选项

单击曲线估计选项，弹出“曲线估计”对话框，如图 3-11 所示。



图 3-11 曲线估计

如图 3-11 所示，系统默认为线性估计，直接单击确定按钮即可生成线性拟合图，如图 3-12 所示。

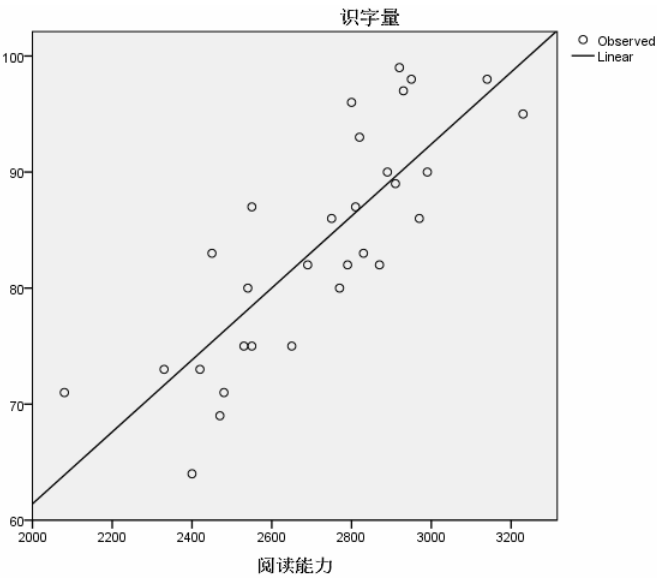


图 3-12 线性拟合图

勾选二次项，单击“确定”按钮，即可生成线性和二次项拟合图，如图 3-13 所示。

图 3-13 中虚线为二次项输出拟合图，实线为直线拟合图。用户可以勾选多个拟合图方法框，如图 3-14 所示的拟合方法选择。则相应的输出结果如图 3-15 所示。

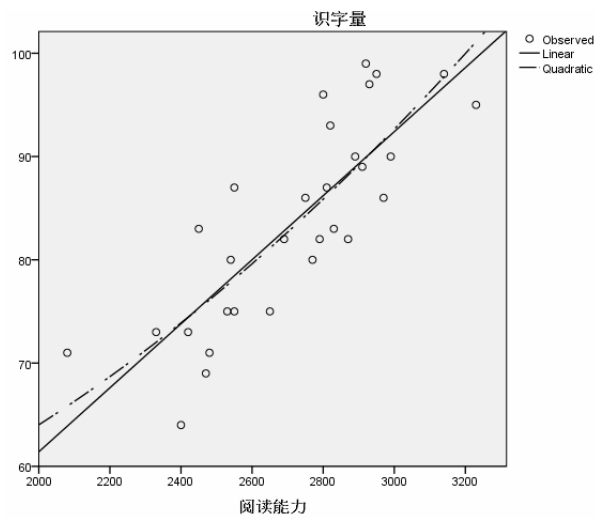


图 3-13 二次拟合图



图 3-14 拟合方法选择

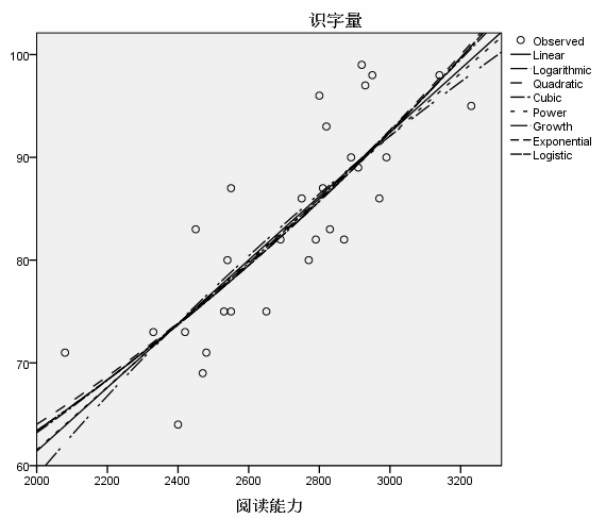


图 3-15 拟合图

由图 3-15 可知，可得到各种方法下的相应的拟合图。用户可以根据相应的拟合好坏程度进行方法的选择，大大地减少了用户的工作量，因此散点图绘制显得尤为重要，SPSS 也相应地为用户提供了更便捷的方法。

3.2.2 直方图绘制

直方图又称质量分布图、柱状图，它是表示资料变化情况的一种主要工具。用直方图可以解析出资料的规则性，比较直观地看出产品质量特性的分布状态，对于资料分布的状况一目了然，便于判断其总体质量分布情况。在制作直方图时，牵涉统计学的概念，首先要对资料进行分组，因此如何合理分组是其中的关键问题。按组距相等的原则进行的两个关键数位是分组数和组距，是一种几何形图表。它是根据从生产过程中收集来的质量数据分布情况，画成以组距为底边、以频数为高度的一系列连接起来的直方型矩形图。

采用如表 3-1 中的数据，进行直方图绘制。

单击“图形”按钮，在该下拉菜单中选择旧对话框|直方图命令，如图 3-16 所示。

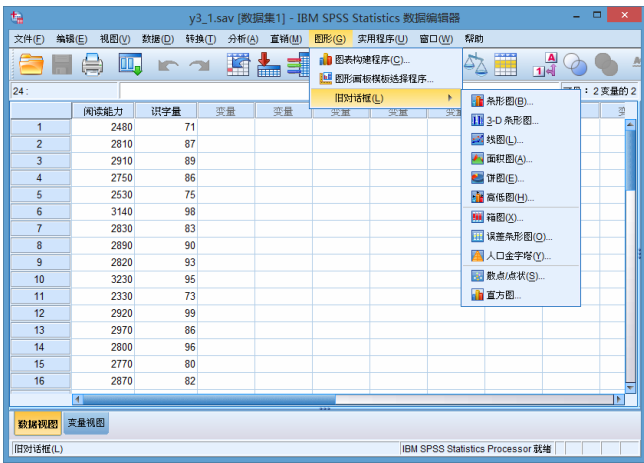


图 3-16 直方图绘制

单击“直方图”按钮后，进行直方图绘制，弹出直方图对话框，如图 3-17 所示。



图 3-17 直方图对话框

勾选“显示正太曲线”选项，单击“确定”按钮，即可得到相应的直方图，如图 3-18 所示。

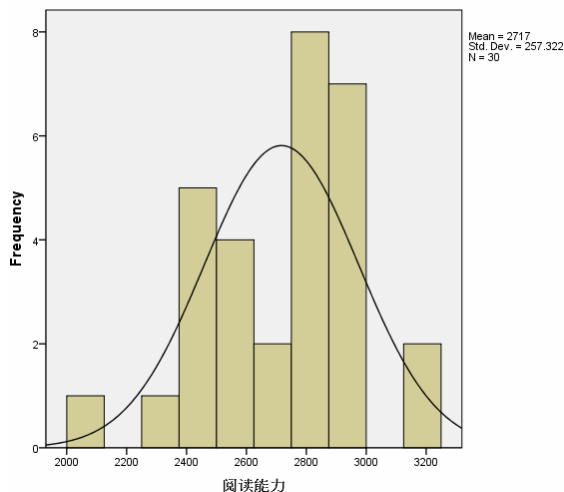


图 3-18 直方图

从图 3-18 中可看出，属于 2400~2600 之间的大约有 9 个数，属于 2000~2200 之间的数有 1 个。因此直方图较为清晰可见。

3.2.3 箱图

一批数据中的异常值值得关注，忽视异常值的存在是十分危险的，不剔除异常值的数据计算分析过程会对结果会带来不良影响。重视异常值的出现，分析其产生的原因，常常成为发现问题进而改进决策的契机。箱形图为用户提供了识别异常值的一个标准。

箱形图（Box-plot）又称为盒须图、盒式图或箱线图，是一种用作显示一组数据分散情况资料的统计图。其绘制须使用常用的统计量，最适宜提供有关数据的位置和分散的参考，尤其在不同的母体数据时其表现更加不同。

具体的箱图选项如图 3-19 所示。

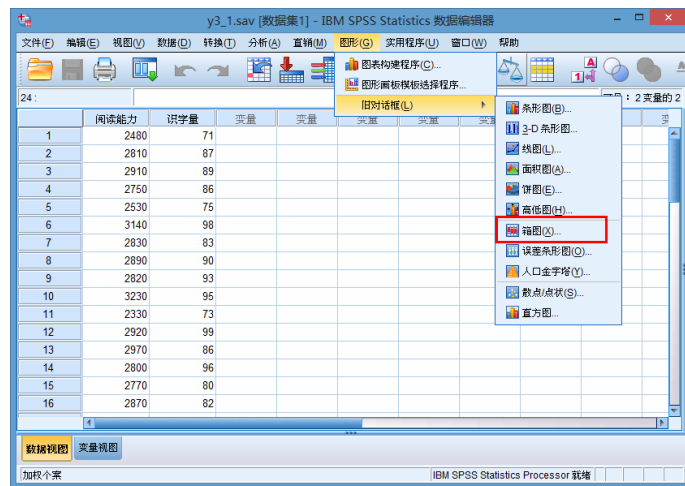


图 3-19 箱图

弹出如图 3-20 所示的对话框，选择简单箱图选项。

在“变量”下的输入框中输入“阅读能力”，在“类别轴”下的输入框中输入“识字量”，具体如图 3-21 所示。

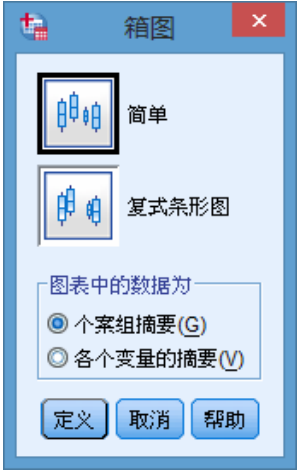


图 3-20 选择简单箱图



图 3-21 定义简单箱图

回到箱图界面，单击“确定”按钮，得到图 3-22。

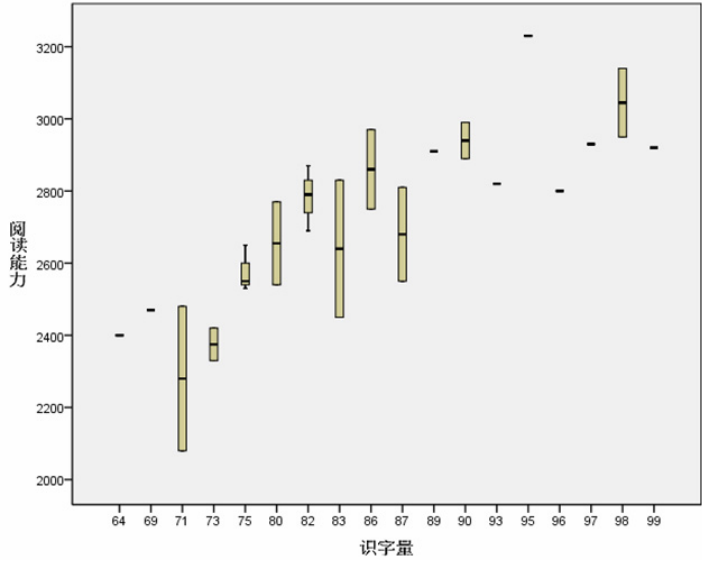


图 3-22 箱图

由图 3-22 可知，属于每一个 case 个案的阅读能力值的区间可以较清晰地看见，用户可以根据这个箱图进行分析阅读能力与识字量之间的说明，较清晰的被用户所接受。

3.3 SPSS 线性回归

回归分析 (Regression Analysis) 是研究一个变量 Y 与其他若干变量 X 之间相关关系的一种数学工具。它是在一组试验或观测数据的基础上, 寻找被随机性掩盖了的变量之间的依存关系。粗略地讲, 可以理解为用一种确定的函数关系去近似代替比较复杂的相关关系。这个函数称为回归函数, 在实际问题中称为经验公式。

回归分析所研究的主要问题就是如何利用变量 X 、 Y 的观察值 (样本), 对回归函数进行统计推断, 包括对它进行估计及检验与它有关的假设等。

在回归按钮下, 有相应的线性回归分析按钮, 如图 3-23 所示。



图 3-23 回归分析

对 x, y 数据进行线性回归分析, 因变量填入 x , 自变量填入 y , 如图 3-24 所示。



图 3-24 回归设置

进行统计量的设置：回归系数置信区间勾选，一般默认为 95%的置信区间， R 方变化求解回归拟合最小误差平方值的变化情况，描述性表明拟合 xy 之间的描述的变量相关性信息值，偏相关性用于检测变量之间的影响因子，共线性诊断用于诊断变量之间的线性关系，如图 3-25 所示。

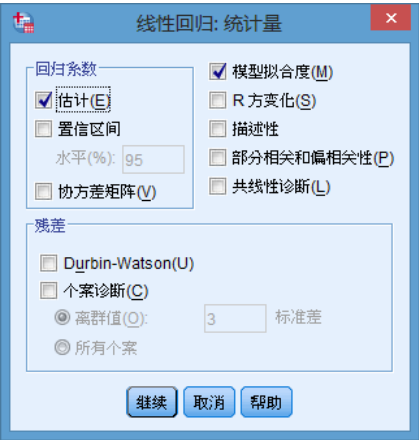


图 3-25 统计量

单击“继续”按钮后，再单击“确认”按钮，即可得到相应的输出，如均方值、方差和变量个数等信息，如表 3-2 所示。

表 3-2 统计特征描述

	Mean	Std. Deviation	N
阅读能力	2717.00	257.322	30
识字量	83.63	9.729	30

从表 3-2 可知，阅读能力总共有 30 个数据，均值为 2717.00，标准差为 257.322，同样识字量均值为 83.63，标准差为 9.729。

相应的相关性系数矩阵如表 3-3 所示。

表 3-3 相关系数

		阅 读 能 力	识 字 量
Pearson Correlation	阅读能力	1.000	.820
	识字量	.820	1.000
Sig. (1-tailed)	阅读能力	.	.000
	识字量	.000	.
N	阅读能力	30	30
	识字量	30	30

在表 3-3 中，由皮尔逊相关系数，可知：

$$\begin{bmatrix} & \text{阅读能力} & \text{识字量} \\ \text{阅读能力} & [1.000 & 0.820] \\ \text{识字量} & [0.820 & 1.000] \end{bmatrix}$$

数据与数据本身相关性为 1，阅读能力与识字量相关系数为 0.820，存在线性相关性。相应的回归模型参数表如表 3-4 所示。

表 3-4 回归模型参数表

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.820a	.672	.661	149.921	.672	57.434	1	28	.000

a. Predictors: (Constant), 识字量
b. Dependent Variable: 阅读能力

表 3-4 中， $R = 0.820$ ， $R^2 = 0.672$ ，自适应 $R^2 = 0.661$ ，累计误差 $e = 149.921$ ，统计量 $F = 57.434$ ，统计量 F 显著性水平 $p = 0.000$ 。因此该自变量和因变量之间存在线性相关。分析得到的均方表如表 3-5 所示。

表 3-5 均方表

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1290894.994	1	1290894.994	57.434	.000a
	Residual	629335.006	28	22476.250		
	Total	1920230.000	29			

a. Predictors: (Constant), 识字量
b. Dependent Variable: 阅读能力

如表 3-5 所示，回归量平方和为 1290894.994，分解后留数平方和为 629335.006，回归后的模型的自由度为 1，残留样本值的自由度为 28，累计样本自由度为 29，统计量 $F = 57.434$ ，统计量 F 显著性水平 $p = 0.000$ 。

相应回归系数如表 3-6 所示。

表 3-6 回归模型系数表

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations		
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part
1	(Constant)	903.338	240.877		3.750	.001	409.924	1396.752			
	识字量	21.686	2.861	.820	7.579	.000	15.824	27.547	.820	.820	.820

由表 3-6 可知，阅读能力与识字量的关系为：

$$y = ax + b = 21.686x + 903.338$$

系数 a 的计算标准差为 2.861，常数 b 的计算标准差为 240.877，系数 a 的置信区间为 $a \in [15.824, 27.547]$ ，常数 b 的置信区间为 $b \in [409.924, 903.338]$ ，该模型参数合理。

为了得到自变量与其他自变量之间是否存在共线问题，则需要共线性诊断。相应的得到共线性诊断如表 3-7 所示。

表 3-7 共线性诊断

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	识字量
1	1	1.994	1.000	.00	.00
	2	.006	17.543	1.00	1.00

a. Dependent Variable: 阅读能力

从表 3-7 中可知，识别量自变量与因变量阅读能力之间存在共线性，特征值为 1.994，大于 1，因此特征较明显。对于二次方关系，特征矢量为 0.006，远远小于 1，因此，不存在高次幂相关。

为了得到系统拟合图曲线模型与实际数值之间的关系，用留数统计表衡量，如表 3-8 所示。

表 3-8 留数统计量

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2291.23	3050.24	2717.00	210.983	30
Residual	-363.035	266.504	.000	147.313	30
Std. Predicted Value	-2.018	1.579	.000	1.000	30
Std. Residual	-2.422	1.778	.000	.983	30

a. Dependent Variable: 阅读能力

模型预测值最大 3050.24，最小值为 2291.23，留数残差最小值为-363.035，最大值为 266.504。

回归模型残差值直方图如图 3-26 所示。

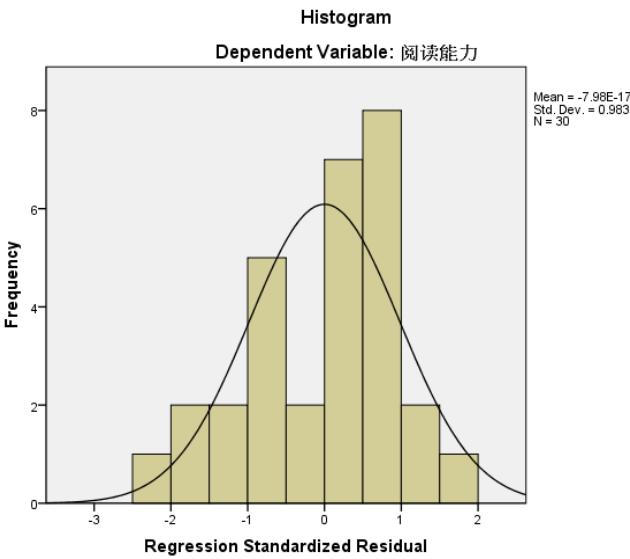


图 3-26 直方图

从图 3-26 中可知，拟合直方图服从正态分布，相应的测量值和实际值位于置信区间内，因此模型较合理。

表 3-9 某企业绩效管理评价指标体系的设计

序 号	指 标 名 称	序 号	指 标 名 称
1	工作层级	2	绩效了解程度
3	绩效合理性	4	考核增加合作
5	考核可行性	6	考核目的发放奖金
7	考核目的确定目标	8	考核目的检查完成情况
9	考核目的职务晋升	10	考核目的培训需求
11	考核目的提高绩效	12	考核目的职业生涯规划
13	考核目的实现战略目标	14	考核目的改编组织文化
15	考核反馈看法	16	考核员工本人
17	考核直接领导	18	考核部门所有员工
19	考核合作员工	20	考核全体员工
21	领导考核成绩	22	绩管制度加强体系培训
23	绩管制度加强评分公式	24	绩管制度加强绩效面谈
25	绩管制度加强申诉渠道	26	绩管制度加强表彰奖励
27	绩效评价业绩程度	28	考核衡量绩效
29	清楚考核要求	30	考核透明度
31	相关部门内考核公平性	32	相对其他部门考核公平性
33	考核促进工作数	34	考核执行情况
35	管理者绩效评价	36	管理者和自己考核结果吻合度
37	对自己考核结果	38	领导和员工对考核结果的沟通
39	领导对员工建议的态度	40	考核结果的应用
41	绩效表述的观点	42	考核内容形成过程
43	哪种考核适合企业总厂	44	适合企业考核方法

输入数据到 SPSS 中，如图 3-28 所示。

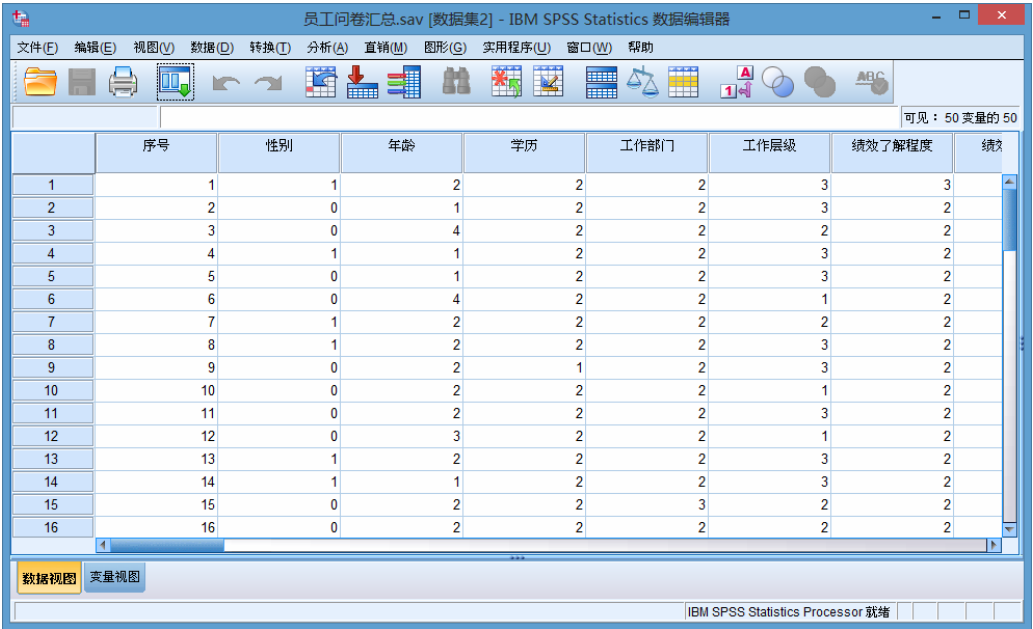


图 3-28 数据输入

单击分析下拉菜单下的降维操作，即可选中因子分析进行因子分析，如图 3-29 所示。

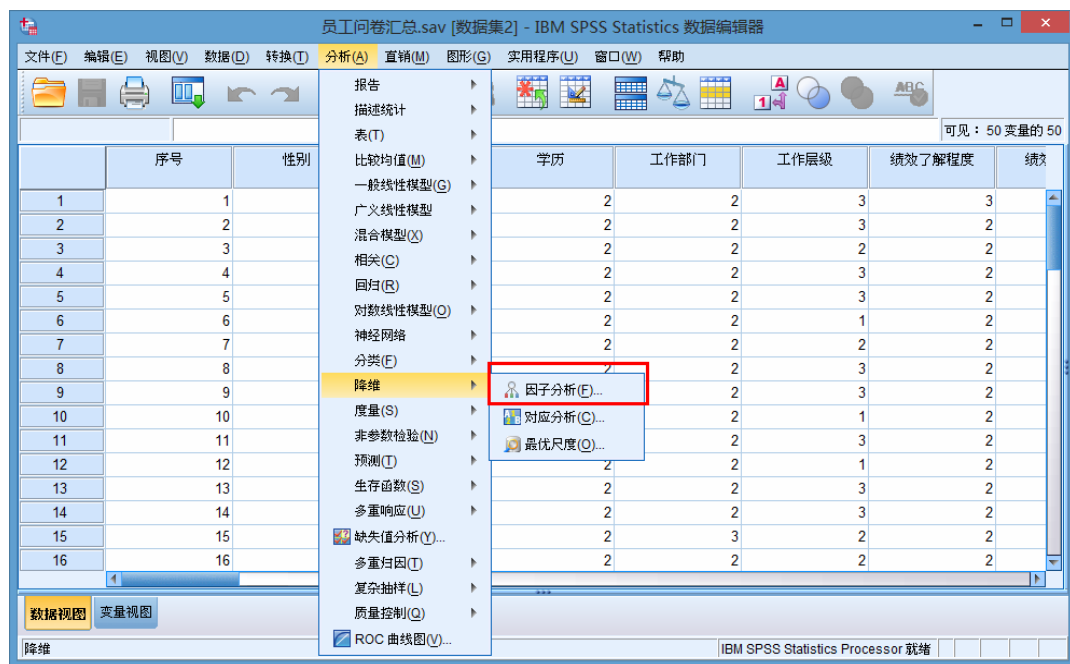


图 3-29 因子分析选项

选择“因子分析”选项，弹出因子分析对话框，如图 3-30 所示。

选择所有的数据项输入到变量框中，单击描述按钮，弹出“描述统计”对话框，勾选“KMO 和 Bartlett 的球形度检验”选框，如图 3-31 所示。



图 3-30 因子分析

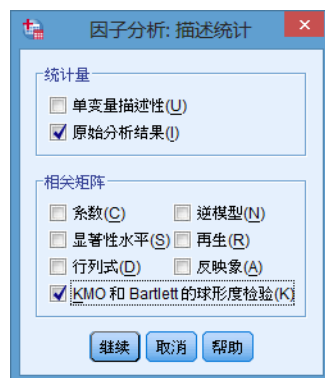


图 3-31 KMO 和 Bartlett 的球形度检验

单击图 3-31 中的“继续”按钮，弹出图 3-30 所示的对话框，单击“抽取”按钮，弹出“抽取”对话框，勾选“碎石图”选框，选择主成分分析方法，如图 3-32 所示。

单击图 3-32 中的“继续”按钮，弹出图 3-30 所示的对话框，单击“得分”按钮，弹出“因子得分”对话框，勾选“显示因子得分系数矩阵”选框，如图 3-33 所示。

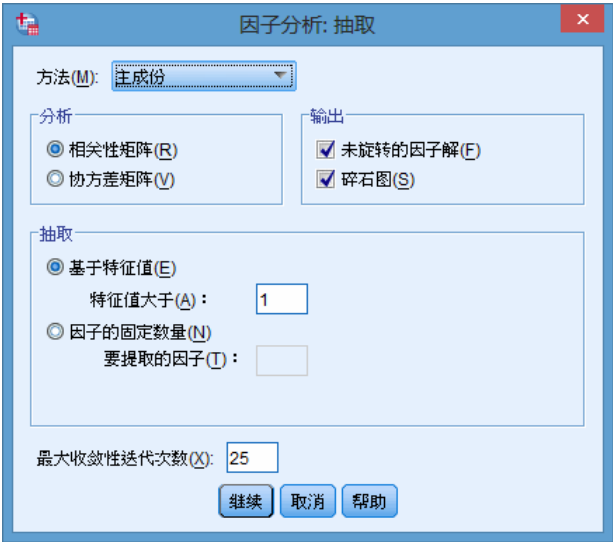


图 3-32 抽取对话框



图 3-33 因子得分

单击因子分析确定按钮，即可得到相应的结果。
得到自变量碎石图如图 3-34 所示。

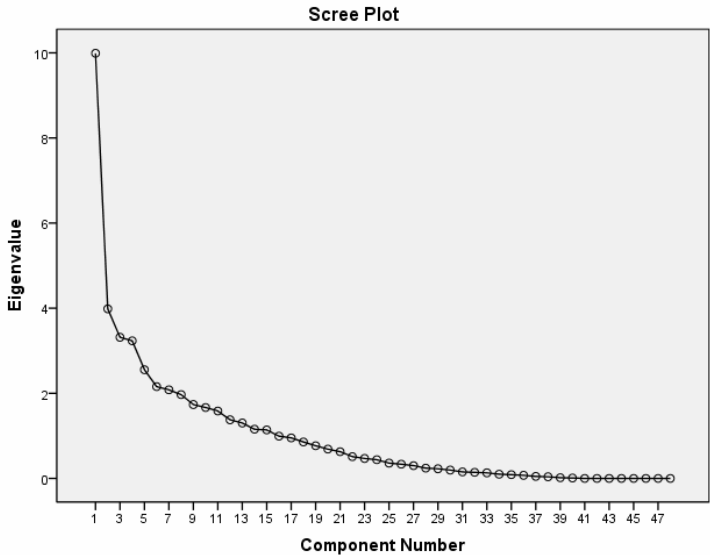


图 3-34 员工数据汇总因子分析碎石图

相应的数据汇总方差中旋转后的贡献率如表 3-10 所示。

表 3-10 员工数据汇总方差中旋转后的贡献率（%of Variance）表（前 15 个公因子）

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	%of Variance	Cumulative %	Total	%of Variance	Cumulative %
1	9.989	20.811	20.811	9.989	20.811	20.811
2	3.984	8.301	29.112	3.984	8.301	29.112
3	3.317	6.910	36.022	3.317	6.910	36.022

续表

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
4	3.233	6.735	42.756	3.233	6.735	42.756
5	2.554	5.321	48.078	2.554	5.321	48.078
6	2.157	4.493	52.571	2.157	4.493	52.571
7	2.081	4.336	56.907	2.081	4.336	56.907
8	1.969	4.102	61.008	1.969	4.102	61.008
9	1.735	3.616	64.624	1.735	3.616	64.624
10	1.666	3.471	68.095	1.666	3.471	68.095
11	1.584	3.301	71.396	1.584	3.301	71.396
12	1.376	2.868	74.263	1.376	2.868	74.263
13	1.301	2.711	76.975	1.301	2.711	76.975
14	1.156	2.408	79.382	1.156	2.408	79.382
15	1.139	2.373	81.755	1.139	2.373	81.755
16	0.999	2.080	83.944			

通过该企业员工和领导层进行调查得到相应的数据,按照提取公因子所解释的方差累计贡献率占整个方差 81%以上的原则,借助碎石图,确定公因子。

针对员工问卷汇总,经 SPSS 因子分析得到 15 个公因子所揭示的方差累计贡献率达到了 81.755%,较好地反映了员工问卷原始指标的信息。经旋转后,15 个公因子所对应的方差贡献率分别为:20.811%、8.301%、6.910%、6.735%、5.321%、4.493%、4.336%、4.102%、3.616%、3.471%、3.301%、2.868%、2.711%、2.408%和 2.373%。

各因子提取量如表 3-11 所示。

表 3-11 各因子提取量

	Initial	Extraction
序号	1.000	.816
考核能衡量绩效	1.000	.858
清楚考核要求	1.000	.772
相对部门内考核公平性	1.000	.823
相对其他部门考核公平性	1.000	.871
考核促进工作数和质	1.000	.797
考核执行情况	1.000	.857
管理者绩效评价	1.000	.833
对自己考核结果	1.000	.887
领导和员工对考核结果的沟通	1.000	.861
领导对员工建议的态度	1.000	.904
考核结果的应用情况	1.000	.778
绩管表述的观点	1.000	.735
考核内容形成过程	1.000	.868
绩管评价业绩程度	1.000	.802
适合瀑电考核方法	1.000	.894
哪种考核适合瀑电总厂	1.000	.844
考核直接领导	1.000	.785

续表

	Initial	Extraction
考核目的发放奖金	1.000	.810
考核目的确定目标	1.000	.824
考核可行性	1.000	.835
考核增加合作	1.000	.772
考核目的职务晋升	1.000	.873
考核目的检查完成情况	1.000	.765
管理者和自己考核结果吻合度	1.000	.846
年龄	1.000	.803
学历	1.000	.861
工作部门	1.000	.799
工作层级	1.000	.740
绩效了解程度	1.000	.771
绩效合理性	1.000	.847
考核目的培训需求	1.000	.759
考核透明度	1.000	.787
考核合作员工	1.000	.869
考核目的提高绩效	1.000	.861
考核目的职业生涯规划	1.000	.852
考核目的实现战略目标	1.000	.768
考核目的改编组织文化	1.000	.731
考核反馈看法	1.000	.921
考核员工本人	1.000	.837
考核部门所有员工	1.000	.823
考核全体员工	1.000	.846
领导占考核成绩比例	1.000	.883
绩管制度加强体系培训	1.000	.791
绩管制度加强评分公示	1.000	.730
绩管制度加强绩效面谈	1.000	.838
绩管制度加强申诉渠道畅通	1.000	.647
绩管制度加强表彰奖励	1.000	.823

在表 3-11 中, SPSS 因子分析得 15 个因子可以充分反映原指标的信息。该因子分析有, 领导对员工建议的态度: 0.914; 考核反馈看法: 0.902; 领导占考核成绩比例: 0.895; 相对其他部门考核公平性: 0.890; 适合瀑电考核方法: 0.886; 考核目的确定目标: 0.885; 哪种考核适合瀑电总厂: 0.883; 考核合作员工: 0.875; 考核执行情况: 0.868; 考核能衡量绩效: 0.863; 领导和员工对考核结果的沟通: 0.857; 绩管制度加强绩效面谈: 0.852; 考核全体员工: 0.849; 考核内容形成过程: 0.846; 学历: 0.846。

根据 SPSS 得到的因子值及公式 $\eta_i = \sum \alpha_p f_{pi}$, 计算出该企业绩效管理的综合得分。其中, η_i 是第 i 个企业的综合得分, α_p 表示总方差中旋转后的贡献率(%of Variance), 即为第 p 个主成分所解释的方差比例, f_{pi} 为因子值。得到, $\bar{\eta}_{\text{员工}}=0.7696$ 。

对于中高层调查结果分析, 按照提取公因子所解释的方差累计贡献率占整个方差 84% 以上的原则, 得到碎石图如图 3-35 所示。

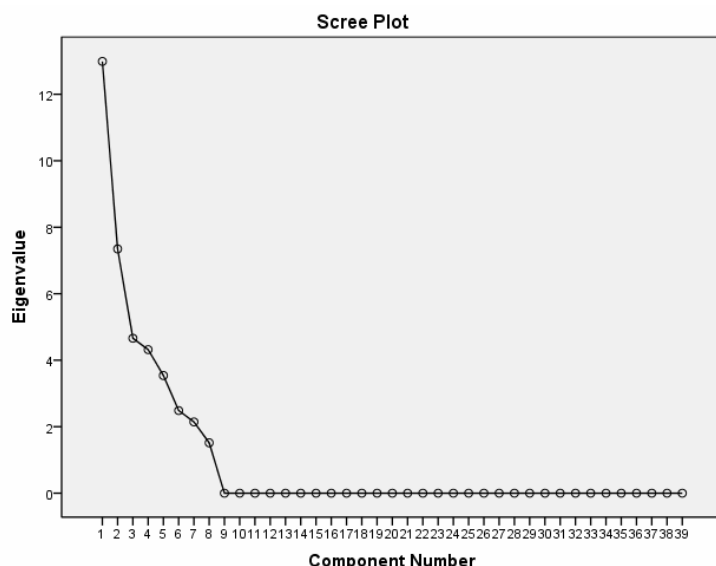


图 3-35 中高层领导数据汇总因子分析碎石图

在图 3-35 中，确定提取出 5 个公因子。5 个公因子所揭示的方差累计贡献率达到了 84.241，较好地反映了原始指标的信息。旋转后：5 个公因子所对应的方差贡献率分别为：33.302%、18.840%、11.951%、11.072%和 9.076%。同理可得到 $\bar{\eta}_{\text{中高层}}=0.6099$ 。

由以上分析可知，该调查数据显示，企业员工绩效管理得分 0.7696 高于中高层领导绩效管理的 0.6099，差别不是很大，因此该企业绩效管理有一定的效果，员工积极性很高。但是也应该注意到，中高层领导应该加强自身建设，合理的应用绩效管理，让绩效管理更加贴近员工本身，提升企业绩效。

3.5 SPSS 系统聚类分析

聚类分析指将物理或抽象对象的集合分组成为由类似的对象组成的多个类的分析过程，它是一种重要的人类行为。聚类分析的目标就是在相似的基础上收集数据来分类。聚类源于很多领域，包括数学、计算机科学、统计学、生物学和经济学。在不同的应用领域，很多聚类技术都得到了发展，这些技术方法被用作描述数据，衡量不同数据源间的相似性，以及把数据源分类到不同的簇中。

聚类与分类的不同在于，聚类所要求划分的类是未知的。

聚类是将数据分类到不同的类或者簇这样的一个过程，所以同一个簇中的对象有很大的相似性，而不同簇间的对象有很大的相异性。

从统计学的观点看，聚类分析是通过数据建模简化数据的一种方法。传统的统计聚类分析方法包括系统聚类法、分解法、加入法、动态聚类法、有序样品聚类、有重叠聚类和模糊聚类等。采用 k-均值和 k-中心点等算法的聚类分析工具已被加入到许多著名的统计分析软件包中，如 SPSS 和 SAS 等。

对表 3-1 中的数据进行聚类分析，每一行表示一个人，一个人的阅读能力和识字量，

进行系统聚类分析，首先载入数据，如图 3-36 所示。

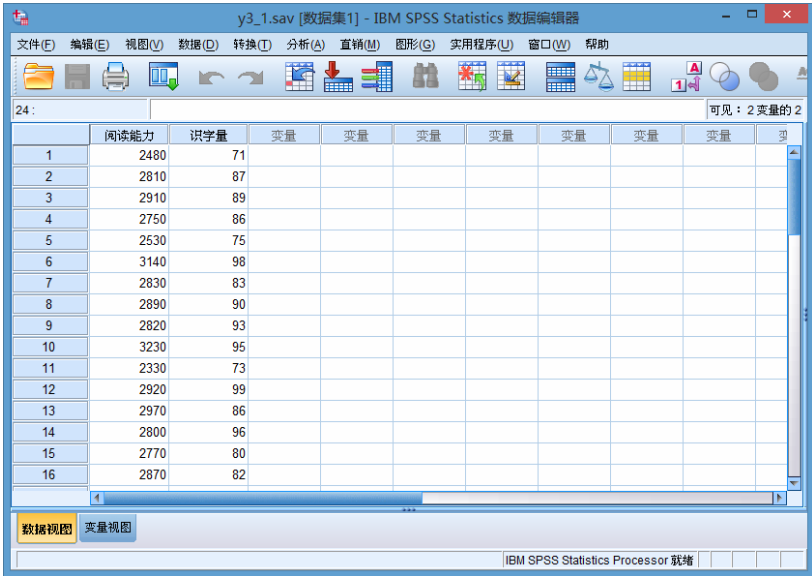


图 3-36 数据载入

选择“分析”|“分类”|“系统聚类”命令，如图 3-37 所示。

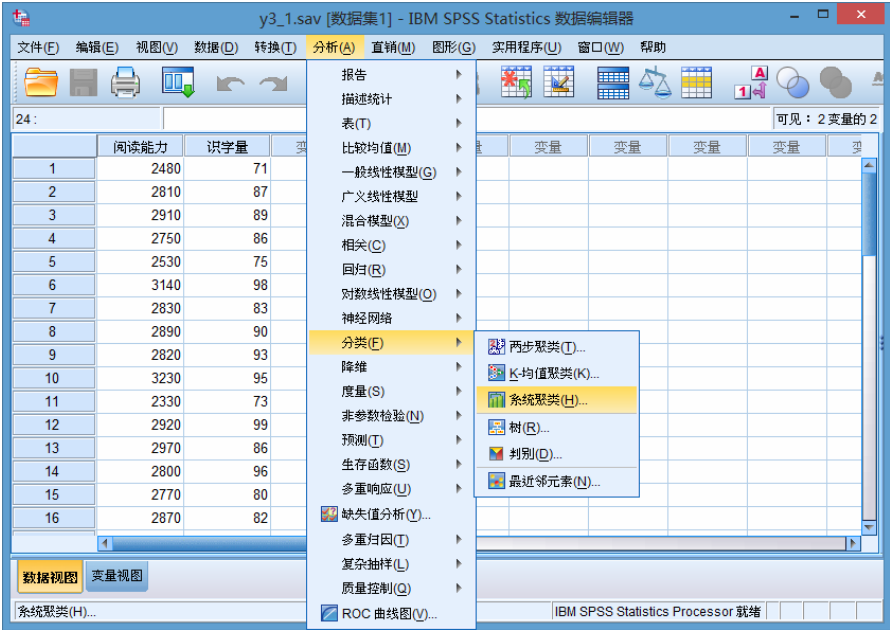


图 3-37 系统聚类选项

单击“系统聚类”按钮后，系统默认为个案分析，即认为数据界面每一行是一个个案，相应的数据即为 30 个个案，选择阅读能力和识字量进行变量窗口，如图 3-38 所示。

单击“绘制”按钮，弹出“系统聚类分析图”对话框，勾选“树状图”选项，如图 3-39 所示。



图 3-38 系统聚类对话框

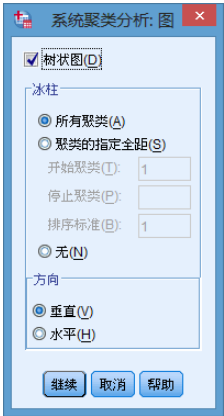


图 3-39 树状图

单击图 3-39 中的“继续”按钮，弹出图 3-38 所示的对话框，单击“方法”按钮，弹出“系统聚类分析方法”对话框，选择聚类方法——组间联接，如图 3-40 所示。



图 3-40 系统聚类分析方法

回到系统聚类分析界面，单击“确定”按钮，即可得到分析结果。
如表 3-12 所示为待分析的数据的统计量，30 个个案，每一个个案都是明显的、有效的。

表 3-12 个案统计量

Case Processing Summary ^{a, b}					
Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
30	100.0	0	.0	30	100.0

a. Squared Euclidean Distance Undefined error #14704 - Cannot open tex

b. Average Linkage (Between Groups)

相应的得到的冰柱图如图 3-41 所示。

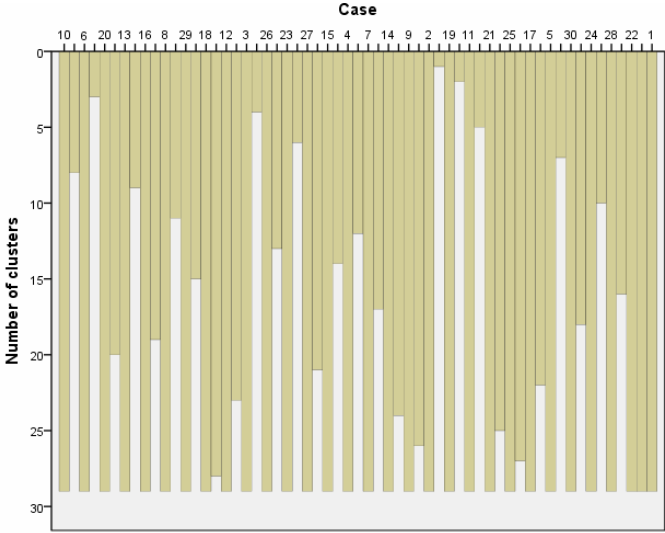


图 3-41 冰柱图

在图 3-41 中，白色的越矮，表示两者之间结合越紧密。例如 12 和 18 个案，2、9 和 14 等个案均是结合紧密的。冰柱图较难看，一般常常采用树状图进行分析，由此得到的树状图如图 3-42 所示。

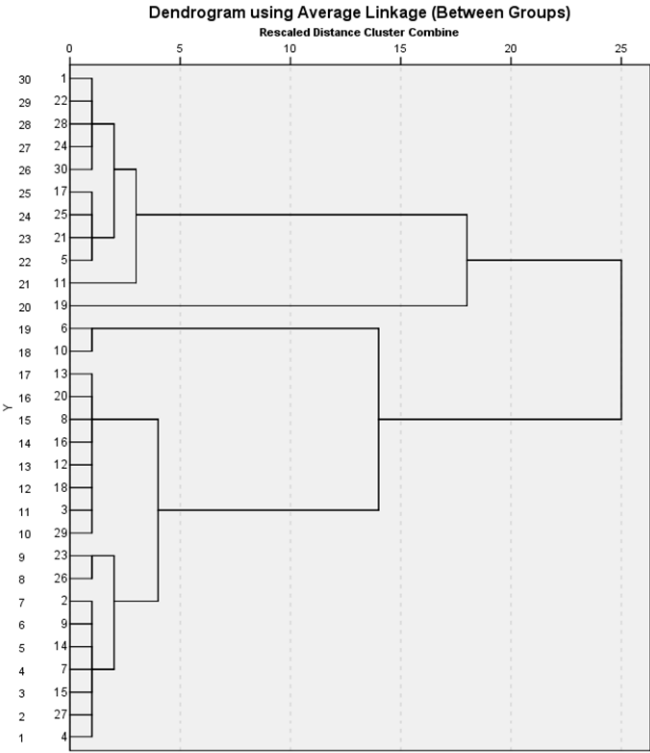


图 3-42 树状图

从图 3-42 中可清楚地看到系统聚类的结果，从每一级来看，系统可聚类为不同的数目，划分的细密些有（2、9、14、7、15、27、4）、（23、26）、（29、3、18、12、16、8、

20、13)、(10、6)、(11)、(19)、(5、21、27、17)、(1、22、28、24、30) 八类，也可以划分为(2、9、14、7、15、27、4、23、26)、(29、3、18、12、16、8、20、13)、(10、6)、(11)、(19)、(5、21、27、17、1、22、28、24、30) 六类；也可以更宽泛地划分为：(2、9、14、7、15、27、4、23、26、29、3、18、12、16、8、20、13)、(10、6)、(19)、(11、5、21、27、17、1、22、28、24、30) 四类。选择组内联接，如图 3-43 所示。



图 3-43 组内聚类

单击“确定”按钮，得到如图 3-44 所示的树状图。

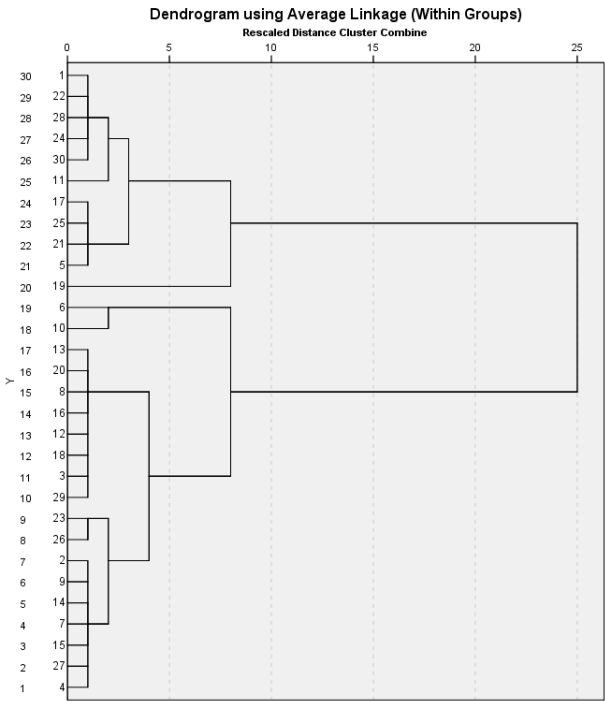


图 3-44 组内聚类树状图

从图 3-44 中可清楚地看到系统聚类的结果,从每一级来看,系统可聚类为不同的数目,划分的细密些有 (2、9、14、7、15、27、4、23、26)、(29、3、18、12、16、8、20、13)、(10、6)、(19)、(5、21、27、17)、(1、22、28、24、30) 6 类;也可以划分为 (10、6)、(19)、(2、9、14、7、15、27、4、23、26、29、3、18、12、16、8、20、13)、(5、21、27、17、1、22、28、24、30) 四类;也可以更宽泛地划分为 (19、5、21、27、17、1、22、28、24、30)、(10、6、2、9、14、7、15、27、4、23、26、29、3、18、12、16、8、20、13、11、5、21、27、17、1、22、28、24、30) 两类。

选择最近邻元素进行聚类,如图 3-45 所示。



图 3-45 最近邻聚类

单击“确定”按钮,得到如图 3-46 所示的树状图。

从图 3-46 中可清楚地看到系统聚类的结果,从每一级来看,系统可聚类为不同的数目,划分的细密些有 (2、9、14、7、15、27、4、29、3、18、12、16、8、20、13)、(10、6)、(11)、(19)、(5、21、27、17、1、22、28、24、30)、(23、26) 六类;也可以划分为 (11、5、21、27、17、1、22、28、24、30)、(10、6)、(19)、(2、9、14、7、15、27、4、29、3、18、12、16、8、20、13、23、26) 四类;也可以更宽泛地划分为 (10、6)、(19)、(2、9、14、7、15、27、4、23、26、29、3、18、12、16、8、20、13、11、5、21、27、17、1、22、28、24、30) 三类。

对比图 3-42、图 3-44 和图 3-46 可知,采用不同的聚类算法,得到的聚类结果有差异,然而主要的大类是没变化的,在此聚类中,(10、6)、(19) 这两类一直比较异常,和其他个案相差很远,常常作为独立空间存在,这也可以作为异常个案进行数据预处理,然后再进行系统聚类。

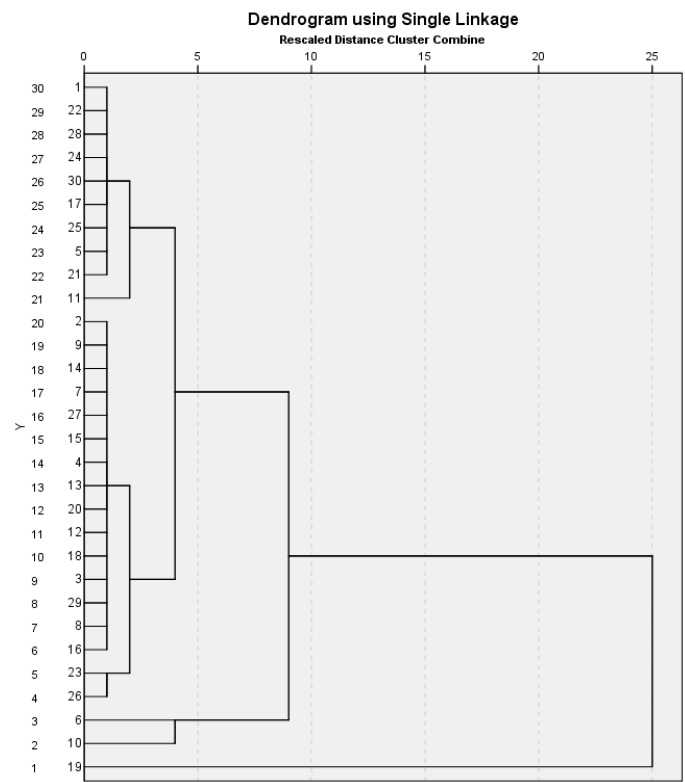


图 3-46 最近邻聚类树状图