中国人民大学统计咨询研究中心
中国人民大学概率论与数理统计研究所　联合推出
教育部重点科研基地应用统计科学研究中心

数据分析系列教材

多元统计分析
——原理与基于SPSS的应用
（第二版）

李静萍　编著

中国人民大学出版社

中国人民大学统计咨询研究中
中国人民大学概率论与数理统计研究
教育部重点科研基地应用统计科学研究中心

数据分析系列教材

多元统计分析
——原理与基于 SPSS 的应用

（第二版）

李静萍　编著

中国人民大学出版社
· 北京 ·

图书在版编目（CIP）数据

多元统计分析：原理与基于 SPSS 的应用/李静萍编著．—2 版．—北京：中国人民大学出版社，2015.4
数据分析系列教材
ISBN 978-7-300-20849-7

Ⅰ.①多… Ⅱ.①李… Ⅲ.①多元分析-统计分析-高等学校-教材 Ⅳ.①O212.4

中国版本图书馆 CIP 数据核字（2015）第 034063 号

数据分析系列教材
多元统计分析——原理与基于 SPSS 的应用（第二版）
李静萍　编著
Duoyuan Tongji Fenxi：Yuanli yu Jiyu SPSS de Yingyong

出版发行	中国人民大学出版社		
社　　址	北京中关村大街 31 号	**邮政编码**	100080
电　　话	010－62511242（总编室）		010－62511770（质管部）
	010－82501766（邮购部）		010－62514148（门市部）
	010－62515195（发行公司）		010－62515275（盗版举报）
网　　址	http://www.crup.com.cn		
	http://www.ttrnet.com（人大教研网）		
经　　销	新华书店		
印　　刷	北京东方圣雅印刷有限公司	**版　　次**	2008 年 6 月第 1 版
规　　格	185 mm×260 mm　16 开本		2015 年 4 月第 2 版
印　　张	12.5 插页 1	**印　　次**	2015 年 4 月第 1 次印刷
字　　数	262 000	**定　　价**	29.00 元

版权所有　侵权必究　　印装差错　负责调换

数据分析系列教材编委会

编委会主任　易丹辉

编委会委员（按姓氏笔画排序）

王惠文　吴喜之　张　波

易丹辉　柯惠新　耿　直

黄登源　谢邦昌

总　序

随着社会经济的不断发展、科学技术的不断进步，统计方法越来越成为人们必不可少的工具和手段。在多年教学过程中，老师们也越来越感到运用统计方法解决实际问题的重要，不少人在探索如何运用统计软件介绍和学习统计方法，思考如何运用这些方法解决实际问题。

自2008年数据分析系列教材出版以来，得到了不少读者的关注，这次丛书在原来的基础上，根据几位老师教学和科研的实践经验，重新策划完成。丛书特别邀请北京航空航天大学王惠文教授加盟，担任《描述统计》一书的撰写。她多年运用描述统计方法解决实际问题，积累了丰富的经验，将为读者正确运用描述统计方法提供参考。由于几位年轻老师不辞辛苦，这套数据分析系列教材在原来的基础上作了调整，更适合读者的实际需要。

在我们看来，掌握统计方法不仅要理论上弄明白，更重要的在于能够正确有效地运用这些方法，分析说明实际问题。这套丛书正是试图利用实际数据，通过统计软件的实际操作，对所能够使用的统计方法加以说明，使读者不仅能够了解相应的统计方法，而且能够通过计算机操作学会运用这些方法处理分析实际数据。希望这套丛书的出版能够为读者提供这样学习的工具。

由于水平有限，难免有不足之处。恳请读者朋友们提出宝贵意见。我们也会循着这样的思路，在教学以及和读者的交流沟通中不断积累、不断提高、不断完善，奉献给读者更多、更好的成果。

感谢为这套丛书的编写付出汗水的研究生，感谢几位认真用心的年轻老师，感谢中国人民大学出版社的大力支持，特别是陈永凤编辑和王伟娟编辑，是她们的努力，让这套丛书在很短的时间里出版。为方便读者，书中的所有例题数据，都将放在中国人民大学出版社的网站上，供读者下载并练习。谢谢读者，希望在数据采集变得越来越容易、大数据时代到来的今天，能够加强沟通和联系，为提高统计方法实际运用的能力和水平共同努力。

易丹辉

试读结束：需要全本请在线购买： www.ertongbook.com

前　言

多元统计分析是近年来发展迅速的统计分析方法之一，广泛应用于自然科学和社会科学的各个学科，是各个领域的研究者和工作者探索多元世界的强有力工具。目前，市面上已经有不少关于多元统计分析方法的教材或专著，其中也不乏从国外引进或翻译过来的同类教材，它们对多元统计的方法原理有比较深入的介绍。但是，对于那些希望在实际研究或工作中应用多元统计方法的研究者而言，急需的是一本既能通俗地介绍多元统计方法原理，又能提供切实的操作指南的参考书。

本书根据读者的上述需求，在简要介绍多元统计方法原理的基础上，侧重于结合实例介绍多元统计方法的应用。在方法的具体实现上，书中采用国内广泛使用的统计软件SPSS，详细介绍多元统计方法在统计软件中的实现以及计算机输出结果的解读。

在内容编排上，本书基本涵盖了常用的多元统计方法，尤其是第十二章所介绍的结构方程模型，是市面上多元统计教材或专著较少涉及的主题；在写作风格上，本书用尽可能通俗的语言阐明各种多元统计方法功能和原理；在案例应用上，本书全部采用现实中的真实数据，尽可能详尽地介绍统计软件的各种操作选项和输出结果。此外，本书在各章的最后一节给出使用各种多元统计方法注意的事项，力求帮助读者正确应用多元统计分析方法。

本书可供经济学、管理学、心理学、生物医学等各个领域的实际工作者参考，同时也可以作为上述有关专业以及统计学专业的高年级本科生或研究生的教材或参考书。

在本书写作过程中，得到了易丹辉教授的悉心指导和大力支持，她对青年教师的关心和支持是本书得以顺利完成的坚强后盾，在此致以衷心的感谢。此外，研究生张剑为本书的案例分析提供了基础素材，付出了大量的劳动，在此也向他表示深深的感谢。当然，文责自负，书中难免有疏漏和错误，全部由编著者承担。

李静萍

目　录

第一章　回归分析 …… 1

第一节　一元回归分析 …… 1

第二节　多元回归分析 …… 10

第三节　回归分析的注意事项 …… 16

第二章　主成分分析 …… 18

第一节　主成分分析的基本模型 …… 18

第二节　主成分求解及其性质 …… 20

第三节　主成分分析的实例分析 …… 23

第四节　主成分分析的注意事项 …… 29

第三章　因子分析 …… 31

第一节　因子分析概述 …… 31

第二节　因子分析的步骤 …… 34

第三节　因子分析的实例分析 …… 38

第四节　因子分析的注意事项 …… 45

第四章　聚类分析 …… 46

第一节　聚类分析方法概述 …… 46

第二节　聚类方法 …… 49

第三节　聚类分析的实例分析 …… 52

第四节　聚类分析的注意事项 …… 64

第五章　判别分析 …… 66
第一节　距离判别法 …… 66
第二节　Fisher 判别法 …… 68
第三节　Bayes 判别法 …… 71
第四节　逐步判别法 …… 72
第五节　判别分析的实例分析 …… 73
第六节　判别分析的注意事项 …… 80

第六章　典型相关分析 …… 82
第一节　典型相关分析概述 …… 82
第二节　典型相关分析的实例分析 …… 85
第三节　典型相关分析的注意事项 …… 90

第七章　对应分析 …… 91
第一节　对应分析概述 …… 91
第二节　对应分析的实例分析 …… 94
第三节　对应分析的注意事项 …… 101

第八章　多维标度分析 …… 102
第一节　多维标度分析概述 …… 102
第二节　多维标度分析的实例分析 …… 105
第三节　多维标度分析的注意事项 …… 110

第九章　广义线性模型 …… 112
第一节　广义线性模型概述 …… 112
第二节　广义线性模型的实例分析 …… 119
第三节　广义线性模型的注意事项 …… 126

第十章　对数线性模型 …… 127
第一节　对数线性模型概述 …… 127
第二节　对数线性模型的实例分析 …… 129
第三节　对数线性模型的注意事项 …… 139

第十一章　生存分析 …… 140
第一节　生存分析概述 …… 140
第二节　非参数方法 …… 142
第三节　参数方法 …… 154

第四节 半参数方法 …… 158
第五节 生存分析的注意事项 …… 168

第十二章 结构方程模型 …… 170
第一节 结构方程模型概述 …… 170
第二节 结构方程模型的实例分析 …… 177
第三节 结构方程模型的注意事项 …… 189

参考文献 …… 190

第一章 Chapter 1 回归分析

回归分析是研究一个因变量与一个或多个自变量是否有关以及关系强度的统计方法。在我们找到变量间的回归关系之后，就可以利用这些关系来描述变量之间的关系。例如，如果确定了居民收入与消费总额的回归方程，就可以了解这两个变量之间的关系，确定居民边际消费倾向的大小。其次，可以利用回归关系对目标变量进行控制。例如，如果找到了商品价格与需求量之间的回归关系，那么通过控制价格，就可以在一定程度上控制需求量。此外，还可以利用回归关系对目标变量进行预测。例如，如果找到了居民收入与消费总额的回归关系，就可以根据居民收入来估计当年的消费总额。

只有一个自变量的回归分析称为一元回归分析，有多个自变量的回归分析称为多元回归分析。本章分别介绍这两种回归分析的基本原理及其应用。

第一节 一元回归分析

一、一元回归分析概述

最简单的回归分析是一元线性回归，即只包括一个因变量 Y 和一个自变量 X：

$$Y_i=\beta_0+\beta_1 X_i+\varepsilon_i,\ i=1,2,\cdots,n \tag{1.1}$$

式中，Y 表示因变量或响应变量（dependent variable；response variable）；X 表示自变量或解释变量（independent variable；explanatory variable）；ε 表示误差项；i 表示第 i 个样本单位，（x_i，y_i）表示（X，Y）的第 i 个观测值。

式（1.1）表现的关系称为线性回归模型（linear regression model）。其中，β_0 和

β_1是模型中的参数（regression parameters），也称为回归系数（regression coefficient）。$\beta_0+\beta_1X_i$反映变量间的线性统计关系，是给定 X_i的条件下，Y 的平均值。将 $Y_i=\beta_0+\beta_1X_i$称为回归函数，回归函数反映了因变量与自变量之间确定性关系的部分。ε_i表示由 X 以外的其他一切因素所引起的变动，是因变量与自变量之间随机性关系的部分。

经典的回归分析有如下假定：

（1）$E(\varepsilon_i|X_i)=0$，$E(Y_i|X_i)=\beta_0+\beta_1X_i$；

（2）$Var(\varepsilon_i|X_i)=\sigma^2$，$Var(Y_i|X_i)=\sigma^2$；

（3）$Cov(\varepsilon_i, \varepsilon_j)=0$，$Cov(Y_i, Y_j)=0$，当 $i\neq j$ 时；

（4）$\varepsilon_i\sim N(0, \sigma^2)$，$Y_i\sim N(\beta_0+\beta_1X_i, \sigma^2)$。

其中，第 1 个假定称为线性回归假定，第 2 个假定称为同方差假定，第 3 个假定称为序列无关假定，第 4 个假定称为正态性假定。

回归分析的基本任务包括根据样本数据对回归参数进行估计和假设检验。

二、回归参数的估计

常用的估计回归参数的方法有两种，即普通最小二乘估计和极大似然估计。

（一）普通最小二乘估计（ordinary least square estimation，OLSE）

对于每一个样本单位，考虑因变量的观测值 y_i与其平均值$\beta_0+\beta_1x_i$的离差，该离差当然越小越好。综合考虑 n 个离差值，定义离差平方和为：

$$Q(\beta_0,\beta_1)=\sum_{i=1}^{n}[y_i-E(Y_i|x_i)]^2=\sum_{i=1}^{n}(y_i-\beta_0+\beta_1x_i)^2 \tag{1.2}$$

所谓最小二乘法，就是寻找 β_0 和 β_1 的估计值 $\hat{\beta}_0$ 和 $\hat{\beta}_1$，使式（1.2）达到最小。

利用微积分知识可以证明，回归参数的最小二乘估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 为：

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{1.3}$$

$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x} \tag{1.4}$$

式中，$\bar{x}$ 为自变量的平均值；$\bar{y}$ 为因变量的平均值。

将回归参数的最小二乘估计代入回归函数，可得到回归方程，即 $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$。$\hat{y}_i$ 是给定 x_i的条件下，因变量平均值的估计值，也称为因变量的拟合值。因变量观察值 y_i与其拟合值$\hat{y}_i$ 之间的离差 $y_i-\hat{y}_i$ 称为残差（residual）。

在回归分析中，需要对残差进行分析，从而诊断经典回归分析的基本假定是否

试读结束：需要全本请在线购买： www.ertongbook.com

成立。

(二)极大似然估计(maximum likelihood estimation,MLE)

极大似然估计是利用总体的分布密度或概率分布的表达式及样本提供的信息建立似然函数,从而求解未知参数估计量的一种方法。

对于连续型随机变量,似然函数就是样本的联合分布密度函数;对于离散型随机变量,似然函数就是样本的联合概率函数。

当总体 X 为连续型分布时,设其分布密度族为 $\{f(x;\theta),\theta\in\Theta\}$。假设总体 X 的一个独立同分布的样本为 x_1,x_2,…,x_n,则其似然函数为:

$$L(\theta;x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i;\theta) \tag{1.5}$$

极大似然估计是在 θ 的所有可能取值中选取使随机样本(X_1,X_2,…,X_n)落在(x_1,x_2,…,x_n)附近的概率最大的 $\hat{\theta}$ 为未知参数 θ 真值的估计值,即极大似然估计量 $\hat{\theta}$ 满足:

$$L(\hat{\theta};x_1,x_2,\cdots,x_n)=\max_{\theta}L(\theta;x_1,x_2,\cdots,x_n)$$

似然函数的概念不仅仅局限于独立同分布的样本,只要样本的联合密度形式已知,就可以应用极大似然估计。

利用经典回归分析的几个基本假定,即可写出回归分析的似然函数,从而求得回归参数的极大似然估计。

三、回归分析的假设检验与拟合优度

当我们得到回归系数的估计后,不能直接利用它进行分析,还需要用统计方法对回归系数进行显著性检验,并对回归方程的拟合效果进行评估。需要注意的是,在进行假设检验时,通常需要作正态性假设,即假定 ε_i 服从正态分布。对系数的检验方法有 t 检验和 F 检验,对方程拟合优度的评价可采用样本决定系数。

(一)t 检验

t 检验是统计推断中一种常用的检验方法,在回归分析中,t 检验用于检验回归系数的显著性,即检验自变量 X 对因变量 Y 的影响程度是否显著。

t 检验的原假设是 H_0:$\beta_1=0$,备择假设是 H_1:$\beta_1\neq0$。

若原假设 H_0 成立,则因变量 Y 与自变量 X 之间并没有真正的线性关系,即自变量 X 对因变量 Y 没有显著影响。

t 检验使用的检验统计量为 t 统计量。给定显著性水平 α,双侧检验的临界值为 $t_{\alpha/2}$。当 $|t|\geqslant t_{\alpha/2}$ 时,拒绝原假设,认为 β_1 显著不为 0,一元线性回归成立;当 $|t|<$

$t_{\alpha/2}$时，不能拒绝原假设，认为 β_1 与 0 没有显著差异，一元线性回归不成立。

（二）*F* 检验

对回归系数显著性的另一种检验方法是 F 检验；F 检验也可对回归方程的显著性进行检验。F 检验根据平方和分解，直接从回归方程的拟合效果检验回归方程的显著性。

总平方和：$SST=\sum_{i=1}^{n}(y_i-\bar{y})^2$

回归平方和：$SSR=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$

残差平方和：$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$

容易得到平方和的分解式为：

$$SST=SSR+SSE \tag{1.6}$$

总平方和 SST 反映因变量的总的波动程度；回归平方和 SSR 是由回归方程确定的，也就是自变量 X 的波动引起的因变量的波动程度；误差平方和 SSE 则是不能用自变量解释的波动，是由 X 之外的不能控制的因素引起的。显然，回归平方和 SSR 越大，回归的效果越好。

F 检验的假设与 t 检验相同，其检验统计量如下：

$$F=\frac{SSR/1}{SSE/(n-2)} \tag{1.7}$$

在正态假设下，当原假设成立时，F 服从自由度为（1，$n-2$）的 F 分布。给定显著性水平 α，F 检验的临界值为 $F_\alpha(1,n-2)$。当 $F\geqslant F_\alpha(1,n-2)$ 时，拒绝原假设，说明回归方程显著，X 与 Y 有显著的线性关系。

（三）样本决定系数

由前述回归平方和与残差平方和的含义可知，回归平方和在总平方和中所占的比重越大，线性回归效果越好，表明回归直线对样本观测值的拟合优度越好。最理想的情况是所有观察值均落在回归直线上，此时 $SSE=0$。如果残差平方和所占的比重大，则说明回归直线对样本观测值的拟合效果不理想，最坏的情况是 $SSR=0$，此时估计回归方程式完全无法预测 y。

将回归平方和与总离差平方和之比定义为样本决定系数，记为 R^2，即

$$R^2=\frac{SSR}{SST}=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{1.8}$$

决定系数 R^2 是一个衡量回归直线对样本观测值拟合优度的相对指标，反映了因变量的波动中能用自变量解释的比例。R^2 的值总是在 0～1 之间，R^2 越接近于 1，拟合优度就越好；反之，说明模型中所给出的 X 对 Y 的信息还不充分，回归方程的效果不好，应进行修改，使 X 与 Y 的信息得到充分利用。

四、一元回归应用实例

我们以对中国 2009 年各地区有效发明专利数量的分析为例，说明一元回归分析的应用。数据为国家统计局网站发布的第二次 R&D 资源清查数据。

（一）变量选择

回归分析的第一步是确定模型中的因变量和自变量。本例使用的数据如图 1—1 所示。其中包括反映 R&D 活动产出的变量——有效发明专利数，以及反映各地区对 R&D 活动所投入的资源的变量，包括项目数、人员投入以及经费投入等。

V1	有效发明专利数	RD项目数	RD人员全时当量	人员全时当量中基础研究人员比重	RD人员	RD人员中博硕士比重	RD经费支出	RD经费支出中人员劳务费比重	RD经费支出中仪器和设备费比重	RD经费支出中企业资金比重
北京	32212	83911	191779	14.18	252676	36.93	6686351	25.04	14.57	36.07
天津	6987	22472	52039	8.34	72599	20.28	1784661	17.98	15.46	76.32
河北	3666	18135	56509	5.65	84601	16.06	1348446	20.55	16.23	75.13
山西	2299	9944	47772	6.34	65147	15.05	808563	19.70	17.55	80.62
内蒙古	736	6264	21676	8.06	31381	16.39	520726	20.17	15.60	79.85
辽宁	11042	29032	80925	8.70	119440	21.56	2323687	17.03	11.29	77.01
吉林	3421	15728	39393	16.12	56428	35.17	813602	18.80	11.45	65.07
黑龙江	5559	19148	54159	12.86	72587	21.00	1091704	20.29	11.52	59.91
上海	24587	50567	132859	10.24	170512	27.50	4233774	29.79	13.17	66.84
江苏	25983	65224	273273	3.23	369403	12.67	7019529	24.73	9.19	81.89
浙江	22664	62145	185069	2.96	239058	12.54	3988367	28.93	13.26	88.68
安徽	5200	23029	59697	9.05	87664	16.61	1359535	18.47	15.08	69.32
福建	3933	21771	63269	5.27	85745	13.41	1353819	26.94	18.22	86.12
江西	1024	15797	33055	6.99	51894	14.23	758936	18.04	13.01	73.88
山东	11330	44251	164620	5.17	233137	14.12	5195920	19.41	14.87	89.07
河南	6187	22347	92571	2.84	132062	11.99	1747599	20.89	16.55	81.36
湖北	9207	36665	91161	7.05	131680	19.51	2134490	21.04	12.79	70.35
湖南	10602	28335	63843	8.09	93806	19.60	1534995	23.16	11.51	79.21
广东	33951	64654	283650	3.06	383524	17.93	6529820	37.36	9.76	88.49
广西	1403	16533	29856	16.05	45049	25.19	472028	24.30	13.96	71.83
海南	194	2534	4210	10.74	6487	22.07	57806	31.67	17.81	38.18
重庆	3573	16140	35005	9.58	53359	21.63	794599	21.82	17.73	77.19

图 1—1 发明专利与 R&D 投入数据

本例中，有效发明专利数为因变量。一个可能的假设是 R&D 的产出与 R&D 的投入有关，投入越多，则产出越多。

在本例中，反映 R&D 活动投入的变量有多个，如果建立一元回归模型，需要从多个变量中选择一个。如果不能通过数据的意义来确定最合适的自变量，一般先计算各个变量与因变量的相关系数，然后根据各变量与因变量的相关性大小来选择自变量。

按照“分析”（Analyze）→“相关”（Correlate）→“双变量”（Bivariate）的路径进入相关分析对话框，然后将所有变量选进“变量”（Variable），点击“确定”（OK），可以得到因变量与其他所有变量的相关系数。本例中，有效发明专利数与各变量的相关系数如表 1—1 所示。由表 1—1 可以看出，除了 R&D 人员中博硕士比重、R&D 经费支出中人员劳务费比重和 R&D 经费支出中企业资金比重以外，其他变量与因变量的相关系数都在 0.05 的显著性水平下具有显著性。由于 R&D 项目数与因变量的相关关系很强，因此下面选择 R&D 项目数为自变量，建立一元回归模型。

在建模之前可以绘制因变量与自变量的散点图，观察变量之间的统计关系的形式。

表 1—1 **有效发明专利数与其他变量的 Pearson 相关系数**

	RD 项目数	RD 人员全时当量	人员全时当量中基础研究人员比重	RD 人员	RD 人员中博硕士比重	RD 经费支出	RD 经费支出中人员劳务费比重	RD 经费支出中仪器和设备费比重	RD 经费支出中企业资金比重
Pearson 相关系数	0.954**	0.930**	−0.383*	0.919**	0.168	0.944**	0.322	−0.480**	0.139
显著性（双侧）	0.000	0.000	0.034	0.000	0.366	0.000	0.077	0.006	0.457

** 在 0.01 水平（双侧）上显著相关；
* 在 0.05 水平（双侧）上显著相关。

按照“图形”（Graphs）→“散点/点状”（Scatter/Dot）→“简单分布”（Simple Scatter）的路径进入散点图窗口，在窗口的“Y 轴”（Y Axis）和“X 轴”（X Axis）中选择因变量与自变量，点击“确定”（OK）后即可生成两者之间的散点图。本例的散点图如图 1—2 所示。由图 1—2 可以看到，有效发明专利数与项目数大体上呈现一种正向的线性关系，可以考虑建立一元线性回归模型。

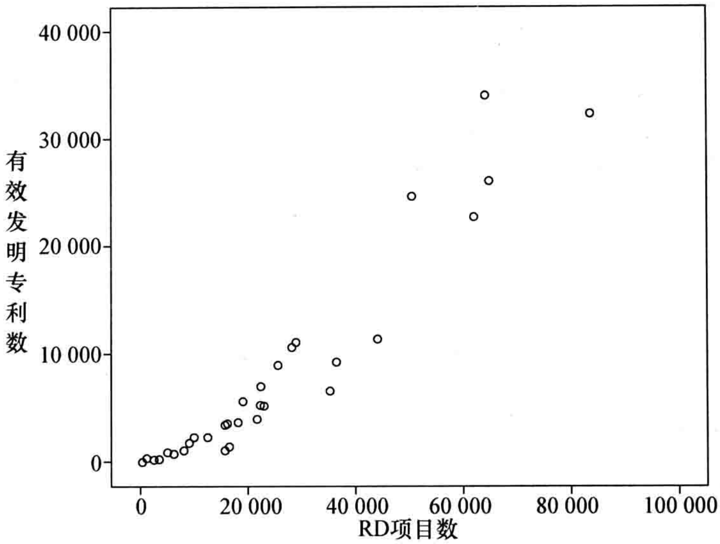

图 1—2 因变量与自变量的散点图

（二）建立一元线性回归模型

在 SPSS 中进行回归模型估计的具体操作步骤为：点选菜单“分析”（Analyze）→“回归”（Regression）→“线性”（Linear regression），出现回归分析对话框，如图 1—3 所示。在变量选择对话框，选择“有效发明专利数”进入“因变量”（Dependent）框，选择“RD 项目数”进入“自变量”（Independent）框，在“方法”（Method）选项中选择“进入”（Enter），表示让选择的自变量进入回归模型，点击“确定”（OK）完成分析。

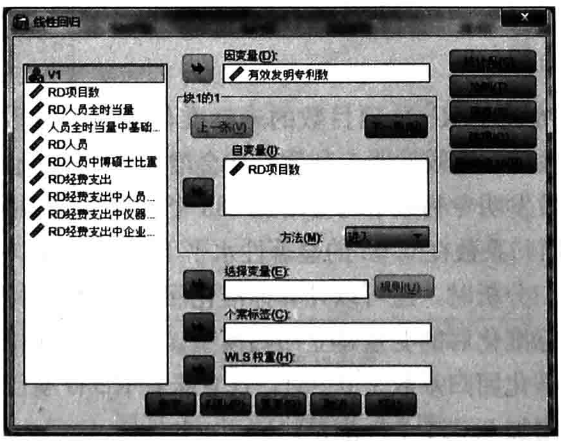

图 1—3 一元回归分析对话框

（三）结果分析

在输出（Output）中可以看到回归模型的一般性统计量表（见表 1—2）。

表 1—2　　回归模型的汇总统计量表

模型	R	R 方	调整 R 方	标准估计的误差
1	0.954[a]	0.911	0.908	2 931.960

a. 预测变量：(常量)，RD 项目数。

由表 1—2 可以看出，回归方程的样本决定系数为 0.911，回归方程的效果很好。

表 1—3 是反映平方和分解的方差分析表。由表 1—3 可以看出，F 统计量为 296.01，伴随的 Sig. 值接近 0，拒绝原假设，可以认为自变量与因变量之间有显著的线性关系。

表 1—3　　方差分析表[a]

模型		平方和	df	均方	F	Sig.
1	回归	2 544 620 119.043	1	2 544 620 119.043	296.010	0.000[b]
	残差	249 295 242.699	29	8 596 387.679		
	总计	2 793 915 361.742	30			

a. 因变量：有效发明专利数。
b. 预测变量：(常量)，RD 项目数。

下面进一步对回归系数进行检验和解释。表 1—4 给出了回归系数的估计值及其显著性检验结果。

表 1—4　　一元回归系数结果[a]

模型		非标准化系数		标准系数	t	Sig.
		B	标准误差	试用版		
1	(常量)	−2 866.931	820.710		−3.493	0.002
	RD 项目数	0.433	0.025	0.954	17.205	0.000

a. 因变量：有效发明专利数。

有效发明专利数关于 R&D 项目数的未标准化的回归系数为 0.433。这意味着，每增加 1 个 R&D 项目，有效发明专利数平均会增加 0.433 个。换言之，每增加 1 000 个 R&D 项目，有效发明专利数平均会增加 433 个。这个未标准化回归系数的 Sig. 值接近 0，所以这个回归系数在 0.05 的显著性水平下是显著的。

另外，进行回归分析时，还可以计算标准化回归系数。对因变量和自变量分别进行标准化，然后对标准化后的变量建立回归模型，所得到的回归系数即为标准化回归系数。本例中，标准化回归系数是 0.954，这意味着 R&D 项目数每增加 1 个标准差，有效发明专利数将增加 0.954 个标准差。

最后，可以根据残差图对回归模型的效果进行评价。在回归分析菜单的“绘制”

试读结束：需要全本请在线购买： www.ertongbook.com