

» 数据分析与模拟丛书

陈彦光 编著

Geographical Data Analysis with Matlab

基于 Matlab 的 地理数据分析

 高等教育出版社
HIGHER EDUCATION PRESS

» 数据分析与模拟丛书

陈彦光 编著

Geographical Data Analysis with Matlab

基于 Matlab 的 地理数据分析

J I Y U M A T L A B D E D I L I S H U J U F E N X I

 高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容简介

本书面向地理问题, 基于 Matlab 软件, 讲述了大量数学方法的应用思路和过程。内容涉及回归分析、主成分分析、因子分析、聚类分析、判别分析、时(空)间序列分析、Markov 链、R/S 分析、线性规划、层次分析法以及人工神经网络建模等方法。通过模仿本书讲授的计算过程, 读者可以加深对有关数学方法的认识和理解, 并且掌握很多 Matlab 的应用技巧。本书最初以北京大学本科生计量地理学的辅助教材形式出现, 但实际上是作者对 Matlab 计算功能深入应用的经验总结。本书中的讲授体例与一般 Matlab 的教科书不同, 计算过程设计为笔者独创, 在国内外其他教科书中未曾见到。

本书虽然是以地理数据为分析对象展开论述的, 但所涉及的内容绝大多数为通用方法。只要改变数据的来源, 书中给出的计算过程完全可以应用到其他领域。本书可供地理学、生态学、环境科学、地质学、经济学、城市规划以及医学、生物学等诸多领域的学生、研究人员和工程技术人员学习或参考。

图书在版编目(CIP)数据

基于 Matlab 的地理数据分析 / 陈彦光编著. —北京: 高等教育出版社, 2012. 7

ISBN 978 - 7 - 04 - 034172 - 0

I. ①基… II. ①陈… III. ①Matlab 软件 - 应用 - 地理信息系统 - 数据 - 分析 IV. ①P208 - 39

中国版本图书馆 CIP 数据核字(2012)第 116756 号

策划编辑 陈正雄 责任编辑 关 焱 封面设计 张 楠 版式设计 马敬茹
插图绘制 黄建英 责任校对 金 辉 责任印制 毛斯璐

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 北京中科印刷有限公司
开 本 787 mm × 1092 mm 1/16
印 张 26.25
字 数 590 千字
购书热线 010 - 58581118

咨询电话 400 - 810 - 0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landrao.com>
<http://www.landrao.com.cn>
版 次 2012 年 7 月第 1 版
印 次 2012 年 7 月第 1 次印刷
定 价 55.00 元 (含光盘)

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 34172 - 00

前 言

数学在任何领域的应用都主要发挥两种功能：一是作为实验或者观察数据的整理手段；二是构造假设、建立模型和发展理论。无论哪种功能的发挥，都会涉及数据的处理和分析。第一种功能自不待言。关于第二种功能，假设与建模最终离不开实证与检验。有人曾将科学研究划分为三个世界：一是现实世界；二是数学世界；三是计算世界。现实世界和数学世界都是非常客观的世界，现实的存在和数学的推理都没有任何随意性可言。可是，计算世界就不一样了，计算世界不可避免人的主观认识。样本的提取、算法的选择和模型的建设等，都要根据研究人员的认识而定。数学方法的第二种功能（假设与建模）可以在数学世界完成，但第一种功能（数据整理）则属于计算世界。理论模型的实证与检测也属于计算世界的问题。要想有效地运用一种数学方法将上述三个世界成功地沟通起来，不仅需要数学原理的深入认识，而且需要对计算过程的熟练掌握。

众所周知，当今科学研究需要三套语言体系：一是交流语言（汉语和英语等）；二是自然语言（数学）；三是计算机语言（算法）。第二套语言体系与第三套语言体系存在一定的内在联系。数学工具是科学人员的“显微镜”和“望远镜”，它可以帮助研究者看得更细、望得更远。要想成功地掌握一门数学方法，至少要熟悉如下五个环节：一是基本原理，也就是一种方法的理论基础和逻辑过程；二是应用范围，即一种方法自身的特长及其功能局限——了解其优势和不足，才能真正有效地运用；三是算法或者运算规则系统，就是一种为在有限步骤内解决数学问题而建立的可重复应用的计算流程体系；四是计算过程，即在一种方法的有效适用范围内，给定一组观测数据，借助一定的算法获取所要求的计算结果；五是典型实例，也就是一种数学方法应用于现实问题的具体案例。假如想要进一步加深对某种数学方法的了解，则还有第六个环节，那就是不同方法的融会贯通。

科学研究中应用较多的是高等数学知识，掌握必要的高等数学原理可以使得研究工作事半功倍。高等数学有所谓“老三高（数学分析、线性代数和概率论与统计学）”和“新三高（拓扑学、泛函分析和抽象代数）”之分，其中以“老三高”的应用最为广泛。以地理学中常用的数学方法——回归分析为例，学习该方法涉及如下过程：在基本思想方面，回归建模就是用数学语言刻画一组变量与某个变量之间的相关关系或者因果关系。关系的强弱通过回归系数表现出来，回归分析的核心问题就是模型参数值的估计。为此，需要一种有效的算法。目前的回归分析算法主要采用误差平方和最小的方法，即所谓最小二乘法。在这个过程中，首先，要采用线性方程组进行描述，理论上用到线性代数的知识；其次，寻求误差平方和最小时的参数估计结果，理论上用到微积分的条件极值方法；在回归结果检验过程中，涉及误差的正态分布

II 前言

思想，这在理论上又用到大量的概率论与统计学原理。可是，虽然很多读者明白上述道理，但在具体应用过程中依然觉得似是而非。究其原因，主要在于不了解计算过程，没有掌握简明易懂的计算范例。

在多年的地理数学方法教学实践中，笔者逐渐总结了一些引导学生学习数学方法的技巧。借助电子表格 Excel、数学软件 Mathcad 或者 Matlab 和统计软件 SPSS 等，可以循序渐进地掌握一些数学方法的计算过程和简明范例。通过这个过程进一步加深对有关数学原理和方法的理解以及对应用领域的认识，进而将不同的方法有机地联系起来。撰写本书的主要目的，就是借助简明的案例帮助读者应用 Matlab 软件进行地理数据分析。通过地理实例学习 Matlab 的使用方法和技巧，通过 Matlab 的运用解决地理学领域的科学问题。Matlab 可以成为学习数学方法的辅助工具，数学知识的拓展又会使得读者加深对 Matlab 的理解。全书的内容分为四大部分：一是相关分析和回归分析，主要讲述线性回归和逐步回归的计算过程以及简单非线性模型的参数估计方法；二是多元统计分析，主要介绍主成分分析、因子分析、聚类分析和判别分析的计算过程；三是时空过程分析，包括时（空）间序列分析和随机过程分析，主要讲述自相关分析、自回归分析、谱分析、小波分析、Markov 链分析和 R/S 分析；四是系统分析，主要讲述层次分析法（AHP）、线性规划求解以及人工神经网络的建模与预测分析方法。

虽然书中讲到许多 Matlab 函数的应用方法，但这不是一本关于 Matlab 的教科书，而是应用 Matlab 软件的数据处理和数学方法运用的教科书。国内讲述 Matlab 的图书汗牛充栋。但是，考察这些图书可发现如下问题：第一，绝大多数图书是似懂非懂地翻译 Matlab 的帮助文件的结果，不能有效帮助读者掌握 Matlab 的使用技巧；第二，基于 Matlab 讲述完整分析案例的图书非常罕见；第三，基于 Matlab 全面、系统地分析地理数据的图书更是鲜见。因此，笔者决定仿照自己的《基于 Excel 的地理数据分析》和《基于 Mathcad 的地理数据分析》书中的教学体例，撰写本书。相对于前两本书，这本书更难撰写。关于 Excel 等，一般不涉及关于函数应用方法的过多讲述，简单说明如何调用有关函数即可。可是，Matlab 的函数带有太多的参数，如果不详细讲解，则应用方法说不清楚；如果讲解过细，则这本书将会非常冗长。最后，笔者还是追求结构的完整，但不求将函数讲述得过度精细。希望读者在应用过程中“学而时习之”，逐步领会 Matlab 函数的调用方法。较之于前两本书，本书有一个明显的特点，那就是弱化了纯粹技术的细节，更新了一些内容，并且加强了对地理数据分析和模型选择的解释。

不同的软件有不同的优势，也有各自的不足。关键在于扬长避短，懂得针对具体问题选择相应的软件。以主成分分析方法为例，采用大型统计分析软件 SPSS 可以很方便地获得全面的计算结果，但是 SPSS 是一个“傻瓜”型软件，其计算过程对读者而言完全是一个“黑箱”。按照固定程序操作该软件，不需要多少数学知识就可以完成有关的统计运算。但是，如果不了解一种方法的计算过程，不知道这些方法的基本原理，即便 SPSS 输出了结果，读者也没有办法给出准确的结果解读。如果读者首先在 Excel 或者 Mathcad 里完成一个简明例子的计算，通过这个过程熟悉主成分分析的数学运算过程，然后再利用 SPSS 开展有关的数据整理和分析，就会主动和透明多了。至于 Matlab，既可以发挥 Excel 或者 Mathcad 的逐步计算功能，也可以

发挥 SPSS 的全程计算功能，非常方便。不过，Matlab 也有一些问题，例如，相关系数平方的校正公式与其他软件采用的公式不同（笔者怀疑是 Matlab 软件制作人员失误）；偏自相关系数的计算也是基于最小二乘法的近似估计，而不是基于 Yule-Walker 方程的递推运算。尽管如此，Matlab 强大的计算功能是目前其他数学计算软件难以替代的，其优越的性能引起了越来越多地理工作者的兴趣。

需要特别强调的是线性回归分析方法。这种方法非常简单，而且是基础的，以致很多读者不重视对该方法的深入学习和潜心练习。实际上，越是简单和基本的数学方法被使用的频率越高，应用范围也越广泛。一些复杂的数学方法，如主成分分析、判别分析、自回归分析、功率谱分析、小波分析、神经网络分析、灰色系统建模和预测分析等，都可以借助线性回归分析来快速入门或者掌握相关的运算技巧。本书讲述了基于回归分析的判别分析建模、自回归建模、R/S 分析建模、神经网络建模和预测等。这样，通过一种简明易懂的数学方法将多种数学方法贯通起来，读者可以通过回归分析了解不同数学方法的理论建设要点。

本书的写作特点是借助简单的例子，从头到尾完整地演示各种数学方法的计算过程和分析思路。读者应从头到尾重复一下书中的计算过程，然后寻找一个类似的例子，自己模仿一遍。在模仿中学习，在思考中消化。通过阅读和操作，可以打开一些数学方法的“黑箱”，了解其内部结构，从而更好地解译运算结果。大致了解之后，就可以借助 Matlab 或者有关统计/数学软件处理自己研究的现实问题了。原则上，本书的每一章都相对独立，如果读者对 Matlab 的基本功能比较熟悉，则从任何一个部分都可以开始学习。

最后，对本书的一些数据处理和模型表现形式给出几点说明。

第一，问题的求解过程是一个连续的计算过程。在一个案例分析过程中，如果下一步用到上一步的计算结果，以函数的形式调用上一步的有关表达式，而不是采用上一步的近似值。Matlab 默认的近似值是小数点后 14 位，因此全书的计算过程是根据 14 位小数完成的。但是，在行文中给出的结果都是根据具体情况截取的近似值。为了与软件中显示的结果相对应，一般保留 4 位小数，而且一概采用等号而非约等号。如果读者利用书中显示的近似值进行验算，前后数值可能不尽一致。顺便说明，Matlab 给出的各种统计量与 Excel 给出的相应统计量数值大体一致，但小数点后 10 位之后可能出现误差。相对而言，Excel 的数据显示更为精确。

第二，在没有特殊说明的情况下，统计检验的显著性水平一律取 0.05，相应的置信度是 95%。采用这个显著性水平有一个好处，那就是在了解样本大小或者自由度的情况下，读者可以方便地估计有关结果的标准误差。

第三，本书涉及两类数学公式：一是说明性的数学公式，采用 Word 的公式编辑器书写；二是演示操作过程的表达式，直接从 Matlab 中复制过来。第一类公式一般分章节编号，形式上符合数学方程的表达规范，允许一个式子中出现多个等号；第二类公式没有编号，往往多个式子排成一行，但每个式子中只有一个等号，这类式子形式上有时不符合数学方程的表达规范（如正、斜体和上、下标的表示）。

本书采用的 Matlab 版本主要是 7.0 和 7.01（参阅本书后记的说明），这两个版本相差无

IV 前言

几。掌握这两个版本的应用方法之后，就可以对其他版本触类旁通。附赠光盘中带有各章使用的原始数据（以 Excel 格式给出）以及笔者编写的 Matlab 计算代码，读者可以调用这些程序，重复书中给出的各种计算过程。Matlab 程序或代码已经调试成功，稍作修改就可以应用于读者的具体问题。此外，光盘中还有相关系数检验、 F 检验、 t 检验、Durbin-Watson 检验、卡方检验以及调和分析的 Fisher 检验临界值表，供读者参考和使用。

作者

2011 年 9 月

国家自然科学基金资助项目（编号：41171129）

目 录

第 1 章 一元线性回归分析	1
1.1 线性回归模型的矩阵形式	1
1.1.1 回归模型的矩阵表示	1
1.1.2 主要统计量的矩阵表示	2
1.2 一元线性回归	4
1.2.1 数据的初步考察	4
1.2.2 第一种模型求解途径——矩阵运算	5
1.2.3 第二种模型求解途径——多项式拟合	7
1.2.4 第三种模型求解途径——调用回归分析程序包	9
1.3 统计检验	13
1.3.1 相关知识的说明	13
1.3.2 主要的统计检验	14
1.4 总体回归估计和预测分析	16
1.4.1 总体回归估计	16
1.4.2 解释和外推预测分析	17
1.5 小结	19
第 2 章 多元逐步回归分析	21
2.1 多元线性回归分析	21
2.1.1 第一种途径——利用矩阵运算	21
2.1.2 第二种途径——调用回归分析程序包	24
2.1.3 统计检验	26
2.2 多重共线性判断	29
2.2.1 VIF 值的第一种计算方法	29
2.2.2 VIF 值的第二种计算方法	32
2.2.3 多元回归分析的变量选择问题	32
2.3 逐步回归分析	34
2.3.1 Matlab 逐步回归功能说明	34
2.3.2 逐步回归的实现	36
2.3.3 回归结果的输出和解读	39
2.4 逐步拟合	42
2.4.1 快速拟合方法	42

II 目录

2.4.2	详细拟合方法	43
2.4.3	几点说明	45
2.5	小结	45
第3章	非线性模型参数估计	47
3.1	常见数学模型表达式	47
3.2	常见实例——一变量的情形	49
3.2.1	指数模型(I)	49
3.2.2	对数模型	56
3.2.3	幂指数模型	59
3.2.4	双曲线模型	63
3.2.5	Logistic 模型(二参数形式)	68
3.2.6	指数模型(II)	72
3.2.7	指数模型与 logistic 模型	75
3.3	常见实例——一变量化为多变量的情形	79
3.3.1	多项式模型	79
3.3.2	二次指数模型	83
3.3.3	三参数 logistic 模型	85
3.3.4	Gamma 模型	94
3.4	常见实例——多变量的情形	97
3.4.1	Cobb-Douglas 生产函数	97
3.4.2	带有交叉变量的回归模型	99
3.5	广义线性拟合	100
3.5.1	广义线性拟合函数	100
3.5.2	典型的例子	102
3.6	方法比较	107
3.7	小结	109
第4章	主成分分析	110
4.1	实例和数据	110
4.1.1	案例数据	110
4.1.2	数据的保存与调用	112
4.2	第一套计算方案	113
4.2.1	详细计算步骤	113
4.2.2	计算程序的整理和结果的输出	120
4.2.3	计算结果的整理	123
4.3	第二套计算方案	124
4.3.1	程序的修改	124
4.3.2	两套方案的比较	125

4.4	第三套计算方案	127
4.4.1	计算程序	127
4.4.2	T 统计量	130
4.5	配套函数的调用	131
4.5.1	从协方差矩阵出发	131
4.5.2	主成分的残差分析	133
4.5.3	Bartlett 检验	134
4.6	结果分析方法	135
4.6.1	结果分析	135
4.6.2	综合评价	136
4.7	小结	138
第 5 章	因子分析	140
5.1	因子分析程序和案例	140
5.1.1	因子分析子程序	140
5.1.2	因子旋转子程序	142
5.1.3	案例与数据	145
5.2	因子模型的主成分解	146
5.2.1	主因子解	146
5.2.2	主因子解的正交旋转	148
5.3	因子模型的最大似然解	149
5.3.1	从原始数据出发	149
5.3.2	从协方差矩阵出发	154
5.3.3	载荷得分双重图	156
5.4	小结	157
第 6 章	层次聚类分析	159
6.1	聚类实例的初步结果	159
6.1.1	实例和数据	159
6.1.2	初步的聚类结果	160
6.2	程序说明与结果解析	161
6.2.1	聚类程序说明	161
6.2.2	聚类结果的解读	168
6.3	效果检验和类别查找	169
6.3.1	聚类效果检测	169
6.3.2	聚类结果的查询	170
6.3.3	聚类结果的比较	171
6.4	距离的选择与处理	174
6.4.1	欧氏距离平方	174

IV 目录

6.4.2	精度加权距离	175
6.4.3	主成分得分与马氏距离	177
6.5	聚类分析结论	178
6.6	小结	179
第7章	判别分析	181
7.1	案例和判别函数	181
7.1.1	数据及其来源	181
7.1.2	判别函数的调用方法	183
7.2	直接判别	184
7.2.1	二分类判别分析	184
7.2.2	三分类判别分析	186
7.3	详细计算过程	188
7.3.1	构造判别函数	188
7.3.2	数值的规范化处理	192
7.3.3	判别函数检验	194
7.3.4	待判样品归类	197
7.4	借助回归分析建立判别函数	198
7.5	聚类-判别联合分析	200
7.6	小结	201
第8章	自相关分析	202
8.1	数据来源和计算公式	202
8.1.1	案例数据来源	202
8.1.2	计算公式	203
8.2	自相关函数(ACF)	204
8.2.1	ACF及语法	204
8.2.2	ACF计算方法	205
8.2.3	ACF检验	207
8.3	偏自相关函数(PACF)	209
8.3.1	PACF函数和语法	209
8.3.2	PACF计算方法1——OLS法	209
8.3.3	PACF计算方法2——蛮力计算法	211
8.3.4	PACF计算方法3——程序计算法	212
8.3.5	结果汇总与PACF检验	214
8.4	自相关分析	215
8.4.1	自相关函数的分析判据	215
8.4.2	ACF和PACF分析	218
8.5	小结	220

第 9 章 自回归分析	221
9.1 样本数据的初步分析	221
9.1.1 案例数据来源和保存	221
9.1.2 数据的初步分析	221
9.2 自回归模型的回归估计	228
9.2.1 一阶自回归模型 $AR(1)$	228
9.2.2 高阶自回归模型 $AR(p)$	231
9.2.3 自回归模型的基本检验	234
9.2.4 预测结果及其比较分析	238
9.3 数据的平稳化及其自回归模型	241
9.3.1 数据平稳化	241
9.3.2 差分自回归	244
9.3.3 检验与预测	246
9.4 小结	247
第 10 章 谱分析	249
10.1 功率谱分析	249
10.1.1 时间序列数据	249
10.1.2 快速 Fourier 变换和频谱分析	250
10.1.3 检验和分析	255
10.1.4 计算程序简化	256
10.2 波谱分析	258
10.2.1 空间序列数据	258
10.2.2 数据准备	259
10.2.3 快速 Fourier 变换和参数估计	260
10.2.4 波谱分析	263
10.3 小结	266
第 11 章 小波分析	268
11.1 数据集和小波工具箱	268
11.1.1 数据集及其预备处理	268
11.1.2 小波工具箱	271
11.1.3 小波分析的基本函数	275
11.2 一维连续小波分析	278
11.2.1 周期长度估计方法之一	278
11.2.2 周期长度估计方法之二	282
11.2.3 随机信号去噪	285
11.3 一维离散小波分析	288
11.3.1 时间序列的压缩与重构	288

VI 目录

11.3.2	离散小波变换	291
11.4	二维小波分析	292
11.5	小结	296
第 12 章	R/S 分析	298
12.1	R/S 分析方法	298
12.1.1	Hurst 指数的定义方式	298
12.1.2	Hurst 指数与自相关系数的关系	299
12.2	编程计算	300
12.2.1	计算 Hurst 指数	300
12.2.2	图像分析	303
12.3	自相关系数和 R/S 分析	305
12.3.1	序列变化的自相关分析	305
12.3.2	分维和功率谱指数的估计	306
12.4	小结	307
第 13 章	Markov 链分析	309
13.1	Markov 链的转移概率矩阵	309
13.1.1	一个简单的例子	309
13.1.2	Markov 链的数学表示	310
13.2	Markov 链分析方法	311
13.2.1	转移概率矩阵的计算	311
13.2.2	自动计算	313
13.2.3	历次转移后的稳定分布	315
13.3	固定向量的计算方法	316
13.3.1	基于特征值和特征向量计算	316
13.3.2	基于线性方程求解计算	317
13.4	小结	318
第 14 章	线性规划	319
14.1	线性规划程序	319
14.1.1	线性规划函数及其输入选项	319
14.1.2	线性规划函数的输出选项	321
14.2	普通规划求解实例	323
14.2.1	实例 1——工业问题	323
14.2.2	实例 2——农业问题	326
14.2.3	实例 3——建筑业问题	327
14.2.4	实例 4——运输业问题	329
14.3	整数规划问题实例	332
14.3.1	一般整数规划	332

14.3.2	0-1 规划	336
14.4	非线性规划及其对偶问题实例	341
14.4.1	非线性规划原模型	341
14.4.2	非线性规划对偶模型	343
14.5	小结	343
第 15 章	层次分析法	344
15.1	问题与模型	344
15.2	计算方法	345
15.2.1	计算目标-准则层单权重	345
15.2.2	计算准则-方案层单权重	348
15.2.3	计算组合权重	351
15.2.4	判断矩阵的调试程序	352
15.3	其他计算途径	354
15.3.1	方根法	354
15.3.2	和积法	355
15.3.3	其他替代方法	356
15.4	结果解释	357
15.5	小结	358
第 16 章	人工神经网络	359
16.1	简单的线性网络	359
16.1.1	单输入-单输出: 对应于一元线性回归	359
16.1.2	多输入-单输出: 对应于多元线性回归	363
16.2	感知器和 M-P 模型	365
16.2.1	感知器的判别功能	365
16.2.2	感知器与自适应网络	369
16.2.3	基于聚类分析的感知器判别	372
16.2.4	基于主成分分析的感知器判别	375
16.2.5	三分类的感知器判别	378
16.3	学习向量量化(LVQ)神经网络	380
16.3.1	LVQ 神经网络的二分类判别	380
16.3.2	LVQ 神经网络的三分类判别	383
16.4	多层神经(BP)网络	384
16.4.1	BP 网络的离散选择	384
16.4.2	BP 网络判别	391
16.5	竞争型网络	394
16.5.1	有目标的先分类后判别	394
16.5.2	无目标的统一分类	396

Ⅷ 目录

16.6 小结	398
参考文献	399
后记	401

第 1 章

一元线性回归分析

线性回归是非常基本的但十分重要的数学建模方法。通过线性回归模型类比，我们可以更好地理解表面看来与回归分析无关的数学方法。利用 Matlab 进行线性回归分析相当方便，可以通过多种方式求解回归模型，包括矩阵运算和调用回归分析程序包。不过，要想得到较多的统计参数估计值，则需要对高等数学原理有比较透彻的了解。否则，Matlab 自动给出的统计量相对有限。同时，如果数学原理比较熟悉，可以熟练地编写 Matlab 文件，则 Excel 和 SPSS 可以给出的各种统计量都可以借助 Matlab 计算出来。下面借助一个非常简单的教学实例，分若干部分介绍 Matlab 的线性回归模型的建设过程。在此之前，首先给出一些基本的回归分析矩阵表达式。

1.1 线性回归模型的矩阵形式

1.1.1 回归模型的矩阵表示

Matlab 是矩阵实验室 (Matrix Laboratory) 的缩写，该软件最初是为基于矩阵的数学实验而设计的。因此，矩阵运算是利用 Matlab 开展回归分析的基本途径。为了使读者更好地理解后面的计算程序，不妨先给出有关回归分析的矩阵表达式。采用矩阵表示和求解，整个分析和计算过程十分简捷。假定有 m 个自变量， n 个样品，则多元线性回归可以表作如下矩阵方程

$$Y = XB + E \quad (1-1-1)$$

其中

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}, \quad B = \begin{bmatrix} a \\ b_1 \\ \vdots \\ b_m \end{bmatrix}, \quad E = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$