

数理统计与 *MATLAB* 数据处理

包研科 李 娜 编著



東北大學出版社
Northeastern University Press

0212/68

2008

数理统计与 MATLAB 数据处理

包研科 李 娜 编著

东北大学出版社

• 沈 阳 •

© 包研科 李 娜 2008

图书在版编目 (CIP) 数据

数理统计与 MATLAB 数据处理 / 包研科, 李娜编著. — 沈阳 : 东北大学出版社, 2008.3

ISBN 978-7-81102-519-4

I. 数… II. ①包… ②李… III. ①数理统计 ②计算机辅助计算—软件包, MATLAB IV. O212 TP391.75

中国版本图书馆 CIP 数据核字 (2008) 第 033993 号

出 版 者: 东北大学出版社

地址: 沈阳市和平区文化路 3 号巷 11 号

邮编: 110004

电话: 024—83687331 (市场部) 83680267 (社务室)

传真: 024—83680180 (市场部) 83680265 (社务室)

E-mail: neuph @ neupress.com

http: // www. neupress. com

印 刷 者: 抚顺光辉彩色广告印刷有限公司

发 行 者: 东北大学出版社

幅面尺寸: 184mm×228mm

印 张: 15.25

字 数: 325 千字

出版时间: 2008 年 3 月第 1 版

印刷时间: 2008 年 3 月第 1 次印刷

责任编辑: 刘乃义

封面设计: 唐敏智

责任校对: 郎 坤

责任出版: 杨华宁

ISBN 978-7-81102-519-4

定 价: 27.00 元

前 言

数理统计是信息处理、科学决策的重要理论与方法,其内容丰富,逻辑严谨,实践性强,应用广泛,是现代管理、科研和工程技术人员必备的基础知识.随着计算机的普及和统计分析软件的发展,数理统计日益走进人们的日常工作和生活.如何使数理统计的学习者兼为统计分析与数据处理的实践者,是高等学校数理统计课程教学改革的重要课题,本书反映了作者在这一方面的思考和实践.

本书是辽宁工程技术大学“十一五”规划教材,源于几年的教学讲义,具有如下特点:第一,结合实例引入基本概念与统计分析方法,注重对问题、解决问题的基本思想的阐述;第二,弱化了公式和定理的纯数学推导,重视统计算法与计算机实现的衔接,将统计分析与 MATLAB 数据处理融为一体;第三,突出了知识的技能化和应用意识的养成.通过对本书的学习,读者不仅可以掌握数理统计的基础内容,而且可以初步掌握 MATLAB 数据处理的基本方法和技巧.

本书可作为工科硕士研究生应用数理统计课程的基础教材、非数学与统计类专业本科高年级学生选修教材,也可作为管理、科研和工程技术人员的参考读物.

在本书出版之际,感谢郭嗣琮教授的推荐,感谢辽宁工程技术大学“十一五”规划教材评审专家的认可,感谢辽宁工程技术大学教务处和理学院领导的关心和支持,感谢在本书编写过程中给予热情帮助的同事们.限于作者水平,缺憾仍存,恳请读者批评指正.

作 者

2008 年 2 月于辽宁工程技术大学

目 录

第 1 章 MATLAB 数据处理入门	1
1.1 数值矩阵的建立与基本操作	1
1.1.1 数值矩阵的建立	1
1.1.2 矩阵的基本操作	4
1.2 基本数学运算与常用函数	6
1.2.1 基本数学运算	6
1.2.2 统计数据处理常用的函数	12
1.3 数据图形化的常用指令与图形的简单修饰	16
1.3.1 数据图形化的常用指令	16
1.3.2 图形的简单修饰	18
1.4 运算流程的控制与指令集的函数化	22
1.4.1 运算流程的控制	22
1.4.2 指令集的函数化	24
1.4.3 M-文件的保护	26
习题 1	26
第 2 章 统计分析的基本概念、工具与推理基础	28
2.1 变量与数据的基本概念	28
2.1.1 变量及其概率分布	28
2.1.2 变量的观测与数据	33
2.2 统计分析的基本工具	36
2.2.1 统计量	36
2.2.2 数据特征的度量及其 MATLAB 函数	37
2.3 统计分析的推理基础	38
2.3.1 常用的统计分布与 α 分位数	38
2.3.2 基于正态分布的常用抽样分布	45
2.3.3 顺序统计量的抽样分布	48
习题 2	49

第3章 统计估计	50
3.1 变量分布形态的估计	50
3.1.1 频率分布表与频率直方图	50
3.1.2 经验分布函数	55
3.1.3 五数概括与 box 图	57
3.2 变量分布参数的估计	60
3.2.1 参数估计的方法	60
3.2.2 估计量的性能分析	66
3.2.3 估计误差的评价与控制	70
习题 3	78
第4章 假设检验	81
4.1 假设检验概述	81
4.1.1 假设检验的思维逻辑	81
4.1.2 假设检验的基本步骤	83
4.1.3 检验的 p 值	85
4.1.4 假设检验中的两类错误与势函数	86
4.1.5 假设检验与区间估计的关系	88
4.2 变量分布参数的检验	90
4.2.1 正态变量均值与方差的假设检验	90
4.2.2 两个正态变量均值与方差的比较	93
4.2.3 非正态变量分布参数的检验	97
4.3 变量分布形态的检验	107
4.3.1 K. Pearson-Fisher 检验	108
4.3.2 Колмогоров-Смирнов 检验	119
4.3.3 正态性检验	123
习题 4	128
第5章 方差分析	131
5.1 方差分析概述	131
5.2 单因子方差分析	133
5.2.1 单因子试验的统计模型及检验方法	133
5.2.2 效应与误差方差的估计	140

5.2.3 重复数相同的方差分析	142
5.2.4 多重比较	144
5.2.5 方差齐性检验	147
5.3 双因子方差分析	149
5.3.1 无交互作用的双因子方差分析	150
5.3.2 有交互作用的双因子方差分析	152
习题 5	156
第 6 章 回归分析	161
6.1 一元线性回归分析	161
6.1.1 一元线性回归模型	161
6.1.2 模型参数的估计	163
6.1.3 回归方程的显著性检验	165
6.1.4 利用回归方程进行预测	166
6.1.5 目标函数可线性化的曲线回归分析	168
6.2 多元线性回归分析	172
6.2.1 多元线性回归模型	172
6.2.2 模型参数的估计	173
6.2.3 回归方程的显著性检验	174
6.2.4 利用回归方程进行预测	176
6.2.5 最优回归方程的选择	177
6.3 偏最小二乘回归分析	181
6.3.1 偏最小二乘回归方法的数据结构与建模思想	182
6.3.2 偏最小二乘回归方法的算法步骤	182
6.3.3 偏最小二乘回归方法的辅助分析	185
习题 6	195
附录 A MATLAB 的基本函数	201
附录 B MATLAB 常用统计分析函数	209
附录 C 正文中缺省的 M-文件	224
参考文献	235

第 1 章 MATLAB 数据处理入门

在统计应用过程中,使用计算机进行数据处理是一种必然的选择.目前,已经有许多大型统计分析软件可供人们使用,最为著名的如 SAS, SPSS 和 SYSTAT 系统等.在这里我们选择 MATLAB 作为统计数据处理工具,不是因为它有更为强大的统计分析能力,而是因为 MATLAB 系统在科学计算领域的通用性,它更适合在专业众多的工科院校普及.

用 MATLAB 进行数据处理,可以在 MATLAB 的指令窗口(Command Window)中进行,也可以在 MATLAB Notebook 环境中进行(这是一种 MATLAB 与 Microsoft Word “无缝”链接的产物,是文字处理、数学计算和图形绘制一体化的工作环境).本章简要介绍在 MATLAB 的工作环境下进行数据处理的基础知识.对于此前尚未接触过 MATLAB 的读者,本章内容可作为入门教程;对于已经了解 MATLAB 的读者,本章内容可作为课余阅读资料,教学中略过.

1.1 数值矩阵的建立与基本操作

1.1.1 数值矩阵的建立

MATLAB 语言主要的数据对象是数值矩阵. MATLAB 语言中,数值矩阵的输入方法有直接输入法、文件装载法、函数生成法.

1.1.1.1 直接输入法

直接输入法是由赋值语句实现的.赋值语句的基本结构是:

$$\text{赋值变量} = \text{赋值表达式}$$

赋值变量通常是用户自定义变量,变量名是由英文字母引导的,由字母、数字和下划线组成;MATLAB 对字母的大小写是敏感的.

用赋值语句建立 m 行 n 列的二维数值矩阵(以下简称矩阵)的基本格式是:

$$\text{矩阵} = [\text{数据列表}]$$

右侧的一对中括号“[]”是矩阵定义符,其中的数据列表排列成 m 行 n 列,每一行用分号“;”区分,行中元素用逗号“,”区分(逗号也可用一个空格代替).

【例 1.1】 建立一个 3×4 的矩阵 A.

```
A = [1,2,3,4; 5,6,7,8; 9,10,11,12]
```

上述指令的运行结果是:

```
A =
     1     2     3     4
     5     6     7     8
     9    10    11    12
```

说明: 输入必须在英文状态下进行. 在定义符“[]”的结尾处(外侧)可以缀加一个分号“;”, 此时仅向 MATLAB 的工作内存输入了一个矩阵 A, 但在当前的工作窗口中不输出(显示)这个矩阵.

在指令窗口中, 输入一条指令后单击“Enter”键即可运行这一指令. 在 Notebook 环境中, 需要用鼠标选中这一指令, 按住“Ctrl”键, 然后单击“Enter”键即可运行这一指令.

【例 1.2】 当矩阵退化为一个数或一个向量时, 可以由一个代数表达式加以定义, 而不必使用矩阵定义符“[]”.

```
e = eps
r = 2 * (3 + 2 * i)
P = (1:2:8) * pi
```

上述指令实际上是向 MATLAB 的工作内存输入了下列数值或向量:

```
e =
    2.2204e-016
r =
    6.0000 + 4.0000i
P =
    3.1416    9.4248   15.7080   21.9911
```

说明: ① 表达式中的 eps, i, pi 都是 MATLAB 的保留常数. MATLAB 常用的保留常数与保留变量如下:

保留常数或变量	常数或变量的意义
eps	机器零阈值, 2.2204×10^{-16}
i, j	虚数单位
pi	圆周率 π
Inf	$+\infty$ 的 MATLAB 表示
NaN	不定式 $0/0$ 或 ∞/∞ 的 MATLAB 表示
ans	预定义缺省输出变量

这些不允许用户在自定义变量时使用.

② 最后一个赋值语句给出了向 MATLAB 输入一个等差数列的方法: 在表达式

$a:d:b$ 中, a 是数列的首项, d 是数列的公差, b 是数列的上界(可能是最后一项或比最后一项大的数)。

1.1.1.2 文件装载法

对于大规模的矩阵, 通常预先编写数据文件存盘, 然后使用“load”语句从数据文件中直接读入。

使用 MATLAB 系统自身的“内存变量编辑器”(Array Editor)编写数据文件十分方便可靠, 操作方法是:

- ① 在指令窗口中向指定的新变量赋“空”矩阵, 如 `byk = []`;
- ② 在“内存变量浏览器”(Workspace)中双击该变量, 启动“内存变量编辑器”;
- ③ 在“内存变量编辑器”弹出的空白表格中, 每一个单元格对应矩阵的一个元素, 填写具体数值;
- ④ 保存该变量为数据文件, 如文件名为 `byk`, 保存到 MATLAB 系统根目录下的 `work` 子目录中(或用户自己的工作目录下. 变量名与数据文件名可以不一致, 一个数据文件也可以包含多个数据变量. 详细的内容请参阅其他 MATLAB 专门教程)。

在需要调用这个数据文件时, 只需运行指令 `load byk` 即可. 也可以利用任何一个文本编辑器(如 Microsoft Excel)编写这个数据文件, 注意要保存为纯文本文件, 如 `byk.txt`, 在需要调用这个数据文件时, 只需运行指令 `load byk.txt`。

1.1.1.3 函数生成法

在一些特殊的场合, 需要用 MATLAB 定义的用来构造特殊矩阵的函数向系统输入数据. 使用 MATLAB 定义的函数称为函数的调用, 这是 MATLAB 应用的一项重要技能. MATLAB 函数调用语句的基本结构是:

[返回变量列表] = 函数名(输入变量列表)

其中, 返回、输入变量列表中均可包含若干个变量, 变量名之间用逗号分隔. 常用的构造特殊矩阵的函数有:

函数及调用格式	函数功能
<code>Z = zeros(r, c)</code>	生成元素全为 0 的 $r \times c$ 矩阵 Z
<code>O = ones(r, c)</code>	生成元素全为 1 的 $r \times c$ 矩阵 O
<code>E = eye(r, c)</code>	生成对角线为 1、其他元素全为零的 $r \times c$ 矩阵 E
<code>D = diag(x)</code>	生成以向量 x 的元素为对角元的对角矩阵 D

在统计研究中, 往往需要构造服从某一特定分布的随机数矩阵, 这方面的函数较多, 在后面的内容中会有这方面的应用, 关于此类函数详细的内容请参阅本书附录 B。

【例 1.3】 特殊矩阵的函数生成。

`Z = zeros(2, 3)` % 生成 2×3 的全零矩阵

O = ones(3) % 生成 3 阶全 1 方阵; 方阵只需输入行数

E = eye(3,4) % 生成 3×4 的对角线为 1 的矩阵

D = diag(1:5) % 生成对角元为 1、2、3、4、5 的 5 阶对角矩阵

上述指令的运行结果是:

Z =

```
0    0    0
0    0    0
```

O =

```
1    1    1
1    1    1
1    1    1
```

E =

```
1    0    0    0
0    1    0    0
0    0    1    0
```

D =

```
1    0    0    0    0
0    2    0    0    0
0    0    3    0    0
0    0    0    4    0
0    0    0    0    5
```

说明: ① 表达式中的百分号“%”是 MATLAB 的注释符, 用来对指令中的某些内容进行说明。“%”必须在英文状态下输入, 其后的内容可以在中文状态下输入, 指令运行时不执行这部分内容。

② **diag** 是一个双向操作函数, 当输入参数 x 是一个向量时, 输出(返回)以这个向量为对角元的对角矩阵; 而当输入参数 x 是一个方阵时, 则返回由这个方阵的对角元构成的列向量。

1.1.2 矩阵的基本操作

下面介绍关于矩阵元素的寻访与修改, 以及矩阵的裁剪与拼接等矩阵操作方法。此类技能是使用 MATLAB 进行数据处理所必须的。

【例 1.4】 矩阵元素的寻访与修改。

A = [1,2,3,4; 2,3,4,5; 3,4,5,6; 4,5,6,7] % 创建一个供操作的矩阵

A23 = A(2,3) % 寻访(取出)A 的第 2 行、3 列交叉位置的元素

A(2,2)=0 % 将 A 的第 2 行、2 列交叉位置元素赋值为零

上述指令的运行结果是:

A =

1	2	3	4
2	3	4	5
3	4	5	6
4	5	6	7

A23 =

4

A =

1	2	3	4
2	0	4	5
3	4	5	6
4	5	6	7

【例 1.5】 矩阵的裁剪(提取某些行、列,或删除某些行、列).

AR3 = A(3,:) % 取出 A 的第 3 行

AC2 = A(:,2) % 取出 A 的第 2 列

AR13 = A(1:2:3,:) % 取出 A 的第 1、3 两行

AR23C14 = A(2:3,4:-3:1) % 取出 A 的第 2、3 行与第 4、1 列交叉位置元素

A(:,4) = [] % 删除 A 的第 4 列, 矩阵的变量名不变

上述指令的运行结果是:

AR3 =

3	4	5	6
---	---	---	---

AC2 =

2

0

4

5

AR13 =

1	2	3	4
---	---	---	---

3	4	5	6
---	---	---	---

AR23C14 =

5	2
---	---

6	3
---	---

```
A =  
    1     2     3  
    2     0     4  
    3     4     5  
    4     5     6
```

由例 1.5 可知, 冒号“:”在 MATLAB 语言的矩阵操作中发挥着剪刀的作用.

【例 1.6】 矩阵的拼接(已知矩阵的扩展, 或几个矩阵合并成一个新矩阵).

```
B = [A, ones(4, 2)] % 在 A 的右边拼接 ones(4, 2)
```

```
C = [A(1:2, :); eye(3)] % 在 A 的 1、2 两行下边拼接 eye(3)
```

```
D = [A(1:2, 2:3), zeros(2); ones(2, 4)] % 在 A(1:2, 2:3) 右接 2 阶零矩阵, 然后下  
接 2×4 全 1 矩阵
```

上述指令的运行结果是:

```
B =  
    1     2     3     1     1  
    2     0     4     1     1  
    3     4     5     1     1  
    4     5     6     1     1
```

```
C =  
    1     2     3  
    2     0     4  
    1     0     0  
    0     1     0  
    0     0     1
```

```
D =  
    2     3     0     0  
    0     4     0     0  
    1     1     1     1  
    1     1     1     1
```

1.2 基本数学运算与常用函数

1.2.1 基本数学运算

MATLAB 数学运算的对象是矩阵, 也就是说, 要理解 MATLAB 中的数学运算, 关

键是把握各种运算符作用于矩阵的规则。

1.2.1.1 矩阵的代数运算

MATLAB 语言提供了如下矩阵代数运算的运算符：

'转置 + 加法 - 减法 * 乘法 ^ 乘幂 \ 左除 / 右除

上述运算符的运算规则遵循了线性代数教程中的相关定义。矩阵代数运算要求维数相容，否则将产生错误信息。这里，需要特别强调的是：

① 矩阵的转置“'”是指矩阵的共轭转置；

② 矩阵的左除“\”和右除“/”的含义是：设 A 是可逆矩阵，则 $AX=B$ 的解是 A 左除 B ，即 $X=A \setminus B$ (若 B 为列向量，则 X 为方程组的解)； $XA=B$ 的解是 A 右除 B ，即 $X=B/A$ (若 B 为列向量，则 X 为方程组的解)。

1.2.1.2 矩阵的标量批处理运算

MATLAB 在需要的时候可以将矩阵视为普通的行列排列整齐的数据集合，通常称为数组。矩阵与数组在形式上是一样的，但却是两个不同的概念。当对一个矩阵(数组)施行标量批处理运算时，这个矩阵就失去了线性代数教程中矩阵的意义而成为一个数组了。MATLAB 语言提供了如下标量批处理运算的运算符：

.' 转置 .* 乘法 .^ 乘幂 .\ 左除 ./ 右除

上述运算符俗称“点运算”，其运算规则是两个数组的对应元素之间的运算。标量批处理运算要求数组的维数相同，否则将产生错误信息。特别地，这里的转置“.’”是非共轭转置。

1.2.1.3 矩阵的关系运算

MATLAB 语言提供了如表 1.1 所示的关系运算符。

表 1.1 MATLAB 语言提供的关系运算符

符号	<	>	<=	>=	==	~=
意义	小于	大于	不大于	不小于	等于	不等于
语法	$A < B$	$A > B$	$A \leq B$	$A \geq B$	$A == B$	$A \sim B$

关系运算是在两个数值之间进行比较，当给定的关系成立时返回数值 1 (表示关系真)，否则返回数值 0 (表示关系假)。当关系运算作用于一个标量与一个矩阵时，是标量与矩阵的每一个元素进行比较，返回一个与参与运算的矩阵同型的由 0 和 1 构成的矩阵；当关系运算作用于两个同型矩阵时，是两个矩阵的对应元素之间进行比较，返回一个由 0 和 1 构成的同型矩阵。

1.2.1.4 矩阵的逻辑运算

MATLAB 语言提供了如表 1.2 所示的逻辑运算符。

表 1.2 MATLAB 语言提供的逻辑运算符

符号		&		~
意义		与	或	非
语法		A & B	A B	~A
A	B			
0	0	0	0	1
1	0	0	1	0
1	1	1	1	0

逻辑运算也是在两个数值之间进行的, 运算过程中将任何非零元素视为 1(真). 当逻辑运算作用于一个标量与一个矩阵时, 运算在标量与矩阵的每一个元素之间进行, 返回一个与参与运算的矩阵同型的由 0 和 1 构成的矩阵; 当逻辑运算作用于两个同型矩阵时, 运算在两个矩阵的对应元素之间进行, 返回一个由 0 和 1 构成的同型矩阵.

MATLAB 语言关于运算优先级的规定与数学中的规定是一致的.

【例 1.7】 两种转置运算的区别.

`H = [(1:3) + (2:4) * i; 1 * i, -1 * i, 3] % 创建一个供操作的矩阵`

`H1 = H'`

`H2 = H. '`

上述指令的运行结果是:

H =

```

1.0000 + 2.0000i    2.0000 + 3.0000i    3.0000 + 4.0000i
0 + 1.0000i        0 - 1.0000i        3.0000

```

H1 =

```

1.0000 - 2.0000i    0 - 1.0000i
2.0000 - 3.0000i    0 + 1.0000i
3.0000 - 4.0000i    3.0000

```

H2 =

```

1.0000 + 2.0000i    0 + 1.0000i
2.0000 + 3.0000i    0 - 1.0000i
3.0000 + 4.0000i    3.0000

```

【例 1.8】 两种乘、除和乘方运算的比较.

① 创建两个供操作的矩阵.

`A = [1, 2, 3; 0, 1, 2; 0, 0, 1]`

`B = [0, 0, 1; 0, 2, 1; 3, 2, 1]`

创建的两个矩阵是:

A =

1	2	3
0	1	2
0	0	1

B =

0	0	1
0	2	1
3	2	1

② 两种乘法运算的比较.

M1 = A * B

M1 = A. * B

这两个指令的运行结果是:

M1 =

9	10	6
6	6	3
3	2	1

N1 =

0	0	3
0	2	2
0	0	1

③ 两种除法运算的比较.

M2 = B/A

N2 = B./A

这两个指令的运行结果是:

M2 =

0	0	1
0	2	-3
3	-4	0

Warning: Divide by zero.

N2 =

0	0	0.3333
NaN	2.0000	0.5000
Inf	Inf	1.0000

④ 两种乘方运算的比较.

M3 = A^2

M3 = A.^2

这两个指令的运行结果是:

M3 =

1	4	10
0	1	4
0	0	1

N3 =

1	4	9
0	1	4
0	0	1

此外, 还应注意标量与矩阵运算的含义: 标量与矩阵的运算是标量与矩阵的每个元素之间的运算.

【例 1.9】 标量与矩阵的运算.

M4 = A * 10

M4 = A. * 10

这两个指令的运行结果是:

M4 =

10	20	30
0	10	20
0	0	10

N4 =

10	20	30
0	10	20
0	0	10

【思考题】 想一想, 下列运算的结果是什么?

① $5/A$, 这个运算有意义吗?

② $5./B$ 和 $B.\setminus 5$, 这两种运算有意义吗? 运算的结果相同吗?

③ $A./B$ 和 $B.\setminus A$, 这两种运算的结果相同吗?

④ A/B 和 $B\setminus A$, 这两种运算的结果相同吗?

【例 1.10】 标量与矩阵、矩阵与矩阵的关系运算.

x = 5

y = 5 * ones(3, 3)