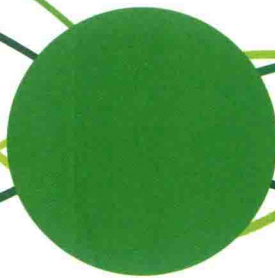
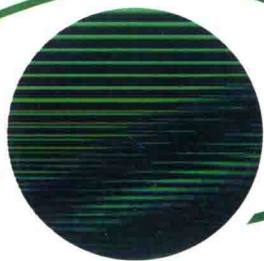




金融数据分析与软件应用系列教材

R 软件与 金融数据分析

主编◎方霞



本书以R软件的操作为线索，从R软件的简介讲起，至R软件的基本操作、R软件编程、作图基础；之后与金融数据分析相结合，重点讲述了债务相关计算、R软件与商业银行风险管理、股票相关计算、R软件与量化投资等内容。本书紧密联系实际，操作性强，既可作为教材，也可供实务界人士参考使用。

 中国金融出版社

本书受浙江省高校人文社会科学浙江工商大学金融学重点研究基地资助

金融数据分析与软件应用系列教材

R软件与金融数据分析

(The Application of R in Finance)

主 编 方 霞

副主编 万 谍 张水泉



中国金融出版社

责任编辑：王效端 王 君

责任校对：张志文

责任印制：张也男

图书在版编目 (CIP) 数据

R 软件与金融数据分析 (R Ruanjian yu Jinrong Shuju Fenxi) /方霞主编. —北京:
中国金融出版社, 2017. 8

金融数据分析与软件应用系列教材

ISBN 978 - 7 - 5049 - 9125 - 6

I. ①R… II. ①方… III. ①金融统计—统计分析—应用软件—教材
IV. ①F830. 2

中国版本图书馆 CIP 数据核字 (2017) 第 187874 号

出版
发行

中国金融出版社

社址 北京市丰台区益泽路 2 号

市场开发部 (010)63266347, 63805472, 63439533 (传真)

网上书店 <http://www.chinafph.com>

(010)63286832, 63365686 (传真)

读者服务部 (010)66070833, 62568380

邮编 100071

经销 新华书店

印刷 北京市松源印刷有限公司

装订 平阳装订厂

尺寸 185 毫米 × 260 毫米

印张 21

字数 474 千

版次 2017 年 8 月第 1 版

印次 2017 年 8 月第 1 次印刷

定价 48.00 元

ISBN 978 - 7 - 5049 - 9125 - 6

如出现印装错误本社负责调换 联系电话 (010) 63263947

编辑部邮箱: jiaocaiyibu@126.com

序 言

近十年来，中国金融业超常规快速发展，互联网金融、私募金融等新金融业态发展迅猛，新型机构不断涌现，市场规模持续扩大。金融市场与行业的发展变化对人才培养提出了新的需求和更高的要求，高校是金融人才队伍培养最重要的阵地，是金融人才的主要来源，应主动适应社会对金融专业人才需求的发展趋势，努力提高人才培养质量。作为典型的应用性学科，高校金融专业人才培养应高度重视实验教学，培养具有创新精神和实践能力的应用型、复合型、创新型人才。在此背景下，浙江工商大学金融学院决定依托软件操作实验课程，促进理论教学与实践教学有机结合，提升学生金融数据分析能力，提高金融人才培养质量。为此，我们决定在全院范围内共开设四门依托软件的金融数据分析课程，供学生修读，涉及 Excel、Matlab、R 语言和 SAS，在多年教学实践的基础上，形成了相对完整的教案。由于国内目前尚缺少以金融数据分析为核心的相关系列实验教材，所以，浙江工商大学金融学院决定在相关课程教案的基础上，组织编写《Excel 与金融数据分析》、《Matlab 与金融数据分析》、《R 语言与金融数据分析》和《SAS 与金融数据分析》系列实验教材。

本系列丛书旨在通过数据处理和软件应用来强化巩固前期部分金融专业知识和基础理论。书稿框架在浙江省高校金融人才培养与教学改革研讨会、钱塘江金融港湾高等金融教育论坛等会议上多次讨论，校内外金融专业教师和银行、证券、保险、私募金融、互联网金融等机构高管、技术骨干提出了宝贵的修改意见。金融学院成立书稿写作协调小组，组织编撰工作，本系列丛书的编写有以下几个特点：（1）注重理论基础与实务前沿、经典与当代相结合，基础能力训练与创新能力培养相结合。（2）与兄弟院校及科研院所相关教师、金融机构技术骨干联合编写。（3）将传统与新兴金融机构的最新案例进行适当简化，既考虑实用性，又兼顾可读性和可操作性。（4）四本书稿在金融理论基础和案例应用等方面既具有统一性，又存在差异性。书稿初稿完成之后，在日常授课中进行试用，并根据学生反馈进一步修改完善，历时两年完成定稿。

如果你觉得金融学相关知识和基础理论枯燥难懂，亦或觉得不知道如何将这些理论应用于实践，那么本系列丛书将教会你如何通过软件实现基于理论的金融数据分析。如果你觉得在大数据时代，不知道如何从庞大的金融数据中提炼出相关信息，那么本系列丛书将会提供一些合理的解决方案，供你获取金融数据背后的逻辑。本系列丛书除了适合金融学类专业本科生和研究生使用，还适合金融机构的相关从业人员使用，同样适合具有一定基础并对金融问题感兴趣的读者使用。

在本系列丛书的撰写过程中，要感谢中国金融出版社教材一部王效端主任的大力支持，还要感谢王君老师在书稿校对期间的细心和耐心。书稿同时得到杭州银行人力资源

部李晓华总经理、中国银行浙江省分行渠道与运营管理部王晔副总经理、蚂蚁金融服务集团平台风险部唐科伟总经理、新华人寿保险股份有限公司浙江分公司人力资源部张立微总经理、浙江核新同花顺网络信息股份有限公司高校事业部张晓伟总监，以及财通证券股份有限公司、永安期货股份有限公司、瑞控股有限公司和杭州龙旗科技有限公司部门负责人的大力帮助。他们为书稿的撰写提供了大量案例和素材，并提出了宝贵的建设性意见。最后，要感谢浙江工商大学金融学院陆霞、闫存、章倍榕、钱超男、成祺炯、颜义人、金惠芳、钱盈、钟强、陈庐、周渐、杨正、程锦文、李中明、李佩琛、刘慧、张妮、俞洁洁、李标标等同学对本系列丛书所做的贡献。

虽然作者在编写丛书的过程中投入了大量精力，力求严谨精确，但是由于自身学识和能力有限，书中难免存在不足和错误之处，真诚地希望广大读者不吝批评指正，或通过电子邮件 fangxia@zjgsu.edu.cn 反馈意见。

钱水土

2017年8月15日

目 录

第1章 R 简介/1

【本章学习目标】/1

- 1.1 R 是什么/1
- 1.2 Rstudio 介绍/2
 - 1.2.1 主界面/2
 - 1.2.2 工作空间 (Environment) 和历史 (History) 窗口/4
 - 1.2.3 命令窗口/5
 - 1.2.4 右下角文件窗口/5
- 1.3 R 的扩展包/6
 - 1.3.1 扩展包的安装/6
 - 1.3.2 扩展包的载入/7
- 1.4 如何获取帮助/7
 - 1.4.1 查看扩展包和函数的信息/7
 - 1.4.2 查看源代码的信息/8
 - 1.4.3 查看网络帮助信息/8
- 1.5 工作空间/9
- 1.6 文件输入和输出/9
 - 1.6.1 文件输入/10
 - 1.6.2 文件存储/10
 - 1.6.3 图形输出/10
- 1.7 数据类型/11
- 1.8 数据结构/12
 - 1.8.1 向量 (vector) /12
 - 1.8.2 矩阵 (matrix) /18
 - 1.8.3 数组 (array) /23
 - 1.8.4 列表 (list) /25
 - 1.8.5 数据框/26
 - 1.8.6 因子/29

【习题】/31

【参考文献】/32

第2章 R 基本操作/33

【本章学习目标】/33

- 2.1 数据的输入/33
 - 2.1.1 手工输入数据/33
 - 2.1.2 从文本文件中读取数据/36
 - 2.1.3 读取 Excel 表格数据/37
 - 2.1.4 从 SAS、SPSS、Stata 统计软件读取数据/39
 - 2.1.5 从网页读取数据/41
 - 2.1.6 从数据库读取数据/42
 - 2.1.7 读取包中数据/43
- 2.2 数据输出/43
- 2.3 特殊数据处理/45
 - 2.3.1 缺失值处理/45
 - 2.3.2 日期数据/46
 - 2.3.3 基本函数及其运算/48
 - 2.3.4 zoo 包简介/49
 - 2.3.5 xts 包简介/51
- 2.4 数据预处理/52
 - 2.4.1 产生新变量/52
 - 2.4.2 修改变量名 (数据标签) /54
 - 2.4.3 变量重新编码/57
 - 2.4.4 行列删除/59
 - 2.4.5 数据排序/60
- 2.5 数据重塑/61
 - 2.5.1 选取数据的子集/61
 - 2.5.2 数据合并/63
 - 2.5.3 长宽格式的转换/68

【习题】/75

【参考文献】/75

第3章 R编程基础/76

【本章学习目标】/76

3.1 流程控制/76

3.1.1 循环执行结构/76

3.1.2 条件执行结构/78

3.2 编写函数/80

3.2.1 泛型函数/80

3.2.2 函数构建/82

3.3 常用R函数/85

3.3.1 统计分析函数/85

3.3.2 回归分析/93

3.3.3 lapply() 系列函数/116

【习题】/123

【参考文献】/123

第4章 R作图基础/124

【本章学习目标】/124

4.1 初建图形/124

4.1.1 图形文件保存和创建/124

4.1.2 高级绘图函数/125

4.1.3 低级绘图函数/128

4.2 图形参数/129

4.2.1 图形类型/130

4.2.2 符号参数/131

4.2.3 线条参数/132

4.2.4 线条和字符颜色: col 参数/133

4.2.5 坐标轴设定与调整/133

4.2.6 添加文本/140

4.2.7 添加图例/142

4.2.8 添加网格和标记/144

4.2.9 图尺寸调整/144

4.3 图形工具/147

4.3.1 高级散点图/147

4.3.2 直方图/151

4.3.3 Q-Q图/153

4.3.4 饼图/156

4.3.5 条形图/158

4.3.6 箱线图/163

4.3.7 克里夫兰点图/164

4.3.8 三维图/167

4.4 多图环境/169

4.4.1 利用 par() 函数/169

4.4.2 利用 layout() 函数/171

4.5 使用高级制图包 (ggplot2 package) /173

4.5.1 从 qplot() 快速入门/173

4.5.2 从 qplot() 到 ggplot() /176

4.5.3 用 ggplot() 例子演示/178

【习题】/181

【参考文献】/181

第5章 债券相关计算/182

【本章学习目标】/182

5.1 债券应计利息的计算/182

5.1.1 日期处理相关函数/183

5.1.2 剩余利息支付次数的计算/185

5.1.3 上一次利息支付日的获取/186

5.1.4 下次利息支付日的获取/187

5.1.5 计息期是否包含闰日的判断/187

5.1.6 日计数因子的计算/188

5.1.7 应计利息的计算/190

5.2 债券内在价值的计算/191

5.3 债券收益率的计算/192

5.3.1 到期收益率/192

5.3.2 实现的收益率/193

5.4 债券久期的计算/195

5.5 债券凸性的计算/197

5.6 债券相关计算举例/198

5.6.1 债券基本信息/198

5.7 应用/202

【习题】/203

【参考文献】/204

第 6 章 R 与商业银行风险管理/205

【本章学习目标】/205

6.1 市场风险度量/205

- 6.1.1 VaR 的定义与计算方法/205
- 6.1.2 VaR 的回测/215
- 6.1.3 VaR 计算的改进：极值分布理论，条件 VaR/217
- 6.1.4 基于风险因子的市场风险管理：多因子风险测算与对冲模型/222

6.2 信用风险管理/224

- 6.2.1 信用风险的成因和度量 (PD、EAD、LGD) /224
- 6.2.2 Merton 违约模型及 KMV 模型/227
- 6.2.3 统计违约风险模型与 PD 预测/233
- 6.2.4 利差风险与违约密度模型/245
- 6.2.5 投资组合信用风险/247

6.3 操作风险度量/252

- 6.3.1 操作风险的定义和评估方法/252
- 6.3.2 损失分布法/254
- 6.3.3 RAROC 方法简介/260
- 6.3.4 模型风险简介/261

6.4 流动性风险/263

- 6.4.1 流动性风险的定义/263
- 6.4.2 流动性风险的度量/263
- 6.4.3 流动性的管理方法/265

【习题】/267

【参考文献】/267

第 7 章 股票相关计算/269

【本章学习目标】/269

7.1 股票行情和历史数据的获取/269

- 7.1.1 雅虎金融数据/269
- 7.1.2 利用 R 获取雅虎金融数据/271
- 7.1.3 利用 R 包 stockPortfolio 获取股票历史收益数据/273

7.2 资产 β 系数的估计/2747.2.1 历史 β 系数和 β 系数的调整/2747.2.2 利用 R 估计股票的历史 β 系数/2767.2.3 β 系数的调整/278**7.3 投资组合优化/280**

- 7.3.1 投资组合优化简介/280
- 7.3.2 资产期望收益率和协方差矩阵的估计/280
- 7.3.3 投资组合的期望收益和风险/284
- 7.3.4 允许卖空下投资组合优化/284
- 7.3.5 不允许卖空下投资组合优化/287
- 7.3.6 利用 stockPortfolio 包进行组合优化/290

【习题】/292

【参考文献】/293

第 8 章 R 与量化投资/294

【本章学习目标】/294

8.1 量化投资概述/294

- 8.1.1 量化投资盈利的理论基础/295
- 8.1.2 量化投资与基本面分析和技术分析的区别与联系/295
- 8.1.3 量化投资理念的来源/296
- 8.1.4 量化投资的优势/297
- 8.1.5 量化策略设计的原则和注意事项/297

8.2 程序化交易简介/298**8.3 常见的量化投资策略介绍/303**

- 8.3.1 配对交易策略/303
- 8.3.2 Fama - French 多因子模型方法与 Alpha 策略/304
- 8.3.3 基于技术分析指标的选股策略/305

8.4 量化策略的回测、评价和改进 (评价指标构建、回溯检验方法) /308**8.5 策略资金管理技术简介/310****8.6 一个简单的策略演示/311**

【习题】/320

【参考文献】/321

图表目录

- 图 1-2-1 Rstudio 下载首页图/2
- 图 1-2-2 RStudio 主界面/3
- 图 1-2-3 RStudio 主界面四个功能区/3
- 图 1-2-4 RStudio 脚本编辑窗口 Tab 键/4
- 图 1-2-5 RStudio 脚本编辑窗口三个主要按钮/5
- 图 1-2-6 RStudio 工作空间和历史窗口/5
- 图 1-2-7 RStudio 文件窗口的 Files 和 Plots/6
- 表 1-5-1 用于管理 R 工作空间的函数/9
- 表 1-7-1 类型转换函数/11
- 表 1-8-1 数据结构/12
- 表 1-8-2 向量中常用函数/17
- 表 2-1-1 read.table 函数的详细参数/36
- 图 2-1-1 读取 Excel 文件/37
- 表 2-1-2 foreign 软件包的函数/40
- 表 2-2-1 函数 write.table 常用参数/44
- 表 2-3-1 R 中的日期代码/47
- 表 2-3-2 R 中的基本函数/49
- 表 2-4-1 向量算术运算符/52
- 图 2-4-1 利用交互式编辑器改变变量名/55
- 表 2-4-2 R 中常见的逻辑运算符/57
- 表 2-5-1 函数 reshape 常用参数/73
- 表 3-1-1 常见控制条件/79
- 图 3-3-1 编辑器显示 GDP 和发电量的数据/91
- 图 3-3-2 编辑器显示 2015 年和 2016 年的同期沥青价格数据/93
- 表 3-3-1 R 表达式中常用的符号/94
- 表 3-3-2 对拟合线性模型非常有用的其他函数/95
- 图 3-3-3 投资占 GDP 比重和储蓄占 GDP 比重关系图/96
- 图 3-3-4 拟合线与投资占 GDP 比重、储蓄占 GDP 比重散点图/97
- 图 3-3-5 回归结果诊断图/98
- 图 3-3-6 加上二次项后回归结果诊断图/98
- 表 3-3-3 Car 包里的回归诊断函数/99
- 图 3-3-7 储蓄占比和投资占比时序图/103
- 表 3-3-4 投资率和储蓄率的单位根检验/105
- 图 3-3-8 加元汇率、加拿大 CPI、美国 CPI 时序图/108
- 表 3-3-5 投资率和储蓄率的单位根检验/109
- 表 3-3-6 购买力平价检验：迹检验/110
- 表 3-3-7 购买力平价检验：最大特征值检验/111
- 图 3-3-9 岭迹图/115
- 图 4-1-1 plot () 函数散点图/126
- 图 4-1-2 plot () 函数回归诊断图/127
- 表 4-1-1 plot () 函数功能/127
- 图 4-1-3 hist () 函数画正态分布直方图/128
- 表 4-1-2 hist () 函数参数说明/128
- 图 4-1-4 point () 函数和 abline () 函数/129
- 表 4-2-1 type 参数指定的图形类型/130
- 图 4-2-1 type 参数指定的图形类型/130
- 图 4-2-2 参数 pch 可指定的绘图符号/131
- 图 4-2-3 参数 cex 为 3 和 1 时点大小对比/131
- 表 4-2-2 cex 参数说明/132
- 表 4-2-3 lty 参数指定的图形类型/132
- 图 4-2-4 参数 lwd/132
- 表 4-2-4 col 参数说明/133

- 图 4-2-5 画彩虹色的饼图/133
- 图 4-2-6 添加标题和坐标轴标签/133
- 图 4-2-7 title () 函数/134
- 图 4-2-8 curve () 函数中添加
ann = FALSE/134
- 表 4-2-5 axis () 参数说明/135
- 图 4-2-9 添加坐标轴/136
- 图 4-2-10 2001—2015 年沪深交易所上市
公司总市值/137
- 图 4-2-11 2001—2015 年沪深交易所上市
公司总市值图调整坐标轴 1/138
- 图 4-2-12 2001—2015 年沪深交易所上市
公司总市值图调整坐标轴 2/139
- 图 4-2-13 添加水平参考线图/140
- 图 4-2-14 用 text () 添加文字/140
- 表 4-2-6 font 参数说明/141
- 图 4-2-15 用 text () 添加文字, 对字体
进行修改/142
- 图 4-2-16 添加数学表达式/142
- 表 4-2-7 legend 参数说明/142
- 图 4-2-17 添加图例/143
- 图 4-2-18 添加网格线和标记点/144
- 图 4-2-19 图边界说明/144
- 表 4-2-8 边界参数说明/145
- 图 4-2-20 图边界参数设置效果差异/145
- 图 4-2-21 中信证券和广聚能源收盘价/146
- 图 4-2-22 图 4-2-21 的图边缘处理图/147
- 图 4-3-1 多组散点图/148
- 图 4-3-2 通过 group 得到多组散点图/148
- 图 4-3-3 多变量散点图/149
- 图 4-3-4 不同因子水平的散点图/150
- 图 4-3-5 散点图添加回归线/151
- 图 4-3-6 散点图添加平滑线/151
- 图 4-3-7 贵州茅台历年股息的简单直方图/152
- 图 4-3-8 对直方图添加密度线/152
- 图 4-3-9 离散型直方图/153
- 图 4-3-10 相对频率直方图/153
- 图 4-3-11 价格的时间序列图/154
- 图 4-3-12 上证指数价格对数收益率的
Q-Q 图/155
- 图 4-3-13 检验卡方分布图/156
- 表 4-3-1 pie () 函数参数介绍/156
- 图 4-3-14 2014 年中国 GDP 中三次产业占比
饼图/157
- 图 4-3-15 2014 年中国 GDP 中三次产业占比
扇形图/158
- 表 4-3-2 barplot () 函数参数介绍/158
- 图 4-3-16 2014 年三大产业 GDP 值的
条形图/160
- 图 4-3-17 2009—2014 年三大产业 GDP 水平
陈列条形图/160
- 图 4-3-18 2009—2014 年三次产业的 GDP 值
分段展示图/161
- 图 4-3-19 2009—2014 年均值条形图/161
- 图 4-3-20 2009—2014 年添加置信区间/162
- 图 4-3-21 2009—2014 年脊柱图/162
- 图 4-3-22 箱线图示例/163
- 图 4-3-23 公募基金复权单位净值增长率
箱线图/163
- 图 4-3-24 不同基金风格与基金业绩的关系
箱线图/164
- 图 4-3-25 克利夫兰点图/165
- 图 4-3-26 克利夫兰点图对比/167
- 图 4-3-27 三维散点图/167
- 表 4-3-3 三维曲面图参数/168
- 图 4-3-28 三维曲面图/169
- 图 4-4-1 用 par () 函数进行图形分割/170
- 图 4-4-2 用 fig () 函数图例/170
- 图 4-4-3 用 fig () 函数进行精细布局/171
- 图 4-4-4 用 layout () 函数进行图形布局/172
- 图 4-4-5 用 layout () 函数进行精细布局/173
- 图 4-5-1 按颜色分组画条形图/174

- 图 4-5-2 按切工分组画直方图/174
- 图 4-5-3 按切工分组画密度图/175
- 图 4-5-4 散点图添加平滑曲线/175
- 图 4-5-5 按颜色分组画扰动点图/175
- 图 4-5-6 按颜色分组画箱线图/176
- 表 4-5-1 常用几何对象图形属性及默认的统计变换/176
- 表 4-5-2 统计变换种类/177
- 表 4-5-3 位置调整种类/177
- 表 4-5-4 位置标度的参数/177
- 表 4-5-5 坐标系统函数种类/177
- 表 4-5-6 网格分面类型/178
- 图 4-5-7 ggplot() 做条形图/178
- 图 4-5-8 ggplot() 做箱线图/179
- 图 4-5-9 ggplot() 做扰动点图/180
- 图 4-5-10 ggplot() 做散点图/181
- 图 5-5-1 单独用久期来衡量债券价格变化所带来的误差/197
- 图 5-6-1 上海证券交易所官方网站“数据”条目/199
- 图 5-6-2 上海证券交易所债券列表一部分, 查询时间: 2016 年 6 月 17 日/200
- 图 5-6-3 上海证券交易所 05 国债(4) 基本信息/201
- 图 5-7-1 利率不变时债券的净价与全价/203
- 图 6-1-1 标准正态分布密度图及 VaR 定义/206
- 图 6-1-2 HS300 指数 2014 年的日收盘价和日收益率时序图/207
- 表 6-1-1 混合时间加权法的例子 ($n=100$, $\lambda=0.96$) /210
- 图 6-2-1 商业违约的双向性示意图/225
- 表 6-2-1 评级与违约概率的关系示例/226
- 表 6-2-2 LGD 与抵押品的关系示例/226
- 图 6-2-2 损失分布示意图/227
- 图 6-2-3 市场风险与信用风险的比较/227
- 图 6-2-4 Merton 模型的违约示意图/228
- 图 6-2-5 ST 超日 2010—2014 年日收盘价图/231
- 图 6-2-6 违约距离计算示意图/233
- 图 6-2-7 违约距离与违约密度的关系示意图/233
- 图 6-2-8 ROC 曲线示意图/236
- 图 6-2-9 CAP 曲线示意图/237
- 图 6-2-10 Logit 模型得出的 ROC 曲线/239
- 图 6-2-11 Logit 模型得出的 CAP 曲线/240
- 图 6-2-12 probit 模型得到的 ROC 曲线/241
- 图 6-2-13 probit 模型得到的 CAP 曲线/242
- 图 6-2-14 SVM 最大间隔法示意图/243
- 图 6-2-15 神经元示意图/244
- 图 6-2-16 神经网络分类得到的 ROC 曲线/244
- 图 6-2-17 神经网络分类得到的 CAP 曲线/245
- 图 6-2-18 公司债违约密度时序图/246
- 表 6-3-1 标准法下的各银行业务的资本要求/254
- 图 6-3-1 实际操作损失分布(以超额收益低于 -5% 记为发生操作损失事件, 操作损失记为股票当日收益率) /258
- 图 6-3-2 模拟出的操作损失分布(模拟次数 1000 次, 频率分布为 Poisson 分布, 损失分布为 GDP 分布) /258
- 表 6-4-1 可用稳定资金和所需稳定资金的种类及计算比重/266
- 图 7-1-1 雅虎财经证券代码查询框/270
- 图 7-1-2 通过雅虎财经获取上证综指历史交易数据/270
- 图 7-1-3 雅虎财经股票历史交易数据选择框/271
- 表 7-2-1 20 只样本股票历史 β 系数估计结果/278
- 表 7-2-2 20 只样本股票的 β 系数估计结果/279
- 图 7-3-1 投资有效边界/280
- 表 7-3-1 历史收益率法估计的样本股票期望收益/282
- 表 7-3-2 历史收益率法估计的 20 只样本股票

- 协方差矩阵的部分内容/282
- 表 7-3-3 单指数模型估计的样本股票月期望收益/283
- 图 7-3-2 允许卖空下资产的前沿组合/287
- 表 7-3-4 3 种方法构建的组的收益和风险/292
- 图 7-3-3 Markowitz 方法得到的 4 只样本股票的可行集和有效边界/292
- 图 8-1-1 最大挫折和最长挫折期示意图/298
- 图 8-2-1 程序化交易策略设计流程图/301
- 图 8-2-2 微软公司股票的日内交易量形态/302
- 图 8-3-1 支持向量回归机示意图/306
- 图 8-6-1 30 分钟突破策略示意图/311
- 表 8-6-1 30 分钟突破策略在股指期货主力合约的表现回测结果/320
- 图 8-6-2 30 分钟突破策略的每日收益分布密度图和资本金累积增长曲线/320
- 图 8-6-3 反转交易策略/321

第 1 章

R 简介

【本章学习目标】

1. 学会下载与安装 R、RStudio 软件，重点掌握 R、RStudio 软件的操作界面。
2. 学会下载和载入 R 软件的扩展包，了解各扩展包的功能，掌握 R 软件的帮助功能。
3. 学会查看并更改 R 的工作目录，掌握 R 脚本的创建、编辑、保存功能并熟悉 R 软件中文件的输入输出过程。
4. 建立 R 语言的环境概念，了解 R 软件包以及函数运行的环境。
5. 建立 R 语言的数据类型的概念，掌握不同数据类型间的转换。
6. 建立 R 语言数据结构的概念，掌握不同数据结构的创建、元素索引及基本运算等。

1.1 R 是什么

随着计算机技术、数据库技术、传感器技术和网络技术的快速发展和广泛应用，自动收集数据的功能得到空前的发展，各个行业所积累数据越来越多，金融行业尤其如此。R 是开发这些数据的很好的选择，因为它提供了大量的算法和方法以及它自由和开放的源码特性为金融分析提供了一体化的解决方案，能够帮助用户完成提取数据、数据重整、金融建模、交易回测等诸多功能等。R 中丰富的模型及分析方法如：线性模型、广义线性模型、非线性回归模型、时间序列分析、经典的参数和非参数检验、聚类和平滑方法等为我们实现金融投资分析的量化提供了方便的途径，此外还有大批灵活的作图程序使数据及模型的可视化成为可能。

那么 R 到底是什么？R 语言的创始人是 Ross Ihaka 和 Robert Gentleman，由于这两位“R 之父”的名字都是以 R 开头，所以就称之为 R 语言。其实 R 语言是一款免费下载的语言，它由一个语言系统（R 语言）和运行环境构成，后者包括图形、调试器（Debugger）、对某些系统函数的调用和运行脚本文件等。R 的主要功能是为人们提供分析、可视化和操作数据的工具，主要包括：数据存储和处理、运算、统计分析、制图功能；最

突出的是 R 提供各种数学计算、统计计算的函数，从而使使用者能灵活地进行数据分析，甚至创造出符合需要的新的自定义的计算方法。

我们可以根据电脑的操作系统下载相应的 R 语言安装文件。下载地址：<http://ftp.ctex.org/mirrors/CRAN/>。此外 R 的一款跨平台 IDE “RStudio 软件”把常用的窗口都整合在一起，对我们在操作时控制整个运行过程非常方便。我们可以在 <http://www.rstudio.org/RStudio> 实现下载并安装。

关于 R 更深入的介绍，请参考官方发行的手册和网站上的学习材料，如官方的六本手册（均可从 R 网站 <http://cran.r-project.org/manuals.html> 上找到）：

- AnIntroductiontoR (R - intro)
- RDataImport/Export (R - data)
- RInstallationandAdministration (R - admin)
- WritingRExtensions (R - exts)
- RInternals (R - ints)
- TheRLanguageDefinition (R - lang)

1.2 RStudio 介绍

与 R 软件相比，RStudio 是一款界面更为友善、更便捷的软件，它无须在命令行和绘图行窗口之间切换，界面可视性更好。使用 RStudio，除需要先安装 R 软件外，还需要安装适合电脑的 Rstudio 软件，下载地址为 <http://www.rstudio.org/download>。



图 1-2-1 Rstudio 下载首页图

在 Rstudio 下载首页选择下载，对于免费用户来讲可以选择 RStudio Desktop，然后选择与电脑操作系统相匹配的 RStudio 软件。

1.2.1 主界面

1. 主菜单。在 Rstudio 软件的最上方是工具栏，该工具栏上有多个命令按钮，单击 File 按钮可以实现文件创建、存取等功能，显示如图 1-2-2 所示，在下拉菜单中我们可以实现创建新文件 RscriptsR 脚本文件、SweavedocumentsSweave 文件、TeXdocuments-Tex 文件等。单击 RScript，建立一个 R 语言的代码文件。和主界面一样，Tab 键在该界面也能实现同样的显示函数和函数参数的功能。

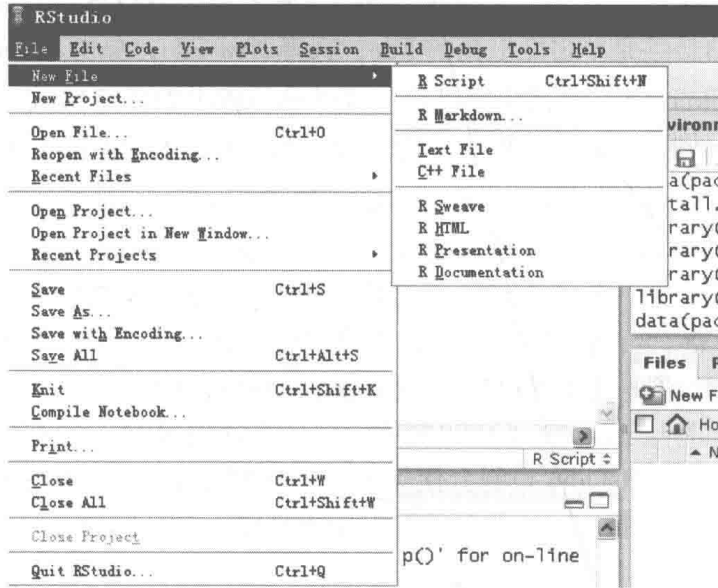


图 1-2-2 RStudio 主界面

2. 四个功能区。Rstudio 界面简单地分为四个窗口，左上角是脚本编辑窗口，右上角是工作空间和历史记录信息，左下角是命令窗口，右下角是绘图、帮助、文件和包管理窗口。

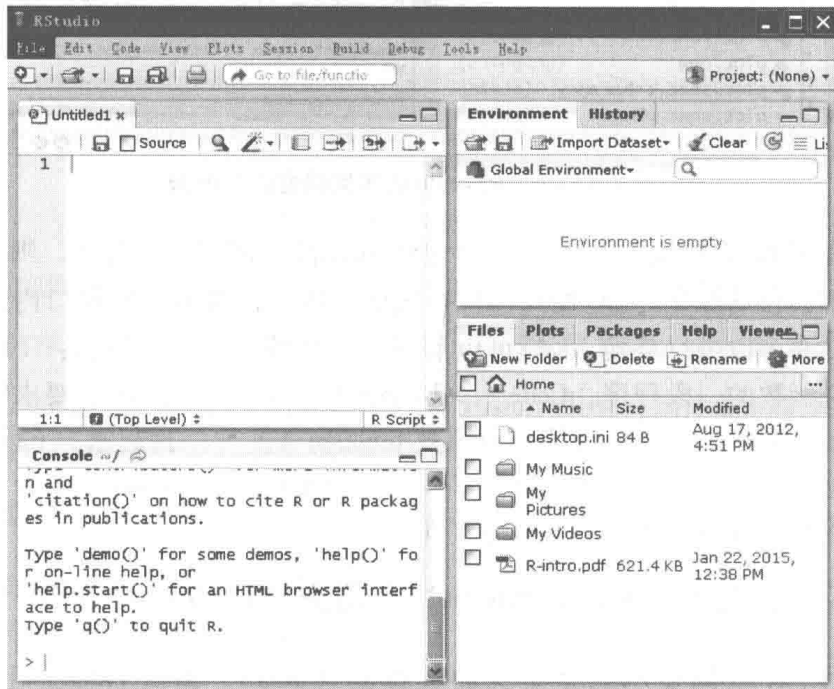




图 1-2-3 RStudio 主界面四个功能区

3. 编写代码和运行。在 Rstudio 中有两种方法可以编写代码，一是直接在左上角的脚本编辑窗口输入代码，根据需要运行所选择的代码；二是在左下角的命令窗口，直接

输入代码，并回车运行。除了前一种方法能进入脚本编辑窗口外，另外还有两种方法能进入脚本编辑窗口：一是单击左上角的 File，选择 New File，单击 R Script；二是在 Rstudio 界面按 Ctrl + Shift + N。写完代码以后，可以单击脚本编辑窗口右上角的  Run 按钮，直接单击该按钮，则只运行鼠标所在代码，如果想运行几行代码，可以先用鼠标选中这些代码行，再单击该按钮。Run 右侧的按钮是  Re - Run，就是重复一次上次的运行。

4. Tab 键：提示功能。无论采用何种方法进行编码，都可以使用 Rstudio 提供的 Tab 键，该键具有提示功能，如：你想输入函数 plot，但忘了该函数的全称，此时你只需要输入 p，再按 Tab 键，会出现所有已安装的程序包中以 p 开头的函数及简要介绍，回车键即可选择。同时，Tab 键还可以显示函数的各项参数，输入 plot ((Rstudio 会自动补上右括号)，按 Tab 键则显示 plot () 的各项参数。当使用 R 进行工作时，可能需要重新执行之前输入的命令，Rstudio 控制台使用方向键回忆之前命令的功能，UP 回忆前一条命令，Down 则是与 UP 相反，这样就可以不用重复输入之前执行过的相同的命令。

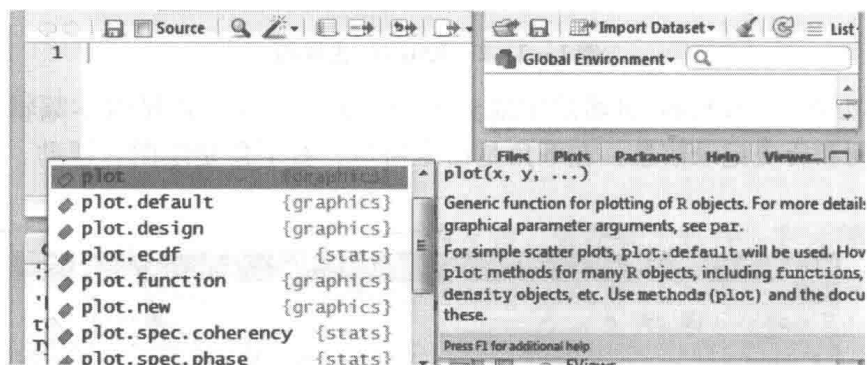



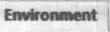


图 1-2-4 RStudio 脚本编辑窗口 Tab 键

5. 脚本编辑窗口上方三按钮。  Source on Save 在该按钮前的方框打钩，则表示自动保存代码，否则不会自动保存。  查找和替换代码功能，在脚本编辑窗口内进行查找你想寻找的代码，并可以进行替换。  可以实现 6 大功能，主要功能为：（1）定义函数。选中需要定义函数的一段程序，Code - > Extract Function，Rstudio 会要求输入自定义函数的名字，输入回车，需要定义的函数则被 function () {} 括起来。（2）定义和取消注释。如需将某段程序注释，选中单击 Code - > Comment/Uncomment Line 即可实现将注释取消。这些功能都可以用右侧的热键实现。

1.2.2 工作空间 (Environment) 和历史 (History) 窗口

 工作空间显示的是定义的数据集 (data)、值 (Value) 和自定义函数 (Function)，Value 和 Function 都是上一次程序运行后，保存在 .RData 文件里面的值，Value 一项中，保存的是程序运行过程中一些变量的值，可以通过鼠标单击，使它们显示出来。而 Function 则是可以方便地找到代码中的函数，同样也可以单击打开查看。import dataset 则是可以导入格式按照 R 语言要求的数据作为数据集的。Clear all 可以将当

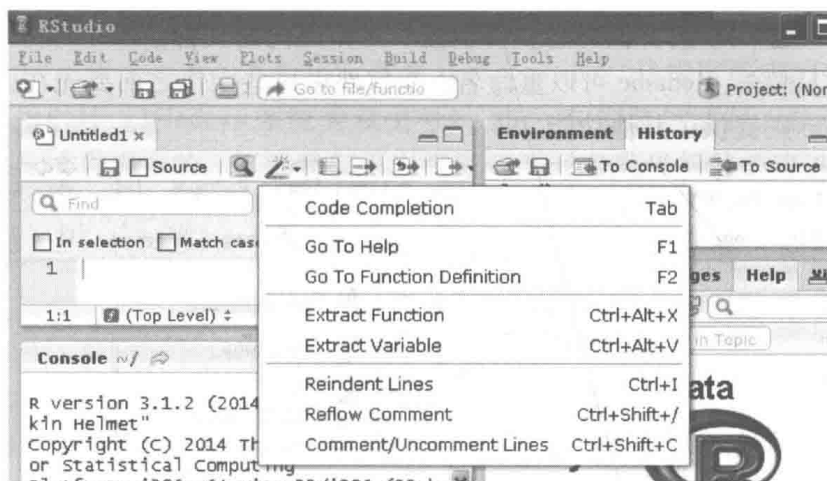


图 1-2-5 RStudio 脚本编辑窗口三个主要按钮

前工作区的 Value 和 Function 清除干净。RStudio 不会自动更新这个工作区的值，如果获得要运行代码的 Value 和 Function，需在运行前调用 Clear 按钮。

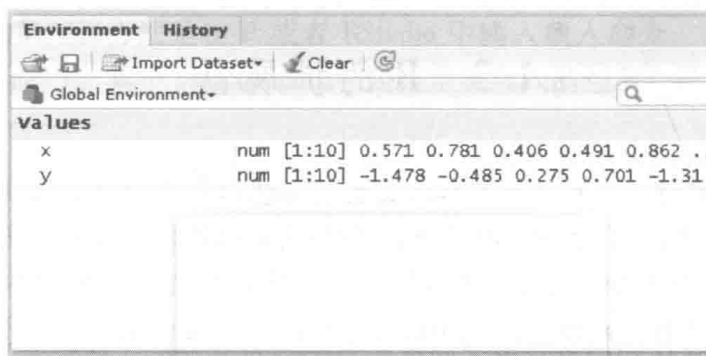


图 1-2-6 RStudio 工作空间和历史窗口

历史窗口显示的是历史操作，显示的代码则是之前运行过的代码，可以保存下来，也可以选择一部分，然后按 To Console 使其进入主控制界面，与重复以前的操作类似。或者按 To Source，将 History 中的代码复制粘贴到脚本区。清除 RStudio 历史信息。

1.2.3 命令窗口

左下角是命令窗口，可以写代码，也能显示程序运行过程和结果，但此处代码无法保存，所以并不建议在此编写程序。

1.2.4 右下角文件窗口

右下角的文件窗口主要输出图形和显示函数的帮助文件。有四个主要的功能区，Files 是查看当前 workspace 下的文件，Plots 则是展示运算结果的图案，Packages 则能展示系统已有的软件包，并且能勾选载入内存。Help 则可以查看帮助文档。