

60个实际案例详解Stata常用统计方法

6个行业应用案例阐述Stata统计分析的方法、思路和分析流程

60道上机练习让读者学练结合，快速掌握Stata统计分析方法

Stata

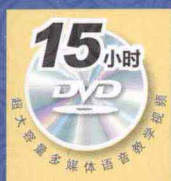
统计分析 with 行业应用 案例详解

(适合Stata 9.0-10.0-11.0-12.0版本)

◆ 张甜 编著 ◆



全面解读Stata在统计分析中的行业应用



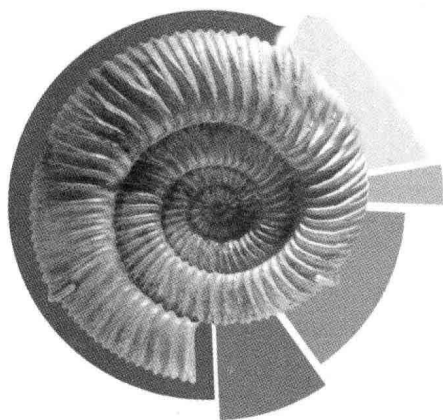
清华大学出版社



Stata

统计分析与行业应用 案例详解

◆ 张甜 编著 ◆



清华大学出版社
北京

内 容 简 介

Stata是公认的应用最广泛的专业数据分析软件之一，因其功能丰富、效率高、操作简便，深受广大用户，尤其是在校师生的青睐。

本书采用先讲解Stata的各个操作功能，再通过综合案例讲述Stata在各个行业中实际应用的思路编写，分为两个部分：第1部分是1~16章，按照统计类型讲述Stata的具体应用；第2部分是17~22章，分行业讲述了Stata的具体应用。各章均附有与正文部分对应的上机操作练习题，目的是着重培养读者的动手能力，使读者在实际练习的过程中能够快速提高应用水平。

本书面向具备一定统计学基础和计算机操作基础的在校各专业的学生，以及企事业单位的相关数据统计分析人员。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。
版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目(CIP)数据

Stata 统计分析与行业应用案例详解 / 张甜编著. --北京：清华大学出版社，2014
ISBN 978-7-302-34510-7

I. ①S… II. ①张… III. ①统计分析—应用软件 IV. ①C819

中国版本图书馆CIP数据核字(2013)第274239号

责任编辑：夏非彼
封面设计：王翔
责任校对：闫秀华
责任印制：李红英

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>，<http://www.wqbook.com>

地 址：北京清华大学学研大厦A座 邮 编：100084

社总机：010-62770175 邮 购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印刷者：清华大学印刷厂

装订者：三河市新茂装订有限公司

经 销：全国新华书店

开 本：190mm×260mm 印 张：36.75 字 数：941千字
(附光盘1张)

版 次：2014年1月第1版 印 次：2014年1月第1次印刷

印 数：1~3500

定 价：89.00元

前 言

Stata 是公认的应用最广泛的专业数据分析软件之一，以功能丰富、效率高、操作简便而著称，主要针对经济、管理、医学、农学、教育、市场研究、社会调查等多个行业和领域。Stata 拥有最具亲和力的窗口，使用者自行建立程序时，软件能提供具有直接命令式的语法，是非常适合进行数据分析的工具软件。本书通过多个实例，详细介绍了 Stata 12.0 在现实生活中的应用。

全书共 22 章，分为如下两个部分。

第 1 部分（第 1~16 章）为 Stata 的各个操作功能在具体实例中的应用。

- 第 1 章介绍了 Stata 12.0 基本窗口以及管理变量与数据，包括 Stata 12.0 窗口说明、数据文件的创建与读取、创建和替代变量、分类变量和定序变量的基本操作、数据的基本操作以及定义数据的子集等。
- 第 2 章介绍了 Stata 制图实例，包括直方图、散点图、曲线标绘图、连线标绘图、箱图、饼图、条形图、点图等。
- 第 3 章介绍了 Stata 描述统计实例，包括定距变量的描述性统计分析、正态性检验和数据转换、单个分类变量的汇总、两个分类变量的列联表分析、多表和多维列联表分析等。
- 第 4 章介绍了 Stata 参数检验实例，包括单一样本 T 检验、两个总体均值差异的假设检验、总体方差的假设检验、总体比率的假设检验等。
- 第 5 章介绍了 Stata 非参数检验实例，包括单组样本的正态分布检验、两独立样本检验、两相关样本检验、多独立样本检验、游程检验等。
- 第 6 章介绍了 Stata 方差分析实例，包括单因素方差分析、多因素方差分析、协方差分析、重复测量方差分析等。
- 第 7 章介绍了 Stata 相关分析实例，包括简单相关分析、偏相关分析等。
- 第 8 章介绍了 Stata 主成分分析与因子分析实例。
- 第 9 章介绍了 Stata 聚类分析实例，包括划分聚类分析和层次聚类分析等。
- 第 10 章介绍了 Stata 最小二乘回归分析实例，包括简单线性回归和多重线性回归等。
- 第 11 章介绍了 Stata 回归诊断分析实例，包括异方差检验、自相关检验、多重共线性检验等。
- 第 12 章介绍了 Stata 非线性回归分析实例，包括非参数回归分析、转换变量回归分析以及非线性回归分析等。
- 第 13 章介绍了 Stata 的 Logistic 回归分析实例，包括二元 Logistic 回归分析、多元 Logistic 回归分析以及有序 Logistic 回归分析等。
- 第 14 章介绍了 Stata 的因变量受限回归分析实例，包括断尾回归分析和截取回归分析。
- 第 15 章介绍了 Stata 时间序列分析实例，包括时间序列分析的基本操作、单位根检验、

协整检验、格兰杰因果关系检验等。

- 第 16 章介绍了 Stata 的面板数据分析实例，包括长面板数据分析和短面板数据分析。

第 2 部分（第 17~22 章）为 Stata 在各个行业中的实际应用。

- 第 17 章介绍了 Stata 在研究城市综合经济实力中的应用。
- 第 18 章介绍了 Stata 在旅游业中的应用。
- 第 19 章介绍了 Stata 在经济增长分析中的应用。
- 第 20 章介绍了 Stata 在原油与黄金价格联动关系研究中的应用。
- 第 21 章介绍了中国上市银行的 ROE 与股权集中度之间关系研究中的应用。
- 第 22 章介绍了 Stata 在农业中的应用。

本书实例典型，内容丰富，有很强的针对性。书中各章不仅详细介绍了实例的具体操作步骤，而且还配有一定数量的练习题以供读者学习使用。读者只需按照书中介绍的步骤一步步地实际操作，就能完全掌握本书的内容。

为了帮助读者更加直观地学习本书，我们将书中实例和练习题所涉及到的全部操作文件都收录到本书的配套光盘中，即“sample”文件夹和“video”文件夹。前者包含书中涉及到的所有 Stata 源文件，后者收录了书中所有实例和练习题的操作录像文件。

本书既可作为数据统计分析的培训教材，也可作为数据统计分析人员的参考书。

本书由张甜编写，此外，参与图书编写和视频制作的还有吕平、王坚宁、高克臻、张云霞、王魁、许小荣、王东、王龙、张银芳、周新国、王松年、张凤琴、陈可汤、陈作聪、聂阳、沈毅、张华杰、彭一明、朱丽云、张秀梅、张玉兰、李爽、田伟和王文婷等，在这里对他们表示感谢。

作者力图使本书的知识性和实用性相得益彰，但由于水平有限，书中错误、纰漏之处难免，欢迎广大读者、同仁批评斧正。

编 者

2013 年 12 月

目 录

第 1 章 Stata 12.0 的基本窗口及管理变量与数据	1
1.1 Stata 12.0 窗口说明	1
1.2 Stata 12.0 数据文件的创建与读取	2
1.2.1 Stata 12.0 数据文件的创建	2
1.2.2 Stata 12.0 数据文件的读取	5
1.3 创建和替代变量	6
1.3.1 创建和替代变量概述	6
1.3.2 相关数据来源	6
1.3.3 Stata 分析过程	6
1.3.4 结果分析	8
1.3.5 案例延伸	10
1.4 分类变量和定序变量的基本操作	11
1.4.1 分类变量和定序变量概述	11
1.4.2 相关数据来源	11
1.4.3 Stata 分析过程	11
1.4.4 结果分析	12
1.4.5 案例延伸	14
1.5 数据的基本操作	14
1.5.1 数据的基本操作概述	14
1.5.2 相关数据来源	14
1.5.3 Stata 分析过程	15
1.5.4 结果分析	16
1.5.5 案例延伸	18
1.6 定义数据的子集	19
1.6.1 定义数据的子集概述	19
1.6.2 相关数据来源	20
1.6.3 Stata 分析过程	20
1.6.4 结果分析	21
1.6.5 案例延伸	22
1.7 本章习题	23

第 2 章 Stata 图形绘制	26
2.1 实例一——直方图	26
2.1.1 直方图的功能与意义	26
2.1.2 相关数据来源	26
2.1.3 Stata分析过程	27
2.1.4 结果分析	28
2.1.5 案例延伸	28
2.2 实例二——散点图	30
2.2.1 散点图的功能与意义	30
2.2.2 相关数据来源	30
2.2.3 Stata分析过程	31
2.2.4 结果分析	32
2.2.5 案例延伸	32
2.3 实例三——曲线标绘图	34
2.3.1 曲线标绘图的功能与意义	34
2.3.2 相关数据来源	34
2.3.3 Stata分析过程	35
2.3.4 结果分析	36
2.3.5 案例延伸	36
2.4 实例四——连线标绘图	38
2.4.1 连线标绘图的功能与意义	38
2.4.2 相关数据来源	38
2.4.3 Stata分析过程	39
2.4.4 结果分析	40
2.4.5 案例延伸	40
2.5 实例五——箱图	41
2.5.1 箱图的功能与意义	41
2.5.2 相关数据来源	42
2.5.3 Stata分析过程	42
2.5.4 结果分析	43
2.5.5 案例延伸	44
2.6 实例六——饼图	44
2.6.1 饼图的功能与意义	44
2.6.2 相关数据来源	45
2.6.3 Stata分析过程	45
2.6.4 结果分析	46
2.6.5 案例延伸	46
2.7 实例七——条形图	47

2.7.1	条形图的功能与意义	47
2.7.2	相关数据来源	48
2.7.3	Stata分析过程	48
2.7.4	结果分析	49
2.7.5	案例延伸	49
2.8	实例八——点图	51
2.8.1	点图的功能与意义	51
2.8.2	相关数据来源	51
2.8.3	Stata分析过程	51
2.8.4	结果分析	52
2.8.5	案例延伸	52
2.9	本章习题	53
第3章	Stata 描述统计	57
3.1	实例一——定距变量的描述性统计	57
3.1.1	定距变量的描述性统计功能与意义	57
3.1.2	相关数据来源	57
3.1.3	Stata分析过程	58
3.1.4	结果分析	59
3.1.5	案例延伸	59
3.2	实例二——正态性检验和数据转换	62
3.2.1	正态性检验和数据转换功能与意义	62
3.2.2	相关数据来源	62
3.2.3	Stata分析过程	63
3.2.4	结果分析	63
3.2.5	案例延伸	64
3.3	实例三——单个分类变量的汇总	66
3.3.1	单个分类变量的汇总功能与意义	66
3.3.2	相关数据来源	66
3.3.3	Stata分析过程	67
3.3.4	结果分析	68
3.3.5	案例延伸	68
3.4	实例四——两个分类变量的列联表分析	68
3.4.1	两个分类变量的列联表分析功能与意义	68
3.4.2	相关数据来源	69
3.4.3	Stata分析过程	69
3.4.4	结果分析	70
3.4.5	案例延伸	71
3.5	实例五——多表和多维列联表分析	71

3.5.1	多表和多维列联表分析功能与意义	71
3.5.2	相关数据来源	72
3.5.3	Stata 分析过程	72
3.5.4	结果分析	73
3.5.5	案例延伸	75
3.6	本章习题	75
第 4 章	Stata 参数检验	78
4.1	实例一——单一样本 T 检验	78
4.1.1	单一样本 T 检验的功能与意义	78
4.1.2	相关数据来源	78
4.1.3	Stata 分析过程	79
4.1.4	结果分析	80
4.1.5	案例延伸	80
4.2	实例二——独立样本 T 检验	81
4.2.1	独立样本 T 检验的功能与意义	81
4.2.2	相关数据来源	81
4.2.3	Stata 分析过程	81
4.2.4	结果分析	82
4.2.5	案例延伸	83
4.3	实例三——配对样本 T 检验	84
4.3.1	配对样本 T 检验的功能与意义	84
4.3.2	相关数据来源	84
4.3.3	Stata 分析过程	85
4.3.4	结果分析	86
4.3.5	案例延伸	86
4.4	实例四——单一样本方差的假设检验	87
4.4.1	单一样本方差假设检验的功能与意义	87
4.4.2	相关数据来源	87
4.4.3	Stata 分析过程	87
4.4.4	结果分析	88
4.4.5	案例延伸	89
4.5	实例五——双样本方差的假设检验	89
4.5.1	双样本方差假设检验的功能与意义	89
4.5.2	相关数据来源	89
4.5.3	Stata 分析过程	90
4.5.4	结果分析	91
4.5.5	案例延伸	91
4.6	本章习题	92

第 5 章 Stata 非参数检验	95
5.1 实例一——单样本正态分布检验	95
5.1.1 单样本正态分布检验的功能与意义	95
5.1.2 相关数据来源	95
5.1.3 Stata 分析过程	96
5.1.4 结果分析	97
5.1.5 案例延伸	97
5.2 实例二——两独立样本检验	98
5.2.1 两独立样本检验的功能与意义	98
5.2.2 相关数据来源	98
5.2.3 Stata 分析过程	98
5.2.4 结果分析	99
5.2.5 案例延伸	100
5.3 实例三——两相关样本检验	100
5.3.1 两相关样本检验的功能与意义	100
5.3.2 相关数据来源	100
5.3.3 Stata 分析过程	101
5.3.4 结果分析	102
5.3.5 案例延伸	102
5.4 实例四——多独立样本检验	103
5.4.1 多独立样本检验的功能与意义	103
5.4.2 相关数据来源	103
5.4.3 Stata 分析过程	104
5.4.4 结果分析	105
5.4.5 案例延伸	105
5.5 实例五——游程检验	106
5.5.1 游程检验的功能与意义	106
5.5.2 相关数据来源	106
5.5.3 Stata 分析过程	107
5.5.4 结果分析	107
5.5.5 案例延伸	108
5.6 本章习题	108
第 6 章 Stata 方差分析	111
6.1 实例一——单因素方差分析	111
6.1.1 单因素方差分析的功能与意义	111
6.1.2 相关数据来源	111
6.1.3 Stata 分析过程	112
6.1.4 结果分析	113

6.1.5 案例延伸	113
6.2 实例二——多因素方差分析	114
6.2.1 多因素方差分析的功能与意义	114
6.2.2 相关数据来源	114
6.2.3 Stata 分析过程	115
6.2.4 结果分析	116
6.2.5 案例延伸	117
6.3 实例三——协方差分析	118
6.3.1 协方差分析的功能与意义	118
6.3.2 相关数据来源	118
6.3.3 Stata 分析过程	119
6.3.4 结果分析	120
6.3.5 案例延伸	122
6.4 实例四——重复测量方差分析	123
6.4.1 重复测量方差分析的功能与意义	123
6.4.2 相关数据来源	123
6.4.3 Stata 分析过程	124
6.4.4 结果分析	125
6.4.5 案例延伸	125
6.5 本章习题	126
第 7 章 Stata 相关分析	129
7.1 实例一——简单相关分析	129
7.1.1 简单相关分析的功能与意义	129
7.1.2 相关数据来源	129
7.1.3 Stata 分析过程	130
7.1.4 结果分析	130
7.1.5 案例延伸	131
7.2 实例二——偏相关分析	133
7.2.1 偏相关分析的功能与意义	133
7.2.2 相关数据来源	133
7.2.3 Stata 分析过程	134
7.2.4 结果分析	134
7.2.5 案例延伸	135
7.3 本章习题	135
第 8 章 Stata 主成分分析与因子分析	137
8.1 实例一——主成分分析	137
8.1.1 主成分分析的功能与意义	137
8.1.2 相关数据来源	137

8.1.3	Stata分析过程	138
8.1.4	结果分析	139
8.1.5	案例延伸	141
8.2	实例二——因子分析	143
8.2.1	因子分析的功能与意义	143
8.2.2	相关数据来源	143
8.2.3	Stata分析过程	144
8.2.4	结果分析	146
8.2.5	案例延伸	166
8.3	本章习题	168
第9章	Stata 聚类分析	169
9.1	实例一——划分聚类分析	169
9.1.1	划分聚类分析的功能与意义	169
9.1.2	相关数据来源	169
9.1.3	Stata分析过程	170
9.1.4	结果分析	171
9.1.5	案例延伸	179
9.2	实例二——层次聚类分析	183
9.2.1	层次聚类分析的功能与意义	183
9.2.2	相关数据来源	183
9.2.3	Stata分析过程	184
9.2.4	结果分析	187
9.2.5	案例延伸	197
9.3	本章习题	206
第10章	Stata 最小二乘线性回归分析	207
10.1	实例一——简单线性回归分析	207
10.1.1	简单线性回归分析的功能与意义	207
10.1.2	相关数据来源	207
10.1.3	Stata分析过程	208
10.1.4	结果分析	209
10.1.5	案例延伸	213
10.2	实例二——多重线性回归分析	214
10.2.1	多重线性回归分析的功能与意义	214
10.2.2	相关数据来源	215
10.2.3	Stata分析过程	215
10.2.4	结果分析	216
10.2.5	案例延伸	221
10.3	本章习题	223

第 11 章 Stata 回归诊断与应对	225
11.1 实例一——异方差检验与应对	225
11.1.1 异方差检验与应对的功能与意义	225
11.1.2 相关数据来源	225
11.1.3 Stata 分析过程	226
11.1.4 结果分析	227
11.1.5 案例延伸	235
11.2 实例二——自相关检验与应对	238
11.2.1 自相关检验与应对的功能与意义	238
11.2.2 相关数据来源	239
11.2.3 Stata 分析过程	239
11.2.4 结果分析	241
11.2.5 案例延伸	247
11.3 实例三——多重共线性检验与应对	248
11.3.1 多重共线性检验与应对的功能与意义	248
11.3.2 相关数据来源	249
11.3.3 Stata 分析过程	249
11.3.4 结果分析	250
11.3.5 案例延伸	254
11.4 本章习题	256
第 12 章 Stata 非线性回归分析	258
12.1 实例一——非参数回归分析	258
12.1.1 非参数回归分析的功能与意义	258
12.1.2 相关数据来源	258
12.1.3 Stata 分析过程	259
12.1.4 结果分析	260
12.1.5 案例延伸	263
12.2 实例二——转换变量回归分析	265
12.2.1 转换变量回归分析的功能与意义	265
12.2.2 相关数据来源	266
12.2.3 Stata 分析过程	266
12.2.4 结果分析	267
12.2.5 案例延伸	273
12.3 实例三——非线性回归分析	273
12.3.1 非线性回归分析的功能与意义	273
12.3.2 相关数据来源	274
12.3.3 Stata 分析过程	274
12.3.4 结果分析	275

12.3.5 案例延伸	280
12.4 本章习题	282
第 13 章 Stata Logistic 回归分析	284
13.1 实例一——二元Logistic回归分析	284
13.1.1 二元logistic回归分析的功能与意义	284
13.1.2 相关数据来源	284
13.1.3 Stata分析过程	285
13.1.4 结果分析	286
13.1.5 案例延伸	291
13.2 实例二——多元Logistic回归分析	293
13.2.1 多元Logistic回归分析的功能与意义	293
13.2.2 相关数据来源	293
13.2.3 Stata分析过程	294
13.2.4 结果分析	295
13.2.5 案例延伸	297
13.3 实例三——有序Logistic回归分析	298
13.3.1 有序Logistic回归分析的功能与意义	298
13.3.2 相关数据来源	298
13.3.3 Stata分析过程	299
13.3.4 结果分析	300
13.3.5 案例延伸	302
13.4 本章习题	304
第 14 章 Stata 因变量受限回归分析	307
14.1 实例一——断尾回归分析	307
14.1.1 断尾回归分析的功能与意义	307
14.1.2 相关数据来源	307
14.1.3 Stata分析过程	308
14.1.4 结果分析	309
14.1.5 案例延伸	312
14.2 实例二——截取回归分析	313
14.2.1 截取回归分析的功能与意义	313
14.2.2 相关数据来源	313
14.2.3 Stata分析过程	314
14.2.4 结果分析	315
14.2.5 案例延伸	317
14.3 本章习题	319

第 15 章 Stata 时间序列分析	321
15.1 时间序列分析的基本操作	321
15.1.1 时间序列分析的基本操作概述	321
15.1.2 相关数据来源	321
15.1.3 Stata 分析过程	322
15.1.4 结果分析	323
15.1.5 案例延伸	327
15.2 单位根检验	328
15.2.1 单位根检验的功能与意义	328
15.2.2 相关数据来源	328
15.2.3 Stata 分析过程	328
15.2.4 结果分析	330
15.2.5 案例延伸	335
15.3 协整检验	336
15.3.1 协整检验的功能与意义	336
15.3.2 相关数据来源	337
15.3.3 Stata 分析过程	337
15.3.4 结果分析	338
15.3.5 案例延伸	341
15.4 格兰杰因果关系检验	345
15.4.1 格兰杰因果关系检验的功能与意义	345
15.4.2 相关数据来源	345
15.4.3 Stata 分析过程	345
15.4.4 结果分析	346
15.4.5 案例延伸	349
15.5 本章习题	350
第 16 章 Stata 面板数据分析	352
16.1 实例一——短面板数据分析	352
16.1.1 短面板数据分析的功能与意义	352
16.1.2 相关数据来源	352
16.1.3 Stata 分析过程	353
16.1.4 结果分析	355
16.1.5 案例延伸	366
16.2 实例二——长面板数据分析	368
16.2.1 长面板数据分析的功能与意义	368
16.2.2 相关数据来源	368
16.2.3 Stata 分析过程	369
16.2.4 结果分析	371

16.2.5 案例延伸	382
16.3 本章习题	383
第 17 章 Stata 在研究城市综合经济实力中的应用	385
17.1 研究背景及目的	385
17.2 研究方法	385
17.3 数据分析与报告	386
17.4 描述性分析	387
17.4.1 Stata分析过程	387
17.4.2 结果分析	387
17.5 相关分析	391
17.6 回归分析	393
17.7 因子分析	398
17.8 因子分析之后续分析	405
17.9 研究结论	406
17.10 本章习题	407
第 18 章 Stata 在旅游业中的应用	409
18.1 研究背景及目的	409
18.2 研究方法	410
18.3 数据分析与报告	410
18.3.1 各城市国内旅游出游人均花费按性别和年龄进行的聚类分析	410
18.3.2 各城市国内旅游出游人均花费按职业进行的聚类分析	417
18.3.3 各城市国内旅游出游人均花费按文化水平进行的聚类分析	424
18.3.4 各城市国内旅游出游人均花费按旅游目的进行的聚类分析	430
18.3.5 各风景区按其自身特点进行的聚类分析	437
18.4 研究结论	444
18.5 本章习题	445
第 19 章 Stata 在经济增长分析中的应用	450
19.1 数据来源与研究思路	450
19.2 描述性分析	451
19.2.1 Stata分析过程	451
19.2.2 结果分析	453
19.3 时间序列趋势图	456
19.3.1 Stata分析过程	456
19.3.2 结果分析	457
19.4 相关性分析	460
19.4.1 Stata分析过程	460
19.4.2 结果分析	461
19.5 单位根检验	463

19.5.1	Stata分析过程	463
19.5.2	结果分析	465
19.6	协整检验	471
19.6.1	Stata分析过程	471
19.6.2	结果分析	472
19.7	格兰杰因果关系检验	474
19.7.1	Stata分析过程	474
19.7.2	结果分析	474
19.8	建立模型	476
19.9	研究结论	478
19.10	本章习题	479
第 20 章	Stata 在原油与黄金价格联动关系研究中的应用	480
20.1	数据来源与研究思路	480
20.2	描述性分析	481
20.2.1	Stata分析过程	482
20.2.2	结果分析	482
20.3	时间序列趋势图	483
20.3.1	Stata分析过程	483
20.3.2	结果分析	484
20.4	相关性分析	487
20.4.1	Stata分析过程	487
20.4.2	结果分析	487
20.5	单位根检验	490
20.5.1	Stata分析过程	490
20.5.2	结果分析	491
20.6	协整检验	496
20.6.1	Stata分析过程	496
20.6.2	结果分析	497
20.7	格兰杰因果关系检验	499
20.7.1	Stata分析过程	499
20.7.2	结果分析	500
20.8	建立模型	501
20.9	研究结论	503
20.10	本章习题	504
第 21 章	Stata 在 ROE 与股权集中度之间关系研究中的应用	505
21.1	研究背景	505
21.2	基本概念与数据说明	506
21.3	实证分析	507