



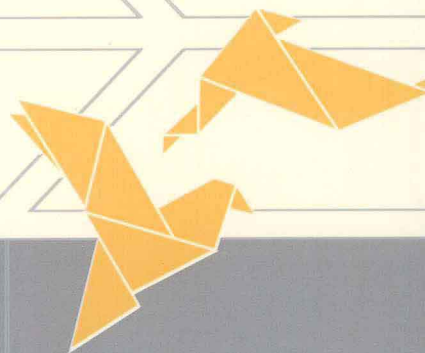
21 世纪高等学校
经济管理类规划教材
高校系列

STATISTICS AND ITS APPLICATIONS

统计学 原理与应用

+ 程建华 洪文 主编
+ 胡本田 杨爱元 龙莹 副主编

理论指导应用，学习有“理”有“据”
追踪前沿思想，紧跟时代脉搏
大数据时代更需要对小数据的把握



ECONOMICS
AND
MANAGEMENT

人民邮电出版社
POSTS & TELECOM PRESS



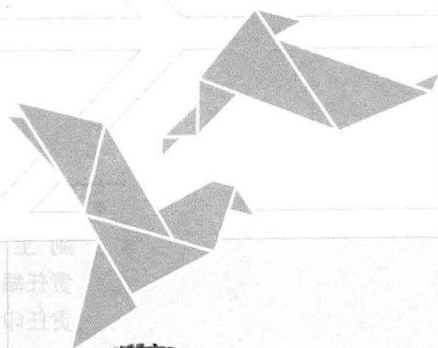
21 世纪高等学校
经济管理类规划教材
高校系列

013955404

STATISTICS AND
ITS APPLICATIONS

统计学 原理与应用

+ 程建华 洪文 主编
+ 胡本田 杨爱元 龙莹 副主编



人民邮电出版社
北京

图书在版编目(CIP)数据

统计学原理与应用 / 程建华, 洪文主编. -- 北京 :
人民邮电出版社, 2013. 9
21世纪高等学校经济管理类规划教材. 高校系列
ISBN 978-7-115-32096-4

I. ①统… II. ①程… ②洪… III. ①统计学—高等
学校—教材 IV. ①C8

中国版本图书馆CIP数据核字(2013)第142960号

内 容 提 要

本书是按照教育部经济学类和工商管理类教学指导委员会制定的经管类核心课程教学大纲要求编写而成的,所有参编人员均结合自己多年从事统计学教学经验,对每一章节教学内容都进行了严格仔细的审定。

本书每章以问题为导向引入该章的教学内容,通过要回答或拟解决问题的讨论,给出理论表述和解决问题的具体方法。在写作过程中自始至终强调“理论从哪里来和到哪里去”的教学理念。在写作方法上突出与统计软件的结合。虽然教材仅以 Excel(非真正意义上的统计软件)作为处理数据和统计计算的工具,目的是希望展示使用软件工具的意义和重要性。在教学内容安排上,既考虑到基础统计学内容的完整性,也考虑到不同专业教学内容的可选择性。为此设计以下主要教学内容:绪论、统计调查与数据整理、统计数据描述、抽样与抽样分布、参数估计、假设检验、方差分析、相关分析与回归分析、时间序列分析以及指数分析等。

本书可作为高等学校全日制经济学类、工商管理类专业本科生和相关职业技术学院的专科生教材,也可作为人文、社会科学以及理工科从事统计分析应用的参考书和阅读文献。

-
- ◆ 主 编 程建华 洪文
副 主 编 胡本田 杨爱元 龙莹
责任编辑 武恩玉
责任印制 彭志环 杨林杰
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京天宇星印刷厂印刷
 - ◆ 开本: 787×1092 1/16
印张: 16.5 2013年9月第1版
字数: 385千字 2013年9月北京第1次印刷

定价: 35.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223
反盗版热线: (010)67171154

前言 Preface



随着计算机技术与应用的发展,统计学最近 20 年得到长足发展,其应用几乎无处不在,统计方法已成为高等院校绝大多数学科领域内研究的基本方法,特别是在 2000 年,教育部经济学教学指导委员会和工商管理类教学指导委员会双双将统计学作为我国高等院校经济学类核心课程和工商管理类核心课程,凸显出统计学这一学科和课程在经济管理相关专业的重要地位。为此,我国各财经类高等院校对于经管类专业的统计学教材在教学内容和教学模式上发生了根本性的改变,尤其是在教学过程中,逐步强调统计方法的实际应用和统计软件对实际数据的处理和统计分析。

目前,国内外统计学教材很多,特色迥异,但即使再好的教科书,也不是对任何一个专业都适用,教材编写的思想与教学人员的教学理念和教学方法密不可分。除此之外,教员还需要针对不同的专业适度调整教学内容和教学方式。在我们自己的教学过程中,尽管所教授的专业是与统计学应用密切相关的经济学类和工商管理类专业,但我们自始至终强调统计学基本原理的来源和出处,把“从哪里来到哪里去”的教学思想渗透到整个教学活动中,即坚持基本原理和应用实践并重的指导思想。高等院校培养的是创新型人才,创新型人才的特点是具有较强独立分析问题和独立解决问题的能力。

早年,经济管理类的学生普遍反映统计学难学,甚至厌恶统计学课程,但今天同类专业的学生(包括研究生)都十分喜欢统计学课程。这两个原因:一是统计学并非他们想象得那样难;二是统计学非常有趣,似乎是在玩数字数据游戏或数字魔方。正因为如此,我们从学生的角度出发,在教材的编写过程中,以“挖掘数据背后有价值的信息”为主线,把统计学教学内容贯穿于教材编写的细节中,既立足于知识点原理的“深”与“透”,同时又着眼于知识的实践应用,知识的应用不是面对实际问题把知识直接“照葫芦画瓢”式的套用,而是侧重于对问题的分析过程,而分析问题的过程则又是对知识原理再次解读的过程,使得知识的应用顺理成章,使读者使用更加得心应手。

即便有好的教学思想,如何把思想表现出来也不是一件容易的事情。为此,在本教材中,我们力图突出如下特点:

(1) 注重知识的基本原理和统计学思想的渗透。

(2) 强调方法的实际运用, 把问题的描述、分析、解决途径以及工具的选择尽可能作为一个完整的过程加以叙述。

(3) 案例和例题的选择紧密结合经济学和工商管理学的知识内容。

(4) 习题的编写少而精, 每章最后给出一道综合习题, 促进学生综合能力的锻炼, 包括逻辑思维、问题的表述、数据处理、公式计算、图形绘制和结果解读等。

本教材按一学期 54 学时教授, 教师可根据统计学与其他课程的衔接以及实际教学效果, 调整教学内容, 所有作业均安排在课后, 教学时数不包括习题课时间, 表 1 拟定的教学时数仅供参考。

表 1 各章教学内容和教学时数计划

章节内容	教学学时数
第 1 章 绪论	2
第 2 章 统计调查与数据整理	4
第 3 章 统计数据描述	4
第 4 章 抽样与抽样分布	4
第 5 章 参数估计	6
第 6 章 假设检验	6
第 7 章 方差分析	6
第 8 章 相关分析与回归分析	8
第 9 章 时间序列分析	6
第 10 章 指数分析	6
教学机动安排	2
合 计	54

本书由安徽大学经济学院和商学院多年从事且至今一直坚守统计学教学第一线的教师集体编写, 程建华和洪文负责全书的内容、版面设计和统稿, 程建华负责第 1 章、第 4 章、第 6 章和第 9 章的编写; 杨爱元负责第 2 章、第 3 章的编写; 胡本田负责第 8 章、第 10 章的编写; 洪文负责第 7 章的编写; 龙莹负责第 5 章的编写。

章节的文字初稿和三稿由编写者自己负责修改, 二稿由安徽大学经济学院统计学专业的研究生赵佳佳、魏琼、任婷、朱圣蓉、方超、许州、丁海婧、徐兵华以及本科生蒋咪、余玲玲、熊贝贝、于戒严、罗昊、金永曦等帮助校对修改, 赵佳佳同学还修改了部分章节的图形, 参与校稿的同学不仅认真修改了文字和公式笔误, 而且作为第一读者提出了许多宝贵意见, 在此向付出辛勤劳动的同学表示衷心感谢。

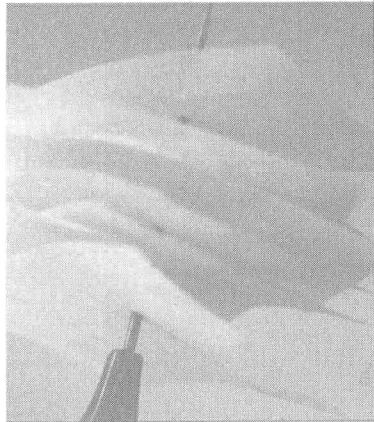
在本书编写过程中, 我们阅读、参考和采纳了所列参考文献中一些著作的观点、思想、方法和部分好的习题与例题, 在此向这些著作的作者表示诚挚的谢意。

鉴于我们的知识水平和教学水平有限, 本书肯定存在错误和内容欠妥之处, 恳请广大读者和同行批评指正。本书文责自负。

程建华

2013 年 3 月

目 录 Contents



第1章 绪论

- 1.1 引言 / 1
- 1.2 统计的涵义 / 2
 - 1.2.1 统计的涵义 / 2
 - 1.2.2 统计方法 / 3
- 1.3 统计的应用 / 4
 - 1.3.1 统计的产生与发展 / 4
 - 1.3.2 统计的应用 / 6
- 1.4 统计学的重要概念 / 8
 - 1.4.1 总体和样本 / 8
 - 1.4.2 标志与指标 / 9
- 1.5 统计软件介绍 / 12
 - 1.5.1 统计软件简介 / 13
 - 1.5.2 Excel简介 / 14
- 本章小结 / 16
- 习题 / 16

第2章 统计调查与数据整理

- 2.1 问题的提出 / 18
- 2.2 数据的计量与类型 / 19
 - 2.2.1 数据的计量 / 19
 - 2.2.2 数据的类型 / 20
- 2.3 统计调查的形式与方法 / 21
 - 2.3.1 数据调查的组织形式 / 21
 - 2.3.2 搜集数据的具体技术 / 24
 - 2.3.3 次级数据的搜集 / 26
- 2.4 调查方案设计 / 27
 - 2.4.1 调查方案的主要内容 / 27
 - 2.4.2 统计调查问卷设计 / 28
- 2.5 统计数据的预处理 / 30
 - 2.5.1 数据的审核与筛选 / 31
 - 2.5.2 数据的排序 / 31

2.6 统计分组与变量数列 / 32

 2.6.1 统计分组的概念 / 32

 2.6.2 频数分布 / 33

 2.6.3 变量数列的编制 / 33

2.7 统计表 / 36

 2.7.1 统计表的作用 / 36

 2.7.2 统计表的构成 / 37

 2.7.3 统计表的种类 / 37

 2.7.4 统计表的设计要求 / 42

2.8 统计图 / 42

 2.8.1 条形图 / 42

 2.8.2 直方图 / 43

 2.8.3 圆形图 / 44

 2.8.4 折线图 / 44

 2.8.5 未分组数据——茎叶图 / 45

 2.8.6 时间序列数据——线图 / 46

2.9 Excel应用 / 46

 2.9.1 利用Excel图表向导制作条形图 / 46

 2.9.2 利用Excel工具制作直方图 / 47

本章小结 / 49

习题 / 50

第3章 统计数据描述

3.1 问题的提出 / 53

3.2 总量指标 / 53

 3.2.1 总量指标的概念 / 53

 3.2.2 总量指标的种类 / 53

 3.2.3 总量指标的计量单位 / 54

3.3 相对指标 / 55

 3.3.1 相对指标的概念和作用 / 55

 3.3.2 相对指标的种类及其计算方法 / 56

 3.3.3 正确运用相对指标的原则 / 60

3.4 平均指标 / 61

 3.4.1 算术平均数 / 61

 3.4.2 调和平均数 / 64

 3.4.3 几何平均数 / 65

 3.4.4 众数 / 66

 3.4.5 中位数 / 67

 3.4.6 众数、中位数和算术平均数的关系 / 69

3.5 标志变异指标 / 70

 3.5.1 变异指标的概念 / 70

 3.5.2 全距 / 70

 3.5.3 平均差 / 71

 3.5.4 标准差与方差 / 72

 3.5.5 变异系数 / 74

3.6 分布的偏态和峰度 / 75

 3.6.1 偏态 / 75

 3.6.2 峰度 / 75

3.7 Excel的描述统计功能 / 76

本章小结 / 78

习题 / 78

第4章 抽样与抽样分布

4.1 问题的提出 / 80

4.2 抽样与抽样方法 / 81

 4.2.1 抽样与统计量 / 81

 4.2.2 抽样方法 / 82

 4.2.3 抽样一般程序 / 84

4.3 抽样分布 / 84

 4.3.1 抽样分布的概念 / 84

 4.3.2 四大重要分布 / 86

4.4 中心极限定理及其应用 / 90

4.5 单样本统计量的抽样分布 / 92

 4.5.1 样本均值的抽样分布 / 92

 4.5.2 样本比例的抽样分布 / 93

 4.5.3 样本方差的抽样分布 / 94

4.6 双样本统计量的抽样分布 / 96

 4.6.1 两样本均值之差的抽样分布 / 97

 4.6.2 两样本比例之差的抽样分布 / 97

 4.6.3 两样本方差之比的抽样分布 / 98

本章小结 / 99

习题 / 99

第5章 参数估计

5.1 问题的提出 / 101

5.2 参数估计概述 / 102

- 5.2.1 参数估计的概念 / 102
- 5.2.2 参数估计的分类 / 102
- 5.2.3 评价估计量的标准 / 103

5.3 一个总体参数的区间估计 / 105

- 5.3.1 总体均值的区间估计 / 105
- 5.3.2 总体比例的区间估计 / 110
- 5.3.3 总体方差的区间估计 / 111

5.4 两个总体参数的区间估计 / 113

- 5.4.1 两个总体均值之差的区间估计 / 113
- 5.4.2 两个总体比例之差的区间估计 / 115
- 5.4.3 两个总体方差比的区间估计 / 116

5.5 样本容量的确定 / 117

- 5.5.1 估计总体均值时样本容量的确定 / 118
- 5.5.2 估计总体比例时样本容量的确定 / 119

本章小结 / 120

习题 / 120

第6章 假设检验**6.1 问题的提出 / 123****6.2 假设检验的基本概念 / 123**

- 6.2.1 统计学上的假设 / 123
- 6.2.2 弃真与存伪错误 / 126
- 6.2.3 显著性水平的设定 / 127
- 6.2.4 检验统计量的构建 / 127
- 6.2.5 接受域与拒绝域 / 128
- 6.2.6 P 值的含义与应用 / 129
- 6.2.7 假设检验的步骤 / 129

6.3 单参数检验 / 130

- 6.3.1 单个总体均值检验 / 130
- 6.3.2 总体比例检验 / 137
- 6.3.3 总体方差检验 / 139

6.4 双参数检验 / 141

- 6.4.1 两个总体均值是否相等的检验 / 141
- 6.4.2 两个总体比例是否相等的检验 / 145
- 6.4.3 两个总体方差是否相等的检验 / 146

本章小结 / 148

习题 / 148

第7章 方差分析**7.1 问题的提出 / 151****7.2 方差分析的基本概念 / 152**

- 7.2.1 方差分析的基本术语 / 152
- 7.2.2 方差分析的基本思想 / 152
- 7.2.3 方差分析中的基本假定 / 153

7.3 单因素方差分析 / 154

- 7.3.1 单因素方差分析的步骤 / 154
- 7.3.2 单因素方差分析的应用 / 157

7.4 双因素方差分析 / 158

- 7.4.1 双因素方差分析概述 / 158
- 7.4.2 无交互作用的双因素方差分析 / 158
- 7.4.3 有交互作用的双因素方差分析 / 163

本章小结 / 165

习题 / 166

第8章 相关分析与回归分析**8.1 问题的提出 / 169****8.2 相关与回归分析的基本概念 / 169**

- 8.2.1 变量之间的关系 / 169
- 8.2.2 相关关系的种类 / 171
- 8.2.3 相关分析与回归分析 / 172

8.3 相关关系的描述与度量 / 173

- 8.3.1 相关表 / 173
- 8.3.2 相关图 / 173
- 8.3.3 相关系数 / 176

8.4 回归分析 / 179

- 8.4.1 一元线性回归分析 / 180
- 8.4.2 一元线性回归模型的估计 / 181
- 8.4.3 一元线性回归模型的检验 / 185
- 8.4.4 一元线性回归模型的预测 / 189

8.4.5 多元线性回归分析与非线性

回归简介 / 192

本章小结 / 195

习题 / 195

第9章 时间序列分析

9.1 问题的提出 / 199

9.2 时间序列分析概述 / 200

9.2.1 时间序列的含义 / 200

9.2.2 时间序列的种类 / 201

9.2.3 编制时间序列的原则 / 203

9.2.4 时间序列的分析方法 / 203

9.3 时间序列的水平指标 / 203

9.3.1 发展水平与平均发展水平 / 203

9.3.2 增长量与平均增长量 / 206

9.4 时间序列的速度指标 / 207

9.4.1 发展速度 / 207

9.4.2 增长速度 / 207

9.4.3 平均发展速度和平均增长速度 / 208

9.5 时间序列的构成要素 / 208

9.5.1 时间序列的构成要素 / 208

9.5.2 时间序列的构成要素模型描述 / 210

9.6 时间序列的要素分解 / 210

9.6.1 长期趋势项的测定 / 211

9.6.2 季节变动项的测定 / 217

9.6.3 循环变动项的测定 / 219

9.6.4 不规则变动项的测定 / 223

本章小结 / 223

习题 / 224

第10章 指数分析

10.1 问题的提出 / 227

10.2 指数的基本概念 / 227

10.2.1 指数概念 / 227

10.2.2 统计指数的作用 / 228

10.2.3 统计指数的分类 / 228

10.3 总指数的编制方法 / 229

10.3.1 简单指数法 / 230

10.3.2 加权指数法 / 231

10.4 几种常用的统计指数 / 240

10.4.1 居民消费价格指数 / 241

10.4.2 证券价格指数 / 242

10.4.3 进出口贸易指数 / 242

10.4.4 企业景气指数 / 243

10.5 指数体系和因素分析 / 243

10.5.1 指数体系 / 243

10.5.2 因素分析 / 244

本章小结 / 250

习题 / 251

参考文献

信息时代的人们几乎每天都要和新闻、信息和数据打交道，这些大量的信息和数据究竟从哪来的？它们是怎样被搜集、整理和计算出来的？它们有没有价值？准确不准确？对一些虚假信息和数据，人们如何去辨别和揭穿？这些问题如今在我们的现实生活中屡见不鲜，例如买卖一支股票，一个理性的投资者不仅需要了解这支股票当天的价格波动，而且还要对这支股票价格波动的历史（新股票除外）、发行这支股票公司的业绩、近期的行业经济状况以及宏观经济基本面等众多方面进行分析。还有，如今的商业公司越来越重视对其产品质量和服务质量的跟踪调查，通常在其售出产品单附页上留有简短的客户问卷调查表，让客户对其产品和服务填写反馈意见，以便更好地搜集改进产品和服务质量的数据资料，提高产品质量和企业的市场竞争力。无论是社会中的个体还是一个组织，对其获得的信息和数据资料做出较为深入的分析，都会为其带来更大的收益或者减少损失，这些“额外”所做的分析工作便是统计工作，其意义和价值不言而喻。

统计工作有时也会令人怀疑，甚至专业权威部门的统计工作或统计报告也不例外。例如，2010年2月25日我国国家统计局发布了《中华人民共和国2009年国民经济和社会发展统计公报》（以下简称《公报》），《公报》发布的当日，公众及新闻媒体便对《公报》中“70个大中城市房屋销售价格上涨1.5%”^①这一统计指标数据提出质疑，因为这与百姓普遍感受到的2009年城市房价涨幅相差甚远。究竟是公众的感受有误，还是公众对国家统计局指标数据的含义解读有误，或者是官方数据统计有误？对公众的质疑，政府部门有责任有义务对这一指标的数据采集样本和口径、数据处理和指标计算方法做出满意的回答，否则将有失政府的公信力。公众的质疑是社会进步的可喜现象，一是公众对国家官方统计数据更为关注，统计素养与责任意识更强；二是政府对外信息公开的透明度增强，政府接受公众质疑有利于未来统计工作水平的进一步提高。

具有法律意识和法制观念的公众对社会公共信息有权利评判，但不应妄下结论，对自己个人所给的判断或结论应该有理有据，若能附于详实的统计资料和证据，那自然更有说服力，如果再具有一定的统计学知识和统计常识，那么个人的行为准则将更加理性和规范。从此角度而言，学习统计学是十分重要又有益的。

^① 中华人民共和国国家统计局，《中华人民共和国2009年国民经济和社会发展统计公报》，2010年2月25日，http://www.stats.gov.cn/tjgb/ndtjgb/qgndtjgb/t20100225_402622945.htm

无论是在普通百姓的日常生活中，还是在国家经济社会的重大活动中，统计学知识的应用、统计活动的开展几乎无处不在。统计思想极其朴素、简单，绝大多数没有系统学过统计学的人，也都经常不知不觉地应用统计学概念、思想、方法和工具，帮助自己思考和解决现实问题，特别是一些较为理性的人，对统计学知识的应用更是恰到好处。比如，一家新的化妆品公司对其产品进行的宣传推广往往会夸大其词，夸大其产品的功效是如何神奇，但理性的人通常会问其神奇功效的证据来源于何处，这一质问实际就是寻求样本、检验总体的过程；再比如，“十赌九输”这几乎是一基本常识，但所有的赌徒都乐此不疲，其原因都是想自己成为极少数的赢家，都想让“赢”的小概率事件在自己身上发生，从而达到一夜暴富的目的。按此说法，那么统计学是一门只需明白一些统计分析思想而无需系统学习就可灵活应用的简单学科。实际上，这只是统计学汪洋大海中的一滴水，统计学思想既朴素又博大精深，它不但包含了丰富的分析方法和灵活多样的应用工具，而且其涉及的内容和应用也非常广阔。正因为如此，许多没有系统学过统计学的人易于对统计学产生片面理解，他们认为统计学极其简单，随便知道点道理就可以了，统计工作就是收集点数据而后对外发布的体力劳动。但从上述引言给出的思考问题我们已知道，统计学不只是采集数据和发布数据（或经济社会信息）这样一个简单工作，多数情形下是通过数据对现实社会中的各类问题进行归纳、总结、推断，并在此基础上尽可能地做出最可靠的决策。因此，我们对统计学需要有一个完整的认识。

统计学是一门采集、整理、展现和分析数据并在此基础上进行科学决策的学科，其目标是探索和把握数据内在的数量规律性。因此，在此意义上，统计就是挖掘数据背后包含的有价值的信息。虽然统计学这一定义是一种描述性的解释，但其思想精髓渗透其中。由于统计学与数据隔离不断，有人则称统计学为“数据的科学”^①，我们则把数据视为统计的“血液”，数据在统计分析和应用过程中不息地流动，自始至终贯穿着统计过程，维系着统计生命。

统计学根据现象的“量”研究现象的“质”，这一解释包含了极其深刻的思想。

1.2.1 统计的涵义

如果把“统计”当作一专业词语来应用，在不同的场景，其涵义或语义是不同的，表达的意思差别明显，“统计”可以表示为统计工作、统计资料和统计学3个不同的涵义。

1. 统计工作

为了认识问题与解决问题，对社会经济现象和自然现象进行数据采集和整理的活动称为“统计工作”（statistical work）。例如，国家人口普查、经济普查，企业市场调查、生产车间采集成品、半成品抽样检验数据都可称为统计工作。

^① 袁卫、庞皓、曾五一、贾俊平. 统计学（第2版）. 北京：高等教育出版社，2005，第2页.

2. 统计资料

统计资料 (statistical data) 是统计工作过程中所取得的各项数字资料、文本资料以及与之相关信息的总称。例如国家统计局各年度、季度、月度的经济社会统计数据, 国家气象局每时每刻的天气、气候数据资料, 企事业单位的生产、经营、人事管理资料等。

统计资料存储模式有电子版的, 也有传统纸质版的, 有些则是两种介质并存。例如国家统计局的年度统计数据既有电子版的光盘数据, 也有纸质版的《中国统计年鉴》。用户根据自己所需, 使用不同存储方式的统计资料或统计数据。统计资料或统计数据的对外发布同样也存在纸质方式和电子版方式。

电子版的统计资料又可分为公共网络版和单一电子介质版, 例如《中国统计年鉴》数据, 不仅在国家统计局 (www.stats.gov.cn) 官方网站上可以查询到, 而且还存有电子版的光盘。

统计资料既有免费使用的, 也有有偿付费使用的。例如, 国家官方对外发布的公共经济社会统计数据和资料都是免费的, 而一些私营公司独立采集到的经济社会调查数据并经过一些整理分析得到的研究报告通常都是有偿使用的。

特别有趣的是, 有些原始统计资料杂乱无章, 看似毫无价值, 但一旦经过整理、分析和深加工, 往往可以获取到“无价之宝”。例如, 股市行情起伏跌宕, 看似无序, 但对一个有经验的投资者来说, 却可以在此无序中发现有价值的信息, 为自己增添财富。在当今信息化时代, 也有人因为详实的“统计资料”跌入深渊。例如, 南京市江宁区房产局局长周久耕因为一次偶然的“炫富”, 被网络人肉搜索 (广义上为“统计资料”) 而锒铛入狱, 这可以说是最典型的“统计资料”抓贪官的案例。当然送贪官进监狱的并非是“统计资料”, 而是个体本身的贪婪所致。所以, 过去所说的“数据就是金钱”现在也不一定完全正确了, “数据也可能是祸害 (即犯罪的证据)”。

3. 统计学

如前所述, 统计学 (statistics) 是一门采集、整理、展现和分析数据, 并在此基础上进行科学决策的学科。统计工作的历史十分悠久, 但统计学作为一门独立的学科分支, 更多的是对各类统计方法的研究、总结和应用, 例如抽样分布、参数估计、假设检验、方差分析、回归分析、实验设计、聚类分析、关联性分析以及时间序列分解等, 都是一些极为常用的统计方法, 统计学在发展中不断完善。统计学的理论基础是概率理论, 而现代统计学已把数学中的测度论、群论、数论等数学方法应用到其中, 使得统计学方法更为丰富和可靠。

1.2.2 统计方法

从统计分析的思想方法角度, 我们可以将统计学分为两大类, 即描述性统计和推断性统计。

描述性统计 (Descriptive Statistics) 研究的是数据采集、整理、分类、汇总、图表表示数据、概括和分析的统计方法。例如, 对某一班级《统计学》期终考试分数的分析, 按优秀、良好、及格和不及格人数的简单汇总和直方图表示、班级平均分的计算、各题型答题的正确率分析等一系列卷面简单分析, 都属于描述性统计分析。

推断性统计 (Inferential Statistics) 研究的是如何利用样本数据来推断总体特征的统计方法。例

如,在产品质量检验中,对产品合格率或优良率的估计在多数情形下不是对每个产品进行全面检测,而是从产品的批次中随机抽取一定的样本进行检测,根据样本检测结果对该批次产品质量做出总体评价。商业企业的客户满意度分析也不是基于对所有客户满意度的问卷调查,而是按一定抽样方案对部分客户作问卷调查,并利用该调查数据结果对客户总体满意程度做出估计。

对于统计方法,我们需要有正确的认识,统计方法获得的结论与逻辑推理获得的结论的途径存在较大差别,任何有效的统计方法只有当这个系统或是所讨论的总体满足方法本身的基本假设时,才能获得可靠的结论,例如古典的线性回归分析方法中的随机扰动项,需要满足同方差、非自相关和服从正态分布等若干假定条件,否则得到的极有可能是伪回归结论。所以,误用统计方法可能会导致描述分析或推断分析严重错误,这个错误将严重影响决策。初学者千万不能把对方法的误用看作是方法本身的问题,这是两个不同的概念。

即使统计方法被正确地应用,结果对于非统计专业人士来说也可能存在解读上的困难。例如,统计数据中显著性的改变可能是由样本的随机变量所致,但这个显著性可能与事先分析的直觉并不完全相符,对结果的解释不仅需要良好的统计学知识,有时还需要较强的专业背景知识。这一典型实例便是收入与学历之间在许多现实情况下并不呈现正相关关系,因为许多高收入者学历并不高,但从整个社会而言,学历高的人群收入则明显高于低学历者。

1.3

统计的应用

1.3.1 统计的产生与发展

统计的产生来自于应用的需求,其发展也是由应用的完善不断获得的,因此了解和掌握统计的应用,可以从统计的产生与发展过程中得到启示,其历史过程可以更好地帮助我们认识统计学的概念和统计的应用。

1. 统计学学派简介

统计学是一门十分古老的学科,它起源于社会经济问题的研究,一般认为其学理研究开始于古希腊的亚里士多德时代,但是将其作为一门学科系统地研究和发展,则是在17世纪中叶的欧洲逐步展开的。对统计学基础理论的探讨,在其历史发展的长河中,形成了国势学派、政治算术学派、数理统计学派和社会统计学派。任何一门学科学派的形成与其代表人物的生活工作背景、知识结构、思想方法密不可分,在学习过程中,可发现各学派对统计学认识的共性和差异性。

(1) 国势学派。该学派的代表人物是德国的康令(H.Conring, 1606—1681)和阿亨瓦尔(G.Achenwall, 1719—1772)。H.Conring第一个在德国黑尔姆斯太特大学以“国势学”为题,讲授政治活动家应具备的知识。G.Achenwall在格丁根大学最早开设了“国家学”课程,其主要著作《近代欧洲各国国势学纲要》讲述的是“一国或多数国家的显著事项”,他用对比分析的方法研究了国家组织、人口、军队、领土、居民职业以及资源财产等一系列问题。该学派在进行国势比较分析中偏

重事物性质的解释，而不注重数量对比和数量计算，但该学派对统计学的最大贡献是它提出了一个世界公认的名词——“统计学（statistics）”。

（2）政治算术学派。该学派的代表人物是威廉·配第（William Petty, 1623—1687）和约翰·格朗特（J.Graunt, 1620—1674）。配第在其著作《政治算术》中，对当时的英国、荷兰、法国之间的财富进行数量上的计算和比较，开创了应用数量方法研究社会经济现象的先河。格朗特的代表性著作《对死亡率公报的自然观察和政治观察》中，对伦敦市 50 多年的人口出生和死亡资料进行了详细的计算和分析，他不仅是统计学的奠基者，更是人口统计学的先驱。政治算术学派在统计发展的历史上有重要的地位，由于它以数量分析为特征，从这一角度评价，其分析思想和方法是统计学的真正起源。

（3）数理统计学派。最初把古典概率论引进统计学的是法国的若干位数学家，如拉普拉斯（P.S.Laplace, 1749—1827）、帕斯卡（B.Pascal, 1623—1662）和费马（Pierre de Fermat, 1601—1665）。帕斯卡和费马将赌博中出现的各种问题归纳为一般的概率原理；拉普拉斯阐明了统计学的大数法则，进行了大样本推断的尝试。与此同时，比利时统计学家、数学家凯特勒（Adolphe Quetelet, 1796—1874）在其著作《社会物理学》中利用大数法则论证了社会生活中的随机偶然现象贯穿着必然的规律性，并运用概率论原理提出了著名的“平均人”概念，计算人类自身各性质标志的平均值，通过“平均人”来探索社会规律。

数理统计学派对统计方法做出了杰出贡献，他们创立的各种统计方法有力地推动了统计学的快速发展。英国的戈赛特（William Sealy Gosset, 1876—1937）利用 t 分布（即著名的 student 分布）建立了“小样本理论”；费希尔（Ronald Aylmer Fisher, 1890—1962）给出了 F 统计量、极大似然估计、方差分析等；内曼（J.Neyman, 1894—1981）和皮尔逊（Egon S.Pearson, 1895—1980）提出了置信区间估计和假设检验；沃尔德（A.Wald, 1902—1950）发明了序贯抽样和统计决策函数；卡尔·皮尔逊（Karl Pearson, 1857—1936）设计的“线性相关系数”是表示两个随机变量相关程度最常用的表达式，并且他还构建了完美的“ χ^2 检验”。

（4）社会统计学派。该学派的主要代表人物有克尼斯（K.G.A.Knies, 1821—1898）、恩格尔（C.L.E.Engel, 1821—1896）和梅尔（G.V.Mayer, 1841—1925）。他们的主要研究对象是社会现象，研究的目的是发掘社会现象的内在联系和相互关系，研究方法是全面调查和抽样调查。著名的“恩格尔系数”是衡量城乡居民生活水准最有效、最客观的统计评价指标。

上述各统计学派对统计学的贡献显示出统计学思想的共性，即从大量的经济社会现象中发现重要的概念和有价值的结论，无论是其利用对比分析的方法、调查方法还是用数量特征和概率大样本理论方法，其研究目标都是一致的。

2. 现代统计学

20 世纪 50 年代以来，统计学的理论、方法和应用进入了一个崭新的阶段，统计理论和方法随着数学学科的发展日益雄厚和丰富，鞅论、各态遍历（ergodic）和非线性时间序列分析相继出现。在研究内容上，贝叶斯统计、非参数统计、探索性数据分析、数据挖掘、面板数据模型、高维统计分析以及高频数据分析成为一个又一个热点。特别是 20 世纪 70 年代以来，电子计算机技术的发展

大大推动了统计学应用的广度和深度，计算机不仅为统计数据的采集、存储、分类、整理和查询带来了极大便利，而且还提供了强大的统计分析工具，同时还将遗传算法、人工神经网络方法等计算方法带进了统计学，产生了统计计算和统计模拟等新的统计学科分支，使得统计学的研究方法和研究内容更为丰富。

借助于现代的计算机技术、信息技术以及空间技术，今天的统计学几乎渗透到每一个学科分支，从社会科学中的经济学、社会学、法学、文献学、历史档案学到自然科学中的生物学、农学、医学、环境科学、工程技术、军事科学以及管理决策等，现实社会呈现的新问题和长期的数据积累本身就为统计学提出众多新的研究课题，在研究和解决实际问题的过程中，则又进一步推动了统计学的发展和升华。

现代统计学最为显著的特征是：直接面对实际问题、统计方法拥有强大的分析工具（一般为统计软件 and 数据处理软件）和详实丰富的数据资料。

1.3.2 统计的应用

因为统计学是关于“数据的科学”，所以统计学的使命便是挖掘数据背后隐藏的有价值的信息。正因为如此，有数据的地方就是统计学的用武之地。关于统计学的应用不能一一举例，统计学与众多学科交叉所形成的学科分支难以有十分清晰的界限，在下面的举例中，我们仅以应用主题为背景作一简单介绍。

1. 医学统计

统计学在医学上的应用是最为普遍的，因为医院每天分布着大量的患有不同疾病的患者，医院每天生成大量的统计资料和统计数据。流行病分析通常都是根据大量患者的体态特征、传播渠道来进行统计分析的；疾病发病原因分析有些是借助于仪器设备进行检测而发掘，有些则是根据患者的体态特征做出初步判断，然后再借助于生化指标进一步确认。

在医学统计中，药物疗效的评价和跟踪分析涉及多元统计分析、寿命生存分析等多种统计分析方法。

【案例分析】“三鹿”奶粉事件的发现

例 1.1 “三鹿”奶粉事件的发现

2008年夏天，江苏省徐州市儿童医院小儿泌尿科医生冯东川在连续为几名肾衰竭患儿动完手术后，心情很沉重，从第三个6个月大的患儿肾脏中取出的结石直径1.5厘米，它会引起便血、脏器坏死，直至婴儿死亡。在婴儿体内发现如此大的结石并且不止一例，这绝不是个偶然的現象。冯东川判断，最有可能的病因，就是孩子的食物。

在连续收治了9例肾结石患儿之后，冯东川跟国内的同行交流，发现许多地方都有类似病例，患儿吃的居然都是同一品牌的奶粉——“三鹿”奶粉。其实，凡是冯东川收治的患儿，他都悄悄告诉家长不要再给孩子吃“三鹿”奶粉了，但是其他医院其他地方的更多的孩子，他们的家长又怎么能知道自己的孩子每天喝下去的奶粉可能会要了孩子的命呢？

冯东川决定以实名向国家质检总局报告其病例发现：“我是一名小儿泌尿外科医生，最近一个月接连诊治了9位双肾结石导致肾衰的患儿，其中六例都吃的是‘三鹿’配方奶粉，希望质检总局能组织专家查明原因。”这只是冯东川医生的初步诊断，由此便揭开了“三鹿”事件的序幕。

详细阅读请查阅：

http://news.cntv.cn/special/fazhirenwutuixuan01/20101204/108260_8.shtml

【案例评述】首先可以肯定的是冯东川是一名具有强烈责任感的医生，而且是一个非常理性的职业人，对其发现的医学现象，他并没有轻易的做出结论，而是通过正常的渠道对自己的发现作进一步权威性的确认；其次，他是一个十分细心的人，他没有依靠仅有6~9个样本点的医学案例放弃其医学观察，他也没有依靠仅仅6~9个样本点的数据就对自己的发现简单地给出统计上的结论，而是依据专业知识做更深入的生化指标检测和医学观察。从此案例我们看到，在实际专业领域，借助统计学的思想、知识、分析方法和工具，可以更快、更有针对性地发现问题，这正是统计学的价值所在。

2. 经济统计

统计在经济学领域的应用极其广泛，微观层面的企业对所售产品的定价决策、市场行情与判断、市场竞争力都需要进行统计分析；中观层面的行业景气分析依赖于统计计算；宏观层面的短期经济形势分析和预测判断通常也都是建立在对经济统计指标数据分析的基础之上的，而国民经济核算是经济统计的重要分支，国家经济统计数据的定期发布都来源于国民经济核算。除此之外，国家经济结构变化、国内外贸易动态、金融证券投资等领域也都离不开统计分析。今天的统计学在经济研究中具有不可替代的作用，统计学对计量经济学的贡献卓著。

3. 产品质量管理

众所周知，摩托罗拉公司推行的六西格玛(6 σ —six sigma)管理是产品质量管理最杰出的典范，其核心思想是利用统计方法将产品质量误差控制在技术条件下的最小范围内，并在生产过程中实施监测和控制，使其产品每一环节的不合格率都降低到最小程度。实际上，统计方法不只是应用在产品生产过程中，在产品的最后成品检验阶段和产品的初始设计阶段，统计方法都发挥着巨大作用，如产品质量检验的抽样方案确定、生产方与使用方对产品合格率判断的风险控制以及产品可靠性和故障率的估计等。

4. 环境统计学

随着经济社会和科学技术发展，人们对环境越来越关注。但对环境的评价则是多角度的，环境优劣包含的因素具有多样性，如天气气候、降水量、土壤成分、水土保持、河流、山林植被、日照时间、地质岩土结构、地质灾害以及地区生物种群变化等，不但如此，对环境数据的采集和跟踪比一般统计数据的采集要复杂困难得多，有些数据无法人工采集，需要利用遥感卫星，即使采集到多项统计指标数据，其数据时间频率也往往不一致。客观需求的迫切性和统计数据的复杂性，使得环境统计学成为当今统计学最热门的研究领域之一。

尽管统计学的应用十分广泛，统计应用可以帮助人们发现经济社会中的问题，或者帮助寻找解

解决问题的方案，但是统计只能完成其学科本身所能胜任的使命，其功能不能覆盖全部。例如，应用统计方法和医学统计数据可以发现，吸烟人群患肺癌的发病率远远高于不吸烟人群，但统计学却不能回答不吸烟人群为什么也会患肺癌。社会上存在这样一个误区，统计学没有多大实用价值，尤其是在一些政治问卷调查中，对一个领导者评价所获取的数据往往存在不真实和不可靠的信息，由此获得的价值评判结论有失公允，甚至获得的结论与公众的实际评价完全相反。其提出的问题是客观存在的，但抽样误差不可避免，由抽样误差导致的错误结论更是不可避免，同时把错误数据带来错误结论归于统计方法其本身就是一个逻辑假定的前提错误，这如同在一个错误的指南针下不能走到正确的目标一样，此与统计学本身并无实质性的关系。

1.4

统计学的重要概念

1.4.1 总体和样本

1. 总体

总体 (Population) 是指根据一定的研究目的统计所要描述与分析的、客观存在的、具有某一共同性质的许多个别单位所构成的整体。构成总体的每一个体单位被称为总体单位，简称单位或个体，它是构成总体的最基本单位。例如，如果要了解和分析某大学新生的来源分布情况，那么该大学所在该年份入学同学的全体就是总体，每一位同学即为个体或单位，其中每一个体具有相同属性，即该大学同一年份入学。

统计总体根据总体单位是否可以计量，分为有限总体和无限总体。

有限总体是指一个统计总体中包含的单位数是有限的。例如上述实例中，某年份大学新生的人数是有限的，且可以计量。对有限总体可以进行全面调查，也可以进行非全面调查。比如，对这些新生家庭经济状况的了解与掌握既可进行全面调查，也可进行非全面的随机抽样调查。

无限总体是指一个统计总体中包含的单位数是无限的。例如，为了建设好城市地下市政工程，需要对某一城市夏季降雨量进行分析，但一个地区夏季的降雨量为一连续随机分布变量，所有降雨量形成的则是一无限总体。对无限总体不能进行全面调查分析，只能抽取一部分样本单位进行分析，并依据样本分析结果推断总体分布特征。

统计总体具有以下 3 个特性：

(1) 同质性，是指构成总体的各个单位必须具有某一个共同的特征和性质。同质性是各个单位构成统计总体的先决条件。例如，了解一个大学新生的基本情况，该总体中所有个体具有一些基本的共同特征或属性，一是个体的身份应属于该大学，二是所有的个体都是同一年入学的。

(2) 大量性，是指总体是由许多单位组成的，仅个别或少数单位不能构成总体。这是因为统计研究的目的是为了描述和把握现象的规律，由于少数个体所呈现的现象具有很大的偶然性，而大量单位的现象综合特征相对稳定，才具有可靠性。因此，现象的规律性只能在大量个体的综合中表现