



张建同 孙昌言 王世进 ©编著

应用 统计学



YINGYONG TONGJIXUE

本书提供配套课件



清华大学出版社

应用统计学

张建同 孙昌言 王世进 编著

清华大学出版社

北 京

内 容 简 介

本书依托目前最为普及的 Excel 软件和社会科学领域最为普及的 SPSS 统计软件为基本工具, 详细介绍了数据收集和整理、统计数据的汇总、概率论基础、抽样分布和参数估计、假设检验、方差分析、一元回归和多元回归、时间序列预测和指数、卡方检验和非参数检验, 以及统计在质量管理中的应用等重要内容。与计算机软件相结合是本书最大的特点, 另一大特点是将统计理论与统计实践紧密联系, 每一章都提供了具体案例和大量例题以辅助教学。

本书体系完善, 布局合理, 实例丰富, 应用性强, 每章均有习题, 并配备了内容丰富的教学课件(下载网址为 <http://www.tupwk.com.cn>)和习题答案, 能够适应统计学的教学需要, 可作为高等院校工科类和经济管理类相关专业的本科、硕士、MBA 和工程硕士等的教学用书, 也可作为从事统计、信息管理、市场调研等实际工作的各类经济管理类人员的参考用书或培训教材。

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

应用统计学/张建同, 孙昌言, 王世进 编著. —北京: 清华大学出版社, 2010.4

ISBN 978-7-302-22087-9

I. 应… II. ①张… ②孙… ③王… III. 应用统计学 IV. C8

中国版本图书馆 CIP 数据核字(2010)第 015360 号

责任编辑: 崔 伟

封面设计: 周周设计局

版式设计: 康 博

责任校对: 胡雁翎

责任印制: 何 芊

出版发行: 清华大学出版社

地 址: 北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

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: 北京市人民文学印刷厂

装 订 者: 三河市李旗庄少明装订厂

经 销: 全国新华书店

开 本: 185×260 印 张: 23.25 字 数: 641 千字

版 次: 2010 年 4 月第 1 版 印 次: 2010 年 4 月第 1 次印刷

印 数: 1~5000

定 价: 32.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题, 请与清华大学出版社出版部联系调换。联系电话: (010)62770177 转 3103 产品编号: 034978-01

前 言

在经济全球化和信息化的环境下,管理决策者面临着高度不确定的外部环境和巨大的风险。应用统计学是定量分析非确定性问题的规律,帮助管理决策者进行科学决策、规避风险、获取最优经济和社会效益的科学方法,已成为现代科学管理中必不可少的强劲工具。

本书在确保知识的系统性和正确性的基础上,尽量使用通俗易懂的语言,阐述应用统计学的基本概念、基本原理及解决问题的基本方法和应用条件,略去了繁琐的推导证明过程,使读者更容易理解和掌握。鉴于计算机是求解统计问题的有效手段,本书中对主要的概率统计方法,都以常用的计算机软件 Excel 和 SPSS 作为求解运算和分析的主要工具,以大幅度提高本课程的实用价值和学生的计算机应用能力。与计算机软件相结合是本书最大的特点。本书另一大特点是将统计理论与统计实践紧密联系,每一章都提供了具体案例和大量例题以辅助教学。

本书内容全面,几乎涉及应用统计学的所有内容。本书是《以 Excel 和 SPSS 为工具的管理统计》(张建同、孙昌言编著,清华大学出版社,2005 年)的修订再版,在基于原书前 6 章内容的基础上,增加了时间序列预测和指数、卡方检验和非参数检验、统计在质量管理中的应用三项重要的内容。对照原书,再版书具体的变化如下:

第 1 章保留了原书的 1.1 节、1.2 节和 1.3 节并增加了统计基本术语和变量的类型,把原书的 1.4 节作为第 2 章。

第 2 章按照分类数据和数值数据分别介绍统计表和统计图,增加了帕累托图和交叉表等新的内容。

第 3 章的内容是在原书第 2 章的基础上增加了四分位数、四分位数极差、Z 值、五数汇总、箱线图、协方差和相关系数等内容。

第 4 章是原书第 3 章并增加了均匀分布的内容。

第 5 章是原书第 4 章 4.1 节、4.2 和 4.3 节的内容。

第 6 章在原书 4.4 节的基础上增加了样本容量确定的内容。

第 7 章是关于单个总体假设检验的问题,参照了原书第 5 章的前四节内容,并增加了考试及格线确定的案例。

第 8 章是关于两个总体假设检验的问题,参照原书 5.5~5.9 节的内容。

第 9 章是原书第 6 章的内容。

第 10 章卡方检验和非参数检验是新增加的内容。

第 11 章是关于一元回归的内容,在原书第 8 章前三节的基础上增加了残差分析的内容。

第12章是关于多元线性回归的内容,在原书8.4和8.5节的基础上增加了多元回归模型的偏 F 检验、虚拟变量、交互作用项和二次回归模型的内容。

第13章是原书第9章的内容。

第14章时间序列预测和指数是新增加的内容。

第15章统计在质量管理中的应用是新增加的内容。

本书体系完善,布局合理,实例丰富,应用性强,可作为高等院校工科类和经济管理类相关专业的本科、硕士、MBA和工程硕士等的教学用书,也可作为从事统计、信息管理、市场调研等实际工作的各类经济管理人士的参考用书或培训教材。其中在章节前没加注*号的,适用于本科和MBA的教学及本科层次的读者,在章节前加*号的,适用于工程硕士研究生的教学和硕士研究生层次的读者。本书每章最后均有习题,并配备了内容丰富的教学课件(下载网址为<http://www.tupwk.com.cn>,亦可访问同济大学经济与管理学院精品课程网站获取更多教学参考资料)和习题答案(教师可发送邮件至 cuiwei80@163.com 索取,我们将及时提供详细答案),以备教学所需。

本书第1~3、5、6、9~12、15章由王世进和张建同编写;第4、7、8章由张建军编写;第13、14章由简泽编写。刘文驰进行了部分校对。全书由张建同总纂并定稿。由于水平有限,书中难免有错误或不足之处,恳请广大读者批评指正,以便本书能不断修正和完善。

张建同

目 录

第 1 章 统计和统计数据收集	1
1.1 统计学概述	1
1.1.1 统计的广泛应用	1
1.1.2 统计与统计学	2
1.1.3 统计研究对象的特点	3
1.1.4 统计学的分类	3
1.2 统计基本术语	4
1.3 数据的收集	5
1.3.1 普遍调查	5
1.3.2 重点调查	6
1.3.3 典型调查	6
1.3.4 抽样调查	6
1.3.5 网上调查	7
1.4 问卷设计	9
1.4.1 合理性	9
1.4.2 一般性	9
1.4.3 逻辑性	10
1.4.4 明确性	10
1.4.5 非诱导性	10
1.4.6 便于整理、分析	11
1.5 变量	11
1.5.1 变量类型	11
1.5.2 度量水平和度量等级	12
习题一	14
第 2 章 统计表和统计图	16
2.1 分类数据的图表	16
2.1.1 汇总表	16
2.1.2 条形图	17

2.1.3	饼图	18
2.1.4	帕累托图	19
2.2	数值数据的整理	22
2.3	数值数据的图表	23
2.3.1	频数分布表	23
2.3.2	其他数值数据统计图	26
2.4	交叉表	35
2.4.1	列联表	35
2.4.2	并行条形图	37
2.5	图表汇总和制作原则	37
	习题二	38
第 3 章	统计数据的描述度量	41
3.1	度量中心趋势的指标	41
3.1.1	算术平均数	41
3.1.2	中位数	42
3.1.3	众数	43
3.1.4	算术平均数、中位数和众数间的关系	44
3.1.5	四分位数	45
3.1.6	五数汇总和箱线图	47
3.1.7	几何平均数	49
3.2	度量离散程度的指标	50
3.2.1	极差	51
3.2.2	四分位数极差	51
3.2.3	平均差	51
3.2.4	方差和标准差	51
3.2.5	变异系数	53
3.2.6	Z 值	53
3.3	度量偏斜程度的指标	53
3.3.1	用标准差为单位计量的偏度系数	54
3.3.2	使用三阶中心矩计量的偏度系数	54
3.4	度量两种数值变量关系的指标	54
3.5	利用 Excel 数据分析功能求各种统计指标	55
	习题三	58
第 4 章	概率论基础	60
4.1	引言	60
	【案例 4.1】 新型洗衣机产品投资生产决策	60

4.2 随机试验与随机事件	62
4.2.1 随机现象	62
4.2.2 随机试验	62
4.2.3 随机事件	62
4.2.4 事件间的关系和运算	63
4.3 概率	65
4.3.1 频率与概率	65
【案例 4.2】 活塞销遭退货的概率有多大	66
4.3.2 条件概率	67
【案例 4.3】 60 岁的人至少能活到 80 岁的概率	68
【案例 4.4】 患癌症的概率	69
4.3.3 事件的独立性	71
4.4 随机变量及其分布函数	72
4.4.1 随机变量	72
4.4.2 随机变量的分布函数	72
4.5 离散型随机变量	73
4.5.1 离散型随机变量的概率分布	73
4.5.2 几种重要的离散型分布	74
【案例 4.5】 应配置多大功率的变压器	74
4.6 连续型随机变量	77
【案例 4.6】 电子产品的寿命分析	78
4.6.1 概率密度	78
4.6.2 几种重要的连续型分布	79
4.7 随机变量的数学期望和方差	84
4.7.1 数学期望	84
4.7.2 方差	86
4.8 大数定律和中心极限定理及其计算机模拟验证	87
4.8.1 大数定律	87
4.8.2 中心极限定理	88
4.8.3 中心极限定理的动态模拟验证	88
4.9 新产品投资决策案例分析	90
4.9.1 投产后各种销售状况下的项目净现值	90
4.9.2 不考虑试生产时的最优决策分析	91
4.9.3 考虑试生产并获取用户试用反馈信息的方案分析	91
4.9.4 追加信息的价值	93
习题四	93

第 5 章 抽样与抽样分布	96
5.1 简单随机抽样和统计量	96
5.1.1 随机样本	96
5.1.2 统计量及抽样分布	99
5.2 其他抽样方法	104
5.2.1 分层随机抽样	104
5.2.2 整群抽样	104
5.2.3 系统抽样	104
5.2.4 方便抽样	105
5.2.5 判断抽样	105
5.3 参数估计	105
5.3.1 参数的点估计	106
5.3.2 点估计的方法	106
5.3.3 估计量的评价标准	107
习题五	108
第 6 章 置信区间估计	110
6.1 基本概念准备	110
6.2 单个正态总体均值和方差的区间估计	111
6.2.1 总体均值 μ 的区间估计	111
6.2.2 总体方差 σ^2 的区间估计	114
6.3 总体比例的区间估计	114
6.4 样本容量确定	116
6.4.1 单个正态总体均值估计的样本容量确定	116
6.4.2 总体比例估计的样本容量确定	117
6.5 两个正态总体的均值差和方差比的区间估计	118
6.5.1 两个正态总体均值差 $\mu_1 - \mu_2$ 的区间估计	118
6.5.2 两正态总体方差比 σ_1^2 / σ_2^2 的区间估计	120
6.6 单侧置信限的估计	120
6.7 区间估计小结	122
习题六	124
第 7 章 单个总体的假设检验	126
7.1 案例介绍	126
【案例 7.1】 新工艺是否有效?	126
【案例 7.2】 机床的加工是否满足要求?	127
7.2 假设检验的基本原理	127
7.2.1 假设检验的基本原理和步骤	127

7.2.2 检验中可能犯的两类错误	129
7.3 单个正态总体均值的检验	129
7.4 单个正态总体方差的检验(χ^2 检验)	131
7.5 单个总体比例的检验	133
7.6 单个总体的假设检验小结	134
习题七	135
第 8 章 两个总体的假设检验	136
8.1 引言	136
【案例 8.1】 哪种安眠药的疗效好?	136
8.2 两个独立正态总体均值的检验	137
8.3 成对样本试验的均值检验	140
8.4 两个正态总体方差的检验(F 检验)	142
8.5 两个总体比例的检验	143
8.6 两个总体的假设检验小结	144
习题八	145
第 9 章 方差分析	147
9.1 引言	147
9.1.1 案例介绍	147
【案例 9.1】 哪种促销方式效果最好?	147
【案例 9.2】 如何确定最优生产工艺条件?	148
9.1.2 方差分析的基本概念	149
9.1.3 方差分析的基本假设条件	149
9.1.4 方差分析的目的	150
9.2 单因子方差分析	150
9.2.1 单因子试验的数学模型	150
9.2.2 方差分析的基本方法	151
9.2.3 检验 H_0 的统计量	151
9.2.4 方差分析表	152
9.2.5 进一步的分析	153
9.3 双因子方差分析	154
9.3.1 不考虑交互作用的双因子方差分析	154
9.3.2 考虑交互作用时的双因子方差分析	157
习题九	161
第 10 章 卡方检验和非参数检验	163
10.1 总体分布的 χ^2 检验	163
10.1.1 检验的基本原理	163

10.1.2	检验的具体步骤	164
10.2	比例差异的 χ^2 检验(独立样本)	166
10.2.1	两个比例差异的 χ^2 检验	166
10.2.2	两个以上比例差异的 χ^2 检验	169
10.2.3	独立性的 χ^2 检验	173
10.3	两个相关样本比例差异检验	175
10.4	两个独立总体的非参数分析: Wilcoxon 秩和检验	177
10.5	单因素方差分析的非参数分析: Kruskal-Wallis 秩检验	181
10.6	小结	183
	习题十	184
第 11 章	一元回归	188
11.1	引言	188
	【案例 11.1】 质量控制应用	188
11.1.1	变量间的两类关系	189
11.1.2	线性回归的数学模型	190
11.1.3	线性回归模型的经典假设条件	191
11.1.4	回归分析的内容和分析步骤	192
11.2	一元线性回归	192
11.2.1	一元线性回归的数学模型	192
11.2.2	参数 β_0 和 β_1 的最小二乘估计	192
11.2.3	最小二乘估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的性质	194
11.2.4	回归方程的显著性检验	194
11.2.5	预测和控制	198
11.3	质量控制应用案例分析	201
11.4	残差分析	204
11.5	曲线回归	207
11.5.1	曲线回归的分析步骤	207
11.5.2	常用曲线的线性化方法	207
	习题十一	212
第 12 章	多元线性回归	214
12.1	多元线性回归的数学模型	214
12.2	参数 β 的最小二乘估计	215
12.3	多元回归模型的显著性检验	218
12.3.1	回归方程的显著性检验	218
12.3.2	回归系数的显著性检验和置信区间估计	219
12.4	预测与控制	222

12.5*	多元回归模型的偏 F 检验	224
12.6*	在回归模型中运用虚拟变量和交互作用项	228
12.6.1	虚拟变量	228
12.6.2	交互作用	230
12.7*	二次回归模型	231
	习题十二	235
第 13 章*	违背经典假设的经济计量模型	240
13.1	引言	240
13.2	异方差	241
13.2.1	异方差的概念	241
13.2.2	异方差产生的原因	242
13.2.3	异方差的后果	244
13.2.4	异方差的识别和检验	244
13.2.5	消除异方差的方法	247
	【案例 13.1】 居民储蓄模型——异方差的检验和处理	250
13.3	自相关	256
13.3.1	自相关的概念	256
13.3.2	产生自相关的原因	257
13.3.3	自相关的后果	258
13.3.4	自相关的识别和检验	259
13.3.5	自相关的处理方法	261
	【案例 13.2】 地区商品出口模型——自相关的检验和处理	264
13.4	多重共线性	267
13.4.1	多重共线性的概念	267
13.4.2	多重共线性的后果	268
13.4.3	产生多重共线性的原因	268
13.4.4	多重共线性的识别和检验	269
13.4.5	消除多重共线性的方法	270
	【案例 13.3】 农业产出模型——多重共线性问题	273
	习题十三	275
第 14 章	时间序列预测和指数	279
14.1	时间序列模型的组成因素	279
14.2	年度时间序列数据的平滑	281
14.2.1	移动平均法	282
14.2.2	指数平滑法	284
14.3	基于最小二乘法的趋势拟合和预测	286

14.3.1	线型趋势模型	286
14.3.2	二次趋势模型	289
14.3.3	指数趋势模型	290
14.3.4	运用第一、第二和百分率差值选择模型	292
14.4	自回归模型用于拟合和预测趋势	294
14.5	时间序列预测季节数据	300
14.6	指数	304
14.6.1	价格指数	304
14.6.2	综合价格指数	306
	习题十四	309
第 15 章	统计在质量管理中的应用	314
15.1	全面质量管理	314
15.2	六西格玛管理	316
15.3	控制图理论	317
15.4	比例的控制图: p -图	319
15.5	极差和均值控制图	324
15.5.1	R -图	324
15.5.2	$\bar{X}$ -图	327
15.6	过程能力	328
15.6.1	顾客满意和规格极限	329
15.6.2	能力指数	330
	习题十五	332
附录 A	泊松分布表	336
附录 B	标准正态分布表	338
附录 C	χ^2 分布表	340
附录 D	t 分布表	344
附录 E	F 分布表	346
附录 F	Wilcoxon 秩和检验的上下临界值	357
附录 G	杜宾-瓦森检验临界值表	358

第 1 章

统计和统计数据收集

本章将介绍统计学的基本知识，统计学的基本术语，统计数据的收集和调查，以及变量的类型。

1.1 统计学概述

1.1.1 统计的广泛应用

人类活动的各个方面都离不开统计工作和统计数据。

在个人生活中，人们的任何决策都依赖于有关的统计信息。无论是报考学校、选择工作单位、购买房屋、进行股票交易、外出旅游，还是购买日常生活用品、生活开支预算等，都离不开有关的统计信息。尽管个人生活中的许多统计信息是以非常简单粗糙的形式出现的，但正是在对过去发生的事情或经验所获得的信息进行综合的基础上，人们才能作出正确的判断和决策。

在政府的层次上，统计更是渗透于每一个部门的管理工作之中。政府部门通过不断收集经济、社会、人口等各个领域的统计数据，在综合分析的基础上对国民经济的发展进行预测、规划、指导和调控，并对全社会提供各种统计信息。

在科学和学术研究的各个领域，包括自然科学、经济学、社会学、体育、医疗卫生、环境保护等领域，都广泛使用统计学方法进行分析和推断。

在商务活动中，统计学有更广泛的应用，经营管理人员在作决策时需要应用统计方法来归纳分析各种可以获得的统计资料(参看如图 1-1 所示的由数据向信息和决策的转换过程)。商务活动中经常使用统计方法的有以下几个方面。

1. 财务分析

企业的财务报告和各種财务分析指标都是对企业的财务状况和经营成果的统计分析结果；管理会计各种方法的应用需要以成本和收益为基础的经营业绩的统计分析；企业的投资和融资决策更离不开各种内外部统计信息的支持。

2. 产品开发

企业产品开发计划的制定,需要对经济发展趋势、商业竞争、顾客需求、财务收支估算等方面的数据进行统计分析,并在此基础上进行产品开发的经济可行性分析。

3. 计划

企业各种计划的制定,都需要对销售、资金、人力资源需求、成本和利润等因素进行预测,而预测就是以过去和当前的统计数据为依据,运用统计推断方法对未来情况进行估计。

4. 市场研究

在市场研究中,需要对消费者的需求偏好及其变化趋势,竞争对手的情况,本企业产品和服务的顾客满意度等方面的数据进行统计分析。

5. 工序管理和质量控制

运用统计分析方法,可以帮助确定影响产品质量的主要因素,制定有效的质量控制标准和工艺规范,达到降低成本、提高生产效率的目标。质量控制和质量管理是企业经营管理中运用统计分析方法最多的领域。

6. 人力资源管理

在企业的人力资源管理中,需要经常使用统计方法来分析人事变动、出勤状况、工作业绩等情况,并在此基础上进行业绩评定、奖励和惩罚、制定有效的激励约束机制等。

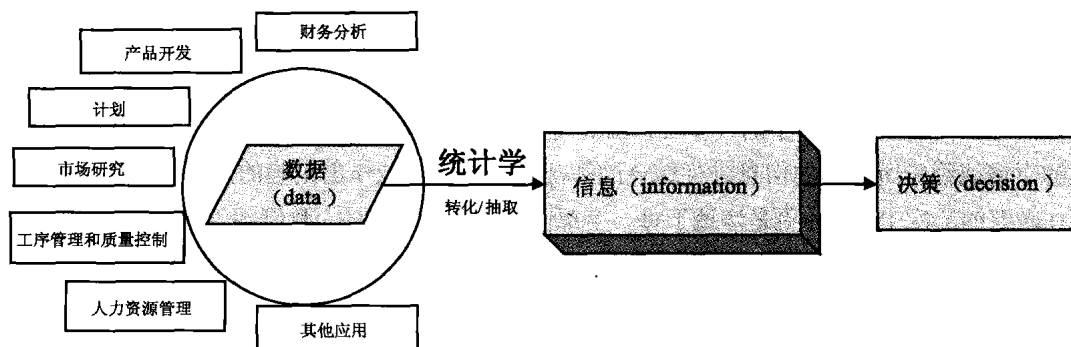


图 1-1 基于统计学的数据转换应用

1.1.2 统计与统计学

关于“统计”一词,可以有三种理解:统计工作、统计资料、统计学。

- 统计工作是指应用科学的方法对调查研究的对象进行数据搜集、整理和分析的全部工作,其成果是形成各种统计数据资料。
- 统计资料是统计工作所取得的成果,包括两大类型:原始资料和再生信息。
- 统计学是关于如何搜集、整理和分析客观现象数量规律的一门方法论的科学,是对统计工作实践加以总结升华而产生的理论,并用以指导统计实践。

1.1.3 统计研究对象的特点

统计研究的对象有以下4个方面的特点。

1. 数量性

统计学研究的对象是客观现象的数量特征和规律性。统计学是通过大量的观察方法来研究事物总体数量方面的特征及其规律性,反映客观事物在一定时间、地点等条件下的数量表现及其发展趋势和变化规律,为人类活动的各种决策提供依据。凡是产生数据的领域,都需要统计工作和统计学;同样,凡是能以数量来描述的事物,都可以作为统计学的研究对象。

2. 总体性

统计学研究的是客观现象总体的数量特征与规律性,而不是个体的量。总体由具有某种相同性质的一定范围内的全体事物组成。例如:对全国农村家庭收入情况进行统计分析,此时全国所有农村家庭的收入就构成研究的总体,但统计学对总体数量规律性的认识是通过对大量个体的观察和分析后获得的。

3. 具体性

统计学研究的对象是一定时间、地点等条件下具体事物的量,而不是抽象对象的量,这是统计学和数学的一个重要区别。因此,统计工作需要具体事物进行调查研究,收集和整理特定时间、地点等条件下所研究对象中大量个体的有关数据是统计分析的基础。

4. 差异(变异)性

组成统计研究对象总体的个体应是有差异的,否则就不需要进行统计分析。客观事物是错综复杂的,受到多种因素的影响,因而不同个体在数量方面必定存在差异,这就需要通过总体中大量的个体进行观察并进行综合分析,才能获得总体的数量特征,如平均值、方差、偏斜度等方面的分布特征。

1.1.4 统计学的分类

根据研究的重点和观察问题的角度不同,统计学主要有以下两种分类方法。

1. 描述统计学和推断统计学

这一分类方法既反映了统计学发展的两个主要阶段,同时也反映了各自不同的侧重。

描述统计学是指研究如何对客观现象进行数量的计量、加工、概括和表示的方法。在20世纪之前,统计学基本上处于描述阶段,描述统计学是统计学的基础。

推断统计学是指研究如何根据样本数据来推断总体的分布情况,概率论是推断统计学的主要理论基础。推断统计学是近代统计学的核心,也是统计学中的主要内容。

2. 理论统计学和应用统计学

理论统计学主要研究统计学的数学原理,它基于概率论的原理,还包括不属于传统概率论的一些内容,如随机化原理的理论、各种估计的原理、假设检验的原理以及一般决策的原理。在统计实践中经常会遇到一些原有的统计方法不能解决的新问题,需要创造新的统计模型和统计分析方法,这就需要统计理论的研究与指导。

将统计学的基本原理应用于各个领域就形成各种应用统计学的分支。它包括适用于各个领域的一般性的统计分析方法,如参数估计、假设检验、方差分析、回归分析等,还包括在某一领域中特定的统计分析方法,如经济领域中的指数分析法等。应用统计学需要既熟悉统计知识又熟悉某一领域业务知识的专门人才,它侧重于阐明统计的基本原理,并将理论统计学的结论作为工具应用于各个领域。

1.2 统计基本术语

变量是事物的特征,是运用统计方法所分析的对象。例如,在商业模型中,销售额、每年的开支和每年的净利润都是企业想要分析的变量。

数据是与变量相关的值。变量可能随时间变动,如某一公司的期望销售额、开支和净利润每年都有所不同。这些不同的值就是与变量相关的数据,或者简单地说,就是统计所要分析的“数据”。

除了时间,变量的数据也可能因其他原因而不同。例如,如果你要分析参加一个大型讲座的人员组成,包括的变量可能有年级、性别和专业。因为班级里每位同学都各不相同,所以这些变量的值也会有所不同。一名学生也许是经济学专业的大一男生,而另一名则有可能是金融学专业的二年级女生。

但需要记住的是,除非赋予实际操作定义,否则所有变量的值(或数据)都是没有任何意义的。而且,进行分析时我们必须非常清楚这些实际操作的定义并一致接受,否则就会产生歧义。例如,对销售额的操作性定义可能会发生这样错误的理解:一个人认为年销售额是指全部连锁店的年销售额,而另一个人则认为是每家店的年销售额。即使是变量中的单个值有时也需要赋予操作性定义。例如年级变量,到底什么是二年级和三年级?

对变量和数据之间区别的理解有助于其他基本术语的学习。

总体是指所研究对象的全体,或者具体指研究对象的某项数量指标的值的全体。

个体是指总体中的每个元素或单元。总体依其所包含的个体总数分为有限总体和无限总体。

样本是从总体中挑选出来用于分析的一部分。这种挑选通常是相对独立的。样本是进行推断统计的依据。

参数是描述总体特征的数值。

统计量是描述样本特征的数值。进行推断统计时,往往不是直接利用样本本身,而是通过基于样本构造样本的适当函数(例如,样本平均值、样本方差等)获得统计量进行分析。