

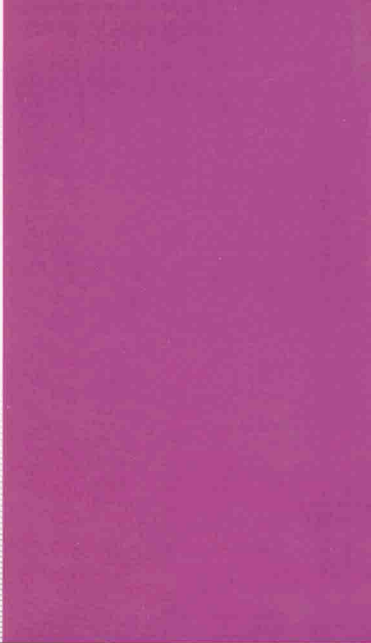
21世纪经济管理精品教材·经济学系列

应用统计学

李卫东 主编



清华大学出版社



应用统计学

清华大学出版社数字出版网站



www.wqbook.com

上架建议：统计

ISBN 978-7-302-37515-9



9 787302 375159 >

定价：36.00元

教学支持说明

尊敬的老师:

您好! 为方便教学, 我们为采用本书作为教材的老师提供教学辅助资源。鉴于部分资源仅提供给授课教师使用, 请您填写如下信息, 发电子邮件或传真给我们, 我们将会及时提供给您教学资源或使用说明。

(本表电子版下载地址: http://www.tup.com.cn/sub_press/3/)

课程信息

书 名			
作 者		书号 (ISBN)	
课程名称		学生人数	
学生类型	<input type="checkbox"/> 本科 <input type="checkbox"/> 研究生 <input type="checkbox"/> MBA/EMBA <input type="checkbox"/> 在职培训		
本书作为	<input type="checkbox"/> 主要教材 <input type="checkbox"/> 参考教材		

您的信息

学 校			
学 院		系/专业	
姓 名		职称/职务	
电 话		电子邮件	
通信地址		邮 编	
对本教材建议			
有何出版计划			

_____年___月___日



清华大学出版社

E-mail: tupfuwu@163.com

电话: 8610-62770175-4903/4506

地址: 北京市海淀区双清路学研大厦 B 座 506 室

网址: <http://www.tup.com.cn/>

传真: 8610-62775511

邮编: 100084

21世纪经济管理精品教材·经济学系列

应用统计学

李卫东 主编

清华大学出版社
北京

内 容 简 介

统计学是收集、整理、描述和分析数据的一门方法论科学,其目的是探索数据内在的数量规律性,它广泛地应用于自然科学、社会科学和日常生活、工作等各领域。本书主要包括描述统计、推断统计、经济管理中常用的统计方法等内容。与其他教材不同,本书突出了对调查相关内容的介绍,强调统计方法与计算机软件的结合。

本书既可作为高等院校财经、管理类各专业高年级本科学士或 MBA 的教材,也可作为广大实际工作者的参考书。

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/李卫东主编. —北京:清华大学出版社,2014

(21世纪经济管理精品教材·经济学系列)

ISBN 978-7-302-37515-9

I. ①应… II. ①李… III. ①应用统计学—高等学校—教材 IV. ①C8

中国版本图书馆CIP数据核字(2014)第170838号

责任编辑:张 伟

封面设计:汉风唐韵

责任校对:宋玉莲

责任印制:何 芊

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 装 者:北京密云胶印厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:18.5 字 数:425千字

版 次:2014年10月第1版 印 次:2014年10月第1次印刷

印 数:1~4000

定 价:36.00元

产品编号:058013-01



前言

在信息经济的今天,我们每天都要面对纷繁复杂的信息,作为数据处理和分析技术的统计方法也越来越广泛地应用于自然科学、社会科学和日常生活、工作等各领域。在过去一段时间内,我国财经管理类专业的统计教材一直沿用苏联的社会经济统计学科体系,主要侧重于经济统计指标体系的介绍。随着我国社会主义市场经济的不断完善与发展,原有的统计学教材已不能满足需要,需要编写符合需求的教材。作为长期从事应用统计学课程教学的一线工作者,对学生的需求有了一定的了解,一直想编写出一本能满足需求的教材。因此我们结合经济管理专业学科特色编写了本书。

本书主要包括描述统计、推断统计、经济管理中常用的统计方法等几部分内容。本书既可作为高等院校财经、管理类各专业高年级本科学生或MBA的教材,也可作为广大实际工作者的参考书。

本书力求体现以下特点。

第一,在内容上,突出了对调查相关内容的介绍。数据收集是统计中的重要内容之一,加强此方面有利于学生掌握对数据的采集和整理的知识和能力。

第二,强调统计方法与计算机软件相结合。本书对常用统计方法的Excel软件处理做了介绍。通过软件的处理,可减轻学生的负担,也有利于对统计思想的理解和把握,提高其运用统计方法分析和解决问题的能力。

第三,通过引用实际案例,便于学生的理解。在教材中我们引用一些案例,通过在经济管理活动、生活、工作中对统计方法的实际应用例子将统计思想、方法、来龙去脉介绍清楚,便于学生的理解。

本书是北京交通大学经济管理学院统计学科组的集体成果,得到了北京交通大学专业改革试点项目基金的支持。本书的出版得到了清华大学出版社的帮助和支持,在此表示感谢。

本书编写分工：李卫东(第一章、第二章、第七章)、龚玉荣(第三章、第四章)、尹常玲(第五章、第六章)、刘似臣(第八章)、卯光宇(第九章、第十章)、周辉宇(第十一章、第十二章)。全书由李卫东任主编,负责全书大纲的设计、书稿的组织和编纂。

书中的不当或疏漏之处在所难免,恳请同行和读者批评指正。

编者

2014年6月



目录

第一章 绪论	1
第一节 统计与统计学	2
一、统计的含义	2
二、统计学	3
三、统计工作、统计资料与统计学的关系	5
第二节 统计学的发展历程	5
一、古典统计学	5
二、近代统计学	6
三、现代统计学	7
第三节 统计学的分科及其与其他学科的关系	8
一、统计学的分科	8
二、统计学与其他学科的关系	10
第四节 统计学的一些基本概念	11
一、总体与总体单位	11
二、样本	12
三、标志与统计指标	12
四、参数与统计量	13
五、变量	13
第五节 Excel 工具介绍	14
一、数据整理功能	14
二、数据函数功能	15
三、数据分析功能	16
思考与练习	16
第二章 统计数据的搜集	18
第一节 统计数据的类型与来源	19
一、统计数据的测量尺度	19
二、统计数据的类型	21

三、统计数据的来源	22
第二节 调查方案设计	27
一、调查目的及任务	28
二、调查对象和调查单位	28
三、调查项目和调查表	28
四、调查的组织形式和搜集数据的方法	29
五、调查时间和调查地点	30
六、调查的组织实施计划	30
第三节 调查问卷设计	36
一、调查问卷的主要类型	36
二、调查问卷的设计原则	38
三、调查问卷的基本结构	39
四、调查问卷的设计过程	40
五、调查问卷问题的设计	42
六、调查问卷答案的设计	45
七、关于问卷设计其他应注意的问题	48
第四节 调查数据的质量	51
一、调查数据质量的评价标准	51
二、非抽样误差的产生及控制	52
三、抽样误差的产生及控制	52
思考与练习	53
第三章 统计数据的整理与展示	54
第一节 数据的预处理	55
一、数据审核与筛选	55
二、数据排序	56
第二节 品质数据的整理与展示	56
一、分类数据的整理与展示	56
二、顺序数据的整理与展示	59
第三节 数值型数据的整理与展示	60
一、数值型数据分组	60
二、数值型数据的展示	62
第四节 用 Excel 进行数据的整理与展示	65
一、用 Excel 作品质数据的频数分布和条形图	66
二、用 Excel 作数值型数据的频数分布和直方图	67
第五节 合理使用图表	69
一、合理使用统计图	69
二、合理使用统计表	69

思考与练习	70
第四章 数据的概括性度量	73
第一节 集中趋势的度量	73
一、众数	74
二、中位数	75
三、平均数	76
四、众数、中位数和算术平均数的关系	77
第二节 离散程度的度量	78
一、全距	78
二、四分位差	78
三、方差和标准差	79
四、离散系数	80
五、 z 分数	80
第三节 偏度与峰度的度量	81
一、偏度及其测度	81
二、峰度及其测度	81
三、用 Excel 作数据概括性度量	81
思考与练习	82
第五章 概率论初步	85
第一节 随机事件及其概率	85
一、随机试验	85
二、随机事件与样本空间	86
三、事件之间的关系及运算	87
四、随机事件的概率	89
五、条件概率与事件的独立性	92
六、全概率公式与贝叶斯公式	93
第二节 随机变量及其概率分布	95
一、随机变量的概念	95
二、随机变量的概率分布	96
三、连续型随机变量及其概率分布	98
四、分布函数	101
第三节 随机变量的数字特征	103
一、数学期望	104
二、随机变量的方差	106
第四节 大数定律和中心极限定理	110
一、切比雪夫不等式	110

二、大数定理	111
三、中心极限定理	112
四、棣莫弗-拉普拉斯中心极限定理	113
第五节 抽样分布	114
一、抽样与统计量	114
二、抽样分布	116
三、样本均值的分布与中心极限定理	119
四、样本均值比例的抽样分布	120
五、两个样本均值之差的分布	121
六、关于样本方差的分布	122
思考与练习	123
第六章 参数估计	124
第一节 参数估计的基本原理	124
一、估计量与估计值	124
二、点估计与区间估计	125
三、评价估计量的标准	126
第二节 一个总体参数的区间估计	127
一、总体均值的区间估计	128
二、总体比例的区间估计	130
三、总体方差的区间估计	131
第三节 两个总体参数的区间估计	132
一、两个总体均值之差的区间估计	133
二、两个总体比例之差的区间估计	137
三、两个总体方差比的区间估计	138
第四节 样本量的确定	140
一、估计总体均值时样本量的确定	140
二、估计总体比例时样本量的确定	140
思考与练习	141
第七章 假设检验	143
第一节 假设检验的基本原理	143
一、假设检验的基本概念	143
二、原假设和备择假设的建立	144
三、假设检验的基本流程	146
四、假设检验中的两类错误及检验力	148
五、双侧检验与单侧检验	150
第二节 一个总体参数的检验	152

一、检验统计量的确定	152
二、总体均值的假设检验	153
三、总体比例的假设检验	156
四、总体方差的检验	157
第三节 两个总体参数的检验	158
一、检验统计量的确定	158
二、两总体均值与两总体比例之差的假设检验	159
三、两个总体比例之差的检验	160
四、两个总体方差比的检验	161
思考与练习	163
第八章 方差分析	166
第一节 方差分析的基本原理	167
一、方差分析的基本概念	167
二、基本假定与基本思路	168
第二节 单因素方差分析	170
一、单因素方差分析数据结构	170
二、单因素方差分析的一般步骤	171
三、关系强度的测量	174
四、方差分析中的多重比较	174
第三节 双因素方差分析	176
一、双因素方差分析及其类型	176
二、无交互作用双因素方差分析	177
三、有交互作用双因素方差分析	180
思考与练习	184
第九章 一元线性回归	187
第一节 变量间关系的度量	187
一、变量间关系的分类	187
二、相关关系的分类	188
三、线性相关关系的分析	189
第二节 一元线性回归模型	191
一、一元线性模型	191
二、最小二乘估计	192
三、经典假设	193
四、Gauss-Markov 定理	194
五、统计推断	194
六、拟合优度	197

第三节 回归模型的预测与残差分析	198
第四节 案例分析	199
思考与练习	201
第十章 多元线性回归	203
第一节 多元线性回归模型	203
一、最小二乘估计	204
二、经典假设	205
三、Gauss-Markov 定理与统计推断	205
四、拟合优度	206
第二节 预测与残差分析	207
第三节 数据的非线性变换	207
第四节 案例分析	208
思考与练习	209
第十一章 时间序列分析	211
第一节 时间序列成分分解	213
一、长期趋势变动	213
二、季节变动	213
三、循环变动	214
四、不规则变动	214
第二节 时间序列的种类与编制	215
一、时间序列的种类	215
二、时间序列的编制	216
第三节 时间序列的分析指标	217
一、发展水平	217
二、平均发展水平	218
三、增长量和平均增长量	220
四、发展速度和增长速度	221
五、平均发展速度和平均增长速度	222
第四节 长期趋势测定	224
一、时距扩大法和序时平均法	224
二、移动平均法	225
三、数学模型拟合法	227
第五节 季节变动的测定	230
一、按月(季)平均法	230
二、移动平均趋势剔除法	231
第六节 循环变动分析与测定	233

一、循环变动及其测定目的	233
二、循环变动的测定方法	233
思考与练习	236
第十二章 统计指数	238
第一节 概述	239
一、统计指数分类	239
二、统计指数的性质	240
三、指数在经济分析中的作用	241
四、指数编制的基本问题	242
第二节 简单指数与加权指数	242
一、简单指数	242
二、加权综合指数	244
三、加权平均指数	248
第三节 平均指标指数	251
一、可变组成指数	252
二、固定组成指数	253
三、结构影响指数	253
第四节 指数体系与因素分析	254
一、指数体系的概念	254
二、指数体系的基本形式与应用	255
第五节 几种常用的价格指数	261
一、零售价格指数	261
二、消费价格指数	264
三、股票价格指数	266
思考与练习	270
附录 常用统计表	272
参考文献	281



应用案例

奥巴马在2012年美国大选中击败罗姆尼,成功连任。在这次成功连任的过程中,有一个重要的工具就是利用统计数据。美国的《时代》杂志撰文指出奥巴马成功筹集数亿美元竞选资金、有效锁定目标选民、即时调整电视广告投放策略并最终赢得选举的秘诀就是大数据。

在总统竞选一开始,竞选活动经理吉姆·梅斯纳(Jim Messina)已经打算要搞一次完全不同的、以数据为驱动力的竞选活动,准备用数据去衡量这场竞选活动中的每一件事情。梅斯纳聘请了一个5倍规模于2008年竞选时的分析部门,并让雷伊德·加尼(Rayid Ghani)出任芝加哥竞选团队总部的首席科学家。这足以看出该竞选团队对统计分析巨大潜力的重视。下面简述数据统计分析在竞选过程中的应用。

在筹集资金方面,起初,竞选团队就创建了一个单一的巨大系统,数据分析团队可以利用数据进行一些测试,分析出哪些类型的人有可能被某种特定的事情说服。数据分析团队可以预测哪些人会在网上捐钱,哪些人会用邮件捐款,甚至对志愿者进行建模分析。竞选办公室的电话名单不仅列出了姓名和电话号码,并且按被说服的可能性和重要性进行了排序。

有相关人员称,民调数据与选民联系人数据每晚都在所有可能想象的场景下被计算机处理、再处理。每天晚上都模拟运行66 000次选举,这些模拟选举推算出了奥巴马在“摇摆州”的胜率,知道赢得这些州的机会在哪,并让他们得以通过这些数据来分配资源。

此外,竞选团队基于背后强大的内部数据库分析,通过建立复杂的模型定位不同的选民,购买了一些非传统类剧集之间的广告时间,替代了在当地新闻时段的传统做法。最后数据统计显示,广告效率相比2008年提高了14%。

在竞选后期,通过数据分析发现社会化新闻网站Reddit中有许多动员的目标,奥巴马决定去社会化新闻网站Reddit回答问题。

由此可以看出,数据的挖掘分析为奥巴马的成功连任立下了汗马功劳,数据统计分析的地位在政治领域逐渐上升。一位官员说,“人们坐在屋里抽雪茄,高喊‘买《60分钟》’的时代已经结束”。政治领域的大数据时代已经到来。

资料来源:<http://it.sohu.com/20121108/n357017969.shtml>

案例中通过数据分析得出结论会涉及许多其他学科的支持,但是统计学作为一门基础学科,其发挥的重大作用理应引起大家的关注。

本章内容从讨论统计与统计学出发,对统计学进行概述性的分析,最后简要介绍Excel在统计分析中的应用。

第一节 统计与统计学

当今世界,人们在社会、经济、科技等各个领域都感受到信息的作用和影响,人类已步入了信息社会。统计作为信息处理的一种重要手段,已成为广泛应用于社会经济、经营管理的殿堂。了解统计知识对每个人来说也都是必要的。比如,在外出旅游时,你需要关心一段时间内的详细天气预报;在投资股票时,你需要了解股票市场价格和上市公司的财务信息;在观看足球比赛时,除了关心进球的多少外,还想了解各个球队的技术统计结果。

通过统计,人们将会得到许多实用、有趣的东西,同时又会提出不少问题,比如,“统计”什么?为什么商品零售物价指数可以反映社会零售商品价格总水平变动?为什么只调查少数人的人均收入就可以了解一地区乃至全国的人均收入?这些都反映了统计的奥秘。

一、统计的含义

“统计”在现代汉语词典中的解释是指对某一现象有关的数据的搜集、整理、计算和分析等。“统计”一词包含三个含义——统计工作、统计资料和统计学,即围绕研究目的和任务,运用科学的统计方法,获取真实的科学的统计数据,通过分析,了解和认识所研究事物的真相。

(一) 统计工作

美国普渡大学的教授墨尔(David S. Moore)认为,统计是从数据中获取信息的科学。我们大家都很熟悉的统计有体育、人口、经济和证券交易等领域数字资料的汇集。在日常生活工作中,“统计”有着多种含义。例如,老师统计上课的学生人数;NBA 篮球比赛中,人们所关注的球员的技术统计——得分、篮板、助攻等方面的数据;工厂在经过一定时间周期后,要对本厂的该时期的效益进行统计,核算其当期完成的产量和产值等。

统计工作中的“统计”具有计数的含义,在这个意义上统计的起源可以追溯到远古时代。

在原始社会,人类的一般技术活动就蕴含着统计的萌芽。四五千年之前,居住在“两河流域”一代的人们就在石头上记录文字与数字。从历史上看,早在奴隶制国家时代,由于政府赋税、徭役、征兵等需要,就对人口、土地、财产等方面进行调查统计。据记载,公元前 27 世纪,古埃及建造金字塔,为征集建筑费,就曾对全国的人口与财产进行调查。《旧约全书》中记载,公元前 10 世纪,犹太国王大卫和所罗门对全国进行了完整的人口和财产调查统计,以此作为实施统治措施的重要依据。我国春秋战国时代,统计已被认为是治理国家、创立基业的重要手段。齐国丞相管仲在《管子》一书中曾提出:“不明于计数,而欲举大事,犹无舟楫而欲经于水险也。”这就是说不善于利用计数而进行宏伟事业,犹如没有船和桨而想渡过激流险滩一样。秦国主持变法的丞相商鞅认为,“强国知十三数”,即他将十三种数据作为反映国情国力的基本指标。由上可知,在“计数”的含义上,统计的应用十分广泛,而且成为历来人们治理国家必要的一项重要工作。发展至今,统计工作已成为国

家进行宏观管理的一项专门职能工作,如我国从地方政府到中央政府都设有专门的统计部门,负责对国家、地区的经济、科技、文化、教育、社会等多方面进行统计信息的收集、整理、加工,统计工作成为信息、咨询、决策职能的重要载体。

(二) 统计资料

在统计工作的基础上,会形成一系列的数字资料,这是统计工作的成果表现,也称统计资料或统计数据,这是“统计”的另一个含义。它与上述的统计工作是紧密相关的,是统计工作的直接结果,反映了一定空间、时间条件下客观总体现象数量特征的数据型资料,因此也是很早就有的。根据历史记载,我国夏禹时代就开始有人口方面的统计数字。随着人类社会的不断进步,统计数字作为一种重要的信息,人们对它的需要越来越多、越来越细化。平时我们翻阅报刊,就可以看到各种统计数字,如股票指数、价格指数、失业率等。国家统计局、有关信息部门每年出版的统计年鉴和相关资料,反映国家或地区的宏观经济、人口、地理等方面基础资料以及科技、教育、文化等方面的发展等情况,这些都是在此意义上的统计。诸如统计表、统计图、统计手册、统计年鉴等统计资料已成为国家的重要战略资源。

另外,“统计”一词还有“统计学”的含义。作为一门科学的“统计学”,它的出现要比统计工作和统计资料晚得多。这是本书要探讨的主要内容。

二、统计学

统计学是关于数据的科学,是一门收集、整理、总结、描述和分析数据,并在分析的基础上作出正确结论和合理决策方法的科学,它所提供的是一套有关数据收集、处理、分析以及解释并从中探索数据的内在规律性的方法。

关于统计学的定义,需要注意以下几点。

(1) 统计学的研究对象是客观事物现象的数量方面,是总体现象的大量数据,而不是个别的单个数据。因为单个数据在没有其他数据作为参考对照时,是说明不了什么问题的。例如,某学生在英语考试中取得 80 分的成绩,仅通过这么一个数字是无法说明该学生的成绩如何的,因为不知道其他学生的考试成绩。只有知道了其他学生的考试成绩,才能对该学生的成绩进行评价,得出他是处于中等水平,还是较高水平,或是较低水平的结论。另外,统计学所研究的客观现象十分广泛,既包括自然现象,也包括社会现象。如前所述,早期统计所研究的问题有人口调查、保险精算等,后来逐渐扩大到社会经济和生物实验等方面。随着人类的实践活动的需要和信息化的发展,各领域都需要研究事物的数量方面,需要联系数量方面的特征来研究事物的本质。因此统计学的应用范围越来越宽,凡能以数量来表现的现象特征均可作为统计学的研究对象。统计方法已渗透到多个科学领域,成为目前最活跃的学科之一。萨维奇曾说:“统计学基本上是寄生的,靠研究其他领域内的工作而生存。这不是对统计学的轻视,这是因为对很多宿主来说,如果没有寄生虫就会死。对于有的动物来说,如果没有寄生虫就不能消化它们的食物。因此,人类奋斗的很多领域,如果没有统计学,虽然不会死亡,但一定会变得很弱。”

(2) 统计学是一门方法论科学,它是由收集、整理、描述和分析数据资料的统计方法