



世纪高等继续教育精品教材·经济管理类通用系列

Tongjixue Jichu

(第三版)

贾俊平 / 编著



中国人民大学出版社

普通高等教育“十一五”国家级规划教材

21 世纪高等继续教育精品教材·经济管理类通用系列

统计学基础

(第三版)

贾俊平 / 编著

中国人民大学出版社
• 北京 •

图书在版编目 (CIP) 数据

统计学基础/贾俊平编著. —3 版. —北京: 中国人民大学出版社, 2010
21 世纪高等继续教育精品教材. 经济管理类通用系列
ISBN 978-7-300-13136-8

I. ①统… II. ①贾… III. ①统计学-成人教育: 高等教育-教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2011) 第 001724 号

普通高等教育“十一五”国家级规划教材

21 世纪高等继续教育精品教材·经济管理类通用系列

统计学基础 (第三版)

贾俊平 编著

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511398 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京宏伟双华印刷有限公司	版 次	2004 年 7 月第 1 版
规 格	170mm×228mm 16 开本		2011 年 2 月第 3 版
印 张	15.75	印 次	2011 年 4 月第 2 次印刷
字 数	259 000	定 价	26.00 元

版权所有 侵权必究 印装差错 负责调换

总序

21 世纪,科学技术发展日新月异,发明创造层出不穷,知识更新日趋频繁,全民学习、终身学习已经成为适应经济与社会发展的基本途径。近年来,我国高等教育取得了跨越式的发展,毛入学率由 1998 年的 8% 迅速增长到 2008 年的 23.3%,已经进入到大众化的发展阶段,这其中高等继续教育发挥了重要的作用。同时,高等继续教育作为“传统学校教育向终身教育发展的一种新型教育制度”,对实现“形成全民学习、终身学习的学习型社会”、“构建终身教育体系”的宏伟目标,发挥着其他教育形式不可替代的作用。

目前,我国高等继续教育的发展规模已占全国高等教育的一半左右,随着我国产业结构的调整、传统产业部门的改造以及新兴产业部门的建立,各种岗位上数以千万计的劳动者,需要通过边工作边学习来调整自己的知识结构、提高自己的知识水平,以适应现代经济与社会发展的要求。可见,我国高等继续教育的发展,既肩负着重大的历史使命又面临着难得的发展机遇。

我国的高等继续教育要抓住机遇发展,完成自己的历史使命,从根本上说就是要全面提高教育教学质量,这涉及多方面的工作,但抓好教材建设是提高教学质量的基础和中心环节。众所周知,高等继续教育的培养对象主要是已经走上各种生产或工作岗位的从业人员,这就决定了高等继续教育的目标是培养能适应新世纪社会发展要求的动手能力强、具有创新能力的应用型人才。因此,高等继续教育教材的编写“要本着学用结合的原则,重视从业人员的知识更新,提高广大从业人员的思想文化素质和职业技能”,体现出高等继续教育的针对性、实用性和职业性特色。

为适应我国高等继续教育发展的新形式、培养应用型人才、满足广大学员的学习需要,中国人民大学出版社邀请了国内知名专家学者对我国高等继续教育的教学改革与教材建设进行专题研讨,成立了教材编审委员会,联合中国人民大学、中国政法大学、东北财经大学、武汉大学、山西财经大学、东北师范大学、

华中科技大学、黑龙江大学等 30 多所高校,共同编撰了“21 世纪高等继续教育精品教材”,计划在两三年内陆续推出百种高等继续教育精品系列教材。教材编审委员会对该系列教材的作者进行了严格的遴选,编写教材的专家、教授都有着丰富的继续教育教学经验和较高的专业学术水平。教材的编写严格依据教育部颁布的“全国成人高等教育公共课和经济学、法学、工学主要课程的教学基本要求”;教材内容的选择克服了追求“大而全”的现象,做到了少而精,有针对性,突出了能力的训练和培养;教材体例的安排突出了学习使用的弹性和灵活性,体现“以学为主”的教育理念;教材充分利用现代化的教育手段,形成文字教材和多媒体教材相结合的立体化教材,加强了教师对学生学习过程的指导和帮助,形象生动、灵活方便,易于保存,可反复学习,更能适应学员在职、业余自学,或配合教师讲授时使用,会起到很好的教学效果。

这套“21 世纪高等继续教育精品教材”在策划、编写和出版过程中,得到教育部高教司、中国成人教育协会、北京高校成人高教研究会的大力支持和帮助,谨表深切谢意。我们相信,随着我国高等继续教育的发展和教学改革的不深入,特别是随着教育部“高等学校教学质量和教学改革工程”的实施,这套高等继续教育精品教材必将为促进我国高校教学质量的提高做出贡献。

杨干忠

前言

本书是在《统计学基础》(第二版)的基础上修订而成的。考虑到读者的特点和实际学习的需要,本次修订保留了第二版的基本内容和风格,但对部分章节作了较大的调整,如删掉了国民经济核算的内容,对指数部分重新进行了编写。总体上说,内容更加简明易懂。

全书共包括9章,第1章介绍统计学的一些基本概念,目的是让读者对统计学有一个基本的了解。第2章介绍数据的收集方法,使读者了解通过什么方法和途径获取所需的统计数据。第3章介绍数据的整理和展示方法,使读者学会数据的基本处理方法和图形展示技术。第4章介绍数据特征的统计描述方法,主要讨论如何使用一些统计量来对数据进行概括性度量。第5章介绍抽样与参数估计,讨论一些常用的概率抽样方法,并重点介绍总体参数的估计方法。第6章介绍假设检验的基本方法,讨论总体均值和总体比例的检验方法。第7章介绍相关与回归分析方法,讨论变量之间关系强度的度量方法,并介绍一元线性回归方法及应用。第8章介绍时间序列的分析和预测方法,主要讨论时间序列的描述性方法以及对时间序列的预测。第9章介绍指数,主要讨论综合指数的编制方法及实际中常用的几种价格指数。

本书的图表制作以及大多数统计计算都是由 Excel 来完成的,建议有能力的读者尝试使用专业的统计软件(如 SPSS 等)进行一些简单的统计分析。

参加本书编写工作的还有沈阳理工大学经济管理学院徐静霞老师。书中部分练习题和所附的常用统计表选自所列参考书目,在此一并致以衷心的感谢。希望读者多提出宝贵意见,并将您的建议及时反馈给我们,以便做出进一步修改。

贾俊平

2011年1月

于中国人民大学统计学院

CONTENTS

目 录

第1章 统计和数据

1.1 统计及其应用领域	2
一、什么是统计学	2
二、统计的应用领域	3
三、统计的误用与滥用	4
1.2 统计数据及其类型	5
一、分类数据、顺序数据、数值型数据	6
二、观测数据和实验数据	7
三、截面数据和时间序列数据	7
1.3 统计中的基本概念	8
一、总体和样本	8
二、参数和统计量	10
三、变量	11
本章小结	12
思考与练习	13

第2章 数据的收集

2.1 统计数据的来源	16
一、数据的间接来源	16
二、数据的直接来源	17
2.2 调查设计	22
一、调查方案设计	22

二、调查问卷设计	23
2.3 统计数据的质量	33
一、统计数据的误差	33
二、统计数据的质量要求	34
本章小结	34
思考与练习	36

第3章 数据整理与展示

3.1 数据的预处理	38
一、数据审核	38
二、数据筛选	38
三、数据排序	42
3.2 品质数据的整理与展示	42
一、频数与频数分布	42
二、品质数据的图示	45
3.3 数值型数据的整理与展示	48
一、数据分组	48
二、数值型数据的图示	50
3.4 使用图表的注意事项	57
本章小结	58
思考与练习	58

第4章 数据分布特征的测度

4.1 集中趋势的度量	64
一、平均数	64
二、中位数和分位数	65
三、众数	67
四、各度量值的比较	68
4.2 离散程度的度量	68
一、极差和四分位差	69
二、平均差	70

三、方差和标准差	71
四、离散系数：比较几组数据的离散程度	74
4.3 偏态与峰态的度量	76
一、偏态及其测度	76
二、峰态及其测度	77
本章小结	78
思考与练习	79

第5章 抽样与参数估计

5.1 抽样与抽样分布	85
一、概率抽样方法	85
二、抽样分布	87
5.2 参数估计的基本原理	91
一、估计量与估计值	91
二、点估计与区间估计	92
5.3 总体均值的区间估计	95
一、大样本的估计	95
二、小样本的估计	97
5.4 总体比例的区间估计	98
5.5 样本量的确定	99
一、估计总体均值时样本量的确定	99
二、估计总体比例时样本量的确定	100
本章小结	101
思考与练习	102

第6章 假设检验

6.1 假设检验的基本原理	107
一、假设的陈述	107
二、两类错误与显著性水平	110
三、检验统计量与拒绝域	112
四、利用 P 值进行决策	114

6.2 总体均值的检验	116
一、大样本的检验方法	117
二、小样本的检验方法	120
6.3 总体比例的检验	123
本章小结	125
思考与练习	125

第7章 相关与回归分析

7.1 变量间关系的度量	130
一、变量间的关系	130
二、相关关系的描述与测度	132
7.2 一元线性回归	137
一、一元线性回归模型	137
二、参数的最小二乘估计	138
三、回归方程的拟合优度	142
四、显著性检验	145
7.3 利用回归方程进行估计和预测	149
一、点估计	149
二、区间估计	150
本章小结	152
思考与练习	153

第8章 时间序列分析和预测

8.1 时间序列及其分解	158
8.2 时间序列的描述性分析	159
一、图形描述	159
二、增长率分析	161
8.3 预测方法的选择与评估	162
8.4 平稳序列的预测	163
一、移动平均法	163
二、指数平滑法	165

8.5 有趋势序列的预测	167
一、线性趋势预测	167
二、非线性趋势预测	170
8.6 复合型序列的分解预测	172
本章小结	177
思考与练习	178

第9章 指 数

9.1 指数的含义与分类	182
9.2 加权指数	183
一、加权综合指数	183
二、加权平均指数	185
三、价值指数与指数体系	186
9.3 几种常用的价格指数	188
一、消费者价格指数	188
二、生产者价格指数	190
三、股票价格指数	190
本章小结	191
思考与练习	192

附录一 练习题参考答案	194
附录二 常用统计表	222
参考文献	236

基础统计知识

第1章

统计和数据

■ 内容提要

本章主要介绍统计这门课程中的一些基本问题，目的是使读者对统计中的一些基本概念有一个正确的理解和认识。具体包括：(1) 统计及其应用领域。介绍统计学的含义及统计的应用领域等。(2) 统计数据及其类型。介绍统计数据及其类型和统计数据与统计方法之间的关系。(3) 统计中的基本概念。介绍统计中常用的几个基本概念，包括总体和样本、参数和统计量以及变量等。

1.1 统计及其应用领域

一、什么是统计学

在日常生活中,我们经常会接触到“统计”这一术语,在有关媒体报道中也经常会看见一些统计数据、图表等。很多人可能难免会将“统计”一词与统计工作联系起来:一提到统计首先想到的是统计工作,想到我们的政府统计机构、企业中从事统计工作的人,等等。这种理解是自然的,但又不全面的。统计活动已有悠久的历史,可以说,自从有了国家就有了统计活动。最初,统计只是一种计数活动,为满足统治者管理国家的需要而收集资料,通过统计计数来弄清国家的人力、物力和财力,作为国家管理的依据。然而在今天,“统计”已发展成为一门数据分析的学科,统计方法也已被应用到科学研究的各个领域。目前,统计学已发展成为具有多个分支学科的大家族。统计学家们给统计学下的定义繁多,其表述大同小异。本书将统计学的定义表述如下:

定义 1.1 统计学 (statistics) 是收集、处理、分析、解释数据并从数据中得出结论的科学。

这一定义表明统计学是一套处理和分析数据的方法与技术。统计分析数据所用的方法大体上可分为描述统计和推断统计两大类。

定义 1.2 描述统计 (descriptive statistics) 是研究数据收集、处理和描述的统计学方法。

描述统计的内容包括如何取得研究所需要的数据,如何用图表形式对数据进行处理和展示,如何通过对数据的综合、概括与分析,得出所关心的数据的特征。

定义 1.3 推断统计 (inferential statistics) 是研究如何利用样本数据来推断总体特征的统计学方法。

推断统计的内容包括参数估计和假设检验两大类。参数估计是利用样本信息推断所关心的总体特征,假设检验则是利用样本信息判断对总体的某个假设

是否成立。比如，从一批灯泡中随机抽取少数几个灯泡作为样本，测出它们的使用寿命，然后根据样本灯泡的平均使用寿命估计这批灯泡的平均使用寿命，或者是检验这批灯泡的使用寿命是否等于某个假定值，这就是推断统计要解决的问题。

统计学是一门关于数据的科学，它研究的是来自各个领域的数据，统计学由一套处理数据的方法所组成，这些方法来源于对数据的研究。离开了数据，统计方法就失去了用武之地，统计学也就失去了存在的意义。然而数据如果不用统计方法去分析，那么它们也仅仅是一堆数据而已，不能得出任何有益的结论。

二、统计的应用领域

目前，统计方法已被应用到自然科学和社会科学的众多领域，统计学也已发展成为由若干分支学科组成的学科体系。可以说，几乎所有的研究领域都要用到统计方法。统计是适用于所有学科领域的通用数据分析方法，只要有数据的地方就会用到统计方法。这里，我们不想列举统计的应用领域，只想通过几个简单的例子说明统计的应用。

【例 1.1】 用统计识别作者。1787—1788 年，三位作者亚历山大·汉密尔顿 (Alexander Hamilton)，约翰·杰伊 (John Jay) 和詹姆斯·麦迪逊 (James Madison) 为了说服纽约人认可宪法，匿名发表了 85 篇著名的论文。在这些论文中，大多数的作者已经得到了识别，但是其中的 12 篇论文的作者身份引起了争议。通过对不同的单词的频数进行统计分析，得出了结论：詹姆斯·麦迪逊最有可能是这 12 篇论文的作者。现在，对于这些存在争议的论文，认为詹姆斯·麦迪逊是作者的观点占主导地位，而且几乎可以肯定这种观点是正确的。

【例 1.2】 用简单的描述统计量得到一个重要发现。费舍 (R. A. Fisher) 在 1952 年的一篇文章中举了一个例子，说明如何由基本的描述统计量的知识引出一个重要的发现。20 世纪早期，哥本哈根卡尔堡实验室的施密特 (J. Schmidt) 发现不同地区所捕获的同种鱼类的脊椎骨和鳃腺的数量有很大不同，甚至在同一海湾内不同地点所捕获的同种鱼类，也有这样的倾向；然而，鳗鱼的脊椎骨的数量却变化不大。施密特从欧洲大陆各地、冰岛、亚速尔群岛以及尼罗河等几乎分离的水域里所捕获的鳗鱼的样本中，计算发现了几乎一样的均值和标准偏差值。由此，施密特推断所有各个不同海域内的鳗鱼是由海洋

中某公共场所繁殖的。后来名为“戴纳 (Dana)”的科学考察船在一次远征中发现了这个场所。

【例 1.3】 挑战者号航天飞机失事预测。1986 年 1 月 28 日清晨, 载有 7 名宇航员的挑战者号进入发射状态。就在发射前, 有冰片牢附在机壳上。几分钟后, 正当电视新闻报道它已进入轨道时, 航天飞机在毁灭性的爆炸中化成碎片, 机上的宇航员尸骨未存。推动航天飞机进入太空的两个固体燃料发动机是由 Thiokol 公司制造的。失事前的一天晚上, Thiokol 公司的经理们和国家航空航天局 (NASA) 就如期发射还是推迟发射产生了争执。天气预报发射时的气温为 31°F 。争执的结果是采纳了 Thiokol 公司经理们的建议: 按计划发射航天飞机。因为他们觉得没有确凿的证据表明低温会对固体燃料火箭推进器的性能产生影响。在失事前, 该航天飞机曾 24 次成功发射。将航天飞机送入太空的两个固体燃料推进器由 6 只 O 形项圈密封。在几次飞行中, 曾发生过 O 形项圈被腐蚀或气体泄漏事故, 这样的事故是极其危险的。前 24 次发射中有一次发动机遭到了永久性破坏。根据 23 次飞行中发生腐蚀或泄漏事故的次数 (因变量 y) 及火箭连接处的温度 (自变量 x) 数据, 进行线性回归得到的回归方程为 $y = 3.698 - 0.04754x$ 。当温度为 31°F 时, O 形项圈发生事故的预计次数为 2.225 次。结果显示连接处的温度与 O 形项圈事故之间有一定的相关性。如果当时那些经理们看到了回归的预测结果, 也许会选择推迟发射。

前两个是统计得以应用并取得成效的例子, 后一个是统计结果未被使用而酿成惨剧的例子。不管怎样, 它们都表明统计在许多领域中都有着广泛的应用。

三、统计的误用与滥用

统计常常被人们有意或无意地滥用, 比如, 错误的统计定义、错误的图表展示、不合理的样本、数据的遗漏或逻辑错误等。这些误用有些是常识性的, 有些是技术性的, 有些则是故意的。作为从数据中寻找事实的统计, 却被有些人变成了歪曲事实的工具。你也许常常看到这样的产品质检报告: “某某产品的抽样合格率是 80%。”乍看上去没什么问题, 但如果事实上只抽查了 5 件产品, 有 4 件合格。那么这样的合格率能说明什么问题呢? 在马路上随便采访几个人, 他们的看法能代表大多数人的观点吗? “调查结果表明……”调查了多少个? 是随机调查的吗? 样本是怎样选取的? 这看上去是在用事实说话, 实际上成了统计陷阱。

在有些人的心目中，数据分析就是寻找支持。他们的心目中可能有了某种“结论”性的东西，或者说他们希望看到一种符合他们需要的某种结论，而后去找些统计数据来支持他们的结论，这恰恰歪曲了数据分析的本质。数据分析的真正目的是从数据中找出规律，从数据中寻找启发，而不是寻找支持。真正的数据分析事先是没有结论的，通过对数据的分析才能得出结论。

此外，统计也往往被作为两个极端使用：一个极端是不懂或不太懂统计的人认为统计没什么用。他们因为不懂统计而瞧不起统计，他们不用或几乎不用统计方法分析数据，即使做些统计分析，也往往是表面上的。走入这一极端的人，他们的决策依据就是自己的大脑：一些杂乱无章的信息组合出的某种直觉。如果他们的决策是正确的，就更增加了他们的自信，更加感到不用统计也挺好；如果他们的决策出了毛病，那么便会找出一大堆理由：市场难测、环境突变、竞争激烈、需求疲软、价格下跌、管理不善、成本上升、出口下降等。总之，决策失误的理由肯定与统计无关。另一个极端是把简单问题复杂化。特别是在管理领域，一些管理者把本来可以用简单方法解决的问题故意复杂化，他们为证明管理的科学性，建立一个别人看不懂的模型，编一大堆程序，输出一大堆数字和符号；他们得出用统计语言陈述的结论，提出一些似是而非的建议。这样的分析往往脱离了管理问题，对实际决策也未必有用。在统计应用中，这两个极端都是不可取的。管理决策中不使用统计几乎不可想象，把简单问题复杂化对管理决策也未必有用。从统计的实际应用来看，简单的方法不一定没用，复杂的方法也不一定有用。正如有的学者所说的，最简单的模型往往是最有用的。统计应该恰当地被应用到它能起作用的地方，不能把它神秘化，更不能把它歪曲，把它作为掩盖事实的陷阱。

1.2 统计数据及其类型

既然统计是研究数据的科学，那么了解统计数据及其类型就显得很有必要。数据（data）是对现象进行计量的结果。比如，对经济活动总量的计量可以得到国内生产总值（GDP）的数据，对股票价格变动水平的计量可以得到股票价格指

数的数据,对人口性别的计量可以得到男或女这样的数据,等等。

统计数据不仅仅是数字,也可以是文字。按照所采用的计量尺度不同,可以将统计数据分为分类数据、顺序数据和数值型数据。统计数据还可以从其他角度进行分类,比如,按照统计数据的收集方法,可以将其分为观测数据和实验数据;按照被描述的对象与时间的关系,可以将统计数据分为截面数据和时间序列数据。

一、分类数据、顺序数据、数值型数据

定义 1.4 只能归于某一类别的非数字型数据,称为分类数据 (categorical data)。

分类数据是对事物进行分类的结果,数据则表现为类别,是用文字来表述的。例如,人口按照性别分为男、女两类;企业按照经济性质分为国有、集体、私营、合资、独资企业等,这些均属于分类数据。

在分类数据中,各类别之间是平等的并列关系,无法区分优劣或大小,各类别之间的顺序是可以任意改变的,分类数据只具有 $=$ 或 $\neq$ 的数学特性。因此,分类数据有时也称为无序分类数据。虽然分类数据只表现为某种类别,但为了便于统计处理,特别是为了便于计算机识别,可以将不同类别用不同的数字或编码来表示,比如用“1”表示男性人口,“0”表示女性人口;用“1”表示国有企业,“2”表示集体企业,“3”表示私营企业等。这些数字只是不同类别的对应代码,并不意味着这些数字可以区分大小或进行任何数学运算。

定义 1.5 只能归于某一有序类别的非数字型数据,称为顺序数据 (rank data)。

顺序数据也是对事物进行分类的结果,但这些类别是有顺序的。比如产品可以分为一等品、二等品、三等品、次品;考试成绩可以分为优、良、中、及格、不及格;一个人的受教育水平可以分为小学、初中、高中、大学及以上;一个人对某一事物的态度可以分为非常同意、同意、保持中立、不同意、非常不同意,等等。这些均属于顺序数据。

顺序数据虽然也表现为类别,但这些类别之间是可以比较顺序的。因此,顺序数据有时也称为有序分类数据。很显然,顺序数据要比分类数据精确一些,它除了具有 $=$ 或 $\neq$ 的数学特性外,还具有 $>$ 或 $<$ 的数学特性。