



新编21世纪远程教育精品教材

• 经济与管理系列 •

统计学教程

(第三版)

主 编 金勇进





新编21世纪远程教育精品教材

• 经济与管理系列 •

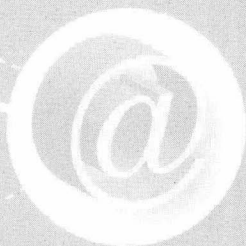
统计学教程

(第三版)

主 编 金勇进

(北京邮电大学)

副主编 王 芳 高 兴 王 芳
编 者 文 欣 田 颖 李 颖 李 颖
审 校 文 欣 田 颖 李 颖 李 颖



中国人民大学出版社

• 北京 •

图书在版编目 (CIP) 数据

统计学教程/金勇进主编. —3 版. —北京: 中国人民大学出版社, 2016. 1
新编 21 世纪远程教育精品教材·经济与管理系列
ISBN 978-7-300-22112-0

I. ①统… II. ①金… III. ①统计学-远程教育-教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2015) 第 270632 号

新编 21 世纪远程教育精品教材·经济与管理系列
统计学教程 (第三版)
主 编 金勇进
Tongjixue Jiaocheng

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)		010-62511770 (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	北京东方圣雅印刷有限公司	版 次	2014 年 11 月第 1 版
规 格	185 mm×260 mm 16 开本		2016 年 1 月第 3 版
印 张	19	印 次	2016 年 1 月第 1 次印刷
字 数	430 000	定 价	39.00 元

新编 21 世纪远程教育精品教材 编委会

(以姓氏笔画为序)

丁兴富	尹伟中	任为民
李林曙	张爱文	陈 丽
郝成义	顾宗连	黄荣怀

作者简介

金勇进，教授，博士生导师，中国人民大学统计学院前院长，教育部重点研究基地“应用统计科学研究中心”主任，教育部统计学专业教学指导分委员会副主任委员，中国商业统计学会会长，高校市场调查与教学研究会会长，国务院政府特殊津贴获得者。主要研究成果：出版专著、译著、教材 30 多部，发表学术论文 150 多篇，获各类教学、科研成果奖 30 多项。

内容简介

本书在第二版的基础上进行了相应的修订，系统地介绍了基本的统计分析方法，具体内容包括数据的收集、数据的描述、概率与概率分布、参数估计、假设检验、方差分析、时间序列分析等。本书在讲授统计方法的同时，注重方法背后统计思想的阐述，并注重统计方法与计算机的结合，语言通俗易懂，既可作为教师讲课的教材，也适用于学员自学。



总序

我们正处在教育史尤其是高等教育史上的一个重大的转型期。在全球范围内，包括在我们中华大地，以校园课堂面授为特征的工业化社会的近代学校教育体制，正在向基于校园课堂面授的学校教育与基于信息通信技术的远程教育相互补充、相互整合的现代终身教育体制发展。一次性学校教育的理念已经被持续性终身学习的理念所替代。在高等教育领域，从 1088 年欧洲创立博洛尼亚（Bologna）大学以来，21 世纪以前的各国高等教育基本是沿着精英教育的路线发展的，这也包括自 19 世纪末创办京师大学堂以来我国高等教育短短一百多年的发展史。然而，自 20 世纪下半叶起，尤其在迈进 21 世纪时，以多媒体计算机和互联网为主要标志的电子信息技术正在引发教育界的一场深刻的革命。高等教育正在从精英教育走向大众化、普及化教育，学校教育体系正在向终身教育体系和学习型社会转变。在我国，党的十六大明确了全面建设小康社会的目标之一就是构建学习型社会，即要构建由国民教育体系和终身教育体系共同组成的有中国特色的现代教育体系。

教育史上的这次革命性转型绝不仅仅是科学技术进步推动的。诚然，以电子信息通信技术为主要代表的现代科学技术的进步，为实现从校园课堂面授向开放远程学习、从近代学校教育体制向现代终身教育体制和学习型社会的转型提供了物质技术基础。但是，教育形态演变的深层次原因在于人类社会经济发展和生活变革的需求。恰在这次世纪之交，人类社会开始进入基于知识经济的信息社会。知识创新与传播及应用、人力资源开发与人才培养已经成为各国提高经济实力、综合国力和国际竞争力的关键和基础。而这些仅仅依靠传统学校校园面授教育体制是无法满足的。此外，国际社会面临的能源、环境与生态危机，气候异常，数字鸿沟与文明冲突，对物种多样性与文化多样性的威胁等多重全球挑战，也只有依靠世界各国进一步深化教育改革与创新，促进人与自然的和谐发展才能得到解决。正因为如此，我国党和政府提出了“科教兴国”、“可持续发展”、“西部大开发”、“缩小数字鸿沟”以及“人与自然和谐发展”的“科学发展观”等基本国策。其中，对教育作为经济建设的重要战略地位和基础性、全局性、前瞻性产业的确认，对高等教育对于知识创新与传播及应用、人力资源开发与人才培养的重大意义的关注，以及对发展现代教

育技术、现代远程教育和教育信息化并进而推动国民教育体系现代化,构建终身教育体系和学习型社会的决策更得到了教育界和全社会的共识。

在上述教育转型与变革时期,中国人民大学一直走在我国大学的前列。中国人民大学是一所以人文、社会科学和经济管理为主,兼有信息科学、环境科学等的综合性、研究型大学。长期以来,中国人民大学充分利用自身的教育资源优势,在办好全日制高等教育的同时,一直积极开展远程教育和继续教育。中国人民大学在我国首创函授高等教育。1952年,校长吴玉章和成仿吾创办函授教育的报告得到了刘少奇的批复,并于1953年率先招生授课,为新建的共和国培养了一大批急需的专门人才。在20世纪90年代末,中国人民大学成立了网络教育学院,成为我国首批现代远程教育试点高校之一。经过短短几年的探索和发展,中国人民大学网络教育学院创建的“网上人大”品牌,被远程教育界、媒体和社会誉为网络远程教育的“人大模式”,即“面向在职成人,利用网络学习资源和虚拟学习社区,支持分布式学习和协作学习的现代远程教育模式”。成立于1955年的中国人民大学出版社是新中国建立后最早成立的大学出版社之一,是教育部指定的全国高等学校文科教材出版中心。在过去的几年中,中国人民大学出版社与中国人民大学网络教育学院合作策划、创作出版了国内第一套极富特色的“新编21世纪远程教育精品教材”。这些凝聚了中国人民大学、北京大学、北京师范大学等北京知名高校学者教授、教育技术专家、软件工程师、教学设计师和编辑们广博才智的精品课程系列教材,以印刷版、光盘版和网络版立体化教材的范式探索构建全新的远程学习优质教育资源,实现先进的教育教学理念与现代信息通信技术的有效结合。这些教材已经被国内其他高校和众多网络教育学院所选用。中国人民大学出版社基于“出教材学术精品,育人文社科英才”理念的努力探索及其初步成果已经得到了我国远程教育界的广泛认同,是值得肯定的。

2005年4月,我被邀请出席《中国远程教育》杂志与中国人民大学出版社联合主办的“远程教育教材的共建共享与一体化设计开发”研讨会并做主旨发言,会后受中国人民大学出版社的委托为“新编21世纪远程教育精品教材”撰写“总序”,这是我的荣幸。近几年来,我一直关注包括中国人民大学网络教育学院在内的我国高校现代远程教育试点工程。这次,更有机会全面了解和近距离接触中国人民大学出版社推出的“新编21世纪远程教育精品教材”及其编创人员。我想将我在上述研讨会上发言的主旨做进一步的发挥,并概括为若干原则作为我对包括中国人民大学出版社、中国人民大学网络教育学院在内的我国网络远程教育优质教育资源建设的期待和展望:

- 新编21世纪远程教育精品教材的教学内容要更加适应大众化高等教育面对在职成人、定位在应用型人才培养上的需要。
- 新编21世纪远程教育精品教材的教学设计要更加适应地域分散、特征多样的远程学生自主学习的需要,培养适应学习型社会的终身学习者。
- 在我国网络教学环境渐趋完善之前,印刷教材及其配套教学光盘依然是远程教材的主体,是多种媒体教材的基础和纽带,其教学设计应该给予充分的重视。要在印刷教材的显要部位对课程教学目标和要求做明确、具体、可操作的陈述,要清晰地指导远程学生如何利用多种媒体教材进行自主学习和协作学习。
- 应组织相关人员对多种媒体的远程教材进行一体化设计和开发,要注重发挥多种媒

体教材各自独特的教学功能，实现优势互补。要特别注重对学生学习活动、教学交互、学习评价及其反馈的设计和实现。

- 要将对多种媒体远程教材的创作纳入到对整个远程教育课程教学系统的一体化设计和开发中去，以便使优质的教材资源在优化的教学系统、平台和环境中的有效的教学模式、学习策略和学习支助服务的支撑下获得最佳的学习成效。

- 要充分发挥现代远程工程试点高校各自的学科资源优势，积极探索网络远程教育优质教材资源共建共享的机制和途径。

中华人民共和国教育部远程教育专家顾问

丁兴富

前言

《统计学教程（第二版）》自 2010 年出版发行以来，受到读者的欢迎。统计学是关于数据的科学，在大数据浪潮滚滚而来的背景下，欲对数据进行更好的挖掘与使用，基本的统计学知识是必不可少的。本书的宗旨是介绍基本的统计方法，同时解读背后其蕴含的统计思想。本书的最大特点是通俗、易懂，没有追求统计方法中的公式推导，而是借助大量生动活泼的例子阐述统计方法的实际应用。与其他同类教科书相比，本书的主要特点表现在：

1. 内容全面，重点突出。本书覆盖了一般统计学教科书中的核心内容，但对一些内容有所侧重。例如，在第二章侧重讨论了误差问题，我们认为统计学的核心问题是对误差的描述和分析；在第三章介绍统计图表的内容后，专门讨论了如何制作一个好的统计图，因为在实践中经常出现统计图误用的情形；在第五章对重要的离散型和连续型随机变量概率分布进行介绍后，自然地引出正态分布衍生的几个重要分布，在此基础上进一步讨论抽样分布，使得逻辑连贯、紧凑；第六章和第七章遥相呼应，把重点集中于一个总体和两个总体的参数估计与检验；第八章讨论的列联表中的相关测量是许多教科书中还没有涉及的内容，但在数据分析中十分必要；第十二章介绍综合评价指数，因为对数据的综合评价是现实中经常遇到的问题。

2. 省略公式推导，强调实际应用。“统计学教程”主要讲授统计方法，方法中的公式推导是统计理论的重要内容。本书没有追求公式推导，而是强调统计方法的实际应用。所以本书中的例题较多，试图通过例题告诉读者统计方法的具体应用；在一些章后介绍了 Excel 的使用也是基于强调应用的考虑，使读者学完之后立即能上手，利用相关软件进行统计问题的处理。

3. 本书内容通俗易懂、生动活泼。每章前都精心挑选一个生动案例，引出该章介绍的内容；作者尽量采用生动的语言，设计“学习导航”“小词典”“你知道吗”“试一试”“想一想”等栏目，介绍重要概念，拓展相应知识，引发读者思考，培养动手能力。各章中也有数量不等的“人物小传”，介绍著名统计学家生平和主要学术贡献，帮助读者在学

习统计方法的同时，对统计学发展史有所了解。每章的最后有“本章小结”，强调该章核心知识点；本书还给出了各章习题的参考答案，便于读者对照检查。

本书特别适合于自学“统计学”课程的读者，也可以作为各类院校“统计学”课程的教材。在本书的写作过程中，中国人民大学统计学院的刘展博士和于力超博士帮助作者整理材料，经作者最后修订形成本书。在此，向刘展博士和于力超博士表示深深的感谢！本书的再版得到了中国人民大学出版社的大力支持，在此一并致以谢意！

金勇进

目 录

第一章 引 论	1
第一节 统计数据与统计学	2
第二节 一些基本概念	5
第三节 统计与统计软件	10
本章小结	11
思考与练习	11
第二章 数据的收集	13
第一节 数据的来源	14
第二节 数据误差	20
第三节 数据文件	24
本章小结	25
思考与练习	26
第三章 数据的描述——数据的直观显示	27
第一节 用统计表描述数据	27
第二节 用统计图描述数据	34
本章小结	47
思考与练习	47
第四章 数据的描述——重要的统计量	51
第一节 集中趋势的描述	52
第二节 离散趋势的描述	62
第三节 偏态与峰度的描述	68
第四节 数据的标准化处理	72
本章小结	74
思考与练习	74

第五章 概率与概率分布	77
第一节 概率的问题	78
第二节 离散型随机变量的概率分布	86
第三节 连续型随机变量的概率分布	91
第四节 抽样分布	100
本章小结	101
思考与练习	102
第六章 参数估计	105
第一节 点估计	106
第二节 区间估计	110
第三节 一个总体情形下参数的区间估计	112
第四节 两个总体情形下参数的区间估计	114
第五节 关于样本量	118
本章小结	121
思考与练习	122
第七章 假设检验	126
第一节 假设检验的基本问题	127
第二节 一个总体参数的假设检验	133
第三节 两个总体参数的假设检验	140
本章小结	148
思考与练习	149
第八章 列联分析	153
第一节 定类数据与列联表	154
第二节 拟合优度检验	158
第三节 独立性检验	163
第四节 列联表中的相关测量	165
第五节 列联分析中应注意的问题	169
本章小结	171
思考与练习	172
第九章 方差分析	175
第一节 基本问题	176
第二节 单因素方差分析	180
第三节 双因素方差分析	185
本章小结	191
思考与练习	192
第十章 相关与回归	197
第一节 相关分析	198
第二节 回归分析	203

第三节 用回归方程进行预测	212
本章小结	215
思考与练习	216
第十一章 时间序列分析	220
第一节 时间序列的描述	221
第二节 时间序列的分解法	229
第三节 时间序列的平滑法	240
本章小结	245
思考与练习	246
第十二章 指 数	248
第一节 基本问题	249
第二节 总指数编制方法	252
第三节 指数体系	259
第四节 几种典型的指数	263
第五节 综合评价指数	269
本章小结	272
思考与练习	273
思考与练习参考答案	276
参考文献	283

统计无处不在

我们每天都可以在报纸上、网络上、电视机中看到、听到各种数据，这些数据是统计研究的结果。这些数据意味着什么？数据是如何收集的？如何进行数据分析以获得直观的结论，从而反映我们感兴趣的变量的现状和发展态势？请看下面一些统计数据 and 统计研究结果：

(1) 2009 年 7 月北京期房商品住宅的销量为 12 840 套，剔除其中的经济适用房和限价房外的期房商品住宅签约套数为 10 862 套，日均销售约 350 套，较 6 月的日均销售 385 套下跌 9.1%。

(2) 央行在公开市场发行了 50 亿元一年期央行票据，发行利率水平较前次发行直降近 95 个基点，为 2.245 9%，达到了 2006 年 5 月份的水平。

(3) 统计分析结果显示，第二个出生的子女没有第一个聪明，第三个出生的子女没有第二个聪明，以此类推。

(4) 2000 年，洛杉矶湖人队 14 名球员的年薪平均数约为 410 万美元，中位数约为 260 万美元。

(5) 全球平均气温近些年总体具有变高的趋势。

这些都是统计学研究的问题和应用统计方法得出的结论。其他类似的股票行情、物价指数、房价波动、汇率变化，乃至交通事故、死亡人数等各种数据围绕着我们，我们生活在数据之中，想避开都难，可以说，有数据的地方就有统计，统计无处不在。

学习导航

- 统计数据的不同类型：按照数据的不同计量尺度、不同收集方法和是否与时间相关划分。
- 统计学的应用领域。
- 统计学的一些基本概念。
- 常用的统计软件。

第一节 统计数据与统计学

一、统计数据的类型

数据是统计的起点,要进行统计分析,首先要获得相关数据。在学习统计学前,有必要了解统计数据的类型。

(一) 按照所采用计量尺度的不同分类

统计数据按照所采用计量尺度的不同可划分为三种类型,即数值型数据、分类型数据和顺序型数据。

1. 数值型数据 (metric data)

数值型数据是指直接使用自然数或度量衡单位进行计量的用数字尺度测量的观察值。

例如,GDP 等的宏观经济运行数据、每天进出海关的旅游人数、某地流动人口的数量等。数值型数据的表现就是具体的数值,统计处理中的大多数都是数值型数据。

2. 分类型数据 (categorical data)

分类型数据是指反映事物类别的数据。分类属性具有有限个(可能很多)不同值,值之间无序。

例如,人口按性别分为男、女两类,受教育程度也可以按不同类别来区分。这种数据通常用频数(率)表示分类的结果,例如,某地区男性常住人口占 52%、女性占 48%。

3. 顺序型数据 (rank data)

只能归于某一有序类别的非数字型数据,称为顺序型数据。

例如,满意度调查中的选项有“非常满意”“比较满意”“比较不满意”“非常不满意”等;再如将学生的考试成绩分为优、良、中、及格、不及格等。

在这三类数据中,数值型数据由于说明了事物的数量特征,因此可归为定量数据(quantitative data),分类型数据和顺序型数据由于定义了事物所属的类别,说明了事物的品质特征,因而可统称为定性数据(qualitative data)。区分数据的类型非常重要,这是因为不同类型的数据在一些情况下,需要用不同的统计方法进行处理。

统计数据按所采用计量尺度的不同的分类如图 1—1 所示。

(二) 按照收集方法的不同分类

按照收集方法的不同,可将统计数据分为观测数据和实验数据两类。

1. 观测数据 (observational data)

观测数据是在没有对事物进行人为控制的条件下,通过调查或观测而收集到的数据。

观测数据主要集中在社会经济领域。例如,对商品零售价格变动水平进行测量可以得到商品零售价格指数,对股票价格变动水平进行测量可以得到股票价格指数等。

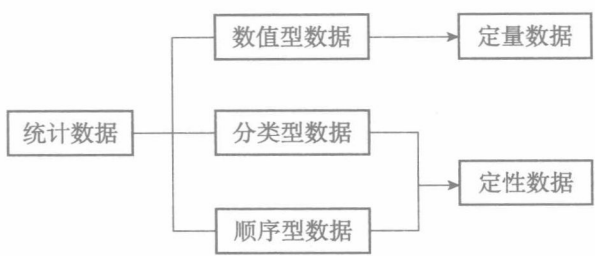


图 1—1 统计数据按所采用计量尺度的不同的分类

2. 实验数据 (experimental data)

实验数据指的是通过在实验中控制实验对象而收集到的数据。

实验数据主要集中在自然科学领域。比如某种新型电池的使用寿命、一种新型降压药疗效的实验数据等。

(三) 按照是否与时间相联系分类

按照是否与时间相联系，还可以将统计数据分为截面数据和时间序列数据。

1. 截面数据 (cross-section data)

在相同或近似相同的时间点上收集的数据，称为截面数据。

截面数据描述了现象在某一时刻的变化情况，它通常是在不同的空间上获得的数据，表 1—1 显示的数据就是截面数据。

表 1—1 2007 年我国部分地区国内生产总值及其构成数据 单位：亿元

地区	国内生产总值	第一产业	第二产业	第三产业
北京	9 353.32	101.26	2 509.40	6 742.66
天津	5 050.40	110.19	2 892.53	2 047.68
河北	13 709.50	1 804.72	7 241.80	4 662.98
山西	5 733.35	269.68	3 438.58	2 025.09
内蒙古	6 091.12	762.10	3 154.56	2 174.46

2. 时间序列数据 (time series data)

在不同时间上收集到的数据，称为时间序列数据。

时间序列数据常用于描述现象随时间变化的情况。表 1—2 显示的数据就是时间序列数据。

表 1—2 2001—2007 年我国国内生产总值及其构成数据 单位：亿元

年份	国内生产总值	第一产业	第二产业	第三产业
2001	109 655.2	15 781.3	49 512.3	44 361.6
2002	120 332.7	16 537.0	53 896.8	49 898.9
2003	135 822.7	17 381.7	62 436.3	56 004.7
2004	159 878.3	21 412.7	73 904.3	64 561.3
2005	183 217.5	22 420.0	87 364.6	73 432.9
2006	211 923.4	24 040.0	103 162.0	84 721.4
2007	249 529.8	28 095.0	121 381.3	100 053.5

二、统计学

统计学是一门收集、整理和分析数据的科学，有数据的地方，就有统计学的应用。收集数据并研究如何得到数据，与之对应的是统计学中的抽样调查和试验设计等理论；整理数据指的是将数据用图或表的形式展现出来，与之对应的是描述统计的方法；分析数据指的是选择适当的统计方法研究数据，并从数据中提取有用信息进而得出结论，更多地对应于推断统计的理论与方法，包含参数估计、假设检验、相关分析、回归分析、时间序列分析等诸多内容。

统计学的应用领域非常广泛，是一门适用于几乎所有学科领域的通用数据分析方法。无论是学术研究还是政府管理，无论是公司或企业的生产经营管理还是人们的日常生活，都离不开统计方法的应用。

表 1—3 列出了统计的一些应用领域，从中可以看到，统计无处不在，统计学是非常有实用价值的。

表 1—3 统计的应用领域

Actuarial work (精算)	Hydrology (水文学)
Agriculture (农业)	Industry (工业)
Animal science (动物学)	Linguistics (语言学)
Anthropology (人类学)	Literature (文学)
Archaeology (考古学)	Manpower planning (劳动力计划)
Auditing (审计学)	Management science (管理科学)
Crystallography (晶体学)	Marketing (市场营销学)
Demography (人口统计学)	Medical diagnosis (医学诊断)
Dentistry (牙医学)	Meteorology (气象学)
Ecology (生态学)	Military science (军事科学)
Econometrics (经济计量学)	Nuclear material safeguards (核材料安全管理)
Education (教育学)	Ophthalmology (眼科学)
Election forecasting and projection (选举预测和策划)	Pharmaceutics (制药学)
Engineering (工程学)	Physics (物理学)
Epidemiology (流行病学)	Political science (政治学)
Finance (金融)	Psychology (心理学)
Fisheries research (水产渔业研究)	Psychophysics (心理物理学)
Gambling (博彩业)	Quality control (质量控制)
Genetics (遗传学)	Religious studies (宗教研究)
Geography (地理学)	Sociology (社会学)
Geology (地质学)	Survey sampling (抽样调查)
Historical research (历史研究)	Taxonomy (分类学)
Human genetics (人类遗传学)	Weather modification (气象改善)

资料来源：贾俊平、何晓群、金勇进：《统计学》，4 页，北京，中国人民大学出版社，2007。

当今世界正进入大数据的时代，全球知名咨询公司麦肯锡称：“数据，已经渗透到当今每一个行业和业务职能领域。”随着信息技术，特别是数据库技术的发展，社会各行业和领域收集、存储数据的能力有了很大的提高，同时积累了大量的数据，这为统计自身理论的进一步发展和在社会各领域的更广泛应用提供了机会。在知识经济和信息时代，统计