

21 世纪通识教育系列教材

21st Century Textbooks of General Education

统计学基础 (第二版)

Fundamental Statistics

贾俊平 © 编著

 中国人民大学出版社

21 世纪通识教育系列教材
21st Century Textbooks of General Education

统计学基础 (第二版)

Fundamental Statistics

贾俊平 © 编著

中国人民大学出版社

图书在版编目 (CIP) 数据

统计学基础/贾俊平编著. —2 版. —北京: 中国人民大学出版社, 2014. 3
21 世纪通识教育系列教材
ISBN 978-7-300-18787-7

I. ①统… II. ①贾… 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 (2014) 第 028991 号

21 世纪通识教育系列教材

统计学基础 (第二版)

贾俊平 编著

Tongjixue Jichu

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	三河市汇鑫印务有限公司	版 次	2010 年 4 月第 1 版
规 格	170 mm×240 mm 16 开本		2014 年 3 月第 2 版
印 张	20 插页 1	印 次	2014 年 3 月第 1 次印刷
字 数	338 000	定 价	36.00 元

版权所有 侵权必究 印装差错 负责调换

· 出版说明 ·

随着信息时代的来临、经济全球化的深入与文化软实力竞争的加剧，重视大学生人文素养与创新能力的培养，提升大学生的综合素质，已成为各国教育改革与发展关注的重点和热点。人们越来越意识到：高等教育不仅要培养大学生良好的专业素质，更重要的是使得他们在走向社会之后拥有长足的自我拓展能力。只有以宽口径、厚基础、复合型为人才培养目标，才能更好地提高我国高等教育的质量，培育出适应现代社会需求的具备公民意识、社会责任感与创新精神的优秀人才。

从中外大学通识教育的实践来看，通识教育是一项系统工程，而课程体系建设始终是推进通识教育的核心任务，教材建设则又是其中的重要环节。为满足广大高校师生对高质量通识教育教材的需求，中国人民大学出版社组织多学科、多领域的专家学者，在广泛调研与深入研讨的基础上，组织编写了这套“21世纪通识教育系列教材”，为推动高等学校通识教育教材建设进行了努力和探索。

本套教材分为人文、政法和经管三大板块，定位为非专业统开课教材，突出“通识”的特色，强调内容阐释的“基础”和“宽度”，力求突破单纯的“专业视域”或“知识视域”，引导学生调整知识结构，拓宽文化视野，以达成人才培养效果上的“宽度”，从而实现高等教育培养复合型人才的目标。

本套教材中的每一本均由该学科领域有影响力的专家学者领衔编写。通识教材的“基础”与“宽度”，需要特别重视教材纲目与内容的适用性、可拓展性和灵活性。唯有在该领域具有丰富教学经验及精深学术水准的名家，方能“取精用弘，由博返约”，编写出体现“通识”特色的高水平教材。

本套教材形式与内容和谐统一，教材内容基础适用，语言简洁生动，并辅以典型、有趣的案例、图表，轻松活泼的栏目和插图等，图文并茂，引人入胜，照顾到青年学生群体的阅读习惯。

作为出版者，我们特别希望通过加强通识教育教材建设，推进高校课程体系的融会贯通，提高学生跨学科、跨文化的理解能力，为学生未来的职业生涯与人生发展奠定良好的知识和能力基础。这套通识教育系列教材只是开始，期望更多的专家学者共襄此事，推进通识教育教学的改革与发展。

中国人民大学出版社

· 前 言 ·

本书是在 21 世纪通识教育系列教材《统计学基础》(第一版)的基础上修订而成的。考虑到读者的特点和实际学习需要,本次修订保留了第一版的基本内容和风格,对时间序列分析和预测一章进行了重新编写,增加了多阶曲线等内容,并对例题和练习题数据做了全面更新。全书各章后思考与练习部分均增加了选择题,并给出了选择题和练习题的答案,供读者学习时参考。

全书共包括 8 章内容。第 1 章介绍统计学的一些基本问题,目的是让读者对统计学有一个基本的了解。第 2 章介绍数据的图表展示方法,使读者学会数据的基本处理方法和图形展示技术。第 3 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 4 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 5 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 6 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 7 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 8 章介绍指数,主要讨论综合指数的编制方法及实践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,建议有能力的读者可尝试使用专业的统计软件(如 SPSS 等)作一些简单的统计分析。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
中国人民大学

· 目 录 ·

第 1 章 统计和统计数据

1.1 统计及其应用领域	2
1.1.1 什么是统计学	2
1.1.2 统计的应用领域	3
1.2 数据的类型	5
1.2.1 分类数据、顺序数据、数值型数据	5
1.2.2 观测数据和实验数据	6
1.2.3 截面数据和时间序列数据	6
1.3 数据来源	7
1.3.1 数据的间接来源	7
1.3.2 数据的直接来源	8
1.3.3 调查方案设计	12
1.3.4 数据质量	13
1.4 统计中的几个基本概念	14
1.4.1 总体和样本	14
1.4.2 参数和统计量	16
1.4.3 变量	17
关键术语	18
思考与练习	18

第 2 章 数据的图表展示

2.1 数据的预处理	24
2.1.1 数据审核	25
2.1.2 数据排序	25
2.1.3 数据筛选	26

2.1.4 数据透视表	29
2.2 品质数据的整理与展示	33
2.2.1 分类数据的整理与图示	33
2.2.2 顺序数据的整理与图示	39
2.3 数值型数据的整理与展示	41
2.3.1 数据分组	41
2.3.2 数值型数据的图示	45
2.4 合理使用图表	57
2.4.1 鉴别图形优劣的准则	59
2.4.2 统计表的设计	59
关键术语	61
思考与练习	61

第3章 数据的概括性度量

3.1 集中趋势的度量	83
3.1.1 平均数	83
3.1.2 中位数和分位数	85
3.1.3 各度量值的比较	87
3.2 离散程度的度量	87
3.2.1 极差和四分位差	88
3.2.2 平均差	88
3.2.3 方差和标准差	90
3.2.4 离散系数: 比较几组数据的离散程度 ...	94
3.3 偏态与峰态的度量	96
3.3.1 偏态及其测度	96
3.3.2 峰态及其测度	97
关键术语	100
思考与练习	101

第4章 抽样与参数估计

4.1 抽样与抽样分布	115
4.1.1 概率抽样方法	115
4.1.2 抽样分布	118

4.2 参数估计的基本原理	124
4.2.1 估计量与估计值	124
4.2.2 点估计与区间估计	125
4.3 总体均值的区间估计	129
4.3.1 正态总体、方差已知, 或非正态总体、 大样本	129
4.3.2 正态总体、方差未知、小样本	132
4.4 总体比例的区间估计	134
4.5 样本量的确定	135
4.5.1 估计总体均值时样本量的确定	136
4.5.2 估计总体比例时样本量的确定	137
关键术语	138
思考与练习	138

第5章 假设检验

5.1 假设检验的基本原理	154
5.1.1 假设的陈述	154
5.1.2 两类错误与显著性水平	158
5.1.3 检验统计量与拒绝域	160
5.1.4 利用 P 值进行决策	162
5.2 总体均值的检验	167
5.2.1 大样本的检验方法	168
5.2.2 小样本的检验方法	174
5.3 总体比例的检验	177
关键术语	179
思考与练习	180

第6章 相关与回归分析

6.1 变量间关系的度量	191
6.1.1 变量间的关系	191
6.1.2 相关关系的描述与测度	193
6.1.3 相关关系的显著性检验	198

6.2 一元线性回归	200
6.2.1 一元线性回归模型	200
6.2.2 参数的最小二乘估计	202
6.2.3 回归直线的拟合优度	206
6.2.4 显著性检验	210
6.2.5 利用回归方程进行估计和预测	213
关键术语	217
思考与练习	217

第7章 时间序列分析和预测

7.1 时间序列及其分解	238
7.2 时间序列的描述性分析	241
7.2.1 图形描述	241
7.2.2 增长率分析	242
7.3 时间序列预测的程序	245
7.3.1 确定时间序列的成分	245
7.3.2 预测方法的选择与评估	247
7.4 平稳序列的预测	249
7.4.1 移动平均法	249
7.4.2 指数平滑法	252
7.5 趋势型序列的预测	254
7.5.1 线性趋势预测	254
7.5.2 非线性趋势预测	256
7.6 复合型序列的分解预测	260
7.6.1 确定并分离季节成分	260
7.6.2 建立预测模型并进行预测	263
7.6.3 计算最后的预测值	265
关键术语	266
思考与练习	266

第8章 指数

8.1 引言	281
--------------	-----

8.2 加权指数	282
8.2.1 加权综合指数	282
8.2.2 加权平均指数	284
8.2.3 价值指数与指数体系	285
8.3 几种常用的价格指数	287
8.3.1 零售价格指数	287
8.3.2 消费者价格指数	288
8.3.3 生产者价格指数	290
8.3.4 股票价格指数	290
关键术语	291
思考与练习	291
附录 用 Excel 生成概率分布表	298
参考文献	306

第 1 章 统计和统计数据

内容提要

本章主要介绍统计中的一些基本问题，目的是使读者对统计学有一个基本的了解和认识。内容包括：

- (1) 统计及其应用领域。介绍统计学的含义以及统计的应用领域。
- (2) 统计数据及其类型。介绍统计数据及其类型以及统计数据与统计方法之间的关系。
- (3) 统计数据的来源。介绍二手数据的来源、原始数据的调查和收集方法、调查方案的设计以及统计数据的误差等问题。
- (4) 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。

你相信这样的一些统计结论吗？

每天你都会看到各种统计数字，但你或许没有仔细想过它们意味着什么。看到下面这些统计研究结果，你会有怎样的看法呢？

● 中国科学院空间环境研究预报中心的专家称，在神舟七号载人航天飞行期间，飞船遭遇空间碎片的概率在百万分之一以下。

● 吸烟对健康是有害的，吸香烟的男性减少寿命 2 250 天。

● 不结婚的男性会减少寿命 3 500 天，不结婚的女性会减少寿命 1 600 天。

● 身体超重 30% 会使寿命减少 1 300 天。

● 每天摄取 500 毫升维生素 C，生命可延长 6 年。

● 身材高的父亲，其子女的身材也较高。

● 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。

● 学生们在听了莫扎特钢琴曲 10 分钟后做的推理测试会比他们听 10 分钟娱乐磁带或其他曲目后做的更好。

● 上课坐在前面的学生平均考试分数比坐在后面的学生高。

看懂这些结论并不困难,但这些结论是怎样得出来的?你相信这些结论吗?学点儿统计学知识有助于你正确理解它们。

1.1 统计及其应用领域

在日常生活中,我们经常会接触到“统计”这一术语,在有关媒体中也经常会看到一些报道使用统计数据、图表等。本章将介绍统计学的一些基本问题,包括统计学的含义、统计数据及其分类、统计中常用的基本概念等。

1.1.1 什么是统计学

统计是处理数据的一门科学。统计学家给统计学下的定义很多,如统计是收集、分析、表述和解释数据的科学;统计是一门收集、分析、解释和提供数据的科学;统计指的是一组方法,用来设计实验、获得数据,然后在这些数据的基础上组织、概括、演示、分析、解释和得出结论。总结上述定义,可以将统计学的含义概括如下:

► **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。

统计学的定义告诉我们,统计是用来处理数据的,统计学是关于数据的科学,它提供了一套有关数据收集、数据处理、数据分析、数据解释并从数据中得出结论的方法。统计研究的是来自各领域的数据。

统计提供的数据分析方法可分为描述统计和推断统计两大类。

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

描述统计的内容包括取得研究所需要的数据、用图表形式对数据进行处理和显示,进而通过综合、概括与分析,得出反映所研究现象的一般性特征。

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

推断统计是研究如何利用样本数据来推断总体特征的统计学方法,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。

1.1.2 统计的应用领域

统计方法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。目前,随着定量研究的日趋重要,统计方法已应用到自然科学和社会科学的众多领域。同时,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。比如,政府部门、学术研究领域、日常生活、公司或企业的生产经营管理中都要用到统计。下面给出统计在工商管理中的一些应用。

(1) 在企业发展战略中的应用。企业发展战略是一个企业的长远发展方向。制定发展战略一方面需要及时了解和把握整个宏观经济状况及发展变化趋势,以便了解市场的变化;另一方面,还要对企业进行合理的市场定位,以便把握企业自身的优势和劣势。所有这些方面都离不开统计,它们需要统计提供可靠的数据、利用统计方法对数据进行科学的分析和预测,等等。

(2) 在产品质量管理中的应用。质量是企业的生命,是企业持续发展的基础。我们在质量管理中同样离不开统计的应用。在一些知名的跨国公司中,六西格玛准则已成为一种重要的管理理念。由此可知,质量控制已成为统计学在生产领域中的一项重要应用。此外,各种统计质量控制图还被广泛用于监测生产过程。

(3) 在市场研究中的应用。企业要在激烈的市场竞争中取得优势,首先必须了解市场;要了解市场,企业就需要作广泛的市场调查,以取得所需的信息,然后对这些信息进行科学的分析,并以此作为生产和营销的依据。这些措施都需要统计的支持。

(4) 在财务分析中的应用。上市公司的财务数据是股民进行投资选择的重要参考依据。投资咨询公司主要是根据上市公司提供的财务和统计数据进行分析,为股民提供投资参考。企业自身的投资也离不开对财务数据的分析,其中就要用到大量的统计方法。

(5) 在经济预测中的应用。企业要对未来的市场状况进行预测,经济学家也常常对宏观经济或某一方面进行预测,他们在进行预测时就要使用各种统计信息和统计方法。比如,企业要对产品的市场潜力作出预测,以便及时调整生产计划,此时就需要利用市场调查取得数据,并对数据进行统计分析。

经济学家在预测通货膨胀时,要利用有关生产价格指数、失业率、生产能力利用等统计数据,运用统计模型进行预测。

(6) 在人力资源管理中的应用。利用统计方法对企业员工的年龄、性别、受教育程度、工资等进行分析,并以此作为企业制定工资计划、奖惩制度的依据。

当然,统计并不仅仅是为了管理,它是为自然科学、社会科学等多个领域发展起来的,并为多个学科提供了一种通用的数据分析方法。从某种意义上说,统计仅仅是一种数据分析的方法,与数学一样,统计是一种工具,它是一种数据分析的工具。

统计的主要作用是帮助我们分析数据。比如,可以利用统计简化繁杂的数据,用图表重新展示数据,再建立数据模型,并进行比较分析,等等。对于一名工商管理人员来说,他会面临企业经营管理的大量数据。这些数据对该管理人员进行管理与决策起到了什么样的作用?一个决策者可能更喜欢看文字性的材料,或者某种结论性的东西,他看到数据就会头痛。也就是说,他“害怕”数字。殊不知,结论性的东西就是来自对数据的分析。

有人认为,统计的全部目的就是让人们看懂数据。实际上,这仅仅是统计的一个方面,统计更重要的功能是数据分析,统计提供了一套分析数据的方法和工具。不同的人对数据分析的理解也会不一样,曲解数据分析是一种常见的现象。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结论,然后再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,我们需要对数据进行分析才能得出结论。

当然,统计不是万能的,它不能解决我们所面临的一切问题。统计能帮助我们进行数据分析,并从分析中得出某种结论,但对统计结论的进一步解释,则需要分析人员具备专业知识。统计不能为管理人员提供所需的一切技巧和方法。当我们把统计用到管理领域时就会发现,统计和管理之间还有差距。许多学过统计的管理人员总觉得还缺少点什么东西,好像统计没有从根本上解决他们的问题。实际上,大部分统计方法都需要某些假设的前提。比如,在建立一个线性回归模型时,首先要假设变量之间的关系是线性的,否则就无从下手。但是,这种假设必须经过严格的检验后,才能应用所建立的模型。作为一个管理人员,他更关心真正有用的信息是什么,结论是什么。

应采取什么样的行动，他并不那么关心严格的统计检验。当然，该管理人员也会检验信息的正确性，但他不一定是通过统计检验，而是通过某种其他方式，如定性的方法，或者干脆就凭直觉。然而，自然科学强调应在不同条件下对非统计方法推导出的假设进行严格的检验。显然，统计在自然科学和管理科学中的作用是不同的。管理人员所需要的功能，统计未必能提供。如果希望在没有外界帮助的条件下找出数据的特征或规律，那么统计对此无能为力。统计是一套分析数据的方法和工具，不能指望统计成为解决所有问题的灵丹妙药。

1.2 数据的类型

统计数据是对现象进行测量的结果。比如，对经济活动总量进行测量可以得到国内生产总值（GDP）数据，对股票价格变动水平进行测量可以得到股票价格指数数据，对人口性别进行测量可以得到男性或女性的相关数据，等等。由于使用的测量尺度不同，统计数据可以分为不同的类型。下面从不同角度说明统计数据的分类。

1.2.1 分类数据、顺序数据、数值型数据

按照所采用的不同计量尺度，我们可以将统计数据分为分类数据、顺序数据和数值型数据。

► **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，该数据表现为类别，是用文字来表述的。分类数据主要是由分类尺度计量形成的。例如，按照性别，人口可分为男性、女性两类；按行业属性，企业可分为医药企业、家电企业、纺织品企业等，这些均属于分类数据。为便于统计处理，我们对于分类数据可以用数字代码来表示各个类别。比如，用“1”表示“男性”，“0”表示“女性”；用“1”表示“医药企业”，“2”表示“家电企业”，“3”表示纺织品企业，等等。

► **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，它是由顺序尺度计量形成的。比如，产品可分为一等品、二等品、三等品、次品等；

考试成绩可以分为优、良、中、及格、不及格等；一个人的受教育程度可以分为小学、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。同样，顺序数据也可以用数字代码来表示。比如，“1”表示“非常同意”，“2”表示“同意”，“3”表示“保持中立”，“4”表示“不同意”，“5”表示“非常不同意”。

► **定义 1.6** 按数字尺度测量的观察值，称为数值型数据 (metric data)。

数值型数据是使用自然或度量衡单位对事物进行测量的结果，其结果表现为具体的数值。在现实中，我们所处理的大多数数据都是数值型数据。

分类数据和顺序数据说明的是事物的品质特征，通常是用文字来表述的，其结果均表现为类别，因而也可统称为定性数据或品质数据 (qualitative data)；数值型数据说明的是现象的数量特征，通常是用数值来表现的，因此也可称为定量数据或数量数据 (quantitative data)。

1.2.2 观测数据和实验数据

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。

► **定义 1.7** 通过调查或观测而收集到的数据，称为观测数据 (observational data)。

观测数据是在没有对事物进行人为控制的条件下得到的，有关社会经济现象的统计数据几乎都是观测数据。

► **定义 1.8** 在实验中控制实验对象而收集到的数据，称为实验数据 (experimental data)。

比如，对一种新药疗效的实验数据，对一种农作物新品种的实验数据。自然科学领域的大多数数据都是实验数据。

1.2.3 截面数据和时间序列数据

按照所描述的现象与时间的关系，可以将统计数据分为截面数据和时间序列数据。

► **定义 1.9** 在相同或近似相同的时间点上收集的数据，称为截面数据 (cross-sectional data)。

截面数据所描述的是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。比如，2010 年我国各地区的国内生产总值数据就是截面数据。