



 **MBA** 精品课程系列教材

应用统计学

APPLIED STATISTICS

主编 贾俊平 谭英平

主审 吴喜之

 中国人民大学出版社



MBA 精品课程系列教材

应用统计学

APPLIED STATISTICS

主编 贾俊平 谭英平

主审 吴喜之

中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

应用统计学/贾俊平, 谭英平主编
北京: 中国人民大学出版社, 2008
(MBA 精品课程系列教材)
ISBN 978-7-300-09718-3

- I. 应…
- II. ①贾…②谭
- III. 应用统计学-研究生-教材
- IV. C8

中国版本图书馆 CIP 数据核字 (2008) 第 141483 号

MBA 精品课程系列教材

应用统计学

主编 贾俊平 谭英平
主审 吴喜之

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 河北涿州星河印刷有限公司

规 格 185 mm×235 mm 16 开本

版 次 2008 年 10 月第 1 版

印 张 17 插页 1

印 次 2008 年 10 月第 1 次印刷

字 数 290 000

定 价 35.00 元

版权所有 侵权必究 印装差错 负责调换

总序

随着经济全球化进程的不断深入，以 MBA 教育为代表的中国管理教育日益受到关注，而 MBA 教育经过 13 年的风雨历程，也取得了长足的发展，引发了人们更加深刻的思考：如何进一步提升中国 MBA 教育的教学质量？

首先，经济全球化对国际化管理人才的素质提出了更高的要求。经济的全球化带来的产业升级和企业的国际化战略将是未来企业求得生存与发展的关键，越来越复杂的经济环境使跨国公司对具有全球性眼光的国际型人才需求日趋增多。经济全球化时代要求企业经营者彻底转变经营思想和经营理念，具备竞争观念、创新意识和开拓精神，具备全球性的敏锐洞察力，通晓多国语言，有雄厚的技术功底和创业精神，有能力管理日益细分和分散的组织。

其次，随着中国市场经济的日益完善，数以万计的民营企业迫切需要自己的企业家和职业经理人，如何使 MBA 教育本土化更是一个迫在眉睫的问题。如果说前 13 年 MBA 教育的发展更多走的是拿来主义的道路，那么现在更需要探索一条本土化的道路。

最后，随着中国企业改革的深化，专业 MBA 教育的时代来临了。现在不仅出现了 MPA、MPAcc 等专业硕士学位，在 MBA 中也向专业细分方向发展；不仅出现了市场营销方向、人力资源管理等专业方向的细分课程，而且出现了涉农管理 MBA、航空管理 MBA、电力资源管理 MBA 等。正是这种专业方向的 MBA 教育，为 MBA 教育提供了更加广阔的市场。

在这样的市场需求下，编写一套国际化与本土化相结合、更加体现专业和行业细分的教材势在必行。中国人民大学出版社在全国 MBA 专业指导委员会的指导下，组织国内外一流专家，编写了“MBA 精品课程系列教材”。这套教材将很好地适应 MBA 教育的发展，也使 MBA 的毕业生更好地适应用人单位的需要，最终提升中国企业的管理水平，促进中国经济的进一步发展。

前 言

统计研究的是数据，提供的是一套通用于所有学科领域的获取数据、分析数据并从数据中得出结论的原则和方法。作为一门应用性很强的学科，多数人的学习目的也主要是应用。但初学者学习统计时面临的主要困惑是学完不会用。问题在于学习过程中多把注意力集中在公式和计算上，而忽视对统计思想的理解。学习统计关键在于理解。记住公式，不等于学会统计；学会计算，不等于会用统计。统计的真谛在于它所体现的思想，在于它所提供的思维方式。统计学家 C. R. 劳教授 (C. R. Rao) 在《统计与真理——怎样运用偶然性》一书中写道：“统计学与其说是收集整理数据引出答案的一组规则，不如说是一种思考或推理的方法”。学好统计的关键是掌握如何运用统计思维来思考问题，而不是简单地记住那些死的统计知识。

初学者对统计课程往往感到畏惧，被书中的公式所吓倒。实际上，抛开公式照样可以学会统计。特别是在计算机应用已经普及的今天，所有的计算都可以由计算机来完成。只要清楚统计方法使用的前提，理解统计方法的实质，要应用统计并不困难。

接到本书的编写任务，总有些犹豫，不知道一本什么样的教材适合于刚刚接触统计学课程的读者。本书在写法上做了一些新的尝试：力图把统计方法的思想用书中标题的形式表达出来，尽管这种表达不一定确切；每种方法都力求用实际问题引出，尽量不使用更专业的统计术语；方法中所用到的公式基本上都放在每章后的附录里；书中例题的解答基本上未给出过程，而是直接使用计算机的输出结果。考虑到读者使用的方便程度，本书使用 Excel 和 SPSS 两个软件。书中所涉及的内容，Excel 能够解决的全部使用 Excel 给出结果，不能解决的（如多元统计分析）则使用 SPSS。对两个软件使用的一些说明放在每章后的附录里。

本书的第 1、2、3、4、5、6、9 章由贾俊平编写，第 7、8、10 章由谭英平编写。十分有幸请到吴喜之教授作为本书的主审，他不仅为本书写作的思路提供了极具建设性的意见，同时仔细审阅了书稿的全部内容，并提出了诸多修改意见。在吴教授的指导下，作者完成了书稿的写作和修改，这无疑使本书增色颇多。十分感谢吴教授对本书的指导和

审阅。

感谢中国人民大学出版社的陈永凤编辑，她对本书的认真校对和颇有效率的工作，才使得本书尽快与读者见面。

贾俊平

2008年5月于中国人民大学

目 录

第 1 章

统计能为你做些什么?	(1)
1.1 统计无处不在	(1)
1.1.1 每个人都离不开统计	(1)
1.1.2 几乎所有的领域都要用统计	(2)
1.2 统计研究什么?	(3)
1.3 统计能做什么和不能做什么?	(5)
1.3.1 统计帮助你分析数据	(5)
1.3.2 统计不能解决你的所有问题	(6)
1.3.3 统计的误用与滥用	(7)
1.4 怎样获得数据?	(8)
1.4.1 变量与数据	(8)
1.4.2 怎样得到一个样本?	(8)
本章附录	(9)
习题	(10)

第 2 章

用图表和统计量看数据	(12)
2.1 用图表描述数据	(12)
2.1.1 用图表展示定性数据	(12)
2.1.2 用图表展示定量数据	(16)
2.2 用统计量描述数据	(23)
2.2.1 用一个值概括一组数据	(23)

2.2.2 找出数据彼此之间的差别	(25)
2.2.3 数据分布的形状	(29)
本章附录	(29)
本章主要公式	(33)
习题	(34)

第3章

用概率分布描述随机变量	(39)
3.1 度量事件发生的可能性	(39)
3.2 随机变量的概率分布	(40)
3.2.1 随机变量及其概括性度量	(40)
3.2.2 离散型概率分布	(42)
3.2.3 连续型概率分布	(44)
3.3 由正态分布导出的几个重要分布	(48)
3.3.1 χ^2 分布	(48)
3.3.2 t 分布	(49)
3.3.3 F 分布	(50)
3.4 样本统计量的抽样分布	(51)
3.4.1 样本均值的抽样分布	(51)
3.4.2 其他统计量的抽样分布	(52)
3.4.3 统计量的标准误差	(53)
本章附录	(53)
本章主要公式	(55)
习题	(56)

第4章

用样本推断总体	(58)
4.1 怎样进行推断?	(58)
4.1.1 用估计量估计总体参数	(58)
4.1.2 用什么方法进行估计?	(59)
4.1.3 用什么样的估计量去估计?	(60)
4.2 估计总体参数	(61)
4.2.1 总体均值的估计	(61)

4.2.2	总体比例的估计	(64)
4.2.3	总体方差的估计	(65)
4.3	检验总体假设	(65)
4.3.1	怎样提出假设?	(66)
4.3.2	依据什么作出决策?	(67)
4.3.3	总体均值的检验	(71)
4.3.4	总体比例的检验	(75)
4.3.5	总体方差的检验	(76)
	本章附录	(78)
	本章主要公式	(80)
	习题	(82)

第 5 章

分类变量对数值变量的影响 (88)

5.1	方差分析解决什么问题?	(88)
5.1.1	比较均值是否相同	(88)
5.1.2	从误差分析入手	(89)
5.1.3	在什么样的前提下分析?	(91)
5.2	考虑一个分类变量的影响	(91)
5.2.1	只考虑一个因子	(91)
5.2.2	关系有多强?	(92)
5.3	考虑两个分类变量的影响	(93)
5.3.1	不考虑交互作用	(93)
5.3.2	考虑交互作用	(95)
	本章附录	(95)
	本章主要公式	(96)
	习题	(97)

第 6 章

利用变量间的关系进行预测 (102)

6.1	变量之间有什么样的关系?	(102)
6.1.1	用散点图描述变量间的关系	(102)
6.1.2	用相关系数度量关系强度	(104)

6.1.3	在总体中也存在这样的关系吗?	(105)
6.2	建立变量之间的数学表达式	(106)
6.2.1	涉及一个自变量的线性回归	(106)
6.2.2	涉及多个自变量的线性回归	(107)
6.3	拟合效果的度量和回归检验	(109)
6.3.1	回归方程拟合得好吗?	(109)
6.3.2	因变量与自变量之间有线性关系吗?	(111)
6.4	所有自变量都有必要放进模型中吗?	(112)
6.4.1	自变量之间相关对模型有什么影响?	(112)
6.4.2	剔除不必要的自变量	(113)
6.4.3	模型有多好?	(115)
6.5	用自变量预测因变量	(116)
6.6	含有定性自变量的回归	(116)
	本章附录	(118)
	本章主要公式	(119)
	习题	(120)

第7章

用少数变量代表多个变量 (128)

7.1	主成分分析	(128)
7.1.1	主成分分析的基本思想是什么?	(128)
7.1.2	如何选择主成分?	(131)
7.1.3	怎样解释主成分?	(133)
7.2	因子分析	(135)
7.2.1	什么是因子分析?	(135)
7.2.2	怎样解释因子分析结果?	(136)
7.2.3	几点说明	(139)
	本章附录	(140)
	本章主要公式	(141)
	习题	(141)

第8章

把对象分成不同的类别 (144)

8.1	聚类分析	(144)
8.1.1	按什么来聚类?	(145)

8.1.2	怎样度量距离远近?	(146)
8.1.3	分层聚类法	(147)
8.1.4	K-均值聚类法	(149)
8.1.5	几点说明	(151)
8.2	判别分析	(152)
8.2.1	判别分析有什么不同?	(152)
8.2.2	距离判别法	(153)
8.2.3	Fisher 判别法	(154)
8.2.4	逐步判别法	(154)
	本章附录	(159)
	本章主要公式	(160)
	习题	(162)

第 9 章

	根据过去的模式预测未来	(166)
--	-------------------	-------

9.1	时间序列的组成要素	(166)
9.2	时间序列预测的程序	(168)
9.2.1	确定时间序列的成分	(168)
9.2.2	选择预测方法并进行评估	(171)
9.3	平滑法预测	(171)
9.4	趋势模型预测	(174)
9.4.1	线性趋势预测	(174)
9.4.2	非线性趋势预测	(175)
9.4.3	自回归模型预测	(181)
9.5	多成分序列的预测	(183)
9.5.1	季节性多元回归预测	(183)
9.5.2	分解预测	(185)
	本章附录	(187)
	本章主要公式	(188)
	习题	(189)

第 10 章

	不依赖于分布的检验	(193)
--	-----------------	-------

10.1	关于非参数检验	(193)
10.1.1	什么时候选择非参数检验?	(193)

10.1.2	预备知识	(194)
10.2	单样本的非参数检验	(196)
10.2.1	符号检验	(196)
10.2.2	Wilcoxon 符号秩检验	(198)
10.3	两样本的非参数检验	(200)
10.3.1	两独立样本的 Wilcoxon (Mann-Whitney) 秩和检验	(200)
10.3.2	匹配样本的中位数检验	(202)
10.3.3	Spearman 秩相关检验	(204)
	习题	(207)
附录		
	各章练习题答案	(212)
	参考书目	(258)

统计能为你做些什么？

统计思维总有一天会像读与写一样成为一个有效率公民的必备能力。

——H. G. 韦尔斯 (H. G. Wells)

1.1 统计无处不在

1.1.1 每个人都离不开统计

了解一些统计学知识对每个人都是必要的。比如，在外出旅游时，你需要关注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的信息，了解某只特定股票的有关财务信息；在观看足球比赛时，除了关心进球数的多少外，你还要知道各支球队的技术统计，等等。要正确阅读并理解下面的统计研究得出的一些结论，就更需要具备一些统计知识。

- 吸烟对健康是有害的。
- 不结婚的男性会早逝 10 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，依此类推。
- 两天服一片阿司匹林会减少心脏病第二次发作的机会。
- 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。
- 怕老婆的丈夫得心脏病的机会较大。
- 学生们在听了莫扎特钢琴曲 10 分钟后所做的推理测试会比他们听 10 分钟娱乐磁带或其他曲目做得更好。

- 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？学点统计知识你就会正确理解它们。

了解一些统计知识，对政策的制定者或企业的管理者来说就更为重要，这不仅有助于他们的决策，还可避免因为不懂统计而弄出笑话来。一个统计办公室的主管与一些统计学者开会，统计学者抱怨从其他部门收到的一些估计值没有给出标准误差（估计时的误差大小，表示估计的精度）。这个主管马上问道：“对误差也有标准吗？”健康部门的一位官员看到一个统计学者提供的报告，报告中提到上一年由于某种疾病，平均 1 000 人中死亡人数为 3.2 人，这位官员对这个数字发生了兴趣。他问他的私人秘书，3.2 个人是如何死法？他的秘书说：“先生，当一个统计学家说死了 3.2 个人时，意味着 3 个人已经死了，两个人正要死”。

有点统计知识就不会闹出这样的笑话来。一位学者写道：“假定你是市场部的新任经理，一次广告活动的统计结果摆到了你面前，声称某个结果是‘统计显著’的。你如何解释这份报告而不是暴露你对该术语的无知？赶快学点统计，这对你和你的事业都非常有用。”^①

1.1.2 几乎所有的领域都要用统计

让你说出哪些领域不使用统计，这很困难，因为几乎找不到一个不用统计的领域；让你说出哪些领域应用统计，同样也很困难，因为几乎所有的领域都应用统计。正如有的学者所说：“举出一个统计没有用途的领域名称比举出一个统计作为其组成部分的领域名称要困难得多”^②。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)、詹姆斯·麦迪逊 (James Madison) 为了说服纽约人认可宪法，匿名发表了著名的 85 篇论文。这些论文中的大多数作者已经得到了确认，但是，其中 12 篇论文的作者身份引起了争议。通过对不同单词的频数进行统计分析，得出的结论是詹姆斯·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·麦迪逊是原创作者的说法占主导地位，而且几乎可以肯定这种说法是正确的。

① [美] G.R. 埃维森、[美] M. 格根著，吴喜之等译：《统计学——基本概念和方法》，2 页，北京，高等教育出版社；海德堡：施普林格出版社，2000。

② [美] 罗伯特·约翰逊、[美] 帕特里夏·库贝：《基础统计学》，11 页，北京，科学出版社，2003。

【例 1.2】 用统计进行产品质量管理。统计在企业产品质量管理中的应用是统计应用的一个重要方面。在统计中， σ 表示一个总体的标准差，它表示的是数据之间的差异程度。比如，在企业生产的产品中，同一种产品没有两个是完全一样的，因为在生产过程中，由于各种因素的影响而使产品质量产生波动。产品的这种差异称为质量的波动性，也正是由于波动性的存在才需要进行质量管理。 6σ 是质量管理中使用的一个术语，它的含义是指偏离正态分布的中心 6 个标准差。它表示在生产过程中缺陷率不超过百万分之三点四，通俗地说，如果生产 100 万个产品，不合格率平均来说不超过 3.4 个。这样的不合格率非常低，以至于可以忽略不计。 6σ 质量管理已成为最新的质量管理理念，近年来，它已成为一些著名国际大企业的质量管理方法，并在这样的管理理念下使企业受益匪浅。例如，实行了 6σ 质量标准，使摩托罗拉公司在 3 年中节省的资金超过 9.4 亿美元。实行 6σ 管理的大公司还有美国通用电气公司 (GE)、宝利来 (Polaroid) 和得州仪器 (Texas Instruments) 等。GE 的前 CEO 杰克·韦尔奇 1999 年 4 月曾说过这样一段话：“ 6σ 培训计划是 GE 下一个世纪领导层得以产生繁衍的园地， 6σ 是我们曾经尝试过的最重要的管理培训方法，它胜过到哈佛商学院就读，也胜过到克顿维尔 (克顿维尔是 GE 公司内部的质量培训部) 进修，它教会你一种与众不同的思维方式。”在推广 6σ 质量管理不到 10 年的时间内，通用电气公司的总市值从世界排名第十位跃升到第二位。

【例 1.3】 用简单的描述量得到一个重要发现。费希尔 (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米特 (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，鳗鱼的脊椎骨数量变化不大。施米特在从欧洲各地、冰岛、亚速尔群岛以及尼罗河等几乎分离的海域里所捕获的鳗鱼样本中，计算发现了几乎一样的均值和标准偏差值。由此，施米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。

这些例子表明统计不仅在许多领域都有着广泛应用，而且在生产、生活、科学研究等各个领域都发挥着日趋重要的作用。

1.2 统计研究什么?

如果你问身边的人，GDP (国内生产总值) 是什么，CIP (消费者价格指数)

又是什么，大家似乎都能说上几句。但要是仔细追问它们究竟代表了什么，就不是每个人都能够说清楚的。统计也是一样。你要问一个人统计是什么，似乎没有人不知道，但你要追问统计究竟是什么，就不是一两句话能够说明白的，要搞清楚统计研究什么就更困难了。

物理学研究的是像热、光、电这类自然现象的运动规律；化学家测定物质的组成及化学元素之间的相互作用；生物学家研究动植物的习性；数学家则在给出的假定之下推演各种命题。这些学科中的每一门都有它自己的问题，而且有解决这些问题的各自的方法，各学科因此而成为一门单独的学科。

统计学（statistics）是一门独立的学科，这似乎没人怀疑。但统计究竟研究什么，可能就有不同的看法。有人认为，统计学是一门独特的学问，没有任何固定的对象。乍听起来似乎难以理解，但仔细想想也许有道理。统计学研究的是来自各领域的的数据，靠解决其他领域的问题而存在和发展。按 L. J. 萨维奇（L. J. Savage）的说法：“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，这是因为对很多寄主来说，如果没有寄生虫就会死。对有的动物来说，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。”^① 这看上去统计似乎被边缘化了，实际上这也正说明了统计在各学科领域的独特地位和作用，也表明了统计作为一门独立学科的存在而具有的特点。

按 C. R. 劳的说法：“今天，统计学已发展成为一门媒介科学，它研究的对象是其他学科的逻辑和方法论——作出决策的逻辑和试验这些决策的逻辑。统计学的未来依赖于向其他学科领域内的研究者正确传授统计学的观点；依赖于如何能够在其他知识领域内将其主要问题模式化。”^② 因此，在他看来，统计学是一门科学、一种工艺和一门艺术这三者的组合。

统计学是一门科学。它提供一套方法和技术，这些方法和技术并不是一成不变的，使用者在给定的情况下必须根据所掌握的专门知识选择使用这些方法，而且，有时还要进行必要的修正。统计方法是通用的数据分析方法，这些方法不是为某个特定的问题领域而构造的。

统计学是一种工艺。如同工业生产过程中的质量控制程序一样，统计方法是在为保证产品达到所希望的质量和保持其稳定性的管理系统中建立起来的。统计方法也能用于控制、减少和考察不确定性。

① [美] C. R. 劳：《统计与真理——怎样运用偶然性》，105页，北京，科学出版社，2004。

② [美] C. R. 劳：《统计与真理——怎样运用偶然性》，41页，北京，科学出版社，2004。

统计学是一门艺术。它提供一种归纳推理的方法，推理就是一种艺术。既然是归纳推理，就不能保证结论百分之百正确，就不能没有争议。怎样让别人看懂并理解统计结论，就要看统计结论表达的技巧和艺术性了。

这些观点听起来有些“哲学”，但它能帮助我们理解统计学是什么，统计学研究什么。统计学研究的是数据，提供的是一套通用于所有学科领域的获取数据、分析数据并从数据中得出结论的原则和方法^①。

1.3 统计能做什么和不能做什么?

1.3.1 统计帮助你分析数据

无论是作为一名管理人员，还是一名研究人员，都会面临大量的数据。这些数据如果不经过分析，也就仅是一堆数据而已，没有什么用处，而有用的信息恰恰藏在数据里面。即使你知道一个地区每个家庭的收入数据，也无法给出这个地区收入状况的一个概括性认识，况且你也难以得到每个家庭的收入数据。抑或许你知道20只灯泡的使用寿命，知道50件产品的合格率，这显然不够，因为你要知道的是这批灯泡的使用寿命、这批产品的合格率。要得到这样的结果，就需要用统计方法去分析。

有人认为统计就是找数据，统计的目的就是让人看懂数据。实际上这仅仅是统计的一个方面，统计更重要的功能是分析数据。统计可以在多个方面帮助你进行数据分析。比如，你可以利用统计简化繁杂的数据，用图表重新展示数据，进行比较分析，建立数据模型进行预测，等等。

不同的人对数据分析的理解也会大不一样，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，在他们的心目中可能早已有了某种“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结论，从数据中寻找启发，而不是寻找支持。真正的

^① 统计方法可分为描述统计和推断统计。描述统计 (descriptive statistics) 研究的是数据收集、处理、汇总、图表描述、概括与分析等统计方法。推断统计 (inferential statistics) 研究如何利用样本数据来推断总体特征的统计方法。比如，要了解一个地区的人口特征，不可能对每个人的特征一一进行测量；对产品的质量进行检验，往往是破坏性的，也不可能对每个产品进行测量。这就需要抽取部分个体即样本进行测量，然后根据获得的样本数据对所研究的总体特征进行推断，这就是推断统计要解决的问题。