高等学校经济统计专业系列教材

统计学（第二版）

费宇　石磊　主编

高等教育出版社

高等学校经济统计专业系列教材

统计学（第二版）

费宇　石磊　主编

高等教育出版社·北京

内容简介

本书介绍统计学的基本理论和方法，主要内容有：统计学基本概念、数据的描述、参数估计、假设检验、方差分析、回归分析、时间序列分析、抽样调查理论和方法、统计指数和统计决策等。

本书结合实例讲解统计学基本理论和方法，并采用统计软件SPSS20.0进行统计计算和分析，可以作为高等院校经济学类和工商管理学类专业本科生的教材，也可以作为从事统计分析的人员和统计工作者的参考书。读者扫描每章后的二维码，可即时检验每章学习效果。

本书配套的教学资源有：教材中的例题、习题和案例的有关数据文件，教学ppt，以及习题和案例的参考答案，具体索取方式参见书后教学支持说明。

图书在版编目（CIP）数据

统计学／费宇，石磊主编. -- 2版. -- 北京：高等教育出版社，2017.1
ISBN 978-7-04-046656-0

Ⅰ. ①统… Ⅱ. ①费… ②石… 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(2016)第262458号

策划编辑 施春花	责任编辑 施春花	封面设计 于文燕	版式设计 杜微言
插图绘制 黄云燕	责任校对 吕红颖	责任印制 尤 静	

出版发行	高等教育出版社	网 址	http://www.hep.edu.cn
社 址	北京市西城区德外大街4号		http://www.hep.com.cn
邮政编码	100120	网上订购	http://www.hepmall.com.cn
印 刷	三河市华润印刷有限公司		http://www.hepmall.com
开 本	787mm×1092mm 1/16		http://www.hepmall.cn
印 张	18	版 次	2010年9月第1版
字 数	440千字		2017年1月第2版
购书热线	010-58581118	印 次	2017年1月第1次印刷
咨询电话	400-810-0598	定 价	36.00元

物料号 46656-00

作者简介

费宇，云南财经大学二级教授，博士生导师，统计学博士，英国曼彻斯特大学博士后，云南省中青年学术技术带头人，云岭教学名师，云南省教学名师，云南省有突出贡献中青年专家。现任中国商业统计学会常务理事，全国经济数学与管理数学学会常务理事，云南省应用统计学会副理事长，云南省统计学会理事。主要从事统计理论与方法、应用统计、数据挖掘和计量经济分析方面的研究。在国内外学术期刊上发表论文40余篇，在科学出版社出版学术专著2部，在高等教育出版社、科学出版社和中国人民大学出版社出版教材4部，获省部级以上奖励10项。

石磊，云南财经大学二级教授，博士生导师，教育部长江学者特聘教授，国家百千万人才工程人选，云南省云岭学者。现任中国统计学会常务理事，中国统计教育学会常务理事，中国数量经济学会常务理事，中国现场统计学会理事，中国现场统计学会生存分析分会副理事长，云南省统计学会副会长，云南省数学学会副理事长，云南省应用统计专业指导委员会常务副主任。主要从事数理统计、应用统计、合作演化理论、计量经济等领域的研究，在国内外统计学期刊发表论文70余篇，出版专著5部。

第二版前言

我们所处的时代是一个数据无处不在的大数据时代，作为数据科学的重要组成部分，统计学在数据分析中扮演了非常重要的角色。统计学是经济类和管理类本科生必修的一门重要基础课程，统计学理论有一定抽象性，涉及的计算比较复杂，很多计算必须借助计算机通过统计软件来完成。所以，统计学是一门大家公认的非常有用但又比较难学的必修课。

本书写作的指导思想是：在不失严谨的前提下，努力突出统计学的应用性特点，统计理论与实际案例相结合阐述统计思想。本书有以下三个主要特点：(1)简明易懂，每章都由一个实际问题(引例)引入本章要介绍的基本内容，结合实例介绍统计理论和方法，方便读者理解统计理论和方法；(2)突出统计软件应用，采用国际通用的统计软件 SPSS20.0 做统计计算和分析，辅助教学，使学生能应用统计软件对实际统计数据进行基本的统计计算和分析；(3)结构合理，侧重介绍统计学的基本理论和方法，减少一些不必要的定理和公式的证明。

本书第一版自 2010 年出版以来，承蒙读者的厚爱，被许多高校作为教材，同时，许多教师和学生给予我们热情的鼓励并对书中有些地方提出了中肯的建议，在此我们表示衷心感谢。

本书第二版由费宇负责修订第 1 章、第 3 章、第 4 章和第 6 章，石磊和鲁筠负责修订第 5 章，雷健敏负责修订第 2 章和第 9 章，孟彦菊和谢佳春负责修订第 7 章，马云玲负责修订第 8 章，王任负责修订第 10 章，最后由费宇和石磊负责全书的统稿。第二版在保持突出应用的风格下，主要做了如下修订：

1. 采用 SPSS20.0 版本代替 SPSS16.0 做统计计算和分析。
2. 更新了原书的很多案例。
3. 重新编写了第 7 章。

在本书的写作过程中，作者参阅了许多国内外的图书资料，并引用了部分例题和习题，在此向有关作者表示感谢。

由于作者水平有限，书中难免有错漏之处，敬请读者批评指正。

费　宇

2016 年 6 月 18 日

于昆明，云南财经大学南院，博远楼 412

第一版前言

统计学是一门研究如何有效地收集数据、整理数据、分析数据并作出有效的估计、推断和预测的方法论学科。

统计学是经济类和管理类本科生必修的一门重要基础课程，统计学理论有一定抽象性，涉及的计算比较复杂，很多计算必须借助计算机通过统计软件来完成。对于这样一门有一定理论性和实践性的课程，学生学习有一定的难度，那么如何提高学生的学习兴趣？如何提高教师的授课效果呢？我们多年的教学经验是：通过实际问题讲解理论是一种好的讲授方法，结合统计软件学习统计学效果非常好。但目前的教材这两方面做得不够，特别是在结合统计软件讲授统计理论和方法方面有待改进。大多数教科书都以 Excel 作为计算软件辅助统计学教学，事实上，Excel 不是一个专业的统计软件，它虽然可以完成一些统计计算和分析工作，但对一些略微复杂的统计分析就无能为力了；此外，Excel 输出的图表也不够美观，因此，选用合适的、操作简单的专业统计软件(比如 SPSS)代替 Excel 辅助统计学教学是现代统计学教学的必然结果。

为了适应现代统计学教学的要求，突出统计软件应用的特点，我们编写了本书。本书有以下三个主要特点：(1)简明易懂，每章都由一个实际问题(引例)引入本章要介绍的基本内容，结合实例介绍统计理论和方法，方便读者理解统计理论和方法；(2)突出统计软件应用，采用国际通用的统计软件 SPSS16.0 做统计计算和分析，辅助教学，使学生能应用统计软件对实际统计数据进行基本的统计计算和分析；(3)结构合理，侧重介绍统计学的基本理论和方法，省略了不必要的定理和公式的证明。

本书由费宇执笔编写第 1 章、第 3 章、第 4 章和第 6 章，石磊执笔编写第 5 章和第 7 章，雷健敏执笔编写第 2 章和第 9 章，马云玲执笔编写第 8 章，王任执笔编写第 10 章，最后由费宇和石磊负责全书的统稿。

在本书的写作过程中，参阅了许多国内外的文献资料，并引用了部分例题和习题，在此向这些文献的作者表示感谢。由于作者水平有限，书中难免有错漏之处，敬请读者批评指正。

此书的出版得到云南财经大学统计与数学学院统计学学科建设基金的支持，得到了高等教育出版社有关编辑的大力支持和帮助，在此表示衷心的感谢！

编　者

2010 年 7 月于昆明

目录

第 1 章　绪论

【引例】 统计数据往往是一个“出新闻”的地方。2010 年 1 月 19 日，在国家统计局网站首页的头条，人们看见了“期待已久”的《2009 年全国房地产市场运行情况》统计报告。

按照统计局公布的这份报告：2009 年 12 月份，全国 70 个大中城市房屋销售价格同比只上涨了 7.8%。特别是城市居民最关心的新建住房销售价格，同比上涨（也就是一年的涨幅）一成都不到，只有 9.1%（仅仅比国家统计局公布的 2008 年涨幅高了 2 个百分点）；二手住房的销售价格，更是只比上年同期“慢慢涨了”6.8%。

就连一些一线大城市，国家统计局公布的 2009 年一年的房价涨幅，同样也“适度而温和”，绝不“吓人”：房价“最热”的北京，它的房屋销售价格，一年也就上涨了 9.2 %，天津为 8.7%，上海、广州和深圳，分别也就 7.4%、8.7%和 18.9%。按照这份“官方的权威统计”，在这全国 70 个大中城市中，房价涨幅一年超过 10%的，仅有六七个城市，很多城市的房价涨幅连 5%都不到。但是作为全国最大房地产专业门户网站之一的搜房网，2009 年曾发布了一个号称“老百姓自己的房价榜”。他们通过对南京市的楼盘和各区域房价的全面统计和加权分析，得出 2009 年 10 月份该市商品房住宅价格比 2010 年年初的 1 月份上涨 34.18%，环比也较上月上涨 6.34%。而与此同时，国家统计局网站上公布的同比涨幅仅为 4.3%，环比也只有 2.4%。另外，据国务院发展研究中心宏观经济研究部的一份报告计算，2009 年全国住宅销售的房价涨幅已经高达 27.28%，整整超过了国家统计局公布的“新建住房销售价格”涨幅的两倍。

各种房价涨幅数据再一次“打架”，引发了“强烈的热议”。若要判断现实中诸多统计数据的合理性，这就要求我们对统计数据和统计学的基本原理有一定的认识。

（资料来源：FT 中文网：http://www.ftchinese.com/）

统计学是一门关于数据的科学，它研究如何有效地收集数据、整理数据和分析数据。本章从统计学的定义出发，介绍统计学的基本概念和统计学中应用广泛的六个分布以及正态总体下的抽样分布，为学习后面的章节做好准备。

1.1　统计数据与统计学

1.1.1　统计数据

在日常生活中，我们可能碰到各式各样的统计数据，请看下面的例子：

【例 1.1】 中国网（china.com.cn）2009 年 10 月 28 日消息，国家统计局上海调查总队近期对 1 000 户城市居民家庭开展了一项有关消费意向和消费观念的专题调查。调查结果显示，上

海市在住房消费领域存在较大消费潜力。调查中，在问及购房意向时，有2.9%的家庭表示年内打算购房，另有9.6%表示在三年内有购房意向。而抽样调查显示，2008年只有1.1%的城市居民家庭实际购房。在有购房意向的被访家庭中，有46.4%表示是结婚用房，49.6%表示为改善居住条件，而作为投资或其他用途的比重很小，只占4%。可见，对房价的稳定预期和刚性需求的持续增长，仍将有力支撑上海房地产市场的需求量。

【例1.2】 2008年美国发表的两项大型临床试验结果显示，维生素及其他抗氧化剂丝毫无助于预防前列腺癌。《美国医学会》杂志在网络版上公布了这一结果：第一项研究是迄今进行过的规模最大的癌症预防对照试验之一，有3.55万名中年男性参加，服用维生素E、硒或安慰剂的时间超过5年。第二项试验历时8年，观察了维生素C和E对近1.5万名男性的影响。两项研究均显示，无论是对前列腺癌，还是所有种类的癌症，这些补充剂都没有预防效果。

【例1.3】 《北京日报》2009年1月12日报道，《2008年中国民生问题调查:食品安全状况最令人担忧》，调查中对城乡居民询问了人身、个人和家庭财产、个人信息隐私、交通、劳动、医疗、食品7个方面的安全感，结果发现在上述7个方面平均有74.6%的人表示“很安全”或“比较安全”。其中人身方面的安全感最高(“很安全”和“比较安全”合计83.2%)，而食品和交通方面的安全感最低，分别只有65.3%和65.7%，认为不安全的人达30%以上。特别值得提及的是，在2006年和2008年的两次调查中，食品安全状况都在各类安全感中排在倒数第一，这说明公众对食品卫生和安全有着长期的担忧。调查的时间：2008年5月至9月；调查样本：此项全国抽样调查覆盖全国28个省市区的134个县(市、区)、251个乡(镇、街道)和523个村(居委会)，共成功入户访问了7 139位年龄在18至69岁的居民，调查误差小于2%，符合统计推论的科学要求。

【例1.4】 据中国国家统计局网公布的消息，2015年11月份，全国居民消费价格总水平同比上涨1.5%。其中，城市上涨1.5%，农村上涨1.3%；食品价格上涨2.3%，非食品价格上涨1.1%；消费品价格上涨1.2%，服务价格上涨2.1%。1—11月平均，全国居民消费价格总水平比上年同期上涨1.4%。

以上4个例子来自于互联网、报纸和杂志的新闻、消息和报道，例子中涉及大量统计数据，只有正确理解这些统计数据，才能真正读懂这些新闻、消息和报道。可见统计数据与我们的日常生活息息相关，不管你愿不愿意，你总要与各种统计数据打交道，这是信息时代的一个重要特征。

1.1.2 统计学

什么是统计学？一般的教科书上给出的定义为：统计学(statistics)是一门研究如何有效地收集数据、整理数据、分析数据，并根据数据作出推断的方法论科学。

按照《不列颠百科全书》的定义，统计是“收集和分析数据的科学和艺术”。这个定义认为统计学既是科学(science)，也是艺术(art)，揭示了统计学的两个重要特点。因为统计方法的应用讲究灵活性，不能教条主义，如果只记住一些数学公式和方法，没有理解公式和方法，在具体问题中不加分析地使用，那么很容易出问题。

本书强调这一点是想提醒读者注意，统计学的理论和方法固然重要，但正确理解统计学的

理论和方法更重要，所以，本书特色之一就是突出统计学的应用特点，深入浅出地介绍统计学的基本理论和方法，使读者在学习之后，能够获得解决和处理统计问题的能力。

从统计学的定义可以看出：统计学是一门关于数据的科学，统计学与统计数据密不可分，统计数据是统计学的研究对象，学习了统计学的基本理论和方法，我们就能够理解日常生活中常见的很多统计数据的含义。比如，学习了第3章参数估计后，就能理解例1.1中基于1 000户城市居民家庭抽样调查得到的结果，“有2.9%的家庭表示年内打算购房，另有9.6%表示在三年内有购房意向”这两个数字的可靠性是很高的；学习了第4章假设检验后，就能知道例1.2“维生素及其他抗氧化剂丝毫无助于预防前列腺癌”，这个结论可以根据统计假设检验得到，如果假设检验的结果在统计上是显著的，那么这个结论应该是可靠的；学习了第8章抽样调查理论与方法后，就容易知道例1.3中“调查误差小于2%”是如何估算出来的；学习了第9章统计指数中的居民消费价格指数后，我们就能理解例1.4中“全国居民消费价格总水平同比上涨1.5%”的确切含义是什么。

1.2 统计学的历史

统计活动的历史非常悠久。在中国，从公元前21世纪夏禹时代开始就出现了记录人口和土地的统计活动；在西方，公元前29世纪左右，古埃及法老每两年就派人清查一次全国的人口，这些都是最初的统计活动。

但是，作为一门独立的学科，统计学的历史却不算很长，一般认为，统计学有两个主要来源，一个是产生于17世纪德国的国势学，另一个是产生于17世纪英国的政治算术。国势学以国家政治社会情况作为研究对象，运用对比的方法来研究各国实力的强弱，国势学派的主要代表人物是康令(H. Conring)和阿亨瓦尔(G. Achenwall)，阿亨瓦尔1749年在《欧洲最主要各国新国势学概要》中首创了一个新的德文词——Statistik，即统计学，代替康令的国势学(Staatenkunde)，此外，国势学派还提出了至今仍为统计学者使用的“统计数字资料”和“数字对比”等术语。政治算术的创始人是英国的配第(W. Petty)，他在1690年出版的《政治算术》一书中以数字资料为基础，采用数量分析方法研究政治问题，第一次提出统计方法并利用统计方法分析数字资料，因此，马克思称他为统计学的创始人。

统计学从18世纪开始与概率论结合，概率论为统计学的进一步发展奠定了坚实的数理基础，促进了统计学理论和实践的繁荣昌盛，到19世纪末已经形成了古典统计学(描述统计学)的主要框架。

20世纪以来，统计学渗透到社会学、生物学、经济学等领域，英国统计学家哥塞特(W. S. Gosset)1908年以Student为笔名在《生物计量学》杂志上发表论文《均值的或然误差》，提出了著名的t统计量，开创了小样本理论先河，使统计学开始由大样本向小样本、由描述统计向推断统计发展。紧接着，20世纪成就最大的统计学家费雪(R.A.Fisher)对t分布、χ^2分布和F分布加以综合研究，还提出了方差分析方法和最大似然估计方法，大大促进了推断统计学的发展。后来，奈曼(J.Neyman)和皮尔逊(E.S.Pearson)提出了系统的统计假设检验理论，并对区间估计做出了系统发展，瓦尔德(A.Wald)提出序贯分析法和统计决策函数理论，进一步丰富了现代统计学的理论，形成了现代统计学(推断统计学)的框架。

1.3 统计学的分类

作为一门方法论学科，统计学可以应用于很多领域，包括生物学、医学、经济学、管理学、政治学、物理学、社会学和心理学等，统计学的分类也五花八门，比如按理论和应用将统计学分为理论统计学和应用统计学，按应用领域将统计学分为生物统计学、经济统计学、卫生统计学、社会统计学等。这些分类比较简单、直观，强调统计学的应用性，这里我们从统计学的研究内容和方法上将统计学分为描述统计和推断统计。

描述统计(descriptive statistics)就是用数字和图表等方法对数据进行总结和展示，揭示数据的基本特征，为进一步的统计推断作准备。而推断统计(inferential statistics)是根据样本数据对总体进行估计、预测和推断，这是现代统计学的核心内容。

下面的例子是描述统计和推断统计应用的实例。

【例 1.5】 2015 年 11 月份，食品价格同比上涨 2.3%，烟酒及用品、医疗保健和个人用品、衣着、娱乐教育文化用品及服务、家庭设备用品及维修服务、居住价格分别上涨 3.8%、2.6%、2.2%、1.2%、0.8%和 0.7%，交通和通信价格下降 1.4%，如图 1.1 所示。

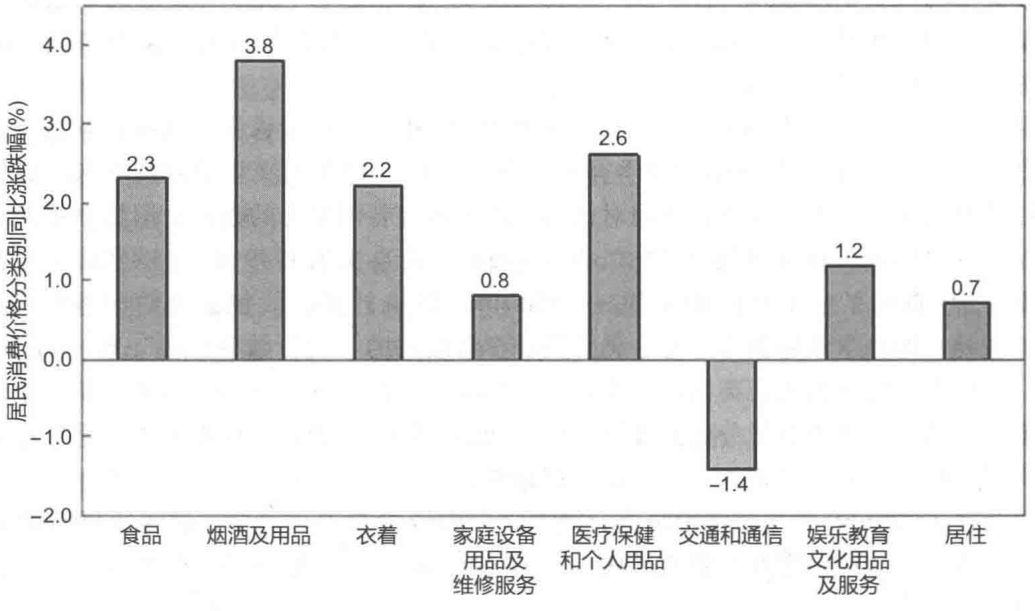

图 1.1 2015 年 11 月份居民消费价格分类别同比涨跌幅示意图

资料来源：国家统计局。

【例 1.6】 2015 年 1—9 月份，全国房地产开发投资 70 535 亿元，同比名义增长 2.6%(扣除价格因素实际增长 4.2%)，增速比 1—8 月份回落 0.9 个百分点。其中，住宅投资 47 505 亿元，增长 1.7%，增速回落 0.6 个百分点。住宅投资占房地产开发投资的比重为 67.3%。

【例 1.7】 《全球最佳 CEO》中介绍，如果根据全球各顶尖首席执行官整个任期内的股东投资回报率和市值变化来排序，那么分析 1995 年至 2009 年间 1 109 位首席执行官后可以发现，拥有 MBA 学位的平均表现要好于那些没有 MBA 学位的。虽然差别并不大，但在统计学上却意

义重大。如果再深入一层研究就会发现，那些在50岁以前就爬到首席执行官位置的，从商学院教育中获益尤其大。实际上，拥有MBA学位，能帮助这些人在榜单上的排名平均提前足足100名。

【例1.8】 零点研究咨询集团、北汽福田汽车股份有限公司、新浪汽车联合发布的《2009福田指数中国居民生活机动性指数研究报告》指出：在参与调查的7个城市中，北京居民上下班或上下学拥堵经济成本为335.6元/月，处在各城市之首。其次是广州和上海，拥堵经济成本分别为265.9元/月和253.6元/月。同时，北京居民上下班时间花费也居高位，道路畅通时平均花费时间40.1分钟，而道路拥堵时则达到62.3分钟。

例1.5描述了2015年11月份食品和非食品商品价格变化情况，而例1.6给出了2015年1—9月份，全国房地产开发投资总额及增长情况，这两个例子属于描述统计的内容。

例1.7和例1.8的结论是根据抽样得到的样本数据估计出来的，这两个例子属于推断统计的内容。

1.4 统计学的基本概念

随机变量、总体、样本和统计量是统计学四个最基本的概念。

1.4.1 随机变量及其分布

统计学是研究随机现象规律性的科学，为了对随机现象的结果进行更简单和直接的定量分析，我们通常引入随机变量这个工具。

随机变量(random variable)是定义在样本空间上的实值函数$X=X(\omega)$，它随样本点ω的变化而变化，它用来描述随机试验的结果。如果一个随机变量只能取有限个或可列个值，则称它为离散型随机变量(discrete random variable)；如果一个随机变量的可能取值充满数轴上的一个区间，则称它为连续型随机变量(continuous random variable)。

【例1.9】 抛一颗六面均匀的骰子，观察出现的点数，则出现的点数X是一个随机变量。

【例1.10】 调查100个顾客，考察顾客对某个品牌笔记本电脑的偏好，记录喜欢这个品牌笔记本电脑的人数X，则X是一个随机变量。

【例1.11】 为了检验某电子产品的质量，检测它的使用寿命(以分钟记)，则产品的使用寿命X是一个随机变量。

【例1.12】 从一大批产品中随机抽取若干个产品，考察次品率X，则X是一个随机变量。

以上4个例子中，例1.9和例1.10中的随机变量X的取值只能是有限个整数，属于离散型随机变量；而例1.11中的随机变量X的取值可以是大于等于0的任何实数，例1.12中的随机变量X的取值可能是0到1之间的任何实数，属于连续型随机变量。

分布函数是描述随机变量分布的重要工具，它既可以用来描述离散型随机变量，也可以用来描述连续型随机变量。

定义1.1 设X是一个随机变量，对任意实数x，称

$$F(x)=P(X\leqslant x) \tag{1.1}$$

为随机变量X的分布函数(distribution function)，称X服从$F(x)$，简记为$X\sim F(x)$。

对于连续型随机变量，通常采用概率密度函数来描述它的分布。下面是概率密度函数的定义。

定义 1.2 设随机变量 X 的分布函数是 $F(x)$，如果存在一个非负函数 $f(x)$，使得对任意实数 x，有

$$F(x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t \tag{1.2}$$

则称 $f(x)$ 是连续型随机变量 X 的概率密度函数(probability density function)，简称为密度函数。

对离散型随机变量，我们往往用概率函数描述它的分布特征。

1.4.2 总体和总体分布

什么是总体？我们研究对象的全体就称为总体(population)或母体，总体中的元素称为个体。如果总体包含的个体的数目是有限的，则称为有限总体；如果总体包含的个体的数目是无限的，就称为无限总体。

【例 1.13】 一批电子元件共 10 万个，研究这批电子元件的平均使用寿命，则该批电子元件的全部使用寿命就构成一个总体，而每个电子元件的使用寿命就是个体。

【例 1.14】 考察某大学一年级新生的身高情况，则全体新生的身高就构成一个总体，而其中每个学生的身高就是个体。

以上两个例子可以看出，统计学中的总体实际是指研究对象的某个(或某几个)数量指标取值的全体，比如例 1.13 中数量指标 X 是使用寿命，而相应的总体是由使用寿命 X 的可能取值的全体构成的集合；而例 1.14 中数量指标 X 是身高，相应的总体是由身高 X 的可能取值的全体构成的集合。由于不同的个体，X 取不同的值，且事先无法准确预言，所以 X 是随机变量，也简单地称 X 是总体，而 X 的分布就称为总体分布(population distribution)。

1.4.3 样本和样本分布

从总体中取出的部分个体构成的集合称为样本，样本中的个体数目称为样本容量。取得样本的过程称为抽样(sampling)，常用的抽样方法有简单随机抽样、分层抽样、整群抽样、等距抽样等。最简单的抽样方法是简单随机抽样，即在取样时，总体中的每个个体入选的机会是相同的，本书如果没有特别说明，我们讨论的样本都是指简单随机抽样得到的简单随机样本(simple random sample)，简称为样本(sample)。采用 $(x_1,\cdots,x_n)$ 记样本容量为 n 的样本，其中，$x_1,\cdots,x_n$ 是相互独立的与总体 X 同分布的 n 个随机变量；$(x_1,\cdots,x_n)$ 的观测值称为样本值，仍记作 $(x_1,\cdots,x_n)$。为了简便起见，有时把样本和样本值统称为样本。

对于简单随机样本 $(x_1,\cdots,x_n)$，若总体 X 的分布函数为 $F(x)$，则样本 $(x_1,\cdots,x_n)$ 的联合分布函数为

$$F(x_1,\cdots,x_n)=\prod_{i=1}^{n}F(x_i)$$

若总体 X 的概率密度函数为 $f(x)$，则样本 $(x_1,\cdots,x_n)$ 的联合概率密度函数为

$$f(x_1,\cdots,x_n)=\prod_{i=1}^{n}f(x_i)$$

若总体 X 是离散型随机变量，其概率函数为 $p(x)=P(X=x)$，则样本$(x_1,\cdots,x_n)$的联合概率函数为

$$p(x_1,\cdots,x_n)=\prod_{i=1}^{n}p(x_i)$$

样本分布 $F(x_1,\cdots,x_n)$、$f(x_1,\cdots,x_n)$或 $p(x_1,\cdots,x_n)$是统计推断的基础。

1.4.4 统计量

抽样获得样本后，根据样本信息推断总体时，通常需要对样本信息进行加工整理，针对不同的问题构造适当的样本函数 $T=T(x_1,\cdots,x_n)$，这种用来推断总体的样本函数称为统计量(statistic)。统计量是用作统计推断的量，所以统计量 $T(x_1,\cdots,x_n)$不能含有未知参数。

设$(x_1,\cdots,x_n)$是来自于总体 $X\sim N(\mu,\sigma^2)$的样本，常见的统计量有以下几个。

1. 样本均值

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{1.3}$$

称为样本均值(sample mean)；它是总体期望 μ 的无偏估计。

2. 样本方差

$$s_n^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{1.4}$$

称为样本方差(sample variance)，其算术平方根 s_n 称为样本标准差(sample standard devication)。修正样本方差

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \tag{1.5}$$

也称为样本方差，其算术平方根 s 也称为样本标准差。因为修正样本方差是总体方差 σ^2 的无偏估计，在实际中，s^2 比 s_n^2 更常用，今后提到样本方差通常是指 s^2。

3. 样本矩

$$A_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k \tag{1.6}$$

和

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k \tag{1.7}$$

分别称为样本 k 阶原点矩和样本 k 阶中心矩。样本矩可以用来估计总体矩，从而获得相应的矩估计。显然，按样本矩的定义，样本均值和样本方差分别是样本一阶原点矩和样本二阶中心矩。

4. 次序统计量

将样本$(x_1,\cdots,x_n)$按 x_i 由小到大排列得到的有序样本$(x_{(1)},\cdots,x_{(n)})$称为样本的次序统计量(order statistic)，其中，$x_{(i)}$为样本的第 i 个次序统计量；$x_{(1)}$称为样本的最小次序统计量，$x_{(n)}$称为样本的最大次序统计量。

5. 样本中位数和样本极差