

统计学

statistics statistics

statistics statistics

张仁寿 主编

 中国统计出版社
China Statistics Press

统计学

statistics statistics

statistics

张仁寿 主编



中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

统计学 / 张仁寿主编. — 北京 : 中国统计出版社,
2018.5

ISBN 978-7-5037-8438-5

I. ①统… II. ①张… 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(2018)第 006779 号

统计学

作 者/张仁寿

责任编辑/梁 超

封面设计/李 静

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://www.zgtjcs.com>

印 刷/河北鑫兆源印刷有限公司

经 销/新华书店

开 本/710×1000mm 1/16

字 数/323 千字

印 张/19.25

版 别/2018 年 5 月第 1 版

版 次/2018 年 5 月第 1 次印刷

定 价/48.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在世界任何地区
以任何文字翻印、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

前

言

在互联网思维、互联网经济、移动互联、电子商务等新兴产业纷纷走上历史舞台的大背景下,大数据(Big Data)时代来临了。大数据时代,统计理论与方法得到前所未有的高度重视,是机遇,更是挑战。尤其是,传统的统计基本方法和数理统计方法将在金融、证券、保险、投资、理财、电子商务、社会管理、科技、医学、生物、生态等领域,得到更为广泛应用。

然而,大数据分析离不开统计学的基本理论和方法。可以说,统计学是大数据分析的基础或基石。这是因为大数据技术中,数据采集、数据处理和数据挖掘都离不开统计学的分类(Classification)、估计(Estimation)、预测(Prediction)、相关性分组或关联规则(Affinity grouping or association rules)、聚类(Clustering)、描述(Description)、复杂数据类型挖掘(Text, Web, 图形图像,视频,音频等)等基本方法。更重要的是大数据分析所依赖的分析方法:假设检验、显著性检验、差异分析、相关分析、T 检验、方差分析、卡方分析、偏相关分析、距离分析、回归分析、简单回归分析、多元回归分析、逐步回归、回归预测与残差分析、logistic 回归分析、曲线估计、因子分析、聚类分析、主成分分析、因子分析、快速聚类法与聚类法、判别分析、对

应分析、多元对应分析(最优尺度分析)等等,都是统计分析方法(诚然,具体的大数据处理方法确实还有很多,在此不一一列举)。

统计学是一门测度、收集、整理、显示和分析统计数据的科学,其目的是探索数据内在的数量规律性,因而也可称统计学为“数据的科学”。为适应普通高校应用统计学教学需要,使广大学生掌握统计思想、理论背景、方法思路和发展前景,提高统计分析能力和解决社会经济的实际问题的能力,组织编写了本教材。为了应对大数据处理及分析的需要,本教材体系编写中,首先是注重实用性,对基本理论方法介绍尽量做到深入浅出,贴近社会生活;其次是内容创新,与其他教材相比,增加了“非参数检验”等内容。这是因为,一方面,非参数检验方法不要求数据的计量尺度局限于数值型数据,对样本数据的总体概率分布也不需要事先设定特定的形式。需要指出的是,在许多参数检验和非参数检验方法都适用的场合下,非参数检验和参数检验方法有着差不多的表现。当数据是非数值类型时,这时候只有非参数检验方法能够适用。由于对数据类型和样本总体概率分布的要求比较宽松,非参数检验方法通常比参数检验更能被广泛地运用。另一方面,大数据中多数属于非数值类型数据。对非数值类型海量数据进行分析检验,只有通过非参数检验方法分析才能获取很多智能的、深入的、有价值的信息,才能获得有巨大价值的产品和服务,或深刻的洞见,最终形成变革或创新之动力。此外,在调查方法中增加了网络调查(如问卷星、微信平台、移动互联网 APP)等信息;在数据处理部分增加了直接插入排序、希尔排序和堆排序等方法。再次,精心制作了配套使用的电子课件,便于教学。

全书共 15 章,第 1 章导论,主要从应用的角度介绍统计方法的基本思想、统计学的基本概念,包括统计总体、样本、参数、统计量、变量等。第 2 章描述统计——表格图形法,主要内容包括:资料的收集、数据的加工处理、数据的展示、数据分布特征的概括与分析。第 3 章描述统计——数值方法,

主要讨论数据分布特征的数值方法、特征及其应用场合,同时也讨论了异常值的检测,两变量间关系的度量等内容。第4章概率基础,主要介绍推断统计学的理论基础,包括随机事件及其概率、概率的性质与运算法则、随机变量及其分布等内容。第5章抽样理论和参数估计,重点对抽样理论和参数估计方法进行介绍。第6章假设检验,主要介绍利用样本数据,对两个对立命题进行分析的方法。第7章两总体均值推断,主要讨论两总体均值之差的区间估计和假设检验以及两总体比例之差的区间估计和假设检验。第8章总体方差的统计推断,主要讨论两个总体均值之差和两总体方差比的区间估计问题。第9章分类数据的统计检验,这一章讨论的统计方法主要用于分类数据的分析,即列联分析或列联表分析。第10章方差分析,主要介绍英国统计学家费希尔爵士创立的方差分析方法(Analysis of Variance, ANOVA)。第11章非参数检验,主要介绍一些非参数检验方法。这一类方法可以应用到之前参数检验方法不能适用的一些场合。第12章相关与回归分析,相关与回归分析是指数值型自变量和数值型因变量之间关系的分析。相关与回归是处理变量之间关系的一种统计方法。第13章统计指数,主要介绍统计指数的基本概念与分类,统计指数的编制原理,以及统计指数在实际工作中的几种应用。第14章时间序列分析,是一种动态数据处理的统计方法。该方法基于随机过程理论和数理统计学方法,研究随机数据序列所遵从的统计规律,解决实际问题。第15章统计综合评价,主要阐述综合评价的理论和方法。

参加本书编写的有张仁寿教授(第1、9章),夏明会教授(第2章),杨林副教授(第3章),邹文理副教授(第4、14章),马健副教授(第5、6、10章),黄小军副教授(第7、8章),张兴发讲师(第11章),范娜讲师(第12章),钟云燕副教授(第13、15章)。由张仁寿教授担任本书的主编,马健、钟云燕担任本书的副主编,负责全书的规划设计、总撰和定稿工作。中央党校研究

生况达同学参加了第1、2、3、9部分章节初稿的资料搜集和撰写。广州大学经济与统计学院研究生李栋同学参加了本书PPT制作,并协助对书稿文字内容进行校对工作。

本书写作过程中得到广州大学校领导的关心和大力支持;得到广州大学教材出版资金的资助。在此表示感谢!本书的顺利出版,还得到中国统计出版社领导和编辑部门的大力支持,在此一一致谢!本书的编写过程中参考了大量国内外同行近年来最新出版的相关教材及科研成果,在此一并致谢!

本教材适用于普通高校统计学专业本科生、经济和管理类各专业本科生教学。也适合于普通高校研究生教学参考书。

限于水平和经验,在教材的编写和编辑过程中缺点和错误在所难免,恳请同行专家和广大读者批评指教。

2018年5月

目

录

第一章 导 论	1
1.1 统计和统计学	1
1.2 统计学在商务和经济中的应用	8
1.3 数据与数据的来源	9
1.4 大数据简介	16
思考与练习	25
第二章 描述统计——表格图形式	27
2.1 统计数据的初步处理	27
2.2 统计数据的汇总与展示	35
2.3 多变量数据的汇总与展示	47
2.4 规范合理使用图表	51
思考与练习	55
第三章 描述统计——数值方法	59
3.1 分布位置集中趋势的度量	59
3.2 分布变异程度的度量	65
3.3 分布形态、相对位置的度量	71

3.4 异常值的检测	73
思考与练习	77
第四章 概率基础	79
4.1 随机事件及其概率	79
4.2 概率的性质与运算法则	83
4.3 离散型随机变量及其分布	86
4.4 连续型随机变量及其分布	95
思考与练习	100
第五章 抽样理论和参数估计	101
5.1 抽样的基本概念	101
5.2 抽样分布	103
5.3 区间估计	108
5.4 估算必要样本容量	116
5.5 总体比例的区间估计	117
5.6 利用统计软件计算置信区间	119
5.7 本章小结	121
思考与练习	122
第六章 假设检验	125
6.1 建立原假设与备择假设	125
6.2 第一类与第二类错误	126
6.3 总体平均数的假设检验—— σ 已知	127
6.4 总体平均数的假设检验——总体标准差 σ 未知	134
6.5 总体比例的假设检验	138
6.6 本章小结	140
思考与练习	141
第七章 两总体均值推断	143
7.1 两总体均值之差的推断	143

7.2 两总体比例之差的推断	154
思考与练习	157
第八章 总体方差的统计推断	160
8.1 一个总体方差的统计推断	160
8.2 两个总体方差的统计推断	163
思考与练习	166
第九章 分类数据的统计检验	168
9.1 分类数据与列联表	168
9.2 拟合优度检验	171
9.3 独立性检验	175
思考与练习	178
第十章 方差分析	180
10.1 方差分析的基本假设	180
10.2 方差分析检验	183
10.3 单因子方差分析的软件实现	187
10.4 多重比较方法	189
10.5 本章小结	192
思考与练习	193
第十一章 非参数检验	195
11.1 符号检验	196
11.2 威尔科克森符号秩检验	198
11.3 曼—惠特尼 U 检验	201
11.4 克鲁斯卡尔—沃利斯检验	203
11.5 秩相关系数	205
11.6 本章小结	207
思考与练习	209

第十二章 相关与回归分析	211
12.1 相关关系的度量	211
12.2 简单线性回归分析	218
12.3 多元线性回归	224
思考与练习	226
第十三章 统计指数	229
13.1 统计指数概述	229
13.2 统计指数的编制方法	233
13.3 指数因素分析	242
13.4 经济指数的编制及应用	254
思考与练习	257
第十四章 时间序列分析	259
14.1 时间序列的分解	260
14.2 时间序列的描述性分析	262
14.3 时间序列的预测	265
思考与练习	268
第十五章 统计综合评价	271
15.1 统计综合评价概述	271
15.2 统计综合评价指标确定与数据预处理	273
15.3 权重的确定与评价结果的综合	279
思考与练习	285
附录 常用统计表	287
参考文献	298

第一章 导 论

统计学为统计活动提供数量研究和认识社会经济规律的科学方法。统计学的核心内容是数据的测度、搜集、整理、归纳、分析的原理和方法。本章主要介绍统计方法的基本思想、统计学的基本概念,包括统计总体、样本、参数、统计量、变量等。

1.1 统计和统计学

1.1.1 走近统计学

统计学是一门测度、收集、整理、显示和分析统计数据的科学,其目的是探索数据内在的数量规律性。正是因为统计学总是在和数据打交道,因而也可称统计学为“数据的科学”。

什么是统计测度,目前学界没有一个统一的定义。国内学者李金昌^①(2015)指出,国内文献资料带有测度这个词的文献不少,但专门针对统计测度(或测度)内涵的讨论几乎没有。一些对社会经济现象进行分析测度的文献,例如收入分配公平性测度、技术效率测度、人力资本测度、产业关联测度等等,所做的测度都是再测度,均不对测度本身进行讨论。李金昌查阅国外文献资料(关键词:Measurement),发现也同样存在这样的问题,只能收集到一些比较零散的表述。如 Ludwik Finkelstein(1975)认为,在我们对事物或现象进行描述时,测度可以被定义为对现实世界中某一现象的个体属性或特征进行量化的过程。Battista Rossi(2007)认为,用以表示测度结果的,是数字或者数字符号。Ludwik Finkelstein(2014)指出,在自然科学技术中,测度的重要性不可否认,它是科学调查和发现必不可少的工具,它可将宇宙中的复杂现象用精确、简洁和普遍的数学语言来描述。根据现有的研究成果,李金昌认为统计测度就是用一定的符号和数字,用一定的形式和载体,对所研究的现象或事物的特征进行量化反映,表现为可

① 李金昌.统计测度:统计学迈向数据科学的基础[J].统计研究,2015(8):1-7.

用于统计分析的数据的过程。

统计测度方法有直接测度和间接测度两大类。其中,直接测度包括计数测度、测量测度和实验测度;间接测度包括定义测度和模型测度。也有学者认为,定义测度既可能是直接测度、也可能是间接测度。计数测度是一种通过观测计数来获得数据的方法,最为简单,一般用于自然测度或时间测度;测量测度是一种根据物理或化学规制对现象进行测量、测算来获得数据的方法,一般用于物理测度或化学测度;实验测度是一种按照科学实验原理,通过观察实验对象在既定条件下的反应来获得数据的方法,一般与测量测度相结合,用于获取科学研究数据;定义测度也可以称之为指标测度,是一种通过探究现象的本质特征和活动规律,归纳出表现其数量特征的范畴,给出统计指标定义(包括内容、口径、计算方法和表现形式等)来获取数据的方法,最常用于价值测度,也用于其他形式的测度。可以说,定义测度方法应用最为广泛,但也最为困难。模型测度是一种根据现象与现象之间的内在联系关系、或者现象自身的发展变化规律,通过建立一定的方程模型来获取数据的方法。

在日常工作和生活中,到处都有统计数据。例如,组织外出旅游活动时负责人要统计一下参加旅游的人数;球类比赛时解说员总要统计竞赛双方的进攻次数和成功率;企业管理人员要掌握生产销售情况和利润额;人们可以通过报刊和电视等获得 GDP(国内生产总值)、CPI(消费者价格指数)和经济增长率等数据;等等。日常工作和生活中的这些数字就是人们所关心的统计数据。

怎样理解统计学是数据的科学呢?下面先看看统计学英文的名词,再举几个例子。

统计学的英文是“Statistics”。它通常有两个含义:当它以单数名词出现时,表示一门学科的名称——“统计学”;当它以复数名词出现时,表示“统计数据”或“统计资料”。“Statistics”一词的英文解释至少可以说明两件事情:

第一,由于统计数据在英文中是以复数形式出现的,表明统计数据不是指个别的单个数字,而是指同类的较多数据。因为单个数字如果不和其他数据进行比较,是不能说明问题的。例如,某个学生在某门课程的考试中得了 85 分,如果仅凭这一个数字,我们很难对这位学生的知识和能力水平作出判断和评价。因为这个 85 分可能是班上的最高分,可能是中等水平的分数,也可能是较低的分数。如果还知道这次考试的平均分数,就可以对这位学生的成绩是高于还是低于平均分数,以及是高多少或者低多少作出评价了。在生产和生活实际问题中,通常可以收集到较多的数据,进而利用统计方法对数据进行加工整理,从而发现数据中的内在联系及数量规律。因此,对数据的科学性分析成为我们下结论、作出决策的

重要依据。

第二,作为单数的统计学和作为复数的统计数据在英文中使用同一名词,显示出二者之间的密切关系。统计学是由收集、整理、显示和分析统计数据的方法组成的,这些方来源于对统计数据的研究,目的也在于对统计数据的研究。离开了统计数据,统计方法乃至统计学就失去了其存在的意义。这正如俗语所说“巧妇难为无米之炊”。这里的“妇”就是掌握统计方法的统计学家或统计工作者,“米”就是统计数据,“炊”就是统计研究或统计工作的目的,即探索数据内在的数量规律性。显然,没有统计数据或没有较好的统计数据,即使很科学的统计方法或很高明的统计学家也难有所作为。

那么,什么是数据内在的数量规律性呢?下面用几个简单的例子来说明。

例 1.1 统计数据会说话:活捉廖耀湘。辽沈战役期间,东北野战军以雷霆之势攻克锦州后,迅速挥师北上与赶来增援的廖耀湘兵团在辽西碰面,一时间双方形成了混战局面。按照林彪的一贯要求,所属各纵队及师团每天都必须上报当日的战况和缴获情况。在一天深夜,值班参谋正宣读某师上报的数据,林彪猛然叫停,问道:“刚才念的在胡家窝棚那场战斗的缴获,你们听到了吗?”林彪见无人出声,便连问了三句:为什么那里缴获的短枪与长枪的比例比其他战斗高?为什么那里缴获和击毁的小车与大车的比例比其他战斗高?为什么在那里俘虏和击毙的军官与士兵的比例比其他战斗高?人们还没来得及思索和回答,林彪就指着地图上的那个点说:“我断定,敌人的指挥所就在这里!”果不其然,廖耀湘这条“大鱼”在胡家窝棚附近被逮个正着。

例 1.2 一个家庭新生婴儿的性别可能是男的,也可能是女的。在过去没有实行计划生育时,有的家庭几个孩子都是男孩,也有的都是女孩。从表面上看,新生婴儿的性别比例似乎没什么规律可循。但如果对新生婴儿的性别进行大量观察,即观察成千上万个或者更多时,就会发现性别比例还是有规律的,即婴儿总数中男孩要多于女孩,大致为每生育 100 个女孩,就有 105 个左右的男孩。这个 $105:100$ 的比例就是新生婴儿男女性别的数量规律性,古今中外都大致相同,它是人类社会长期遗传与发展的结果。因为人类社会要延续、要发展,就要保持男女人数的大致相同。那么有人会问,新生儿男多于女,不是性别不平衡了吗?是的,新生儿时男多于女会出现不平衡,但男孩的死亡率高于女孩,到了中青年时,男女人数就大致相同了。进入中老年后,男性的死亡率仍然高于女性,导致男性的平均预期寿命比女性短,长寿的男性要少于女性。从一个国家看,如果没有人为的 B 超检查和堕胎等干扰,其规律是婴幼儿时男性略多于女性,中青年时男女人数大致相同,老年时女性又略多于男性。这样既保证人类在中青年结婚生育时性别的

大致平衡,又使得在人口总数上男女也大体相当,有利于人类社会的进化和发展。对人类性别比例的研究是统计学的起源之一,也是统计方法探索的最早的数量规律之一。

例 1.3 我们都做过掷硬币和掷骰子的游戏,都知道随机地掷一次硬币或骰子不可能事先确定出现正面还是反面或某个点数,也就是说个别游戏或试验中充满了不确定性或偶然性。机会游戏或赌博正是利用了这种不确定性和偶然性才能够吸引人。但当我们进行大量观察,即不断做重复试验时,就会发现掷一枚均匀硬币出现正面和反面的比率会大体相同,即 $0.5:0.5$ 。试验的次数越多,出现正面和反面的可能性就越接近 $1/2$ 这一稳定的数值。同样,在掷一枚均匀骰子时,出现 1 至 6 点中任一点数的可能性也都接近 $1/6$ 。这里的 $1/2$ 和 $1/6$ 就是掷硬币和掷骰子出现某一特定结果的概率,也就是我们探索的数量规律性。

例 1.4 在进行农作物试验时,如果其他试验条件相对固定,我们会发现某种粮食作物的产量会随某种施肥量的增加而增加。当开始增加施肥量时,产量增加较快。以后增加同样的施肥量,粮食产量的增加量逐渐减少。当施肥量增加到一定数值量,产量不再增加。这时如果再增加肥料,产量反而会减少。这一施肥量与粮食产量的数量关系就是我们要探索的规律性。当我们从大量试验数据中用统计方法探索到施肥量与粮食产量之间的数量关系时,就可以考虑肥料的费用并选择最佳的施肥效果了。探索事物之间的数量因果关系是统计学的核心任务。

例 1.5 与例 1.4 相似但有不同的是商品广告费用与销售额的关系。一般而言,随着广告费用的增加,商品的知名度和销售额会相应增加。对于不同的商品,投入相同的广告费用带来的销售额显然是不尽相同的。即使是同一种商品,投入的第一个 50 万元广告费、第二个 50 万元广告费和第三个 50 万元广告费等所带来的销售增量也是不同的。统计方法就是要从广告费与销售额的数据中找出其内在的数量规律,从而确定最佳广告策略。

例 1.6 天津汽车制造厂与丰田公司合资生产的威驰(Vios)汽车安装的是 8A 型发动机,排量为 1.342L,公司声称其手动挡汽车每百公里油耗(等速情况下)不超过 5L。要检验购买的车辆是否达到了说明书上的节油标准,就需要计算在高速路上匀速行驶的平均油耗。假定你的汽车平均每百公里油耗为 5.5 L,那么你的汽车是属于质量达标汽车中恰巧由于运气不好碰上的个例呢,还是该批产品本身就不合格呢?统计方法可以作出检验并给予回答。

例 1.7 某企业开发出一种新的化妆品,在正式投产之前,需要根据市场需求情况制定其价格和销售策略等。该企业委托某市场调查公司在全市 200 万户家庭中抽取 1000 户家庭作为样本,免费赠送给这 1000 户家庭试用,然后了解该化妆

品的销售前景。如何科学地从该市 200 万户家庭中抽出 1000 户家庭进行调查,并由这 1000 户家庭反馈的市场信息科学地推断出全市 200 万户家庭对该化妆品的购买意愿这一数量规律,是统计工作者的任务。

1.1.2 统计方法的基本思想

为什么统计方法能够通过对数据的大量观察和处理,进而研究和探索出其内在的数量规律性呢?这是由客观事物本身的特点和统计方法的特性共同决定的。从客观事物方面来说,根据辩证法的基本原理,任何客观事物都是必然性与偶然性的对立统一。同样,任何一个数据也都是必然性与偶然性共同作用的结果,是二者作用的对立统一。必然性反映了事物本质的特征和联系,是比较稳定的,因而它决定了事物的内在本质是有规律可循的。偶然性反映了该事物每个表现形式的差异。如果客观事物只有必然性一个方面的特征,事物的表现形式就会比较简单,就可以比较容易把握它的规律性。正是由于偶然性的存在,造成了事物的表现形式与必然性和规律性发生偏移,从而形成了表面形式的千姿百态,形成数据表现形式的千差万别。这样,必然性的数量规律性就被掩盖在表面的差异之中了。前面举出的 7 个例子,本身都存在必然性的数量规律,但就每个新生儿的性别、每次掷硬币的结果、每次施肥带来的增产、每次广告增加的销售额、每辆汽车的百公里耗油量、每个家庭对化妆品的购买意愿都是不同的,是有差异的,其表现形式也是充满偶然性的。但每个例子本身都是有规律可循的,应用统计方法就可以从偶然性中探索到内在的、本质的数量规律。从统计方法来看,统计学提供了一系列的方法,专门用来收集数据、整理数据、显示数据的特征,进而分析和探索(或推断)出事物总体的数量规律性。当然,如果事物本身的规律比较简单,所用的统计方法也就相对容易;如果事物本身的规律错综复杂,所用的统计方法也就相对复杂。以上的 7 个例子中,前三个例子比较简单,用描述统计的方法就可以解决问题。而后面 4 个例子就需要应用比较复杂的推断统计方法,如后面章节中的抽样与抽样分布、假设检验和相关与回归分析等。至于选择什么统计方法,取决于有什么类型的数据以及要解决什么问题。

1.1.3 统计学的基本概念

1. 总体

总体(Population)是人们研究的所有基本单位(通常是人、物体、交易或事件)的总和。例如,总体可以包括:①中国的全部人口;②广州市的选民总人数;③华为智能手机的所有客户数;④广汽丰田一年的所有汽车产量;⑤201×年春运期间铁路销售的总的车票数量;⑥华为公司在 201×年的销售额;⑦201×年中国发生的总的群体事件次数等。其中,①至③这三个总体是人的总和;④和⑤这两个总

体是物体的总和；⑥这一总体是交易的总和；⑦这一总体是事件的总和。同时也可以看出，每一个总和都包括了研究总体的所有单位。

2. 样本

样本(Sample)是总体的一部分单位。例如，一家上市公司正在接受审计调查，审计人员没有必要对该公司年度内的所有 55400 张发票全部审查，只需随机抽查 100 张发票的样本即可，审计人员通过这 100 张样本发票计算的差错率就可对全部 55400 张发票的差错率进行推断。样本也可以理解中从总体中抽取的部分单位。

3. 参数

参数(Parameter)是用来描述总体特征的概括性数字度量，它是研究者想要了解的总体的某种特征值。研究者所关心的参数通常有总体平均数、总体标准差、总体比例等。在统计中，总体参数通常用希腊字母表示。比如，总体平均数用 μ (Mu) 表示，总体标准差用 σ (Sigma) 表示，总体比例用 π (Pi) 表示，等等。由于总体数据通常是不知道的，所以参数是一个未知的常数。比如，我们不知道某一地区所有人口的平均年龄；不知道一个城市所有家庭的收入的差异；不知道一批产品的合格率，等等。正因为如此，所以才进行抽样，根据样本计算出某些值去估计总体参数。

4. 统计量

统计量(Statistic)是用来描述样本特征的概括性数字度量。它是根据样本数据计算出来的一个量，由于抽样是随机的，因此统计量是样本的函数。研究者所关心的统计量主要有样本平均数、样本标准差、样本比例等。样本统计量通常用英文字母来表示。比如，样本平均数用 $\bar{x}$ (读作 x - bar) 表示，样本标准差用 s 表示，样本比例用 p 表示，等等。由于样本是已经抽出来的，所以统计量总是知道的。抽样的目的就是要根据样本统计量去估计总体参数。比如，用样本平均数($\bar{x}$)去估计总体平均数(μ)，用样本标准差(s)去估计总体标准差(σ)，用样本比例(p)去估计总体比例(π)，等等。除了样本均值、样本比例、样本方差这类统计量，还有一些是为统计分析的需要而构造出来的统计量，比如用于统计检验的 Z 统计量、 t 统计量、 χ^2 统计量、 F 统计量，等等，它们的含义将在后面相关的章节中再作介绍。

5. 变量

在研究总体时，重点关注的是总体单位具有哪些重要的特征或属性，我们把这些特征称为变量(Variable)。变量是总体中个体单位所具有的特征或特性。“变量”的名称是针对总体中每一基本单位的属性都存在着差异而言的。例如，被