21世纪高等院校
经济管理类规划教材

Applied Statistics

应用统计学

◎主编　颜节礼　唐建荣

西安电子科技大学出版社
http://www.xduph.com

21 世纪高等院校经济管理类规划教材

应用统计学

主编　颜节礼　唐建荣

西安电子科技大学出版社

内 容 简 介

本书以大学本科、专科学生为读者对象，介绍了经济、管理和社会学研究中的主要统计方法，适合作为统计学的入门教材。本书从统计数据的搜集和整理开始，介绍了统计中重要的数据特征值分析、统计图以及统计表的制作；在概率论基础知识的基础上，通过理论讲解、例题分析以及软件操作的方式，重点讲解了推断统计中的参数估计、假设检验、卡方检验、方差分析以及相关回归分析的理论和方法。为了增强本书的可读性和适用范围，例题和课后的复习题涉及经济学、管理学、财务、自然以及生物医学等多个方面，希望能对读者掌握具体的统计数据处理方法和应用统计方法解决实际问题有所帮助。

图书在版编目(CIP)数据

应用统计学/颜节礼，唐建荣主编.
—西安：西安电子科技大学出版社，2016.1
21世纪高等院校经济管理类规划教材
ISBN 978-7-5606-3990-1

Ⅰ. ① 应… Ⅱ. ① 颜… ② 唐… Ⅲ. ① 应用统计学—高等学校—教材 Ⅳ. ① C8

中国版本图书馆CIP数据核字(2016)第002839号

策　　划　高　樱
责任编辑　阎　彬　王文秀
出版发行　西安电子科技大学出版社(西安市太白南路2号)
电　　话　(029)88242885　88201467　　邮　　编　710071
网　　址　www.xduph.com　　电子邮箱　xdupfxb001@163.com
经　　销　新华书店
印刷单位　陕西华沐印刷科技有限责任公司
版　　次　2016年1月第1版　2016年1月第1次印刷
开　　本　787毫米×1092毫米　1/16　印张17
字　　数　414千字
印　　数　1～3000册
定　　价　29.00元
ISBN 978-7-5606-3990-1/C
XDUP　4282001-1

前　言

随着大数据时代的到来，统计学作为现代科学研究的重要方法论工具，其作用更加突出。在生物医学研究、社会学研究以及经济决策中，人们已经越来越依赖于现代统计方法。从纷繁复杂的数据中得到科学的结论、进行科学的决策，这无论对于理论研究者还是社会实践者而言都是必备的基本素质。

本书是统计学的入门教材，我们在写作过程中有两个基本原则。第一个原则是注重统计方法的思维逻辑而非简单的数理推导。统计学中的数学公式很多，每一种统计方法背后都隐藏着归纳推理的认知逻辑，数理推导仅是实现过程，对于初学者而言，正确理解思维逻辑要比数理推导更为重要，毕竟仅仅注重数理模型而误用统计学方法的事例并不鲜见。第二个原则是注重统计方法在具体软件中的应用。作为初学者，操作也很重要，能够将所学的统计方法在软件中学以致用，不仅是写作《应用统计学》这本书的目的，还能够激发读者进一步学习更复杂的统计方法的兴趣。

从初等统计学的教学要求而言，本书的写作主要包括三大块内容：第一部分为1～4章，是统计学的基础知识，主要讲解统计数据的类型、来源和描述统计数据分布的基本方法；第二部分为5～8章，主要讲解基于概率论知识的推断统计，包括区间估计和假设检验的基本思想和操作方法；第三部分为9～12章，主要讲解研究变量关系的基本方法，包括检验品质变量相关性的卡方检验，检验品质变量与数值变量关系的方差分析，检验数值变量之间关系的相关回归。在本书最后一章我们介绍了统计指数方法，这一章相对其他章节是独立的，但由于指数是经济分析中的一个重要工具，因此在写作时我们将这一部分内容也包含在内。

本书面向统计学初学者编写，可作为经济管理、社会学相关专业的专业教科书，也可作为工科类、医学类的相关学生学习统计方法的专业参考书。

本书的编写工作主要由颜节礼老师完成，主审工作由唐建荣老师完成，软件的应用部分得到了杨一兵老师的帮助和支持。在此向在本书写作过程中提供帮助的各位同事表示真诚的感谢。

本书的出版得到了西安电子科技大学出版社的大力支持，在此对出版社相关工作人员表示诚挚的感谢！

限于能力和时间，书中难免有疏漏或不当之处，还望各位同仁和读者多多批评指正，我们将努力修改完善，以期再版时修订。

编　者

2015 年 8 月

目　录

第1章 绪 论

统计能告诉我们什么？下面我们通过生活中的几个例子来简要说明。

假如有人问你，上海市民的收入状况如何，你该怎样回答？这时统计数据和数字的对比比任何仅用语言文字的描述更加具象和易于理解。根据国家统计局上海调查总队调查，2014年上海市城镇居民人均可支配收入为47 710元。那么这一数据能告诉我们什么信息？这一数据反映的收入水平是高还是低？实际上统计数字的信息永远是通过数据的比较反映出来的。从纵向对比来看，这一数据较2013年上海市城镇居民的人均可支配收入43 851元增长8.8%；从横向对比来看，2014年全国城镇居民平均可支配收入是19 867.2元，上海市的水平是全国平均水平的2.4倍。除此之外，也可以与上海市农村居民家庭的人均可支配收入21 192元比，上海市城镇居民家庭是农村居民家庭人均可支配收入的2.25倍；当然也可以列出全国内地其他31个省、直辖市、自治区的城镇居民可支配收入进行比较，从比较中反映上海市城镇居民可支配收入的状况。如果我们需要进一步理解这一数据所反映的信息，可以结合结构分析，即可支配收入47 710元这一数据是如何构成的？根据调查，人均可支配收入47 710元中，工资性收入为30 629元，占64.2%；经营净收入为2345元，占4.9%；财产性收入为846元，占1.8%；转移性收入为13 890元，占29.1%。当然，也可以结合图形(如饼图)，把这一信息更加直观地展示出来。

其实，城镇居民人均可支配收入这一数据是经过对调查得到的原始数据进行加工处理得来的。那么如果对原始数据进行进一步分析，并关心收入这一变量的分布，我们可以得到处于不同收入水平的家庭占比，即通常我们所讲的收入分配。通过对原始调查数据的分析，不仅能掌握可支配收入的平均值，还可以得到收入分配的内部差异程度。如果需要进一步了解收入分配的影响因素，可以通过对收入这一变量与其他变量之间是否存在相关关系进行探索，如个人收入与个人的受教育程度是否相关，个人收入是否与个人所处的行业相关等。

总而言之，从以上例子中，我们发现通过数据的比较才能将统计数据所包含的信息充分展示出来，我们最关心的就是统计数据分布、构成以及变量和变量之间的关系。那么，在社会科学、自然科学中我们所研究的所谓“规律”，其本质就表现为变量与变量之间的关系。从哲学意义上讲，事物之间内在的、本质的、普遍的联系在现实中一定表现为变量与变量的关系，研究变量关系几乎是所有学科普遍的实证逻辑。

1.1 什么是统计学

1.1.1 统计学的定义

“统计”一词在我们的生活中经常出现，我们每天都会在电视、报纸或者网络上看到一些关于经济、社会、医疗等领域的统计数据。如2015年1月份，全国居民消费价格总水平同比上涨0.8%、2014年末我国劳动年龄人口91 583万人，占人口总数的67%等。通常，在有些人眼中统计学或者说统计只不过是数据的搜集汇总而已，其实这只是统计一词最原始的含义，统计学或者统计不仅仅是枯燥的统计数据，每一个统计数据背后实际反映的是活生生的经济、社会现象。现代科学无论是自然科学还是社会科学已经离不开统计学，统计学已经深深渗透到这些学科的研究当中，通过揭示统计数据的某些特性，往往能发现隐藏在现象背后的客观规律。

在生物遗传规律中，基因组合规律已经被现代人所熟知。然而在基因尚未被认知之时，现代遗传学之父孟德尔(Gregor Johann Mendel，1822－1884)的豌豆试验利用豌豆子代各种花色出现的几率，推断出基因组合的遗传规律，并提出了生物的性状是由遗传因子(Gene)控制的观点。这一研究方法是典型的通过分析观察数据来揭示隐藏在数据背后的遗传规律。

在人文社会科学领域内，一个很有趣的例子就是根据统计规律判断保存在牛津大学Bodelian图书馆的一首新诗的作者为莎士比亚。1985年11月，研究莎士比亚的学者泰勒从自1775年以来就保存在Bodelian图书馆的收藏中发现了写在纸片上的九节新诗。然而新诗只有短短429个字，也没有记载谁是诗的作者。这首诗会是莎士比亚的作品吗？最终答案由两个统计学家给出。统计学家Thisted和Efron(1987)利用统计的方法研究了这个问题，得出结论：这首诗用词的风格(规范)与莎士比亚的风格非常一致。这个研究纯粹基于统计学基础，他们根据莎士比亚所有著作的用词总数884 647个(其中31 534个是不同的)，统计了31 534个词的用词频率，经过与其他几位人们认为有可能的作家比较，发现新诗的用词频率与莎士比亚的用词频率最为接近。在这个研究的背后有一个潜在的逻辑——作者隐藏在用词频率这一统计数字背后的写作风格是基本固定的。

那么，我们如何给统计学下一个准确的定义呢？其实，给出统计学定义的学者很多，只是不同的定义侧重点不同而已。根据《不列颠百科全书》的定义：统计学是研究如何测定、收集、整理、归纳和分析反映客观现象总体数量的数据，以便给出正确认识的方法论科学。这一定义，比较清晰地说明了统计学的研究对象为数据，其本质是根据数量关系来说明客观规律。通常在统计学中所研究的数量不同于数学中的数字，统计学中的数据有明确的含义，所研究的数量关系一般包括数量大小、数量结构、数量的横向比较和纵向(动态)比较以及数量之间(不同变量)的关系。

根据《新韦氏国际英语大词典(第3版)》给出的统计定义：统计是一门收集、分析、解释和提供数据的科学。这一定义比较清晰地说明了统计的研究过程，即统计工作的四个阶段：统计数据的收集、统计数据的整理和分析、统计数据所反映的可能存在的统计规律的解释以及提供决策支持的统计数据。认识统计工作的一个完整的过程对于学习统计学理论

和利用统计方法解决实际问题都很有帮助。统计学家 Mario F. Triola 也给出类似的定义：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。这一定义更侧重于在自然科学研究中的统计测量和分析。

为了更好地理解统计学，在这里有必要说明三个相互关联的词：统计学、统计工作和统计资料。“统计”一词兼有上述三层含义，在不同的场合可能指三者之一。统计工作就是社会实践中搜集、加工、整理、分析统计数据，解释统计数据背后隐藏的客观规律的社会实践活动；统计学是学科范畴，是统计工作的理论总结和方法性的指导，对于统计数据搜集、整理、分析有一套科学规范的方法，是通过系统学习可以尽快掌握的理论知识体系；而统计资料就是统计工作的具体成果，包含统计数据、统计图表、统计分析咨询报告等。

1.1.2 统计学的研究方法

1. 描述统计和推断统计

统计学具体的数据处理方法多种多样，但总体而言，统计学的研究方法可归为两类：描述统计(descriptive statistics)和推断统计(inferential statistics)。描述统计是推断统计的基础，对总体规律的推断必须先从对样本数据的描述统计出发。

描述统计是研究数据收集、整理和描述的统计学分支。其主要内容包含：数据收集、数据整理、数据展示和数据的描述性分析。其目的主要是描述数据的分布特征、展示数据之间的关系和规律。

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分支。它通过观察样本数据而推断总体数据的分布形态或者总体变量之间的关系。其主要方法包括参数估计和假设检验。推断统计一般包含 5 个要素或者过程：

(1) 确定研究对象，研究对象也就是总体；

(2) 确定所要研究的总体变量；

(3) 收集样本数据；

(4) 对样本数据进行描述性统计；

(5) 根据样本数据的描述性统计，结合数理统计知识对总体进行推断以及对推断进行可靠性度量。

20 世纪初，概率论被引进统计学，产生了推断统计方法，其重要贡献就是使我们能利用正确选出的样本，对每一种统计推断提供一种可靠性度量。

2. 统计学研究方法的逻辑基础

作为一门方法论学科，统计学通过测定、收集、整理数据，对数据进行比较，归纳客观世界事物之间的普遍联系，统计学就是“依据反映客观现象总体数量关系的数据，以便给出正确认识”的方法论。因此，有必要从认识论的角度，分析统计学这种方法论的思维逻辑。归纳推理与演绎推理是人们在认识客观世界过程中两种紧密联系、互相依赖、互为补充，又有着显著区别的逻辑思维过程。从认识论的角度观察统计学的研究方法，统计学更偏重于归纳推理。

演绎推理是从一般到特殊(个别)的推理过程，依靠人们先前积累的一般性理论知识指导推论得出新的结论。典型的演绎推理形式即三段论，包括：大前提、小前提和结论。例

如，三角形的内角和等于180°(大前提)，图形ABC是三角形(小前提)，因此图形ABC的内角和等于180°(结论)。演绎推理的正确性受到大前提正确性的影响，如果大前提的正确性存疑，那么结论的可靠性就受到质疑。历史地看待人类的认识活动，马克思辩证唯物主义的认识论认为实践是人类认识的基础，人们认识活动的起点一定是前人认识活动的结果。演绎推理中的大前提必然是前人认识的结果，我们又如何保证前人认识活动的结果确切无误呢？人类的科学史告诉我们，科学的发展实际上是在不断否定前人认识活动结果中前进的。

归纳推理是从特殊到一般的推理过程，通过生产实践中观察到的具体个案，经过思维加工抽象出一般性的结论。例如，“天鹅都是白色的”这一结论就是通过多次观察的个案(样本)归纳出的一般性的结论。然而由于这种归纳是不完全归纳，因此可能会得出错误的结论，但是在人类认识和生产实践中完全归纳是非常困难的，因此统计就需要根据样本数据通过不完全归纳法得出有一定应用价值的结论，例如，根据样本实验田提供的产量和施肥量之间的关系推断出施肥量对产量存在正的贡献。显然，统计方法就是归纳推理，这种归纳推理是有一定风险的。

两种逻辑思维方法虽存在上述区别，但它们在人类认识世界的实践活动中是相依相济的。通常作为演绎推理一般性知识的大前提必须借助于归纳推理从具体的经验中概括出来，从这个意义上我们可以说，没有归纳推理也就没有演绎推理。当然，归纳推理也离不开演绎推理。例如，归纳活动的目的、任务和方向是归纳过程本身所不能解决和提供的，这只有借助于理论思维，而这本身就是一种演绎活动。而且，单靠归纳推理是不能证明必然性的，因此，在归纳推理的过程中，人们常常需要应用演绎推理对某些归纳的前提或者结论加以论证。从这个意义上我们也可以说，没有演绎推理也就不可能有归纳推理。

因此，演绎推理所得出的结论需要通过客观观察数据的验证，即现实是不是如此。归纳推理所得出的结论也需要演绎推理的结论来解释，即为什么是如此。两者结合才能得出较为可靠的结论。而统计学在认识客观实践中更多的是担当归纳推理的角色，尤其是推断统计，就是由样本规律归纳得到总体一般规律的思维过程。

1.1.3 避免误用统计学

通过分析统计学的思维逻辑，我们知道要正确地运用统计学就必须注意避免将客观世界简化成数学符号。认识到统计认识世界的过程绝不是枯燥的数字堆砌，统计学中的数字不同于数学中的数字，它们都是活生生的有血有肉的数字，统计学的经验模型也离不开每一门具体学科，统计模型一定建立在具体学科定性研究的基础之上。有统计学家提出过这样的观点：统计学基本上是寄生的，靠研究其他领域内的工作而生存，由解决其他领域内的问题而存在并兴旺发达。这也准确地说明了统计学与其他具体学科之间的关系。

数学可以离开其他学科而自成体系地发展，但是如果统计学离开了其他具体学科则成为无源之水。统计学所构造的经验模型反映的一定是经济社会和自然现象中的内在联系。例如，经典的回归分析一定要能够找到被解释变量与解释变量之间的因果联系，包括多元统计分析中的路径分析也反映事物内部的因果联系；因子分析中提取出的每个因子必须有具体的经济内容等。这些都反映了统计的这一特征。统计学的研究绝不可以将客观世界简化成数学符号的联系。

在西方有一句谚语："谎言，该死的谎言，统计数字(Lies，damned lies，and statistics)"，主要描述数字的说服能力，特别是用来讽刺一些使用统计数字支持、但毫无说服力的分析报告，以及人们倾向于贬低那些不支持其立场的统计结论。如果我们抛开了统计数据背后的经济实质或忽视了分析数量关系后面的内在因果关系，则会得到许多啼笑皆非的结论。例如，每天喝3杯以上咖啡的人心脏病发病几率会增加50%(数据是观察数据还是实验数据？这能说明喝咖啡和心脏病发作之间的因果关系吗?)、在教堂结婚的比例越高则人均预期寿命越低(由英国学者根据历史数据研究发现两者之间存在负相关，但是这种数据上的负相关又能说明什么因果关系?)等不一而足，如果不能解释这些现象之间的因果关系，那么这些分析也仅仅是茶余饭后的笑话而已。

1.2 统计学的产生和发展

如果将统计理解为关于数据的搜集、整理、分析和解释，那么可以说从结绳记事起就有了人类统计活动的萌芽，随着人类经济活动不断发展，统计实践活动也不断地展开。随着统计实践活动的发展，作为对于统计实践活动的理论总结，这期间统计学也出现了不同的研究流派。

1.2.1 政治算术学派

政治算术学派产生于17世纪的英国，其代表人物是资产阶级古典经济学家威廉·配第(1623—1687)。马克思高度评价配第的贡献，他认为："配第是政治经济学之父，在某种程度上可以说是统计学的创始人"，并称他为"最有天才的和最有创建性的经济研究家"。

15～17世纪的欧洲，在重商主义的理论指导下，认为财富或价值来自于流通领域，政策上主张国家通过贸易来积累国库的贵重金属。最早荷兰成为最强大的国家，后起的资本主义国家英国、法国在国内市场基本饱和的情况下，为了扩展到国际市场，和荷兰先后发生了战争，争夺殖民贸易等利益。在这样的历史条件下，威廉·配第为了给战争中的英国人鼓气，分析了三国(英国、法国、荷兰)间的经济实力，还提出了发展本国产业、开拓国际市场、增强殖民掠夺等政策和建议。在其书中大量运用了数字表达的方法，包括运用数字、重量、尺度计量和简单直观的图表。配第这种用数据分析说明经济问题的方法被认为是统计学不同于其他学科的本质特征，配第也被认为是经济统计学派的创始人。

1.2.2 国势学派

国势学派产生于18世纪的德国。由于该学派主要以文字记述国家的显著事项，故又称之为记述学派。其主要代表人物是海尔曼·康令(H. Conring，1606－1681)和阿亨华尔(G. Achenwall，1714－1772)。康令第一个在德国黑尔姆施泰特大学以"国势学"为题讲授政治活动家应具备的知识。阿亨华尔在哥廷根大学开设"国家学"课程，其主要著作是《近代欧洲各国国势学纲要》，书中讲述"一国或多数国家的显著事项"，主要用对比分析的方法研究了解国家组织、领土、人口、资源财富和国情国力，比较了各国实力的强弱。因在德文中"国势"与"统计"词义相通，后来正式命名为"统计学"。该学派在进行国势比较分析

中，偏重对事物性质的解释，而不注重数量对比和数量计算，但却为统计学的发展奠定了经济理论基础。但随着资本主义市场经济的发展，对事物量的计算和分析显得越来越重要，该学派后来发生了分裂，分化为图表学派和比较学派。

1.2.3 数理统计学派

数理统计学派由19世纪比利时数学家凯特勒(A. Quetelet)开创。他率先将概率论引入了统计学，开创了统计研究方法的新领域。在其著作中，凯特勒最先提出用数学中的大数定律作为分析社会经济现象的一种工具，通过大量观察个体事物的随机性来研究事物整体的必然规律。

在凯特勒之后，经过几代统计学家的研究，尤其是在以费雪(R. A. Fisher，1890—1962)为代表的一批学者的努力下，建立了相关回归分析、假设检验、χ^2(读作卡方)检验和t分布理论，使数理统计学逐渐成为一门独立的、完整的学科。

在现代统计学的发展中，一方面每个学派在不断地独立发展，另一方面各个领域内研究思想和研究方法也相互借鉴。因此，这也产生了另一个问题，由于社会统计专门研究社会问题，而数理统计学既研究社会问题也研究自然现象，如何对统计学的研究领域加以界定就产生了争议。同时，统计学到底是一门实质性学科还是一门方法论学科也是争论的焦点。然而现代统计学正是在不断争论中向前推进的。无论是现代自然科学还是社会科学，许多学科问题的研究已经离不开统计方法，统计学方法和理论也在解决其他学科问题的过程中不断丰富。

1.3 统计学研究中的基本概念

1.3.1 总体和样本

总体(population)是指客观存在的、在同一性质基础上结合起来的许多个别单位的整体，即研究对象的集合，是根据研究目的和要求所确定的研究事物的全体，总体也称母体。

总体是由每一个单位或个体组成的，构成总体的所有个体叫做总体单位。例如，考察某厂生产的灯泡的使用寿命，该厂生产的所有灯泡为总体，每个灯泡为一个个体。当总体中所含个体总数有限时，称为有限总体，否则，称为无限总体。

样本(sample)是在研究中根据需要从总体中抽取的部分单位组成的集合。统计分析的目的就是要对总体的特征、不同总体间的差异等做出推断，然而在实践中全面了解总体的情况往往难以办到。例如，要了解某一地区的年平均降雨量，总体本身就是无限总体。或者尽管是有限总体，但进行全面调查从经济上是不可行的，如要观测生产线上灯泡的平均使用寿命，不可能对所有灯泡进行破坏性试验，记录每一个灯泡的使用寿命。所以往往通过观测部分个体，以获得总体的信息。为了使样本能够正确反映总体情况，对总体的范围要有明确的规定，总体的范围称为抽样框；在抽取样本的过程中，必须遵守随机原则；样本的观察单位还要有足够的数量(样本容量)。

1.3.2 参数和统计量

参数是描述总体特征的量，包括反映总体数据集中趋势的总体平均数，如某一人群的平均体重、平均身高等；还有反映总体数据变异程度的总体方差；反映不同总体变量相关关系的相关系数等。

统计量是描述样本特征的量，如样本平均数、样本方差、样本相关系数等。统计量可以由样本观测值计算得到，因而是样本观测值的函数。一般来说，每一个总体参数都有一个对应的样本统计量。因而由样本推断总体也可以理解为由统计量推断参数，见图1-1。

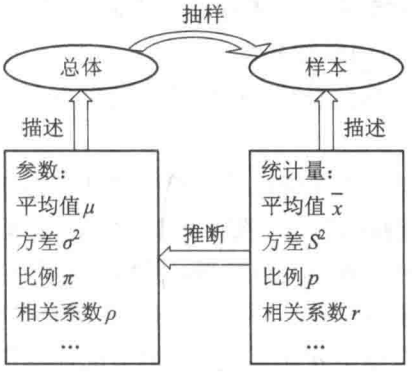

图1-1 总体与样本、参数和统计量之间的关系

1.3.3 变量和变异

变量是指总体单位的标志，如人的性别、年龄，企业的产值、利润额等。总体单位的标志值在各单位间是有差异的，这种存在差异的性质叫做变异。变异的大小反映了总体在某个标志上的同质性或差异程度的大小。变量值和变量值之间的差异、变量值之间的关系、变量值的分布规律都是统计研究的客观对象。

1.4 统计学研究中变量的分类

统计学所研究的变量，根据不同的分类标准有不同的分类方法。不同类型的变量其表现形式和特征不同，其统计数据处理的方法存在很大差异，这也就是统计数据分类的必要性。

1. 按照统计数据的计量方法分类

按照统计数据的计量方法分，统计学中所研究的变量可分为分类变量、顺序变量和数值型变量。

在社会经济研究中，很多变量只表示事物的分类而不能用数字来表示，例如，性别用男、女表示；宗教信仰用基督教、佛教等表示；大学生最喜欢的运动项目用游泳、足球等表示。这类变量称为分类变量。

顺序变量一般情况下也用文字来表示，不同的是这类变量值存在优劣先后之分。例如，客户对酒店餐饮服务的质量评价可以是非常满意、满意、一般、不满意和非常不满意。这种变量的结果本身存在优劣之分。一般情况下，这类变量在问卷设计中可采用两分法

(即满意、不满意)、三分法(满意、一般、不满意)或者更多的五分法、七分法等。又如对学生的综合评价可分为优、良、中、及格和不及格等。

数值型变量是最常见的变量。数值型变量可以用数值来表示，如身高、体重、收入等。

在统计研究中，计量层次较高的变量可以当做计量层次较低的变量使用，但是会损失一些信息。例如，在研究某地区的家庭收入状况时，根据当地经济发展水平，家庭年收入“小于 20 000 元”可以定义为“低收入”；“收入 20 000 到 49 999 元”可定义为“中低收入”；“50 000 元到 149 999 元”可定义为“中等收入”；“150 000 元到 499 999 元”可定义为“中高收入”；“500 000 以上”定义为“高收入”。这就是一个数值型变量用作顺序变量的例子。但是低层次计量的变量一般不能用作高层次计量的变量。

2. 按照统计数据反映的时间特征分类

按照统计数据反映的时间特征分类，统计学中所研究的变量可分为横截面数据、时间序列数据及面板数据。

横截面数据是在同一时间、不同统计单位相同统计指标组成的数据。横截面数据是按照统计单位排列的。因此，横截面数据不要求统计对象及其范围相同，但要求统计的时间相同。也就是说必须是同一时间截面上的数据。例如，2015 年某城市家庭收入抽样数据由 10 000 个家庭组成，统计对象和范围不同，但是对于“家庭收入”这一指标的“口径”必须相同，家庭的顺序对于数据组织也不存在影响。传统的统计分析一般认为横截面数据来自一个总体分布未知但一致的总体，一些经典的统计估计方法也是基于这一假设。

时间序列数据是在不同时间点上收集到的数据，这类数据反映某一事物、现象等随时间的变化状态或程度。例如，通常会按照年份将某一地区的 GDP 排成一列，此时统计数据是按照时间先后排列，在分析时也不能打乱排序。对于时间序列数据，可以认为每期的观察值来自于一个未知分布总体的随机观察，但是，对于不同期数据来自分布形态相同的总体这一假设往往不成立，因此时间序列分析必须发展出新的分析统计方法。时间序列数据直观的表示方法一般是在坐标平面上，横轴表示时间而纵轴表示变量数值。

面板数据是横截面数据与时间序列数据综合起来的一种数据类型。其有时间序列和截面两个维度，当这类数据按两个维度排列时，是排在一个平面上，与只有一个维度的数据排在一条线上有着明显的不同，整个表格像是一个面板，所以把 panel data 译作“面板数据”。面板数据的优点显而易见：既可以观察同一单位在不同时间上变量的变化趋势，也可以分析同一时间上不同单位变量值的分布规律。

3. 按照统计数据取得的方式分类

按照统计数据取得的方式分类，统计学中所研究的变量可分为观察数据和试验数据。

试验数据往往是在某些可控条件下，通过试验取得的数据；而观察数据通常是在不可控条件下通过观察取得的数据，社会经济科学研究中更多的要依赖观察数据。两种数据在分析得到结论时的差别是显而易见的。在试验中，试验环境是受到严格控制的，数据的产生一定是某一约束条件下的结果，往往通过控制条件变量观察结果变量的变化，一般来说，结果变量的变化除随机因素外可以主要解释为由条件变量变化引起的，在自然科学研究中试验的方法应用非常普遍。而在社会经济研究中，结果变量的变化受多种不可控变量甚至是未知因素的影响，解释结果变量变化的原因要相当慎重，避免错误地归结因果关系。

本章小结

一、本章主要概念

本章主要概念包括：描述统计和推断统计，数据，定性数据和定量数据，观察数据和试验数据，总体和样本，参数和统计量。

二、本章主要方法

(1) 根据统计数据类型的定义，判断社会经济研究中所关心的变量属于哪一类。

(2) 识别社会经济研究中的总体、样本、参数和统计量。

本章复习题

一、简答题

1. 应用统计的应用可以分为哪两部分？

2. 推断统计包含哪五个要素或过程？

3. 统计数据类型是如何分类的？

4. 统计学的目的是什么？

5. 推断统计学的主要贡献是什么？

二、单项选择题

1. 指出下面的哪一个数据属于分类数据(　　)。

A. 年龄　　B. 工资　　C. 汽车产量

D. 购买商品的支付方式(现金、信用卡、支票)

2. 指出下面的哪一个数据属于顺序数据(　　)。

A. 年龄　　B. 工资　　C. 汽车产量

D. 员工对企业某项制度改革措施的态度(赞成、中立、反对)

3. 某研究部门准备在全市200万个家庭中抽取2000个家庭，据此推断该城市所有职工家庭的年人均收入，这项研究的统计量是(　　)。

A. 2000个家庭　　B. 200万个家庭

C. 2000个家庭的人均收入　　D. 200万个家庭的人均收入

4. 了解居民的消费支出情况，则(　　)。

A. 居民的消费支出情况是总体　　B. 所有居民是总体

C. 居民的消费支出情况是总体单位　　D. 所有居民是总体单位

5. 统计学研究的基本特点是(　　)。

A. 从数量上认识总体单位的特征和规律

B. 从数量上认识总体的特征和规律

C. 从性质上认识总体单位的特征和规律

D. 从性质上认识总体的特征和规律

6. 一家研究机构从IT从业者中随机抽取500人作为样本进行调查，其中60%的人回