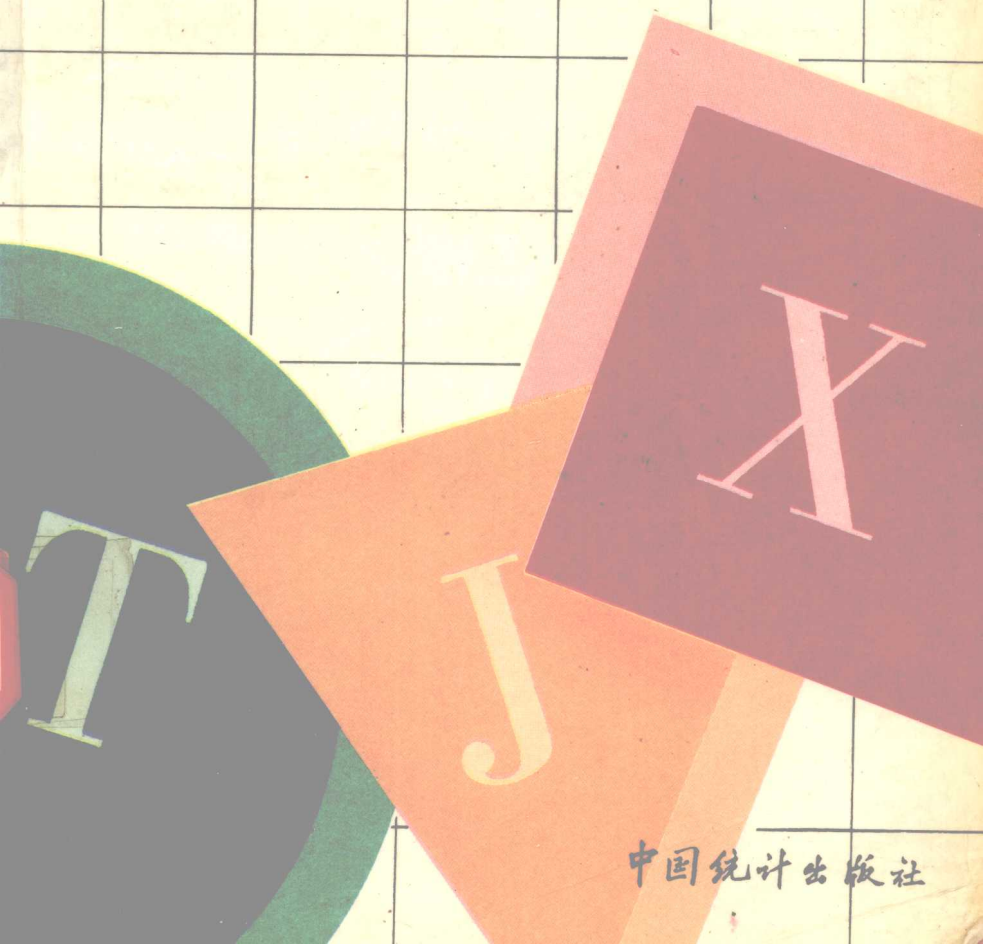


29.9
W36

统计学

王寿安 主编



中国统计出版社

统 计 学

王寿安 主编

中国统计出版社

(京)新登字 041 号

统 计 学

TONGJIXUE

王寿安 主编

*

中国统计出版社出版

新华书店经销

北京顺义振华印刷厂印刷

*

850×1168 毫米 32 开本 13.5 印张 34 万字

1994 年 2 月第 1 版 1994 年 2 月北京第 1 次印刷

印数: 1—4696

ISBN 7-5037-1450-6/C · 848

定价: 14.00 元

前 言

作为认识社会现象和自然现象有力武器的统计学，是一门历史悠久、发展迅速、前景广阔的方法论科学。它的研究范围之广泛，决定了它的内涵之丰富。对于刚刚结束文化基础知识学习进入高等院校学习专业知识的大学生和既无统计理论亦无统计实践但具有高中文化程度而欲学习统计基本知识的读者，究竟应该向他们提供什么样的教科书为宜呢？我们认为这是一个关系到我国统计理论水平和统计工作质量如何尽快与世界同步、如何使统计尽快适应改革开放的需要、促进我国现代市场经济体制的建立和完善的大事情。启蒙教育（或普及教育）是严肃的、神圣的。向读者提供的知识必须通俗而不失科学，浅出而又能深入，尽管是凤毛麟角而又可以领略全貌。不致于在学习了厚厚的“社会经济统计学原理”和众多的“部门统计”之后仍然对于“什么是统计”不甚了然，好似莽莽草原被群马一番腾踏，固然痕迹斑驳，却是杂乱无章，颇有“人生识统计糊涂始”之感。教学两难的教训和形势发展的需要是我们撰写这本《统计学》的动机和目的。

问题的关键在于向读者提供一本“通用”的统计学教科书，能够使读者基本了解“什么是统计学”。为此，我们在编写本书时着重注意到了以下四个问题；或者说，本书具有以下四个特点：

第一，统计学的研究对象——随机变量的统计规律。

研究对象是产生和形成一门科学的基础。统计学的研究对象是随机变量的统计规律。客观事物在量的表现上分为两类：数学变量和随机变量。数学变量是随机变量的特殊情形，而随机变量则带有普遍性、绝对性。所以，比较数学而言，统计学对客观世界的认识作用更为广泛。也正是在这个意义上，我们才说“统计是认识世界的有力武器”。

世界是物质的，物质是运动的。统计学通过随机变量刻画物质世界的运动规律（即统计规律），其结果便形成各种各样的统计分布。统计分布就是描述随机变量统计规律的数学模型。而统计分布又是通过一系列特征数来表征的，所以，认识了统计分布的某些个主要特征数，便在一定程度上认识了统计分布，也就是说，在一定程度上认识了客观事物量的规律性，进而认识其本质规律，从而完成统计学认识客观世界的根本目的。因此，“统计学的研究对象是随机变量的统计规律”这个命题是统计学的纲，而其他内容则是统计学的目。抓住了这个纲，不仅使读者易于掌握统计学的全部内容，更重要的是便于读者理解“什么是统计”。正因为如此，我们在编写本书时，不仅在第一章着重强调论述统计学的研究对象是随机变量的统计规律，而且在全书各章节中均以随机变量为研究对象展开讨论。

第二，统计学的功能——统计推断——与概率论。

那么，统计学是通过什么手段来认识随机变量的统计规律进而完成认识客观事物的本质规律这一根本任务的呢？我们说，统计学主要通过统计推断实现对于客观世界的认识。如果说，古典（或传统）统计学的功能主要是描述事物的局部和现状的话，那么，现代统计学的功能则主要是推断事物的总体和未来。因此，现代统计学比较古典统计学是一个巨大的进步，或者说是一个质的飞跃。统计意义上的总体即是总体分布（或曰理论分布），而人们所能认识的往往是样本分布（或曰经验分布）。推断就是通过样本分布来认识总体分布。如前所述，分布是通过其特征数来表征的，所以，推断实质上就是通过样本特征数来认识总体特征数。

统计推断既然是通过样本特征数来认识总体特征数，那么就决定了其逻辑方法必然是归纳法而不是演绎法。因为归纳法就是由个别的或特殊的情形求导一般化结论的逻辑方法。归纳法的性质决定了由统计推断所得到的一般化结论（即总体认识）必然只有若干成数的可靠性，即所得之总体认识具有一定的概率。由此

可知为什么在学习统计推断知识之前必须学习概率的一般理论。概率论是统计推断的指导思想和理论基础。概率论的引进使得统计学发生了一次革命。没有概率论就没有统计推断，没有统计推断就没有现代统计。明白了这一点，不仅使读者易于掌握统计学的结构体系，更重要的还在于能够帮助读者领会统计学的内容实质。

第三，实用性。

考虑到统计学是一门方法论科学，因此在编写中基本上回避了对于公式的推导，而将重点放在统计方法的定性阐释和具体运用上，以期使读者通过对于案例的学习而将认识上升到理性的高度，并能举一反三，触类旁通，将所学书本方法用以解决实际问题。

第四，系统性。

由于本书为普及性教材，读者为初学者，我们认为有必要按照统计工作的三个阶段——统计调查、统计整理、统计分析——来安排和介绍统计学的有关内容。这样作有两个好处：首先是有益于读者理解“什么是统计”——从内涵方面讲，统计就是通过样本信息推断总体；从外延方面讲，统计包含统计理论、统计资料、统计工作三个方面的内容，而统计工作又分为统计资料的搜集（调查）、统计资料的整理和统计分析三个阶段。其次是有益于读者了解统计推断之前的基础工作——统计资料的搜集与整理。准确、全面、及时的统计资料是使统计分析（主要是统计推断）的结论具有科学性的基础。因此，初学者必须了解和掌握统计资料搜集和整理的一般方法。

诚然，统计分析是统计学的主要内容，构成统计学的主体。统计分析方法如春天之花苑，千姿百态，繁荣兴旺，并且还在不断发展。本书所介绍的统计分析内容，只是其中最基本的一小部分。

本书除主编王寿安同志编写第一章、第四章外，参加编写的还有向书坚同志（编写第二章、第三章、第七章、第九章）和刘

建龙同志（编写第五章、第六章、第八章）；邹顺华教授主审。

我们的观点是：只有一门统计学，它是以随机变量的统计规律为其研究对象的方法论科学；它只有初级、中级、高级之分，而不应有“社会经济统计学原理”与“数理统计学”之分；至于所谓“部门统计”，实在只是一定社会形态下政府管理社会经济活动的一些统计规定。在这样一种认识前提下，我们企图向读者提供一本通用方法论性质的统计学。对于初学者而言，除了使其初步了解统计学的有关基础知识而外，还想使其掌握一些常用基本统计方法。我们力求使本书在内容和体系上具有新意，易学且实用。但心有余而力不及，错误与不足在所难免，恳请初学者及广大同仁不吝赐教。

编者

一九九三年十月

目 录

第一章✓ 绪论	(1)
第一节 统计学的哲学认识	(1)
第二节 统计学的传统方法	(13)
第三节 统计学的几个基本概念	(22)
第二章✓ 统计资料的搜集与整理	(30)
第一节 统计资料搜集方案	(30)
第二节 搜集统计资料的组织形式与方法	(33)
第三节 统计资料整理	(40)
第三章 频率分布的特征值	(57)
第一节 集中趋势特征值	(57)
第二节 离散趋势特征值	(84)
第三节 偏度与峰度	(96)
第四章 概率与概率分布	(105)
第一节 随机事件与概率	(105)
第二节 随机变量及其概率分布与数字特征	(122)
第三节 几种常见的随机变量分布形式	(133)
第四节 极限定理	(148)
第五章 抽样估计	(152)
第一节 样本单位的抽选方法	(152)
第二节 抽样分布	(165)
第三节 参数估计	(187)
第六章 假设检验	(209)
第一节 参数假设检验与非参数假设检验	(209)
第二节 总体平均数的假设检验	(217)
第三节 总体成数与总体方差的假设检验	(230)
第四节 常用非参数统计检验方法	(235)

第七章 ✓ 相关分析与回归分析	(247)
第一节 相关分析	(247)
第二节 一元线性回归分析	(264)
第三节 多元线性回归分析	(292)
第八章 ✓ 时间数列分析	(308)
第一节 时间数列动态描述指标	(308)
第二节 时间数列长期趋势的测定方法	(331)
第三节 季节波动及循环波动的测定	(348)
第九章 ✓ 统计指数	(361)
第一节 统计指数概述	(361)
第二节 统计总指数编制方法之一——加权综合法	(364)
第三节 统计总指数编制方法之二——加权平均法	(382)
第四节 指数体系与因素分析	(391)
附表	(403)

第一章 绪 论

第一节 统计学的哲学认识

统计学是一门以推断为其特征的方法论科学。它产生于统计工作实践,由于概率论的引进而发生质的飞跃,从而使统计学由古典阶段进入近代阶段。随着人类社会的发展和科学技术的进步,作为方法论的统计学还在继续发展,不断丰富。

统计学之所以发展如此迅速,生命力如此强大,主要原因在于它在应用上的广泛性和内容上的多科性。我们称统计学是一门方法论科学,因为它是认识客观现象(包括社会现象和自然现象)的有力工具。通过各种不同的指标和各种不同的具体方法对各种不同的客观现象进行观察、计量和分析,从而形成统计学的各种分类,如社会经济统计学、数理统计学,以及各种分科的应用统计学。

然而,不论何种统计学分科,均具有共同的基础,即具有共同的基本原理和基本方法;这些基本原理和基本方法,便构成统计学的基础知识。本书的宗旨,就是企图向读者传授统计学的基础知识,以使大家对统计学有一个基本认识,从而奠定进一步研究统计理论和进行统计实践的基础。

统计学是一门融多种学科于一体的交叉学科,其间蕴含着丰富的哲学思想。为便于初学者领会其实质,掌握其方法,我们拟从哲学的角度出发,对统计学的涵义、研究对象和特征进行阐述。

一、统计学的涵义与研究对象

(一)“由部分推及全体”的思想,贯穿统计学的始终;由此出发认识统计学的主要功能——推断。

人们要改造客观世界，首先必须认识客观世界，把握客观世界的规律性。然而，由于时空条件和其他诸方面条件的限制，人们对于客观事物的了解，只能遵循由部分及整体的认识路线。

统计学是对事物的量的规律性进行定量分析的方法论科学。它所面对的，往往是单位数目很大、甚至是无限的总体。但是，我们不可能（或者不必要）观察总体的全部单位。根据由部分及整体的认识路线，只能（或者只须）观察一部分单位——样本，以样本所提供的不全面的信息为依据，运用一定的数学模型对样本的不确定性给予测度和归纳，从而对总体进行一定可信程度的推断。因此，我们可以初步定义统计学如下：

所谓统计学，就是以样本信息为依据，运用数学模型推断总体的方法论科学。

可见，统计学所遵循的对客观事物的认识思想，是由部分及整体的思想。在这个基础上，才易于理解和掌握统计学的基本原理和方法（诸如频率与概率及其关系，用统计量推断总体特征数，由经验分布求得理论分布，等等），才不难理解统计学的主要功能就是推断。

（二）从“客观世界是运动的，运动是有规律性的”思想出发，认识统计学的研究对象——随机变量。

·世界是物质的，物质是运动的，运动是有规律性的。物质的运动规律分为两类，即运动规律（亦称动态规律）和统计规律（亦称概率规律）。

运动规律，是指事物本质的规律，它毫无例外地适用于事物的所有个体。例如自由落体，便是遵循运动规律的典型例子，任何自由落体都遵循如下运动规律：

$$S = \frac{1}{2}gt^2$$

下面对统计规律进行一般说明。

1. 定义 所谓统计规律，简言之，就是通过对随机变量的大

量观察所呈现出来的事物的集体性规律，又称大数恒静规律。

集体性规律由事物的集体性质所规定，它与事物的单一个体的性质时而偶合，时而近似，时而简直没有什么联系。

2. 随机变量及其特性

如果在观察或试验中，某一个变量可能取得各种数值，但事先不能预知其确切数值，则称这种变量为随机变量。

随机变量有两个特性：

第一，变异性。这是说，随着试验或观察次数不同，随机变量可能取得不同数值，即其数值是在不断发生变化的。

第二，不定性。即在每次试验或观察之前，随机变量的数值是不能事先确定的。例如，“日用灯泡的寿命”便是一个随机变量，因为不同的灯泡有不同的寿命，并且不可预知每只灯泡寿命的确切数值。

我们把随机变量的上述特性叫做随机性。事物的随机性是由其各个个体来体现的。

3. 产生随机性的原因——次要因素（随机扰动因素）

客观世界作用于事物各个个体的因素分为两类：基本因素和次要因素。基本因素决定着事物的运动规律，而次要因素使事物呈现统计规律。人们所能认识因而能够控制的是基本因素，而大量的次要因素人们未能认识，因而也不能控制。但是，只要存在次要因素的影响，就必然会有所表现，其中包括量的方面的表现，这种量就是随机变量。所以，随机变量是表明随机现象的重要信息，是刻画随机现象的变量。通过对随机变量的研究，可以发现事物的统计规律，从而开辟了认识客观事物的另一条重要途径。例如，我们可以用下述模型来描述工业产品成本：

$$y = a + \frac{b}{x} + \epsilon$$

式中， a 为固定成本，主要包括工资、原材料、燃料及辅助材料费用； b 为待摊费用。在一定的待摊费用下，产量 x 越大，则 $\frac{b}{x}$ 越小。

一般说来， x 是可以控制的，因而可以事先确定其数值。因此，从理论上讲，因变量 y （产品成本）是一个可以预先知道其确切数值的变量。但事实上，企业在一定时期所计得的实际成本往往在根据 a 和 $\frac{b}{x}$ 所算得的成本（姑且称之为理论成本）附近波动，其波动程度因时因地而异。原因是在生产经营活动中存在着大量的随机因素（随机扰动）。我们用随机变量 ϵ 来描述这些随机因素。正是 ϵ 使得工业产品实际成本 y 呈现随机性。

那么， ϵ 的构成如何？其主要方面是什么？它在未来某一时期的数值会是多大？所有这些，均不能作数学控制；因此，无法预知其确切数值。于是，人们采用大量观察法，在对观察资料进行整理之后，即可发现影响成本的随机因素 ϵ 的主要方面，比如说是废、次品的损失费用。然后再对这个主要方面进行观察，便可发现 ϵ 渐渐稳定在、或者说静止在某一数值上。这就是所谓大数恒静规律，亦即随机变量的统计规律。根据这个特定的统计规律，我们就可以在一定程度上对工业产品实际成本的一般水平进行推断（或预测）。

4. 运动规律与统计规律的关系

基本因素和次要因素在时间和空间上的同一性，决定了自然界和社会经济领域一切事物的发展变化均表现为偶然现象（随机现象）和必然现象相互交织的统一过程。这就是辩证唯物主义认为事物的必然性和偶然性都是客观存在的、都是对立统一的根本出发点。必然性是在事物的内部联系中必然会产生东西，如灯泡必有寿命，工业产品必有成本，等等，这是事物的本质；偶然性是在事物的发展变化过程中表现出来的不具确定性的东西，它使事物在量的表现上呈现差异，而不能使事物发生本质的变化。所以，偶然性是必然性的表现形式和补充，必然性必定通过偶然性表现出来，二者密不可分。科学研究的目的，就是要发现反映事物本质的客观规律，这就必须排除偶然性的掩盖和干扰，为此必

须首先认识这种偶然性。于是，统计学应运而生。所以，统计学不是研究事物本质的客观规律，而是透过事物的随机现象，通过对随机变量的研究，从而发现事物的统计规律性，并将其应用于客观规律的认识和把握方面。诚然，运动规律和统计规律是两种不同性质的规律，这是它们对立的一面；但它们同时又具有统一的一面，那就是二者在完成对客观事物的认识上相辅相成、互相促进、互为补充。

现在，我们可以对统计学作如下更进一步的认识：

统计学是遵循由部分及整体的认识路线，依据样本信息，应用数学模型，采用归纳推理方法，研究随机变量的统计规律的方法论科学。

（三）在质与量的辩证统一思想下，正确认识统计学的意义和作用。

一切科学都是在研究事物的质的过程中产生和发展的。统计学也是一门密切联系着事物的质来研究事物的随机变量统计规律的方法论科学。为此，必须克服两种偏向。

第一种偏向：认为既然统计学是一门运用数学模型研究事物随机变量的方法论科学，那么联系事物的质就成为多余，更无“密切”可言。于是，不管事物的性质，不论区别不同事物的数量界限，而进行随心所欲的统计假设；滥用数学模型和统计方法，对样本资料进行纯数学论证和推断，以至得出与事物本质相悖的结论。这使得统计学的本来面目和真实价值蒙受极大损害。

第二种偏向：对统计学研究随机变量的意义缺乏正确和足够的认识；而且，认为随机变量不具普遍性。比如，在社会主义条件下，政府对于许多经济现象可以运用行政手段进行宏观调控，因而能够预知其数量变化。所以，这种变量就不是随机变量。对此，我们认为：

第一，所谓对事物的数量变化可以进行控制，因而能够预知其确切数值，只是在对该事物的性质已经有了基本认识和在可行

的条件下，为了更加全面深入地认识事物本质，研究事物的发展变化规律而采用的一种方法。但是，由于主观和客观两方面的原因，使得人们对事物在其发展变化过程中客观存在并且起着影响的许多因素不可能全部认识或者认识不够，因而不能对其进行有效控制和预知其确切数值。所以，只能作为随机变量运用统计学的理论和方法处理。按照马克思主义的哲学思想，人们对于客观事物的认识，永远不会停止在一个水平上；所谓可知，只能是有限的，暂时的，相对的。因此，不应该也不可能在此与不可知之间划一条绝对的界限。所以，不能断言客观世界任何事物的变化过程中绝对不存在随机现象。

第二，事实上，随机变量和数学变量作为对立统一体存在于一切事物的发展变化过程中。二者的区别仅仅在于：谁起主要作用，谁起次要作用。即便是物理学描述自由落体运动规律的公式 $s = \frac{1}{2}gt^2$ ，其数学变量 t 也并非毫无随机成分，如测时器本身的误差，观测者的观测误差，等等，都会使 t 带有不同程度的随机色彩。只是由于这些随机因素对 s 的影响微不足道，因此可以不必“计较”。所以，那种认为统计学是资产阶级唯心主义形而上学的东西，因而一概反对和摒弃的认识和态度是不足取的。统计学作为一门方法论科学，不仅应用于自然科学领域，而且应用于社会经济领域，其本身并不具阶级性和社会性。随着生产力的发展和社会的进步，统计学在各个方面，尤其是在社会经济领域的应用，必将愈益广泛而深入。

（四）在事物都是相互联系、相互依存的思想下，认识统计学与其它有关科学的关系。

1. 统计学与数学的关系

首先阐述统计学与数学的分支——概率论的关系。

概率论是以现实世界客观存在的随机变量为其研究对象的一门数学分科。这一命题包含两个要点：

第一，概率论与统计学的研究对象均为随机变量。并且，概率论的逻辑思维形式和有关基本原理为统计学所必需。这就确立了概率论在统计学中的重要地位，决定了概率论与统计学的密切关系。

第二，概率论属于数学，而非统计学。

(1) 研究对象是区分不同科学的主要标志，但不是唯一标志。尽管研究对象相同，但由于研究目的和研究方法不同，从而也就形成了不同的科学概念。

(2) 概率论的研究目的在于建立描述和解释随机现象的统计规律性的数学模型。统计学的目的则是利用数学模型（其中包括概率模型），根据样本而推断总体。其主要内容是：①拟定统计观察方法；②分析统计资料。它所研究的主要问题是：①未知分布函数的估计；②未知分布参变数的估值；③统计假设的检验。

(3) 在研究问题的方法上，概率论运用演绎法，统计学运用归纳法。

总之，概率论与统计学是两门科学，而不是一门科学。它们是一种利用和被利用的关系。因此，称这类教科书为“概率论与统计学”是确切的；而所谓“概率统计”是含混的，不科学的。

其次说明统计学与经典数学的区别。

统计学区别于经典数学的主要标志在于统计学具有推断功能。这是统计学得以产生、赖以生存和发展的基础。在统计学问题中，一切结论都由推断而来；因此，一切结论都具有概率。经典数学则仅仅研究数学变量的运动轨迹，揭示客观事物的运动规律。在数学问题中，一切结论都由演绎而来；因此，一切结论都似乎绝对肯定。

举一个说明统计学与数学的区别的生动例子。

利用多元函数的极值理论，根据某一试验的 n 对观测值建立函数关系（经验方程），在其理论根据和方法步骤上，微积分学与统计学几乎完全一致。但是，在微积分学中，只要求出了参数 a ，

b ，就确定了经验公式 $y=a+bx$ ，于是便完成了其最终任务。它也承认经验点与所作经验公式的轨迹一般都存在着偏差，运用这个线性模型进行预测实际上存在着误差，然而微积分学本身无法刻画这个误差。因此，只好假定这个线性函数是唯一确定的。

事实上，这个线性函数不是唯一确定的。因为线性函数的参数 a, b 不是唯一确定值，而是估计值；从而，函数值 y 也是一个估计值。所以，必须对 a, b, y 进行一定概率保证下的区间估计，于是就必须首先求导 a, b, y 所服从的分布并求得分布的主要特征数——期望和方差。所有这些，只有利用统计学的理论和方法才能圆满解决，而微积分学则是无能为力的。

2. 社会经济统计学与数理统计学的关系

统计学之所以有社会经济统计学与数理统计学之分、之争，一是由于社会由初级而高级发展之客观过程使然，二是因为不同国情所致，不必引以为怪，亦不必强求统一。在发达国家（如美国），只有一种统计学，其体系大同小异，只有因内容难易不同而存在初、中、高级之区别。在我国，虽然长期以来存在两种统计学，即社会经济统计学与数理统计学，但是，随着生产力水平的提高和社会的进步，统计学理论与实践的发展和变化趋势亦是很明显的：最初几乎是社会经济统计学占统治地位；后来是社会经济统计学与数理统计学并存；到现在，数理统计学的研究和应用已经有了长足的进步。所谓社会经济统计学，就是运用数理统计学的基本原理与方法研究社会经济领域的随机变量。这种应用随着社会经济的发展和科学技术的进步必将逐渐广泛和深入，直至二者高度相融，合二而一。而在目前，我们必须正视社会经济统计学与数理统计学并存的客观事实。

数理统计学与社会经济统计学的共同之处主要表现在：第一，数理统计学与社会经济统计学的研究对象都是事物的量的规律性。这里所指的量，均为随机变量；因而这里所指的规律性，当然都是统计规律性。