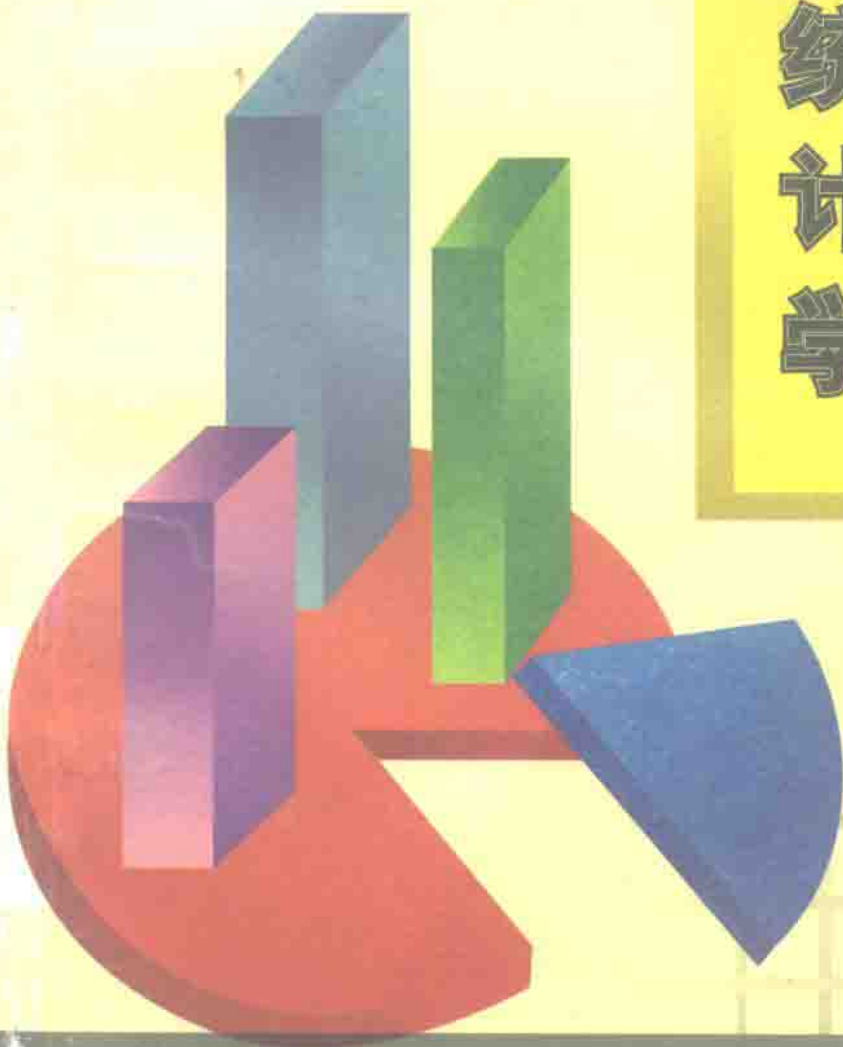


赵喜仓 路正南
吴向阳 编著

现代统计学

企业管理出版社



X I A N D A I T O N G J I X U E

图书在版编目(CIP)数据

现代统计学/赵喜仓等编著. -北京:企业管理出版社,1996.3

ISBN 7-80001-648-X

I. 现… II. 赵… III. 统计学 IV. C8

中国版本图书馆 CIP 数据核字(96)第 02891 号

现代统计学

赵喜仓 路正南 吴向阳 编著

企业管理出版社出版

(社址:北京市海淀区紫竹院南路 17 号 100044)

*

新华书店北京发行所发行

北京邮电大学印刷厂印刷

*

850×1168 毫米 32 开 8.625 印张 223 千字

1996 年 4 月第 1 版 1996 年 4 月第 1 次印刷

印数:7000 册

定价:12.80 元

ISBN 7-80001-648-X/F·646

序

理论指导实践,方法服务于实践是所有学科研究的目的,统计学研究也不例外。我国自改革开放以来,随着经济体制的转变和社会主义市场经济的建立,统计实践发生了巨大变革,对统计理论与方法也有了更新更高的需求,统计学的研究不能落后于实践,更要超前于实践,这就是我们统计教学与研究工作者在统计理论与方法领域中所肩负着的历史任务。

可喜的是我们统计学界的第三代人正在茁壮成长起来,他们扎根于不同学派交融的、世界与传统杂交系的肥沃土壤之中,加之这一代人的刻苦勤奋,必然会融多方精华于一体,开出时代的花朵,结出新世纪的果实。

本书便是他们努力攀登成功之路的有益试探。我读之后,觉得至少有如下几方面是应该肯定的:

第一,从结构上,突破了传统模式,突出了方法论性质,在内容上也有所创新。全书紧紧围绕数据的采集、传输、加工处理及开发利用这一体系展开,可分为三大部分。第一部分基本统计理论,包括总论、描述统计和推断统计三章,是全书的基础理论部分。总论中首先论述统计学的内涵,统计学的发展以及与计算技术相互渗透的辩证关系,统计必将向广度、深度和精密度发展的趋势,揭示了统计在认识世界、改造世界、促进经济发展和社会进步中的巨大作用。其次,介绍了统计数据的一般问题,包括统计数据的来源、统计数据与统计工作的关系、统计数据的质量及其控制、统计数据的计量、获取和分类问题。并阐述了统计学中的几个基本范畴,包括总体与总体单位、标志与指标、变异与变量。在第二章描述统计部分,将传统的内容(统计调查、统计整理、总量指标、相对指标、平均指标与变异指标、统计图表)浓

缩为一章。第三章推断统计并不是概率论与数理统计的重复,由于本书是以概率论为先修课程,只列举了对概率及概率分布的基本要求,以免重复,重点放在了参数估计与假设检验方面。第二部分统计方法,包括时间序列分析、指数、抽样法、相关与回归、平衡法、统计推算与预测、实验设计共7章,这是全书的核心部分。第三部分国民经济核算体系的基本原理,是本书的重要组成部分,主要适用于非统计专业教学。

第二,全书内在逻辑严密,结构清晰,语言精炼,通俗实用,可配合常用统计软件包进行辅助教学,有利于提高学生的综合技能,提高统计研究中融经济学理论、统计方法论及计算机应用于一体的能力,以适应市场经济对统计人才的需求。

由于作者试图在研究中有所创新,难免在探讨中有不足之处,如对总体分类的研究、对统计数据准确性的理解、对统计发展方向的探讨等等,均需进一步研究。

陈 东

1995年12月

前 言

读者只要稍加留意,就可以发现本书在整体框架和涉及内容的深度、广度上与面市的统计学教材相比给人一种耳目一新的感觉。

众所周知,统计学是一门关于统计资料收集、显示、描述和分析方法的学科。统计学包含的具体内容一般分为统计基础理论和统计应用方法两大部分。统计基础理论通常由统计学原理与数理统计学构成;统计应用方法通常包括社会经济(人口、资源、工商业等)的专业的统计调查、统计指标设计、计算分析方法,以及国民经济核算体系等。如果我们用以上的统计学的理论与方法结构来衡量,《现代统计学》在内容安排与取舍上是全面的、完整的、适度的。

《现代统计学》这本书在处理基本概念、基本理论方面是严谨的,而在一些设计方法的论述方面则有详有略,这是合理的。因为后继的课程如抽样调查、工业统计及各个部门统计、多元统计分析,以及经济预测与决策等,对涉及的统计方法还有更加深入、细致的研究与讨论。《现代统计学》的作者在设计本书结构框架时,注意到作为统计学的基础教材,它在整个专业课程体系中的地位和作用。

涉及经济与管理专业的各类教材,都面临不断变化的国内经济的改革浪潮,也面临不断深入的与国际经济接轨的开放形势,统计学专业教材也不例外,必须注意到教材内容的先进性、适用性、可操作性,以及与统计国际惯例的相

近性、一致性。本书在数据的计量尺度,田口实验设计方法,以及我国新国民经济核算体系(NEASC)等内容阐述上都体现了这些特色。

计算机的广泛应用推动了统计的现代化和现代化统计的兴起。计算机这一先进的工具已经贯穿统计活动的始终,并覆盖统计资料的调查、整理、加工的全过程,而且为数理统计的分析计算提供强有力的支持。作为本书附录的常用统计软件包的简介,也是本书的一大特色。只要有高等学校专业计算机应用课程的适当配合,读者熟悉和掌握这些软件是完全可能的。

有特色、有创新的教材是会受到读者欢迎的。这就是我对《现代统计学》的信念和期待。

李光久

1995年12月

目 录

第一章 总论	(1)
第一节 统计学的产生和发展.....	(1)
第二节 统计数据的一般问题.....	(4)
第三节 统计研究中的几个基本范畴	(13)
第二章 描述统计	(19)
第一节 统计数据调查	(19)
第二节 统计数据整理	(23)
第三节 集中趋势的度量	(29)
第四节 离中趋势的度量	(40)
第五节 统计图表	(45)
第三章 推断统计	(56)
第一节 概率与概率分布	(56)
第二节 参数估计	(56)
第三节 假设检验	(71)
第四章 时间序列分析	(82)
第一节 时间序列概述	(82)
第二节 时间序列的分析指标	(85)
第三节 长期趋势分析	(96)
第四节 季节变动的测定.....	(103)
第五章 指数	(110)
第一节 指数的意义和种类.....	(110)
第二节 综合指数.....	(112)
第三节 平均数指数.....	(118)

第四节	指数体系和因素分析·····	(123)
第五节	指数数列·····	(132)
第六章	抽样调查·····	(136)
第一节	抽样调查的意义和作用·····	(136)
第二节	抽样调查的基本概念·····	(138)
第三节	抽样调查的设计与实施步骤·····	(140)
第四节	抽样推断的一般原理·····	(143)
第五节	抽样调查的基本组织形式·····	(147)
第七章	相关与回归分析·····	(163)
第一节	相关分析的意义与种类·····	(163)
第二节	单相关分析·····	(165)
第三节	一元线性回归分析·····	(169)
第四节	多元线性回归分析·····	(173)
第八章	平衡分析·····	(175)
第一节	平衡分析法的意义和作用·····	(175)
第二节	平衡表法·····	(176)
第三节	帐户法·····	(179)
第四节	投入产出法·····	(183)
第九章	统计推算与预测·····	(189)
第一节	统计推算·····	(189)
第二节	统计预测·····	(194)
第十章	实验设计·····	(207)
第一节	实验设计技术的创立和发展·····	(207)
第二节	实验设计的几个基本概念·····	(208)
第三节	正交实验设计·····	(210)
第四节	田口实验设计方法原理·····	(220)
第十一章	国民经济核算体系的基本原理·····	(224)
第一节	国民经济核算体系概述·····	(224)
第二节	国民经济核算体系的描述原理·····	(226)

第三节 国民经济核算体系的一些基本概念和核算原则	(232)
第四节 我国国民经济核算体系的总体设计和基本内容	(235)
附录 1 常用统计软件包简介	(243)
附录 2 随机数字表	(255)
附录 3 正态分布概率表	(257)
附录 4 T 分布表	(260)
附录 5 常用正交表	(261)
参考书目	(263)
后记	(264)

第一章 总 论

第一节 统计学的产生和发展

一、统计学的定义

统计学是一门研究总体现象定量认识方法的科学，其目的在于探索客观现象内在的数量规律性。统计探索客观现象数量规律性的过程如图 1-1 所示。

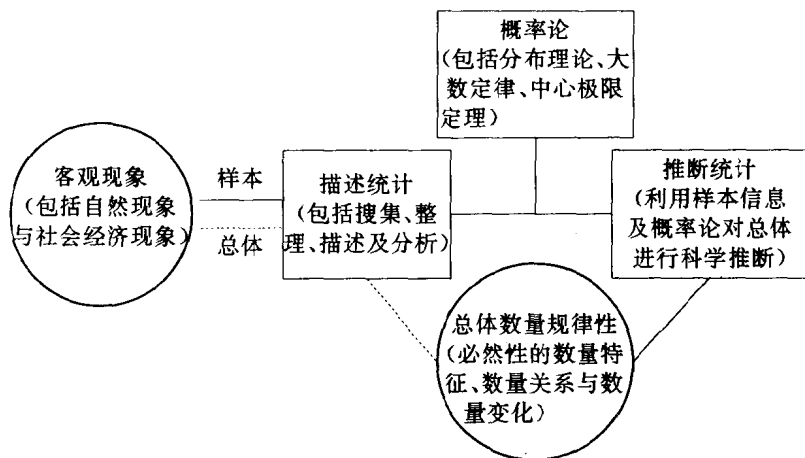


图 1-1 统计探索客观现象数量规律性过程框图

“统计”一词通常有统计工作、统计资料、统计学三种含义，它们既相区别，又相联系。统计工作是统计实践活动，包括设计制定统计指标体系，搜集整理资料和对资料进行分析研究，以及建立和加强统计组织管理等方面的工作。统计资料是统计工作的最终成果，指经过搜集、整理和计算分析以后所得的能够反映总体现象数量方

面的各种统计数据，常常以统计图表、统计年鉴等形式表现。统计学则是对统计实践活动经验的理论概括，它源于并不断发展于实践，又高于实践，指导统计实践。

二、统计学的发展历程

统计作为一种实践活动，在我国至少可以追溯到公元前 2000 多年前的大禹治水时期，在国外公元前 3000 多年前埃及建造金字塔时就已经有了人口普查的雏形。而统计成为一门科学却不过数百年的历史，统计学的发展大致经历了三个阶段。

第一阶段：统计学的初创阶段。随着商品经济的萌芽和发展，在 17 世纪的欧洲，先后产生了统计学的三个来源：德国的“国势学”、英国的“政治算术”及法国的“概率论”。从英国威廉·配第（William Petty）（1623—1687）的“政治算术”、英国约翰·格朗特（John Graunt）（1620—1674）的“人口统计”、德国海爾曼·康令（Hermann Corning）（1606—1681）的“国势学”和法国帕斯卡尔（Blaise Pascal）（1623—1662）的“古典概率”论研究到 19 世纪末英国卡尔·皮尔逊（Karl Pearson）（1857—1936）的“矩估行”与“ X^2 检验”，经过两个半世纪的漫长时间，描述统计与概率论的基本内容才逐渐形成。

第二阶段：统计学方法体系基本确定阶段。从 20 世纪初英国戈赛特（W. S. Gosset）（1876—1936）的“小样本 t 分布”，经过费希尔（R. A. Fisher）（1890—1962）的“F 分布”、“极大似然估计”、“方差分析”与“试验设计”，内曼（J. Neyman）和小皮尔逊（E. S. Pearson）的“置信区间估计”与“假设检验”到 20 世纪 40 年代末沃尔德（A. Wald）（1902—1950）的“统计决策函数”与“序贯抽样”，推断统计得到了长足的发展，概率论体系也日臻完善。

第三阶段：统计方法与应用研究全面发展阶段。从本世纪 50 年代起，统计学受计算机、信息论等现代科学技术的影响，新的研究领域层出不穷，如多元统计分析、探索性数据分析（EDA）、现代时间序列方法、刀切法与自助法、响应面设计、投影寻踪、图像处理、

人工智能等等。据美国学者统计，现代统计学以指数型加速度发展，新的研究分支每隔 17 年就会增加 1 倍。在这一阶段，涌现出了一大批杰出的统计学家，统计应用的领域覆盖了自然科学、农业科学、医药科学、工程与技术、人文与社会科学的各个学科，且其实用领域仍在不断拓宽。统计学的广泛应用，不仅促进了统计方法本身的发展，而且极大地推动了所应用领域学科的发展。

三、统计与计算机

计算机对现代统计发展的重要性怎么说也不过分。计算机技术引入统计学是统计科学发展史上的又一里程碑，其意义之深远当不亚于 A·凯特勒（1796—1874）当日之将概率论引入统计学和本世纪 20 年代以后国民经济核算体系的建立引起的统计调查方法和统计内容的革命。计算机对数据的高效传输、加工处理及辅助决策能力，增强了统计在帮助人们认识世界和改造世界过程中的作用，有利于统计的信息、咨询、监督职能的充分实现，成为统计生产力的重要组成部分。由于统计理论日趋复杂化，使得其应用需要完成的计算更加困难，因此，从这一意义上说，统计理论的发展和电子计算机的发展相互促进。每当统计学家提出描述和运用数据决策的新方法，计算机科学家便会相应地推出更新、更有效的实现这些运算的方法；反过来，随着功能更强的计算技术的出现，也激励统计研究人员探索新的更加精确的统计分析方法。在此顺便指出，统计决策实际上就是数据信息的运用；统计数据的准确性是统计工作的生命；计算机输出结果的质量取决于其输入数据的质量，“Garbage in, garbage out”（输入的是垃圾，输出的也是垃圾）就是这个道理。

总之，统计学随着统计认识的不断发展而向广度、深度和精度三方面发展，电子计算机对统计学发展的影响愈益巨大。随着电子计算机的发展和应用以及学科间的交叉渗透，统计应用的领域将会更加广阔，统计学将会有更大的发展。

第二节 统计数据的一般问题

统计数据是反映客观事物数量特征的数字信息，是统计科学研究的基本原料。对统计数据来源、分类、准确性及其计量尺度的探讨，是有效地测度客观现象、迅速获取有用信息以及提高统计认识水平的重要方面。

一、统计数据的来源及分类

1. 统计数据的来源

统计数据的来源无外乎亲自调查搜集和利用已发表的汇编材料两大渠道。就我国来说，目前主要来源于各类出版物、国家统计局部门和地方各级统计部门的统计报告制度以及专门组织的调查（如人口普查、全国工业普查等）。就世界各国和联合国经济统计工作的实际情况看，统计数据大致来源于四个方面：（1）各国政府统计机构连续编制的反映其全国及各地区国民经济和社会发展全貌的统计月报和年鉴；（2）各国社会经济科学研究者和政府社会经济研究所编写的社会经济问题专著或政府专业报告；（3）各国大学和研究机构进行科学研究而专门调查取得的专题性研究报告；（4）联合国等国际统计机构（或社会经济机构）编制的各种统计月报和年鉴。

2. 统计数据的分类

统计学家在其研究工作中通常利用横截面数据、时间数列以及纵列数据。

（1）横截面数据。指在同一时间（或时期）对不同单位（如个人、家庭、企业或国家）观测同一组变量的变化过程而得到的数据。例如 1993 年我国部分地区城镇居民家庭平均每人全年生活费支出数据，如表 1—1 所示。

（2）时间序列。指在不同时间（往往是等间隔的时间）对同一单位观测同一组变量的变化过程而得到的数据。例如中国社会经济发展的有关数据，如表 1—2 所示。资料来源同表 1—1。

表 1—1 1993 年我国部分地区城镇居民家庭平均每人全年生活费支出

单位:元

地区	生活消费支出	食品	衣着	医疗保健	交通和通讯	教育	住房
全国	2110.81	1058.20	300.61	56.89	80.63	95.35	52.93
北京	2939.60	1404.66	471.89	66.73	143.55	109.91	65.82
上海	3530.07	1867.90	413.83	67.87	199.28	92.49	64.07
江苏	2310.51	1141.36	313.46	32.50	98.24	84.41	73.83
浙江	2855.92	1411.66	364.94	80.33	145.02	89.12	83.71
山东	1946.83	895.77	349.20	48.30	54.18	97.90	48.46
广东	3777.43	1847.33	267.93	103.66	298.03	171.55	143.20
贵州	1876.24	1019.15	248.59	45.99	45.00	83.70	33.49
云南	2186.30	1066.99	289.71	86.87	73.97	91.01	51.28
甘肃	1679.74	851.77	283.61	61.34	44.82	72.08	26.03
青海	1869.60	970.45	281.86	87.15	53.09	112.49	10.45
新疆	1834.72	907.34	294.73	50.17	70.10	84.53	33.21

资料来源:《中国统计年鉴》(1994)

表 1—2 中国社会经济发展的一些情况(1985—1993 年)

年份	GNP (亿元)	GNP 指数 (上年=100)	总人口 (万人)	人口自然 增长率 (%)	全国零售 物价总指数 (上年=100)	全国居民 消费水平 (元)	进出口 总 额 (人民币亿元)
1985	8557.6	112.8	105851	14.26	108.8	403	2066.7
1986	9696.3	108.1	107507	15.57	106.0	447	2580.4
1987	11301.0	110.9	109300	16.61	107.3	508	3084.2
1988	14068.2	111.3	111026	15.73	118.5	635	3821.8
1989	15993.3	104.4	112704	15.04	117.8	694	4155.9
1990	17695.3	104.1	114333	14.39	102.1	723	5560.1
1991	20236.3	108.2	115823	12.98	102.9	803	7225.8
1992	24378.9	113.4	117171	11.60	105.4	947	9119.6
1993	31342.3	113.2	118517	11.45	113.2	1148	11271.0

(3)纵列数据。指在不同时间对同一组单位进行调查所得到的结果,它是横截面数据和时间序列混合起来的数据。例如表 1—3 给出的我国部分地区育龄妇女平均初婚年龄。资料来源同表 1—1。

表 1—3 我国部分地区育龄妇女平均初婚年龄

单位:岁

地区 \ 年份	1989	1990	1991	1992	1993
北京	24.45	24.38	24.44	24.28	24.77
上海	24.62	24.58	24.31	24.01	24.52
江苏	21.92	22.05	22.37	22.55	23.07
山东	22.91	23.10	23.37	23.75	24.22
广东	22.72	22.79	23.06	23.40	23.63
四川	21.29	21.42	21.53	21.71	21.82
贵州	21.52	21.77	21.85	22.20	22.11
青海	21.03	21.05	21.42	22.21	21.76

二、统计工作过程与统计数据

统计资料(数据)是统计工作的最终成果,是人类劳动的结晶。对统计工作过程的了解,有助于把握统计认识活动的全貌,把握具体统计资料的获取及开发应用过程。

统计是对一定总体的定量认识活动,就一次统计认识活动而言,其工作全过程包括统计设计、统计调查、统计整理、统计分析四个阶段,是从“定性——定量——定性”的过程。实事求是地反映客观现实,准确、及时、全面、系统地提供统计资料是对统计工作的基本要求。

统计设计是指根据统计研究对象的性质和研究目的,对统计研究对象的各个方面和统计工作的各个环节所做的系统考虑和安排。统计设计贯穿于统计工作的始终,其科学性直接影响着整个统计工作的效率。

统计调查是指根据统计研究的目的要求，运用各种科学的调查方法或方法组合，有序地搜集反映客观现象数量特征的统计资料的过程。统计调查为统计整理和统计分析提供基础资料，是决定统计工作质量的重要环节。

统计整理是指根据统计研究的条件和任务，对统计调查资料进行系统的加工和汇总，通过信息的合成和分解，取得能够表明客观现象总体特征的综合资料的工作过程。统计整理是统计工作的中间环节，是统计分析的前提。

统计分析是指对经过加工整理的统计资料进行分析研究和开发应用的过程，包括对客观现象数量规律性的描述、对客观现象的分析评价及前景预测等。统计分析是统计工作的决定性阶段。

统计工作是统计认识活动的具体化。经过统计调查和统计整理，人们取得了从静态与动态、从结构与功能、从系统与环境诸方面反映总体的数据资料，实现了从对个体认识到总体认识的过渡，为实现对客观总体的简单描述到数量规律性及科学推论奠定了基础。经过统计分析，通过对总体的观察研究，概括出个体的共性、数量规律性与差异性。这一升华过程的完成，取决于统计数据的准确性。

三、统计数据的准确性

统计数据的准确性是任何社会无法回避的问题，也是统计实践中最疾首的问题。国外有人曾将世间的说谎者依其罪孽轻重分为三阶，即：“说谎者，该入地狱的说谎者和统计学家。”统计学家列为世界上最大的说谎者。这种公众对统计的评价，指出了在统计研究和应用中存在着某些误用甚至有意乱用统计的问题，也指出了研究影响统计数字质量的根本原因及科学理解统计数据准确性的重要性。

统计认识中主、客体及不同利益主体间的矛盾，是社会矛盾在统计工作中的反映，它集中表现在统计数据准确性上。解决这一矛盾的有关理论是统计立法和统计组织的依据。对这一问题的深入研究，应引起足够的重视。在此，简要谈谈对统计资料准确性的理解。

统计资料的准确性是一个广义的范畴，应做全面的理解。首先应立足趋势性原则。通常所说的统计数据的准确性，其前提是把趋势把准，并不是包览无遗地绝对地准确。其次，准确性是相对的，误差是准确性的一面镜子。“实事求是”就是要通过大量反映客观实际的相对准确的数字，探寻隐藏在其背后的规律性，如果追求绝对的准确，忽视了统计的管理工具的性质，往往会陷于为统计而统计之中。第三，搞准统计数据，并不是要求全面搞准，而是要求把涉及国计民生、重大国情国力、影响社会经济运行及研究对象主要方面的数据搞准。第四，数据的准确性是与进行统计调查的经济性、时效性等联系在一起的，应用联系的观点看待准确性。第五，层次性。统计数据的准确性是指基础数据的准确性，应与用户、与不同的研究对象、不同的分析目的联系起来，管理部门把趋势搞准，具体搞核算的人应把数字搞准。

四、数据的计量尺度

统计数据是对客观现象进行测度的结果，它以多种形式出现，其所载信息量的大小取决于数据量度的形式。要明确数据的计量尺度（即测度的层次），首先要理解测度的含义。

1. 测度的含义及构成

从一般意义上讲，所谓测度是指按照某种法则给物体和事物分派一定的数字或符号，亦即把某种物体、事件或现象的非数状态转化为一种数字符号状态的过程。测度由测度的客体、数字或符号、分派数字或符号的法则三大基本要素构成。

（1）测度的客体。即所要测度的对象，通过测度可以刻划客体的某一（或某些）特征或属性。例如，对人进行测度，其属性或特征有性别、年龄、身高、职业等等。

（2）数字或符号。是对测度客体的某一特征或属性给予的一种标识。例如，我们用 1.75 米标识出一个人的身高，用“0”（男）或“1”（女）标识出一个人的性别，等等。

（3）测度法则。是对测度客体分派数字符号的统一标准或准则。