

官方统计数据 质量评估与控制

陶用之 著

新华出版社

前 言

统计数据质量是统计工作的生命 提高统计数据质量是官方统计工作的重要任务，也是统计理论研究和实践探索的重要课题。

“九五”期间我在主持国家教委人文社会科学研究规划项目《官方统计数据质量控制》研究中，思路主要从统计体制、统计调查方法和统计法制方面出发探讨统计数据质量控制的理论与方法 经过进一步思考，感到这种认识还存在局限性。统计数据质量是统计系统活动的产物，影响统计数据质量的因素有的来自统计系统内，有的来自统计系统外，就统计系统内而言，有统计技术的因素和非统计技术的因素，有统计方法的因素和人为的因素 就其现状而言，人为的干扰、破坏、弄虚作假等因素十分突出。决定的因素是统计活动的数据生成能力 因此，应该从统计活动的过程研究统计数据质量。

本书还认为，质量是全世界的共同语言，21世纪是质量的世纪，官方统计数据质量要跟上世界质量管理发展的趋势，应贯彻 ISO9000 标准，建立统计数据质量体系。

目 录

前言	(1)
1 导论	(1)
1.1 官方统计数据质量评估与控制的研究对象和方法	(1)
1.2 官方统计数据质量评估与控制研究的发展	(3)
1.3 本书研究的问题和结构安排	(5)
2 统计数据质量	(7)
2.1 统计数据质量	(7)
2.2 影响统计数据质量的因素	(13)
3 中国官方统计数据质量	(17)
3.1 旧中国统计数据质量	(17)
3.2 1949—1976 新中国统计数据质量	(23)
3.3 改革开放以来中国统计数据质量	(27)
4 统计数据质量评估	(38)
4.1 统计数据质量评估	(38)
4.2 统计数据质量评估方法	(43)
4.3 统计数据质量评估方法应用	(55)
5 官方统计数据质量控制	(64)
5.1 官方统计数据质量控制的基本问题	(64)
5.2 现代质量管理的基本思想和 ISO 标准	(69)

5.3	统计数据质量控制的基本原理	(74)
5.4	统计数据质量保证体系要素	(76)
5.5	统计数据质量保证体系控制要点	(79)
6	官方统计数据质量保证体系的建立与实	(91)
6.1	官方统计数据质量保证体系建立与实施的准备	(91)
6.2	统计数据质量保证体系总体设计	(94)
6.3	编写统计数据质量体系文件	(96)
6.4	统计数据质量保证体系的运行	(101)
7	普查统计数据的质量保证和控制	(104)
7.1	工业普查数据质量保证和控制	(104)
7.2	农业普查数据质量保证和控制	(111)
7.3	人口普查数据质量保证和控制	(121)
8	官方统计数据质量评估与控制的国际比较	(129)
8.1	世界各国统计数据质量	(129)
8.2	世界各国统计数据质量评估与控制研究内容	(133)
8.3	世界各国统计数据质量的法律保证	(136)

1 导论

- 1.1 官方统计数据质量评估与控制的研究对象和方法
- 1.2 官方统计数据质量评估与控制研究的发展
- 1.3 本书研究的问题和结构安排

本章作为全书的开始，需要介绍一下官方统计数据质量评估与控制的研究对象、方法、结构安排。

1.1 官方统计数据质量评估与控制的研究对象和方法

1.1.1 官方统计数据质量评估与控制的研究对象

官方统计数据质量评估与控制主要研究统计数据质量的评估和控制体系，它属于统计方法与管理学、系统科学的交叉形成的统计数据质量理论，即统计数据质量管理控制理论。

官方统计数据质量评估与控制究竟有那些内容需要研究呢？

(1) 统计数据质量的涵义。对于什么是统计数据质量，人们一般理解为准确、及时、全面就是质量，这是质量的狭义定义，它还有广义的定义

(2) 影响统计数据质量的因素。长期以来，人们比较注意影响统计数据质量的技术因素研究，对统计技术、统计误差的研究，对影响数据质量的非技术因素缺乏研究，重视不够。

(3) 统计数据质量现状。社会对统计数据质量的“水分”看法颇多，怎样看待当前的统计数据质量是应该回答的问题。

(4) 统计数据质量评估。统计数据质量评估是对统计数据质量开展分析和评价，估计其可靠性，包括评估方法和对一些主要官方统计数据的评估。

(5) 统计数据质量控制和保证体系。官方统计要树立以数据质量为中心，建立统计数据质量控制和保证体系，对统计数据质量实施全程质量管理、全域质量管理和全员质量管理，通过让社会 and 统计部门自身满意的管理，以及为实施这种管理需要的组织结构、程序、过程和资源，建立有效的官方统计体制。

(6) 统计数据质量的国际比较。

1.1.2 官方统计数据质量评估与控制的研究方法

统计数据质量控制研究方法有两大特征：一是统计方法的应用，统计方法是研究所有社会经济问题，如经济增长、刺激消费、社会保障等问题的一个基本方法，但与其他的社会经济问题相比，统计数据质量评估与控制更依赖于统计方法的应用与分析，对统计误差、不同统计方法应用与数据质量、数据评估等都要运用多种统计方法。在统计数据质量评估与控制研究中，统计方法既是研究的内容，也是一种方法的运用；二是把系统分析方法与统计学、经济学、管理学相结合，形成统计数据质量管理控制的方法。以前人们研究统计数据质量，多从误差、数据的准确性、抗干扰方面进行研究，在统计数据质量控制方法研究中，对于统计数据质量，应该从以下四个方面来考虑：

(1) 官方统计数据质量的标准研究。即对统计数据质量提出质量目标和质量标准,按照 ISO9000 标准组织统计数据的调查和数据处理。

(2) 官方统计数据质量的全程质量管理研究。即对从统计数据质量统计调查开始,直到最后对外发布统计数据的每一个环节实施严格的质量管理和监督。

(3) 官方统计数据质量的全域质量研究。统计数据生成是一个复杂的系统,对于系统的各个要素,要明确职责,合理分工,周密安排,科学管理,对数据质量实施最优化管理。

(4) 官方统计数据质量的全员质量管理。通过提高职业道德教育和质量管理技术培训工作,不断提高所有统计人员的综合素质和统计数据质量意识。

1.2 官方统计数据质量评估与控制研究的发展

从 20 世纪初叶开始,人们在研究统计误差时开始关注官方统计数据质量问题,其过程大致如下所述

(1) 20 世纪初至 40 年代,这个时期统计数据质量主要围绕调查数据的准确性、样本的代表性问题。官方统计数据质量也是围绕不同调查方法的误差等问题。第一次世界大战后,实际工作者们在许多应用中都发现各种非抽样误差如测量误差、无回答误差等对全部调查误差的影响比抽样误差的影响更重要、更明显,这就激发了人们对非抽样误差的研究和讨论。而这种研究实际上是随着政府进行大规模统计数据搜集而展开的,并在 30 年代~40 年代取得了重大的进展。当时最有影响的是美国国情普查局进行人口普查时对非抽样误差的研究,以及印度政府进行粮食产量估计时对非抽样误差的研究。其中美国国情普查局对非抽样误差的研究主要包括:1) 在普查报告中提出统计资料产生矛盾和

误差的可能原因；2)对普查数据的误差及准确性进行分析；3)研究如何利用一种“单独复核的方法”来计算误差。复核的内容包括：调查范围误差，如人口、住户和农场的遗漏，重复错列的误差，填报内容误差，如调查表中具体问题的漏报或误报等。

(2) 20世纪50年代至70年代，这个时期官方统计数据质量研究是围绕着统计调查误差模型的建立及由此展开的专项研究展开。调查误差模型是以整体的方式测定抽样误差和非抽样误差的影响。对调查模型的研究最早始于40年代末，1946年汉森(M. H. Hansen)和赫维茨(W. N. Hurwitz)提出了这个问题。50年代研究取得成果。60年代由汉森和赫维茨等首次提出完整的调查误差模型。这个模型对调查过程中误差的控制、评估奠定了重要的分析基础。

(3) 20世纪70年代末至今，官方统计数据质量的含义扩大，对统计数据质量的研究发展到如何建立有效的保证、控制、评估体系。随着官方统计数据搜集系统的日益制度化，抽样调查地位的确立及调查模型的建立，使对统计数据质量的研究趋势是研究调查设计的规范化、调查误差与调查费用的平衡等内容。统计数据质量含义得到扩展，由单纯的统计数据准确性扩展到比数据准确性更为丰富的内涵。戴伦纽斯(T. Dalenius)提出统计数据质量的“测量向量”，包括统计数据的准确性、经济性、保密性、相关性、时效性、数据详细程度等，他还提出数据质量的综合控制应包括设计调查控制和调查实施控制两部分。1980年联合国统计局出版的《统计组织手册》中，对官方统计资料提出八项要求，进一步明确了政府统计数据质量的综合含义。1988年美国田纳西 Oak Ridge 统计数据质量专题研讨会上，提出了统计数据质量环的一般内容，提出了统计数据质量管理方法体系的标准化。此后，不仅美国，而且世界上许多其他国家如意大利、瑞典的农业普查中都进行了统计数据质量的综合控制。相应地，如

何进行统计数据质量综合控制也成为重要的研究课题。

(4) 我国官方统计数据质量研究。新中国成立前，很少开展对官方统计数据质量研究。新中国成立后，官方统计数据质量研究主要围绕统计报表制度的实施展开，问题集中在如何保证统计数据质量准确性、及时性、全面性方面。

改革开放以来，我国统计在统计内容范围、统计方法技术、统计组织体系等各个方面都获得了前所未有的发展。特别是经济、社会、科技“三位一体”统计指标体系的建立健全，新国民经济核算体系的实施，统计调查方法体系的改革，统计信息处理技术手段的提高，1982年第三次人口普查中开始实施的全面质量管理方法，1995年提出“一靠科技，二靠法制，归根到底靠人才”的提高统计数据质量的措施等，这些统计实践活动的发展使统计数据在设计质量、统计调查的准确性、及时性、全面性、系统性等方面都有很大的提高，而且也推动了人们对统计数据质量的理论研究。这包括如何进行统计数据事后质量评估的研究和对统计数据全面质量管理的研究。

1.3 本书研究的问题和结构安排

1.3.1 问题的提出

官方统计数据质量控制是一个纷繁复杂的过程，目前对这个过程的研究，尚处在起步阶段，其中对数据质量的评估，影响数据质量的非技术因素以及建立统计数据质量体系的认识与现实有较大的偏差。

笔者认为，对中国官方统计数据质量应有一个基本的判断，不能曲解统计数据质量。在中国共产党领导下，官方统计数据质量基本上是可信的。我们处在一个科技革命日新月异、经济飞速

发展的时代。统计数据质量的含义也在不断扩展，为了适应信息技术革命和新经济带来的变化，应该树立统计数据质量的新观念，以满足社会对统计数据质量的更高要求。

提高统计技术水平，减少统计技术因素和非统计技术因素对数据质量的影响，是统计数据质量得到保证的一个重要方面。同时，更重要的是实行统计数据质量管理，贯彻 ISO9000 标准，建立统计数据质量保证体系，使统计数据事后检查、评估，变成从数据质量设计、调查、数据处理控制，保证数据质量。

我国质量控制理论研究一些领域处于国际领先水平，统计数据质量管理应借鉴和学习，发展我国官方统计数据质量管理和控制研究

1.3.2 结构安排

本书共分八章。第二章和第三章分别介绍了官方统计数据质量的一些基本问题，对影响数据质量的因素进行分析，对我国统计数据质量作一个基本评价；第四章介绍数据质量评估方法，如何对官方统计主要数据进行评估；第五章、第六章和第七章对统计数据质量控制体系进行研究，提出要按照产品和服务质量的概念，用 ISO9000 标准建立和实施统计数据质量体系，并且对我国在工业普查、农业普查、人口普查中实施统计数据质量控制进行分析；第八章是统计数据质量控制的国际比较研究，比较、借鉴国外统计数据质量控制理论和实践，提高我国统计数据质量水平。

2 统计数据质量

2.1 统计数据质量

2.2 影响统计数据质量的因素

统计数据质量是官方统计的生命。

影响统计数据质量的因素很多，但并非捉摸不定和不可控制。

2.1 统计数据质量

国民经济和社会活动离不开各种各样、形形色色的统计数据和统计数据服务。无论是统计数据还是提供统计数据服务，都必须满足人们的某种需要，满足程度不同，就产生了质量问题。统计数据质量的优劣不仅直接影响统计部门的形象，而且还会影响国民经济管理和政府决策，影响企业生产经营活动和人民生活。统计数据质量已被认为是一个国家管理水平高低的标志。

2.1.1 统计数据质量的涵义

什么是统计数据质量？长期以来，人们给统计数据质量下了不同的定义。

当今，质量已成为全世界的共同语言，质量代表一个国家的形象，一个民族的精神。我们国家在实现现代化过程中，把质量第一作为一个长期的战略方针，把提高质量作为一项基本国策。因此，应该从质量的深刻内涵认识统计数据质量。

质量的一般含义是随着社会发展变化和质量管理活动的深化而发展的。

何谓质量？美国质量管理专家米兰博士把新产品质量定义为“适应性”，即从用户角度看质量。另一位专家克劳斯比把产品质量定义为产品符合规定要求的程度，即从企业管理者的角度来看质量。现代管理科学把质量定义为反映实体满足明确和隐含需要的能力的特性总和，涵盖了产品的“适应性”和符合“规定性”两方面的内容。

现代管理对质量的定义给我们的启示是：

(1) 质量涵义具有广泛性，适用于各行各业。定义中的实体指可以单独描述和考虑的事物，可以是产品（包括硬件、软件、流程性材料和服务），也可以是过程、系统、人、活动、组织等。从这个意义上说，质量的意义也符合统计数据的质量。

(2) 质量涵义具有全面性。定义中的质量为满足“需要”，这种需要是一种全方位的，着重强调的是满足消费者和社会的需要。这种需要包括“明确的需要”和“隐含的需要”，明确的需要指标准、技术规范、法律、法律中已明确规定的，通过图样、技术要求、工艺文件、管理文件所体现出来，满足消费者和社会需要。反映统计数据质量要求的法律和法规有：统计法、统计法

实例细则、统计法实施条例、标准和管理文件有各种统计调查方案、汇总方案等。

(3) 质量的涵义具有动态性，定义中的“需要”会随着时间的推移、科技的进步、社会的发展、人们认识水平的提高而逐渐变化。因此，必须定期适时评价这种需要的变化，注重研究质量的时间内涵、可信性，开展对质量要求的跟踪和动态评估，并相应修改有关规范和要求。

(4) 质量的涵义从追求功能性到功能与经济性并重，通常讲质量特征是将质量要求量化的一组技术性能指标，如功能特性（如强度、精度等）、合用性、可信性、环境、适应性等，质量涵义的发展要求从追求功能性到功能与经济性并重。在市场经济下，要讲究质量成本，不应盲目追求高性能、高可靠性，而宜于具有适度的质量。因为质量的完善一般意味着成本的增高，应该从价值工程的角度来考虑适度的质量。

关于统计数据质量，人们对它有不同的理解。

在国内，对统计数据质量具体理解有以下几种代表性的意见：

王强、许亦频（1990）指出，人们对统计数据质量的理解是：1）统计数据质量是指统计数字本身的质量与统计数字为社会服务的优劣程度；2）指反映客观实际所达到的准确程度，同时，还有时效性和适用性等要求，统计数字质量与统计工作质量是两个不同的概念；3）包括三方面的含义：客观性、一致性、目的性；4）是统计方法、统计制度、统计观念和统计素质的集中表现。仅讨论统计数字质量的含义是不够的，应该把它放到统计工作的质量中来讨论，应该包括准确性、时效性、适用性、经济性。

李全昌（1998）指出，统计数据质量是一个综合性的概念，从使用者的要求看，取决于准确性、及时性、完整性、有用性和

简便性五个方面；从形成过程看，取决于搜集质量、加工整理质量、存储保管质量、分析质量和开发研究质量五个环节；从经济效益看，取决于投入产出之间的对比关系。

权贤佐（2001）认为，从统计学的角度看，准确是将误差控制在一定区间上的准确而不是数字上的 $1+1=2$ ，以客观性、可靠性、可信度来评估统计数据质量，比准确性更科学、更符合统计工作实际。

在国外，统计数据质量日益受到人们的关注。

联合国统计机构（1980）对官方统计机构提出要求，对官方统计数据质量特性的规定是：1）统计工作必须充分理解使用者进行决策和研究的需要；2）统计服务的方向应针对许多种使用者而不只是一类使用者；3）统计资料要通过指标的相互联系，而形成有机的体系；4）统计资料要保持历史连贯性，以系统性的时间数列反映历史变动；5）统计资料应当及时收集和加工、及时公布；6）应当保障提供调查资料的被调查者（居民、公私企业）的利益或秘密；7）统计机构要公正客观而不受任何偏见的影响；8）为了统计资料的准确性和及时性，统计机构应当有能胜任的业务和行政领导。

国际货币基金组织自 1995 年以来出台一套数据发布标准，目的是推动各国统计数据来源的透明度、规范各国统计制度方法，改进各国统计数据质量，标准对统计数据质量规定为准确性、连贯性和是否按国际标准编制。

菲律宾大学统计学教授 B.T.Onate 博士（1995）提出，农业普查和抽样调查的数据质量标准包括：精度（可以接受的最低误差），准确性（最小的非抽样误差），关联性（适用于决策），及时性（在最短的时间内取得数据），一致性（在序列上的可比性），最低成本（在满足前五个标准的前提下，用最经济的方式取得数据）

笔者认为,从官方统计参与国民经济活动过程和结果看,统计数据是统计工作的产品,统计数据质量应遵从一般质量的定义,即反映实体满足明确或隐含需要的能力的特性总和。

2.1.2 统计数据质量的特征

正确理解统计数据质量的涵义,不仅有助于人们认为什么是统计数据质量,提高数据的质量意识,消除不必要的误解,也有助于完善统计方法,提高统计组织管理水平,生产合格的统计产品,满足社会需要。

首先,关于统计数据质量的特性。统计数据满足社会使用需要所具备的属性和特征称为质量特性。根据不同的使用要求和使用目的,不同的统计数据具有不同的质量特性。有内在特性,如客观性、可靠性、可信度等;有外在的特性,如时效性、使用时间、所属时间等;有经济特性,如调查、整理成本等。把各种统计数据的质量特性综合起来,可以将它们归纳成可靠性、时效性、经济性等三个方面。

(1) 可靠性 可靠性指统计数据要能反映客观实际,在收集、整理和传递过程中客观、公正,没有人为的干扰和篡改。可靠性是统计数据的可信程度。统计工作搜集大量现象的数据,目的是为了反映事物本身的一种基本状况,一种存在,即通过数量说明质量。只有客观、实事求是地说明事物的存在,才是可信的、可靠的。

(2) 时效性 时效性指统计数据的时间价值。时间是统计数据的一个构成要素,统计数据质量包括数据的时间概念,如资料所属的时期长短是否一致,时点是否统一等。此外,统计数据的形成和提供的时效和速度也会构成统计数据的质量。

(3) 经济性 统计数据从调查方案设计,搜集资料调查、数

据加工汇总到提供统计数据都需要投入经费，形成统计数据成本。统计数据在满足需要和使用的前提下，所花费成本越低，数据的质量越高。

其次，关于统计数据质量的评价标准。统计数据质量的定性化或定量化就形成评价标准。长期以来，我国官方统计数据质量评价只有定性要求，没有形成定性与定量结合的标准。翟振武等（1989）提出，统计数据质量标准包括效度和信度，效度是测度某一种现象或观念的标识或统计指标是否得当，是否真实反映研究的客观事实。效度不高，数字再准确也不能认为质量好，这种数字不符合研究的目的。信度又称可靠性，指统计数据的可靠程度，具有完全信度的统计信息是客观公正的，没有收集和传递过程中被扭曲的信息。

再其次，要有科学的统计数据质量观。对官方统计数据的质量要求，人们长期以来形成一种评价的尺度，就是准确、及时、全面。近年来，社会上对统计数据质量，主要是对它的真实性、准确性问题提出疑问，这一方面反映人们对统计数据和统计数据反映的社会经济问题的关注，另一方面也反映了相当一部分人对统计数据质量的认知还停留在传统的认识上。评价统计数据质量，应该有科学的数据质量观：从人的认识讲，统计数据的准确是相对的，不可能绝对准确；从统计方法看，统计数据都有误差，用误差大小、可靠性评价统计数据的质量比准确性更科学，更符合统计工作的实际；从调查成本讲，搜集数据必须讲究投入产出效益，不看需要和使用的要求对数据的准确性加大投入是对社会资源的一种浪费。

2.2 影响统计数据质量的因素

2.2.1 影响统计数据质量的因素分类

影响统计数据质量的因素很多，人们可以从不同的角度认识这些因素。包括：(1) 从行为角度划分为有意识的行为和无意识的行为；(2) 从数据的形成过程分成搜集、加工、整理、存储和数据发布四个环节；(3) 从制度因素分为非正式约束、正式约束和实施机制。

2.2.2 从人们的行为角度划分，影响统计数据质量的因素有无意识的行为和有意识的行为

无意识的行为主要指由于人们的认识水平、工作方式和手段、管理体制、统计制度、统计方法的限制，影响统计数据的质量，例如，统计人员工作中的疏忽（如录入、抄写失误等）造成统计数据错漏，统计调查时被调查者合作程度低，不如实填写调查问卷等；统计调查制度和方法落后，难以收集到所需要的数据和资料等。

有意识的行为是有人为了某种目的和需要，有意识地、故意制造虚假数据影响统计数据质量。有这样一副对联，上联为：“上压下，层层加码，马到功成”；下联为：“下瞒上，层层加水，水到渠成”。这副对联形象地描述了普遍存在的扭曲的统计行为。有意识的行为影响统计数据质量包括：篡改，为了小集团或个人的利益，凭借职权和工作上的便利条件，擅自修改真实的统计资料；虚报，有些单位的领导，为了能得到国家补贴、救灾款等，虚报农民的人均收入、受灾面积等；瞒报，为了能多要投资项