



2005.03: 统计数据质量涵义之我见

2005.03.31 16:48:19

统计数据质量涵义之我见

文 / 张 芳

摘要: 本文从质量的一般涵义出发, 提出了全面系统的统计数据质量概念, 指出统计数据质量是一个具有丰富内涵的综合性概念, 具体包括统计数据的内容质量、表述质量及约束标准三大方面。这个概念具有评价的可操作性, 也为质量管理提供了更多的方便。

随着我国社会主义市场经济体制的逐步建立, 社会各界对统计信息的需求日益广泛, 对统计数据质量的关注更多, 要求更高。同时, 统计所面临的内部环境和外部环境都发生了巨大的变化, 这对传统统计方法下的统计数据质量既是挑战, 又是机遇。为了更好地满足社会经济发展过程中社会各界对统计数据的需求, 有必要对统计数据质量进行系统深入的研究。

一、质量的涵义

“质量”是在现代社会中被广泛使用的词语之一, 它没有单一、固定的概念, 通常在不同情况下表示不同的意思。质量的一般涵义经历了物质产品质量的狭义概念到物质产品质量、服务质量及各行各业质量的广义概念的认识过程。传统的“质量”概念, 主要关注物质产品的物理性质、客观的特征, 例如“持久耐用”、“坚固”等, 仅仅用于表明一件产品是好或者不好。然而, 随着工业化进程的加快和社会文明程度的提高, 人们将“质量”定义为“满足不同顾客的个性化需求”, 因此, “质量”一词也具备了一些主观的特征, 而且这些主观的特征越来越引起人们更多的关注。虽然不同学者或学术机构给“质量”一词以不同的定义, 但下面两种是较为典型的。一是J u a r n和G r g n a在1980年提出的: 所谓“质量”即指“适合使用”(fitness for use)。他们对质量的定义强调, 产品或服务的质量应该满足用户的期待和需要。而国际标准化组织(ISO)对“质量”的定义更明确的阐述了这一点。1986年, ISO(8402)提出: “质量”是指“产品或服务所具备的满足明确或隐含需求能力的特征和特性的总和”。

二、统计数据质量的涵义

传统的统计数据质量仅仅指其准确性, 通常用统计估计中的误差来衡量。但“质量”的概念被拓宽以后, “统计数据质量”的概念也有必要拓宽。从ISO关于质量的定义出发, 把用户的需求作为衡量统计数据质量高低的首要因素, 那么可以把统计数据质量定义为“影响统计数据满足用户需求的特征”。但由于用户多种多样, 且其对统计数据的使用目标也不一样, 从而对统计数据质量的要求也不一样, 因此, 上述定义还不是一个具有可操作性的概念。但从这个定义出发, 充分考虑不同用户的需求, 可以提出一个系统而全面的统计数据质量概念(不仅仅指准确性)。

笔者认为, 统计数据质量是一个具有丰富内涵的综合性概念。具体来说, 它包括统计数据的内容质量、表述质量及约束标准这三大方面。

(一) 统计数据的内容质量

统计数据的内容质量是统计数据最基本的特征，它包括相关性、准确性与及时性。一旦缺少了其中任何一个，统计数据就失去了转化为信息的性质和基本作用。因此，这三个特征也可称为统计数据质量的主要特征。

1. 相关性

相关性是指统计机构所生产的数据是否正是用户感兴趣的统计数据。统计数据的相关性反映了它满足用户需求的程度，它与所提供的可利用数据是否关注了对用户来说最重要的主题有关。由于对相关性的评价是主观的，会随用户需求目标的变化而改变，所以统计机构所要做的是平衡不同用户的互相矛盾的需求目标，在给定的资源条件限制下，尽可能的满足大部分用户的大部分需求。

2. 准确性

准确性指观测值或估计值与未知的真值之间的距离（接近程度），通常用统计误差来衡量。它是统计数据质量的基础和核心内容，也是传统的“统计数据质量”概念所考虑的主要问题。一般来讲，误差分为系统误差和随机误差两部分，有时也用引起不准确性的主要潜在原因（如抽样误差、无回答误差等）来分类描述。完全准确的测量经常受到成本的限制，有时甚至是不可能的。所以关键的是误差是否已降低到用户可以接受的地步。

3. 及时性

与用户需求相关的准确的统计数据如果没有在用户做出决策之前传递给用户，那么该数据对用户来说，是没有用的。所以，及时性也是统计数据能否满足用户需求的重要特征。如果该现象本身变化比较迅速，则对该类统计数据的及时性要求高；如果该现象本身变化比较缓慢，则对及时性要求不高。

（二） 统计数据的表述质量

对统计数据质量来说，仅考虑其内容方面的质量是不够的。统计数据需要表述，特别是将某个统计数据同其他相关统计数据相互联系地加以表述时，要考虑表述的质量问题。比如单个数据的内容是正确的，但表述不清晰、不充分，就会影响整套数据的质量，甚至引起误解。所以，统计数据质量必须考虑其表述质量。统计数据的表述质量包括可比性、可衔接性和可理解性。

1. 可比性

可比性是指同一项目的统计数据在时间上和空间上的可比程度。这要求统计的概念和方法在时间上保持相对稳定，在不同地区使用统一的统计制度方法和分类标准，保持统计数据的口径范围、计算方法在时间上一致衔接，在地区之间可比。

2. 可衔接性

可衔接性是指同一统计机构内部不同统计调查项目之间、不同机构之间以及与国际组织之间统计数据的衔接程度。这要求全国范围内所有专业统计项目在统一的统计框架体系、分类标准下，按统一的方法编制统计数据，在统计调查和数据加工整理中使用统一的方法和程序，同时采用国际统计标准，如联合国1993年SNA的框架体系等。

3. 可理解性

可理解性是指统计数据便于用户正确理解并使用的程度。统计数据是提供给用户使用的，如果某些用户不能理解，看不懂统计数据和统计分析报告，当然也就谈不上使用数据。为了恰如其分地使用从统计机构那里得到的统计数据，用户必须了解他们所获得数据的性质。这就要求统计机构在提供统计数据的同时附带提供对数据的补充说明，如提供隐含在有关概念下面的说明、已使用的分类法、数据收集和加工过程中所使用的方法以及统计机构自身对数据质量的评价。

（三） 统计数据的约束标准

在实现统计数据目标的过程中，除了注意统计数据的内容质量和表述质量这两方面外，还必须注意以下两项具有普遍意

义的约束标准，这也体现了统计数据的质量特征。

1. 可取得性

可取得性是指用户从统计部门取得统计数据的便利程度。对于有用的统计数据，用户必然要考虑：能得到哪些数据，如何得到这些数据。因此，统计数据必须以一种用户能够使用（搜寻方便）而且能够负担的形式提供给用户。这要求提供统计数据时，必须列明用户从统计机构可以取得的统计数据内容，同时要应用先进便捷的统计数据服务方式，使用户取得数据更为便利。

2. 有效性

有效性是指统计数据的利用所产生的效益要大于提供该数据的成本。如果情况相反，则提供这种数据对提供方和使用方来说都是不值得的。虽然目前统计数据的效益和成本特别是前者不容易准确地计量，但是，保持这样一种基本的指导思想是十分必要的。这要求在统计数据的其他质量不受大的影响的前提下，尽可能降低统计数据的生产费用，提高效率。

三、统计数据质量性质之间的相互关系

上述统计数据质量性质的八个方面是相互联系的，而非彼此完全独立。例如，准确性和及时性会影响信息的相关性，可理解性会直接影响信息的可比性和可衔接性，准确性与及时性有冲突，与当前相关和与历史数据保持一致可比有时也是一对矛盾，关于数据准确性的信息会成为可解释性的一种要求。注意到数据质量各方面的这些相互联系，会给统计机构的数据质量管理提供一个有用的基础。譬如，不同的用户对同一统计产品可能也有不同的质量要求，有的侧重于准确性，有的则更关注其及时性，因此，从满足用户需求的角度来看，没有统一的质量评价标准。所以，政府统计机构首先要做的，就是关注主要用户的主要需求目标。

另外，值得注意的是，对用户来说，如果没有统计机构所提供的附加信息（或源数据），统计数据的许多重要性质并非显而易见。光看数据本身，并不能推断该数据的准确性（同其它来源的数据进行比较也许对此有帮助）。但统计机构能得到源数据及调查整理过程中所使用方法的第一手资料，因此，它有义务提供对数据准确性的测度。如果没有得知隐含的概念、分类及加工过程中使用的方法，可比性和可衔接性也无法直接看出。只有及时性和可取得性能够由用户直接看出。

此外，相关性、可比性、可衔接性与可取得性是对一整套统计产品来说的，并不是针对单个产品而言。统计数据的相关性依赖于其它有效的信息，因此，只有对整个计划进行评价，才能得到所谓数据质量相关性的评价。同样的道理，可比性与可衔接性也是如此。对整个统计机构来说，大部分统计产品都是通过一个共同的传播体系传递，因此，可取得性也是一套产品的共同特征。另一方面，准确性、及时性和可解释性可以看成单个产品的特征。虽然，在这里，每一个产品也许都使用了在整个项目中普遍使用的工具和方法。

至此，一个系统而全面的统计数据质量概念可以定义为：影响统计数据满足用户需求的特征，其中包括内容方面的准确性、及时性、相关性；表述方面的可比性、可衔接性、可理解性；约束标准方面的可取得性和有效性。衡量单个或一整套统计数据产品质量的高低，可以从定义中涉及的八个方面综合考虑。

（作者单位：湖南商学院）

附件