

变异系数

—— 一个衡量离散程度简单而有用的统计指标

文 / 王文森

在分析写作过程中，我们常常会碰到这样的问题：如何定量地说明一个总体各单位的不同时期集中或分散的发展变化趋势，或不同总体同一时期的发展差异程度？比如我们要分析某地区工业企业的盈利水平在五年内是趋于平均化还是趋于分化，又如要对比分析2001年广东与广西两省内部的经济发展均衡程度，等等。在这里，笔者介绍一个既简单常用又能说明问题的统计指标——变异系数。

变异系数的概念

变异系数，就是标准差系数，有的书上也称差异系数、离散系数，都是同一个意思。

变异系数是反映总体各单位标志值的差异程度或离散程度的指标，是反映数据分布状况的指标之一。其含义是总体各单位的标准差与其算术平均数对比的相对数，其计算公式为：

$$V = \frac{\sigma}{\bar{x}}$$

上式中，V代表变异系数，σ代表标准差， \bar{x} 代表算术平均数。

大家知道，我们一般用平均数来反映客观现象总体各单位某一数量标志的一般水平，如平均年龄、平均收入等。但平均数只能反映总体的一般数量水平，不能说明总体各单位标志值的数量差异程度，不能揭示其离散程度和集中趋势，这时候就要用到平均差、标准差和变异系数（标准差系数）等变异指标。下面图示可帮助大家理解平均数和变异系数的含义。

图1

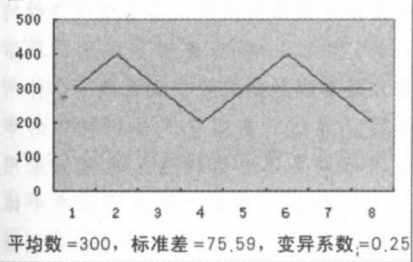
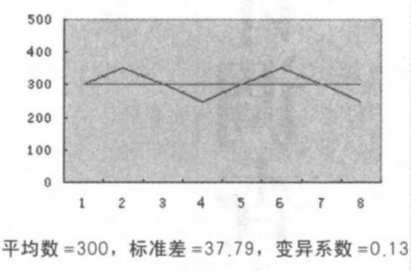


图2



上述两个数据序列的平均数都是300，但标准差和变异系数不同，序列1的变异系数为0.25，而序列2的变异系数为0.13。表现在图上，就是图1的数据波动幅度比较大，而图2的数据波动幅度比较小，也即序列1各数据之间的差距大于序列2。变异系数正是用来描述这种数据波动程度。

标准差

在进一步了解变异系数之前，需要先解释标准差。标准差是总体各单位标志值与其算术平均数离差平方和的算术平均数的平方根，通常用σ表示。标准差的平方称为方差，用σ²表示。标准差的一般计算公式为：

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N}}$$

式中N代表数据个数，x代表各单位标志值， \bar{x} 代表各单位标志值的平均数。

下表列出5个企业产值数据，以此为例计算企业产值的标准差。

企业	产值(千元)x	离差(x- \bar{x})	离差平方和(x- \bar{x}) ²
A	200	-100	10000
B	250	-50	2500
C	300	0	0
D	350	50	2500
E	400	100	10000
合计	1500	0	25000

根据上表，可计算出标准差σ：

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N}} = \sqrt{\frac{25000}{5}} = 70.7$$

标准差的局限性。标准差以算术平均数为中心，它可以反映一个总体全部标志值的离散程度。但它是一个绝对指

标，当用其来对同一总体的不同时期或不同总体进行对比时，缺乏可比性。原因是：

第一，当总体平均水平不同时，用绝对差异指标不可比。例如当比较广东与新建两地的人均收入差异程度时，若计算得出两地的人均收入标准差都是500元，这时我们不能简单地认为两省的人均收入差异程度是一样的，因为500元对两省人均收入的意义大不一样。

第二，绝对指标受量纲（计量单位）的限制，导致横向数据不可比。比如在比较中国和美国各地人均GDP的差异程度时，就存在一个不同货币计量单位之间的换算问题，如果采用相对差异指标，就可以消除这种限制。因此，在不同水平的总体之间、不同量纲的总体之间，需要采用变异系数来比较标志值变动程度的大小。变异系数越小，说明标志值离散程度越小，总体各单位的差距越小。

变异系数的计算

由上面的内容可知，计算出一个数据序列的标准差和平均数，就可得出变异系数。上例中，

$$\sigma = 70.7, \quad \bar{x} = 300,$$

$$V = \frac{\sigma}{\bar{x}} = \frac{70.7}{300} = 0.24$$

变异系数的计算较简便，利用计算器或有关的统计软件就可以得出结果。如在EXCEL软件中，利用“工具”菜单中的“数据分析”工具就可计算出标准差、平均数和变异系数等一些常用的统计指标。

在实际计算时，若有必要考虑各数据的不同重要性，则通常采用加权平均的形式。比如，在计算各地区人均GDP的变异系数时，一般应以各地区的人口数作为权数进行加权评价。用公式表示为：

$$V = \frac{1}{\bar{x}} \cdot \sqrt{\sum [(x - \bar{x})^2 \cdot \frac{P_i}{P}]}$$

式中 v 、 x 和 \bar{x} 的意义同上, p_i 代表第 i 地区的人口数, p 代表全部地区的人口数总和。

变异系数的应用

变异系数是一个虽简单但用途广泛的指标,普遍地用于反映某一总体各单位标志值的差异程度、集中或离散状况。国内外学者普遍利用变异系数来分析地区差距问题。下面笔者举出在写作过程中用过的几个例子,其他可以举一反三。

例1:工业50强发展差距分析。

在一篇分析广东工业50强的文章中,利用变异系数来说明大企业发展差距拉大。引文:“从50强企业的发展变化看,强势企业间的差距有进一步拉大的趋势。1998年50强企业销售收入的变异系数为0.58,1999年的变异系数为0.60,显示强势企业间的差距也在拉大。”

例2:广东省地区发展差距分析。

利用1995年和2000年广东省各市人均GDP数据,计算得出1995年人均GDP的变异系数为0.16,2000年人均GDP的变异系数为0.18(未作加权)。计算结果表明,“九五”期间,广东省各地区的发展速度快慢不均,发展差距有所拉大。如果将改革开放以来全省人均GDP的变异系数都计算出来,则可进一步分析地区差距的发展变化趋势。

分析差异状况的其他指标

除变异系数外,还有其他反映总体各单位标志值差异状况的统计指标,如极差、极值比率、基尼系数、Theil熵等。

极差是数据序列中极大值和极小值之间的差额,极值比率是极大值与极小值之比。这两个指标在分析中也比较常用,比如算工业50强中的第一名和最后一名的差距或比率,分析其实力差距。基尼系数将在后面章节中具体介绍。Theil熵最早是由Theil于1967年提出来的,用于研究国家之间的收入差距,该指标相对复杂,这里不做叙述。

(作者单位:广东省统计局)

统计调查方法的

文/王天营

哲学思考

统计学作为一门关于如何收集数据、整理数据和分析数据的方法论科学,其之所以成为科学就在于其以世界是可以认知的观点为基本出发点,在于其所阐述的各种统计方法中无不处处闪烁着哲学的思想火花,更在于通过各种统计方法的应用,能帮助人们对五彩缤纷的世界有更深刻的认识。尽管统计学是从量的方面认知社会、自然、科技等现象的方法论体系,而且每一具体的方法都蕴涵着朴素的哲学思想,但是本文仅就收集数据中几种基本调查方法所体现的哲学意义加以探讨,以便使人们能更自觉、科学地使用这些统计调查方法,并强化人们对运用这些统计调查方法所得数据的尊重意识。

统计调查方法与哲学的基本范畴
统计调查是在一定的调查总体内收

集数据的一项实践活动,调查总体由有限或无限个总体单位所构成,因此,就统计调查的范围来看,统计调查方法有全面调查和非全面调查之分。顾名思义,全面调查就是对调查总体中全部总体单位一一进行调查,而非全面调查仅对调查总体中部分总体单位进行调查。第二,就非全面调查而言,按照从总体中选取部分单位方法的不同,又可以分为重点调查、典型调查和抽样调查等。重点调查就是仅对调查总体中调查者认为比较重要的单位进行的调查。典型调查就是仅对调查总体中调查者认为比较典型的单位进行的调查。抽样调查就是根据随机性原则从调查总体中选取部分单位进行的调查。第三,从抽样调查的具体组织实施来看,抽样调查又有简单随机抽样调查、系统抽样调查、分层(组)抽样调查、整群抽样调查和多阶段抽样调查等。

唯物辩证论通常要研究有限与无限、部分与整体、现象与本质、原因与结果、特殊与一般、可能性与现实性、相对性与绝对性、量变与质变、偶然性与必然性、形式与内容等基本哲学范畴。而对这些问题的基本观点与态度,不仅影响着一个人对周围世界的根本态度,而且对一门学科的发展同样有着基础性的影响。例如,对各种统计方法的衍生与应用就有着至关重要的引领作用。

统计调查方法的哲学意义

1. 全面调查与非全面调查的哲学意义。从全面调查与非全面调查的含义来看,它体现着“部分与整体”、“有限性与无限性”、“可能性与现实性”等辩证关系。事实上,只要承认通过对总体中足够多的个体的认识,就可以达到对总体一定程度上的正确认识,那么,人们在一定的场合无论采用全面调查获取数据,还是采用非全面调查获取数据都可以达到认识总体数量规律性的目的。其二,客观事物(现象)通常都可以看成是由有限个或无限个个体所构成。显然对有限个个体所构成的事物,可以采用全面调查,也可以采用非全面调查。但是,对由无限个个体所构成的事物,则必须采用非全面调查