关于对标志变异指标概念的重新认识

孔建新[1] 孔璐[2] 何光伟[3]

1 云南会泽者海 2-24-6 统计科研小组 云南会泽者海镇（654211）

E-mail：kongfanjx@163.com

2 广东省中山火炬开发区职业技术学院 中山市（528437）

E-mail：yudianlulu@163.com

3 广东药学院中山校区 中山市（528458）

E-mail：weiweihonest@163.com

摘 要：统计分析的中心内容是数据变异程度的度量和分解。然而目前随机变量标志变异指标的概念与内涵则存在一定的缺陷，已经满足不了把数据间存在的差异解释清楚的需要。本文在前期对西格玛内涵拓展的研究成果以及对变差重新认识的基础上进行综合归纳，同时将标志变异指标的内涵也相应地进行拓展，进而对其概念重新认识。

关键词：标志变异指标，变差，方差，西格玛

中图分类号：O212.2

0．引言

统计学界的教育大家张尧庭先生认为：“ 无论是一元统计或多元统计，统计分析的中心内容都是数据变异程度的度量和分解，从而解释变异的来源与影响它的因素是否重要、重要的程度如何。 ”并说：“ 能把数据间存在的差异解释清楚，统计就学好了。[1] ”在统计实践中由于随机变量变异程度受多方面的影响，使之在统计工作中需要根据统计对象所处的实际情况来进行统计分析。以便把数据间存在的差异解释清楚。随着科技的进步，计量手段的提高，所采集的数据已经达到一定的精度。然而传统习惯上所应用的变异指标内容在统计的实际工作中已经远远不能满足现代统计分析的要求，为此需要根据统计研究对象的实际情况来拓展变异指标体系的具体内涵。目的是：把统计研究对象数据间存在的差异解释清楚。

需要提及的重要问题是：平均值应用的广泛性使之在统计的实际应用中客观地存在一定的误区。“ 平均数是统计分析中使用频率最高的用来描述数据分布集中程度的统计指标，平均数的特点是容易受到数据取值中极端值的影响，这就使得平均数看似精确的背后往往隐藏着未知的陷阱[2]。 ”由此说明平均值的使用一定要考虑统计研究对象的适用条件，滥用带来的结果必然会出现矛盾。所以“ 统计方法的正确应用有助于我们认清事物的真象、发现事物变化的数量界限、揭示事物发展的内在规律。相反，统计方法的错误使用，将造成事实的扭曲、读者的误解，甚至决策的失败。统计在许多应用领域都存在着不同程度的误用[2]。 ”由此说明：随机变量与平均值离散程度的变异指标在统计实际工作中是不能概括随机变量与其它位置特征值离散程度的变异指标。这是本课题所探讨问题的重要统计背景。

2009 年 1 月 10 日在南京召开的第六届海峡两岸统计与概率学术研讨会中国统计学会会长李德水先生指出：“ 统计学是一门收集和分析数据，并根据数据进行推断和决策的科学，统计方法是自然科学、工程技术、社会经济等各个研究领域进行数量分析的基本手段。 [3] ”在自然科学与社会科学的许多现象中，细微变化的统计分析将需要统计方法的不断进步和更新。这对于随机变量与不同位置特征值的变异指标体系需要重新进行严密的逻辑梳理，使其在统计实践中发挥更有效的作用。这是统计数据科学发展的需要，也是应用数学发展的必然

结果。

本文将随机变量与不同位置特征值的变异指标分为三个层次来进行探讨，通过不同的涵义来进行归纳，从而进一步拓展标志变异指标的内涵，以完善其概念。

1．不同位置特征值的基本概念

“我们在研究频数直方图时看到，统计观测值有一种集中的趋势，即在某个数值附近的频数比较大，而在远离该值的地方频数比较小，这种趋势集中的数值称为统计观测值的位置特征。位置特征有多种估计方法，如：平均数、中位数、众数等[4]。”

以上三种常见的位置特征值在统计实践中用于不同的研究对象其作用不同。平均数是统计观测值的分布中心[5]（与位置中心数是不同的概念）；众数（峰值）是统计观测值的集众位置；中位数是统计观测值顺序排列中间位置的那个数。

在统计实践工作中位置特征值并不限于常用的三种。还客观存在其它多种，它们与常用的位置特征值在实践应用中反映着不同的征象和发挥不同的作用，共同构成完整的分布数组间的数据逻辑关系。其相互间的联系和不同的作用为统计的分析和判断提供有力的依据。

在《关于单峰分布中心数位置特征的初步讨论》[6]一文中已经详细论述了不同位置特征值的基本概念。本文进行补充与完善，规范地重述如下：

众数(mode)：也称峰值，它是单峰分布的数学期望值（expected value）。是总体分布中频数出现次数最多的标志值，它能直观地说明客观现象分布状况的集众趋势，不受极端值的影响。它的位置可以在整个分布区间的任意一点上，其位置变动很大。在单侧规范产品的质量指标分布中，当它等于目标值时，其位置处在分布一边的端点上，分布正好是正态分布的一半。

平均值(mean value)：是指在同质总体内将各单位某一数量标志的差异抽象化，用以反映总体在具体条件下的一般水平。也称平均指标[7]。它受极端值的影响较大，在分布区间它的位置围绕分布中心值波动，其位置也是变动的，变动的范围比众数小得多。

中位数(median)：现象总体中各单位标志值按大小顺序排列，居于中间位置的那个标志值就是中位数[7]。在数组中它的位置也是围绕分布中心数波动，其位置变动的范围比平均值还小。

标准值(standard value)：标准规定的值，在双侧规范产品中它是最好水平值，等同目标值。在单侧规范产品中标准规定的值是质量要求的最低水平值，等同合格值。在一般情况下标准值在分布中位置是稳定不易变动的。（在此特别说明：本文的标准值特指双侧规范产品最好水平值，即目标值。单侧规范产品标准规定仅指上或下规范值。即：合格值。标准值在单侧与双侧的涵义完全不同，为不至于与双侧的标准值混淆，所以单侧规范产品质量指标不存在标准值的表述，标准值特指双侧规范产品质量指标的目标值。）

目标值(target value)：是产品质量指标最好水平值，在双侧规范产品中它的位置处在规范限内的中心点（在特殊的情况下，在公差的配合中上或下限值的标准要求不一致，此时标准值的位置就不在规范限内的中心点上。），等同标准值；在单侧规范产品中它的位置处在合格值（规范值）对应的另一端点上。

规范值（norm value)：也称合格值（qualified value)它是标准规定的最低水平值，在双侧规范产品中它在标准值（目标值）的两边。在单侧规范产品中它仅位于目标值对应的另外一端，分为上规范值和下规范值。

中心数分为分布中心数与规范中心数两类，分别是：

分布中心数（distribution center value）：是指在同质总体的一组数据中最大标志值与最小标志值之间处于中心位置的数。虽然它受极端的影响较大，但是位置变动则不大。

规范中心数（standard center value)可分为两种：

一种是双侧规范条件下的规范中心数：是指上下规范限内中心位置的数，等同标准值或目标值。当上下规范限被确定后，其位置稳定不会变动。它不受规范限以外极端值的任何影响。

另一种是单侧规范条件下的规范中心数：是指上或下规范限与目标值之间中心位置的数。当上或下规范限和目标值被确定后其位置稳定不会变动。它同样不受规范限以外极端值的任何影响。

分布中心数与规范中心数的区别是：前者受极端的影响，后者不受极端的影响。在此顺便一提，这两个位置特征值在产品质量等级的判定中有着重要的不同作用。

统计研究对象所收集到的一组随机变量一旦被确定后，其分布状况及位置特征值也随之被确定。这个确定是事后通过统计计算才能得知。然而在统计预测和推断中人们很难事先知道统计研究对象具体的分布状况，也就不容易知道可变的位置特征值在随机变量分布中的具体位置所在，所以提出不同位置特征值对于统计分析必将有一定的现实意义。

2. 变差、方差、西格玛的基本概念与定义

从“ 变差无处不在[1] ”说明变差是分析数据和统计方法不可缺少的重要的基础参数之一。方差由变差引出，西格玛则由方差引出。本文综合前期的研究成果，从变差的基础入手探讨变异指标的三个层次的具体内涵及其所包括的基本范畴。

2.1 变差的基本概念与定义

变差的基本概念在《解读随机变量与不同位置特征值的变差》[8]已经作了详细的论述，有必要再重述如下：

从网上百度百科查到变差（variation）有如下解释：

1.变差又称回差，是指仪表在上行程和下行程的测量过程中，同一被测变量所指示的两个结果之间的偏差。

2.在机械结构的检测仪表中，由于运动部件的摩擦，弹性元件的滞后效应和动态滞后的时间影响，使测量结果出现变差。

3.测定值是一个以概率取值的随机变量，多次测定所得到各次测定值通常都是参差不齐的，其间的差异称为变差，是反映测定结果稳定性的一个重要标志。

4.变差既可能是由于随机因素，也可能是由于试验条件的改变而引起的。如果是前者引起的，则属于试验误差，反映了测定结果的精密度；如果是后者引起的则属于因素效应，反映了测定条件对测定结果的影响，变差大小可用偏差平方和表示。[9]

通过上述对变差的解释，其基本概念可以理解为：变差属于统计对象标志变异指标的范畴。

变差所指的范畴应该是：随机变量与不同位置特征值距离之差和的平均值及全距。

全距指的是数组或分布值域中最大值与最小值之差，是变差的最大值。

综上所述可以给出变差的基本概念，定义为：在统计数据中，随机变量与所对应数据的不同位置特征值离差绝对值和的平均值。称为变差。以不同位置特征值为中心，随机变量分别从左或右两边与所对应数据的不同位置特征值离差绝对值和的平均值。称为左变差和右变

差。全距是最大的变差。

2.2 方差的基本概念与定义

单峰分布概率密度函数的建立引出了左右期望差的新概念，必然地还引出相应的左右期望方差的新概念。变差与西格玛内涵的扩展使得方差的范畴也相应地随之扩展。

传统方差基本概念是：随机变量与平均值离差平方和的平均值。

方差新的概念，定义为：在统计数据中，随机变量与所对应数据的不同位置特征值离差平方和的平均值。统称为方差。以不同位置特征值为中心，随机变量分别从左或右两边与所对应数据的不同位置特征值离差平方和的平均值。称为左方差和右方差。

2.3 西格玛的基本概念与定义

西格玛的新概念在《关于对西格玛内涵拓展的研究》[10]已经作了详细的论述，重述一下。

传统西格玛基本概念是：随机变量与平均值离差平方和的平均的平方根。

西格玛新的基本概念，定义为：在统计数据中，随机变量与所对应数据的不同位置特征值离差平方和的平均的平方根。称为西格玛。以不同位置特征值为中心，随机变量分别从左或右两边与所对应数据的不同位置特征值离差平方和的平均的平方根。称为左西格玛和右西格玛 [10] 。

综合以上三个层次变异指标的定义得出，任何一组随机变量与不同位置特征值离散程度的变异指标将形成变差、方差、西格玛这样一套指标系统，从属于标志变异指标的统计范畴。

3. 变差、方差、西格玛包括的具体内容及计算方法

3.1 变差包括的具体内容及计算方法

变差（variation）[8]包括：

平均值差、标准值差、目标值差、期望值差（等价众数值差和峰值差）、中位值差、分布中心值差、规范中心值差。

对应以上变差以及客观存在数组（分布）的不对称性又可分解为：左右平均值差、左右标准值差、左右目标值差、左右期望值差（等价左右众数值差和左右峰值差）、左右中位值差、左右分布中心值差、左右规范中心值差。

以下分别给出定义和计算方法。

平均值差（mean variation）(等价于传统的平均差)：随机变量与平均值之差取绝对值之和的平均数。符号记为：Vm 。

左平均值差（left mean variation）：小于等于平均值的随机变量与平均值之差取绝对值之和的平均数。符号记为：Vm- 。

右平均值差（right mean variation）：大于等于平均值的随机变量与平均值之差取绝对值之和的平均数。符号记为：Vm_+ 。

计算方法：

令：随机变量频数之和为 n；

小于平均值的频数为 n_1；

大于平均值的频数为 n_2；

等于平均值的频数为 n_3。

设：平均值左边的频数为 n_-；

平均值右边的频数为 n_+；

则：$n_- = n_1 + n_3 \div 2$

$n_+ = n_2 + n_3 \div 2$

满足：$n = n_- + n_+ = n_1 + n_2 + n_3$

根据以上条件给出计算平均值差 Vm、左平均值差 Vm-、右平均值差 Vm_+的公式如下：

$$平均值差：Vm = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|,\quad (i=1,2,\cdots,n) \tag{1}$$

$$左平均值差：Vm- = \frac{1}{n_-}\sum_{i=1}^{n_-}|x_i - \bar{x}|,\ x_i \le \bar{x}\quad (i=1,2,\cdots,n-) \tag{2}$$

$$右平均值差：Vm_+ = \frac{1}{n_+}\sum_{i=n_-+1}^{n}|x_i - \bar{x}|,\ x_i \ge \bar{x}\quad (i=n-+1, n-+2,\cdots,n) \tag{3}$$

标准值差（standard variation）：随机变量与标准值之差取绝对值之和的平均数。符号记为：Vs 。计算方法与（1）式等同，只需将标准值的符号：B 取代（1）式中的平均值 $\bar{x}$ 即可。

左标准值差（left standard variation）：小于等于标准值的随机变量与标准值之差取绝对值之和的平均数。符号记为：Vs- 。计算方法与（2）式等同，只需将标准值的符号：B 取代（2）式中的平均值 $\bar{x}$ 即可。

右标准值差（right standard variation）：大于等于标准值的随机变量与标准值之差取绝对值之和的平均数。符号记为：Vs_+ 。计算方法与（3）式等同，只需将标准值的符号：B 取代（3）式中的平均值 $\bar{x}$ 即可。

目标值差（target variation）：随机变量与目标值之差取绝对值之和的平均数。符号记为：Vt 。计算方法与（1）式等同，只需将目标值的符号：M 取代（1）式中的平均值 $\bar{x}$ 即可。

左目标值差（left target variation）：小于等于目标值的随机变量与目标值之差取绝对值之和的平均数。符号记为：Vt- 。计算方法与（2）式等同，只需将目标值的符号：M 取代（2）式中的平均值 $\bar{x}$ 即可。

右目标值差（right target variation）：大于等于目标值的随机变量与目标值之差取绝对值之和的平均数。符号记为：Vt_+ 。计算方法与（3）式等同，只需将目标值的符号：M 取代（3）式中的平均值 $\bar{x}$ 即可。

期望值差（target variation）：随机变量与期望值之差取绝对值之和的平均数。符号记为：Vt 。计算方法与（1）式等同，只需将期望值的符号：μ 取代（1）式中的平均值 $\bar{x}$ 即可。

左期望值差（left target variation）：小于等于期望值的随机变量与期望值之差取绝对值之和的平均数。符号记为：Vt- 。计算方法与（2）式等同，只需将期望值的符号：μ 取代（2）式中的平均值 $\bar{x}$ 即可。

右期望值差（right target variation）：大于等于期望值的随机变量与期望值之差取绝对值之和的平均数。符号记为：Vt_+ 。计算方法与（3）式等同，只需将期望值的符号：μ 取代（3）式中的平均值 $\bar{x}$ 即可

众数值差（mode variation）：随机变量与众数值之差取绝对值之和的平均数。符号记为：Vmo 。计算方法与（1）式等同，只需将众数值的符号：Mo 取代（1）式中的平均值 $\overline{x}$ 即可。

左众数值差（left mode variation）：小于等于众数值的随机变量与众数值之差取绝对值之和的平均数。符号记为：Vmo- 。计算方法与（2）式等同，只需将众数值的符号：Mo 取代（2）式中的平均值 $\overline{x}$ 即可。

右众数值差（right mode variation）：大于等于众数值的随机变量与众数值之差取绝对值之和的平均数。符号记为：Vmo_+ 。计算方法与（3）式等同，只需将众数值的符号：Mo 取代（3）式中的平均值 $\overline{x}$ 即可。

中位值差（median variation）：随机变量与中位值之差取绝对值之和的平均数。符号记为：Vme 。计算方法与（1）式等同，只需将中位值的符号：Me 取代（1）式中的平均值 $\overline{x}$ 即可。

左中位值差（left median variation）：小于等于中位值的随机变量与中位值之差取绝对值之和的平均数。符号记为：Vme- 。计算方法与（2）式等同，只需将中位值的符号：Me 取代（2）式中的平均值 $\overline{x}$ 即可。

右中位值差（right median variation）：大于等于中位值的随机变量与中位值之差取绝对值之和的平均数。符号记为：Vme_+ 。计算方法与（3）式等同，只需将中位值的符号：Me 取代（3）式中的平均值 $\overline{x}$ 即可。

分布中心值差（distribution center variation）：随机变量与分布中心值之差取绝对值之和的平均数。符号记为：Vc 。计算方法与（1）式等同，只需将中心值的符号：C 取代（1）式中的平均值 $\overline{x}$ 即可。

左分布中心值差（left distribution center variation）：小于等于分布中心值的随机变量与分布中心值之差取绝对值之和的平均数。符号记为：Vc- 。计算方法与（2）式等同，只需将中心值的符号：C 取代（2）式中的平均值 $\overline{x}$ 即可。

右分布中心值差（right distribution center variation）：大于等于分布中心值的随机变量与分布中心值之差取绝对值之和的平均数。符号记为：Vc_+ 。计算方法与（3）式等同，只需将中心值的符号：C 取代（3）式中的平均值 $\overline{x}$ 即可。

规范中心值差（norm center variation）：随机变量与规范中心值之差取绝对值之和的平均数。符号记为：Vcn 。计算方法与（1）式等同，只需将规范中心值的符号：Cn 取代（1）式中的平均值 $\overline{x}$ 即可。

左规范中心值差（left norm center variation）：小于等于规范中心值的随机变量与规范中心值之差取绝对值之和的平均数。符号记为：Vcn- 。计算方法与（2）式等同，只需将规范中心值的符号：Cn 取代（2）式中的平均值 $\overline{x}$ 即可。

右规范中心值差（right norm center variation）：大于等于规范中心值的随机变量与规范中心值之差取绝对值之和的平均数。符号记为：Vcn_+ 。计算方法与（3）式等同，只需将规范中心值的符号：Cn 取代（3）式中的平均值 $\overline{x}$ 即可。

在以上变差中，目标值差需要注意统计对象应用的条件。在双侧规范的条件下存在三个变差，如以上所述。在单侧规范的条件下仅有一个变差，即：左目标值差或右目标值差。因为在单侧规范的条件下目标值（最好水平值）是在数组（分布）位置的端点上。当规范上限

时，目标值为最小值，存在右目标值差，即：大于等于目标值的随机变量与目标值之差取绝对值之和的平均值。当规范下限时，目标值为最大值，存在左目标值差，即：小于等于目标值的随机变量与目标值之差取绝对值之和的平均值。应用时需要充分理解。[8]

3.2 方差包括的具体内容及计算方法

方差（variance）包括：

平均方差、标准方差、目标方差、期望方差（等价众数方差和峰值方差）、中位方差、分布中心方差、规范中心方差。

对应以上方差以及客观存在数组（分布）的不对称性又可分解为：左右平均方差、左右标准方差、左右目标方值差、左右期望方差（等价左右众数方差和左右峰值方差）、左右中位方差、左右中心方差、左右规范中心方差。

以下分别给出定义和计算方法。

平均方差（mean variance）：随机变量与平均值离差平方和的平均值。 符号记为：$\sigma^2 m$ 。

左平均方差（left mean variance）：小于等于平均值的随机变量与平均值离差平方和的平均值。符号记为：$\sigma^2 m_-$ 。

右平均方差（right mean variance）：大于等于平均值的随机变量与平均值离差平方和的平均值。符号记为：$\sigma^2 m_+$ 。

计算方法所给出的条件与计算变差相同，具体公式如下：

$$\text{平均方差：}\sigma^2 m=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,\quad (i=1,2,\cdots,n) \tag{4}$$

$$\text{左平均方差：}\sigma^2 m_-=\frac{1}{n_-}\sum_{i=1}^{n_-}(x_i-\bar{x})^2,\ x_i\le\bar{x}\quad (i=1,2,\cdots,n_-) \tag{5}$$

$$\text{右平均方差：}\sigma^2 m_+=\frac{1}{n_+}\sum_{i=n_-+1}^{n}(x_i-\bar{x})^2,\ x_i\ge\bar{x}\quad (i=n_-+1,n_-+2,\cdots,n) \tag{6}$$

以下不同的方差计算方法与平均值差的计算方法等同，只需将具体的符号取代（4）（5）（6）式中的平均值$\bar{x}$即可。

标准方差（standard variance）：随机变量与标准值离差平方和的平均值。符号记为：$\sigma^2 s$ 。

左标准方差（left standard variance）：小于等于标准值的随机变量与标准值离差平方和的平均值。符号记为：$\sigma^2 s_-$ 。

右标准方差（right standard variance）：大于等于标准值的随机变量与标准值离差平方和的平均值。符号记为：$\sigma^2 s_+$ 。

目标方差（target variance）:随机变量与目标值离差平方和的平均值。符号记为：$\sigma^2 t$ 。

左目标方差（left target variance）：小于等于目标值的随机变量与目标值离差平方和的平均值。符号记为：$\sigma^2 t_-$ 。

右目标方差（right target variance）：大于等于目标值的随机变量与目标值离差平方和的平均值。符号记为：$\sigma^2 t_+$ 。

期望方差（target variance）：随机变量与期望值离差平方和的平均值。符号记为：$\sigma^2 t$ 。

左期望方差（left target variance）：小于等于期望值的随机变量与期望值离差平方和的

平均值。符号记为：σ^2t- 。

右期望方差（right target variance）：大于等于期望值的随机变量与期望值离差平方和的平均值。符号记为：$\sigma^2 t_+$ 。

众数方差（mode variance）：随机变量与众数离差平方和的平均值。符号记为：σ^2mo 。

左众数方差（left mode variance）：小于等于众数值的随机变量与众数离差平方和的平均值。符号记为：σ^2mo- 。

右众数方差（right mode variance）：大于等于众数值的随机变量与众数离差平方和的平均值。符号记为：$\sigma^2 mo_+$ 。

中位方差（median variance）：随机变量与中位值离差平方和的平均值。符号记为：σ^2me 。

左中位方值差（left median variance）：小于等于中位值的随机变量与中位值离差平方和的平均值。符号记为：σ^2me- 。

右中位方差（right median variance）：大于等于中位值的随机变量与中位值离差平方和的平均值。符号记为：$\sigma^2 me_+$ 。

分布中心方差（distribution center variance）：随机变量与分布中心值离差平方和的平均值。符号记为：σ^2c 。

左分布中心方差（left distribution center variance）：小于等于分布中心值的随机变量与分布中心值离差平方和的平均值。符号记为：σ^2c- 。

右分布中心方差（right distribution center variance）：大于等于分布中心值的随机变量与分布中心值离差平方和的平均值。符号记为：$\sigma^2 c_+$ 。

规范中心方差（norm center variance）：随机变量与规范中心值离差平方和的平均值。符号记为：σ^2cn 。

左规范中心方差（left norm center variance）：小于等于规范中心值的随机变量与规范中心值离差平方和的平均值。符号记为：σ^2cn- 。

右规范中心方差（right norm center variance）：大于等于规范中心值的随机变量与规范中心值离差平方和的平均值。符号记为：$\sigma^2 cn_+$ 。

在以上方差中，目标方差需要注意统计对象应用的条件。在双侧规范的条件下存在三个方差，如以上所述。在单侧规范的条件下仅有一个方差，即：左目标方差或右目标方差。

3.3 西格玛包括的具体内容及计算方法

《关于对西格玛内涵拓展的研究》[9]讨论了西格玛（Sigma）的范畴并对所包括的具体内容及计算方法作了详细的论述。本文不再赘述。

其具体内容包括：

期望差（expected deviation）、左期望差（left expected deviation）、右期望差（right expected deviation）。

均值差（mean deviation）、左均值差（left mean deviation）、右均值差（right mean deviation）。

中位差（median deviation）、左中位差（left median deviation）、右中位差（right median deviation）。

众数差(mode deviation)、左众数差(left mode deviation)、右众数差(right mode deviation)。

峰值差（peat deviation）、左峰值差（left peat deviation）、右峰值差（right peat deviation）。

标准差（standard deviation）、左标准差（left standard deviation）、右标准差（right standard deviation）。

目标差（target deviation）、左目标差（left target deviation）、右目标差（right target deviation）。

分布中心差（center deviation）、左分布中心差（left center deviation）、右分布中心差（right center deviation）。

规范中心差（norm center deviation）、左规范中心差（left norm center deviation）、右规范中心差（right norm center deviation）。

在单峰分布的条件下，以上术语的概念中，期望差与众数差和峰值差完全等价，称谓不同意义一致。

4. 变差、方差、西格玛三者的关系

综上所述表明变差、方差、西格玛三者的基本要素完全一致，都是随机变量与不同位置特征值离差程度的变异指标。从以上三者的定义和计算公式可知它们的不同之处及其相互之间的关系。

变差：离差距离和的平均（1 阶距）；

以此为基础，可以得到第二层次方差：离差平方和的平均（2 阶距）；

方差开平方后得到第三层次西格玛。

上述说明了变差、方差、西格玛三者的关系是：变差是基础，是 1 阶距。方差在变差的基础上是 2 阶距。方差与西格玛的关系则是：西格玛是方差的正平方根；方差是西格玛的平方。

它们在统计分析中必然有着不同的意义和作用。这将在后续的课题中详细论述。

5. 标志变异指标重新认识的基本概念

由岪诗松教授主编的《统计手册》解释了“标志变异指标是反映总体中各单位标志值分布特征的重要综合指标，又称标志变动度。它反映总体中各单位标志值的差异程度、变动范围或离散程度。其作用在于衡量平均数代表性的典型和现象的均衡性。常用的标志变异指标有全距、平均差、标准差、标准差系数、偏度和峰度六种[11]。”

以上点明了标志变异指标“其作用在于衡量平均数代表性的典型和现象的均衡性。”若仅仅是“衡量平均数代表性的典型和现象的均衡性。”这单一的作用，在现代的统计工作中是远远不够的，所以对传统的标志变异指标的解释存在一定的缺陷。由此说明标志变异指标注入新的内涵是统计发展的需要。基于这样的新认识。综上所述对标志变异指标作补充性的解释如下：

标志变异指标是反映总体中各单位标志值分布特征及变异程度的重要综合指标，又称标志变动度。它反映总体中各单位标志值的与不同位置特征值的差异程度、变动范围或离散程度。其作用在于衡量针对不同位置特征值代表性的大小和不同统计对象的均衡性。常用的标志变异指标有变差、方差、西格玛、变差系数、西格玛系数、偏度[12]和峰度七大类。

有关变差系数、西格玛系数所对应其拓展的内涵还需要在后续的研究中进行论述。

6. 结论

在概率统计中，要求对于研究对象的所有基本事件都要全部包括在其中。统计分析的实

践经验也要求，凡是统计研究对象所涉及到的所有因素也需要全部考虑在其中。作为传统的标志变异指标，随机变量与单一的平均值离散程度的变异指标是不可能把数据间的差异解释清楚的。将随机变量与不同位置特征值离散程度的变异指标以三个层次形成的一个系统的指标体系注入进标志变异指标的内涵中，这是对标志变异指标的重新认识。目的是：针对统计分析对象的不同要求采用不同的变异指标，共同将数据间存在的差异解释清楚。

参考文献

[1] 龚凤乾《化神奇为平易——张尧庭统计教育思想研究》[J] 统计研究 2008.9 第 25 卷（203 期） P93-94 页

[2] 王琪延 白日荣《统计在法律中的应用与展望》[J] 统计研究 2008.5 第 25 卷（199 期）P104 页

[3] 石庆焱 李伟 丁孜《第六届海峡两岸统计与概率学术研讨会综述》[J] 统计研究 2009.3 第 26 卷（209 期） P108 页

[4] 唐国兴《高等数学》（二）第二分册：概率统计[M] 武昌珞珈山 武汉大学出版社 1991.1

[5] 张公绪 孙静《质量工程师手册》[M] 北京市海淀区紫竹院南路 17 号 企业管理出版社 2002.2

[6] 孔建新 孔璐 何光伟《关于对单峰分布中心数位置特征的初步讨论》[OL] http://www.paper.edu.cn 2008.10.15

[7] 李洁明 祁新娥《统计学原理》[M] 上海市国权路 579 号 复旦大学出版社 2007.6 第四版

[8] 孔建新 孔璐 何光伟《解读随机变量与不同位置特征值的变差》[OL] http://www.paper.edu.cn 2009.0

[9]《变差 百度百科》[OL] http://baidu.com/viem/1855209.htm

[10] 孔建新 孔璐 何光伟《关于对西格玛内涵拓展的研究》[OL] http://www.paper.edu.cn 2008.12.10

[11] 峁诗松 《统计手册》[M] 北京东黄城根北街 16 号 科学出版社 2003.1

[12] 何光伟 孔璐 孔建新《单峰分布偏度计算公式改进的研究》[OL] http://www.paper.edu.cn 2009.05.07

About to symbol variation target concept new understanding

Kongjianxin[1] Konglu[2] Heguangwei[3]

1 Yunnan Hui Zhe Zhe Hai, 2-24-6 statistics scientific group
Zhe Hai, Hui Zhe Yun Nan 654211 Email: kongfanjx@163.com

2 Zhongshan Torch Development Zone in Guangdong Province Vocational and Technical College .Zhongshan City, 528437 E-mail：yudianlulu@163.com

3 Guangdong College of Pharmacy, Zhongshan District, Zhongshan City 528458
E-mail：weiweihonest@163.com

Abstract: Statistical analysis's central content is the data variation degree measure and the decomposition. However the present random variable symbolized that the variation target's concept and the connotation have certain flaw, already could not meet the difference explanation clear needs which has the data. This article knew to the variation carries on the synthesis induction in the earlier period to the Sigma connotation development's research results as well as in the foundation which, simultaneously will symbolize that the variation target's connotation also correspondingly carries on the development, then to symbol variation target concept new understanding.

Key Words: Symbol variation quota；Variation；Variance；Sigma