左右方差计算公式的推导与应用

孔建新
（云南省会泽老年科技工作者协会，云南 会泽 654211）

摘要：随机变量与不同位置特征值离差平方和的平均值是方差的新概念。频数分布客观存在的不对称性使不同位置特征值两边必然存在不相等的左右方差。计算公式涉及到频数比率的分配问题。对此进行推导并通过实例进行演算。

关键词： 方差；频数比率；平均方差；左平均方差；右平均方差
中图分类号：O211.3

Left-variance and right-variance of computing formula inference and application

KONG Jianxin
(Yunnan province Huize county elderly association of science and technology, Yunnan Huize 654211)

Abstract: Random variables with characteristic values at different positions and the mean value deviation is the variance of the new concept. Frequency distribution of asymmetry objective characteristic value of different positions on both sides there is not necessarily equal the between left-variance and right-variance. Formula involves the allocation of the frequency ratio. Derivation and examples of this for the application of compute.

Key words: variance;frequency ratio;mean variance;mean left-variance;mean right-variance

0 引言

传统方差是一个内容非常单一的概念，仅仅是指随机变量与平均值离差平方和的平均值。而在实践中统计研究对象的广泛性和统计特征的多样性使统计对象标志值的随机变量频数分布的位置特征值就不仅仅只是平均值、众数、中位数三种。统计对象根据质量特征的不同可以分为单侧规范、双侧规范、多侧规范三种。它们的分布规律不同，所对应统计对象质量指标的要求不同，使其被关注的位置特征值还包括有期望值（等同峰值与众数）、标准值、合格值、分布中心值、规范中心值等等。随机变量与不同位置特征值离差平方和的平均值就成为方差被扩展后的具体新内容，统计研究需要具有的丰富内涵被赋予了方差新的概念。又由于随机变量频数分布客观存在的不对称性就存在左右方差[1]。在计算左右方差时涉及到以对应位置特征值左右两边频数和对应频数的比率如何进行合理的左右分配问题。在前期的研究课题中虽然已经提出并进行过简单的推论，但是并没有在实际应用的计算公式中进行深入的讨论，尤其是没有把频数与对应频数的比率问题联系起来进行深究。

通过对左右方差计算公式的推导，解决左右频数的合理分配问题，以此为基础简化方差、左方差、右方差的计算公式。左右方差的提出与计算公式的简化为统计过程进行方差分析开辟一条新的思考之路。

本文以 2006 年全国人口变动情况抽样调查样本数据[2]为例，应用推导的新公式进行实际的运算。从计算平均方差、左平均方差、右平均方差的结果验证频数比率分配的正确性和简洁性。为下一步扩展计算左右变差、左右西格玛提供依据。同时为以后设计计算机的左右变差、左右方差、左右西格玛的计算函数软件程序奠定理论基础。

作者简介：孔建新（1950.6-），男，质量工程师、高级统计师，研究方向：高斯分布. E-mail: kongfanjx@163.com

1 平均方差、左平均方差、右平均方差、期望值的定义

随机变量与不同位置特征值离差平方和的平均值是与传统“方差”内涵完全不同的新概念。传统“方差”的概念所包括的内容仅仅是指：随机变量与平均值离差平方和平均值。它仅是新概念方差所包含内容的其中之一：平均方差。所以平均方差与传统“方差”是一个完全等价的概念，这是由它们具有相同的定义所决定。又由于平均值左右两边不一定对称，所以平均值两边的随机变量分别与平均值的离散程度就不一定相等，左平均方差与右平均方差的概念由此引出。而传统的“方差”一般说来是一个被限制在正态分布假设条件下的概念，所以就不存在左右“方差”之分。本文就以传统“方差”的概念为基础，还其真正名称：平均方差，连同客观存在不对称的左右平均方差分别给出具体的定义。

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

左平均方差（mean left-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：$m\sigma_-^2$ 。

右平均方差（mean right-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：$m\sigma_+^2$ 。

以下推导涉及到期望值的概念需要定义如下：

期望值（expectation value）：随机变量的频数分布在单峰的条件下是分布曲线最大值点上的取值。符号记为 μ 。满足随机变量 $x=\mu$ 时,$P(\mu)=(\sqrt{2\pi}\,\sigma_-)^{-1}$ 或 $P(\mu)=(\sqrt{2\pi}\,\sigma_+)^{-1}$ 为 $P(x)$ 的最大值的条件。在正态分布的条件下平均值等于期望值，在偏斜分布的条件下平均值与期望值无关。

2 平均方差、左平均方差、右平均方差的计算公式与推导

由于平均方差与传统方差是完全等价的概念，平均方差计算公式就与传统方差的计算公式完全一致。不再赘述。左平均方差、右平均方差的计算公式则有必要进行推导。

在单峰的条件下任意偏斜分布以期望值为垂线可看成是左右两个不同曲线正态分布的一半。容易理解，左边旋转 180^O 形成左正态分布，右边旋转 180^O 形成右正态分布。两个不同的正态分布其期望值等于平均值，简化平均方差的计算公式就此得出：

$$m\sigma^2=\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]\div n \qquad (1)$$

（1）式的对称性可以化简为下式：

$$m\sigma^2=\left[\sum_{i=1}^{n/2}(x_i-\bar{x})^2\right]\div(n\div 2) \quad x_i\le\bar{x} \text{ 或 } x_i\ge\bar{x} \qquad (2)$$

化简的公式（2）是为确定左右平均方差的计算公式提供理论的依据。其依据还在于正态分布的性质之一“正态曲线对于纵轴是对称的。曲线在任一 z 值上的高度，正好与曲线在该值的负数上的高度相同。[3]”从以上导出简化公式（2）式等价（1）式，得出结论：随机变量等于平均值的频数二分之一属于左边，二分之一属于右边。

由以上结论得出左右平均方差计算公式的表达式。

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：$n = n_- + n_+ = n_1 + n_2 + n_3$

根据以上设定的条件计算左平均方差：$m\sigma_-^2$，右平均方差：$m\sigma_+^2$ ，平均方差：$m\sigma^2$ 的公式如下：

左平均方差：$m\sigma_-^2 = \left[\sum_{i=1}^{n_-}(x_i - \overline{x})^2\right] \div n_- \quad x_i \le \overline{x}$ （3）

右平均方差：$m\sigma_+^2 = \left[\sum_{i=n_-+1}^{n}(x_i - \overline{x})^2\right] \div n_+ \quad x_i \ge \overline{x}$ （4）

平均方差：$m\sigma^2 = \left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right] \div n$ （5）

以上计算公式是从正态分布条件下进行的推论，就是说：

在正态分布的条件下（2）式的左右频数各为总体频数的二分之一，从而推出等于平均值的随机变量的频数一半属于左边，一半属于右边。因为 $n_1 = n_2$ 。

以上推导是以期望值（峰值）为中心演变为正态分布曲线进行的推导，随机变量频数还可以从任意分布形态来说明以上的结论。

任意随机变量的频数分布在不对称的条件下，平均值不等于期望值，以平均值为垂线分为左右两部分，左边向右旋转 180^{O} 就成为对称的分布曲线（对称但不一定是正态分布曲线。）根据由此形成对称的分布曲线容易推出：曲线在平均值的右边任意 z 值上的高度，正好与平均值左边曲线在对应该值点上的高度相同。计算该对称曲线的平均方差应用的公式为（1）式。由于对称，等价（1）式的公式必然是（2）式。又由于此对称的分布曲线是由左边向右旋转 180^{O} 形成的，所以是左平均方差。

同理，右边也一样。

综合以上几方面的推论得出如下结论：

随机变量的频数分布在任何条件下，随机变量与不同位置特征值计算左右方差，等于位置特征值随机变量的频数一半属于左边，一半属于右边。

以上得出的结论还有一个充分的理由：不论如何进行分布频数的分组，就是说不论组距如何确定，所在组的标志值一般都是按组中值来确定的。组距内的频数与组距的区间数值虽然不一定均匀，但是作为位置特征值所在组，取的是组中值，所以位置特征值的频数按二分之一平均分属左右两边是合理的选择。

以上（3）（4）（5）式是简单的计算表达式，在实践应用中则存在一个加权的问题，即每一随机变量 x_i 都有对应的权数 f_i，所以计算方差就存在加权的以下公式。

左平均方差：$m\sigma_-^2 = \left[\sum_{i=1}^{n_-}(x_i - \overline{x})^2 f_i\right] \div n_- \quad x_i \le \overline{x}$ （6）

右平均方差：$m\sigma_+^2 = \left[\sum_{i=n_-+1}^{n}(x_i - \overline{x})^2 f_i\right] \div n_+ \quad x_i \ge \overline{x}$ （7）

平均方差：$m\sigma^2 = \left[\sum_{i=1}^{n}(x_i - \overline{x})^2 f_i\right] \div n$ （8）

3 平均方差实例计算的启示

以《中国统计年鉴》2006 年全国人口变动情况抽样调查样本数据[2]列表如下。

表 1 2006 年全国人口变动情况抽样调查样本数据

Tab. 1 Changes in the 2006 national population sample survey data

项目	年龄		人口数	比重	人口数（人）		比重（ % ）		性比（%）
序号	组距	组中值	人	%	男	女	男	女	男/女
1	0-4	2	60556	5.0774	33121	27435	2.7771	2.3003	120.73
2	5-9	7	70588	5.9185	38942	31646	3.2651	2.6534	123.06
3	10-14	12	89136	7.4737	48057	41079	4.0294	3.4443	116.99
4	15-19	17	105023	8.8057	55481	49542	4.6518	4.1539	111.99
5	20-24	22	76160	6.3857	37271	38889	3.1250	3.2607	95.84
6	25-29	27	74110	6.2138	35868	38242	3.0074	3.2064	93.79
7	30-34	32	93398	7.8310	45819	47580	3.8417	3.9894	96.30
8	35-39	37	113952	9.5544	56033	57920	4.6981	4.8563	96.74
9	40-44	42	115781	9.7077	57276	58505	4.8024	4.9054	97.90
10	45-49	47	76496	6.4139	38248	38247	3.2069	3.2068	100.00
11	50-54	52	90607	7.5970	45643	44964	3.8270	3.7700	101.51
12	55-59	57	68277	5.7247	34539	33737	2.8959	2.8287	102.38
13	60-64	62	48886	4.0989	24947	23939	2.0917	2.0072	104.21
14	65-69	67	39996	3.3535	20460	19536	1.7155	1.6380	104.73
15	70-74	72	32692	2.7411	16142	16550	1.3534	1.3876	97.53
16	75-79	77	20632	1.7299	9850	10781	0.8259	0.9039	91.36
17	80-84	82	10825	0.9076	4604	6221	0.3860	0.5216	74.01
18	85-89	87	4142	0.3473	1561	2581	0.1309	0.2164	60.48
19	90-94	92	1130	0.0947	366	764	0.0307	0.0641	47.91
20	95≤	97	279	0.0234	75	204	0.0063	0.0171	36.76
—	总计	—	1192666	100	604303	588362	50.6683	49.3317	102.71

根据表 1 选择女性人口年龄及频数资料计算平均值及平均方差如下：

表 2 2006 年全国人口女性年龄分布样本资料计算表

Tab. 2 Female population in 2006 sample data computation of the age distribution

分组序号	年龄		人口	比率	与平均值离差	离差平方	加权离差平方	比率离差平方	年龄比率
	组距	组中值	人	%					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	0-4	2	27435	4.6629	-35.00	1225	33607875	57.12	0.09
2	5-9	7	31646	5.3787	-30.00	900	28481400	48.41	0.38
3	10-14	12	41079	6.9819	-25.00	625	25674375	43.64	0.84
4	15-19	17	49542	8.4203	-20.00	400	19816800	33.68	1.43
5	20-24	22	38889	6.6097	-15.00	225	8750025	14.87	1.45
6	25-29	27	38242	6.4997	-10.00	100	3824200	6.50	1.75
7	30-34	32	47579	8.0867	-5.00	25	1189475	2.02	2.59
8	35-39	37	57919	9.8441	0.00	0	0	0.00	3.64
9	40-44	42	58505	9.9437	5.00	25	1462625	2.49	4.18
10	45-49	47	38248	6.5007	10.00	100	3824800	6.50	3.06
11	50-54	52	44964	7.6422	15.00	225	10116900	17.19	3.97
12	55-59	57	33738	5.7342	20.00	400	13495200	22.94	3.27
13	60-64	62	23939	4.0687	25.00	625	14961875	25.43	2.52
14	65-69	67	19536	3.3204	30.00	900	17582400	29.88	2.22
15	70-74	72	16550	2.8129	35.00	1225	20273750	34.46	2.03
16	75-79	77	10782	1.8325	40.00	1600	17251200	29.32	1.41
17	80-84	82	6221	1.0573	45.00	2025	12597525	21.41	0.87
18	85-89	87	2581	0.4387	50.00	2500	6452500	10.97	0.38
19	90-94	92	764	0.1299	55.00	3025	2311100	3.93	0.12
20	95≤	97	204	0.0347	60.00	3600	734400	1.25	0.03
—	总计	—	588363	100.00	—	—	242408425	412.00	36.24
平均方差			—	—	—	—	412.00	412.00	—

表 2 第（8）列是按传统加权计算方差的公式得出的结果，本列总计栏数据为加权离差平方和 242408425，除以人口总计 588363 得到平均方差 412.00 。表 2 第（9）列比率离差平方是第（7）列离差平方乘以对应第（5）列人口比率再除以 100 得出的结果。表 2 第（9）列之和正好与按传统离差平方和的平均值完全一致。通过表 2 的计算给出一个重要的启示：以随机变量的频数对应计算的比率作为权数与传统以频数作为权数计算的结果完全一致。同理第（10）列年龄比率是第（3）年龄组中值乘以对应第（5）列人口比率再除以 100 得出的结果。其合计得到平均年龄 36.24 ，得出 2006 年全国人口变动情况抽样调查样本数据女性平均年龄为 37 岁（36.24 岁超过 36 岁按 37 岁计）。根据上述得出的结论，需要将表 2 第（5）列平均值所在的第 8 行的比率 9.8441%平均分为 4.92205%分别归属于左右两边。通过计算，左边的比率之和为 51.5620% ；右边的比率之和为 48.4380% 。

通过表 2 第（9）列按比率离差平方的计算得出方差的较传统的简洁得多，由此可以对方差给出另外一种定义：随机变量与不同位置特征值频数比率离差平方之和。相应的平均方差定义为：随机变量与平均值频数比率离差平方之和。计算公式以（6）（7）（8）式为基础，将公式中权数（频数）f_i 表示为比率（注：此时的比率不再是百分比，下同）。进一步简化如下：

左平均方差：$m\sigma_{-}^{2} = \sum_{i=1}^{n_-}(x_i-\bar{x})^2 f_i \quad x_i \le \bar{x}$ （9）

右平均方差：$m\sigma_{+}^{2} = \sum_{i=n_-+1}^{n}(x_i-\bar{x})^2 f_i \quad x_i \ge \bar{x}$ （10）

平均方差：$m\sigma^{2} = \sum_{i=1}^{n}(x_i-\bar{x})^2 f_i$ （11）

以上按频数比率离差平方之和计算的方差为计算左右方差提供理论的依据。就是说计算左右方差是可以应用频数比率离差平方之和的方法来计算的。需要说明的问题是：这个频数比率是按独立的各自左右两边来计算。推论略。以上清楚表明，公式（9）（10）（11）对应（6）（7）（8）式更为简洁。

4 左平均方差、右平均方差的计算

对表 2 女性人口年龄及频数资料进行必要的整理，计算左右平均方差列表如下：

表 3 2006 年全国人口女性年龄分布样本资料计算表

Tab. 3 Female population in 2006 sample data computation of the age distribution

分组序号 (1)	年龄 组距 (2)	年龄 组中值 (3)	人口 人 (4)	比率 % (5)	与平均值离差 (6)	离差平方 (7)	左边比率离差平方 (8)	右边比率离差平方 (9)
1	0-4	2	27435	9.0434	-35.00	1225	110.78	—
2	5-9	7	31646	10.4314	-30.00	900	93.88	—
3	10-14	12	41079	13.5408	-25.00	625	84.63	—
4	15-19	17	49542	16.3305	-20.00	400	65.32	—
5	20-24	22	38889	12.8189	-15.00	225	28.84	—
6	25-29	27	38242	12.6057	-10.00	100	12.61	—
7	30-34	32	47579	15.6834	-5.00	25	3.92	—
8	35-39	37	28959.5	9.5459	0.00	0	0.00	—
—	左边合计		303371.5	100.00	—	—	399.99	—
8	35-39	37	28959.5	10.1615	0.00	0	—	0.00
9	40-44	42	58505	20.5287	5.00	25	—	5.13
10	45-49	47	38248	13.4208	10.00	100	—	13.42
11	50-54	52	44964	15.7773	15.00	225	—	35.50
12	55-59	57	33738	11.8382	20.00	400	—	47.35
13	60-64	62	23939	8.3999	25.00	625	—	52.50
14	65-69	67	19536	6.8549	30.00	900	—	61.69

15	70-74	72	16550	5.8072	35.00	1225	—	71.14
16	75-79	77	10782	3.7833	40.00	1600	—	60.53
17	80-84	82	6221	2.1829	45.00	2025	—	44.20
18	85-89	87	2581	0.9056	50.00	2500	—	22.64
19	90-94	92	764	0.2681	55.00	3025	—	8.11
20	95≤	97	204	0.0716	60.00	3600	—	2.58
—	右边合计		284991.5	100.00	—	—	—	424.80

表 3 是依据上述的结论：随机变量等于平均值的频数需要平均分配给左右两边，所以将平均年龄 37 岁所在组的频数 57919 分为 28959.5 分别计入左右两边。分别计算左右两边独自的比率。表 3 第（8）列左边比率离差平方和第（9）列右边比率离差平方分别是对应表 3 第（5）列乘以第（7）列再除以 100 的结果。

根据公式（9）（10）（11）具体计算根据表 2、表 3 的计算结果得出：

表 3 左平均方差：$m\sigma_-^2 = \sum_{i=1}^{n-}(x_i-\bar{x})^2 f_i = 399.99$

表 3 右平均方差：$m\sigma_+^2 = \sum_{i=n-+1}^{n}(x_i-\bar{x})^2 f_i = 424.80$

表 2 平均方差：$m\sigma^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 f_i = 412.00$

对应平均方差的平方根的平均西格玛称为：平均差。（为与传统概念的平均差相区别，将传统的平均差称为：平均变差）

左平均差：$m\sigma_- = [\sum_{i=1}^{n-}(x_i-\bar{x})^2 f_i]^{1/2} = 399.99^{1/2} = 20.00$

右平均差：$m\sigma_+ = [\sum_{i=n-+1}^{n}(x_i-\bar{x})^2 f_i]^{1/2} = 424.80^{1/2} = 20.61$

平均差：$m\sigma = [\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i]^{1/2} = 412.00^{1/2} = 20.30$

以上根据计算公式结合统计计算表能够清晰反映计算的全过程。即：计算公式是根据，统计表反映的是计算过程。从而使数学表达式与计算结果简洁明了。充分反映了应用 Excel 进行数学运算具有巨大的优越性。

5 左平均方差、右平均方差与整体平均方差关系的验证与推广

从以上表 2 和表 3 计算统计表得出的结果表明：计算方差的权数可以用频数也可以用比率来计算，其结果是完全一致的。

将频数和比率两种权数应用前期的研究成果“加权符：$\hat{+}$ [4]”来进行验证左平均方差和右平均方差与平均方差的关系。

即：$m\sigma^2 = m\sigma_-^2 \hat{+} m\sigma_+^2$，

用频数作权数，从表 3 已知左边频数之和为 303371.5 ，右边频数之和为 284991.5 。

有：$412.00 = 399.99 \hat{+} 424.80$

$$\text{验证等式右边} = \frac{399.99\times303371.5}{303371.5} \hat{+} \frac{424.80\times284991.5}{284991.5}$$

$$= \frac{399.99\times303371.5+424.80\times284991.5}{303371.5+284991.5}$$

$$= \frac{121344150+121064275}{588363}$$

$$=\frac{242408425}{588363}$$

$$=412.00$$

验证等式右边结果等于左边。

用比率作权数，从表 2 第(5)计算已知左边的比率为 51.562%；右边的比率为 48.438%。

有：$412.00=399.99\widehat{+}424.80$

$$验证等式右边=\frac{399.99\times51.562}{51.562}\widehat{+}\frac{424.80\times48.438}{48.438}$$

$$=\frac{399.99\times51.562+424.80\times48.438}{51.562+48.438}$$

$$=\frac{20624+20576}{100}$$

$$=\frac{41200}{100}$$

$$=412.00$$

验证等式右边结果等于左边。

通过以上分别用频数和比率作权数验证左右平均方差与平均方差的关系，说明平均方差是左平均方差和右平均方差的加权平均值。其权数可以是频数也可以是频数的比率。

以上应用加权符：$\widehat{+}$ 验证了左右方差与方差的关系。同理，推广应用于验证西格玛与左右西格玛的关系。即验证：左平均差和右平均差与平均差的关系。

已知：左平均差 $m\sigma_{-}=20.00$；右平均差 $m\sigma_{+}=20.61$；平均差 $m\sigma=20.30$

验证：$m\sigma=m\sigma_{-}\widehat{+}m\sigma_{+}$ 即：$20.30=20.00\widehat{+}20.61$

用比率作权数，已知左边的比率为 0.51562；右边的比率为 0.48438。

$$验证等式右边=\frac{20.00\times0.51562}{0.51562}\widehat{+}\frac{20.61\times0.48438}{0.48438}$$

$$=\frac{20.00\times0.51562+20.61\times0.48438}{0.51562+0.48438}$$

$$=\frac{10.3124+9.9831}{1}=20.2955\approx20.30$$

验证等式右边结果等于左边。

验证结果表明，西格玛是左西格玛和右西格玛的加权平均值。加权符：$\widehat{+}$ 应用于左右方差和左右西格玛来求取整体方差和整体西格玛都是通用的。

6 左右方差提出的理论依据

统计学界教育大家张尧庭先生的统计思想："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。[5]"根据张教授的统计思想将变异指标方差注入新的内涵，扩展研究随机变量与不同位置特征值的差异，并将方差分解为左右两部分就能把数据间存在的差异解释得更清楚一些。分部（左右）描述则是范剑青教授局部建模的优点在于可以大大降低误差的新理念。[6] 正如应用数学大师林家翘教授所述："自然界的事物基本上都很简单，所有的基础原理及主要问题都可以用数学

方式表达，这是应用数学家的一个信仰。"他归国八年以来绝少面对媒体，但几乎每一次出现在镜头前他都会强调这一点——应用数学的意义在于揭示自然界和社会实际问题的规律。[7]左右方差的提出正是揭示随机变量与不同位置特征值变异规律性。所依据的也正是应用数学大师们的理论。左右方差计算公式的改进实际上是统计发展的需要，也是数学发展的必然结果。

7 左右方差计算公式的推广应用

在统计实践中应用直方图表达随机变量的频数分布与表达随机变量的频率分布的图形是完全一致相同的。由此证明了计算方差的权数可以应用频数也可以应用频数比率。频数与频率在统计分析中具有不同的作用。频数反映的是随机变量值出现的具体频次，而频率反映的是随机变量值出现的具体频次在样本总数中所占的比率。频率是在频数的基础上进一步的统计结果，在进行统计分析中是一项不可缺少的重要指标，同时也是进行统计推断和统计预测的重要依据。既然频率是对频数再整理，应用频率简化计算方差的公式是一进步。在统计实践中，一些统计对象的频数非常大，应用频率来描述就简单得多。如：人口普查中人口年龄频数的分布就是其中的实例。

在前期研究的课题中，有关随机变量与不同位置特征值离散程度的变异指标被分为变差、方差、西格玛三个层次来反映。[1]由于随机变量的频数分布客观存在的不对称性，就存在左右变差、左右方差、左右西格玛。左右方差简化公式的提出使其推广应用于以上变异指标的计算中。

各指标的具体的代表符号分别如下：

变差 δ、左变差 δ_-、右变差 δ_+；方差 σ^2、左方差 σ_-^2、右方差 σ_+^2；西格玛 σ、左西格玛 σ_-、右方差 σ_+。

随机变量 x ，频数比率 f ，频数组序 n ，左频数组序 n_- ，右频数组序 n_+ 。

以期望值 μ 代表不同的位置特征值。通用计算公式分别表达如下：

变差：$$\delta = \sum_{i=1}^{n} |x_i-\mu| f_i \tag{12}$$

左变差：$$\delta_- = \sum_{i=1}^{n_-} |x_i-\mu| f_i \quad x_i \le \mu \tag{13}$$

右变差：$$\delta_+ = \sum_{i=1}^{n_+} |x_i-\mu| f_i \quad x_i \ge \mu \tag{14}$$

方差：$$\sigma^2 = \sum_{i=1}^{n} (x_i-\mu)^2 f_i \tag{15}$$

左方差：$$\sigma_-^2 = \sum_{i=1}^{n_-} (x_i-\mu)^2 f_i \quad x_i \le \mu \tag{16}$$

右方差：$$\sigma_+^2 = \sum_{i=1}^{n_+} (x_i-\mu)^2 f_i \quad x_i \ge \mu \tag{17}$$

西格玛：$$\sigma = \left[\sum_{i=1}^{n} (x_i-\mu)^2 f_i\right]^{1/2} \tag{18}$$

左西格玛：$$\sigma_- = \left[\sum_{i=1}^{n_-} (x_i-\mu)^2 f_i\right]^{1/2} \quad x_i \le \mu \tag{19}$$

右西格玛：$$\sigma_+ = \left[\sum_{i=1}^{n_+} (x_i-\mu)^2 f_i\right]^{1/2} \quad x_i \ge \mu \tag{20}$$

左右方差的提出与推广应用在统计过程中为进行方差分析开辟一条新的思路，并将其引向深入起到一定的积极作用

8 平均变差、左平均变差、右平均变差的计算与其关系的验证

随机变量与不同位置特征值的标志变异指标体系除方差和西格玛外，还有最为基础的变差。借此也进行验证。根据表 1 资料和计算变差公式（12）（13）（14）计算平均变差及平均左右变差见表 4：

表 4 2006 年全国人口女性年龄分布样本资料计算表

Tab. 4 Female population in 2006 sample data computation of the age distribution

分组序号 (1)	年龄 组距 (2)	年龄 组中值 (3)	人口人 (4)	左右比率 (5)	整体比率 (6)	离差绝对值 (7)	离差频数比率离差 (8)	左边频数比率离差 (9)	右边频数比率离差 (10)
1	0-4	2	27435	0.0904	0.0466	35	1.6320	3.17	—
2	5-9	7	31646	0.1043	0.0538	30	1.6136	3.13	—
3	10-14	12	41079	0.1354	0.0698	25	1.7455	3.39	—
4	15-19	17	49542	0.1633	0.0842	20	1.6841	3.27	—
5	20-24	22	38889	0.1282	0.0661	15	0.9915	1.92	—
6	25-29	27	38242	0.1261	0.0650	10	0.6500	1.26	—
7	30-34	32	47579	0.1568	0.0809	5	0.4043	0.78	—
8	35-39	37	28960	0.0955	0.0492	0	0.0000	0.00	—
—	左边合计		303372	1.0000	0.5156	—	8.7209	16.91	—
8	35-39	37	28959.5	0.1016	0.0492	0	0.0000	—	0.00
9	40-44	42	58505	0.2053	0.0994	5	0.4972	—	1.03
10	45-49	47	38248	0.1342	0.0650	10	0.6501	—	1.34
11	50-54	52	44964	0.1578	0.0764	15	1.1463	—	2.37
12	55-59	57	33738	0.1184	0.0573	20	1.1468	—	2.37
13	60-64	62	23939	0.0840	0.0407	25	1.0172	—	2.10
14	65-69	67	19536	0.0685	0.0332	30	0.9961	—	2.06
15	70-74	72	16550	0.0581	0.0281	35	0.9845	—	2.03
16	75-79	77	10782	0.0378	0.0183	40	0.7330	—	1.51
17	80-84	82	6221	0.0218	0.0106	45	0.4758	—	0.98
18	85-89	87	2581	0.0091	0.0044	50	0.2193	—	0.45
19	90-94	92	764	0.0027	0.0013	55	0.0714	—	0.15
20	95≤	97	204	0.0007	0.0003	60	0.0208	—	0.04
—	右边合计		284992	1.0000	0.4844	—	7.9586	—	16.43
	总计		588363	—	1.0000	—	16.68	—	—

根据公式及表 4 计算得出计算结果如下：

平均变差：$m\delta = \sum_{i=1}^{n} \left| x_i - \bar{x} \right| f_i = 16.68$

左平均变差：$m\delta_- = \sum_{i=1}^{n-} \left| x_i - \bar{x} \right| f_i = 16.91$

右平均变差：$m\delta_+ = \sum_{i=1}^{n+} \left| x_i - \bar{x} \right| f_i = 16.43$

验证左右 $m\delta = m\delta_- \hat{+} m\delta_+$ 即：$16.68 = 16.91 \hat{+} 16.43$

用比率作权数，根据表 4 已知：左边的频数比率为 0.5156；右边的频数比率为 0.4844 。

$$\text{验证等式右边} = \frac{16.91\times 0.5156}{0.5156} \hat{+} \frac{16.43\times 0.4844}{0.4844}$$

$$= \frac{16.91\times 0.5156 + 16.43\times 0.4844}{0.5156+0.4844}$$

$$= \frac{8.7188+7.9587}{1}$$

$$= 16.6775$$

$$\approx 16.68$$

验证等式右边结果等于左边。

验证结果表明，平均变差是左平均变差和右平均变差的加权平均值。通过对变差、方差、西格玛与对应左右变差、左右方差、左右西格玛的验证，证明应用加权符：$\widehat{+}$ 都能表达它们各自的关系。进而说明应用频数比率作为权数是一种比频数作为权数更为简洁的方法。

9 结论

统计对象的广泛性与统计特征的多样性使随机变量的频数分布具有诸多的不确定性。又由于统计研究对象的复杂性，使之随机变量具有不同的位置特征值，针对统计研究对象的不同，随机变量与不同位置特征值离散程度的变异指标就不仅仅只限于随机变量与平均值一种单一的变异指标。况且随机变量的不对称性是广泛存在的客观现象，所以还存在左右的变异指标。平均方差和左右平均方差仅仅是方差新概念的一个具体的内容，通过实例对左右平均方差的计算，推导出有关频数的科学分配，从而推导出较传统更简洁的计算方法。可推广应用到随机变量与其它位置特征值离散程度指标三个层次的计算中。为统计过程进行方差分析提供新的思考元素。本文的演算过程为将来计算机设计左右变差、左右方差、左右西格玛的计算函数软件奠定了理论基础。

[参考文献] (References)

[1] 孔建新,孔璐,何光伟.关于对标志变异指标概念的重新认识[OL].中国科技论文在线， 2009.5.18.
[2] 中国统计年鉴（2006） [OL] http://www.sei.gov.cn/try/hgjj/yearbook/2007/indexCh.htm
[3] [美] G·H·维恩堡，J·A·休息麦克 D·奥尔特曼.数理统计初级教程[M] 常学将,胡文明,王明生. 等译 太原:山西人民出版社 1986.8 P157 页
[4] 孔建新. 关于对左期望方差和右期望方差与期望方差数学关系表达符的探讨[OL].中国科技论文在线, 2010.4.21.
[5] 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.
[6] lixing.范剑青:把数学作为解决社会问题的工具 [OL] 经济学论坛-中国经济学教育科研网 2008-2-13
[7] 刘文嘉.林家翘:大师之忧[N].光明日报 2010 年 5 月 7 日.第 1 版.