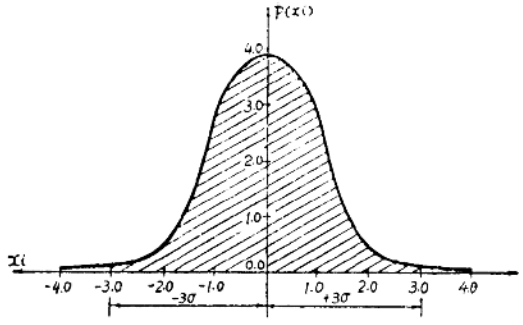

TIYU TONGJI XUE

体育统计学

北京体育学院编译室

目　　录

前　　言

有效地安排运动训练过程，是体育运动的基本任务之一。

要解决这个问题，必须全面研究运动员机体发展的规律和各种体育手段对运动员机体的影响。

训练课的总方案即便安排得合理，也要根据运动员的个人情况作相当大的变动。这就是说，随时随地都要考虑到运动员的个人特点、生物特点和心理特点、年龄、性别、以及训练课的目的和任务。

此外，目前对运动员机体发展的一般规律和特殊规律，以及它们之间的相互作用，研究得还很不够，所以有效地安排训练过程，是体育运动的一个很复杂的问题。这需要有渊博的知识，丰富的经验，相当的直觉力和创造性。要研究这个问题，就要从质量和数量上对涉及运动员生活和活动的一切现象进行分析。

数学的手段和方法应用极为广泛，在体育科学中也是不可缺少的。教练员和体育教师收集了大量数据（教育学的、心理学的，组织管理学的，生理学的、医学的等等）。对这些数据要加以研究，并在研究这些数据的基础上：拟定合理的体育训练方法，预测运动成绩，探索潜在过程和潜在现象之间的相互关系，掌握运动员专项技能成长的动态，了解体育手段对运动成绩的影响，分析运动员在各训练阶段的活动，求体育活动的参数，从医学和教育学上找根据——总之，在体育运动中用数学方法来解决的问题不限于这些。

竞技运动是如此，群众性体育也不例外。各年龄组和职业组的训练必须严格根据训练课的目的、社会任务、从事者的健康状况、兴趣和需要来安排。因人而异地选择负荷和探索合理的训练形式，都得进行科学研究和实验，这就必须借助数学的手段和方法。

在数学的各分支中应用最广的就是数理统计方法。数理统计学是体育学院的一门必修课，大多数体育研究生、体育科研人员、体育教师和教练员在工作中都得使用数理统计方法。

控制论在体育运动中已广泛使用。训练的过程实际上就是对运动员机体适应各种外界作用的控制过程，所以控制论对运动训练能起很重要的作用。

在建立控制体育运动的组织机构体制时也得应用控制论和信息理论，因为这里主要的任务是根据专门的通过信息提供的参数进行控制。

目前，在科研中利用积分微分方程式进行数学分析的方法已越来越多了。这种方法无论就问题的解决和问题的提出都是十分复杂的。用积分微分方程式描述研究的过程要求科研人员博学多才，有专门的数学素养，有很好的直觉能力，对体育运动有全面而深入的了解，对所研究的问题有足够的了解。在体育科学研究中伴随采用数学分析方法会产生许许多多逻辑方面和技术方面的困难，但是这一切并没有阻挡住科研人员。近来，有很多采用数学分析法写成的著作已问世。

可以设想，今后体育运动的各方面都将采用愈益复杂的数学方法进行数量上的研究。今后的教练员和运动员一定会学到更高深的数学理论，而各种应用数学分支都将成为解决体育运动的多种问题所不可缺少的工具。

就是在今天，体育运动文献中的数理统计学的思想和方法已经比比皆是，不懂数理统计学的教练员已不能提出和解决现代体育运动中的问题了。

本书介绍一些最常用的数理统计学方法。内容深入浅出，通俗易懂，供没有专门学过数学和没有听过体院数理统计学课的教练员和体育教师使用。

本书对体育院校师生、研究生和科研人员以及一切从事体育运动数量研究的人员都有参考价值。

第一章　如何处理原始统计资料

本章讲的是平均数方法，即变量数列。这是数理统计学中最常用的一部分。变量数列有助于说明一个总体，即具有共同特征的一群数据。

一般说，教练员和体育教师经常会收集到这样的群数，即具有一定特征，但凌乱而未经概括化的许多指标。譬如，脉搏频率、克服阻力的肌力、重复某种练习的次数、运动成绩、人体测量数据等。这种凌乱的群数不会给教练员以任何有价值的信息，因为它还不成体系。因此，教练员必须把这种原始数据变为一定的数学体系，使其参数能详细说明这个体系的性质和保证得到有用的信息。

利用这个体系，就能及时调整训练过程，预测运动成绩，规定一些合理的要求和解决体育运动中的其他很多问题。

在谈变量数列的原理之前，介绍几个数学符号。不掌握这些符号，就不能判断变量数列的特征，也无法解决其他数学问题。

几　个　数　学　符　号

大家知道，初等数学用数字进行运算，而高等数学用字母进行运算。这是因为高等数学所研究的是一些更为复杂、更为概括的过程和现象。

数理统计学也用字母符号进行运算。现举例说明使用字母符号的必要性。

例 1　收集到了一些运动员的比赛成绩：4′02″；4′01″；4′05″；4′08″；4′10″——五名二级自行车运动员 3 公里的成绩（男）；

11″8；12″0；12″3——三名100米赛跑运动员的成绩（女）；

69.40米；70.12米；71.20米；74.00米——四名一级标枪运动员的成绩（男）。

从上例中可看出，每一群数都是一些在性质上互不相同的指标。

每一群数我们都用一个符号表示（用拉丁字母表中最常用的字母）。自行车运动员的成绩用符号X表示，赛跑运动员的成绩用Y表示，标枪运动员的成绩用Z表示。为了不必把每一群数的所有数字都详细列出来，只用一些符号就可以了。可以说："把全部X相加"来代替4′02″+4′01″+4′05″+4′08″+4′10″；"用5除全部Z"等等。从该例中可看出，如果每一群数有很多数字的话，那么为了简便起见，用符号来代替是很必要的。

在一群数中每个数的顺序位置用字母旁的数字表示。例如，自行车运动员的成绩可表示为：4′02″是X_1，因为4′02″在用X符号表示的那个群数中居首位；4′04″是X_2，因为4′04″位于第二；4′05″是X_3；4′08″是X_4；4′10″是X_5。赛跑运动员的成绩表示为：11″8——Y_1；12″0——Y_2；12″3——Y_3等等。标枪运动员的成绩表示为：60.40米——Z_1；70.12米——Z_2；71.20米——Z_3；74.00米——Z_4。

还有用字母表示的指数。最常用的是"i"和"n"。

指数"i"表示有一定顺序数列的任何一个数。例如，Z_i表示Z群数中的任何一个数。在上面引用的例子中Z_i既是Z_1 = 69.40米，又是Z_2 = 70.10米，又是Z_3 = 71.20米和Z_4 = 74.00米。

指数"n"表示在每一群数中最后一个数。例如，符号X_n表示群数X中的最后一个数。例如在上述例子中第5个数为最后一个数，$X_n = X_5 = 4'10''$。

如果需要把一群数相加，那便用和的符号Σ（这是希腊字母表中的大写字母）。

符号Σ右面是要相加的那一群数的符号。例如，ΣX_i表示群数X相加。

在符号Σ下面应该写加法开始时的那个数的指数，在Σ上面则是加法结束时的那个数的指数。例如，$\sum_{i=1}^{i=4} x_i$ 表示：群数X所有数从第1个数到第4个数都要相加。

在有些情况下，为了方便起见，$\sum_{i=1}^{i=4} x_i$ 可用 $\sum_{1}^{4} x_i$ 代替，或 $\sum_{i=1}^{i=n} z_i$ 用 $\sum_{1}^{n} z_i$ 代替等。

有时需要将一群数，特别是一大群数相加，同时不逐一按顺序相加。在这种情况下就得用2个数、3个数……之和再相加。

例如，将自行车运动员的成绩，除第3个数外，都加起来，那就成为：

$$\sum_{i=1}^{i=2} x_i + \sum_{i=4}^{i=5} x_i$$

$\sum_{i=1}^{i=n} x_i$ 或 $\sum_{i=1}^{i=n} y_i$ 的和可以作为一个数，再作进一步的处理。

例如，要把赛跑运动员们的成绩之和除以K，就应写成为：

$$\frac{\sum_{i=1}^{i=n} y_i}{k}$$

或者要把标枪运动员的成绩之和乘以C，就应写为：

$$C \cdot \sum_{i=1}^{i=n} z_i$$

可见，当数目多时，用符号代替数字进行运算就简便多了。可以随意选择字母作为符号和指数（常用符号Σ除外），当然最好用拉丁字母，其次是希腊字母。

下面我们就利用上述的几个符号，开始讲变量数列。

变 量 数 列 的 排 列

具有共同特征的一群数称为总体。

如上所述，原始的运动统计资料只是一群凌乱的数字，无助于教练员了解运动训练的现象或过程。因此，必需把一群数字变成一个体系，从其中获得所需要的信息。

排列变量数列，正好就是组成一定的数学体系。现举例说明。

例2　测定了34名滑雪运动员训练后脉搏恢复正常所需要的时间（秒）：

81；78；84；90；78；
74；84；85；81；84；
79；84；74；84；84；
85；81；84；78；81；
74；84；81；84；85；
81；78；81；81；84；
84；84；78；81。

显然，从这一群数中我们得不到任何信息。

为了排列变量数列，先要按数的大小排列顺序，或从小往大排，或从大往小排。例如，从小往大排的结果为：

74；74；74；74；
78；78；78；78；78；78；
81；81；81；81；81；81；81；81；81；
84；84；84；84；84；84；84；84；84；84；84；
85；85；85；
90。

从大往小排的结果为：

90；
85；85；85；
84；84；84；84；84；84；84；84；84；84；84；
81；81；81；81；81；81；81；81；81；

78；78；78；78；78；78；

74；74；74；74。

经这样排列之后，可以清楚地看出，这群数的外形是参差不齐的，有的数字多次重复。要想一日瞭然地看出每个数重复多少次数，还得改变一下书写形式。例如，可将从小往大排的形式改成为：

74——4；

78——6；

81——9；

84——11；

85——3；

9——1。

左边的数字，表示运动员脉搏恢复时间，右边的数字表示脉搏恢复时间的重复次数。

我们可以根据上面介绍使用数学符号的方法，用字母X来表示这一群测定的数字。考虑到这群数是从小往大排列的，因此可以写成：X_1——74秒；X_2——78秒；X_3——81秒；X_4——84秒；X_5——85秒；X_6——Xn——90秒，这里每个数据都可用符号Xi表示。

我们用字母n表示重复次数。那末，$n_1=4$；$n_2=6$；$n_3=9$；$n_4=11$；$n_5=3$；$n_6=n_n=1$，而每个重复数字都可用n_i表示。

测定的数据总数是34。这就是说，例中各个n之和等于43。用符号表示：

$$\sum_{i=1}^{i=n} n_i = 34$$

现在我们用一个字母n表示这个和。那末，该例的原始资料可写成如表1所示(表1)。

这群数字是根据教练员在研究开始时所得到的凌乱的数据整理出来的。

这群数有一定的体系，其参数能说明测定的性质。

测定结果的数(Xi)叫做变量(变数)；n_i——重复次数，叫做频数；n——全部频数之和，是总体量。

所得的整个体系叫做变量数列，有时叫做经验数列或统计学数列。

不难看出，也可能出现各个频数都等于1，即$n_i=1$的变量数列，就是说，在这群数据中每个数据只测到1次。当然，这是极个别的情况。

变量数列也可用图表表示。制变量图表时，首先要在水平轴和垂直轴上做出按一定比例的标度。

表1

X_i	n_i
74	4
78	6
81	9
84	11
85	3
90	1
	n=34

例如，可先在水平轴上标出脉搏恢复时间的值(X_i)，选择适当的单位长度相当于1秒。标出这些值时，要稍许离开两轴的交接点O，从70秒开始。

在垂直轴上标出频数值(n_i)，使任选的单位长度相当于频数单位。

制图的条件已具备了，现在就开始把所获得的变量数列用图表示。

将第一对数$X_1 = 74$，$n_1 = 4$列入统计图表中：在Xi轴上$X_1 = 74$；并从此点作一条垂直线，而在n轴上$n_1 = 4$，并从这点作水平线与垂直线相交。这两条线都是辅助线，因此，在图表中用虚线表示。图中两线的交点就是$X_1 = 74$和$n_1 = 4$之比（图示点）。依此类推，作出图中其余各点。然后各交点分别用一段一段的直线联结起来。为了使该图有封闭的外形，将尽端的两个点与水平轴上相邻的点分别用线段联结起来。

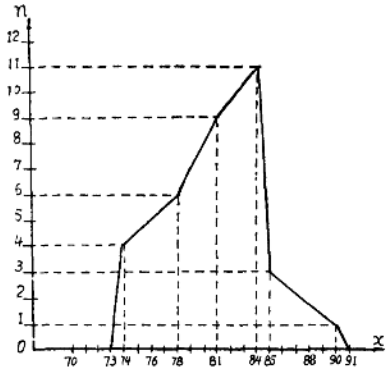

图1　　变量数列图

这样做出的图表就是变量数列的图（图1）。

十分清楚，每个变量数列都可做成图。

从图1中可见：1、在测试的运动员中脉搏恢复时间为84秒的为数最多；2、其次是81秒；3、最小数为74秒，最大数为90秒。

因此，做完一组试验后，应该把所获得的数按大小排列，组成一定的数学体系的变量数列。为醒目起见，可将变量数列用图表示。

上述的变量数列中每个变量都是用一个数表示的，这种数列可称为离散数列。此外，还有间隔数列。在间隔数列中每个变量是用间隔（组限），即从…到…的数值界限表示的。

让我们再次分析例2。假设数列的间隔为5秒，那我们就会看出，脉搏恢复时间从74秒到74秒＋5秒＝79秒的，共4＋6＝10人；从79秒到79秒＋5秒＝84秒的，共9＋11＝20人；从84秒到84秒＋5秒＝89秒的共3人；从89秒到89秒＋5秒＝94秒的1人。这样就得出如下的间隔数列（表2）：

表2

X_i	n_i
74—79	10
79—84	20
84—89	3
89—94	1
	n＝34

下面我们再举几个组成变量数列的例子。

例3　12名射击运动员卧射10发，成绩（分）分别为：

94；91；96；94；94；

92；91；92；91；95；

94；94。

为了组成变量数列，先将这些数据按从小到大的顺序排列为：

91；91；91；

92；92；

94；94；94；94；94；

95；

96。

然后组成变量数列（表3）。

表3

X_i	n_i
91	3
92	2
94	5
95	1
96	1
	n = 12

例4　26名中小学生投篮。每个学生只许试投12次，投中的次数分别为：

7；9；10；6；0；

8；8；10；0；7；

9；4；2；9；8；

12；5；5；0；6；

9；6；8；10；2；

9。

然后，按从小到大的顺序排列为：

0；0；0；

2；2；

4；

5；5；

6；6；6；

7；7；

8；8；8；8；

9；9；9；9；9；

10；10；10；

12。

然后组成变量数列（表4）。

表4

X_i	n_i
0	3
2	2
4	1
5	2
6	3
7	2
8	4
9	5
10	3
12	1
	n = 26

在运算熟练和测量的数据不多的情况下，可用口算方式，将数据按大小排列，直接得出变量数列。

变量数列的参数

变量数列的一些主要特征是：算术平均数、平均差、离势、标准差、变异系数、众数和中位数。

变量数列中最能说明特征和最重要的是算术平均数。它能表明数列中所有各测量数的平均水平。

还有简单的算数平均数（或不加权均数），它是所有变量X_i之和除以总体量n所得的商数。

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots\cdots + Xn}{n}$$

$$= \frac{\sum_{i=1}^{i=n} x_i}{n} \qquad (1)$$

加权平均数是计算变量频数（权数）的值，其算法如下：

1．每个变量（X_i）用其频数（n_i）相乘，即为：（$X_i n_i$）

2．将所得之积（$X_i n_i$）从第一个到最后一个数加起来，即为：$\sum_{i=1}^{i=n} x_i n_i$

3．所得之和 $\sum_{i=1}^{i=n} x_i n_i$ 用总体量n除，即为：$\frac{\sum_{i=1}^{i=n} x_i n_i}{n}$

我们用 $\overline{X}$ 表示算术平均数，因此求加权平均数公式如下：

$$\overline{X} = \frac{\sum_{i=1}^{i=n} x_i n_i}{n} \qquad (2)$$

换句话说，公式是用符号说明运算的顺序。

现在根据例2求算术平均数，并另加一栏（表5）。

$$\overline{X} = \frac{\sum_{i=1}^{i=n} x_i n_i}{n} = \frac{2762}{34} = 81.24\text{秒} \approx 81\text{秒}$$

从此例中可见，第三栏是每个变量与其频数相乘之积，而其下面的数则是各积之和。算术平均数不在表内。

表5

X_i	n_i	$X_i n_i$
74	4	296
78	6	468
81	9	729
84	11	924
85	3	255
90	1	90
	n = 34	2726

应该注意到，求出的算术平均数一般都不是整数，应当化整。化整的程度（保留小数点后一位或两位数等）要取决于测量的精确度。最好使算术平均数化整的程度与变量化整的程度一致起来。因此，上例中所获得的算术平均数 $\overline{X} = 81.24$ 秒，可化为整数，$\overline{X} = 81$ 秒因为变量都是整数。

求第3例中的算术平均数（表6）：

$$\overline{X} = \frac{\sum_{i=1}^{i=n} x_i n_i}{n} = \frac{1118}{12} = 93.1 \approx 93\text{分}$$

求第4例中的算术平均数（表7）：

$$\overline{X} = \frac{\sum_{i=1}^{i=n} x_i n_i}{n} = \frac{169}{26} = 6.5 \approx 6\text{次}$$

表6

X_i	n_i	$X_i n_i$
91	3	273
92	2	184
94	5	470
95	1	95
96	1	96
	n = 12	1118

表7

X_i	n_i	$X_i n_i$
0	3	0
2	2	4
4	1	4
5	2	10
6	3	18
7	2	14
8	4	32
9	5	45
10	3	30
12	1	12
	n = 26	169

如果在变量数列中所有的频数都等于1，即每种测量都是1次，那么$X_i n_i$相乘，就失去意义了，因为任何一个数乘1，其值不变。在这种情况下应直接将所有的变量数相加 $\sum_{i=1}^{i=n} x_i$ 然后用总体量n除其和：

$$\bar{X} = \frac{\sum_{i=1}^{i=n} x_i}{n} \qquad (1)$$

实际上这就是大家所熟知的按公式（1）求简单的算术平均数的法则。

算术平均数能反映变量数列的主要特征，但不能反映变量数列的所有特征，更不能反映现实变量按不同方式与平均水平相比较的事实。因此，还要采用变量数列的其他参数，这些参数能说明变量对算术平均数的分散程度。这一参数就是平均差d。

现在利用例2将变量数列表扩充，求平均差d（表8）。

表8

X_i	n_i	$X_i n_i$	$X_i - \bar{X}$	$\lvert X_i - \bar{X} \rvert$	$\lvert X_i - \bar{X} \rvert n_i$
74	4	296	− 7	7	28
78	6	468	− 3	3	18
81	9	729	0	0	0
84	11	924	+ 3	3	33
85	3	255	+ 4	4	12
90	1	90	+ 9	9	9
	n = 34	2762			100

$$\bar{X} = \frac{2762}{34} \approx 81\text{秒}$$

$$d = \frac{100}{34} = 2.9\text{秒}$$

从例中看出，在第四栏里是每个变量X_i减去所得的算术平均数$\bar{X}=81$秒之差。例如，74－81＝（－7）；78－81＝（－3）；81－81＝0。依此类推。在该栏里的数字分别说明，现实变量偏离算术平均数的程度。如，第一个运动员(X_1)脉搏恢复时间为78秒，比全组平均指标少7秒；第二个运动员($X_2=78$秒)——少3秒；最后一个指标$X_n=90$秒，比全组平均指标多9秒等。

弄清楚每个变量与平均数的差（$X_i-\bar{X}$），最好还要弄清楚所有测量数与平均值的平均差。为此，要求出所有差数的算术平均数。其方法与求所有变量的算术平均数一样。

但是，这样相加，$\sum_{i=1}^{i=n}(X_i-\bar{X})\cdot n_i$ 将会得到一个不大的数（甚至是0），因为正数与负数相加时要抵消。为了避免这种情况，必须把（$X_i-\bar{X}$）的正号或负号去掉，再与频数相乘$|X_i-\bar{X}|\cdot n_i$。因此，所有差数的算术平均数应按下列步骤求出：

1．每个差数(去掉符号)$|X_i-\bar{X}|$乘以频数(n_i)，即为

$$|X_i-\bar{X}|\cdot n_i$$

2．将所得的各积相加，即为

$$\sum_{i=1}^{i=n}|X_i-\bar{X}|\cdot n_i$$

3．将所得的和除以总体量，即为

$$\frac{\sum_{i=1}^{i=n}|X_i-\bar{X}|\cdot n_i}{n}.$$

故求平均差的公式为：

$$d=\frac{\sum_{i=1}^{i=n}|X_i-\bar{X}|\cdot n_i}{n} \qquad (3)$$

在上例中的平均差$d=2.9$秒，就是说，对该组来说，脉搏恢复的平均数是$\bar{X}=81$秒，而其平均差（包括正的和负的）为2.9秒。这组测量数的特征一般写为：（81 ± 2.9）秒。

与81秒的离差大于2.9秒者，那就意味着，这一运动员与该组其他运动员脉搏恢复时间的差别较大，他不能代表该组的特征。

例3的平均差也可按表9求出（表9）：

这组的特征为93 ± 1.5分，就是说，对该组来说，平均差为1.5分，不论是正的还是负的。

在例4中（表10）：

表9

X_i	n_i	$X_i \cdot n_i$	$X_i - \bar{X}$	$\lvert X_i - \bar{X} \rvert$	$\lvert X_i - \bar{X} \rvert \cdot n_i$
91	3	273	− 2	2	6
92	2	184	− 1	1	2
94	5	470	+ 1	1	5
95	1	95	+ 2	2	2
96	1	96	+ 3	3	3
	n = 12	1118			18

$$\bar{X} = \frac{1118}{12} = 93.1 \approx 93\text{分}$$

$$d = \frac{18}{12} = 1.5\text{分}$$

表10

X_i	n_i	$X_i \cdot n_i$	$X_i - \bar{X}$	$\lvert X_i - \bar{X} \rvert$	$\lvert X_i - \bar{X} \rvert n_i$
0	3	0	− 6	6	18
2	2	4	− 4	4	8
4	1	4	− 2	2	2
5	2	10	− 1	1	2
6	3	18	0	0	0
7	2	14	+ 1	1	2
8	4	32	+ 2	2	8
9	5	45	+ 3	3	15
10	3	30	+ 4	4	12
12	1	12	+ 6	6	6
	n = 26	169			73

$$\bar{X} = \frac{169}{26} = 6.5 \approx 6\text{次}$$

$$d = \frac{73}{26} = 2.8 \approx 3\text{次}$$

对这组中小学生平均的投篮命中数$\bar{X}$ = 6 次，平均差为 3 次。

在求平均差时，许可有些误差，这是由于去掉了变量前的正负号。为了避免这种误差，可用另一种求与算术平均数离散程度的方法，即求分散度和标准差。

大家知道，任何一个数的平方都得正数。因此，应求出带正、负号的差数平方：$(X_i - \bar{X})^2$，这样就避免了符号不一的现象。为了求这些平方，我们需先求算术平均数，因为标准差表示现实变量在平均数周围的离散程度。求标准差的方法：

1．差数平方$(X_i - \bar{X})^2$乘其频数n_i，即为$(X_i - \bar{X})^2 \cdot n_i$；

2．所得的各积相加 $\sum_{i=1}^{i=n}(X_i-\bar{X})^2\cdot n_i$

3．用总体量除所得之和，

即 $\dfrac{\sum_{i=1}^{i=n}(X_i-\bar{X})^2\cdot n_i}{n}$

这样求得的数叫做σ^2（西格玛方）。它表示变量在平均数周围的分散度，用下列公式求出：

$$\sigma^2=\frac{\sum_{i=1}^{i=n}(x_i-\bar{X})^2\cdot n_i}{n} \qquad (4)$$

在变量数列中每个变量都出现一次，即所有频数$n_i=1$的情况下，$(Xi-\bar{X})^2$乘ni，就失去了意义，这时求分散度的公式为：

$$\sigma^2=\frac{\sum_{i=1}^{i=n}(X_i-\bar{X})^2}{n} \qquad (5)$$

方差本身和其数值都是平方数，因此不便于进一步运算。为此，需要求方差的平方根。这个数值叫做标准差σ（西格玛）。

$$\sigma=\sqrt{\sigma^2} \qquad (6)$$

标准差的实际意义与平均差是一样的，但其值不同，因为是通过不同方法求得的。求方差和标准差都常用下列的表格。

在例2中（表11）：

表11

X_i	n_i	$X_i\cdot n_i$	$X_i-\bar{X}$	$\lvert X_i-\bar{X}\rvert^2$	$\lvert X_i-\bar{X}\rvert^2 n_i$
74	4	296	−7	49	156
78	6	468	−3	9	54
81	9	729	0	0	0
84	11	924	+3	9	99
85	3	255	+4	16	48
90	1	90	+9	81	81
	n=34	2762			438

$$\bar{X}=\frac{2762}{34}=81.24\approx 81\text{秒}$$

$$\sigma^2\ \frac{438}{34}=12.9\text{秒}^2$$

$$\sigma = \sqrt{12.9秒^2} = 3.6秒.$$

在这种情况下，这组的特征为（81±3.6）秒。

在例3中（表12）：

表12

X_i	n_i	$X_i \cdot n_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 n_i$
91	3	273	− 2	4	12
92	2	184	− 1	1	2
94	5	470	+ 1	1	5
95	1	95	+ 2	4	4
96	1	96	+ 3	9	9
	n = 12	1118			32

$$\bar{X} = \frac{1118}{12} = 93.1 \approx 93分$$

$$\sigma^2 = \frac{32}{12} = 2.66分^2$$

$$\sigma = \sqrt{2.66分^2} = 1.6分$$

该组特征为：（93±1.6）分

在例4中(表13)：

表13

X_i	n_i	$X_i \cdot n_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 n_i$
0	3	0	− 6	36	108
2	2	4	− 4	16	32
4	1	4	− 2	4	4
5	2	10	− 1	1	2
6	3	18	0	0	0
7	2	14	1	1	2
8	4	32	2	4	16
9	5	45	3	9	45
10	3	30	4	16	48
12	1	12	6	36	36
	n = 26	169			293

$$\bar{X} = \frac{169}{26} = 6.5 \approx 6次$$

$$\sigma^2 \frac{293}{26} = 11.2次^2$$

注：附录中4表和5表是平方表和平方根表。

$$\sigma = \sqrt{11.2次^2} = 3.34 \approx 3次$$

该组特征为：（6 ± 3）次。

为了求出间隔数列的变量指标，应该将间隔数列变为离差数列。为此，在每个组距（间隔）中求出其平均值（即组中点——译注）——把它看做该组变量的代表值，然后按上述公式进行计算。

还有一种类似的求标准差σ的方法，它的依据是6西格玛定理。举例如下：

$$\sigma = \frac{Xmax - Xmin}{6} \qquad (7)$$

σ——标准差；Xmax——数列中最大的变量；Xmin——最小的变量。

用公式（7）求出的标准差只是近似值。

在例2中：

$$\sigma = \frac{Xmax - Xmin}{6} = \frac{90秒 - 74秒}{6} = \frac{16秒}{6} = 2.66秒$$（准确值为3.6秒）。

在例3中：

$$\sigma = \frac{Xmax - Xmin}{6} = \frac{96 - 91}{6} = \frac{5}{6} = 0.9分$$（准确值为1.6分）。

在例4中：

$$\sigma = \frac{Xmax - Xmin}{6} = \frac{12 - 0}{6} = \frac{12}{6} = 2次$$（准确值为3次）。

很清楚，各例中按公式（7）所得的标准差都是近似值。

为了求算术平均数，常采用假設平均数的方法。在变量数字繁多，并运算困难的情况下，用这种方法特别简便。在用这种方法时先假设变量中的某一个数为平均数，然后求校正量，即真平均数与假设平均数之间的差异。这个校正量就是变量与假设平均数的离差之加权总和，除以总体量：

$$b = \frac{\sum_{i=1}^{i=n} (X_i - A) \cdot n_i}{n} \qquad (8)$$

b——计算假设平均数与真正平均数之间差异的校正量；

X_i——变量；

A——假设平均数；

n_i——组频数；

n——总体量。

算术平均数就是假设平均数与所得的校正量之和：

$$\bar{X} = A + b = A + \frac{\sum_{i=1}^{i=n} (X_i - A) \cdot n_i}{n} \qquad (9)$$

现用具体例子说明。

例5　16名滑雪运动员滑雪5公里的成绩（X_i）。求这些成绩的算术平均

数（表14）：

表14

X_i	n_i	$X_i - A$	$(X_i - A)\cdot n_i$
22′10″	2	−6′20″	−12′40″
24′00″	2	−4′30″	− 9′00″
27′20″	7	−1′10″	− 8′10″
28′30″	3	0	0
29′00″	1	+0′30″	+ 0′30″
30′40″	1	+2′10″	+ 2′10″
	n=16		−27′10″

将变量 A = 28′30″ 作为假设平均数。第三栏里是变量与假设平均数的差值，为$(X_i - A)$，第四栏里是它们的加权值，为$(X_i - A)\cdot n_i$。变量与假设平均数之间差的和为 $\sum_{i=1}^{i=n}(X_i - A)\cdot n_i = 27'10''$。校正量为：

$$b = \frac{\sum_{i=1}^{i=n}(X_i - A)\cdot n_i}{n} = \frac{-27'10''}{16} \approx -1'42''$$

现用公式9求算术平均数：

$$\bar{X} = A + b = 28'30'' - 1'42'' = 26'48''; \quad \bar{X} = 26'48''$$

现在用一般方法求算术平均数。为此，将原始资料变换为单位统一的数字，例如以秒为单位（表15）：

表15

X_i	X_i（秒）	n_i	$X_i\cdot n_i$
22′10″	1330	2	2660
24′00″	1440	2	2880
27′20″	1640	7	11480
28′30″	1710	3	5130
29′00″	1740	1	1740
30′40″	1840	1	1840
		n=16	25730

$$\bar{X} = \frac{25730}{16} = 1608 = 26'48''$$

不出所料，两种算法得出的结果是一样的。