统计推断理论基础及其应用

利　　人编

群众出版社

一九八二年·北京

统计推断理论基础及其应用

利　　人编

群众出版社出版　新华书店北京发行所发行
贵州新华印刷厂印刷
787×1092毫米　16开本　30印张　713千字　插页4
1982年5月第1版　1982年5月贵州第1次印刷

统一书号：13067·58　定价：4.70元
印数：0001—5600册

目　　录

第一章　基本概念

第二章　统计推断理论总说

第三章　参数估计

第四章　检验方差的统计假设

第一章　基本概念

1.1　随机现象，统计推断中的基本术语

人们在社会实践中到处都会遇到随机现象．例如，对某轴承的直径进行测量时，每次测量的数值可能一样，也可能不一样，也可能有某些值是一样的．这样，每次所得的测量数字，我们统称做测量（或观察）结果，而这每次观察结果就是随机的；又如，对某目标进行射击，有可能命中目标，也有可能不命中目标．因此每次射击的结果就是随机的．为什么每次测量的轴承直径值不一样呢？又为什么每次射击可能命中也可能不命中呢？按照哲学观点，无原因的现象是不存在的，而一切事物的运动都是和其周围其他事物互相联系着和互相影响着．再如，设计一门火炮，其射程为 100 公里．如果你要想用它打中1000公里的目标，这是不可想象的．然而，即便是要打 100 公里的目标，某次射击也可能命中目标，也有可能不命中目标．那么，某次射击不命中目标（假定按正常射击规则进行射击），必定是其周围的许多随机因素作用的结果，也就是说，周围的这许多因素决定各次给定现象（如射击）发展的相异性．由于这种相异性，使得观察结果发生某种差异，使得被研究的现象离开了规律性，或者说使得规律性有某种偏差．这种由于外部联系而引起的被研究对象离开规律性的偏差，我们就叫它随机现象．

如此说来，随机现象是客观存在的，与我们是否明了现象的原因完全无关．随机现象（或称随机性规律，偶然性规律）与客观的必然性规律不同的地方，就是后者是一种固定的、有规律的现象联系，前者却不是一种内在的、不可避免的从一定过程中有规律地引申出来的，而是可以这样、也可以那样形成起来的．因此，随机性是以外部联系和表面为其范围，而必然的规律性则是以内部的、决定性的、主要的联系和关系为其范围的，这些联系和关系决定着事物的发展和进程．

科学的认识与科学的研究，要求我们把必然的规律性同随机性严格区别开来，要求我们从似乎是一团乱麻的现象中找出它的必然的规律性．我们研究随机现象的目的，在于学会预见随机因素的作用，揭示以随机性出现的必然的规律性，并在实践活动中考虑随机性因素的影响和应用所发掘出来的必然的规律性．

现在我们来介绍一些在统计推断中经常遇到的术语和基本概念．

实现在一定条件下给定的随机现象的观察，我们叫做实验．实验的结果，我们叫做事件，可以以定量表示，也可以以定性表示．例如射击，命中目标是一个事件，不命中目标也是一个事件，测量轴承的直径所得之量小于某数 a 的数值也是一个事件．

在实验中可以取可能数值序列中的某一个值的实验结果的任何定量表示，叫做随机变量．

在给定实验结果中一定会出现的事件叫做必然事件，一定不会出现的事件叫做不可能事

件，而可能出现也可能不出现的事件叫做随机事件．

我们所研究的对象的全体称做总体，总体中的每个单元称做个体或元素，总体中元素或个体的特征表示称做标识．

为了研究总体，通常是从总体中抽取部分的个体加以研究．我们把抽取出来的这部分的全体叫做子样（或抽样，或样本）．

1.2　数据表和分布图

在从事数据处理的实际工作中，我们首先要遇到观察得来的原始数据．例如，某工厂生产一批铆钉．在这批成品中，我们抽出 200 个样本，检查每个铆钉头的直径．所得数据如表1.1．

表1.1　　200个铆钉头的直径（以毫米为单位）

13.39	13.43	13.54	13.64	13.40	13.55	13.40	13.26
13.42	13.50	13.32	13.31	13.28	13.52	13.46	13.63
13.38	13.44	13.52	13.53	13.37	13.33	13.24	13.13
13.53	13.53	13.39	13.57	13.51	13.34	13.39	13.47
13.51	13.48	13.62	13.58	13.57	13.33	13.51	13.40
13.30	13.48	13.40	13.57	13.51	13.40	13.52	13.56
13.40	13.34	13.23	13.37	13.48	13.48	13.62	13.35
13.40	13.36	13.45	13.48	13.29	13.58	13.44	13.56
13.28	13.59	13.47	13.46	13.62	13.54	13.20	13.38
13.43	13.35	13.56	13.51	13.47	13.40	13.29	13.20
13.46	13.44	13.42	13.29	13.41	13.39	13.50	13.48
13.53	13.34	13.45	13.42	13.29	13.38	13.45	13.50
13.55	13.33	13.32	13.69	13.46	13.32	13.32	13.48
13.29	13.25	13.44	13.60	13.43	13.51	13.43	13.38
13.24	13.28	13.58	13.31	13.31	13.45	13.43	13.44
13.34	13.49	13.50	13.38	13.48	13.43	13.37	13.29
13.54	13.33	13.36	13.46	13.23	13.44	13.38	13.27
13.66	13.26	13.40	13.52	13.59	13.48	13.46	13.40
13.43	13.26	13.50	13.38	13.43	13.34	13.41	13.24
13.42	13.55	13.37	13.41	13.38	13.14	13.42	13.52
13.38	13.54	13.30	13.18	13.32	13.46	13.39	13.35
13.34	13.37	13.50	13.61	13.42	13.32	13.35	13.40
13.57	13.31	13.40	13.36	13.28	13.58	13.58	13.38
13.26	13.37	13.28	13.39	13.32	13.20	13.43	13.34
13.33	13.33	13.31	13.45	13.39	13.45	13.41	13.45

此表内的数据取自[27]

从这样的数据表很不容易看出 200 个铆钉头直径的变化情况．为了能从数据表上看到铆钉头直径的分布情况，我们可以列出如下的数据表（表 1.2）．

表1.2

毫米	铆钉数		毫米	铆钉数		毫米	铆钉数	
13.10			13.30	丅	2	13.50	正一	6
13.11			13.31	正	5	13.51	正一	6
13.12			13.32	正丅	7	13.52	正	5
13.13	一	1	13.33	正一	6	13.53	止	4
13.14	一	1	13.34	正丅	7	13.54	止	4
13.15			13.35	止	4	13.55	下	3
13.16			13.36	下	3	13.56	下	3
13.17			13.37	正一	6	13.57	止	4
13.18	一	1	13.38	正正	10	13.58	正	5
13.19			13.39	正丅	7	13.59	丅	2
13.20	下	3	13.40	正正丅	12	13.60	一	1
13.21			13.41	止	4	13.61	一	1
13.22			13.42	正一	6	13.62	下	3
13.23	丅	2	13.43	正止	9	13.63	一	1
13.24	下	3	13.44	正一	6	13.64	一	1
13.25	一	1	13.45	正丅	7	13.65		
13.26	止	4	13.46	正丅	7	13.66	一	1
13.27	一	1	13.47	下	3	13.67		
13.28	正	5	13.48	正止	9	13.68		
13.29	正一	6	13.49	一	1	13.69	一	1
							总计：200	

从表 1.2 可以一目了然地知道，直径等于13.13 毫 米 的有 1 个，13.24 毫米的有 3 个，13.40 毫 米的有12个，等等．按此数据表，将直径作为横坐标，铆钉数作为纵坐标，在此坐标系内以点号代表铆钉数描绘出的图形，我们称之为次数分布图．

数据表1.2或次数分布图虽然比数据表1.1较明显，但是还不及累积次数分布图更能说明问题．现列出表1.3以求累积次数．

根据表 1.3 的第一栏和第三栏便能够作出图1.1.

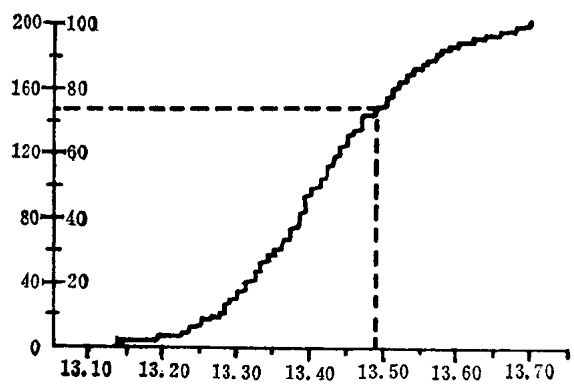

图1.1　累积次数分布图

图 1.1 称做累积次数分布图．其横坐标代表铆钉头直径，以毫米为单位，而纵坐标轴左边的刻度代表铆钉数，右边刻度代表累积次数的百分数，即右边刻度 100 相当于左边刻度 200，通常把这种刻度值称做累积百分数．依据累积次数分布图不难看出：当选定一个横坐标值 x 时，它所对应的纵坐标值 y，若依左边刻度计算，就表示小于或等于 x 毫米的铆钉数；若依右边刻度计算，就表

表1.3

直径（毫米）	次数	累积次数	直径（毫米）	次数	累积次数	直径（毫米）	次数	累积次数
13.13	1	1	13.35	4	59	13.51	6	161
13.14	1	2	13.36	3	62	13.52	5	166
13.18	1	3	13.37	6	68	13.53	4	170
13.20	3	6	13.38	10	78	13.54	4	174
13.23	2	8	13.39	7	85	13.55	3	177
13.24	3	11	13.40	12	97	13.56	3	180
13.25	1	12	13.41	4	101	13.57	4	184
13.26	4	16	13.42	6	107	13.58	5	189
13.27	1	17	13.43	9	116	13.59	2	191
13.28	5	22	13.44	6	122	13.60	1	192
13.29	6	28	13.45	7	129	13.61	1	193
13.30	2	30	13.46	7	136	13.62	3	196
13.31	5	35	13.47	3	139	13.63	1	197
13.32	7	42	13.48	9	148	13.64	1	198
13.33	6	48	13.49	1	149	13.66	1	199
13.34	7	55	13.50	6	155	13.69	1	200

示小于或等于x毫米的铆钉数占总数（200个）的百分数．例如，图上对应于横坐标$x=13.49$的纵坐标$y=148$（累积次数）或74%（累积百分数），这就是说有148个（或74%）的铆钉头直径小于或等于13.49毫米．

这里值得提及的是：如果在所取的x处曲线有跳跃，则对应于此x的y应为曲线上的上端的坐标；反之，当我们先取纵坐标y，并且在所取y处曲线有跳跃时，则对应于此y之x应取为曲线上的左端的横坐标．

据以上所述，虽然数据表1.2（或次数分布图）和累积次数分布图都比原始数据表1.1显明、可用，但是，我们也看到利用数据表1.2(或次数分布图)和累积次数分布图都很费时间，并且在实际应用中并不必如此精确．因此，一般地说来，当样本数目超过25时，总是将观察得来的数据（样本）先行分组．这既可以节省时间，又能保证必要的精确性．其方法如下：

就表1.2所提供的数据而论，我们可以选取适当的组距将这些数据加以等分．例如，我们取组距为0.05来对数据作12等分：

13.095－13.145，13.145－13.195，13.195－13.245，13.245－13.295，

13.295－13.345，13.345－13.395，13.395－13.445，13.445－13.495，

13.495－13.545，13.545－13.595，13.595－13.645，13.645－13.695．

这些组的特点是：

（i）组距各为0.05；

（ii）每组的两端，其最后一位数字是5，并比原始数据多一位小数；

（iii）前一组后面的值等于后一组前面的值．

这些组各自的中点值（或组中值）为：

13.12，13.17，13.22，13.27，13.32，13.37，13.42，13.47，

13.52，13.57，13.62，13.67.

按上述分组方法所得的结果，我们也可以作出关于200个铆钉头直径的分布表如下（表1.4）.

表1.4

组端值	组中值（毫米）	铆钉数	累积次数	次数频率	累积频率
13.095－13.145	13.12	2	2	0.010	0.010
13.145－13.195	13.17	1	3	0.005	0.015
13.195－13.245	13.22	8	11	0.040	0.055
13.245－13.295	13.27	17	28	0.085	0.140
13.295－13.345	13.32	27	55	0.135	0.275
13.345－13.395	13.37	30	85	0.150	0.425
13.395－13.445	13.42	37	122	0.185	0.610
13.445－13.495	13.47	27	149	0.135	0.745
13.495－13.545	13.52	25	174	0.125	0.875
13.545－13.595	13.57	17	191	0.085	0.955
13.595－13.645	13.62	7	198	0.035	0.990
13.645－13.695	13.67	2	200	0.010	1.000

根据表1.4的数据便可作矩形柱的图形如下(图1.2).

矩形柱图1.2是利用组中值与对应的铆钉数画出的，矩形柱底的中点是组中值，底宽是组距，即0.05，因此矩形柱底的两端值就是组端值，而矩形柱高是铆钉数.

另外，又可以利用表1.4中第一栏中的组右端值和第四栏的累积次数作出累积多角形(图1.3).

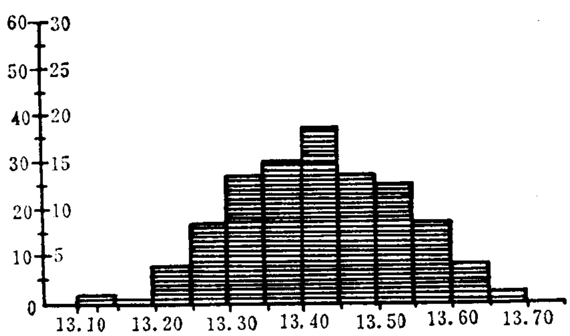

图1.2 矩形柱图

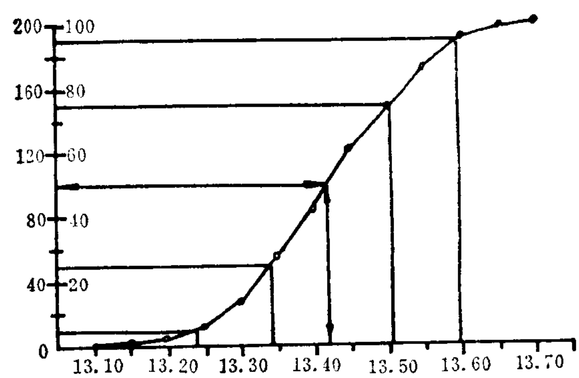

图1.3 累积多角形

为了说明累积多角形的作用，我们要用到如下一些概念.

如果通过百分数轴（即纵坐标右边的刻度）上任何值p画一条平行于横坐标轴的横线而与累积多角形相交，再由此交点引平行于纵坐标轴的竖线而与横坐标轴相交于一点，记作x_p，则x_p称做“百分之p值”.

百分之50值，即 $x_{50\%}$，称做“中值”(或称做中位数)；百分之25值，即 $x_{25\%}$，称做“下中值”(或称下四分位数)；百分之75值，即$x_{75\%}$，称做“上中值”(或称上四分位数)．

现在我们来说明累积多角形的作用．例如从图1.3上可求得百分之50值,即$x_{50\%}=13.42$，也就是说，大约有百分之50的铆钉头直径是小于或等于 13.42 毫米．而百分之 5 值和百分之95值，即$x_{5\%}=13.24$和$x_{95\%}=13.59$，就是说，有百分之 5 的观察结果（铆钉头直径值）是小于或等于13.24和有百分之95的观察结果是小于或等于13.59（或者说有百分之 5 的观察结果是大于 13.59 的)．显然,区间的对分是由百分之50值，即由 $x_{50\%}=13.42$ 来实现的．因此知道，落在区间 $x_{5\%}<x\leqslant x_{50\%}$ 内的观察结果有百分之45，而落在区间 $x_{50\%}<x\leqslant x_{95\%}$ 内的观察结果也有百分之45，所以落在区间 $x_{5\%}<x<x_{95\%}$ 内的结果有百分之90．又从图1.3上不难求得百分之25值，即“下中值” $x_{25\%}=13.34$毫米，而百分之75值，即“上中值” $x_{75\%}=13.51$ 毫米．

这样我们就看到累积多角形的作用了．事实上累积多角形就是用子样分布（或经验分布）代替总体分布的曲线．这些问题将在节1.5中讨论．

1.3 均值，方差，协方差

1.3.1 均值的定义和计算

在统计学中，所谓观察结果的均值，其定义是：所有观察值的和除以观察值的个数．

如果我们令观察结果为 x_1，x_2，$\cdots$，x_n，并以 $\bar{x}$ 表示均值，则

$$\bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i. \tag{1.3.1}$$

如果在 x_1，x_2，$\cdots$，x_n 中有许多相同的,设其中有 k 个不同的数值 $y_1, y_2, \cdots, y_k$，而每个 y_i 出现的次数记作 f_i($i=1, 2, \cdots, k$)，则均值 $\bar{x}$ 可表示为

$$\bar{x}=\frac{f_1y_1+f_2y_2+\cdots+f_ky_k}{f_1+f_2+\cdots+f_k}=\frac{\sum\limits_{i=1}^{k}f_iy_i}{\sum\limits_{i=1}^{k}f_i}. \tag{1.3.2}$$

显然,(1.3.2)是(1.3.1)的一个特例,因为在(1.3.2)中，若令$f_i\equiv 1$ ($i=1, 2, \cdots, k=n$)，则 (1.3.2) 就是 (1.3.1)．通常我们把表达式 (1.3.2) 称作加权平均数．从 (1.3.2) 这个表达式不难看到，如果诸 y 值是某随机变量的观察值的话，那么 $\bar{x}$ 就是这个随机变量的数学期望值．因此，我们说，在统计学中,“均值”、“平均数”、“数学期望”、“期望值”这些术语都是同义词．

例1.1　今对10个铆钉头的直径进行测量,我们得到10个观察结果分别为13.40，13.41，13.49，13.42，13.48，13.29，13.33，13.46，13.50，13.35毫米，则它们的均值为

$$\bar{x}=\frac{13.40+13.41+\cdots+13.50+13.35}{10}\approx 13.413\ (\text{毫米}). \tag{*}$$

例1.2　如果我们要求得到表 1.1 中的 200 个铆钉头直径测量值的均值，此时可以利用

表1.3的数据代入加权计算公式（1.3.2）就得

$$\bar{x}=\frac{1\times 13.12+1\times 13.14+\cdots+12\times 13.40+\cdots+1\times 13.69}{200}=13.4165. \qquad (**)$$

由此不难看到，当所讨论的样本个数很多时，上面计算均值的方法相当费时间。在节1.2中作图的时候，我们曾提出一种分组方法. 在求算均值的过程中，特别是当样本较多（一般为数在25以上）时，利用分组法也是特别简单的. 设x_i代表组中值，f_i代表落在$\left(x_i-\frac{1}{2}h,\ x_i+\frac{1}{2}h\right)$当中的样本（或观察值）数，其中$h$代表组距. 现在列出表1.5以示求均值的方法.

表1.5

组中值(x_i)	次数	f_ix_i
x_1	f_1	f_1x_1
x_2	f_2	f_2x_2
$\vdots$	$\vdots$	$\vdots$
x_k	f_k	f_kx_k
总计	n	$\sum f_ix_i$
$\bar{x}$	$\frac{1}{n}\sum_{i=1}^{k} f_ix_i$	

例1.3　今取节1.2中对于200个铆钉头直径的观察结果为例，我们来计算均值. 按表1.5列表1.6计算如下：

将这个结果同（**）结果比较，可见经过归组算得的$\bar{x}$精确到第三位小数.

下面我们讲述用移动坐标法以精简计算的方法. 令a和b是两个任意选定的常数，定义一个新的变量Z_i，它与样本值x_i存在如下的关系

$$x_i=a+bZ_i,\quad i=1,2,3,\cdots,n. \qquad (1.3.3)$$

表1.6

组中值 x_i	次数 f_i	f_ix_i
13.12	2	26.24
13.17	1	13.17
13.22	8	105.76
13.27	17	225.59
13.32	27	359.64
13.37	30	401.10
13.42	37	496.54
13.47	27	363.69
13.52	25	338.00
13.57	17	230.69
13.62	7	95.34
13.67	2	27.34
总计	200	2683.10
$\bar{x}$		13.4155

表1.7

观察值 x_i	$x_i-13.40$ Z_i
13.40	.00
13.41	.01
13.49	.09
13.42	.02
13.48	.08
13.29	−0.11
13.33	−.07
13.46	.06
13.50	.10
13.35	−.05
总和	0.13
$\bar{Z}$	0.013

因为 $$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{n}\sum_{i=1}^{n}(a+bZ_i)=a+b\frac{1}{n}\sum_{i=1}^{n}Z_i,$$

所以有 $$\bar{x}=a+b\bar{Z}. \tag{1.3.4}$$

我们首先讨论 $b=1$ 这一情况．此时有

$$x_i=a+Z_i,\quad \bar{x}=a+\bar{Z}. \tag{1.3.5}$$

由此可知，若在每一个变量 x_i 上加一常数（或减一常数）a，则所得新变量 Z_i 的均值等于 x_i 的均值加（减）同一常数 a．此法对于分组后求均值也是正确的．今举一例说明．

例1.4　今利用公式(1.3.5)来求例1.1中给定的10个铆钉头直径的均值．

现在取常数 $a=13.40$ 毫米，应用关系式 $x_i=a+Z_i$ 可得表1.7的值．

由上表并根据 $\bar{x}=a+\bar{Z}$ 则得

$$\bar{x}=13.40+0.013=13.413.$$

此结果与(*)是一样的，但是计算工作量大大减小了．

现在讨论当 $b\neq 1$ 时的情况．我们知道，当样本数很大（一般当超过25）时，应用分组方法计算均值(见例 1.3)是方便的．但是，此时我们应用关系(1.3.3)和(1.3.4)，其中 $b\neq 1$，能够更为简便．现在举一个例子说明．

例1.5　计算表1.1中给定的 200 个观察值的均值．

为了进行计算，首先应对原始数据加以分组，因而要适当地选取组距 h．其次是选取适宜的常数 a 和 b，这两个常数的选取要视原始数据的变动情况和组距的大小而定．比如说，关于 a 的选取问题，一般说来是选取观察值 x_i 中的某一个值，而 b 一般选取值是小于或等于组距．现在取 $a=13.40$，$b=h=0.05$．据此可以根据表1.4的第二、三栏列表1.8并计算．

以上我们讨论了关于均值计算的各种简便方法．如果能够灵活应用这些结果，就可以使计算精简至我们所希望的程度．

表1.8

组中值 x_i	次数 f_i	$x_i=13.40+0.05Z_i$ Z_i	f_iZ_i
13.12	2	-5.6	-11.2
13.17	1	-4.6	-4.6
13.22	8	-3.6	-28.8
13.27	17	-2.6	-44.2
13.32	27	-1.6	-43.2
13.37	30	-0.6	-18.0
13.42	37	0.4	14.8
13.47	27	1.4	37.8
13.52	25	2.4	65.0
13.57	17	3.4	57.8
13.62	7	4.4	30.8
13.67	2	5.4	10.8
总计	200		67.0
$\bar{Z}$			0.335
$\bar{x}=13.40+0.05\times0.335=13.4168$			

1.3.2　方差的定义和计算

在统计学中，经常遇到的另一个重要参数就是子样方差．设某一随机变量 X 的观察值（子样）为 $x_1, x_2, \cdots, x_n$，则子样方差定义作

$$S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}. \tag{1.3.6}$$

其中 $\bar{x}$ 是子样均值

$$\bar{x}=\frac{\sum_{i=1}^{n} x_i}{n}.$$

计算 S^2 的时候，在一般的情况下，可以不使用公式（1.3.6）的形式，而使用比较容易计算的形式. 事实上，由于

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})^2&=\sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)\\&=\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2\\&=\sum_{i=1}^{n}x_i^2-2\bar{x}n\bar{x}+n\bar{x}^2\\&=\sum_{i=1}^{n}x_i^2-n\bar{x}^2,\end{aligned}$$

又因为

$$n\bar{x}^2=\frac{(n\bar{x})^2}{n}=\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n},$$

所以

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2. \tag{1.3.7}$$

代入（1.3.6）便得

$$S^2=\frac{1}{n-1}\left\{\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2\right\}. \tag{1.3.8}$$

为了说明问题和计算上的方便，我们引入记号. 例如对于特定的 x 记 $S_x^2=S^2$，$S(x^2)=\sum_{i=1}^{n}x_i^2$，和$S(x)=\sum_{i=1}^{n}x_i$. 这样，公式（1.3.8）可以改写作

$$S_x^2=\frac{1}{n-1}\left\{S(x^2)-\frac{1}{n}(S(x))^2\right\}. \tag{1.3.9}$$

另一方面，当所研究的总体的元素个数很多时，在这种场合（一般为数在25以上），求其方差是很费时间的，因此需要采用分组的方法. 如果以 x_i 代表组中值，f_i 代表在$\left(x_i-\frac{1}{2}h,\ x_i+\frac{1}{2}h\right)$当中的子样个数，则

$$S_x^2=\frac{1}{n-1}\left\{\sum_{i=1}^{k}f_ix_i^2-\frac{1}{n}\left(\sum_{i=1}^{k}f_ix_i\right)^2\right\}. \tag{1.3.10}$$

具体计算时，可按表1.9进行.

表1.9

组中值 x_i	次数 f_i	$f_i x_i$	$f_i x_i^2$
x_1 x_2 $\vdots$ x_k	f_1 f_2 $\vdots$ f_k	$f_1 x_1$ $f_2 x_2$ $\vdots$ $f_k x_k$	$f_1 x_1^2$ $f_2 x_2^2$ $\vdots$ $f_k x_k$
总　　计	n	$S(x)$	$S(x^2)$
S_x^2	$\frac{1}{n-1}\left\{S(x^2)-\frac{1}{n}(S(x))^2\right\}$		

下面举例说明这个计算过程.

例1.6　设某工厂为了防止铁片生锈而在其表面上镀锌. 今从一批成品中抽取75片样品, 而检查每片上附着锌的重量. 所得数据如表1.10.

表1.10　**75片样品上附着锌的重量（克）**

1.47	1.60	1.53	1.56	1.44	1.62	1.60	1.58	1.39	1.35	1.52
1.38	1.32	1.65	1.53	1.77	1.73	1.62	1.62	1.38	1.55	1.70
1.47	1.53	1.46	1.53	1.60	1.42	1.47	1.44	1.38	1.60	1.45
1.34	1.47	1.37	1.48	1.34	1.58	1.43	1.64	1.51	1.44	1.49
1.64	1.46	1.53	1.56	1.56	1.50	1.63	1.59	1.48	1.54	1.61
1.54	1.50	1.43	1.57	1.42	1.53	1.60	1.55	1.67	1.57	1.34
1.54	1.64	1.47	1.75	1.60	1.57	1.57	1.63	1.47		

现在来求由表1.10给出的75个数据的方差. 为此首先进行分组, 例如取组距为0.05, 则可得组中值和次数的结果（表1.11中的头两栏）. 于是也可计算出 $f_i x_i$ 和 $f_i x_i^2$ 值（表1.11的后两栏）. 最后, 可按（1.3.10）求出S_x^2值, 列在表1.11的下部.

表1.11

组中值 x_i	次数 f_i	$f_i x_i$	$f_i x_i^2$
1.30	1	1.30	1.6900
1.35	5	6.75	9.1125
1.40	6	8.40	11.7600
1.45	13	18.85	27.3325
1.50	8	12.00	18.0000
1.55	17	26.35	40.8425
1.60	14	22.40	35.8400
1.65	7	11.55	19.0575
1.70	1	1.70	2.8900
1.75	3	5.25	9.1875
总　　计	$n=75$	$S(x)=114.55$	$S(x^2)=175.7125$
S_x^2	$\frac{1}{75-1}\left\{175.7125-\frac{1}{75}(114.55)^2\right\}=0.01020$		

在前一节内，曾经利用移动坐标法以精简关于均值的计算。移动坐标法同样适用于方差的精简计算。事实上，引入一个新变量 Z_i，设 a 是个常数，则 x_i 与 Z_i 的关系可记作

$$x_i = a + Z_i. \tag{1.3.11}$$

由于

$$x_i - \bar{x} = Z_i - \bar{Z},$$

也就是说，均值的偏差与坐标原点无关。这样对上式两边同时求平方和可得

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(Z_i - \bar{Z})^2.$$

因此知道

$$S_x^2 = S_Z^2. \tag{1.3.12}$$

如果令

$$x_i = a + bZ_i, \tag{1.3.13}$$

其中 a 和 b 是两个常数，则可证得

$$S_x^2 = b^2 S_Z^2. \tag{1.3.14}$$

计算时可按表1.12进行之。

表1.12

组中值 x_i	次数 f_i	$x_i = a + bZ_i$		
		Z_i	f_iZ_i	$f_iZ_i^2$
x_1 x_2 $\vdots$ x_k	f_1 f_2 $\vdots$ f_k	Z_1 Z_2 $\vdots$ Z_k	f_1Z_1 f_2Z_2 $\vdots$ f_kZ_k	$f_1Z_1^2$ $f_2Z_2^2$ $\vdots$ $f_kZ_k^2$
总计	n		$\sum_{i=1}^{k} f_iZ_i$	$\sum_{i=1}^{k} f_iZ_i^2$
$(S(Z))^2$	$\left(\sum_{i=1}^{k} f_iZ_i\right)^2$			
S_Z^2	$\frac{1}{n-1}\left\{S(Z^2) - \frac{1}{n}[S(Z)]^2\right\}$			
S_x^2	$b^2S_Z^2$			

例1.7　仍以例1.6中的镀锌片数据为例。

取 $a = 1.50$，$b = 0.05$，则得表1.13。

比较表1.11和表1.13的 S_x^2 结果，不难看出它们是一致的。

最后指出，在讨论均值和方差时，应当知道，方差永远是名数，即它的单位是均值单位的平方，换句话说，它只能测定单位相同的数列离中程度或说偏离均值的程度。