常用农业试验的统计分析方法

阮 寿 康 编

河北省植保土肥研究所印

一九七七年十月

毛主席语录

科学技术这　　　　　打，而且必须打好。

要解决问题，还须作系统的周密的调查工作和研究工作，这就是分析的过程。提出问题也要用分析，不然对着模糊杂乱的一大堆事物的现象，你就不能知道问题即矛盾的所在。

对情况和问题一定要注意到它们的数量方面，要有基本的数量分析。任何质量都表现为一定的数量，没有数量也就没有质量。

凭客观存在的事实，详细地占有材料，在马克思列宁主义一般原理的指导下，从这些材料中引出正确的结论。

用不同的方法去解决不同的矛盾。

在实践中不断开辟认识真理的道路。

编 者 按

伟大领袖和导师毛主席教导我们，对于一切事物要做到"**胸中有'数'。这就是说，对情况和问题一定要注意到它们的数量方面，要有基本的数量分析。任何质量都表现为一定的数量，没有数量也就没有质量。**"如果我们"**不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无'数'，结果就不能不犯错误。**"。因此我们必须做"基本的调查"，"基本的分析"，才能把我们从事的革命工作搞好。

农业工作和其他革命工作一样，为了摸清情况，决定相应的措施促使农业大上，就必须作大量的调查研究和科学实验。对于所得到的大量数据则应在辩证唯物论的指导下进行深入细致的"**去粗取精、去伪存真、由此及彼、由表及里**"的分析研究，才能使大量零乱而无意义的数据成为有用的资料，进一步指导我们的工作。

生物统计法这个工具可以帮助我们解决这个问题。因此我们编写了《常用农业试验的统计分析方法》这本小册子，以适应群众性科研工作发展的需要，希望能在"**农业学大寨**"的伟大群众运动中，为促进农业大上起到一点有益的作用。

由于我们在数学基础理论和实践经验方面都很缺乏，因此缺点和错误是难免的，衷心希望得到广大同志的批评和指正。

阮 寿 康

目　录

一、引　言

二、几个主要统计值的意义和计算方法

三、二数相关关系的度量

四、净相关、复相关和多元回归

五、卡方测定

六、常用农业试验的统计分析

七、附　　录

八、附　　表

常用农业试验的统计分析方法

一、引　言

1．生物统计研究的主要内容：

任何事物都有它的数量和质量两个方面，都是量和质的对立统一体。事物的发展都是从量变开始的，量变达到一定程度就会引起质变，产生新现象。因此我们对任一现象的研究都要分析“量”和“质”两方面。统计学就是在质和量的密切联系中研究现象的数量方面的表现的科学。

在农业生产中，我们研究的对象是生物群体。这些群体都是由众多的个体所组成的。每一个个体不仅具有群体的特征（共性），同时个体之间又存在着差异（个性）。例如小麦品种甲是高秆品种，品种乙是矮秆品种，它们就是两个不同的群体。甲群体中的每一个个体都具有植株较高的特征而乙群体中的每一个个体都具有植株较矮的特征，所以甲乙两个群体的质是不同的。但是这两个群体的任一群体中的每一个个体的植株高度也不尽相同，有的较高，有的较矮。这是生物群体中普遍存在的现象。现在，产生了一系列的问题。如果我们要说明甲、乙两个小麦品种的株高（或穗长、穗粒重、产量等）这个数量特征，我们就必须进行实际的度量。但是在我们对同一群体的不同个体度量后，就发现这些数据之间有差异，甚至少数个体与大多数个体之间的差异还较大。那么究竟怎样说明这个群体的株高特征呢？怎样从调查所得的一大堆数据中，找出简单而具明确意义的能代表群体特征的概念来呢？又如何比较不同群体的差异呢？另外，在我们度量任一群体时，由于群体很大，对所有个体都进行度量是不可能的。所以只能抽取一定数量的样本进行度量。那么样本的代表性又如何呢？样本表现的特征是否能够代表总体的特征呢？辩证唯物论还告诉我们：一切事物都不是孤立存在的，而是互相联系互相制约的。但是一切事物和他事物之间的关系有的密切，有的就不够密切。例如植物病害的发展速率受许多因素的影响，而这些因素的影响作用及其程度的大小就不相同。那么怎样指示它们之间的关系，并从而选取主要因素来进行病害流行速率的预测呢？诸如此类的问题很多，都是统计分析研究的内容。但概括来看统计在农业科学研究中的作用主要是帮助我们：（1）整理分析调查所得的大量数据，反映现象所具有的特征。如产量水平的高低、变异的幅度（稳定性）等。（2）判断调查或试验结果的可靠性。（3）研究分析变量间的相互关系。（4）在进行试验设计和调查时减少误差和估计误差。

2．运用统计分析时应注意的问题：

在运用统计分析时应注意四个问题：

一是要从大量的现象出发。这是因为生物现象是非常复杂的，每一种现象的具体数量表现是许多因素影响的结果，因此具有一定的偶然性。如果仅从个别现象推断真实情况，

容易导致错误的结论。

二是这些现象应是本质上相同的。这是因为把不同质的现象混在一起研究，所得的结果是没有意义的，是不能说明什么问题的。但是现象的同质性是相对的概念，应依我们调查研究的目的来定。例如我们要调查某一地区的小麦生产水平，那么不论品种等条件是否相同，所有小麦田都可作为调查统计对象。这时所有麦田都是性质相同的。但是如果我们要调查的是不同小麦品种的产量水平，那么同一品种是性质相同的，而不同品种的性质不同，因此要分品种进行调查统计。如果把不同品种的地块混在一起作产量统计，所得的结果就不能说明品种间产量水平的差异。这是显而易见的。

三是必须注意资料的可靠性。就是说应当用严格的试验处理和适宜而尽可能精确的调查方法来取得第一手的资料。并对取得的大量资料进行研究，“去粗取精，去伪存真”，然后进行统计分析。

四是应在生物学理论的指导下应用统计分析。例如一种害虫的发育需要一定的温度，过低则产生滞育。因此在研究温度对该害虫发育进程的影响时，应当考虑发育起点温度的问题。否则也易导致错误结论。另外在统计分析了现象的数量表现后，应在生物学理论的指导下，进一步研究质的特征。因为每种数量都是一定质的量。进一步研究质的特征才能深入理解生物现象的本质，使数量统计起积极的作用。例如分析了小麦品种的产量表现后，可对品种的产量水平和稳定性有所了解。如果进一步去研究这一品种高产稳产数量特征的内因，就可能揭发现象的本质，进一步指导我们寻求其他措施促使量变走向质变而获得更高的产量或更好的品质。而这些也必须是在生物学理论研究的指导下才能达到的。

3．几个常用数学名词简介：

（1）方程式、常数和变数：

方程式可理解为条件等式。所谓解方程式就是寻求适应这一条件的答案，数学上通称为“解”。当然，方程式的幂次不同，方程式可以有一个、两个或多个解。但是除了方程式的解以外，其他值都不适应这一条件。例如$x-2=0$，$6x^2-11x+4=0$都是方程式。前一例中只有$x=2$时，而后一例中只有当$x=\frac{1}{2}$和$\frac{4}{3}$时，才能使等号两边相等。例中x的值因条件不同而不同，被称为变数。而（-2）、6、（-11）、4则不因条件不同而改变其值，被称为常数。

（2）函数、自变数和依变数

如　$y=2x+3$　就是一个函数式。它的含意是：如果给x一个特定值，就相应得到一个y值。亦即y的值受x值的影响而变。这时我们就说y是x的函数，在微积分学中写作　$y=f(x)$，其中x是自变数而y是依变数。

（3）恒等式：恒等式与方程式不同，并不需要求解。恒等式是等量的不同表示。式中的变数无论取何值，等式两边恒等。如　$(a+b)^2=a^2+2ab+b^2$。

二、几个主要统计值的意义和计算方法

（一）样本集中性的度量：——算术平均数（简称平均数）

1．平均数的意义：

平均数是研究数量变异的最重要的和最常用的统计值之一。它反映的是总体特征的典型水平。用它可对同类群体（如小麦产量、某种病虫害造成的损失等）进行空间（不同地区）的或时间（不同年度、不同生育期等）上的比较。例如甲队的小麦产量水平是平均亩产700斤而乙队的是800斤，可知乙队的小麦产量水平高于甲队。

那么用总产（即总数）来比较而不用平均数比较不是更简单吗？是的，确实是比较简单，但是只有在统计个体数（样本）相等的条件下才能用总数进行比较。如果个体数不同，用总数比较会导致错误结论。例如上例中甲乙两队均种麦100亩，则甲乙两队的小麦总产依次为70000斤和80000斤，结论与前相同。又如甲队种麦100亩而乙队种麦只有50亩，则甲队小麦总产仍为70000斤而乙队小麦总产则为50×800＝40000斤，如果仅用总产比较单产的水平，岂不是得到了与事实相反的结论了吗？可见应用平均数的重要性。

2．总数与平均数的关系：

如果N个个体的观察值是$x_1x_2x_3\cdots\cdots x_n$，则总数应为$(x_1+x_2+x_3+\cdots\cdots+x_n)$；而平均数应为$\frac{x_1+x_2+x_3+\cdots\cdots+x_n}{N}$。在统计上常用$\Sigma x$代表x的总和（Σ是希腊字母，读作西格马，代表总和的意思），用Σf代表观察个体（样本）的总数（$\Sigma f=N$），而用$\bar{x}$代表x的平均值，因此总数与平均数的关系可用下式表示：$\bar{x}=\frac{\Sigma x}{\Sigma f}$。有时又用T表示总数，所以总数与平均数的关系又可写为：$\bar{x}=\frac{T}{N}$或$T=N\bar{x}$。

例如求2、4、9三个观察值的平均数。则

$$\bar{x}=\frac{\Sigma x}{\Sigma f}=\frac{2+4+9}{3}=5$$

3．在观察值多时平均数的计算方法（加权平均法）：

如用x代表观察值。用f代表观察值出现的次数（常称频度）。则可将所有观察值列为下面的频度表。

X	f	fx
x_1	f_1	f_1x_1
x_2	f_1	f_2x_2
⋮	⋮	
x_n	f_n	f_nx_n
	$\Sigma f=N$	$\Sigma fx=T$

式中x_1代表第一种观察值，f_1代表第一种观察值出现的次数，f_1x_1则表示第一种观察值的小总数，f行各值相加总和为$\Sigma f = N$，即观察的总次数，fx行各值相加总和为$\Sigma fx = T$，即总数，因此：

$$\bar{x} = \frac{T}{N} = \frac{\Sigma fx}{\Sigma f}$$

例一：在某生产队的棉田中于定苗后检查因枯萎病的死苗率。每样点检查200株苗，共检查200个样点，其中死苗率为6%、8%、10%、12%、14%的样点数依次为40、65、55、35、和5个。求平均死苗率。则：

x（死苗率%）	f（样点数）	fx（小总数）
6	40	240
8	65	520
10	55	550
12	35	420
14	5	70
	200	1800

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{1800}{200} = 9$$ 即死亡率为9%

注意：（1）样本大小是一样的。如不一样不能这样计算，此时则应用$\frac{\text{总死苗数}}{\text{调查总苗数}} \times 100$公式计算。

（2）不能用出现的各类死苗率的数字，直接相加平均计算。因为各类死苗率的出现次数（即权重）不同。

例二：两种病情指数计算公式的来源：

1. 严重度用百分率分级标准时平均病情指数的计算公式：

$$\bar{x}\text{（病情指数\%）} = \frac{f_0x_0 + f_1x_1 + f_2x_2 + \cdots\cdots + f_nx_n\text{（级别与相应各级出现的个体数的乘积之和）}}{N\text{（调查总各体数）}}$$

这个公式是直接由加权平均法得来的，推导如下：

严重度（x、%）	个体数（f）	fx
x_0	f_0	f_0x_0
x_1	f_1	f_1x_1
x_2	f_2	f_2x_2
⋮	⋮	⋮
x_n	f_n	f_nx_n
	Σf	Σfx

$$\bar{x}(\%)=\frac{\Sigma fx}{\Sigma f}=\frac{f_0x_0+f_1x_1+f_2x_2+\cdots\cdots+f_nx_n}{N}$$

2、严重度用级别表示，而平均病情指数仍以百分率表示时的计算公式是：

$$\bar{x}(\text{病情指数}\%)=\frac{0\cdot f_0+1\cdot f_1+2\cdot f_2+\cdots\cdots n\cdot f_n}{nN}\times100=\frac{\Sigma fx}{nN}\times100$$

这个公式是先求平均级别然后转化为百分率而得来的，推导如下：

x（级别）	f（出现次数）	fx
0（相当x_0）	f_0	$0\cdot f_0$（相当f_0x_0）
1（相当x_1）		$1\cdot f_1$（相当f_1x_1）
2（相当x_2）		$2\cdot f_2$（相当f_2x_2）
⋮	⋮	⋮　⋮
n（相当x_n）	fn	$n\cdot f_n$（相当f_nx_n）
	$N=\Sigma f$	Σfx

$$\bar{x}(\text{平均级别})=\frac{\Sigma fx}{\Sigma f}=\frac{0\cdot f_0+1\cdot f_1+2\cdot f_2+\cdots\cdots nf_n}{N}\cdots\cdots(1)$$

但是分级标准中的最高级别n，相当百分率分级中的100，所以再把$\bar{x}$（平均级别）转化为$\bar{x}$（以百分率表示的病情指数，即病情指数，%）时，应有：

$\bar{x}$（平均级别）：n（最高级别）＝$\bar{x}$（病情指数，%）：100

所以：$\bar{x}(\text{病情指数}\%)=\dfrac{\bar{x}(\text{平均级别})}{n}\times100$

将（1）式代入，即得：

$$\bar{x}(\text{病情指数}\%)=\frac{\Sigma fx}{nN}\times100=\frac{0\cdot f_0+1\cdot f_1+2\cdot f_2+\cdots+nf_n}{nN}\times100.$$

这里应当指出，如果采用的严重度分级标准不同，应当分别选用相应的公式来计算病情指数。而由于用简单级别分级时，每级都是用该级的最高被害严重度计算的，所以都化成以百分率表示的病情指数进行比较时，用简单级别计算所得的病情指数应比用百分率表示严重度时计算所得的病情指数要偏高些。举一实例如下：如以0＝无病；1＝25%以下被害；2＝25—50%被害；3＝50—75%被害；4＝75—100%被害，则有：

X(%)	f	fx	X(级)	f	fx
0	20	0	0	20	0
10	10	100	1	10	10
30	20	600	2	20	40
60	40	2400	3	40	120
80	10	800	4	10	40
	100	3900		100	210

$\overline{X}(\%)=\frac{\Sigma fx}{\Sigma f}=\frac{3900}{100}=39$

$\overline{X}(级)=\frac{\Sigma fx}{\Sigma f}=\frac{210}{100}=2.1$

$\overline{X}(级\rightarrow\%)=\frac{\Sigma fx}{n\cdot N}\times 100$

$=\frac{210}{4\times 100}\times 100=52.5$

即 $\overline{X}(\%)=39\%$

即 $\overline{X}(\%)=52.5\%$

因此，在同一类试验调查中，或协作调查研究中以及需要进行年度间或地区间比较时，应事先商定统一的调查分级方法和标准，采用相适应的病情指数计算公式，否则将影响正确结论的获得。

4．用假定平均数求平均数的方法：

用此法可简化计算，节省时间。其原理是利用在一数列中对每一数字同加、同减、同乘、同除一个常数（假定的平均数）后，使原数列中的数字简化，然后计算简化后的数列的平均数，再用简化数列的平均数求真正的平均数。真正平均数（$\bar{x}$）、假定平均数（c）与简化数列的平均数（M）间的关系如下：

（1）在原数列中，对每一数字同加一数：

$$M=\frac{(x_1+c)+(x_2+c)+\cdots\cdots+(x_n+c)}{N}=\frac{(x_1+x_2+\cdots\cdots+x_n)+NC}{N}$$

$$=\frac{x_1+x_2+\cdots\cdots+x_n}{N}+C=\bar{x}+C$$

所以：$\bar{x}=M-C$

（2）在原数列中，对每一数字同减一数：

$$M=\frac{(x_1-c)+(x_2-c)+\cdots\cdots+(x_n-c)}{N}=\frac{(x_1+x_2+\cdots\cdots+x_n)-NC}{N}$$

$$=\frac{x_1+x_2+\cdots\cdots+x_n}{N}-C=\bar{x}-C$$

所以：$\bar{x}=M+C$

（3）在原数列中，对每一数字同乘一数：

$$M=\frac{cx_1+cx_2+\cdots\cdots+cx_n}{N}=\frac{c(x_1+x_2+\cdots\cdots+x_n)}{N}=C\bar{x}$$

所以： $\bar{x}=\frac{M}{C}$

（4）在原数列中，对每一数字同除一数：

$$M=\frac{\frac{x_1}{c}+\frac{x_2}{c}+\cdots\cdots+\frac{x_n}{c}}{N}$$

$$=\frac{x_1+x_2+\cdots\cdots+x_n}{CN}=\frac{\bar{x}}{C}$$

所以： $\bar{x}=CM$

在农业工作中，常用同减或同除一个常数来简化计算。以少数数字举列如下：

如度量某一菌系在培养基上的生长速度，共作五次，结果分别为

10，10.5，11.0，9.5，8.5毫米/天，求平均生长速度。

X(m.m/天)	X－C（C＝10）	$\frac{X}{C}$（C＝5）
10	0	2
10.5	0.5	2.1
11.0	1.0	2.2
9.5	-0.5	1.9
8.5	-1.5	1.7
49.5	-0.5	9.9

常法： $\bar{x}=\frac{49.5}{5}=9.9$ m.m/天

用 $\bar{x}=M+C=\frac{-0.5}{5}+10=-0.1+10=9.9$ m.m/天

用 $\bar{x}=CM=5\times\frac{9.9}{5}=9.9$ m.m/天

在数字多时，用此法更可节约时间。

（二）样本变异性（离中性）的度量——标准差。

1．平方和、变量和标准差：

如上所述，生物群体间的差异，可用代表群体水平特征的平均数来比较。但是任一生物群体中的个体不都是完全一致的，群体内同样存在着差异，这种差异是衡量群体特征的另一重要标志。例如在两个小麦品种中，需要选择一个进行种植时，我们除了要评定它们的产量水平高低外，还要看它们的产量的稳定性，就是这样的问题。评选病虫防治措施时，在技术评定中同样也遇到这个问题。

对于群体变异性，即离中性的变量，曾经提出过不少办法。

用离均差不能解决这个问题。因为离均差的和〔$\Sigma(x-\bar{x})$〕为零。这点很容易证明，因为$\Sigma(x-\bar{x})=\Sigma x-\Sigma\bar{x}=\Sigma x-N\bar{x}=T-T=0$。

用离均差绝对值的和〔$\Sigma|(x-\bar{x})|$〕，可以进行比较，但只能用于样本数相同的情况下，而且不能进一步作深入的统计分析。用离均差绝对值总和的平均数$\left[\frac{\Sigma|(x-\bar{x})|}{N}\right]$，解决了样本大小不一致的问题，但对进一步深入分析的问题也未能解决。最后提出了利用标准差这个统计值来度量变异性。

标准差（常用S.D.或σ符号表示，σ为Σ的小写）的计算公式是：

$$S.D.=\sigma=\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$$

式中$\Sigma(x-\bar{x})^2$为离均差（又称对平均数的偏差）平方之和，简称平方和，

N 为观察次数。

$$\frac{\Sigma(x-\bar{x})^2}{N}$$

为离均差平方和的算术平均数，称为变量，常用V表示。在计算标准差、变量时均需先计算平方和。

2．在观察值少的情况下，平方和、变量、标准差的计算方法：

（1）冗法：即按原公式定义直接计算的方法。以上述菌系生长速度为例计算如下：

X	$(X-\bar{X})$	$(X-\bar{X})^2$
10	0.1	0.01
10.5	0.6	0.36
11.0	1.1	1.21
9.5	-0.4	0.16
8.5	-1.4	1.96
49.5		3.70

$$\bar{X}=\frac{49.5}{5}=9.9 \text{ 毫米/天}$$

$$V=\frac{\Sigma(x-\bar{x})^2}{N}=\frac{3.7}{5}=0.74$$

$$S.D.=\sqrt{V}=\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}=\sqrt{0.74}=0.86\text{毫米}$$

（2）利用恒等式$\Sigma(x-\bar{x})^2=\Sigma X^2-\frac{T^2}{N}$计算平方和，然后计算：

恒等式是这样导出的：

$$\Sigma(x-\bar{x})^2=\Sigma(x^2-2\bar{x}x+\bar{x}^2)=\Sigma x^2-2\bar{x}\Sigma x+\Sigma\bar{x}^2$$
$$=\Sigma x^2-2\bar{x}(N\bar{x})+N\bar{x}^2=\Sigma x^2-N\bar{x}^2$$

$$= \Sigma x^2 - N \cdot \left(\frac{T}{N}\right) = \Sigma x^2 - \frac{T^2}{N}$$

仍以上列计算如下：

X	X^2
10	100.00
10.5	110.25
11.0	121.00
9.5	90.25
8.5	72.25
49.5	493.75

$$\Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{T^2}{N} = 493.75 - \frac{(49.5)^2}{5} = 493.75 - \frac{49.5\times 9.9}{1}$$

$$= 493.75 - 490.05 = 3.7$$

$$V = \frac{\Sigma(x-\bar{x})^2}{N} = \frac{3.7}{5} = 0.74$$

$$S.D. = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}} = \sqrt{0.74} = 0.86\text{毫米}$$

3．在观察值多的情况下，平方和、变量、标准差的计算方法：

（1）冗法：以度量200个小麦穗穗长的结果为例。

x(穗长.公分)	f(频度)	fx	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
7.0	10	70.0	−1.0	1.00	10.00
7.5	15	112.5	−0.5	0.25	3.75
8.0	150	1200.0	0	0	0
8.5	15	127.5	0.5	0.25	3.75
9.0	10	90.0	1.0	1.00	10.00
	200	1600.0			27.50

$$\bar{x} = \frac{1600}{200} = 8\text{公分}$$

$$v = \frac{\Sigma f(x-\bar{x})^2}{N} = \frac{27.5}{200} = 0.1375$$

$$S.D. = \sqrt{0.1375} = 0.37\text{公分}$$

(2)利用恒等式：

因为：$\sum f(x-\bar{x})^2=\sum f(x^2-2\bar{x}x+\bar{x}^2)=\sum fx^2-2\bar{x}\sum fx+\bar{x}^2\sum f$

$$=\sum fx^2-2\bar{x}\cdot N\bar{x}+N\bar{x}^2=\sum fx^2-N\bar{x}^2$$

$$=\sum fx^2-\frac{T^2}{N}$$

所以可以用上列恒等式先计算平方和，再计算变量和标准差，仍以上例进行计算如下：

x(穗长.公分)	f(频度)	fx	fx^2
7.0	10	70.0	490.00
7.5	15	112.5	843.75
8.0	150	1200.0	9600.00
8.5	15	127.5	1083.75
9.0	10	90.0	810.00
	200	1600.0	12827.50

$$\sum f(x-\bar{x})^2=\sum fx^2-\frac{T^2}{N}=12827.5-\frac{(1600)^2}{200}$$

$$=12827.5-12800=27.5$$

$$V=\frac{\sum f(x-\bar{x})^2}{N}=\frac{27.5}{200}=0.1375$$

$$S.D.=\sqrt{0.1375}=0.37\text{公分}$$

4．标准差的含义和计算时应注意的问题：

如上所述标准差是用来表明总体的变异程度的最常用的方法，它是总体各个变数的离均差的平方总和的算术平均数的平方根。

在计算时用 $S.D.=\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$ 公式。 但此式适用于样本数量大时。

根据数理统计原理，如样本数量小时，变量和标标差的正确计算应作必要的修正。合理的计算公式，应用自由度（N－1）代替N作为除数。变量与标准差的计算公式应改为：

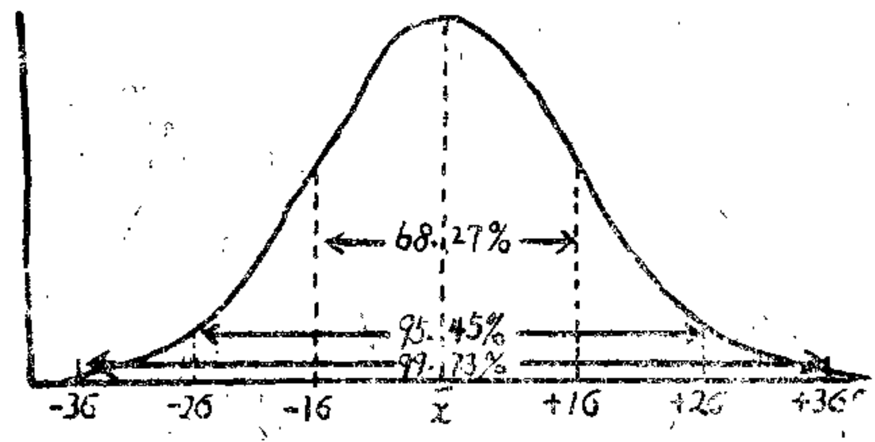
$$V=\frac{\Sigma(x-\bar{x})^2}{N-1} \quad 与 \quad S.D.=\sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}}$$

上面在观察值少时计算变量与标准差的例子中，因未引入自由度的概念，故沿用

$$V=\frac{\Sigma(x-\bar{x})^2}{N} \quad 与 \quad S.D.=\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$$

两个适用于大样本的公式。在实际应用时，应根据样本大小选用适当公式。习惯上以样本数等于30为区分标准。在样本数大于30时，计算变量和标准差用N除。在样本数小于30时，用N－1除。

标准差是度量群体变异性（或分散性）的统计值。数理分析证明，如所研究现象的分布形式是常态分布，或近似于常态分布时，在$\bar{x}\pm$ S.D.的范围内包括总体的68.27%的单位，在$\bar{x}\pm 2$ S.D.的范围内将包括总体单位数的95.45%，而有99.73%的总体单位总数包括在$\bar{x}\pm 3$ S.D.的范围内。如下图所示：



（三）变异系数：

如上所述标准差是反映总体中各个变量的分散程度或变异程度的数值。在平均数相同的两个或多个群体进行比较时，标准差大的标志群体的变异性大；相反，标准差小的则标志群体的变异性小。

但是，在比较不同群体时，如果它们的平均数不等，甚至相差很大时，直接比较标准差就不能正确说明群体间变异性的大小。因为标准差是个绝对值，它受平均数大小的影响。例如比较两个生产队小麦产量的稳定性，甲队小麦平均产量为800斤，标准差为40斤，而乙队小麦平均产量为400斤，标准差也是40斤。可见甲队小麦各地块间产量的平均差相当于平均数的5%，而乙队的为10%，甲队各地块间小麦产量的变异性（差别）要比乙队的小。或者说甲队的小麦比较稳产，平衡增产搞得较好。在这里我们可以看出，在对不同群体的变异性进行比较时，如果它们的平均数不同时，不能直接比较标准差，而应当把平均数和标准差结合起来考虑，才能求得正确结论。为了解决这个问题，提出了变异系数这个统计值。

所谓变异系数，就是以平均数为准则，看标准差是平均数的百分之几。计算公式为：

$$\text{变异系数（C.V.）}=\frac{\text{标准差（}\sigma\text{）}}{\text{平均数（}\bar{x}\text{）}}\times 100.$$

例如比较甲乙两小麦品种的穗长及其变异情况。各度量100穗。结果及分析如下：（以x示甲品种，y示乙品种）。