生物統計学基础

（牧医专业）

江
1978.7.22，
于农垦大学。

黑龙江农垦大学牧医系

一九七六年九月

毛主席语录

胸中有数。这是說，对情况和問题一定要注意到它們的数量方面，要有基本的数量的分析。任何质量都表現为一定的数量，沒有数量也就沒有质量。我們有許多同志至今不懂得事物的数量方面，不懂得基本的統計，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”，結果就不能不犯錯誤。

目 录

練习一 資料整理和次数分布

一、数字的初步整理方法——分类和分组

生物统计学是用数学逻辑来解释生物现象的一种方法，在进行畜牧生产科研的时候，工作中经常遇到许多数字，就必须用生物统计的方法，对数字进行整理，归纳和分析。

整个统计过程，包括调查测量（收集资料）、综合归纳（资料整理）及分析研究三个阶段。收集来的原始资料常是大量的、看不出规律的数字，不经过必要的整理，不容易看出任何问题，因为一个数字来自于一次测量，只能说明事物的个别现象，对来自大量测量记录的数字要经过正确处理，才能揭示事物的规律性。分类时，要把同类数字加以归类，而把不同性质的数字区别开来。例如对家畜体尺进行统计时，要考虑到家畜的类别、品种、产地、性别、年龄等等而分类，假若把幼畜与成畜的体尺混在一起统计，必然会得出荒谬的结论。

分类工作完成后，可以将同一变异特征的变异数列分组，分组步骤如下：

1．最大数、最小数及全距的确定

先由变异数列中找出最大数与最小数，最大数与最小数之差即为全距，全距即最大数与最小数之间的距离，它表示该变异数列的变异辐度。

2．组数的确定：

由样本得到的大量数字资料，必须分为若干组。组数过少，所得结果误差较大，组数过多，则计算繁杂。在一般情况下，组数是根据总次数的多少而定的，分组时可参考以下关系：

当 n＝40—60	分为6—8组
n＝60—100	分为7—10组
n＝100—200	分为9—12组
n＝200—500	分为12—17组

3．组距（i）的确定：

依组距和中价排列起来的数列称为组距数列。如果一个变异数列的总次数不多，全距又不长，而变异数间的距离又相同时，就可以不用组距和中价，而按相同的变异划为一组，这种数列称为分组数列。非连续性变异资料，如猪的产仔数，鸡的产卵量等，它所包含的变数全部是整数，可以构成分组数列。上下限间的距离、

组距是相邻二组间的距离，组距大小，按下列公式决定：

$$组距(i)=\frac{最大数-最小数}{组数}=\frac{全距}{组数}$$

如所得数值为小数，可进成整数，便于计算，各组之间的组距必须相等。

4．组限的确定：

组限是一组的两极端数，组内的较小数称下限，较大数称上限。上限与下限的中间数称中价，（上限＋下限）÷2＝中价，决定组限时，应注意不要使一个变数既能落于这一组，又能落于另一组，同时由第一组下限到最末组上限，必须包括全部变数。为了方便计算，应当尽量使中价为整数。

即上、下限的平均数，即用一个中价都来代替上、下限间的各个数、

二、分组示例

已测量成年蒙古母牛35头的前管围记录如下（单位：厘米）：

15.5 16 15.5 16 16 15.5 15.5 15.5 14 16.5 16 16 17.5 16
15.5 16.5 16 16.5 15 17.5 16 15.5 16 15.5 15.5 17 16 15.5
15.5 15.5 18 16.5 17.5 16.5 16

1．从资料中可以直接看出：最大数＝18，最小数＝14，所以全距＝18－14＝4。

2．决定组数：因总次数为35，所以可分为5组。

3．决定组距：$i=\frac{\text{全距}}{\text{组数}}=\frac{4}{5}=0.8\simeq 1$

4．决定组限：根据最小数为14，组距为1的情况，可以用下列三种组限的任何一种：

组　限	中　价	组　限	中　价	组　限	中　价
13.5—14.49	14	14.0—14.99	14.5	13.01—14.0	13.5
14.5—15.49	15	15.0—15.99	15.5	14.01—15.0	14.5
15.5—16.49	16	16.0—16.99	16.5	15.01—16.0	15.5
16.5—17.49	17	17.0—17.99	17.5	16.01—17.0	16.5
17.5—18.5	18	18.0—18.99	18.5	17.01—18.0	17.5

三、次数分布表

分组完成后，还要进一步整理，将数列中的变数根据组限进行归纳。找出各组中所包含的各变数出现的次数，这样就能对事物得出初步了解，同时为以后的计算创造条件，这种表示各组次数分布情况的数字表称为次数分布表。

1．计算示例：100头肉用猪20天内增重记录如下（单位：市斤）：

30.4 35.6 15.4 19.7 22.8 25.2 27.5 29.6 32.4 35.8 30.8 39.1
15.2 13.5 35.0 24.4 29.2 16.7 36.9 26.6 25.3 30.9 14.4 34.6
20.2 21.5 22.0 27.7 33.7 25.8 33.9 24.8 13.3 38.2 37.5 37.2
17.9 4.5 31.4 26.3 26.4 27.3 8.2 14.2 14.5 30.5 32.7 29.7
31.6 29.4 25.0 22.4 28.7 6.8 18.4 29.1 32.0 19.2 26.9 16.0
21.3 45.5 44.0 8.9 18.5 23.5 23.8 24.2 35.6 36.2 17.5 40.0
39.5 38.3 29.3 20.9 23.2 33.3 24.4 17.6 45.6 40.6 16.4 42.3
19.9 43.6 32.9 55.8 38.6 34.9 22.9 24.8 28.2 28.4 22.5 27.4
32.1 25.9 20.5 38.1

最大数＝55.8，最小数＝4.5

全距＝55.8－4.5＝51.3

决定分为9组，组距 $i=\frac{51.3}{9}\simeq 6$

确定组限后，求出每组中价，将每个变数依顺序判明应当归哪一组，就在哪一组划出符号，直到将全部变数归纳完毕，次数总和应等于全部变数的总数。

表 1—1　　　　100头肉用猪20天增重次数分布表

组　限(i=6)	中　价(x)	只 数 归 纳 符 号	次　数(f)
3 — 8.99	6	𤴓	4
9 — 14.99	12	正	5
15 — 20.99	18	正　正　正　一	16
21 — 26.99	24	正　正　正　正　下	23
27 — 32.99	30	正　正　正　正　正	25
33 — 38.99	36	正　正　正　丅	17
39 — 44.99	42	正　丅	7
45 — 50.99	48	丅	2
51 — 57	54	一	1
Σ			100

2．次数分布图：为了能更清楚地看出问题，可以在次数分布表的基础上作次数分布图。一般有三种：

（1）柱形图：

画出直角座标，纵轴等距离标出次数，横轴依次标出组距（各组下限或各组上限），然后以横轴各组组距的宽度与纵轴所标各组次数的高度构成长方形，即得次数分布的柱形图。

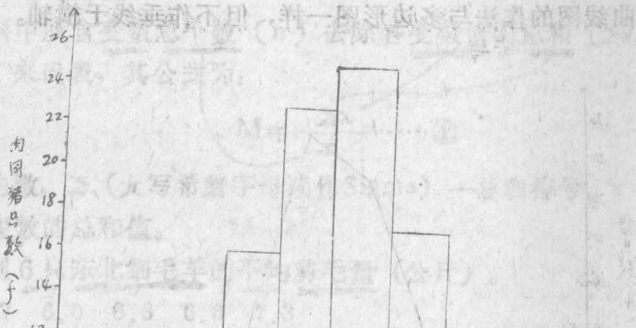


图 1—1　100头肉用猪20天增重次数分布柱形图

（2）多边形图：同样画一直角座标，纵轴代表次数，横轴代表中价，然后根据各组的次数和中价找出交叉点，各点相互联线，并从各交叉点作直线垂直于横轴，就是多边形图。

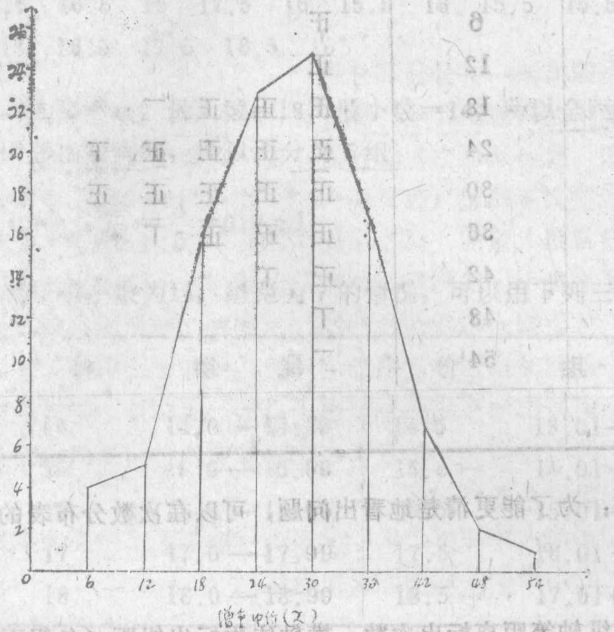
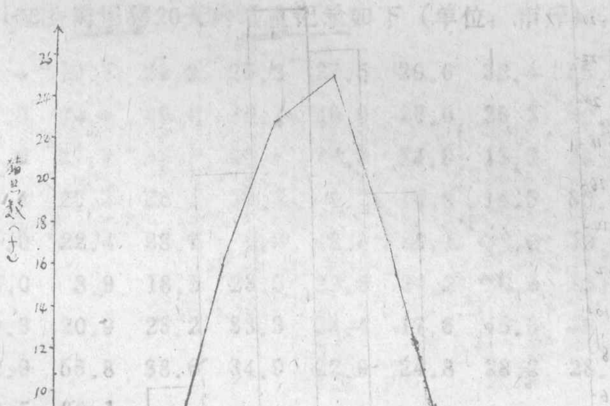


图1—2　100头肉用猪增重次数分布多边形图

（3）曲线图：曲线图的作法与多边形图一样，但不作垂线于横轴。



作业，有100头三河公犊牛出生体重记录如下（单位：公斤）：

43	21	24.5	24	26	32	29.5	39	32	32	45	23.5
27	28	32.5	30	38	31.5	29	33	50	22	25	29
38.5	28	30	30.5	32	33.5	42	26	26	30.5	28.5	31.5
28.5	38.5	29.5	33	40	27	28	29.5	32.5	29.5	31	30
39	30.5	39	28	29	38	31	33	30	34	33.5	34.5
31	35	33	31	35.5	36	33	34.5	36.5	35.5	32	35.5
36	36.5	32	34	35	34	32.5	36.5	31.5	33	34.5	31.5
33	35.5	34	33	35	34.5	33.5	36	35	36.5	34	33.5
35	34	35	31								

根据以上数据进行分组，作出次数分布柱形图及曲线图。

練习二　集中性的测定——計算平均数

一、集中性测定的意义

原始资料经过整理作出统计表和统计图之后，能够显示出资料所具有的初步意义，但是难以简明地指出变异特征的要点，更难以和其他变异特征或样本相比较，所以还要求进一步研究资料分布的集中性和离中性。测定集中性最常用的方法是平均数，平均数代表了一个资料中各变数分布的中心位置，它有两个作用：一是反映了资料的集中性，即说明了资料的一般情况，二是作为资料的代表，与另一样本的代表进行比较。

二、平均数的测定方法

平均数是资料中所含变数总个数（n）去除各变数值的总和（Σx或Σfx）所得之商，常以字母M（或$\bar{x}$）来代表，其公式为：

$$M=\frac{\Sigma x}{n}\cdots\cdots①$$

此处M—平均数，Σ（大写希腊字母读作Sigma）—总和符号，x—变数，n—变数的总个数，Σx—n个变数的总和值。

例如：求下列6只东北细毛羊的平均剪毛量（公斤）。

4.7　5.7　6.0　6.3　6.6　7.3

解：变数总个数n＝6

n个变数的总和$\Sigma x=4.7+5.7+6.0+6.3+6.6+7.3=36.6$（公斤）

$$M=\frac{\Sigma x}{n}=\frac{36.6}{6}=6.1(\text{公斤})$$

三、平均数的特性

下面介绍平均数的几种重要特点，在进行大样本的计算时，其中有些特性能简化计算手续。

1．求平均数时只要知道资料中各变数值的总和与变数的总个数即可计算。

2．每个变数都对平均数有影响，尤其是两极端的变数对平均数的影响最大。

3．各变数与平均数之差（即离均差）的总和永远等于零。

$$\Sigma(x-M)=0$$

如以前述六只东北细毛羊为例，求得各变数离均差总和为：

$$\Sigma(x-M)=(-1.4)+(-0.4)+(-0.1)+(0.2)+(0.5)+(1.2)=0$$

表2—1　　　　6只东北细毛羊剪毛量离均差表

剪毛量（变数x）	离均差（x—M）
4.7	−1.4　(4.7−6.1=−1.4)
5.7	−0.4
6.0	−0.1
6.3	0.2
6.6	0.5
7.3	1.2
特性1.　$\Sigma x=36.3$ $M=\frac{\Sigma x}{n}=\frac{36.6}{6}=6.1$（公斤）	特性2　合计：　−1.9 +1.9 0

M'=1

假定均数法 Σ(x−M')

3.7
5.0
4.7
5.3
5.6
6.3

$\Sigma(x-M')=30.6$

$M=\frac{\Sigma(x-M')}{n}+1$

$=5.1+1=6.1$

这一特性很重要，可以用来简化标准差的计算手续。

4．资料中各变数被任意一数值处理（加、减、乘、除）时，等于原平均数被此数值处理。

即：　$\frac{\Sigma x}{n}=M$　　　$\frac{\Sigma(x-M')}{n}=M-M'$　　$\frac{\Sigma(x-0)}{n}=M-0$

（变数减一个数等于平均数减一个数）

仍以6只东北细毛羊剪毛量为例：

设M'＝1时

$\Sigma(x-M')=(4.7-1)+(6.0-1)+(5.7-1)+(6.3-1)+(6.6-1)+(7.3-1)$

$=3.7+5.0+4.7+5.3+5.6+6.3=30.6$（公斤）

n＝6，M＝6.1（公斤）

代入　　$\frac{\Sigma(x-M')}{n}=M-M'$

得：$\frac{30.6}{6}=6.1-1$，　即：$5.1=5.1$

这个特性应用于我们以后要讲的平均数的假定均数法。

5．各变数与平均数之差的平方和（即离均差的平方和）为最小。

即：　$\Sigma(x-M)^2<\Sigma(x-A)^2$

仍以6只东北细毛羊剪毛量为例：

由下表可见：

当M＝6.1公斤时，各变数离均差平方和为$\Sigma(x-M)^2=3.86$。

当A＞M，设A＝6.2时，各变数离A值差平方和为：$\Sigma(x-A)^2=3.92$。

当A＜M，设A＝6.0时，各变数离A值差平方和为：$\Sigma(x-A)^2=3.92$。

说明无论A＞M或A＜M，各变数离A值差平方和$\Sigma(x-A)^2$值都大于$\Sigma(x-M)^2$值。这个特性可以用来简化标准差的计算手续。

表2—2　　6只东北细毛羊剪毛量离均差平方和最小表

剪毛量 (变数x)	离均差 (x—M)	离均差 平方 $(x-M)^2$	当A＞M，设A＝6.2 离A值差 (x—A)	离A值差的 平方 $(x-A)^2$	当A＜M，设A＝6.0 离A值差 (x—A)	离A值差的 平方 $(x-A)^2$
4.7	−1.4	1.96	−1.5	2.25	−1.3	1.69
5.7	−0.4	0.16	−0.5	0.25	−0.3	0.09
6.0	−0.1	0.01	−0.2	0.04	0	0
6.3	0.2	0.04	0.1	0.01	0.3	0.09
6.6	0.5	0.25	0.4	0.16	0.6	0.36
7.3	1.2	1.44	1.1	1.21	1.3	1.69
36.6		3.86	−0.6	3.92	+0.6	3.92

四、平均数的计算

1．小样本的计算：

因小样本所含变数个数比较少，所以一般直接由资料代入公式 $M=\frac{\Sigma x}{n}$ 计算平均数，在生物统计学中称为普通算术平均数，如前述6只东北细毛羊剪毛量的计算方法。

2．大样本资料平均数的计算：

变数超30个时，可以作为大样本。大样本因所含个数较多，可将变数按一定组距分组，作出次数分布表，其变数的总和，即为资料内有相同数值的各变数（在连续性变数资料中即各组的中价x）与次数（f）的乘积和。所以，用变数的总个数n去除有相同数值的各变数（或各组中价）乘以次数乘积和Σfx所得的商，就是平均数。这种求平均数的方法在生物统计学中称为次数分配法。所得的平均数称为加权平均数，在计算上又可分为①普通法，②简捷法。

（一）普通法

计算公式为：

$$M=\frac{\Sigma fx}{n}\cdots\cdots\cdots\cdots\cdots\cdots②$$

M—平均数，Σ—总和符号，f—分组次数，x—同样数值的变数或中价，n—变数的总次数$n=\Sigma f$。

以某单位80只3月龄哈白母猪为例：

表2—3　　80只3月龄哈白母猪体重（斤）

50.3	64.5	50.6	51.7	51.0	50.0	48.0	41.5	55.0	53.7	53.6	49.5
52.0	52.0	62.0	61.2	53.7	48.2	51.0	67.0	53.0	53.0	48.5	48.0
53.6	50.5	52.3	50.9	53.0	47.0	46.0	51.8	46.0	54.5	50.5	56.0
63.0	50.0	50.0	51.5	58.0	45.2	51.3	56.5	61.2	53.0	56.0	57.1
53.0	51.5	57.2	66.5	59.0	54.0	58.2	61.5	49.0	52.8	52.4	50.0
47.6	52.0	56.5	51.2	43.6	49.0	47.0	50.7	43.0	44.0	41.5	50.5
46.5	41.5	60.6	46.2	48.5	56.0	54.0	61.0				

计算步骤：

（1）先将资料整理作出次数分布表（第Ⅰ、Ⅱ、Ⅲ、Ⅳ项），见表2—4。

表2—4　80只3月龄哈白母猪体重的次数分布及平均数计算表（单位：市斤）

组　别	3月龄哈白母猪体重		母猪的只数	次数 × 中价
	组　距（i）	中　价（x）	（次　数f）	（fx）
Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ
1	41 — 44.9	43	6	258
2	45 — 48.9	47	13	611
3	49 — 52.9	51	26	1326
4	53 — 56.9	55	20	1100
5	57 — 60.9	59	6	354
6	61 — 64.9	63	7	441
7	66 — 68.9	67	2	134
			80	4224

（2）根据第Ⅲ、Ⅳ栏求第Ⅴ栏，即将每一组的次数f乘以该组的中价x，求出的fx值列入第Ⅴ栏，如：

第一组　$fx=6\times43=258$

第二组　$fx=13\times47=611$　　………………依次类推。

（3）将每一项fx值加在一起，求出$\sum fx$值

$\sum fx=258+611+1326+1100+354+411+134=4224$（斤）

（4）代入公式②求平均数。

$\because$已知$\sum fx=4224$，$n=80$

$$\therefore M=\frac{\sum fx}{n}=\frac{4224}{80}=52.8\text{（斤）}$$

（5）结论：80只3月龄哈白猪平均体重为52.8斤。

（二）简捷法（假定平均数法）

用普通法求平均数时，如果中价（x值）和次数（f）的数字越大，计算就越繁杂，容易发生计算错误。用假定平均数法，可以加速计算速度，减少计算错误的发生。具体算法有下列二种：

A．假定平均数的非等级法：（把中价变为一通过减一个假定平均数）（简捷法）

公式为：

$$M=M'+\frac{\sum f\beta}{n}\cdots\cdots\cdots\cdots\text{③}$$ 算出来后，同时，得加一个假定平均数、

M—平均数，M′—假定平均数，$\sum$—总和符号，f—分组次数，β—组中价与假定平均数的差（$x-M'$），n—总个数。

此公式是由平均数第四个特性推导而来的。

$$\because\ \frac{\sum(x-M')}{n}=M-M'$$

假定均数！ 离假定均数差

$$\therefore\quad M = M' + \frac{\sum(x - M')}{n} \cdots\cdots$$ 计算方法比计算平均数要小，故计算简单。但计算的结果是差数，算出平均数后，要加上假定平均数，才是真平均数。

$$\text{即}\quad M = M' + \frac{\sum f\beta}{n} \cdots\cdots ③$$

计算步骤：仍以80只３月龄哈白母猪为例。

（１）先将资料整理作出次数分布表（第Ⅰ、Ⅱ、Ⅲ、Ⅳ项）见表２～５。

表２—５　80只３月龄哈白母猪体重平均数计算表（单位：市斤）

组别	３月龄哈白母猪体重		小母猪只数	离假定均数差	离假定均数差乘次数
	组距（i）	中价（x）	（f）	（$\beta = x - M'$）	（$f\beta$）
Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ
1	41 — 44.9	43	6	−8	−48
2	45 — 48.9	47	13	−4	−52
3	49 — 52.9	51	26	0	0
4	53 — 56.9	55	20	4	80
5	57 — 60.9	59	6	8	48
6	61 — 64.9	63	7	12	84
7	65 — 69	67	2	16	32
	$\sum$		80		144

（２）设假定平均数M′。在次数分布表中任一中价（x）均可设其为假定平均数，但为计算简便起见，一般常设次数分布表中位置适中或次数最多的中价为假定平均数。因此，在例中设假定平均数M′＝51。

（３）分别求出各组中价与假定平均数的差（$\beta = x - M'$）列入第Ⅴ栏。如第一组离假定平均数差$\beta = 43 - 51 = -8$。

（４）以每一组的次数 f 乘相对的β值，求出各组的$f\beta$值列入第Ⅵ栏。如第一组的$f\beta = 6 \times (-8) = -48$。

（５）将各组的$f\beta$值总和在一起，求出$\sum f\beta$值。

$$\sum f\beta = (-48) + (-52) + 0 + 80 + 48 + 84 + 32 = 144$$

（６）代入公式③求平均数。

$$\because M' = 51,\ \sum f\beta = 144,\ n = 80$$

$$\therefore\quad M = M' + \frac{\sum f\beta}{n} = 51 + \frac{144}{80} = 51 + 1.8 = 52.8\ \text{（市斤）}$$

（７）结论：80只３月龄哈白猪平均体重为52.8市斤。

B．假定平均数等级法：（把中价再变小，一边是再除以组距，）

用假定平均数的非等级法求平均数时，已较普通法简捷得多，但当中价是小数或组距比较大时，仍会使计算繁杂。为使计算更简便，各中价与假定平均数的差可以用组距为单位来算，也就是用组距 i 除之，$d = \frac{x - M'}{i}$。最后，离假定均数差的总和的平均值（$\frac{\sum fd}{n}$）

还要再乘以组距才能成为假定均数与真正均数之间的校正数。

$$M=M'+\frac{\sum fd}{n}\cdot i \cdots\cdots\cdots\cdots\cdots ④$$

M—平均数，M′—假定平均数，∑—总和符号，f—次数，n—变数总个数，i—组距，d—以组距为单位的各中价与假定平均数之差，$d=\frac{x-M'}{i}$

计算步骤：仍以80只3月龄哈白母猪为例。

（1）先将资料整理作出次数分布表（第Ⅰ、Ⅱ、Ⅲ、Ⅳ项），见表2—6。

表2—6　80只3月龄哈白母猪体重平均数计算表（单位：市斤）

组别	3月龄哈白母猪体重		小母猪只数	离假定均数差	离假定均数差乘次数
	组距（i）	中价(x)	(f)	$d=\frac{x-M'}{i}$	fd
Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ
1	41 — 44.9	43	6	−2	−12
2	45 — 48.9	47	13	−1	−13
3	49 — 52.9	51	26	0	0
4	53 — 56.9	55	20	1	20
5	57 — 60.9	59	6	2	12
6	61 — 64.9	63	7	3	21
7	65 — 69	67	2	4	8
	∑		80		36

$\frac{全距}{组数}=\frac{69-41}{7}=4$

（2）设假定平均数M′=51。

（3）分别求出各组中价与假定均数之差，以组距除之，即得各组的d值，列入第Ⅴ项，如第一组 $d=\frac{43-51}{4}=\frac{-8}{4}=-2$。

（4）用每组的次数乘相对的d值，求出的fd列入第Ⅵ项，如第一组的fd=6×(−2)=−12。

（5）将各组的fd值总和在一起，求出∑fd值。

（6）代入公式④求平均数。

已知　M′=51，∑fd=36，n=80，i=4

$$\therefore\quad M=M'+\frac{\sum fd}{n}\cdot i=51+\frac{36}{80}\times 4=51+1.8=52.8\ (市斤)$$

（7）结论：80只3月龄哈白母猪平均体重52.8市斤。

求大样本的平均数，上面三种方法计算结果是一致的，其中以假定平均数的等级法最简捷，应用最广泛。

作　业

以练习一的100头三河牛公犊出生体重次数分布表为基础，用假定均数法求平均体重。

練习三　离中性的测定——計算标准差和变異系数

一、离中性测定的意义

平均数可以代表资料各变数的集中情况，但不能说明各变数的分散情况，平均数相同的数列，可能分散情况互不相同，例如下列两个资料：

（1）6只考力代羊母羔的初生体重为：

7　7　7　8　9　10　　　M＝8斤

（2）6只蒙古羊母羔的初生体重为：

4　4　6　8　13　13　　　M＝8斤

可以看出，考力代母羔初生重的变异范围是7～10斤之间，平均数与每个变数的差最多是2斤。而蒙古羊母羔初生重变异范围是4～13斤之间，平均数与每个变数之差为4～5斤。所以，我们还要用“标准差”这个概念去说明数列的离中情况，它有两个作用：一是说明所统计的数据的离中程度。标准差是测定各变数离开其平均数距离大小的一个指标，凡一群数值分散性大的，其标准差也大，分散性小的，标准差也小。二是反映平均数在数列里的代表能力，标准差大的，说明各变数分布不集中，其平均数的代表性就弱，反之，标准差小的，平均数的代表性就强。

因为差数有正负，所以要平方。
因为是各个变数的差数之和，平均后成为一个数。

二、标准差的测定方法

标准差是各变数与平均数之差的平方的平均数的开方。常以希腊字母σ表示。

经过开方，再还原为一个差数。

1．小样本资料的标准差计算：

A．普通法：计算公式为：

$$\sigma=\pm\sqrt{\frac{\sum(x-M)^2}{n-1}}\cdots\cdots\cdots\cdots\cdots\cdots①$$

要是用平均数自乘之和除以自由度
参p.26、36，　即离均差平方和（平方和）。

σ—标准差，M—平均数，x—变数，∑—总和符号，n—总次数，n—1自由度（注一）

以6只东北细毛羊剪毛量为例。（P.5.）

计算步骤：

（1）先将资料整理出依次表，并依此表列出平均数和标准差计算表（表3—1）。

表3—1　　6只东北细毛羊剪毛量标准差计算表（单位：公斤）见P.5.例、

剪　毛　量 （变　数 x）	离　均　差 （x－M）	离均差平方 $(x-M)^2$
4.7	－1.4　6.1公斤	1.96
5.7	－0.4	0.16
6.0	－0.1	0.01
6.3	0.2	0.04
6.6	0.5	0.25
7.3	1.2	1.44
∑x＝36.6	$\Sigma(x-M)^2$	＝3.86

（2）求平均数：

$$M=\frac{\Sigma x}{n}=\frac{36.6}{6}=6.1\text{(公斤)}$$

（3）分别求出离均差，列入第Ⅱ栏内。

（4）分别将各项离均差自乘，列入第Ⅲ栏内。

（5）总和每项离均差自乘值，求出$\Sigma(x-M)^2=3.86$。

（6）代入公式①

已知$\Sigma(x-M)^2=3.86$，$n=6$

$$\sigma=\pm\sqrt{\frac{\Sigma(x-M)^2}{n-1}}=\pm\sqrt{\frac{3.86}{6-1}}=\pm\sqrt{\frac{3.86}{5}}=\pm\sqrt{0.772}=\pm0.8786\text{(公斤)}$$

（7）结论：6只东北细毛羊剪毛量标准差为±0.8786公斤。

B. 简捷法（变数自乘法）

用普通法求标准差时，要经过求平均数，求离均差，求离均差平方等三个步骤，为了简化计算手续，可以用代数推导，把上述公式①简化为下述变数自乘法的计算公式。

$$\because \Sigma(x-M)^2=\Sigma x^2-\frac{(\Sigma x)^2}{n}\quad\text{(注二)}$$

$$\therefore \sigma=\pm\sqrt{\frac{\Sigma x^2-\frac{(\Sigma x)^2}{n}}{n-1}}\cdots\cdots\cdots\cdots②$$

注一 自由度的意义：计算平均数M时，每个变数都对M有影响，因此变数间彼此是独立的，故用n个变数计算M时就有n个自由度；计算标准差时，使用了$\Sigma(x-M)^2$，要考虑每个变数与平均数的比较，由于n个x与M比较中只有n－1个独立的比较，所以计算标准差时只有n－1个自由度，故除数用n－1。如当有两个变数10和20时，M＝(10＋20)÷2＝15，此时只有一个独立的比较，进行10与15的比较，就决定另一个必然是20与15的比较，所以说只有n－1个独立的比较，或者说自由度是n－1。

注二 $\Sigma(x-M)^2=\Sigma x^2-\frac{(\Sigma x)^2}{n}$的求证：

$$(x-M)^2=x^2-2Mx+M^2$$

$$\because M=\frac{\Sigma x}{n}$$

$$\therefore \Sigma(x-M)^2=\Sigma x^2-2M\Sigma X+nM^2$$

$$=\Sigma x^2-2\Sigma x\left(\frac{\Sigma x}{n}\right)+n\left(\frac{\Sigma x}{n}\right)^2$$

$$=\Sigma x^2-2\frac{(\Sigma x)^2}{n}+\frac{(\Sigma x)^2}{n}$$

$$=\Sigma x^2-\frac{(\Sigma x)^2}{n}$$

计算步骤：

（1）先将资料整理出依次表和标准差计算表（表3—2）

表3—2　　6只东北细毛剪羊毛量标准差计算表（单位：公斤）

剪毛量（变数X）	变数平方（X^2） (或变数自乘)
4.7	22.09
5.7	32.49
6.0	36
6.3	39.69
6.6	43.56
7.3	53.29
$\Sigma x=36.6$	$\Sigma x^2=227.12$

（2）将变数自乘，列入第二项。

（3）求出变数总和$\Sigma x=36.6$

（4）求出变数平方总和$\Sigma x^2=227.12$

（5）代入公式②

已知　$\Sigma x=36.6$　$\Sigma x^2=227.12$　$n=6$

$$\sigma=\pm\sqrt{\frac{\Sigma x^2-\frac{(\Sigma x)^2}{n}}{n-1}}=\pm\sqrt{\frac{227.12-\frac{36.6^2}{6}}{6-1}}$$

$$=\pm\sqrt{\frac{227.12-\frac{1339.56}{6}}{5}}=\pm\sqrt{\frac{227.12-223.26}{5}}$$

$$=\pm\sqrt{\frac{3.86}{5}}=0.8786\text{（公斤）}$$

（6）结论：6只东北细毛羊剪毛量标准差为±0.8786公斤。

2．大样本的标准差计算：

A．普通法：

$$\sigma=\pm\sqrt{\frac{\Sigma f(x-M)^2}{n}}\cdots\cdots\cdots\cdots③$$

σ——标准差　Σ——总和符号　f——分组次数

x—该组中价　M——平均数　n—总次数

以100头肉用猪20天增重资料为例（见表3—3）

表3—3　　100头肉用猪20天内增重（市斤）

35.6	15.4	19.7	22.8	25.2	27.5	29.6	32.4	35.8	39.1
15.2	13.5	35.0	24.4	29.2	16.7	36.9	26.6	30.9	14.4

34.6	24.4	20.2	21.5	22.0	27.7	33.7	33.9	24.8	13.3
38.2	34.9	37.5	37.2	17.9	4.5	26.3	26.4	27.3	8.2
14.2	20.5	30.5	32.7	29.7	29.4	25.0	22.4	28.7	6.8
18.4	38.1	32.0	19.2	16.0	21.3	45.5	44.0	8.9	18.5
23.5	14.5	24.2	30.4	30.8	25.3	25.8	31.4	31.6	26.9
33.3	29.1	35.6	36.2	17.5	40.0	38.5	38.3	29.3	20.9
23.2	23.8	17.6	45.6	40.6	16.4	42.3	19.9	43.6	32.9
38.6	55.8	22.9	24.8	28.2	28.4	22.5	27.4	32.1	25.9

（1）先将资料整理作出次数分布表（表3—4）（第Ⅰ、Ⅱ、Ⅲ、Ⅳ栏）

（2）根据次数分布表列出平均数和标准差计算表（第Ⅴ、Ⅵ、Ⅶ、Ⅷ栏）

表3—4　　100头肉用猪20天内增重平均数和标维差计算表

组别	20天内增重 组距（i）	20天内增重 中价（x）	次数（f）	次数×中价（fx）	离均差（x—M） 27.3	离均差平方（x—M）²	次数×离均差平方 f（x—M）²
Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ	Ⅶ	Ⅷ
1	3—8.9	6	4	24	—21.3	453.69	1814.76
2	9—14.9	12	5	60	—15.3	234.09	1170.45
3	15—20.9	18	16	288	— 9.3	86.49	1383.84
4	21—26.9	24	23	552	— 3.3	10.89	250.47
5	27—32.9	30	25	750	2.7	7.29	182.25
6	33—38.9	36	17	612	8.7	75.69	1286.73
7	39—44.9	42	7	294	14.7	216.09	1512.63
8	45—50.9	48	2	96	20.7	428.49	856.98
9	51—57	54	1	54	26.7	712.89	712.89
	Σ		100	Σfx =2730			$\Sigma f(x-M)^2$ =9171.00

（3）求平均数　$\Sigma f = n$

已知　$\Sigma fx = 2730$　　$n = 100$

$$\therefore M = \frac{\Sigma fx}{n} = \frac{2730}{100} = 27.3 \text{（斤）}$$

（4）分别求出各变数的离均差（x—M），列入第Ⅵ栏，如第一组

$x - M = 6 - 27.3 = -21.3$

（5）将各组离均差平方列入第Ⅶ栏，如第一组 $(x-M)^2 = (-21.3)^2 = 453.69$

（6）将各组次数与相对的 $(x-M)^2$ 值的乘积列入第Ⅷ栏，如第一组 $(x-M)^2$

$= 4 \times 453.69 = 1814.76$

（7）将各组 $f(x-M)^2$ 总和在一起，求出 $\Sigma f(x-M)^2 = 9171.00$

（8）代入公式③求标准差