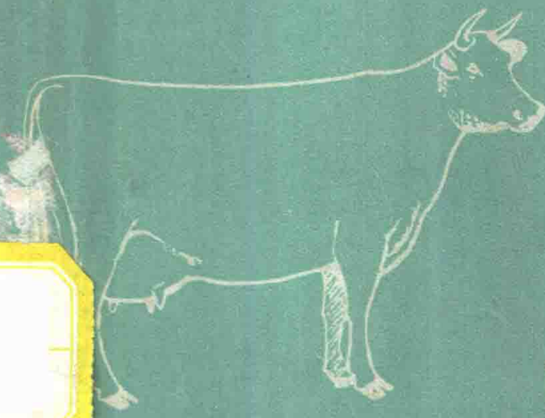
畜牧兽医专著选譯

统计方法在畜牧上的应用

〔日本〕山田淳三 著



上海科学技术出版社

統計方法在畜牧上的应用

〔日本〕 山 田 淳 三 著
刘 瑞 三 譯
刘 祖 洞 校

上 海 科 学 技 术 出 版 社

内 容 提 要

本文譯自日本«畜产の研究»杂志中一篇連載讲座，內容是介紹統計学的基本概念及其在畜牧上的应用，文字浅显易懂，可作为畜牧兽医工作人員初步掌握与应用統計方法的入門书。有关教学和科学研究人員也可参考。

畜産家のための推計学の使い方
«畜産の研究»連載講座
山田淳三
養賢堂

統計方法在畜牧上的应用
刘瑞三 譯 刘祖洞 校

上海科学技术出版社出版（上海瑞金二路 450 号）
上海市书刊出版业营业許可証出 093 号

上海大众文化印刷厂印刷 新华书店上海发行所发行

开本 850×1156 1/32 印张 3 18/32 排版字数 91,000
1965 年 6 月第 1 版 1965 年 6 月第 1 次印刷
印数 1—3,500

統一书号 16119·527 定价(科六) 0.50 元

前　　言

統計学在本世紀初期由 Fisher 創始(他也是群体遺传学創始人之一),到现在不过五十多年。专門研究这門学科的学者相当多,理論亦甚深奧,因而精通統計学并非易事。由于統計学在实际应用中价値甚大,所以我們可仅从实际应用方面来領会它。理解应用方法而能应用的人,虽在某些理論方面不十分精通,但作到正确地使用統計方法常沒有什么困难。因此,我們从使用者的立场(作者并非統計学专家,仅仅是个使用者),以实例为主来解說統計学的方法在畜牧方面的应用,其中所引实例全是从本杂志上(《畜产の研究》——譯者注)发表过的論文中引用来的。

本文的叙述方式,想先用一定篇幅說明定义与基本概念,之后再引用实例、范例来說明主要問題,并力求浅显易懂。

目　　录

一、平均数与标准偏差

在研究統計之前，我想先說明統計方法上最基本的平均数与标准偏差。例如，計算某一群10只白鼠的体重，得表1-1。这一群数字表现什么呢？先計算出它的平均数。計算的方法就是把所有的量数加起来，再用次数（或例数）来除。表1-1的例则为：

$$(19+21+17+20+23+19+20+21+20+20)\div 10=20$$

表1-1　　白鼠体重

个体編号	体重（克）
1	19
2	21
3	17
4	20
5	23
6	19
7	20
8	21
9	20
10	20
合計	200
平均	20

为方便計，用符号来表示数字。每个量数之值用$(x_1, x_2, x_3 \cdots\cdots x_n)$来表示，于是，平均数（均数）则为：

$$\bar{x}=\frac{1}{n}(x_1+x_2+x_3+\cdots+x_n),$$

式中n表示次数（例数）。$(x_1+x_2+x_3+\cdots+x_n)$用符号表示则为：

$$\sum_{i=1}^{n} x_i=x_1+x_2+x_3+\cdots+x_n \quad （簡写作\sum x）$$

因而平均数公式为

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \quad \left(\text{简写作}\frac{1}{n}\sum x\right)$$

$(x_1, x_2, x_3, \cdots\cdots, x_n)$ 的意思是，x_1 就是第一項量数，x_i 表示 i 項的量数。$\sum$讀作 (sigma)，意思就是总和。因此，$\sum_{i=1}^{n}x_i$ 的意思是从第一項量数加到 n 項量数。

以上所求得的平均数便叫算术均数（或加法均数），此外还有几何均数（或乘法均数）及調和均数，其公式分别如下：

几何均数 $\quad G=\sqrt[n]{x_1\times x_2\times x_3\times\cdots\times x_n}$

調和均数 $\quad \frac{1}{H}=\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}$

这些方法在畜牧統計上使用不多，因而从略。

从表 1-1 白鼠例中所求得的均数为 20 克。那么仅用这个均数能否表示这群白鼠的情形呢？再举例来說明，表 1-2 所求出的均数也是 20 克，与表 1-1 的均数相等。仅从这一点来看，我們不能貿然作出結論說，这两群白鼠的情形是相同的。因此，还应当进一步研究这两表的实际情形。稍一观察就能看出，表 1-1 的离差少（全距小）而表 1-2 的离差多（全距大），因而除均数外，还有必要用其他方式来描写这一群量数的状况。表示离差（即全距）的最簡单方法之一，就是求最大数与最小数的相差。表 1-1 全距为 $23-17=6$，表 1-2 全距为 $25-16=9$。但是，用此表仅知最大的量数与最小的量数的相差，中間的量数是怎样分布的，其中关系如何并不清楚，故此法亦不甚完备。

再进一步考虑使用另一种方法。先找出均数，再求各量数与均数的相差，这便是偏差。从表 1-1 及表 1-2 的例子作成表 1-3。再把各偏差总加起来，正負相消后，其結果应等于零。从以下公式即可明白：

$$\begin{aligned}&(x_1-\bar{x})+(x_2-\bar{x})+(x_3-\bar{x})+\cdots+(x_n-\bar{x})\\=&x_1+x_2+x_3+\cdots+x_n-n\bar{x}=0\end{aligned}$$

表1-2　白鼠体重

个体編号	体重（克）
1	20
2	17
3	25
4	16
5	19
6	23
7	21
8	20
9	21
10	18
合計	200
平均	20

表1-3　表1-1及表1-2中，白鼠的体重及偏差、偏差平方

表1-1的白鼠				表1-2的白鼠			
个体編号	体重	偏差差	(偏差差)²	个体編号	体重	偏差差	(偏差差)²
1	19	−1	1	1	20	0	0
2	21	1	1	2	17	−3	9
3	17	−3	9	3	25	5	25
4	20	0	0	4	16	−4	16
5	23	3	9	5	19	−1	1
6	19	−1	1	6	23	3	9
7	20	0	0	7	21	1	1
8	21	1	1	8	20	0	0
9	20	0	0	9	21	1	1
10	20	0	0	10	18	−2	4
合計 (絕对值)	200	0 (10)	22	合計 (絕对值)	200	0 (20)	66
平均	20	1.0	2.2	平均	20	2.0	6.6

若以符号$\sum$表示之，則为：

$$\sum(x-\bar{x})=\sum x-n\bar{x}=\sum x-n\cdot\frac{\sum x}{n}=0$$

根据上式总和既等于零，則不能表示差异情形，因而改用不計符号，即仅取各偏差的絕时值相加之和，以次数(例数)除之，所求得的这个数值便叫平均偏差。表 1-1 例中絕对值相加就是 10，表 1-2 例中絕对值相加就是 20。此数值随量数增大而增大。表 1-1 及表 1-2 中，其平均偏差分别为 1 克及 2 克。其公式为：

平均偏差

$$\bar{d}=\frac{1}{n}\{|x_1-\bar{x}|+|x_2-\bar{x}|+|x_3-\bar{x}|+\cdots+|x_n-\bar{x}|\}$$

$$=\frac{1}{n}\sum(|x-\bar{x}|)$$

总之，这种求平均偏差的方法，无論从实际应用上，或从統計理論上来看，已經是陈旧而繁瑣的方法了。

现在我們再研究一种不計符号而以各偏差自乘(平方)的方法。偏差自乘积的总和便叫偏差平方和(簡称平方和)。为求出平均值，用次数去除平方和，其商便叫方差(variance)。如在表 1-1 及表 1-2 例內計算的，分别为：2.2 及 6.6。其公式如下：

方差 $$\sigma^2=\frac{1}{n}\{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2\}$$

$$=\frac{1}{n}\sum(x-\bar{x})^2$$

在計算方差时，曾把各偏差自乘，如上例所得結果 2.2 及 6.6，我們认为还不够明确。只有把方差再开平方，还轉原位，才更清楚。这个数值在統計上象平均数一样使用最广。这个数值就叫标准偏差。表 1-1 及表 1-2，分别为$\sqrt{2.2}=1.5$克，$\sqrt{6.6}=2.6$克，其公式如下：

标准偏差 $$\sigma=\sqrt{\frac{1}{n}\sum(x-\bar{x})^2}$$

求偏差絕对值及用开方的方法都象上述那样簡单，也許还有人不相信。当然象在上述中所举的例子，其絕对值确实簡单。今还想逐步叙述如何对大量数字也能象簡单数字那样計算。要正确說明平方的理論，那是属于数学方面的学問，这里不作繁冗的解释。总而言之，希望大家明了开平方是个极方便的方法。

与标准偏差相关联的还有一种表现偏差的方法。一般說来，标准偏差是表示量数間变异情况的。量数分散，则标准偏差大；量数密集，则标准偏差小。在平均数不同、单位不同的数例相比較时，就用均数去除标准偏差，就是把标准偏差化作均数的百分比，就便于比較了。这个数值便叫变差系数。其公式为：

变差系数　　$C.V.=\frac{\sigma}{\bar{x}}$

表1-1及表1-2中的变差系数，分别为：

$$\frac{1.5}{20}=0.075(7.5\%),$$

$$\frac{2.6}{20}=0.13(13\%)$$

这个数值便于平均数不同、单位不同的数例相比較，但如忘記了这个数值所根据的平均数和标准偏差，就容易得出錯誤的結論。

二、平均数与标准偏差的計算方法

綜前所述，从平均数求出标准偏差，結合实际資料，用上述方法，任何人都会感到方便。但从平均数計算方差时，平均数如为整数，那当然方便；若遇到循环小数，究竟小数留到第几位才适当？或者次数过多时，用偏差自乘也比較麻煩。如果使用不要平均数的計算方法，这些問題就不存在了。为了解决上述問題，玆介紹下述方法。先看公式，再說明原理。

$$\sigma^2=\frac{1}{n}\sum(x-\bar{x})^2=\frac{1}{n}\sum(x^2-2\bar{x}x+\bar{x}^2)$$

$$=\frac{1}{n}(\sum x^2-2\bar{x}\sum x+n\bar{x}^2)$$

$$=\frac{1}{n}\left(\sum x^2-2\frac{\sum x\sum x}{n}+n\cdot\frac{(\sum x)^2}{n^2}\right)$$

$$=\frac{1}{n}\left(\sum x^2-\frac{(\sum x)^2}{n}\right)$$

在此公式中，我們先用各量数自乘，再用各数的总次数去除自乘积的总和，就得出平方和。用 n 除平方和，得出方差(均方)。在这个公式中，$\frac{(\sum x)^2}{n}$ 就叫校正数。以后在方差分析时，时常用到，这个数值很有用，应牢記在心。

为了說明公式，我們用表 1-1 例計算出表 2-1。首先，把体重自乘，在(体重)^2 項下逐一写上数值，并在項末記上总和。于是，校正数为，$200^2/10=4000$，平方和($S.S.$)为 22，方差为 2.2，标准偏差为 1.5。

表 2-1　　表 1-1 白鼠的标准偏差

个体編号	体重	(体重)^2
1	19	361
2	21	441
3	17	289
4	20	400
5	23	529
6	19	361
7	20	400
8	21	441
9	20	400
10	20	400
合計	200	4022

校正数 $=\frac{200^2}{10}=4000$　$S.S.=4022-4000=22$　$\sigma^2=\frac{22}{10}$　$\sigma=\sqrt{2.2}=1.5$

这样計算，大家也許认为不簡单，然而同下述方法結合起来，就会感觉容易了。

这个方法适用于求平均数及标准偏差。我想先以实际資料来說明。手边有一本畜牧学杂志（J. Animal Science），从中引用一个猪的例子来求平均数与标准偏差，作成表 2-2。首先，从体重中选出 200 作为假定平均数（或均数），分别从各量数中减去 200，求出 C 項的数值，再根据这个数值求出平均数与标准偏差。于是平均数、标准偏差分别为 33.1，与 20.9。B 項各数值的平均数为 $33.1+200=233.1$，C 項数值的标准偏差仍为 20.9。这样計算非常簡便。欲証明之，大家不妨亲自試試。

表2-2　母猪体重

A	B	C	D
猪的編号	体重	$B-200$	C^2
60	246	46	2116
61	252	52	2704
62	230	30	900
63	273	73	5329
64	236	36	1296
70	244	44	1936
71	229	29	841
72	211	11	121
73	201	1	1
74	209	9	81
合計		331	15325

C: 平均数 $\frac{331}{10}=33.1$　　校正数 $\frac{331^2}{10}=10956.1$

$S.S.=15325-10956.1=4368.9$

$\sigma^2=\frac{4368.9}{10}=436.89$　　$\sigma=\sqrt{436.89}=20.9$

B: 平均数 $33.1+200=233.1$　$\sigma=20.9$

在此例中，各項数值都是从以 200 为假定平均数計算出来的。但是从理論上讲，究竟选哪个数值为宜呢？例如，也可选用負数

(不减而相加)。这样負数相加也便于找出眞正数值。实际上还是自己斟酌,以选用最简单的数值为宜。为了在C項不出现負数,在此例中选 200 比較适当。在計算自乘时,还可使用自乘表,这就极便利了。从自乘表上返轉来又可查出平方根。假使用电动計算机,求自乘积就更方便。标准偏差、平均数均可笔算,50个左右的数字用自乘表或算盘求标准偏差也极方便。从以上所举实例来計算,只需五分钟就能把表制出来。

以上所說的方法是：从各个量数中减去一个假定的数值，来求平均数与标准偏差。在另外的情况下，也可利用其他的方法。例如，有一方法是把各个量数用一个假定的数字来乘（或用一假定数字来除)，再根据所得数值求出平均数与标准偏差，但乘过的数(或除过的数)要再被除（或再乘过)，才可求出原来的平均数与标准偏差。这个方法与前述方法不同之点就是：标准偏差必須还原。例如,把前述方法与第二种方法結合起来計算,試看表 2-3:在B項各量数中,最后的数字，以 0 或 5 为最多，所以我們使用这样的数值。先从B項各数值中分别减去 123,得出(C),然后再把(C)二倍之，就便利地得出 (D) 数。最后根据 (D) 求平均数与标准偏

表 2-3　　十只豚鼠的体重

A	B	C	D	E
个体編号	体　重　(克)	$B-123$	$C\times 2$	D^2
1	123.5	0.5	1	1
2	143.0	20.0	40	1600
3	144.5	21.5	43	1849
4	138.5	15.5	31	961
5	131.0	8.0	16	256
6	140.0	17.0	34	1156
7	129.5	6.5	13	169
8	145.0	22.0	44	1936
9	133.5	10.5	21	441
10	135.5	12.0	25	625
合　　計			268	8994

差，分别为：26.8 与 13.46。于是原来的，即

B 的平均数为：　$26.8\div 2+123=136.4$

B 的标准偏差为：　$13.46\div 2=6.73$

其計算过程如下：

$$D:\qquad \bar{x}=\frac{268}{10}=26.8$$

$$\text{校正数}\quad \frac{268^2}{10}=7182.4$$

$$S.S.=8994-7182.4=1811.6$$

$$\sigma^2=\frac{1811.6}{10}=181.16$$

$$\sigma=\sqrt{181.16}=13.46$$

$$B:\qquad \bar{x}=26.8\div 2+123=136.4$$

$$\sigma=13.46\div 2=6.73$$

綜合上述方法得知求眞正数值是依据：

	减去一定数值时	乘了一定数值时
求平均数时	加上該数	用該数除
求标准偏差时	原数不动	用該数除

此表的意思是：减去或乘过的，那就分别加上或除过，即减去的就加，乘过的就除。結合两种方法求平均数时，就有两步手續，必須先减去、乘过然后还原。必須依序进行，不能顛倒。

結束上項前，倘請注意一点。方差（均方）的定义就是：用次数来除平方和，其公式如下：

$$\sigma^2=\frac{1}{n}(x-\bar{x})^2$$

然而我們在实际应用时，却常以无偏方差来代替这种方差。求无偏方差的方法为：

$$u^2=\frac{1}{n-1}\sum(x-\bar{x})^2$$

公式中，以 $n-1$（称这数为自由度）来除所得之値，而不用 n。如果例数多，方差与无偏方差就非常接近。关于使用无偏方差的理

由，在总体这一概念未說明前，是无法解释的。

例数多的情况中(大样本)

使用前述方法計算50个以内的例数，大概不会有何困难，但是我們有必要計算50以上，100或1,000个例数时，用前述方法，不但多費时間，而且在計算过程中，計算上发生錯誤的可能性也增大。假使允許有細微錯誤，这里准备使用一种既簡单而又少差錯的簡捷方法。

表2-4　　100头中約克夏母猪的体重，修訂单位为平均数130公斤，标准偏差10公斤

个体编号	体重	个体编号	体重	个体编号	体重	个体编号	体重	个体编号	体重
1	125	21	118	41	134	61	135	81	117
2	121	22	146	42	130	62	142	82	128
3	119	23	141	43	118	63	107	83	138
4	127	24	129	44	131	64	144	84	118
5	139	25	131	45	133	65	143	85	111
6	103	26	130	46	120	66	135	86	130
7	135	27	153	47	129	67	143	87	142
8	137	28	131	48	128	68	122	88	142
9	132	29	139	49	140	69	119	89	130
10	140	30	145	50	139	70	130	90	131
11	115	31	113	51	129	71	125	91	138
12	121	32	129	52	134	72	125	92	132
13	157	33	126	53	130	73	141	93	119
14	148	34	125	54	137	74	130	94	121
15	133	35	133	55	123	75	130	95	133
16	130	36	124	56	134	76	136	96	123
17	130	37	114	57	127	77	117	97	124
18	116	38	136	58	112	78	141	98	124
19	149	39	122	59	120	79	136	99	133
20	127	40	126	60	128	80	147	100	126

表2-4的实际資料是經过訂正的。訂正后，平均数为130公斤，标准偏差为10公斤。初看此表，也許认为它不过是煩厌的数字表

而已。现在来說明那个簡捷方法。先把体重分为15～20个左右的等距的組，即先作出次数分布表，以組的中央值作为各組的代表值。实际計算如下：

（1）分組：先从选出的最大的数值中减去最小的数值，再以15～20来除，即：

$$157-103=54 \qquad 54\div15=3.6。$$

这样就以“3”为組距，組距尽量以奇数为宜。

（2）用3公斤的組距，把体重分組，从最小的个体减去組距的$\frac{1}{2}$$\left(\text{即3公斤的一半，或}\frac{3}{2}\text{公斤}\right)$，这就成为組距3公斤的最小的組，即：

$$103-\frac{3}{2}=101.5$$

因此，第一組即为101.5～104.5（A）

（3）数清各組所含的个体数（B）

（4）把正中这一組当作0，0以上写作-1，-2，-3……0以下的写作1，2，3，……。尽量把正中的量数作为0，計算起来才方便（即C）

（5）分别計算　$C\times B$，$C^2\times B$（即D，E）

（6）計算B、D、E各項的合計

B項即次数（这里是100）

D項即假定总和

E即假定平方和（1122）

使用这些数值来計算假定（C的）平均数与标准偏差，分别为表中所列：平均数为0，标准偏差为3.350。真正平均数（即A的）是以組距来乘假定平均数，再加上以C为0的那一組的平均数，如公式$\left(\frac{128.5+131.5}{2}=130\right)$，那么，

平均数为：　$0\times3+\dfrac{128.5+131.5}{2}=130$，

标准偏差为：$3.350\times3=10.05$