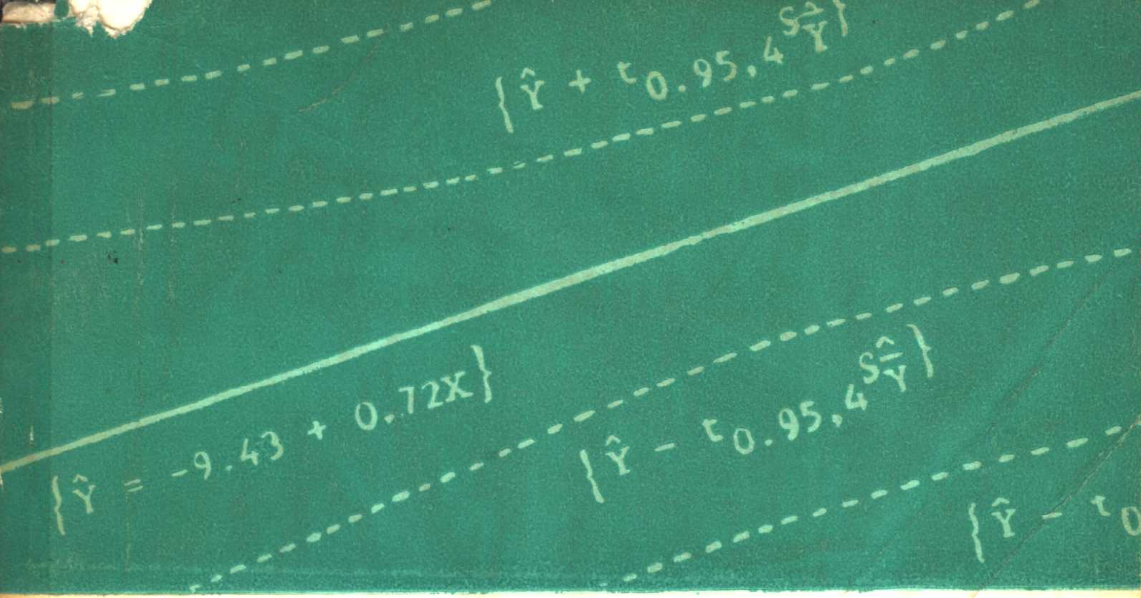

农业科学生物统计

〔美〕耿旭 F.J.希尔斯著
高明尉 张全德 胡秉民译

农 业 出 版 社

农业科学生物统计

〔美〕耿旭 F. J. 希尔斯著
高明尉 张全德 胡秉民译

农业出版社

责任编辑　张本云

农业出版社出版（北京朝阳区枣营路）

新华书店北京发行所发行　　农业出版社印刷厂印刷

850×1168毫米 32开本　8 印张　191 千字

1988年8月第1版　1988年8月北京第1次印刷

印数　1—2,300册　定价 2.85 元

ISBN 7-109-00053-2/S·38

译 者 序

本书根据美国加利福尼亚大学戴维斯分校农学及草原科学系耿旭博士及F. J. Hills博士共同编著的《农业科学生物统计》一书1982年版本译出。原书系该校农科各系大学生修习生物统计课的教材。本书主译人1982—1983年间在美征得作者同意将该书译为中文出版。

本书的特点是内容安排合理，取材简练，层次分明，叙理清楚，在某些重要概念问题上能够做到深入浅出，提纲挈领，在各章之后附有总结及相当数量的练习作业以供学习者课外巩固知识之用。本书可供农业院校师生作教学参考书，也适于广大农业试验研究人员在设计试验、整理分析农业试验数据时作参考。

本书第一至第九章由高明尉译，第十及第十一章由张全德、胡秉民翻译，并由高明尉对全书进行统一校订。原书若干明显笔误及印刷错误，均由译者在翻译过程中代为订正，不另一一注明。

一九八四年七月

目　录

第 一 章

统计符号与体制

§1.1　单向分组的综和符号

在统计计算中要求有一个简明的符号系统，这样就可免去采用复杂的公式来反映数学运算。

一个试验单位的可以计量的性状，叫做变量。一般用X、Y、Z来代表变量。变量的个体观察值叫做变量所取的值，简称变量值（variates）。变量可以是离散的（不连续的），它只能取有限个值；也可以是连续的，即在给定的区间内，它可取任何值。例如，树上的叶片数是离散型变量，而植株的高度和重量就是连续型变量。

现设有 r 个变量值：

Y_1，Y_2，$Y_3\cdots$，Y_r，

它们的和可以表示为：

$$Y_1+Y_2+Y_3+\cdots Y_r=\sum_{j=1}^{r}Y_j$$

上式中的希腊字母 Σ 表示对各项求和。综和符号 $\sum\limits_{j=1}^{r}$ 指明了在 j 区间内被加的变量值是从第 1 个加到第 r 个。当被加的变量值起讫范围已明确时，下标 j 的起讫界限就可以省略，整个运算可简单表示为 ΣY。下举一例说明如何使用综和符号：

$$\sum_{j=1}^{3}Y_j=Y_1+Y_2+Y_3$$

$$\sum_{j=3}^{5} Y_j^2 = Y_3^2 + Y_4^2 + Y_5^2$$

$$\sum_{j=1}^{2} Y_j(Y_j - 1) = Y_1(Y_1 - 1) + Y_2(Y_2 - 1)$$

如果运算中涉及到常量，例如a与b，则有

$$\begin{aligned}\sum_{j=1}^{r} (Y_j^2 + 2aY_j + b) = & \ Y_1^2 + 2aY_1 + b \\ & \ Y_2^2 + 2aY_2 + b \\ & \ Y_3^2 + 3aY_3 + b \\ & \ \cdots\cdots\cdots\cdots \\ & \ \cdots\cdots\cdots\cdots \\ & \ \cdots\cdots\cdots\cdots \\ +) & \ \underline{Y_r^2 + 2aY_r + b} \\ = & \ \Sigma Y_j^2 + 2a\Sigma Y_r + rb\end{aligned}$$

从上举一例中，可以得出以下几条规则：

1.常量的和为 r 乘以常量。

2.常量乘以变量后的和等于该常量乘以变量的和。

3.两项或两项以上的和的综和，等于各项分别综和后的和。

应注意 ΣY^2和 $(\Sigma Y)^2$有区别。

例：1—1设 $Y_1 = 3$，$Y_2 = 4$，$Y_3 = 6$，则有：

$$\Sigma Y^2 = 3^2 + 4^2 + 6^2 = 61$$

$$(\Sigma Y)^2 = (3 + 4 + 6)^2 = 169$$

§1.2 双向分组的综和符号

为了便于计算，通常都把试验数据列成双向表，其符号如表1—1所示。

在这份双向表中，用 Y_{ij} 表示一般的变量值，其中 i 表示从第 1 行到第 k 行，j 表示从第 1 列到第 r 列。例如，第 2 行 第 3

列的变量值为 Y_{23}。变量值的下标写成黑点时，表示对该行或该列的所有变量值求和。$Y_{1\cdot}$ 表示第 1 行的各变量值的和，即

$$\sum_{j=1}^{r} Y_{1j}$$

表 1—1　双向表中的变量值

行	列					合计	均数
	1	2	3	…	r		
1	Y_{11}	Y_{12}	Y_{13}		Y_{1r}	$Y_{1\cdot}$	$\bar{Y}_{1\cdot}$
2	Y_{21}	Y_{22}	Y_{23}		Y_{2r}	$Y_{2\cdot}$	$\bar{Y}_{2\cdot}$
3	Y_{31}	Y_{32}	Y_{33}		Y_{3r}	$Y_{3\cdot}$	$\bar{Y}_{3\cdot}$
·	·	·	·		·	·	·
·	·	·	·		·	·	·
·	·	·	·		·	·	·
k	Y_{k1}	Y_{k2}	Y_{k3}	…	Y_{kr}	$Y_{k\cdot}$	$\bar{Y}_{k\cdot}$
合计	$Y_{\cdot 1}$	$Y_{\cdot 2}$	$Y_{\cdot 3}$	…	$Y_{\cdot r}$	$Y_{\cdot\cdot}$	
均数	$\bar{Y}_{\cdot 1}$	$\bar{Y}_{\cdot 2}$	$\bar{Y}_{\cdot 3}$	…	$\bar{Y}_{\cdot r}$		$\bar{Y}_{\cdot\cdot}$

至于 $Y_{\cdot\cdot}$，则表示全体变量值的和，即

$$\sum_{i=1}^{k}\sum_{j=1}^{r} Y_{ij}。$$

第 1 行的均数为 $\bar{Y}_{1\cdot}(=Y_{1\cdot}/r)$，全体变量值的总均数为 $\bar{Y}_{\cdot\cdot}(=Y_{\cdot\cdot}/rk)$。

表 1—2 所列数值可供练习。

表 1—2　按双向分组的数值

行(k=4)	列(r=3)			$Y_{i\cdot}$	$\bar{Y}_{i\cdot}$
	1	2	3		
1	8	2	5	15	5.0
2	2	3	2	7	2.3
3	5	1	1	7	2.3
4	1	4	6	11	3.6
$Y_{\cdot j}$	16	10	14	40	
$\bar{Y}_{\cdot j}$	4.0	2.5	3.5		3.3

若所加的变量值的起讫界限很明确，则双重综和符号$\sum_{i=1}^{k}\sum_{j=1}^{r}$通常可简化为Σ。例如：

$$\sum_{i=1}^{k}\sum_{j=1}^{r}Y_{ij}=\Sigma Y_{ij} \quad 以及 \quad \sum_{i=1}^{k}\sum_{j=1}^{r}Y^2_{ij}=\Sigma Y^2_{ij}$$

对于表 1—2 的数值，有：

$$\Sigma Y_{ij}=8+2+\cdots\cdots+1+6=40,$$

及 $\Sigma Y^2_{ij}=8^2+2^2+\cdots\cdots+1^2+6^2=190$。

在以上各式中，点子“…”表示对所有变量值进行运算，一直计算到最后 2 个变量值为止。

§1.3 阶　　乘

在许多问题中都需用到有序的几个正整数连乘的符号。例如：

$$1\cdot2\cdot3\cdot4\cdot5\cdot6=720$$

称为“6 的阶乘”，通常用符号6! 加以表示。一般地，$n!=n(n-1)(n-2)\cdots3\cdot2\cdot1$。为了求得一致，对 0! 定义为等于 1 。下列阶乘形式在计算上很有用处：

$$8!=8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1=8\cdot7!$$
$$=8\cdot7\cdot6!=8\cdot7\cdot6\cdot5! \text{ 等等。}$$

此外还有：

$$\frac{n!}{r!}=\frac{n(n-1)(n-2)\cdots(r+1)r!}{r!}=n(n-1)(n-2)\cdots(r+1)$$

式中 r 为小于 n 的正整数。

总　　结

1. 变量可以用 X、Y 或 Z 加以表示。
2. $\sum_{j=1}^{r}Y_j$ 表示对从 Y_1 到 Y_r 的 Y 值求和。

3. $Y.$ 表示从 Y_1 到 Y_r 的Y 值均数，即 $Y./r$。

4. $\sum_{i=1}^{k}\sum_{j=1}^{r}Y_{ij}$ 表示对从第 1 列至 r 列以及从第 1 行至 k 行各变量值求和。此式可简写为 ΣY_{ij} 或 $Y..$。

5. $\overline{Y}..$ 表示双向表中全体变量值的均数，即 $Y../rk$。

6. 综和规则：

a) $\sum_{j=1}^{r}c = rc$，式中 c 为常量。

b) $\sum_{j=1}^{r}cY_j = c\Sigma Y_j$。

c) $\sum_{j=1}^{r}(Y_j^2 + 2aY_j + b) = \Sigma Y_j^2 + 2a\Sigma_j + rb$。

练　习

1. 设有一组测定值：1,3,5,7,9,求：

a) r, b) $\sum_{i=1}^{3}Y_i$, c) $\sum_{i=1}^{3}(i+Y_i)$,d) $\sum_{i=1}^{5}Y_i$, e) $\sum_{i=1}^{5}(i)(Y_i)$

2. 设有一组测定值：2,4,6,8,10,验证

a) $\sum_{i=1}^{5}2Y_i = 2\sum_{i=1}^{5}Y_i$

b) $\sum_{i=1}^{5}(2Y_i)^2 = 4\sum_{i=1}^{5}Y_i^2$

c) $\sum_{i=1}^{5}(Y_i^2 + 2Y_i - r) = \sum_{i=1}^{5}Y^2 + 2\sum_{i=1}^{5}Y_i - 25$

3. 运用代数的方法证明 $\Sigma(Y-\overline{Y}) = 0$，并利用表1—2第 1 列的数值加以验证。

4. 运用代数的方法求证表1—1双向表中数值的下述关系：

$$\sum_{j=1}^{r}Y._j = \sum_{i=1}^{k}Y_i$$

并利用表1—2的数据加以验证。

5.利用表1—2中的数据计算以下各式的值：

a) $\sum_{i=1}^{4}\sum_{j=1}^{3}(Y_{ij}-\bar{Y}_{..})^2$ (56.67)

b) $\sum_{i=1}^{4}\sum_{j=1}^{3}Y^2_{ij}-\frac{\left(\sum_{i=1}^{4}\sum_{j=1}^{3}Y_{ij}\right)^2}{rk}$ (56.67)

c) $r\sum_{i=1}^{4}(\bar{Y}_{i.}-\bar{Y}_{..})^2$ (14.94)

d) $k\sum_{j=1}^{k}(\bar{Y}_{.j}-\bar{Y}_{..})^2$ (4.68)

e) $\sum_{i=1}^{4}\sum_{j=1}^{3}(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..})^2$ (37.35)

6.对修习某门课程的学生加以计数并分类，若获得如下资料：

年级 \ 性别	男	女	合　计
大　三	30	20	50
大　四	50	40	90
研究生	40	30	70
合　计	120	90	210

a) 指出以下各符号的数值：

$Y_{2.}, Y_{1.}, Y_{2.}$ 及 $Y_{..}$

b) 计算 $\bar{Y}_{2.}, \bar{Y}_{.2}$ 以及 $\bar{Y}_{..}$ (45, 30, 35)

c) 计算 ΣY^2_{ij} (7900)

d) 计算 $\Sigma(Y_{ij}-\bar{Y}_{..})^2$ (550)

第 二 章

数据的处理与综和

§2.1 频数分布

通常人们把总体（Pupolation，在英语中又作人口解）一词理解为一个国家或者一个城市的人口数，但是从统计意义上而言，加利福尼亚州居民所构成的总体可以有各种各样的总体，例如该州居民的身高总体、体重总体以及每年收入额总体等等。从统计学上来说，总体是指有关个体的某一具体性状的观察值的全体集合。总体是用某些叫做**参数**的常量来表征的。在生物科学中我们所研究的总体大多数都很大，不能一一加以测定，因此只能从中抽取**样本**，据以获得有关总体的信息。根据样本对参数所作估计值叫做**统计量**。

为了从一个大样本中得到信息，应该先编制成频数表，这是数据分析的第一步。编制频数表时必须把观察值的变幅（即最大值减去最小值）划分为相等大小的许多组，并记下归入各个组的频数。用这种方法所得到的各个组的频数，连同各个组的组中值一起，叫做**频数分布**。表 2—1 列示 100 株甜菜的不同蔗糖浓度，这份资料是 F. J. Hills 在加利福尼亚州的伍德兰收集到的。

为了进一步将其制成频数表（表 2—2），现将整理步骤加以编号列示如下，同时将这些步骤应用于表 2—1 的数据中去。

（1）确定变幅R。将数据中的最大数值减去最小值。对表 2—1而言，有 $R = 17.2 - 4.4 = 12.8$。

（2）选择组数 k，并把数据归入到各个组。

(a) 对大多数资料而言，以划分 8 到 20 组为宜，也即8⩽

表2—1　100个甜菜块根的蔗糖浓度（%鲜重）

甜菜株号	蔗糖（%）	甜菜株号	蔗糖（%）	甜菜株号	蔗糖（%）	甜菜株号	蔗糖（%）	甜菜株号	蔗糖（%）
1	11.8	21	12.8	41	14.9	61	11.6	81	15.1
2	13.1	22	15.3	42	15.0	62	12.2	82	14.9
3	9.2	23	12.6	43	12.1	63	7.5	83	12.6
4	8.7	24	16.1	44	12.6	64	13.4	84	14.1
5	12.9	25	17.2	45	13.0	65	14.7	85	11.4
6	13.7	26	13.5	46	14.1	66	14.2	86	9.4
7	9.6	27	11.9	47	14.4	67	14.0	87	12.4
8	13.7	28	16.7	48	13.1	68	15.1	88	15.0
9	8.5	29	9.6	49	13.3	69	6.5	89	9.4
10	15.7	30	15.1	50	15.0	70	8.7	90	12.9
11	14.1	31	14.6	51	10.1	71	11.0	91	13.4
12	11.9	32	10.4	52	12.4	72	13.0	92	10.6
13	16.7	33	13.4	53	10.8	73	9.2	93	6.5
14	7.4	34	14.6	54	11.3	74	7.0	94	11.0
15	10.0	35	10.5	55	6.3	75	13.2	95	11.9
16	4.4	36	8.6	56	15.7	76	9.0	96	11.8
17	13.2	37	15.2	57	14.3	77	14.0	97	12.6
18	13.8	38	11.1	58	15.0	78	13.2	98	9.5
19	9.1	39	14.5	59	12.5	79	15.0	99	12.2
20	11.9	40	12.1	60	11.8	80	13.8	100	8.2

表 2—2　根据表2—1数据所编成的频数表

组别	组　限	组中值	频度计数	频数(f)	累计频数
1	4.05—5.55	4.8	/	1	1
2	5.56—7.05	6.3	////	4	5
3	7.06—8.55	7.8	////	4	9
4	8.56—10.05	9.3	卌 卌 ///	13	22
5	10.06—11.55	10.8	卌 卌	10	32
6	11.56—13.05	12.3	卌 卌 卌 卌 ////	24	56
7	13.06—14.55	13.8	卌 卌 卌 卌 ///	23	79
8	14.56—16.05	15.3	卌 卌 卌 //	17	96
9	16.06—17.55	16.8	////	4	100