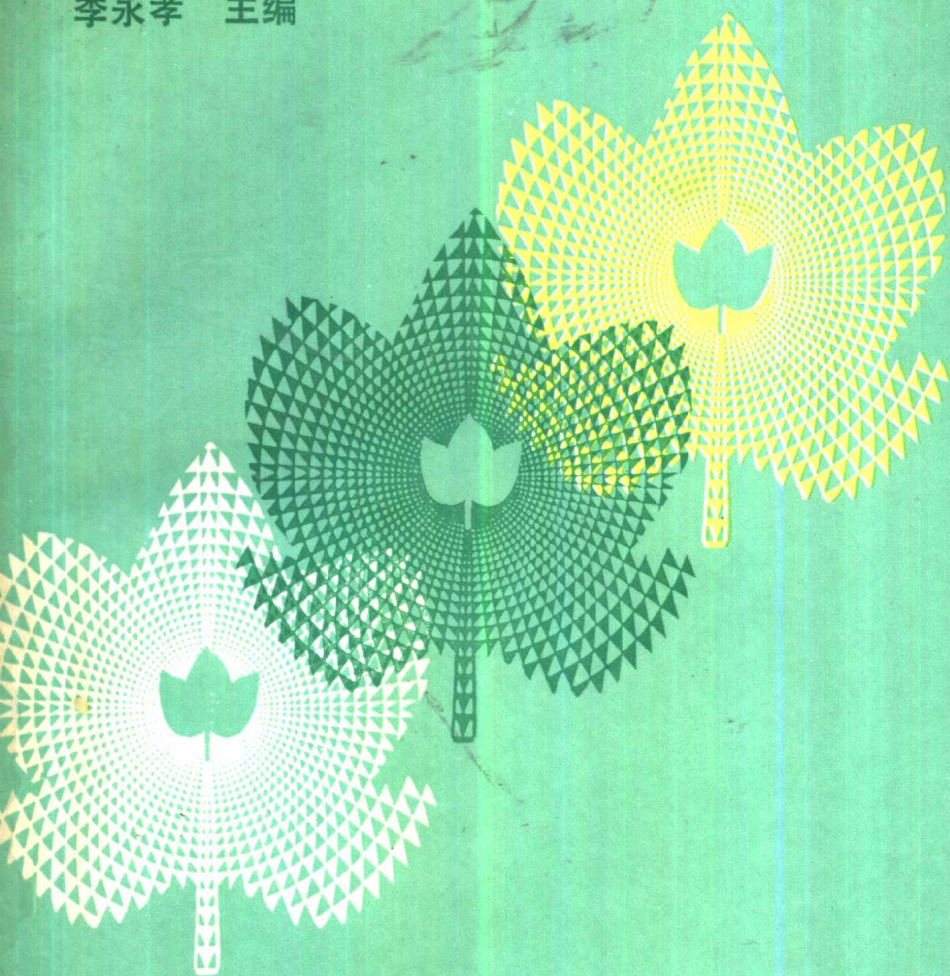


李永孝 主编



农业应用生物统计

山东科学技术出版社

农业应用生物统计

李永孝 主编

山东科学技术出版社

主编 李永孝

编著 李永孝 张存良 李佩珽

邢 邯 李翠梅 丁发武

农业应用生物统计

李永孝 主编

*

山东科学技术出版社出版
(济南市玉函路)

山东省新华书店发行

山东新华印刷厂潍坊厂印刷

*

787×1092毫米32开本 17.5印张 1插页 371千字
1989年11月第1版 1989年11月第1次印刷

印数: 1—3000

ISBN 7—5331—0588—5/S·95

定价 6.30 元

前 言

随着农业科研和生产的发展,生物统计的应用越来越广泛。近几年来,虽然有些生物统计书籍面世,但由于内容不全,公式冗长,远不能满足科研工作者、农业院校师生及农业技术干部的迫切需要。鉴于此,编写了这本《农业应用生物统计》。

本书以应用为主,由浅入深,在简要、系统介绍统计分析的基本概念和运算方法的基础上,叙述了曲线回归、二元一次线性回归和一元二次非线性回归分析;多元回归,增加了一般生物统计书中较少见的正交多项式回归和多元正交多项式回归的应用分析;第十章至第十二章介绍了逐步回归的分析及应用、通径分析及其应用、正交试验设计及分析。为了对农业科技人员探索农作物高产、优质最佳农艺方案提供设计和统计分析方法,编写了回归的正交设计和回归的旋转设计及其统计分析方法。根据育种工作者的需要,编写了作物品种稳定性分析及其应用等。

本书根据多年从事农业科研工作的实践,以多次为各地农业技术干部举办的生物统计学习班讲稿为基础,参阅了国内外的有关文献。书中尽量避免繁琐的数学推导和冗长的文字叙述,尽可能多地介绍农业科研、生产的实例分析及应用。

李永孝 于山东省农业科学院

1988年12月

目 录

第一章 基础知识	1
第一节 基本概念	1
第二节 资料的整理	2
第三节 样本特征数	9
第四节 概率.....	16
第五节 概率分布.....	18
第二章 统计推断	32
第一节 统计假设测验.....	32
第二节 参数估计.....	55
第三章 χ^2 检验	65
第一节 卡平方 (χ^2) 的概念与分布	65
第二节 适合性检验.....	69
第三节 独立性检验.....	77
第四节 样本方差的同质性检验.....	83
第四章 方差分析及应用	87
第一节 方差分析的意义与原理.....	87
第二节 单向分组资料的方差分析.....	89
第三节 两向分组资料的方差分析.....	92
第四节 多因素试验的方差分析	102
第五节 方差分析的数学模型	117
第六节 三因素试验的方差分析	120
第七节 裂区设计试验结果的方差分析	136

第五章 一元回归和相关	147
第一节 回归和相关的概念	147
第二节 直线回归方程的建立	148
第三节 直线回归方程的应用举例	152
第四节 利用回归方程进行预测	155
第五节 相关系数和决定系数	161
第六章 协方差分析	164
第一节 单向分组资料的协方差分析	164
第二节 两向分组资料的协方差分析	176
第七章 曲线回归分析	182
第一节 $\hat{y} = a + \frac{b}{x}$ 的回归方程	182
第二节 曲线方程 $\hat{y} = \frac{1}{a + bx}$	188
第三节 曲线方程 $\hat{y} = \frac{x}{a + bx}$	192
第四节 指数曲线方程 $\hat{y} = ae^{bx}$	196
第五节 曲线方程 $\hat{y} = axe^{bx}$	199
第六节 幂函数方程 $\hat{y} = ax^b$	203
第七节 生长曲线 $\hat{y} = \frac{k}{1 + ae^{bx}}$	207
第八章 二元一次线性回归及一元二次非线性回归分析	212
第一节 二元一次线性回归分析	212
第二节 一元二次回归方程	218
第三节 二元一次线性回归的矩阵解法	223
第九章 多元线性回归方程及二元二次回归方程	230
第一节 用消元法解多元线性回归方程	231
第二节 用高斯法解多元回归方程	237

第三节	用矩阵法求多元回归方程	247
第四节	回归系数的检验	264
第五节	二元二次回归方程	268
第六节	正交多项式回归方程	275
第七节	多元正交多项式回归方程	282
第八节	复相关和偏相关	289
第十章	逐步回归分析	298
第一节	最优回归方程的概念及常用公式	298
第二节	逐步回归应用实例	300
第十一章	通径分析及其应用	313
第一节	通径分析的基本理论	313
第二节	通径分析的应用	316
第十二章	正交试验设计及统计分析	328
第一节	基本概念	328
第二节	正交试验设计	332
第三节	正交试验结果分析	337
第四节	正交试验注意事项	343
第十三章	回归设计及应用分析	347
第一节	一次回归正交设计	347
第二节	二次回归正交设计	357
第三节	回归的旋转设计	377
第四节	二次通用旋转组合设计的统计分析	392
第十四章	多品种多点试验结果的统计分析	409
第一节	品种多点试验结果的统计分析	409
第二节	多年多点试验结果的统计分析	421
第十五章	品种稳定性的测定及应用	436
第一节	Eberhart and Russell 模式在作物品种稳定性分 析中的应用	436

第二节 Tai 法在作物品种稳定性分析中的应用	454
第三节 作物产量的基因型 $\times$ 环境互作效应的分析及应用	465
附表	476
参考文献	548

第一章 基础知识

第一节 基本概念

一、总体与个体

具有共同性质的个体所组成的集团，称为总体，亦可简述为研究对象的全体。如研究泰山1号小麦的株高，所有泰山1号小麦品种的株高构成一个总体。同样，所有泰山1号小麦品种的穗重，可构成另一个总体。

组成总体的每一个单元（元素）称为个体。总体的性质取决于个体的性质，是一个与一定的数量指标相联系的概念。一般构成总体的所有个体数目以 N 表示，称为总体容量。

若构成总体中的个体为有限个，则称为有限总体，如一块玉米田的所有果穗，一群牲畜中的头数等。若构成总体的个体为无限多个，则称为无限总体，如某品种的粒数、粒重等。

二、样本

在研究总体时，由于总体的无限性或测量方法具有破坏性，往往不允许对总体中的全部个体一一测量，而多采用从总体中随机地抽取若干个体的方法，用以估计所有研究的总体，这部分个体称为样本，其数目以 n 表示，称为样本容量。生物统计通常规定：

$n < 30$ 为小样本

$n \geq 30$ 为大样本

三、返置抽样

凡每个已被抽取调查观测过的个体再归入总体内，以后有再度被抽取调查之可能的抽样方法称返置抽样。凡每个已被抽取调查观测过的个体不再归入总体内，称不返置抽样。前一种抽样方法，每一个抽样个体从总体中被抽取的机会是相等的，后者是不相等的。

四、真值、估计值与抽样误差

假定对一块小麦田进行测产，麦田包括 N 个抽样单位，从这些单位产量可计算总体平均产量，这是真值。倘从总体随机抽出几个单位构成样本，从样本可计算样本平均数，这是估计值。从样本所得的估计值与总体的真值是有差异的，这个差异一般称之为抽样误差。很明显，当抽样误差很小时，样本估计值可以代表总体真值；反之，则估计值是没有作用的。所以，抽样要求从总体中获得有代表性的样本，尽量减小抽样误差，从而获得可靠的估计值。真值与估计值的差数又称为代表性误差，一般抽样调查研究中真值是未知数，因此，表示抽样误差是用样本平均数的标准误差。

第二节 资料的整理

通过田间试验的观察、测定和记载，可以得到大量的试验资料。对于这些资料，必须按照一定的程序进行科学的整理和分析，才能透过现象看本质，把感性认识逐步上升到理性认识。所以，资料的整理和分析是田间试验的重要组成部分。

分，也是深入认识客观事物的一个重要步骤。

一、次数（频数）分布表

如果数据资料所包含的观测值较少，那么资料不加整理就能看出数据的变化趋势；如果数据资料中所含的观测值很多（如有数百个观测值），则不加整理很难直接看出数据的变化趋势。如果把这些观测值按数值大小进行分组，制成次数分布表，就可以看出资料的集中和变异情况，从而对资料得到一个初步的概念。

次数分布表的编制方法，因随机变量的性质不同略有差异。

（一）离散型随机变量次数分布表的编制

现以某大豆品种的单株节数为例，简述其方法。

〔例 1—1〕从某大豆品种中随机抽取 100 株，数其单株节数，试制定次数分布表（见表 1—1）。

表 1—1 100 株大豆单株节数（节）

17	18	16	17	18	17	16	17	15	19
16	17	18	17	19	16	20	17	15	17
17	18	17	15	18	17	19	18	17	16
19	17	18	18	17	15	18	17	16	17
20	17	19	18	15	17	16	16	17	16
14	16	15	17	19	18	16	17	19	18
16	17	19	18	16	20	17	15	18	16
15	18	20	17	19	16	18	15	17	16
18	17	19	16	17	14	18	15	17	20
15	16	19	18	16	17	18	19	18	17

表 1—1 资料为离散型随机变数，把这些变数加以归类，

可分为 14、15、16、17、18、19、20 七组，组与组相差为 1 节，整理后的次数分布表见表 1—2。

表 1—2 某大豆品种株节数频数

单株节数 (x)	14 15 16 17 18 19 20	$\Sigma (n)$
次 数 (f)	2 11 19 30 21 12 5	100

从表 1—2 中，可以很直观地看出所取样本的变化范围以及取值的分布规律，每株节数在 14~20 节之间变化，多数集中在 17 节左右，小于 15 或大于 19 节出现的次数比较少。

(二) 连续型随机变量次数分布表的编制

连续型随机变量资料的整理不同于离散型随机变量资料的整理方法，因其取值可以充满一个区间，即为该区间上的任一实数。就一个具体取值来说，其发生的可能性是很小的，所以，必须先确定组数、组距、各组的上下限，然后按观测值的大小来归组。

〔例 1—2〕200 个掖单 2 号穗重（见表 1—3）。

1. 求极差 R ：

$$R = \max[x_i] - \min[x_i],$$

式中： R 为数据中的最大值与最小值之差。本例

$$R = 351 - 155 = 196 \text{ (克)}$$

2. 确定组数 (m) 和组距 (L)：组数 m 的确定，以便于反映数据的统计规律性为出发点，既考虑到资料整理的目的是为了简化资料，从中得到规律性的认识，又必须保持资料的真实面貌。一般规定：

表 1—3 报单 2 号 200 个穗的穗重 (克)

248	230	251	230	274	243	219	301	330	216	251	321	300	250	218
232	283	231	270	287	275	212	262	313	281	286	209	280	259	284
308	234	242	233	236	203	278	305	222	171	199	292	321	205	299
331	248	345	161	248	270	289	198	196	331	243	305	295	224	236
217	218	261	213	244	243	322	229	293	263	196	237	187	252	247
293	351	272	270	264	273	196	213	264	312	275	155	274	219	192
251	191	243	325	209	317	256	325	242	217	216	255	263	317	212
239	230	277	325	234	315	209	259	229	229	338	303	280	260	274
231	245	221	343	232	295	261	222	228	236	263	288	165	268	166
297	300	286	240	176	241	264	342	252	266	274	255	238	270	275
260	274	299	198	278	329	245	220	300	250	260	204	222	235	173
205	319	312	284	312	212	261	285	305	333	209	269	230	250	243
201	303	221	204	281	219	222	313	245	277	193	231	247	304	246
279	279	265	297	264										

$n < 50$, 分 5 ~ 6 组

$n > 100$, 分 10 ~ 20 组

本例 $n = 200$, $m = 15$,

$$\text{组距 } L = \frac{R}{m} = \frac{196}{15} = 13.06 \approx 13.$$

3. 确定组限 $[q_i, \bar{q}_i]$: 组限为各组的边界值, 其数值小的为下限 q_i , 数值大的为上限 $\bar{q}_i$. 组限确定的方法是首先确定第一组的下限 q_1 . 其 q_1 的求法是:

$$q_1 = \min[x_i] - \frac{L}{2} = 155 - 6.5 = 148.5$$

$$\bar{q}_1 = q_1 + L = 148.5 + 13 = 161.5$$

于是分成如下 16 组: 148.5 ~ 161.5, 161.5 ~ 174.5 ... 343.5

~356.5。因数值间是连续的，所以组与组之间也是连续的。每组的下限为闭区间，上限为开区间，即 $[\theta_i, \bar{\theta}_i)$ 。前一组的上限为下一组的下限，即观测数值恰好等于每组上限那个数值，就将其排入该组的下一组。

4. 求组中值 $\bar{x}_i$ ：每一组的中心点谓之组中值。其表达式为：

$$\bar{x}_i = \frac{\theta_i + \bar{\theta}_i}{2}$$

$$\text{如本例, } \bar{x}_1 = \frac{\theta_1 + \bar{\theta}_1}{2} = \frac{148.5 + 161.5}{2} = 155$$

$$\bar{x}_2 = 155 + 13 = 168$$

$$\bar{x}_3 = 168 + 13 = 181$$

.....

5. 统计各组频数 f_i' ：频数是观测值在某一取值区间出现的次数，记作 f_i' 。统计各组频数的方法是：按变量所取值逐个在相应的组内划线，每一取值划一线，最后统计出变量在各组区间中出现的次数，一般用“正”字划线统计次数。例如，本例第二组(161.5~174.5)共包含 165、166、173、176 四个观测值， $f_i' = 4$ ，由此便可列出次数分布表（见表 1—4）。

列出频数分布表以后，就能够比较清楚地看出数据波动的规律。频数分布表表明，200 个掖单 2 号的单穗重量在 149~351 克之间变化，多数集中在 201~330 克之间。

二、次数分布图

试验资料除用次数分布表来表示外，还可以画成图形，以便更形象地表明次数分布的情况，这类图形叫做次数分布

表 1—4

200 个掖单 2 号穗重频数分布表

组 距	组中值 $\bar{x}$	划 记 号 数	频数 f_i
148.5~161.5	155	丁	2
161.5~174.5	168	正	4
174.5~187.5	181	丁	2
187.5~200.5	194	正 正	10
200.5~213.5	207	正 正 一	11
213.5~226.5	220	正 正 正 正	20
226.5~239.5	233	正 正 正 正 下	23
239.5~254.5	246	正 正 正 正 正 一	26
252.5~265.5	250	正 正 正 正	20
265.5~278.5	272	正 正 正 正 下	23
278.5~291.5	285	正 正 正	15
291.5~304.5	298	正 正 正 下	18
304.5~317.5	311	正 正 正 正	19
317.5~330.5	324	正 正 一	11
330.5~343.5	337	正	4
343.5~356.5	350	丁	2

图。次数分布图使用最普遍的图形有直方形图和多边形图。

(一) 直方形图

直方形图适用于表示连续性变数的次数分布。现以表 1—4 200 个掖单 2 号单穗重量分布表为例说明。该表有 16 组，在横轴上 (x 轴) 分 16 等份，每一等份代表一组，第一组的上限即为第二组的下限，如此依次类推。在纵轴 (y) 上标定次数，查最多一组的次数为 26，故在纵轴上分为 26 个等份，每等份代表一次，以组距为底边划出高度为次数的矩形，使得图 1—1。这种图在统计上叫做直方形图。

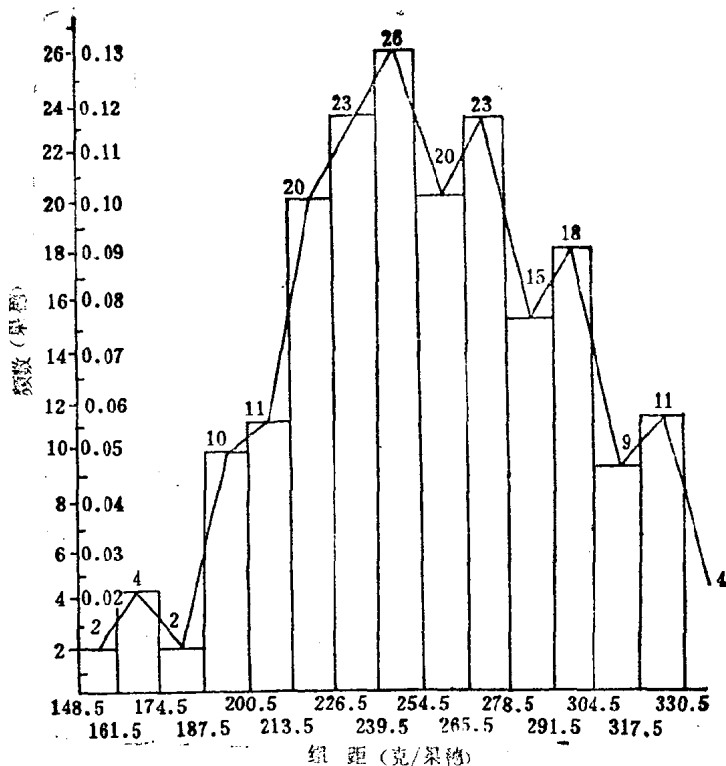


图 1—1 200 个玉米果穗穗重分布的直方形图和多边形图

(二) 多边形图

多边形图也是表示连续性变数资料的一种普遍方法，以每组的组中值表示在横坐标上，纵坐标表示次数，在直角坐标系中点绘上 $(\bar{x}_i, f_i')$ 各点，然后将各点依次连接起来，即为多边形图（见图 1—1）。如上例中玉米掖单 2 号单穗重的多边形图，以 $\bar{x}_i$ （组中值）为横坐标， f_i' （每组频数）

为纵坐标，在直角坐标系中绘出各点，再把各点连接起来即为图 1—1。

一般情况下，次数分布以直方图表示较佳。但是，如果在一幅图上同时表示几个次数分布，则宜用多边形图。间断性变数的次数分布通常不用多边形图，因为其组与组之间是不连续的。

第三节 样本特征数

在统计工作中，除了用次数分布表或次数分布图整理和概括数据外，还要用数学语言来描述数据基本特征的数量指标，反映次数分布的统计特征，这些数量指标即为样本特征数。样本特征数主要有平均数、方差、标准差、变异系数及标准误。

一、平均数（中心趋向的度量）

平均数是数量资料的代表值，表示资料中观测值的中心位置，并且可以作为资料的代表与另一组资料相比较，借以明确二者相差的情况。平均数的种类较多，其中最常用的是算术平均数。此外，尚有几何平均数、中位数和众数。

（一）算术平均数

1. 算术平均数的概念及表示方法：一个数量资料中各观测值的总和除以观测值个数所得的商数，称为算术平均数。样本平均数用 $\bar{x}$ 表示，总体平均数用 μ 表示。

2. 计算公式及计算方法：

（1）样本算术平均数的计算公式：