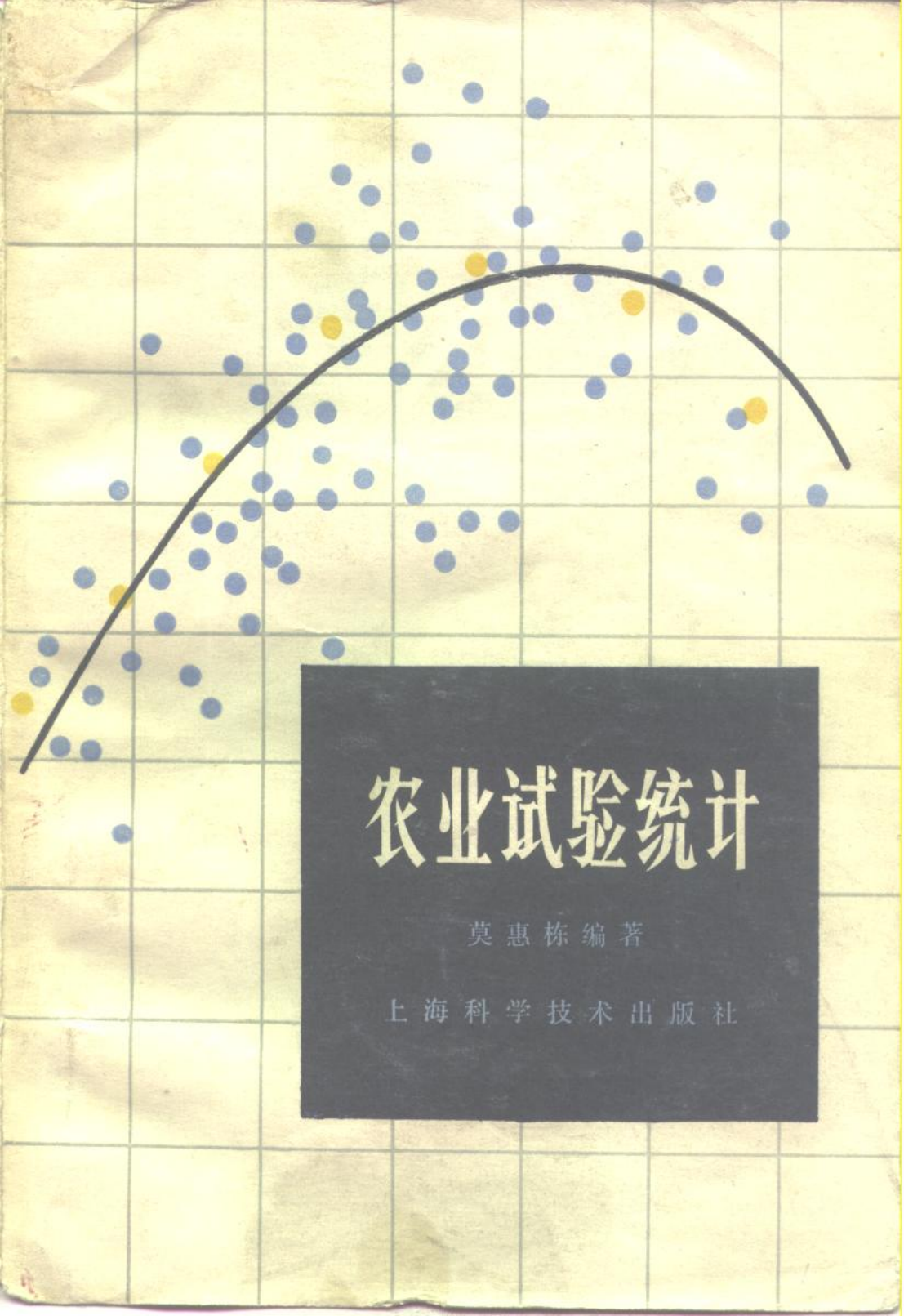


The background of the book cover is a light yellowish-tan grid. Scattered across the grid are numerous small dots, mostly blue, with a few yellow ones. A thick, solid black parabolic curve is drawn across the upper half of the grid, starting from the left edge, peaking in the center, and ending on the right edge. The dots are distributed both above and below this curve.

农业试验统计

莫惠栋 编著

上海科学技术出版社

农业试验统计

莫惠栋 编著

上海科学技术出版社

农业试验统计

莫惠栋 编著

上海科学技术出版社出版

(上海瑞金二路 450 号)

新华书店上海发行所发行 江苏扬中印刷厂印刷

开本 850×1150 1/32 印张 20 字数 524,000

1984 年 1 月第 1 版 1984 年 1 月第 1 次印刷

印数: 1—14,500

统一书号: 16119·772 定价: (科五) 2.20 元

内 容 提 要

本书系统地介绍了农业试验统计的原理和方法。全书共分 14 章,在扼要阐明次数分布、描述统计、理论分布和抽样分布的基础上,着重讨论了统计推断,次数资料的测验,非参数测验,方差分析基础和常用试验设计的方差分析,线性回归和相关分析,协方差分析,非线性回归分析,多项式回归分析,多元(线性和非线性)回归和相关分析。而对回归和相关的分析方法,尤为侧重。本书比较密切地联系我国农业试验研究的实践,附有各类例题 167 道,习题 89 道。内容安排力求循序渐进,由浅入深,深入浅出,通俗易懂。本书可供从事农业科技和教育工作的同志阅读,亦可作为有关专业大学生和研究生的学习材料,对于农业科研单位人员尤有重要参考价值。

前 言

农学和生物学试验研究所得的绝大部分数据,都不是常量,而是具有一定概率分布的变量。因此,统计方法就成为收集、整理、分析和解释这些数量资料的一个必不可少的重要工具。在国际上,近几十年来,统计方法不仅已在生物学领域内广泛而深入地渗透,而且在物理学、化学、天文学、地学、社会学、工学等领域内,也已得到日益普遍的应用。但在我国,介绍科研中的统计原理和方法的书籍,尤其是适合农业科技工作者阅读的统计书籍,却相当缺少。这和我国农业现代化的要求是不相适应的。

本书是以笔者历年来为农业院校的大学生、研究生、进修教师和农业科技干部所编的各种试验统计讲义为基础,再参考近年来国内、外的一些有关专著撰写而成。希望能对统计的基本原理和方法及其在农学和生物学方面的应用,提供一个稍为完整的轮廓,对于我国农业科研水平的提高起到一点推动和促进的作用。统计方法是呆板的,而实践资料则是无比生动而丰富的。因此在编写上,笔者特别注意了处理特殊和一般的关系,交代清楚基本概念和注意事项,使读者便于举一反三。另外,考虑到从事农业试验研究工作同志的数学基础一般不如理工科方面的同志,笔者在抽象内容的表达上也作了努力:一方面尽量避免冗长的数学推导,另一方面则引入为说明原理和方法的例题和习题 256 道。这些例题和习题,涉及作物栽培、畜禽饲养、遗传育种、植物生理、植物保护、土壤肥料、果树蔬菜等专业,为农业科技工作者所喜闻乐见,相信有助于他们从具体到抽象的思维发展。对于自学本书的读者,如能复

做例题,再做习题,也可帮助解决在原理上可能遇到的疑难之点。

在本书辑成的过程中,南京农学院马育华教授曾给予指引,并提供了许多参考文献;江苏农学院曹显祖、朱庆森等同志曾赐予大量试验资料,季穆如同志帮助绘制了全部插图,胡雪华、姜长鉴等同志协助了计算和核对。另外,本书的例题和习题,绝大部分都是实际试验结果,但因陆续摘录的时间长达20余年,原作者已难以一一追忆罗列。对于这些同志的支持和帮助,笔者一并表示衷心的感谢。

由于水平限制和涉猎资料有限,本书难免会有许多缺点和错误,敬请读者批评指正,以利日后的修订。

莫惠栋 于江苏农学院 1983年2月

目 录

绪 论	1
§ 0.1 农业试验的任务	1
§ 0.2 统计方法的功用	2
第 1 章 变数的次数分布	5
§ 1.1 次数分布的意义	5
§ 1.2 次数分布表	5
§ 1.3 次数分布图	9
§ 1.4 频率分布和累积频率分布	11
习题	13
第 2 章 平均数和变异数	14
§ 2.1 变数分布的特征	14
§ 2.2 算术平均数	14
§ 2.3 其他平均数	16
1. 几何平均数	16
2. 调和平均数	18
3. 众数	19
4. 中位数	20
§ 2.4 标准差和变异系数	22
1. 标准差	22
2. 变异系数	25
§ 2.5 由次数表求平均数和标准差	25
§ 2.6 变量的线性数学模型	28
习题	29
第 3 章 概率和理论分布	30
§ 3.1 事件和概率	30

1. 事件和概率的意义	30
2. 事件的相互关系	32
3. 计算事件概率的法则	32
§ 3.2 二项分布	34
1. 二项总体和二项分布的意义	34
2. 计算二项分布概率的方法	35
3. 二项分布的平均数和标准差	39
§ 3.3 普松(Poisson)分布	41
§ 3.4 正态分布	43
1. 正态分布的方程和性质	43
2. 标准正态分布及其累积函数	45
3. 计算正态分布一定区间概率的方法	48
4. 实际次数分布配合正态曲线	52
§ 3.5 二项分布的正态近似	54
习题	57
第 4 章 统计数的分布	59
§ 4.1 随机抽样和无偏估计	59
§ 4.2 样本平均数的分布	62
§ 4.3 样本平均数的差数的分布	66
§ 4.4 学生氏 t 分布	69
§ 4.5 样本方差的分布	72
§ 4.6 χ^2 分布和 F 分布	73
1. χ^2 分布	73
2. F 分布	76
习题	78
第 5 章 统计推断	80
§ 5.1 统计推断的意义和内容	80
§ 5.2 统计假设测验	81
1. 统计假设测验的步骤	81
2. 统计假设测验的几何意义	84
3. 两尾测验和一尾测验	84
4. 统计假设测验的两种错误	86
§ 5.3 单个平均数的假设测验	89

1. 单个平均数假设测验的方法	89
2. 单个平均数的线性数学模型	90
§ 5.4 两个平均数的假设测验——组群比较	90
1. 组群比较假设测验的方法	90
2. 组群比较的线性数学模型	93
3. 总体方差不相等时的组群比较	95
§ 5.5 两个平均数的假设测验——成对比较	95
1. 成对比较假设测验的方法	95
2. 成对比较的优点	97
3. 成对比较的线性数学模型	98
§ 5.6 参数的区间估计及与假设测验的关系	98
1. 总体平均数 μ 的区间估计	98
2. 总体平均数差数 $(\mu_1 - \mu_2)$ 的区间估计	99
3. 一尾测验的总体平均数及其差数的区间估计	100
§ 5.7 容许区间	101
§ 5.8 样本容量	103
1. 平均数的样本容量	103
2. 平均数差数的样本容量	104
§ 5.9 关于方差的统计推断	106
1. 单个方差的统计推断	106
2. 两个方差的统计推断	108
3. 多个方差的统计推断	109
§ 5.10 关于二项成数的统计推断	110
1. 单个二项成数的统计推断	110
2. 两个二项成数的统计推断	112
3. 二项成数的样本容量	114
习题	115
第 6 章 次数资料的测验	118
§ 6.1 次数资料和 χ^2 分布	118
§ 6.2 次数资料的适合性测验	119
1. $k=2$ 的次数资料	120
2. $k \geq 3$ 的次数资料	122
3. 配合理论分布的次数资料	125

§ 6.3	次数资料的齐性测验	126
§ 6.4	次数资料的独立性测验	128
1.	2×2 表的独立性测验	128
2.	$2 \times C$ 表的独立性测验	131
3.	$R \times C$ 表的独立性测验	134
习题		135
第 7 章	非参数测验	137
§ 7.1	非参数方法的意义	137
§ 7.2	符号测验	137
§ 7.3	秩和 W 的分布与秩和测验原理	141
§ 7.4	组群比较的秩和测验	144
§ 7.5	成对比较的秩和测验	146
1.	成对比较的秩和分布	146
2.	成对比较的秩和测验示例	148
习题		149
第 8 章	方差分析基础	151
§ 8.1	方差分析的意义	151
§ 8.2	单向分组资料(每组样本容量相等)的方差分析	152
§ 8.3	多重比较	157
1.	最小显著差数测验法	157
2.	Duncan 氏新复极差测验法	159
3.	Tukey 氏固定极差测验法	163
4.	Dunnett 氏最小显著差数测验法	164
5.	多重比较方法的选择	165
§ 8.4	单一自由度的独立比较	166
§ 8.5	线性数学模型和期望均方	170
§ 8.6	单向分组资料(每组样本容量不等)的方差分析	174
§ 8.7	两向分组资料的方差分析	176
§ 8.8	系统分组资料的方差分析	180
§ 8.9	变量转换	188
1.	方差分析的基本假定	188
2.	变量转换的类别	188
3.	转换资料的分析	190

习题	192
第9章 常用试验设计的方差分析	196
§ 9.1 常用试验设计的类型	196
1. 根据试验因素的多少分	196
2. 根据试验单元的排列方式分	200
§ 9.2 单因素随机区组试验的方差分析	202
§ 9.3 二因素随机区组试验的方差分析	207
1. 按固定模型分析	214
2. 按随机模型分析	222
§ 9.4 三因素随机区组试验的方差分析	223
§ 9.5 拉丁方试验的方差分析	235
1. 单个拉丁方试验的方差分析	235
2. 多个拉丁方试验的方差分析	239
§ 9.6 裂区试验的方差分析	245
1. 二裂式裂区试验的方差分析	245
2. 三裂式裂区试验的方差分析	255
§ 9.7 品种区域化试验的方差分析	259
1. 多地品种区域试验的方差分析	260
2. 多年多地品种区域试验的方差分析	267
§ 9.8 试验抽样的方差分析和抽样方案设计	278
1. 试验抽样的方差分析	278
2. 试验抽样方案的设计	282
§ 9.9 缺值估计	293
1. 缺值估计的需要	293
2. 缺值估计的原理	294
3. 随机区组试验的缺值估计	294
4. 拉丁方试验的缺值估计	297
5. 二裂式裂区试验的缺值估计	299
习题	301
第10章 线性回归和相关分析	306
§ 10.1 相关关系和散点图	306
1. 相关关系的定义	306
2. 自变数和依变数	307

3. 相关系统计分析的任务	308
4. 相关关系的散点图	309
§ 10.2 线性回归方程及离回归标准误	310
1. 线性回归方程	310
2. 回归方程的估计标准误	313
§ 10.3 线性回归模型和假设测验	316
1. 线性回归模型	316
2. 线性回归的假设测验	317
§ 10.4 线性回归的置信区间	321
1. 条件总体平均数 $\mu_{Y/X}$ 的置信区间	321
2. 条件总体观察值 Y 的预测区间	322
3. 置信区间和预测区间的图示	323
4. 回归截矩 α 和回归系数 β 的置信区间	325
§ 10.5 具有重复观察值的线性回归分析	327
§ 10.6 两个线性回归方程的比较	331
§ 10.7 相关系数和相关模型	334
1. 决定系数和相关系数	334
2. 线性相关模型	338
§ 10.8 相关系数的统计推断	340
1. 相关系数的假设测验	340
2. 相关系数的置信区间	343
3. 两个样本相关系数的比较	344
§ 10.9 相关和回归的关系及计算程序	345
1. 相关和回归的关系	345
2. 相关和回归的计算程序	347
§ 10.10 由相关表计算相关和回归	347
1. 相关表的编制	347
2. 由相关表计算相关和回归	348
§ 10.11 相关和回归的应用注意	353
习题	356
第 11 章 协方差分析	360
§ 11.1 协方差分析的意义	360
§ 11.2 单向分组资料的协方差分析	361

§ 11.3	回归系数的假设测验	369
§ 11.4	处理平均数的回归矫正和矫正平均数的假设测验	370
1.	处理平均数的回归矫正	370
2.	矫正平均数 $\bar{y}_{i(X=\bar{x})}$ 的假设测验	373
3.	矫正平均数 $\bar{y}_{i(X=\bar{x})}$ 的多重比较	375
§ 11.5	协方差分析的简化	377
§ 11.6	协方差分析的线性数学模型和基本假定	379
§ 11.7	两向分组资料的协方差分析	383
§ 11.8	系统分组资料的协方差分析	388
	习题	393
第 12 章	非线性回归分析	396
§ 12.1	非线性回归分析的意义	396
§ 12.2	非线性方程 $\hat{Y} = \frac{a+bX}{X}$	398
§ 12.3	非线性方程 $\hat{Y} = \frac{1}{a+bX}$	406
§ 12.4	非线性方程 $\hat{Y} = \frac{X}{a+bX}$	413
§ 12.5	非线性方程 $\hat{Y} = ae^{bX}$	419
§ 12.6	非线性方程 $\hat{Y} = aXe^{bX}$	423
§ 12.7	非线性方程 $\hat{Y} = aX^b$	428
§ 12.8	非线性方程 $\hat{Y} = ae^{bX^2}$	432
§ 12.9	对称 S 形曲线(正态累积函数)	436
§ 12.10	不对称 S 形曲线	442
§ 12.11	Logistic 生长曲线 $\hat{Y} = \frac{K}{1+ae^{-X}}$	446
§ 12.12	其他具有两个统计数的非线性方程	453
	习题	458
第 13 章	多项式回归分析	465
§ 13.1	多项式回归模型	465
§ 13.2	多项式方程回归统计数的计算	467
1.	多项式回归统计数的正规方程组	467
2.	解正规方程组的方法	468
§ 13.3	多项式方程的基本性质和初步选择	472

§ 13.4	多项式方程的统计选择	474
1.	多项式方程统计选择原理	474
2.	多项式方程统计选择示例	476
3.	多项式回归的计算程序	481
§ 13.5	由相关表计算多项式回归方程	481
§ 13.6	正交多项式回归	487
1.	正交多项式回归分析原理	487
2.	正交多项式回归分析程序	490
3.	正交多项式回归分析示例	491
§ 13.7	相关指数	496
	习题	497
第 14 章	多元回归和相关分析	499
§ 14.1	多元线性回归模型	499
§ 14.2	多元回归统计数的正规方程组和解	502
§ 14.3	多元回归和偏回归的假设测验	510
1.	多元回归的假设测验	510
2.	偏回归系数的假设测验	511
§ 14.4	由相关表计算多元线性回归方程	514
§ 14.5	正交多元线性回归方程	518
§ 14.6	最优多元线性回归方程的统计选择——逐步回归方法	521
1.	逐个淘汰不显著自变数的回归方法	522
2.	逐个选入显著自变数的回归方法	528
§ 14.7	多元回归中自变数的相对重要性——通径系数	537
1.	通径系数的意义	537
2.	通径系数的计算	539
3.	通径系数的类型和应用	544
§ 14.8	多元相关和偏相关	545
1.	复相关系数及其假设测验	545
2.	偏相关系数及其假设测验	549
3.	偏相关(回归)和单相关(回归)的区别	554
§ 14.9	多元非线性回归方程	556
1.	由 X_i 的转换值 X'_i 组成的多元非线性回归方程	556
2.	由 X_i 和 Y 的非线性形式表示的多元非线性回归方程	556

3. 多元多项式回归方程	558
习题	569
附表	572
表 1 正态分布的概率密度函数表	572
表 2 正态分布的累积函数表	573
表 3 t 分布的临界 t_{α} 值表	576
表 4 χ^2 分布的临界 χ^2_{α} 值表	577
表 5 F 分布的 $F_{0.05}$ (上)和 $F_{0.01}$ (下)值表	578
表 6 正态总体的容许系数 l 值表	588
表 7 符号测验 n_{α} 值表	589
表 8 秩和测验 W 表	591
表 9 秩和测验 W' 表	591
表 10 Duncan's 新复极差测验 $SSR_{0.05}$ 和 $SSR_{0.01}$ 值表	592
表 11 q 测验的 $q_{0.05}$ 和 $q_{0.01}$ 值表	594
表 12 Dunnett's 测验的 $Dt_{0.05}$ 和 $Dt_{0.01}$ 值表	596
表 13 成数 p 的反正弦转换表	597
表 14 r 与 R 的 0.05 和 0.01 临界值表	600
表 15 r 值与 z 值转换表	601
表 16 百分率和概率单位(P)转换表	603
表 17 正交多项式系数 c_j 值表	607
主要统计符号注释	615
主要参考文献	620

§ 0.1 农业试验的任务

农业试验是为了提高农业生产力而进行的一种自觉的、有计划的科研实践。

农业试验的任务,首先在于解决农业生产中需要解决的问题。例如,某地区水稻白叶枯病流行,为了解决这个问题,就需要进行多方面的农业试验,如:①对各种防治病害措施作出鉴定,提供生产上利用;②征集抗病的水稻品种,通过比较试验,供生产上择优选用;③以抗病品种为亲本进行杂交,通过育种试验,选出抗病高产的优良基因型,以取代生产上的感病品种等。同样,为了制订某作物的科学管理规程或某家畜的饲料配方,以充分发挥品种的增产潜力,就需要进行一系列的栽培试验或饲养试验。

农业试验通常是在易于控制的较小的空间中进行的。因而,它有可能最大限度地排除各种非研究因素的干扰,将需要研究的问题充分地突出起来;同时,它又可以向各个方面试探解决问题的最佳方案,而不致造成大的损失。所以,农业试验能够较有效率地解决农业生产中存在的一些问题。由于农业生产有着较强的地区性,而生产水平又是在不断发展的,会不断出现新的情况和问题,故这类直接为农业生产服务的试验,在各个农区都是面广量大的。

农业试验也是解决农业科学提出的问题的有效手段。这是由于农业试验并不完全依赖于农业生产。它可以通过控制或改变某些条件,提供生产中不能或不易自然发生的新的事实,以造成新的

科学观念或科学假定;也可以根据一定的科学观念或科学假定,设计出相应的试验,来检验这些观念或假定的正确性。例如,已成为动、植物育种的重要基础的分离定律、自由组合定律和连锁交换定律,就是通过控制条件下的生物杂交试验而得出科学观念,又被从这种观念出发而进行的大量再试验所证实的。所以,从根本上说,农业科学的发展要以农业生产的发展为基础,这是毋庸置疑的。但是,要使农业科学走在农业生产的前头,却又非侧重于农业试验不可,这也是没有疑问的。

从农业试验和农业生产的关系来说,农业试验通常都可看作是农业生产的先行和准备。“先行”体现了农业试验的探索性和先进性,“准备”则体现了农业试验的目的性。为了迅速发展我国的社会主义农业生产和农业科学,必须大力加强农业试验研究工作。

§ 0.2 统计方法的功用

在农业试验中,为了深刻认识研究对象或过程的表现和规律,往往都要进行系统的观察和测定。而观察测定的项目,大多表现为一定的数量。因而统计方法就成为做好农业试验的一种必不可少的工具。统计方法的基本功用有:

(1) 提供整理和描述数据的科学方法: 试验数据通常都是有变异的,例如同一组条件下 100 个玉米果穗的穗重数据,往往各不相同。在统计上,将这种性状或特性的有变化的数据叫做变数(variable),而变数的每一个具体值则叫做变量(variate)或观察值(observed value)。在试验实践中,可能要分别研究每一个变数,也可能要联合研究某两个或两个以上的有关变数。由于每一变数包含较多的变量,要是将每个变量都罗列起来,往往庞杂零乱,而又不能说明任何问题。统计方法提供了整理资料、化繁为简的科学程序,以及由众多变量归纳出几个确能描述该变数特征的数值的计算方法,使试验者从少数的特征数或一些简单的图表,就能了解到数据中蕴藏的信息。