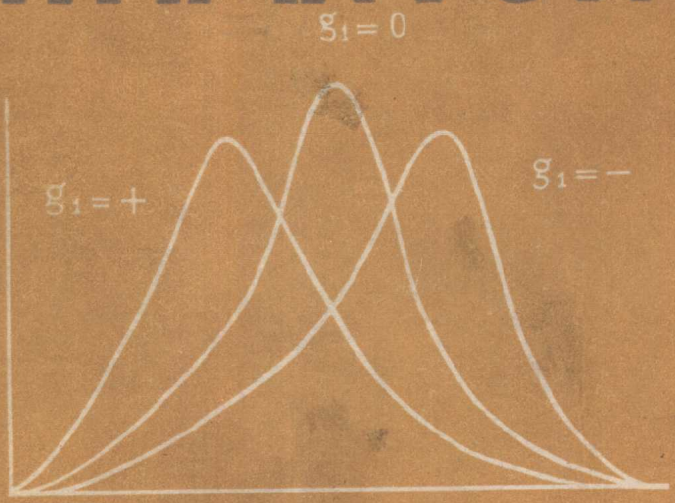
试验统计

SHIYANTONGJI

马育华编著



农业出版社

试 验 统 计

马育华 编著

农 业 出 版 社

内 容 简 介

本书叙述试验统计的基本原理及其计算方法。总共十八章。第一、二、十一、十二共四章介绍科学试验设计与田间试验的基本方法与分析。第三、四、九、十、十四、十五共六章阐述一般描述统计和相关方法，其中包括简单回归、相关，复回归、相关，曲线回归、相关以及次数资料分析。自第五章起论述有关统计推断方法，包括概率知识，理论分布，抽样分布，统计假设测验以及方差、协方差分析等共五章。第十六、十七章介绍简易统计方法与非参数性统计方法，这些方法简便易懂，适于当前农业科技工作的应用。第十八章讨论试验过程中的取样技术。最后三章乃一般统计方法书所较少介绍的内容。

本书可作为高等农业院校的田间试验与统计方法课程的参考书，也可供一般农业研究院所研究人员和农村科技人员的参考。

试 验 统 计

马育华 编著

农业出版社出版（北京朝内大街130号）

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　27 印张　701 千字

1982 年 1 月第 1 版　1982 年 1 月北京第 1 次印刷

印数 1—6,000 册

统一书号 16144·2177　定价 4.90 元

序

自然科学的试验方法，也是一种实践、认识、再实践、再认识的过程，它的步骤包括：第一，观察；第二，设立假设；第三，设计与进行试验以获得数据，并通过分析以验证假设的正确性；第四，推论，即肯定或否定这个假设，并综合地总结提高为科学理论。在此过程中，设计试验与分析试验数据的方法占非常重要的位置，而这种方法即构成了“试验统计”的内容。因此，也可以说，“试验统计”是一门应用统计方法于设计试验与分析试验数据的科学，也就是考验假设是否正确的一种工具。所以，它是每一个从事试验研究工作者必不可少的知识。

统计方法应用于农业与生物学试验，近年来在试验设计、取样技术和遗传统计等方面有新的发展。本书在统计方法方面阐述了试验数据的整理与初步分析、理论分布、统计推断、相关、非参数性统计方法与简易统计方法等六个部分，其中非参数性统计与简易统计两章，乃近代统计学的新内容，对于简化试验数据的统计处理具有重要意义。试验设计方面阐述了田间试验方法、顺序设计以及随机设计中的随机区组、拉丁方、交叉式设计、裂区与条区等等。取样技术方面阐述了纯随机取样、分层取样、顺序取样和多级取样等方法的设计与分析。遗传统计方面因内容丰富，拟留待以后另编写专书出版。

编写本书的目的旨在较为详尽地阐述试验统计的方法和理论，以供读者参考应用。但为了科学系统性关系及避免内容庞杂，故未过多涉及数理统计的内容，仅引出几个主要统计数如c、t、F与x^2

等抽样分布的推导，以供参考。

书中例题较多，皆涉及农业与生物学试验问题，以解释统计方法与原理的应用。例题所用数据分为两类：一为试验资料，目的在于说明统计方法对试验结果的实际应用，但实际资料计算繁难，不易用来阐明统计原理与计算步骤；为此，另一类则采用假定数据，目的在于简捷地说明统计方法。

本书乃从编者二十多年来讲授试验统计课程的讲稿整理的，1966年将要出版，因故未能问世。修订此书，增加了新的内容，但仍保留原版章节。限于水平，书中谬误之处，请读者批评指正。

马育华　1978年

此为试读，需要完整PDF请访问：www.ertongbook.com

目　录

序
第一章　引言……1
第一节　科学试验与统计方法……1
第二节　试验设计……3
第二章　田间试验的基本方法……7
第一节　田间试验的任务、性质与种类……7
第二节　田间试验的试验方案的拟订……11
第三节　田间试验误差的控制……15
一、试验地的选择……16
二、控制土壤差异的方法……18
第四节　田间试验应注意的事项……25
第三章　试验资料的统计处理……32
第一节　试验资料的性质……32
第二节　依次表和次数分布表……33
第三节　次数分布的图示方法……40
第四章　样本平均数与标准差……45
第一节　算术平均数……45
一、平均数的任务和种类……45
二、算术平均数的计算方法……46
三、算术平均数的性质……52
第二节　两极差与标准差……53
一、两极差（全距）……54
二、标准差与方差……54
第三节　变异系数……63
第四节　中数与四分位距……65
第五章　概率、二项分布与泊松分布……68
第一节　事件的概念……68

第二节 概率的概念 ……71
一、概率的定义……71
二、概率的计算……75
第三节 重复试验的概率分布 ……79
第四节 二项分布 ……85
一、二项式概率分布 ……85
二、二项分布的平均数与标准差……89
三、二项分布的形状和正态接近 ……93
第五节 泊松分布 ……95
第六章 正态分布与抽样分布 ……101
第一节 正态分布的方程式和特性 ……101
第二节 连续性变数的概率分布 ……106
第三节 非正态性的形式和测验 ……115
第四节 抽样分布及其试验 ……121
第五节 样本平均数分布 ……125
第六节 样本平均数的差数分布 ……132
第七节 二项总体抽样 ……135
第七章 统计假设测验……140
第一节 统计假设测验的基本原理 ……140
一、统计假设 ……140
二、统计假设测验的基本方法 ……142
三、假设测验的两尾测验与一尾测验……147
四、假设测验的两类错误 ……148
第二节 参数的区间估计（置信距与置信限）……153
第三节 平均数假设测验 ……160
一、样本平均数与总体平均数的比较……160
二、两个样本平均数的比较 ……162
第四节 平均数的假设测验当总体σ为不知条件下（t—测验）……165
一、样本平均数和总体平均数的比较……168
二、在配偶试验条件下两个样本平均数的比较 ……170
三、两个样本平均数的比较 ……173
四、两个样本具有显著差异的标准差，其平均数的比较……176
第五节 变异度的假设测验 ……178

一、样本方差与总体方差的比较 …… 181
二、两个样本方差的比较 …… 184
三、大样本变异度的假设测验 …… 188
第六节　百分数（成数）资料的测验 …… 189
一、百分数或成数测验 …… 190
二、百分数的差数测验 …… 194
第七节　泊松分布数据的测验 …… 197
第八章　方差分析 …… 204
第一节　方差分析的基本原理与计算步骤 …… 204
一、变异各个部分的平方和与自由度的分裂以及方差估计 …… 205
二、统计假设的显著性测验——F 测验 …… 208
三、两个平均数间差数的假设测验——t 测验 …… 216
第二节　多重比较 …… 217
一、对试验处理与一个对照的比较 …… 218
二、对试验处理间所有可能各对差数的比较 …… 220
三、对各个类型的对比测验 …… 225
第三节　单方面分类的分析 …… 228
第四节　方差分析的数学模型与基本假定 …… 235
一、方差分析的线性可加模型 …… 235
二、方差分析的基本假定 …… 238
三、试验效应的三种统计模型 …… 241
第五节　两方面分类的分析 …… 246
一、单独观察值 …… 246
二、重复观察值 …… 251
三、效应的统计模型 …… 256
第六节　方差分析的若干有关问题 …… 260
一、对不符合基本假定数据的处理 …… 260
二、多个样本方差的比较 …… 264
三、两种线性模型的期望均方和（参数估计） …… 267
第九章　直线回归与相关 …… 272
第一节　直线回归 …… 273
一、散布图 …… 273
二、直线回归方程与回归系数 …… 275

三、直线回归的数学模型和基本假定……278
第二节 回归的假设测验与置信区间 ……280
一、估计值的抽样分布 ……280
二、回归系数的假设测验与置信限估计 ……282
三、两个回归系数的比较 ……290
第三节 依变数重复试验的回归分析 ……292
一、部分试验点重复试验 ……293
二、全部试验点均进行重复试验 ……296
第四节 直线相关 ……301
一、相关系数 ……301
二、相关系数的假设测验与置信限……306
三、两个相关系数比较与多个相关系数比较 ……308
四、组内相关 ……311
第五节 分组资料的回归与相关 ……314
一、相关表的编制……314
二、相关表内 b 和 r 值的计算 ……315
第十章 次数资料的分析……320
第一节 χ^2值的定义与计算……320
一、χ^2的定义……320
二、χ^2的分布及假设测验 ……321
三、χ^2分布与 c，t 和 F 分布的关系……324
第二节 适合性的测验 ……326
一、（1×2）表……326
二、（$1\times n$）表……328
三、χ^2分析的再分裂方法 ……329
第三节 独立性的测验 ……332
一、（2×2）相依表 ……333
二、（$2\times J$）相依表 ……335
三、（$r\times c$）相依表 ……336
第四节 同质性的测验 ……338
一、几组杂交后代分离结果的同质性——异质性χ^2计算……338
二、（2×2）相依表的同质性测验 ……341
三、相关系数的同质性 ……343

第五节　应用分散指数以测验小样本资料 …… 345
第六节　最大偏差方法用于适合性测验 …… 348
第十一章　田间试验常用的设计 …… 352
第一节　田间试验设计的基本原理与简单举例 …… 352
一、完全区组设计 …… 356
二、不完全区组设计 …… 358
第二节　顺序排列的试验设计 …… 360
一、顺序排列设计的两种类型 …… 360
二、康士坦丁纳夫的对比设计 …… 362
第三节　完全随机试验设计 …… 366
一、完全随机试验设计的特点 …… 366
二、完全随机设计的统计分析 …… 368
三、完全随机设计的数学模型 …… 371
第四节　随机区组设计 …… 372
一、随机区组设计的特点 …… 372
二、随机区组的随机方法与田间布置 …… 372
三、随机区组试验的统计分析 …… 376
四、随机区组试验的缺区估计 …… 381
五、随机区组试验的数学模型 …… 386
第五节　拉丁方设计 …… 389
一、拉丁方设计的特点 …… 389
二、拉丁方设计的随机方法与应用时的要点 …… 390
三、拉丁方试验的统计分析 …… 392
四、拉丁方试验的缺区估计 …… 398
五、拉丁方试验的数学模型 …… 402
第六节　交叉式设计 …… 403
第十二章　复因子试验的设计 …… 405
第一节　复因子试验在农业试验中的重要意义 …… 405
第二节　裂区试验的设计与分析 …… 409
一、裂区设计的各种排列 …… 410
二、再裂区试验设计 …… 414
三、裂区设计的统计分析要点与误差估计 …… 415
四、裂区试验的统计分析例题 …… 416

五、裂区设计在应用时的要点和缺区估计 …… 426
六、裂区试验的数学模型 …… 430
第三节 条区试验的设计与分析 …… 432
一、条区设计的几种排列方式 …… 433
二、条区设计的统计分析与标准差估计 …… 435
三、条区试验的统计分析例题 …… 436
四、应用条区设计应注意的要点 …… 439
第十三章 协方差分析 …… 441
第一节 协方差分析的原理 …… 441
第二节 用协方差分析来控制试验误差 …… 452
第三节 协方差分析的基本假设、假定与注意要点 …… 464
第十四章 复回归、复相关与部分相关 …… 466
第一节 复回归方程与部分回归系数 …… 467
第二节 部分回归系数的显著性测验与置信限 …… 475
第三节 复相关系数与部分相关系数 …… 482
第四节 应用高斯乘数以分析回归问题 …… 488
第五节 行列式方法计算复回归方程与复相关系数、部分回归系数 …… 496
第十五章 曲线回归与相关 …… 513
第一节 曲线关系的类型 …… 513
第二节 对数曲线（指数曲线）的配合——直线转换方法 …… 516
第三节 多项式曲线的配合 …… 522
第四节 用均衡多项式配合回归曲线 …… 530
一、用ξ_γ表配合多项式法 …… 534
二、用总和法配合多项式 …… 539
第五节 随手曲线的配合及其应用 …… 545
第六节 相关指数 …… 549
第七节 相关比率——相关表的曲线相关方法 …… 554
第十六章 简易的统计方法 …… 563
第一节 平均数的估计值 …… 564
第二节 变异数的估计值 …… 571
第三节 两个平均数的比较 …… 580
第四节 平均数的置信限 …… 585

此为试读，需要完整PDF请访问：www.ertongbook.com

第五节　两个变异数的比较 ……589
第六节　几个样本平均数的比较——极差分析 ……591
第七节　随机区组的极差分析——两方面分类的分析 ……596
第八节　拉丁方试验的极差分析 ……600
第九节　两个因子试验（复因子试验）具有重复观察值的极差分析……604
第十节　两个因子试验排列为区组设计的极差分析 ……609
第十一节　裂区试验的极差分析 ……613
第十二节　回归系数与回归方程式 ……619
第十七章　非参数性的统计方法 ……625
第一节　游程测验 ……626
一、游程的概念……626
二、游程数的抽样分布 ……628
三、非随机性测验……632
四、两个样本分布的测验 ……634
第二节　符号测验 ……636
第三节　Wilcoxon 配偶数据符号等级测验 ……642
第四节　等级和测验 ……645
一、等级和测验应用于两个样本分布的比较 ……646
二、等级和测验用于比较复杂的试验资料 ……649
三、用等级和测验以比较几个样本的分布 ……651
四、等级和测验的多重比较 ……655
第五节　中数测验 ……656
一、两个样本的中数测验 ……657
二、k 个样本的中数测验 ……658
第六节　随机性测验 ……660
一、邻差均方和测验……661
二、连续相关 ……663
第七节　等级相关 ……666
第八节　契比歇夫不等式 ……669
第九节　中数的置信限 ……672
第十节　非参数性的宽容限度 ……673
第十一节　非参数性统计方法应用时的要点 ……677

第十八章 试验过程中的取样技术 ……680
第一节 取样目的与取样的统计问题 ……680
第二节 估产取样技术 ……682
第三节 取样技术的一些研究——取样单位的面积、形状及其比较 ……684
第四节 分层取样 ……688
第五节 顺序取样 ……692
第六节 取样方法的种类与取样误差估计 ……695
第七节 纯随机取样平均数的取样误差 ……705
第八节 纯随机取样成数或百分数的取样误差 ……711
第九节 纯随机取样确定必要的抽样数目的方法 ……714
一、确定连续性数据的必要抽样数目公式 ……715
二、确定成数数据的必要抽样数目公式 ……717
第十节 分层取样的统计分析 ……720
一、分层取样关于平均数的取样误差 ……720
二、分层取样关于成数数据的取样误差 ……729
三、分层取样确定抽样数目的方法 ……734
第十一节 多级取样的设计 ……740
第十二节 多级取样的统计分析 ……749
一、多级取样的平均数与标准差 ……749
二、方差分析以估计取样误差 ……750
三、三级取样方法 ……753
四、确定抽样单位数目的方法 ……758
五、化学分析上抽样误差的控制 ……763
附录一 试验用的统计表 ……767
附录二 统计参考书 ……839
索引 ……845

第一章 引　言

第一节 科学试验与统计方法

统计学是应用于试验过程中所获得的观察资料的数学。任何从事于科学试验的农业工作者必须掌握统计方法。

统计方法对科学试验有哪些重要意义呢？

科学试验所得的数量资料往往是大量的，而且是复杂、凌乱无章的，怎样把这些资料整理和精简，以几个简单的数值表示其全部资料结果，使试验工作者容易领会其主要内容和结论，这显然是很重要的。再者，试验的原始资料往往包括有一些不十分确切的或者粗糙的内容，通过统计处理可以去伪存真，把最重要的和真实的结果提炼出来，从而可以深入研究彼此的关系。毛泽东同志在《实践论》中指出："要完全地反映整个的事物，反映事物的本质，反映事物的内部规律性，就必须经过思考作用，将丰富的感觉材料加以去粗取精、去伪存真、由此及彼、由表及里的改造制作工夫，造成概念和理论的系统，就必须从感性认识跃进到理性认识。"我们应当把毛泽东同志的理论作为思想武器，运用到试验统计工作的实际中去。

任何试验都不可能在十分有把握的控制条件下进行，这意味着在除了所要研究的因子外，尚有许许多多外来因子干扰试验的进行，因而产生了"试验误差"，影响试验的结果。例如同一个品种的小麦种子播在田里所长成的植株，其生长发育的表现不会完全相同；同一品种的大豆种子在各次油分分析中获得相异的结果。可见，无论在

能变性较大的生物和农业试验中，或者在控制条件较佳的化学、物理试验中，试验误差是普遍存在的。统计方法帮助我们估计试验误差的数量，进一步把试验结果与误差估计进行比较，因而可以判断试验的表面效应，是归属于所研究的试验处理的直接结果，或者仅是由于试验误差的影响而发生的。这样一种“测验”令人分清某一种试验结果是不是真实的、或者是建立在不充分的证据上、或者仅仅是虚假的。

科学试验的目的不仅是研究事物的本质，更重要的是研究事物间的相互联系。例如，研究大豆品种的产量，必须联系到品种的植株特征和产量因素如荚数、粒数、百粒重等以及生理特性等等来进行研究，发现它们之间的相关，这样才能真正明了丰产品种的特征，从而掌握这些特征而有把握地来选育丰产品种。这种研究事物（指变数）之间相关的性质以及相关程度的方法是统计方法的一部分，而这些方法提供试验工作者以科学预见的作用。

由于在试验过程中存在着试验误差，因此很明显的我们在确定一个品种的产量时，不能以一株植株的生产率做为该品种产量的真值，我们必须测定大量植株产量的群体以获得其真实产量。换句话说，要从试验中发现规律性，必须从研究总体着手，而不能仅以其个体结果下结论。所谓总体乃指具有相似特征、性质的大量个体的群体。统计方法所研究的对象是总体并非个体。从总体中发现的规律称统计规律。自然科学中若干原理都是通过总体研究建立的，例如气体运动原理、自然选择原理等等。尽管每个气体分子的运动速度无法可知，但在大量分子时在一定温度与压力下都有它一定的规律性。在试验研究中如果忽略所获结论的统计性质，那么我们在解释试验结果时就难免发生错误。但是，研究总体往往是不可能的，而且是过于费时费事的，因而需要从这个总体随机抽取一部分个体，把这部分个体组成为一个样本。从样本研究着手，以所获样本的结论来推断其总体可能的结果。一切试验都是从样本进行研究，从这

些试验所获结论都是用来推断总体的。因此不懂得统计推断方法和每个试验结论的统计性质，试验工作者就一定会犯错误的论断。

做一个试验不仅要懂得试验误差的存在和估计，而且必须要求在其所进行的试验中得到一个没有“偏性”的试验误差估计，并且使这个试验误差为最小。要做到这点，我们又必须懂得试验设计。统计方法是进行正确的试验设计所不可缺少的。

第二节 试验设计

田间试验的试验设计内容，主要是讨论如何在试验田上布置试验小区的原理和方法，使得有一个试验误差的估计，而且这个误差估计是没有“偏性”与最小的。怎样做到这点呢？让我们从品种比较试验举例说明这一概念。

如果要确定一个品种的真实产量，在试验田种上一个小区，小区面积 1 平方米，所得产量以推测每亩产量，这样做是不够精确的。因为只有一个小区产量数字，没有方法估计试验误差，所以科学试验的第一个原理是试验必须有重复，意思是这一品种多种几个小区，从这些小区的产量差异可以估计出试验误差的大小，这样我们可以从这一个品种的样本产量结果以确定其真实产量所在的范围。倘要比较两个品种的产量，看看孰优孰劣，在试验田上的一小区上种上 A 品种，另一小区上种上 B 品种，如果所有外来因子都假定全部控制，那么这样比较也许是可以的。但是事实上没有这种理想的条件，试验误差仍然是存在的，因此这种试验方法是不合理的。假定扩大这个试验，比较 4 个品种的产量，这时考虑到外来因子的干扰，其中尤其是土壤肥力的差异性，因此采用每个品种重复种上 4 个小区，这样在试验田上则需有 16 个试验小区，这里出现了怎样布置这 16 个小区的问题。

现将这 16 个试验小区安置在 4 条地段上，每条地段的土壤肥

力有很大差异，其中隔着一小条的排水沟，每条地段的4个小区可称为一个区组。倘若我们把这4个品种A、B、C和D如图1·1布置在这4个区组上，这种设计显然是荒谬的。因为名义上每个品种重复了4个小区，实际上这4个小区连接在一条地段上象一个大区一样，而且倘出现了差异，则这种差异可能是由于区组间肥力的差异，亦有可能是由于品种间的差异。

略好一点的布置方法是将这4个品种完全随机分配在这16个小区上，丝毫没有主观的选择，这样可以将16个卡片上标上4个A、4个B等等，然后一一放在一个小布袋内，完全混杂后，一一地抽取出来，一一地填写在4个区组里，其排列结果如图1·2所示。

采用随机排列使得由此估计的试验误差排除了发生“偏性”的可能，这是科学试验的第二个原理，但是每个品种的平均产量却引进了误差，因为有两个A品种和两个C品种小区发现在第一个区组、两个B品种小区在第二个区组等等。这样品种本身又包括有和第一种设计一样的区组间的差异，这样一个试验设计还是不合乎要求的。

为着使品种的差异和区组的差异能分开来，采用每个区组内各小区安排不同品种，如图1·3所示，这样作显然就会使试验误差降低了。这种把不同品种布置在同一个区组内的作法，称为地区控制，它是科学试验的第三个原理。

详细研究图1·3的布置，发现在4个区组内，A品种常和B品种为邻，同样B和C为邻，而C和D为邻。这种设计称为顺序排列设计，在估计试验误差方面仍是有“偏性”的，因此仍不是一个很好的设计。

但是，倘按图1·3每个区组的4个品种采用随机排列，即可获得如图1·4所示的结果，这种设计看出每一个品种都重复地占有4个小区；所有用来比较的品种在每个区组中出现了同样的次数，而

此为试读，需要完整PDF请访问：www.ertongbook.com