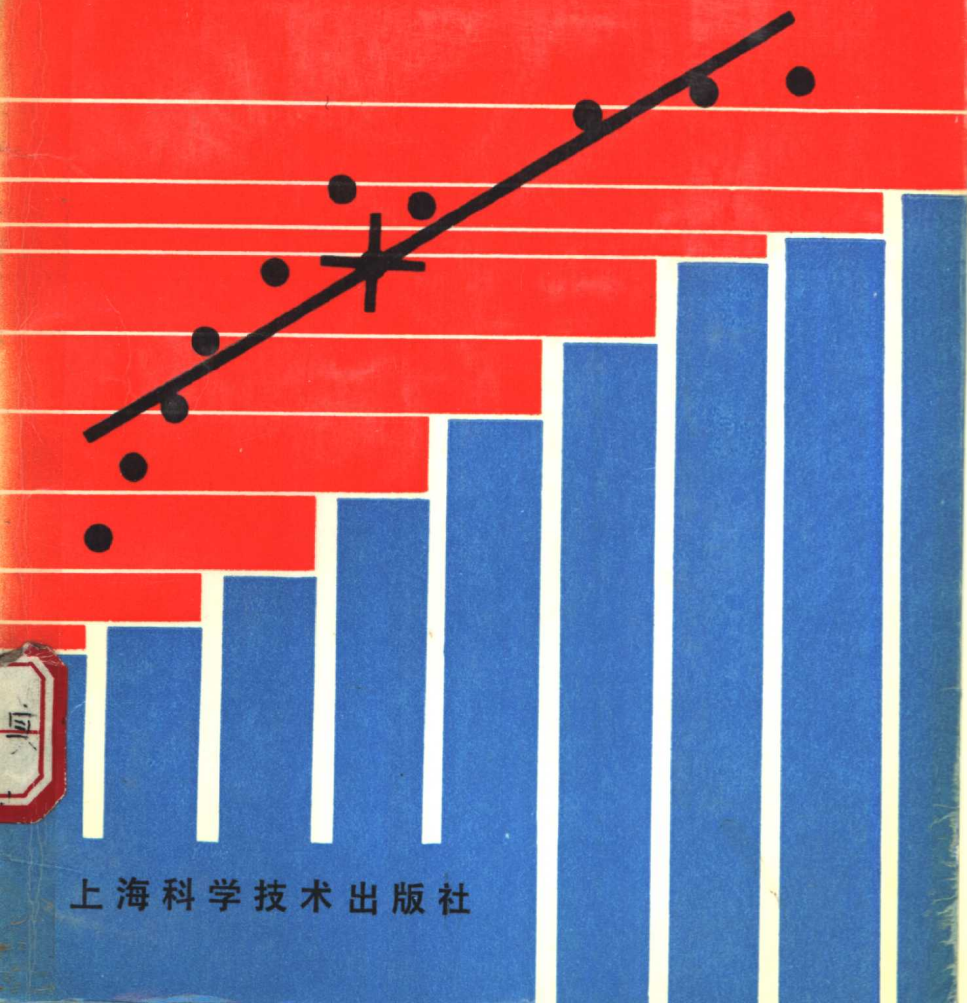


实用蚕桑生物统计学

钟伯雄 徐俊良 编著



上海科学技术出版社

实用蚕桑生物统计学

钟伯雄 徐俊良 编著

上海科学技术出版社

(沪)新登字108号

实用蚕桑生物统计学

钟伯雄 徐俊良 编著

上海科学技术出版社出版

(上海瑞金二路450号)

本书由上海发行所发行 浙江农业大学印刷厂印刷

开本787×1092 1/32 印张13.5 字数292000

1992年4月第1版 1992年4月第1次印刷

印数1—2500

ISBN 7-5323-2755-8/S.303

定价: 6.70元

内 容 提 要

本书共分十一章，在简要阐明蚕桑生物统计的基本概念、理论分布和抽样分布的基础上，着重介绍了蚕桑试验的常用设计、统计推断、大样本的统计分析、方差分析基础、常用试验设计的方差分析、简单相关分析、线性回归分析及品种区域化试验统计分析方法。本书与蚕桑教学、科研、生产密切联系，内容编排循序渐近、由浅入深，论述严谨，条理清楚，通俗易懂。可作为蚕桑专业本科生、专科生和研究生的教材，亦可供从事蚕桑教育、科研、生产技术部门人员阅读参考。

前 言

当前,蚕桑科学工作者已越来越多地采用生物统计分析方法解释所得的研究结果,以期得到合理的结论。为了适应这一形势的需要,笔者根据多年从事蚕桑生物统计学教学和科学研究工作的体会,编写了本书。

本书简要阐述了蚕桑生物统计的基本概念、原理、理论分布和抽样分布,着重介绍了蚕桑试验的常用设计方法,针对桑树试验比养蚕试验更难控制外部环境这一情况,所以较多地介绍了桑树试验设计的方法。本书还详细地介绍了统计推断、大样本的统计分析、方差分析基础、常用试验设计的方差分析、简单相关和线性回归分析。基于目前对品种比较试验采用生物统计分析的现象越来越普遍,特在最后一章详细介绍了品种区域化试验方差分析方法。此外,在撰写过程中参考了近年来国内外许多有关的文章和书籍,希望能较准确地介绍生物统计的基本原理和方法及其在蚕桑科研中的应用。为了使读者在实际工作中能举一反三,融会贯通,在内容的安排上,特别注意特殊和一般的关系,对基本概念和注意事项作了详细的介绍,同时辅以一些实例,以帮助理解、提高。

由于水平有限,书中难免有谬误和不足之处,敬请读者批评指正。

在本书编写出版过程中,浙江农业大学蚕学系和有关部门的领导 and 同志们给予了真诚的帮助和支持,并得到了陆星垣教授、刘权教授、胡秉民教授和张全德副教授的大力支持,在此一并表示衷心感谢。

编 著 者

1990年12月于浙江农业大学

目 录

第一章 绪论	(1)
第一节 蚕桑试验统计概述.....	(1)
第二节 常用统计术语.....	(1)
一、总体、样本和随机样本.....	(1)
二、变数、参数和统计数.....	(2)
三、处理因素、环境因素和水平.....	(3)
四、单因素试验、复因素试验和综合因素试验.....	(4)
五、单点和多点试验、一年和多年试验.....	(4)
六、处理效应、简单效应、主要效应和交互作用效应	(4)
七、系统误差、随机误差和过失误差.....	(7)
第三节 蚕桑试验的要求.....	(8)
一、试验的目的性.....	(8)
二、试验结果的可靠性.....	(9)
三、试验条件的代表性.....	(10)
四、试验结果的重复性.....	(10)
第二章 蚕桑试验设计	(11)
第一节 试验因素和水平的确定.....	(12)
一、因素数的确定.....	(12)
二、水平数的确定.....	(13)
三、唯处理因素差异原则的贯彻.....	(13)

第二节 试验误差的来源及其控制	(14)
一、试验材料的差异及其控制	(14)
二、操作和管理技术的差异及其控制	(14)
三、外界环境条件的差异及其控制	(15)
四、试验地的差异及其控制	(15)
第三节 试验设计的基本原则	(17)
一、重复	(17)
二、随机	(18)
三、局部控制	(20)
第四节 控制环境差异的小区技术	(21)
一、小区容量	(21)
二、小区形状	(23)
三、重复次数	(23)
四、区组设置	(25)
五、对照区	(26)
六、保护区	(27)
第五节 常用试验设计	(27)
一、完全随机设计	(27)
二、随机区组设计	(28)
三、拉丁方设计	(29)
四、裂区设计	(31)
五、正交试验设计	(36)
第三章 主要特征数的含义和计算方法	(40)
第一节 集中性特征数	(40)
一、算术平均数	(40)
二、几何平均数	(43)
三、中位数	(43)
四、众数	(43)
第二节 离散性特征数	(44)

一、极差.....	(44)
二、方差.....	(44)
三、标准差.....	(46)
四、变异系数.....	(46)
第三节 异常数据的取舍.....	(47)
第四章 理论分布和统计数分布	(49)
第一节 预备知识.....	(49)
一、随机事件、频率与概率.....	(49)
二、事件间的关系及其概率运算.....	(53)
第二节 理论分布.....	(56)
一、二项分布.....	(56)
二、普松分布.....	(59)
三、正态分布.....	(60)
第三节 统计数分布.....	(67)
一、随机抽样和无偏估计.....	(67)
二、样本平均数的分布.....	(71)
三、样本平均数的差数分布.....	(72)
四、学生氏 t 分布.....	(73)
五、 χ^2 (卡方) 分布.....	(76)
六、 F 分布.....	(79)
第五章 统计推断	(82)
第一节 统计假设测验.....	(82)
一、统计假设测验的步骤.....	(83)
二、统计假设测验的几何意义.....	(84)
三、两尾测验和一尾测验.....	(85)
四、统计假设测验的两种错误.....	(86)
第二节 单个样本平均数的假设测验.....	(89)
第三节 两个样本平均数的假设测验.....	(90)

一、成组比较资料假设测验方法	(91)
二、总体方差不相等时的成组比较	(94)
三、成对比较资料假设测验方法	(95)
四、成对比较的优点	(97)
第四节 参数的区间估计及其与假设测验的关系	
.....	(98)
一、总体平均数的区间估计	(99)
二、总体平均数差数的区间估计	(100)
三、一尾测验的总体平均数及其差数的区间估计	(101)
四、参数区间估计与假设测验显著性的关系	(102)
第五节 二项成数的统计推断	(102)
一、单个样本二项成数的统计推断	(103)
二、两个样本二项成数的统计推断	(105)
第六节 样本容量	(108)
一、单个样本平均数假设测验的样本容量	(109)
二、成组比较和成对比较假设测验的样本容量	(110)
三、二项成数假设测验的样本容量	(111)
第六章 次数资料的统计分析	(113)
第一节 次数资料及其卡方(χ^2)检验	(113)
第二节 适合性检验	(114)
一、 $k=2$ 的次 数 资 料	(115)
二、 $k \geq 3$ 的次 数 资 料	(116)
三、拟合性检验	(118)
第三节 独立性检验	(119)
一、 2×2 表的独立性测验	(119)
二、 $2 \times C$ 表的独立性测验	(121)
三、 $R \times C$ 表的独立性测验	(123)
第七章 大样本的统计分析方法	(126)

第一节 分布形态的描述	(126)
一、次数分布表描述	(126)
二、次数分布图描述	(130)
第二节 平均数和标准差的计算	(132)
第三节 正态分布的拟合性检验	(134)
一、偏斜系数和峰态系数法检验	(134)
二、皮尔逊 χ^2 检验	(136)
第四节 大样本估计总体的精确度	(138)
第八章 方差分析基础	(140)
第一节 方差分析的意义	(140)
第二节 单向分组资料(每组样本容量相等)的方差分析	(141)
第三节 多重比较	(146)
一、学生氏最小显著差数法(LSD法)	(147)
二、邓肯氏新复极差法(SSR法)	(149)
三、多重比较结果的表示方法	(151)
第四节 方差分析的线性模型和期望均方	(154)
一、固定模型	(155)
二、随机模型	(156)
三、混合模型	(156)
第五节 单向分组资料(每组样本容量不等)的方差分析	(157)
第六节 两向分组资料的方差分析	(160)
一、组内只有单个观测值资料的方差分析	(160)
二、组内设有重复观测值资料的方差分析	(166)
第七节 变量转换	(173)
一、方差分析的基本假定及其检验	(173)
二、数据变换	(181)

第九章 常用试验设计的方差分析	(193)
第一节 随机区组试验设计的方差分析	(193)
一、单因素随机区组试验设计的方差分析	(193)
二、二因素随机区组试验设计的方差分析	(197)
三、三因素随机区组试验设计的方差分析	(204)
第二节 拉丁方试验设计的方差分析	(212)
一、单个拉丁方试验设计的方差分析	(212)
二、复拉丁方试验设计的方差分析	(216)
第三节 裂区试验设计的方差分析	(222)
一、裂区设计的各种排列	(222)
二、二裂式裂区试验设计的方差分析	(224)
三、时间裂区试验设计的方差分析	(231)
四、裂区组随机区组试验设计的方差分析	(240)
五、裂区组拉丁方试验设计的方差分析	(250)
六、三裂式裂区试验设计的方差分析	(259)
第四节 正交试验设计的方差分析	(262)
一、正交试验设计结果分析的依据	(262)
二、试验结果的方差分析	(263)
三、正交试验设计分析的注意点	(278)
第十章 简单相关和线性回归分析	(280)
第一节 相关模型和回归模型的定义	(280)
第二节 相关分析	(282)
一、相关关系和散点图	(282)
二、线性相关模型和相关系数	(284)
三、相关系数的计算	(286)
四、相关系数的统计假设测验	(289)
五、相关系数的置信区间	(294)
六、两个样本相关系数的比较	(295)

第三节 线性回归分析	(297)
一、线性回归方程	(297)
二、回归方程的估计标准误差	(302)
三、线性回归模型	(305)
四、线性回归的统计假设测验	(306)
五、线性回归的置信区间	(309)
六、两个线性回归方程的比较	(312)
第四节 决定系数	(315)
第五节 相关与回归的关系及其应用注意点	(317)
一、相关与回归的关系	(317)
二、相关和回归的应用注意点	(319)
第十一章 品种区域化试验的方差分析	(322)
第一节 品种多点试验结果的统计分析	(323)
一、品种多点试验分析步骤	(323)
二、品种多点试验结果统计分析示例	(326)
第二节 品种多点多年(期)试验结果的统计分析	(338)
一、品种多点多年(期)试验分析步骤	(338)
二、品种多点多年(期)试验结果统计分析示例	(343)
附表	
表一 随机数字表	(355)
表二 标准拉丁方表	(358)
表三 常用正交表	(359)
表四 舍掉可疑数据的界限值表	(383)
表五 正态分布累积函数表	(384)
表六 正态离差 u 值表(两尾)	(388)
表七 t 分布的临界 t_{α} 值表	(389)
表八 χ^2 分布的临界 χ^2_{α} 值表	(390)

表九 F 分布的 $F_{0.05}$ (上)和 $F_{0.01}$ (下)值表.....	(392)
表十 邓肯氏新复极差测验5%和1% SSR值表.....	(404)
表十一 百分数 p 的反正弦 ($\sin^{-1}\sqrt{X}$) 转换表.....	(408)
表十二 相关系数 r 的临界 r_i 值表.....	(412)
表十三 r 值和 Z 值转换表.....	(413)
主要参考文献.....	(415)

第一章

绪 论

第一节 蚕桑试验统计概述

蚕桑试验统计,是利用数理统计的原理和方法,分析和解释蚕桑试验所得到的结果。具体的内容包括:①科学的试验设计;②数据的整理和描述;③由样本推断总体;④分析试验误差,得出真实的处理效应;⑤分析试验数据的相关关系。

蚕桑科学研究方法可分为两种:①纯调查研究,即在自然条件的状态下,考察生物自发现象和其变化规律;②试验,即在人为控制的条件下,有目的地考察生物对不同状态的反应或作用,以认识和揭示生物的内在规律。由于试验对条件的设置既是可调的,又是精密的,故能促使发生一些自然条件下不能发生的现象,得到一些具有普遍意义的结论,所以,试验是科学研究的主要方法。

第二节 常用统计术语

一、总体、样本和随机样本

总体是指具有相同性质的变量组成的集合,或者说是某

一变数的全部可能值的集合。总体可分为无限总体和有限总体两类。无限总体所包含的变量往往是无限的或可以设想为无限的；有限总体包含的变量是有限的，或可数的。如春蕾品种的产卵量这一总体，是指春蕾这一品种无数可能的产卵量的集合，所以是一无限总体；而1990年春期浙江农业大学蚕种场饲养的春蕾品种的产卵量，因指明了时间、地点，故包含的变量数是可数的，所以是有限总体。

试验研究的目的，在于认识总体的表现和规律，但由于总体很庞大，即使是有限总体，其包含的变量也常常很多，所以在实际研究中，几乎都是通过对总体的一个部分加以研究，以此来推断总体，这一个部分，就称为样本。为了使样本能无偏地代表总体，在取样时，必须保证使总体中的每一变量都有同等机会被取作样本，这样取得的样本，叫做随机样本。本书除特殊说明外，所有样本均指随机样本。

统计上，常常把变量数小于或等于30的样本称为小样本，而大于30的样本称为大样本。

二、变数、参数和统计数

试验所得数据都是有变异的，如同一匾的400条蚕所结的茧子，其全茧量各有差异。统计学将这种同一性质或特征的、具有变化的一组数据，叫做变数或随机变数，而每一个具体数值叫做变量或观测值。上述400个茧子的全茧量，就称为全茧量变数，而每一个数据则称为全茧量变量。

由总体的某一变数计算得到的某一数值，反映了总体的特征，如总体平均数，反映了总体的集中情况。这些特征数，称为参数。常用希腊字母表示。

由样本的某一变数计算得到的某一数值，反映了样本的

特征,如样本平均数,反映了样本的集中情况,称为统计数。常用拉丁字母表示。统计数是参数的估计值。

三、处理因素、环境因素和水平

每一个试验都有相应的研究目的,其相应的衡量指标称为试验指标。在各试验中,指标的选择是各不相同的。如蚕品种比较试验,选用的指标有收茧量、茧丝长、抗病性等;而桑品种比较试验,选用的指标则为产叶量、叶质、抗逆性等。影响试验指标的各种条件称为因素。一个试验指标所涉及的因素是很多的。如影响桑叶产量指标的因素,有自然因素、技术因素和社会因素等。自然因素方面有温、光、热等;技术因素方面有土、肥、水等;社会因素方面有体制、劳力、价格等。这些因素还可再细分为若干较小的因素,如光因素可再分为光强、光长、光质等;肥因素可再分为肥料种类、施肥数量、施肥时期、施肥方法等。这些影响桑叶产量的因素,总称为生产因素。

生产因素当然都能对试验指标有影响,但如对生产因素不加选择地进行全面研究,试验规模太大,现有的条件难以具体实施。因此,通常的做法总是在一个试验中只研究一部分生产因素,这部分因素就称为处理因素或试验因素。而那些暂不研究的生产因素,称为环境因素或试验背景。

显然,处理因素必须是:①可控的;②在数量或质量上具有不同的等级。处理因素的不同等级,称为水平。例如,桑园施氮肥试验,施肥量为20、25、30、35、40 kg。这里的处理因素就是施氮肥,而20、25、30、35、40 kg,则是施氮肥这个因素按数量水平划分的5个等级,即5个水平。又如,蚕品种比较试验,有A、B、C、D、E、F 6个品种,即蚕品种是处理因素,而A、B、C、D、E、F是品种这个因素按质量(即遗传背景)水

平划分的6个等级。

四、单因素试验、复因素试验和 综合因素试验

只研究一个处理因素的试验,称为单因素试验。因素内每一个水平就构成了一个试验处理,简称处理。研究二个及二个以上处理因素的试验,称为复因素试验。各因素的不同水平的组合构成了处理,或称处理组合。如有一个两因素的试验,A因素为蚕品种,分为3个水平,即 A_1 、 A_2 、 A_3 ,B因素为不同叶位的桑叶,也是3水平,即 B_1 、 B_2 、 B_3 ,这个试验的处理组合即为9个,具体是 A_1B_1 、 A_1B_2 、 A_1B_3 、 A_2B_1 、 A_2B_2 、 A_2B_3 、 A_3B_1 、 A_3B_2 、 A_3B_3 。

从复因素试验的处理组合中,按一定的规律挑选一定数量的处理组合加以实施,这种试验称为综合因素试验。如正交试验。

五、单点和多点试验、一年和多年试验

仅在一个地点进行的试验称为单点试验;同一试验同时在多个地点进行,称为多点试验。

仅在一年内进行的试验,称为一年试验;同一试验在多个年份反复实施,称为多年试验。

六、处理效应、简单效应、主要效应和 交互作用效应

在同等试验背景下,处理平均数的差数称为处理效应。如一个施氮肥试验,不施氮肥的桑叶平均产量 $\bar{x}_1 = 160\text{kg}$,施氮肥的桑叶平均产量 $\bar{x}_2 = 200\text{kg}$,因而施氮肥对不施氮肥的