

高等院校农业科学、动物科学、生命科学和林业科学专业教材

生物统计学

SHENGWU TONGJIXUE

孟宪勇 王 晶 刘 彭 主编



海洋出版社

高等院校农业科学、动物科学、生命科学和林业科学专业教材

生物统计学

SHENGWU TONGJIXUE

孟宪勇 王 晶 刘 彭 主编



海洋出版社

2016年·北京

内 容 简 介

生物统计学融合了概率论、生物学与数学等知识,是生命领域不同专业学生都应该掌握的重要工具之一。

全书共分为 10 章、2 个附录,主要内容包括统计学概念、试验设计、资料(数据)的描述性统计分析、概率论及常见分布、统计推断、非参数统计、方差分析、正交设计、直线相关与回归、多元线性回归与相关等。附录部分为常见统计软件简介和本书相关的各种统计数值表。

适用范围:高等院校农业科学、动物科学、生命科学和林业科学相关专业课教材。

图书在版编目(CIP)数据

生物统计学/孟宪勇,王晶,刘彭主编.--北京:海洋出版社,2016.6

ISBN 978-7-5027-9490-3

I. ①生… II. ①孟… ②王… ③刘… III. ①生物统计 IV. ①Q-332

中国版本图书馆 CIP 数据核字(2016)第 123410 号

总 策 划: 刘 斌

责任编辑: 刘 斌

责任校对: 肖新民

责任印制: 赵麟苏

排 版: 海洋计算机图书输出中心 晓阳

出版发行: **海洋出版社**

地 址: 北京市海淀区大慧寺路 8 号(707 房间)印
100081

经 销: 新华书店

技术支持: 010-62100059

发 行 部: (010) 62174379 (传真) (010) 62132549

(010) 62100075 (邮购) (010) 62173651

网 址: <http://www.oceanpress.com.cn/>

承 印: 北京朝阳印刷厂有限责任公司

版 次: 2016 年 6 月第 1 版

2016 年 6 月第 1 次印刷

开 本: 787mm×1092mm 1/16

张: 14.75

字 数: 354 千字

印 数: 1~3000 册

定 价: 35.00 元

本书如有印、装质量问题可与发行部调换

前 言

生物统计学是高等院校农业科学、动物科学、生命科学、林业科学等专业开设的核心课程之一,是生物科学家对生物实验进行科学分析所需掌握的重要数据分析工具。生物统计学是一门方法论科学,是指导学生进行科学实验并得出科学结论的有力武器。

为了满足教学需要,确保学生培养质量,国内已出版了数种版本、各种名称的优秀生物统计学教材。随着科技的发展和知识更新的加快,生物统计学教材也需要不断补充、更新和完善,出于此目的,我们重新编写了生物统计学教材,这也是我们教学实践经验的总结。

本教材系统地介绍了生物统计学的基本理论、方法及试验设计的基本原理。全书共分10章,包括绪论、试验设计、资料(数据)的描述性统计分析、概率论及常见分布、统计推断、非参数统计、方差分析、正交设计、直线相关与回归和多元线性回归与相关。通过学习,学生将能较好地掌握生物统计的基本思想和各种数据分析方法,提高分析问题的能力。

本教材的特色是内容全面、完整、有新意;叙述深入浅出,通俗易懂,注重思想,注重应用;在附录中介绍了常用统计软件,强调了统计软件在现代数据分析中的重要作用。

本教材是生物统计学的基础教程,可作为高等院校农业科学、动物科学、生命科学、林业科学等专业本科生教材使用,对广大实际工作者也极具参考价值。

本教材由山东农业大学孟宪勇、王晶、刘彭老师共同编写而成。各章执笔人分别为:孟宪勇(第1,2,7,8章及附录1)、王晶(第6,9,10章)、刘彭(第3,4,5章及附录2)。

在编写过程中,我们参阅了大量资料和著作,吸收了同行们辛勤劳动的成果,在此向他们表示衷心感谢。感谢海洋出版社为本书出版给予了大力协助。感谢山东农业大学信息科学与工程学院数学与信息系全体同仁的支持与帮助。

由于编者水平所限,不妥之处在所难免,敬请读者批评指正。

编者

2016年4月

目 录

第一章 绪论	1
第一节 统计学	1
第二节 生物统计学在科学实践中的地位	2
一、数据与生物试(实)验	2
二、生物统计学的地位	3
第三节 生物统计学的作用与学习方法	3
一、生物统计学的内容	3
二、生物统计学的作用	4
三、生物统计学的学习方法与要求	5
习 题	5
第二章 试验设计	6
第一节 试验设计概述	6
一、试验设计的意义、任务和作用	6
二、基本概念	7
第二节 试验计划和试验方案的设计	9
一、试验计划	9
二、试验方案	10
三、试验方案的拟订	11
第三节 试验误差及其控制	13
一、试验误差的概念和类型	13
二、试验误差的来源	13
三、试验误差的控制途径	14
第四节 试验的评价	14
一、试验计划的评价	14
二、试验结果和结论的评价	15
第五节 试验设计的基本原则	15
一、重复	15
二、随机化	16
三、局部控制	17
第六节 常用试验设计方法	18
一、完全随机设计	18

二、随机区组设计	19
三、拉丁方设计	22
四、系统分组设计	25
五、裂区设计	26
习 题	27
第三章 资料(数据)的描述性统计分析	28
第一节 试验资料(数据)的整理	28
一、资料(数据)的类型	28
二、随机变量(数)的类型	29
第二节 用图表描述资料	29
一、统计表	29
二、统计图	32
第三节 用统计量描述资料	37
一、位置的描述	37
二、变异的描述	40
三、分布形状的描述	43
习 题	44
第四章 概率论及常见分布	46
第一节 概率论基础知识	46
一、概念	46
二、事件及事件间的关系	47
三、概率的计算法则	47
四、概率分布	48
第二节 几种常见的理论分布	49
一、二项分布	50
二、泊松分布	50
三、正态分布(normal distribution)	51
第三节 抽样分布	54
习 题	58
第五章 统计推断	59
第一节 假设检验概述	59
一、假设检验的原理	59
二、假设检验的步骤	60
三、双侧检验与单侧检验	61
四、假设检验的两类错误	62
第二节 样本平均数的假设检验	63

一、单个样本平均数的假设检验	63
二、两个样本平均数的假设检验	64
第三节 样本频率的假设检验	68
一、单个样本频率的假设检验	69
二、两个样本频率的假设检验	69
第四节 样本方差的假设检验	71
一、单个样本方差的假设检验	71
二、两个样本方差的假设检验	72
三、多个样本方差的假设检验	73
第五节 参数估计	74
一、参数估计的原理	74
二、单个总体平均数的估计	75
三、两个总体平均数差数的估计	75
四、单个总体频率的估计	77
五、两个总体频率差数的估计	77
习 题	77
第六章 非参数统计	80
第一节 符号检验	80
一、单个样本的符号检验	80
二、两个样本的符号检验	81
三、需要说明的几个问题	83
第二节 秩和检验	85
一、成组数据比较的秩和检验	85
二、配对数据比较的秩和检验	87
三、需要说明的几个问题	88
第三节 χ^2 统计量及 χ^2 检验	89
一、 χ^2 检验的基本原理	90
二、 χ^2 检验中需要注意的问题	91
第四节 适合性检验	91
一、适合性检验的理论基础及主要步骤	91
二、离散型分布的适合性检验	92
三、连续型分布的适合性检验	94
第五节 独立性检验	95
一、独立性检验的基本原理及一般程序	96
二、 2×2 列联表的独立性检验	97
三、 $n \times m$ 列联表的独立性检验	99
四、两个需要说明的问题	100
习 题	101

第七章 方差分析	103
第一节 单因素方差分析	103
一、分析基本方法	104
二、多重比较	108
三、线性模型	111
四、数据变换	113
第二节 双因素无交互作用的方差分析	114
一、双因素无重复试验模型与统计假设	114
二、平方和分解	116
三、显著性检验	116
四、期望均方	117
五、多重比较	117
六、随机效应模型	119
第三节 双因素有交互作用的方差分析	119
一、双因素等重复试验的统计(数学)模型与统计假设	119
二、多重比较	123
三、混合随机效应模型	123
第四节 常用单因素试验设计结果的统计分析	124
一、随机区组设计单因素试验结果的方差分析	124
二、拉丁方设计的单因素试验结果的统计分析	125
第五节 常用两因素试验设计结果的统计分析	127
一、两因素随机区组试验结果的方差分析	127
二、两因素系统分组设计试验结果的统计分析	131
三、两因素裂区设计试验结果的统计分析	133
习 题	135
第八章 正交设计	140
第一节 正交设计试验	140
一、正交设计的基本思想	140
二、正交表	142
三、正交设计的基本步骤	143
第二节 正交设计试验结果的统计分析	147
一、直观分析法	147
二、方差分析法	149
习 题	152
第九章 直线相关与回归	154
第一节 回归与相关概述	154
一、回归与相关的概念	154

二、回归与相关的分类	155
三、回归与相关的作用	155
第二节 一元线性回归	156
一、一元线性回归方程的建立	156
二、一元线性回归方程的显著性检验	159
三、一元线性回归方程的应用	163
第三节 直线相关	166
一、相关关系与相关系数	166
二、相关系数的显著性检验	167
三、有关应用问题的讨论	169
第四节 能线性化的曲线回归	170
一、曲线回归分析概述	170
二、能直线化的曲线类型	172
三、曲线回归实例	174
习 题	178
第十章 多元线性回归与相关	180
第一节 多元线性回归	180
一、多元线性回归方程的建立	180
二、多元线性回归的统计推断	182
三、多元线性回归的区间估计及预测	186
第二节 多项式回归	189
一、多项式回归概述	189
二、应用实例	190
第三节 多元相关	191
一、复相关	191
二、偏相关	193
三、偏相关和简单相关的关系	195
习 题	195
附录 1 常用统计软件简介	197
附录 2 附表	202

第一章 绪论

第一节 统计学

每个人都离不开统计,了解一些统计学知识对每个人都是必要的。比如,在外出旅游时,你需要关心一段时间内的旅游信息;在投资股票时,你需要了解股票市场价格的信息,了解某只特定股票的有关财务信息;在观看 NBA 篮球赛时,除了关心进球数多少,你还想知道各球队的技术统计;等等。统计无时不在,无处不有。

统计学 (statistics) 在《大英百科全书》中的定义是:统计学是“用以收集数据、分析数据和由数据得出结论的一组概念、原则和方法”。更确切地说,统计学是“研究如何获取数据、如何分析数据、如何解释数据,从数据中提取信息,寻找规律性的科学”。统计学的研究对象是数据;研究任务是分析数据、提取信息,因此,有数据的地方就需要统计学。统计学的使用范围几乎覆盖了社会科学和自然科学的各个领域,甚至被用于工商业及政府的情报决策之上。统计学实际问题用到了大量的数学 (主要是数理统计学) 知识,当然也不能离开所论问题的专门知识。随着大数据 (Big Data) 时代的来临,统计学的面貌也逐渐改变,与信息、计算等领域密切结合,成为数据科学 (Data Science) 的重要主轴之一。

统计学是一门古老的科学,迄今已有 2300 多年的历史,统计学的实践活动则可追溯到更早。统计学的产生与发展是和生产的发展、社会的进步紧密相连的。20 世纪以前,描述性统计占主导地位。描述性统计就是收集大量数据,并进行一些简单的运算,譬如求平均值、百分比,或用图表、表格把它们表示出来。我国古代就有钱粮户口的统计,并在实践中产生了许多思想和方法,例如,春秋时期齐国在管仲的调查思想指导下,有了第一次经济普查,并且使用了经济折算概念。在西方,古希腊的亚里士多德撰写“城邦政情”,其内容包括各城邦的历史、行政、科学、艺术、人口、资源和财富等社会和经济状况的比较分析。“城邦政情”式的统计研究延续了近 2000 年,这些统计工作都与国家实施统治有关。自 17 世纪以来,西方一些著名学者的工作促进了统计方法、数学计算和逻辑推理的结合,分析社会经济问题的方式更加注重运用定量分析的方法,如高斯等人在误差方面的研究工作,正态分布也因此被称为高斯分布。与此同时,研究不确定性的数学分支——概率论也在这个时期蓬勃发展起来。19 世纪末 20 世纪初,出现了以概率论为基础的数理统计学,并成为各个学科的研究工具。我国在 2011 年由国务院学位委员会将统计学定为一级学科。

统计学的英文词 statistics 最早源于现代拉丁文 statisticum collegium (国会) 以及意大利文 statista (国民或政治家)。德文 Statistik, 即“统计学”,最早由 G. Achenwall 于 1749 年使用,代表对国家的数据资料进行分析的学问,也就是“研究国家的科学”。1787 年,英国学

者 E.A.W.Zimmerman 据语音把 “statistik” 译成英语 “statistic”。19 世纪, 统计学传到日本, 日本的学者将其译成 “统计学”。

统计方法通用于所有的科学领域, 而不是为某个特定的问题领域而构造的。统计学延伸到不同的领域, 形成了不同的分支方向, 如延伸到生命科学、医学、心理学、地理学等领域, 相应地形成生物统计学、医学统计学、心理统计学、空间统计学等。

将统计学应用于生物科学就称为生物统计学 (Biostatistics) 或生物计量学 (Biometrics), 包括科学研究设计, 资料的搜集、整理, 综合归纳、表达及分析等方面的内容。任何科学研究都离不开调查或试验, 进行调查或试验首先必须解决的问题是: 如何合理地进行调查或试验设计。生物统计学既是统计学的分支, 也是应用数学、数量生物学的分支, 主要是应用数理统计的原理和方法处理生物学中的各种数据 (数量资料), 从而透过现象揭示生物学本质的一门科学, 是科学研究与实践应用的基础工具。

第二节 生物统计学在科学实践中的地位

一、数据与生物试 (实) 验

数据是信息的载体。统计分析离不开数据, 没有数据统计方法就成了无米之炊。数据收集就是取得所需要的数据。数据的收集方法可分为两大类: 一是观察方法; 二是试验方法。观察方法是通过调查或观测而获得数据, 试验方法是在控制试验对象条件下通过试验获得数据。生物科学研究中的试验就是生物试验。

生物试验是为了提高生产力而进行的一种自觉的、有计划的科研实践。生物试验的首要任务在于解决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试验, 比如: ①对各种防治病害措施做出鉴定, 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。

生物试验也是解决生物科学研究问题的有效手段。这是由于生物试验并不完全依赖于生产。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条件, 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证实的。所以, 从根本上说, 生物科学的发展要以生产的发展为基础, 但是, 要使生物科学走在生产的前头, 却又非侧重于生物试验不可。

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

二、生物统计学的地位

在生物科学研究中，经常会遇到许多数量方面的问题。图 1-1 表示统计学在科学研究中涉及数据收集与分析的各个层面都发挥着重要作用，贯穿于最初提出问题直到最后得出结论的始终。此图区分了两种类型的研究：试验性的和调查性的。在试验性的研究中，各种因素经常受到控制并且固定在事先设定的水平上，在试验的每轮实施中保持不变。在调查性的研究中，有许多因素无法控制，但是，却可以对它们进行记录与分析。本书将侧重于试验性的研究，当然，书中介绍的许多分析方法也同样适用于调查性的研究。

上述两类研究中的关键环节就是数据。所有数据在收集过程中受到各种各样偶然因素的干扰，难免产生变异。这种变异可能产生于基因型之间固有的差异，或者是环境条件变化引起的随机差异，或者是测量仪器读数中的测量误差，或者是许多其他已知及未知因素影响的结果。

利用统计学方法进行试验设计，能够有效消除已知来源的偏差，控制未知来源的偏差，确保试验能够提供目标对象有关性状的精确数据，防止采用不经济的设计造成不必要的试验资源浪费。同样地，利用统计学方法进行试验数据分析，能够简要展示试验结果，获得关于生物学现象的推断性结论。

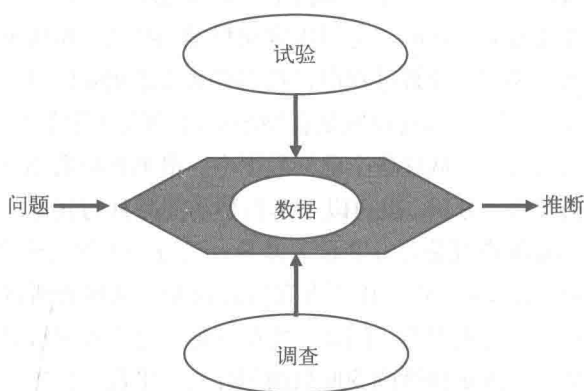


图 1-1 科学研究中统计学的关键作用环节

第三节 生物统计学的作用与学习方法

一、生物统计学的内容

生物统计学的内容包括试验设计和统计分析。试验设计是指应用数理统计的原理与方法，制订试验方案，选择试验材料，合理分组，降低试验误差，利用较少的人力、物力和时间，获得多而可靠的数据资料。

统计分析是指应用数理统计的原理与方法对数据资料进行分析与推断，认识客观事物的本质和规律性，使人们对所研究的资料得出合理的结论。这是生物统计的又一重要任务。由

于事物都是相互联系的,统计不能孤立地研究各种现象,而必须通过一定数量的观察,从这些观察结果中研究事物间的相互关系,揭示出事物客观存在的规律。

统计分析与试验设计是不可分割的两部分。试验设计需要以统计分析的原理和方法为基础,而正确设计的试验又为统计分析提供了丰富可靠的信息,两者紧密结合推断出合理的结论,不断地推动应用生物科学研究的发展。

二、生物统计学的作用

现代生物统计学已在科学研究和生产中得到极为广泛的应用。其作用主要体现在以下4个方面。

(1) 提供科学的试验设计方法。科学的试验设计可用较少的人力、物力和时间取得丰富的可靠的试验数据,尽量降低试验误差,从试验所得到的数据中能够无偏地估计处理效应和试验误差的估值,以便从中得出正确的结论。相反,设计不周,不仅不能得到正确的试验结果,而且还会带来经济上的损失。因此,在开展任何一项生物试验之前,都必须科学地进行试验设计,包括样本容量的确定、抽样方法的挑选、处理水平的选择、重复数的设置以及试验的安排等,都必须严格遵循“随机、局部控制、重复”这一试验三原则。

(2) 提供科学的试验分析方法。在生命科学试验过程中,常常可以获取大量的非常复杂的第一手资料,如何透过纷繁复杂的信息得出客观科学的结论,抓住蕴含在其中的生物学本质规律呢?在数据收集、整理、分析过程中,我们必须根据实际资料,选取一套科学而严密的生物统计学分析方法。例如,通过研究某新型小麦品种的产量特征,可以获得不同植株、不同地区、不同批次的结穗数。从这些杂乱的数据中,很难直接看出其规律性,如果采用生物统计学方法对其进行整理、分析,就可以了解新型小麦品种与其他小麦品种产量之间的关系,以及不同地区该小麦的产量是否存在显著差异,为进一步深入研究提供科学依据。

(3) 正确评价试验结果的可靠性。由于存在试验误差,从试验所得的数据资料必须借助于统计分析方法才能获得可靠的结论。例如,某农场要研究两种饲料对肉用仔鸡增重及饲料收益的影响。选择同品种及体重接近的500只肉用仔鸡,半数饲以甲种饲料,半数饲以乙种饲料,8周龄后称其体重并结算饲料消耗,分析比较这些资料,从中得出结论。这就要运用统计分析方法,以决定两群鸡体重及饲料消耗的差异,究竟属于本质原因造成的,抑或属于机遇造成的,即判断是由于不同饲料造成的,还是由于其他未经控制的偶然因素所引起的。分析之后才能做出比较正确的结论。

(4) 为学习其他课程提供基础。我们要学好遗传学、育种学等学科,就必须学好生物统计学。例如,数量遗传学就是应用生物统计方法研究数量性状遗传与变异规律的一门学科,如果不懂生物统计学,也就无法掌握遗传学。此外,阅读中外科技文献也常常会碰到统计分析问题,也必须要有生物统计的基础知识。因此,生物科学工作者都必须学习和掌握统计方法,才能正确认识客观事物存在的规律性,提高工作质量。

生物统计学在生物科学研究中虽然有着重要的作用,但却不是万能的。因为生物科学中很多现象受物理学、化学、生物学规律的支配。所以,生物科学研究工作,应该在生物学相关专业理论的指导下进行。

总之,生物统计学是一种很有用的工具,正确使用这一工具可以使科学研究更加有效,使生产效益更高,它是每位从事生物科学工作者都必须掌握的基础知识。

三、生物统计学的学习方法与要求

生物统计学是数学与生物学相结合的一门交叉学科,所包含的公式很多,在学习中,首先,要弄懂统计的基本原理和基本公式。要理解每一个公式的含义和应用条件,可以不必深究其数学推导。其次,要认真地做好习题作业,加深对公式及统计步骤的理解,能熟练地应用统计方法。第三,应注意培养科学的统计思维方法。生物统计意味着一种新的思考方法——从不确定性或概率的角度来思考问题和分析科学试验的结果。第四,必须联系实际,结合专业,了解统计方法的实际应用;平时要留意书籍和杂志中的表格、数据及其分析和解释,以熟悉表达方法及应用。

习 题

1. 统计学的基本内容是什么?生物统计学的基本内容是什么?统计学与生物统计学关系如何?
2. 生物统计学的功能有哪些?
3. 试从生物科学研究的过程看生物统计学在科学实践中的地位。
4. 生物统计学的作用有哪些?

第二章 试验设计

不少人认为统计的主要任务是数据处理和分析。实际上,统计的第一步便是试验设计(有时要通过观察研究,但其思想依然是试验设计的思想)。

试验设计(Design Of Experiment, DOE)是统计学的一个分支,是进行科学研究的重要工具,它包含有丰富的内容,其数据分析方法主要是方差分析。试验设计是由英国统计学家费歇(R.A.Fisher)于20世纪20年代为满足农业科学试验的需要而提出的。试验设计的主要目的是在尽可能少的人力、物力、时间和经费的条件下通过试验获得满足一定要求的数据,达到试验目的。也就是说,通过合理科学的设计,一方面提高试验效率,一方面保证数据质量。

第一节 试验设计概述

一、试验设计的意义、任务和作用

试验在生产和科学研究中具有重要地位,如何做试验,这里面大有学问。试验工作做得好,试验次数不多,就能达到预期目的;试验工作做得不好,就会事倍功半,甚至劳而无功。如果要最有效地进行科学试验,必须用科学方法来设计。

试验设计,广义理解是指试验研究课题设计,也就是整个试验计划的拟定。主要包括课题的名称,试验目的,研究依据、内容及预期达到的效果,试验方案,试验单位的选取、重复数的确定、试验单位的分组,试验的记录项目和要求,试验结果的分析方法,经济效益或社会效益估计,已具备的条件,需要购置的仪器设备,参加研究人员的分工,试验时间、地点、进度安排和经费预算,成果鉴定,学术论文撰写等内容。而狭义的理解是指试验单位的选取、重复数目的确定及试验单位的分组,也就是收集样本数据的计划。生物统计学中的试验设计主要指狭义的试验设计。

“凡事预则立,不预则废”。试验设计的任务是在研究工作进行之前,根据研究项目的需要,应用数理统计原理,作出周密安排,力求用较少的人力、物力和时间,最大限度地获得丰富而可靠的试验数据,通过分析得出正确的结论,明确回答研究项目所提出的问题。如果设计不合理,不仅达不到试验的目的,甚至会导致整个试验失败。因此,能否合理地进行试验设计,关系到科研工作的成败。

试验设计在科学研究实践中能发挥重要作用,主要有:

- (1) 大大缩短试验周期,降低试验成本,能用尽可能少的试验获得尽可能多的信息。
- (2) 可以正确估计、预测和有效控制、降低试验误差,从而提高试验的精度。
- (3) 可以分析试验因素对试验指标影响的规律性,找出主要因素,抓住主要矛盾。

(4) 可以了解试验因素之间相互影响的状况。

(5) 可以较快地找出优化的生产条件或工艺条件, 确定优化方案。

(6) 通过对试验结果的分析, 可以明确为寻找更优生产或工艺条件、深入揭示事物内在规律而进一步研究的方向。

二、基本概念

1. 试验指标 (experimental index)

在试验设计中根据研究的目的而选定的用来衡量或考核试验效果的质量指标, 也叫观察项目、响应变量 (response variable) 或输出变量。试验指标作为试验研究过程中的因变量, 常作为试验结果特征的一种判据, 例如, 可以把株高作为判断灌水有无促进植物生长的依据, 株高就是反映灌水作用的指标。试验观察指标要选准。因为它关系到试验结果能否回答所研究的问题。如对几个甘薯品种作比较的试验来说, 用鲜重作为产量指标就不妥当, 因为不同的品种含水量会不同, 如果以干物重或切干率作指标就较合适。生物体有多种性状和特征, 为了从多方面说明问题, 观察指标不应该只选一个就满足, 但也不宜选得太多。究竟应当选哪些指标? 这要从它对回答所研究的问题有无用处来考虑, 即所谓实用性。例如, 株高和植株干重可以作为表达植物生长的指标, 但是在研究作物品种的抗虫性上, 这两个指标毫无用处, 而幼苗有无茸毛则是研究品种抗虫性的十分有用的指标。

从观察对象 (性状或特征) 的性质上说, 指标可分为定性指标和定量指标两类。定性指标显示观察对象的属性, 即质的规定, 如施药后个体反映出有效与无效、受害与未受害、死亡与存活等; 定量指标则显示观察对象的量, 即量的规定, 如产量、株高、茎粗、土壤容量、含水量等。

从定性指标观察所获得的资料叫定性资料; 从定量指标观察所获得的资料叫定量资料。定性资料大都是以观察对象中出现或不出现某一属性的比率来表达的。它没有计量单位, 因为属性一般不能度量, 只能一个一个地计数。例如, 对一些植株清点出有病的有多少株, 无病的又有多少株, 然后计算出有病株的株数占观察的所有株数的百分比。定性资料通过数出出现某一属性的个体数而取得, 因而常常叫计数资料。定量资料是用测量指标所得的数量来表达, 因此又叫计量资料或测量资料。它有计量单位, 如度量衡或时间等单位。这是它和定性资料在表达方式上的不同点。一般说来, 计量资料要比计数资料精确, 可从中得出较为有把握的结论。而且由于它比较精确, 即在多次重复测量所得结果之间的变异较小, 而且在抽样观察中所抽取的样本含量可以较少。因此试验中最好选用可获得计量资料的指标。

在一个试验中对同一项指标的观察标准要统一。用文字叙述观察的结果要简明扼要。同时, 一个试验中既可以选用单指标, 也可以选用多指标, 这由专业知识对试验的要求决定。如农作物品种比较试验中, 衡量品种的优劣、适用或不适用, 围绕育种目标需要考察生育期、丰产性、抗虫性、耐逆性等多种指标。当然一般田间试验中最主要的常常是产量这个指标。各种专业领域的研究对象不同, 试验指标各异。在设计试验时要合理地选用试验指标, 它决定了观察记载的工作量。过简则难以全面准确地评价试验结果; 过繁又增加许多不必要的浪费。试验指标较多时还要分清主次, 以便抓住主要方面。

2. 试验因素 (experimental factor)

试验中凡对试验指标可能产生影响的原因或要素都称为因素,也称为因子(factor)、影响因子或输入变量。因素就是生产和科学研究过程中的自变量,常常是造成试验指标按某种规律变化的原因。如猪的每日增重量受饲料的配方、猪的品种、饲养方式、环境温湿度等诸方面的影响,这些都是影响猪的日增重的因素。它们有的是连续变化的定量因素,有的是离散状态的定性因素。

由于客观原因的限制,一次试验中不可能将每个因素都考虑进去。我们把试验中所研究的影响试验指标的因素称为试验因素,把除试验因素外其他所有对试验指标有影响的因素称为条件因素,又称实验条件(experimental conditions)。如在研究增稠剂用量、pH值和杀菌温度对豆奶稳定性的影响时,增稠剂、pH值和杀菌温度就是试验因素。除这3个因素外的其他所有影响豆奶稳定性的因素都是条件因素。它们一起构成了本试验的试验条件。试验因素常用大写字母A, B, C, ……表示。

3. 因素水平 (level of factor)

试验因素所处的某种特定状态或数量等级称为因素水平,简称水平。因素的水平可以是定性的,如不同品种,具有质的区别,称为质量水平;也可以是定量的,如喷施生长素的不同浓度,具有量的差异,称为数量水平。数量水平不同级别间的差异可以等间距,也可以不等间距。因素水平用代表该因素的字母加添足标1, 2, ……来表示。如 $A_1, A_2, \dots, B_1, B_2, \dots$ 。

4. 试验处理 (experimental treatment)

事先设计好的实施在试验单位上的一种具体措施或项目称为试验处理,简称处理。在单因素试验中,实施在试验单位上的具体项目就是试验因素的某一水平。如进行饲料的比较试验时,实施在试验单位(某种畜禽)上的具体项目就是喂饲某一种饲料。所以进行单因素试验时,试验因素的一个水平就是一个处理。在多因素试验中,实施在试验单位上的具体项目是各因素的某一水平组合。例如,进行3种饲料和3个品种对猪日增重影响的两因素试验,整个试验共有 $9(3 \times 3=9)$ 个水平组合,实施在试验单位(试验猪)上的具体项目就是某品种与某种饲料的结合。所以,在多因素试验时,试验因素的一个水平组合就是一个处理。事实上,在数理统计中,一个处理被称为一个总体。实施在试验单位上的具体项目就是样本。

5. 试验单位 (experimental unit)

在试验中能接受不同试验处理的独立的试验载体叫试验单位,也称为试验单元。它是试验中实施试验处理的基本对象。如在田间试验中的试验小区;在生物、医学实验中的小白鼠、医院病人等;在畜禽、水产试验中,一只家禽、一头家畜、一只小白鼠、一尾鱼等。试验单位往往也是观测数据的单位。

6. 全面试验(complete experiment)

试验中,对所选取的试验因素的所有水平组合全部给予实施的试验称为全面试验。全面试验的优点是能够获得全面的实验信息,无一遗漏,各因素及各级交互作用对试验指标的影响剖析得比较清楚,因此又称为全面析因实验(factorial experiments),亦称全面实施。但是,当试验因素和水平较多时,试验处理的数目会急剧增加,因而试验次数也会急剧增加。当实验还要设置重复时,试验规模就非常庞大,以致在实际中难以实施。如3因素,每个因