

# 田间试验与统计分析

——全国高等农林专科统编教材

● 朱明哲 主编

● 农学专业用

● 农业出版社



全国高等农林专科统编教材

# 田间试验及统计分析

朱明哲 主编

农学专业用

农 业 出 版 社



(京)新登字060号

封面设计 蒋志南

全国高等农林专科统编教材

**田间试验及统计分析**

朱明哲 主编

\* \* \*

责任编辑 徐建华

农业出版社出版 (北京市朝阳区农展馆北路2号)

新华书店北京发行所发行 兰州新华印刷厂印刷

787×1092mm 16 开本 15.125印张 329千字

1992年5月第1版 1992年5月兰州第1次印刷

600册 定价 4.00 元

109-02154-8/S·1418

主 编 朱明哲（河南技术师范学院）  
副主编 陆绍椿（莱阳农学院）  
编写者 张全德（浙江农业大学）  
王福亭（河南农业大学）  
周广余（熊岳农业专科学校）  
崔秀珍（河南技术师范学院）  
审稿人 范 谦（主审，河南农业大学）  
高明尉（浙江农业大学）

## 前 言

本书是全国高等农林专科学校农学专业的专业基础课教材，本教材的编写是以专科教育的培养目标、业务规格和农学专业的教学计划为依据来确定编写的指导思想和处理教材的广度的。在编写原则上，着重加强三基本：“基本理论、基本知识、基本技能”，保证三性：“科学性、针对性、实践性”，突出三面向：“面向现代化，面向世界，面向未来”。

根据教学计划规定，针对课程的基本要求，本着降低理论深度，加强适用性，将教材划分为10章，其中前8章为必修内容（按60学时编写）。后两章主要考虑当前农业科学技术的发展，对试验资料的统计分析水平的要求日益提高的需要，增加曲线回归、多元回归与通径分析（按14学时编写）作为选修内容，可供有条件的学校开设专业补充课时选用。

精选教材内容、突出专科特色，如波松分布、拉丁方、三因素随机区组、再裂区试验设计、协方差分析及田间抽样等内容，不适于专科讲授，未编入教材；有关数学模型部分，亦很少在教材中出现。此外将田间试验设计与方差分析合在一章编写，不另分立章节，可免内容重复，增强系统性，有利于章节间的有机衔接。由相关表计算回归和相关内容，予以删节，另增添有重复观察值的回归分析。对理论深、难度大，而又必不可少的内容，如概率理论，只着重介绍其基本知识，对理论分布与统计数分布及假设测验章节，尽量从基本理论、基本概念上下功夫，采取从具体到抽象，从现象到理论的编写手法，反复推敲，删繁就简，力求做到深入浅出，化难为易，对抽象而又重要的概念，尽可能利用直观的信息结构图象表示。为使学生切实掌握基本概念和计算技能，力求联系实际应用去阐明统计方法的基本原理，有助于学生从具体到抽象逻辑思维的发展，达到学以致用目的。

总之，在内容安排上注意抓住主线，削枝强干，突出重点，以精求质，特别强调着重于统计方法的应用，不计较数学公式的推导，力求由浅入深，通俗易懂。

每章内容后均附有习题，它对熟悉和掌握本门课程是非常必要的。

考虑当前农业科学试验中运用统计方法的要求与日俱增，广大农业科技人员迫切希望有一本深入浅出、具体实用的《田间试验及统计分析》书籍，为此对本教材的编写，立足于既能满足高等农林专科农学专业学生的需要，也要照顾社会教育、职业教育及广大农业科技人员的需要。由于从事农业试验的科技人员，数学基础一般不如理工科方面的同志，因此从教材的适用性与适应性考虑，应着眼于具有高中水平的数学基础和一定的农学知识的读者。

在本书辑成过程中，陈付贵、胡瑞法同志协助计算和核对，在此表示衷心感谢。

由于编写人员的业务水平有限，加以编写时间仓促，错误及欠妥之处在所难免，恳请广大师生和各方面的读者不吝指教，以便再次修订时改进。

编 者

1990年4月

# 目 录

|                                 |    |
|---------------------------------|----|
| <b>第一章 试验数据的收集、整理与特征数</b> ..... | 1  |
| 第一节 数据的收集 .....                 | 1  |
| 一、总体与样本 .....                   | 1  |
| 二、数据收集的途径 .....                 | 1  |
| 第二节 资料的整理 .....                 | 3  |
| 一、试验资料的性质 .....                 | 3  |
| 二、次数分布表 .....                   | 3  |
| 三、次数分布图 .....                   | 6  |
| 第三节 资料的特征数 .....                | 7  |
| 一、平均数 .....                     | 7  |
| 二、变异数 .....                     | 10 |
| 习题 .....                        | 12 |
| <b>第二章 概率和概率分布</b> .....        | 14 |
| 第一节 事件和概率 .....                 | 14 |
| 一、事件和概率的概念 .....                | 14 |
| 二、计算概率的基本法则 .....               | 16 |
| 第二节 二项分布 .....                  | 18 |
| 一、二项总体和二项分布的概念 .....            | 18 |
| 二、二项分布概率的计算 .....               | 19 |
| 三、二项分布的形状和参数 .....              | 20 |
| 第三节 正态分布 .....                  | 22 |
| 一、正态分布的概念 .....                 | 22 |
| 二、正态分布曲线的特性 .....               | 23 |
| 三、正态分布的标准化 .....                | 25 |
| 四、正态分布的概率计算 .....               | 26 |
| 第四节 抽样分布 .....                  | 29 |
| 一、随机抽样和无偏估计 .....               | 29 |
| 二、样本平均数的抽样分布 .....              | 31 |
| 三、样本平均数差数的抽样分布 .....            | 34 |
| 四、 $t$ 分布 .....                 | 36 |
| 习题 .....                        | 37 |
| <b>第三章 总体平均数的假设测验</b> .....     | 38 |
| 第一节 假设测验的基本原理 .....             | 38 |
| 一、假设测验的意义 .....                 | 38 |
| 二、假设测验的推理步骤 .....               | 38 |

|                                   |    |
|-----------------------------------|----|
| 三、两尾测验和一尾测验 .....                 | 40 |
| 第二节 单个平均数的假设测验和区间估计 .....         | 42 |
| 一、单个平均数的假设测验 .....                | 42 |
| 二、总体平均数的区间估计 .....                | 44 |
| 三、假设测验与区间估计的关系 .....              | 46 |
| 第三节 两个平均数的假设测验与区间估计 .....         | 47 |
| 一、成组数据的比较 .....                   | 47 |
| 二、成对数据的比较 .....                   | 53 |
| 第四节 百分数的假设测验与区间估计 .....           | 56 |
| 一、单个样本百分数的假设测验和区间估计 .....         | 56 |
| 二、两个样本百分数的假设测验和区间估计 .....         | 59 |
| 习题 .....                          | 62 |
| 第四章 卡平方 ( $\chi^2$ ) 测验 .....     | 64 |
| 第一节 卡平方 ( $\chi^2$ ) 的概念和测验 ..... | 64 |
| 一、卡平方 ( $\chi^2$ ) 的概念 .....      | 64 |
| 二、卡平方 ( $\chi^2$ ) 的显著性测验 .....   | 65 |
| 三、卡平方 ( $\chi^2$ ) 的连续性矫正 .....   | 65 |
| 第二节 适合性测验 .....                   | 66 |
| 第三节 独立性测验 .....                   | 69 |
| 一、 $2 \times 2$ 表的独立性测验 .....     | 69 |
| 二、 $2 \times c$ 表的独立性测验 .....     | 71 |
| 三、 $r \times c$ 表的独立性测验 .....     | 73 |
| 第四节 方差的同质性测验 .....                | 74 |
| 一、单个样本方差的假设测验 .....               | 75 |
| 二、两个样本方差的同质性测验 .....              | 76 |
| 三、多个样本方差的同质性测验 .....              | 77 |
| 习题 .....                          | 78 |
| 第五章 田间试验的基本原理和方法 .....            | 80 |
| 第一节 田间试验的意义、种类及基本要求 .....         | 80 |
| 一、田间试验的意义 .....                   | 80 |
| 二、田间试验的种类 .....                   | 81 |
| 三、田间试验的基本要求 .....                 | 82 |
| 第二节 田间试验的误差及其控制途径 .....           | 82 |
| 一、田间试验的误差 .....                   | 82 |
| 二、控制误差的途径 .....                   | 83 |
| 第三节 田间试验的方法步骤 .....               | 87 |
| 一、田间试验计划的制订 .....                 | 87 |
| 二、试验地的准备和田间区划 .....               | 88 |
| 三、种子准备 .....                      | 88 |
| 四、播种 .....                        | 89 |
| 五、栽培管理 .....                      | 89 |

|                                         |     |
|-----------------------------------------|-----|
| 六、田间试验的观察记载和测定 .....                    | 89  |
| 七、收获及脱粒 .....                           | 90  |
| 习题 .....                                | 91  |
| <b>第六章 单因素试验设计与统计分析</b> .....           | 92  |
| 第一节 单因素完全随机设计与方差分析 .....                | 92  |
| 一、完全随机试验的意义和优缺点 .....                   | 92  |
| 二、完全随机设计试验结果的方差分析 .....                 | 92  |
| 第二节 完全随机区组设计与方差分析 .....                 | 102 |
| 一、完全随机区组的设计方法 .....                     | 102 |
| 二、完全随机区组设计试验结果的方差分析 .....               | 102 |
| 三、完全随机区组试验的缺值估计 .....                   | 106 |
| 第三节 对比法和间比法试验设计及其统计分析 .....             | 107 |
| 一、对比法试验设计及其统计分析 .....                   | 107 |
| 二、间比法试验设计及其统计分析 .....                   | 108 |
| 习题 .....                                | 110 |
| <b>第七章 多因素试验设计与统计分析</b> .....           | 112 |
| 第一节 两因素随机区组试验设计与方差分析 .....              | 112 |
| 一、两因素随机区组试验设计方法 .....                   | 112 |
| 二、两因素随机区组试验结果的方差分析 .....                | 113 |
| 第二节 裂区试验设计与方差分析 .....                   | 119 |
| 一、裂区试验的设计 .....                         | 119 |
| 二、裂区试验的方差分析 .....                       | 120 |
| 第三节 条区试验设计与方差分析 .....                   | 126 |
| 一、条区试验设计 .....                          | 126 |
| 二、条区试验的方差分析 .....                       | 127 |
| 第四节 一年多点随机区组设计的综合分析 .....               | 133 |
| 一、试验误差方差的同质性测验 .....                    | 134 |
| 二、各试点误差均方同质性测验 .....                    | 135 |
| 三、多点品种区试的联合分析 .....                     | 136 |
| 习题 .....                                | 139 |
| <b>第八章 直线回归和相关</b> .....                | 142 |
| 第一节 回归相关的概念 .....                       | 142 |
| 第二节 直线回归方程 .....                        | 143 |
| 一、回归关系的散点图 .....                        | 143 |
| 二、配合直线回归方程的理论依据 .....                   | 143 |
| 三、直线回归方程计算实例 .....                      | 145 |
| 第三节 直线回归的显著性测验 .....                    | 147 |
| 一、 $F$ 测验 .....                         | 147 |
| 二、 $t$ 测验 .....                         | 150 |
| 第四节 直线回归方程预测值的区间估计 .....                | 151 |
| 一、 $y$ 的总体平均值 $\mu_{y_x}$ 的置信区间估计 ..... | 151 |



|                                     |     |
|-------------------------------------|-----|
| 二、总体观察值单个 $y$ 的预测区间估计               | 153 |
| 三、 $\mu_{y_x}$ 的置信区间和单个 $y$ 的预测区间图示 | 154 |
| 第五节 有重复观察值的直线回归分析                   | 155 |
| 一、建立直线回归方程                          | 158 |
| 二、回归方程的显著性测验                        | 159 |
| 第六节 两个直线回归方程比较的差异显著性测验              | 160 |
| 第七节 直线相关                            | 162 |
| 一、相关系数的定义                           | 162 |
| 二、相关系数的基本公式                         | 162 |
| 三、相关系数的特性                           | 163 |
| 四、相关系数的计算                           | 164 |
| 五、相关系数的显著性测验                        | 164 |
| 六、两个相关系数比较时的差异显著性测验                 | 165 |
| 七、相关和回归的关系                          | 167 |
| 习题                                  | 168 |
| 第九章 曲线回归                            | 170 |
| 第一节 可线性化的曲线回归                       | 170 |
| 一、对数函数曲线                            | 170 |
| 二、指数函数曲线                            | 172 |
| 三、幂函数曲线                             | 175 |
| 四、Logistic生长曲线                      | 178 |
| 第二节 多项式曲线                           | 182 |
| 一、二次多项式回归方程的建立                      | 183 |
| 二、二次多项式回归方程的显著性测验                   | 185 |
| 三、绘二次多项式回归图示                        | 185 |
| 习题                                  | 185 |
| 第十章 多元线性回归                          | 188 |
| 第一节 多元线性回归方程的建立                     | 188 |
| 一、多元线性回归方程                          | 188 |
| 二、多元线性回归正规方程组及其求解方法                 | 188 |
| 三、多元线性回归实例                          | 193 |
| 四、多元线性回归方程的估计标准误                    | 198 |
| 第二节 多元线性回归的假设测验                     | 199 |
| 一、多元线性回归关系的假设测验                     | 199 |
| 二、偏回归系数的假设测验                        | 200 |
| 三、自变数的重要性的取舍                        | 202 |
| 第三节 通径分析                            | 207 |
| 一、通径系数的基本概念                         | 208 |
| 二、通径系数的意义和应用                        | 209 |
| 三、通径分析的步骤和方法                        | 209 |
| 习题                                  | 213 |
| 主要参考文献                              | 215 |

|                               |     |
|-------------------------------|-----|
| 附表                            | 216 |
| 附表 1 10000个随机数字               | 216 |
| 附表 2 累积正态分布 $F_N(X)$ 值表       | 220 |
| 附表 3 正态离差 $u$ 值表(两尾)          | 222 |
| 附表 4 学生氏 $t$ 值表(两尾)           | 223 |
| 附表 5 $\chi^2$ 值表(一尾)          | 224 |
| 附表 6 5%(上)和1%(下)点 $F$ 值(一尾)表  | 225 |
| 附表 7 Duncan's新复极差测验5%和1%SSR值表 | 231 |

# 第一章 试验数据的收集、整理与特征数

农业试验工作者在对农业试验中某种性状表现进行统计分析之前，首先要针对试验的全群合理地收集资料，并将收集到的资料加以整理，算出某些重要的特征数字，以作为进一步统计分析并据以作出科学结论的基础。

## 第一节 数据的收集

在试验中，如果我们要对研究对象的所有个体逐个进行观测是不可能的，因为全群的个体往往很多，全部研究要花费很多人力物力，所以是不经济的，也是不必要的。通常可以在全群中抽取若干个体作为代表来研究，这就是抽样调查。

### 一、总体与样本

具有相同性质的个体所组成的集团称为总体。确切地说，是根据研究目的确定的、符合指定条件的全部观测对象称为总体。组成总体的每一个基本单元称为总体单位，总体又分无限总体与有限总体两类：无限总体包含的单位数是无穷多的，例如鲁玉4号玉米在生产上推广种植了一千万亩，假如我们要研究鲁玉4号玉米总体的株高，那么这一千万亩鲁玉4号的所有植株的株高就构成了一个株高的无限总体；有限总体包含的单位数则是有限的，例如鲁玉4号玉米种植在一个试验小区中，虽然有几百株，但与无限总体相比，则是极有限的，它们成为一个小集团，故称有限总体。无论是无限总体或有限总体，其包含的单位数目称为总体容量，均用 $N$ 表示之。

总体是我们研究的对象，但通常只能从总体中抽出一部分总体单位来研究，以推断总体的全貌，这些被选为代表部分的观测值叫作样本，而抽取样本的过程称为抽样。样本中包含的单位数称为样本容量，例如从总体中抽取20个个体构成一个样本，则其样本容量为20。一般可用 $n$ 表示样本容量，当 $n$ 较大时称大样本， $n$ 较小时称小样本，在农业试验中，常以30作为样本大小的分界线，即 $n \geq 30$ 为大样本， $n < 30$ 为小样本，但这并不是绝对的，应视具体的研究对象而异。样本容量 $n$ 与有限总体容量 $N$ 的比值 $n/N$ 称为抽样分数，这在抽样时有用。

样本的概念还应包括试验结果所取得的数据，例如一个小麦比较试验，每个品种种4个小区，这4小区的产量数据，实际上就是这个品种产量的一个样本，其样本容量 $n$ 等于4。

由总体全部个体所得观测值算得的总体特征数称为参数，用希腊字母表示之，如总体的平均数为 $\mu$ ，总体的标准差为 $\sigma$ 。从样本的各个体算得的特征数称为统计数，用拉丁字母表示，如样本平均数为 $\bar{x}$ 、样本标准差为 $s$ 。

### 二、数据收集的途径

数据的收集，一般通过试验与调查两个途径：

(一) 调查 调查有两种,一种是普查,另一种是抽样调查,在农业科学研究中,普查虽然有其重要意义,但大量的调查工作还是抽样调查,而生物统计所涉及的主要内容,正好是有关试验和抽样调查的原理和方法。

抽样调查是一种非全面调查,它只是根据一定的原则,对整个研究总体抽取一部分个体作为样本进行观测和统计处理,而后根据分析结果对总体作出推断和估计。实践证明,正确的抽样调查,不仅可以节省人力、物力,而在与统计方法结合的情况下,可以作出相当准确的推断。

大多数农业科学涉及的是有限总体的抽样调查,一些理论性无限总体的抽样调查,可以通过设计试验收集数据。

从有限总体抽样,一般有机械抽样、典型抽样和随机抽样三种方法。

1.机械抽样 也叫系统抽样或顺序抽样,它的特点是:在总体中每隔若干个单位抽取一个样本单位。例如在一块小麦田中纵横每隔若干距离抽取一个  $1\text{ m}^2$  的面积作为样块,以便根据样块产量来估测麦田亩产量,这种方法就叫机械抽样法,比较简单易行,但是如果各总体单位的排列有某种周期性或某种不明的趋向性变化时,样本就可能出现偏差。所以只有在各总体单位的排列是随机性时,方可用机械抽样抽取样本,一般田间试验中所用的对角线法、三点或五点取样法等,都属于机械抽样法。

2.典型抽样 也称经验抽样,主要根据调查工作者的经验和判断能力,有目的地在总体中挑选有代表性的典型单位或单位群,由于抽样时未考虑到随机化,所以其调查结果无法估算抽样误差。

3.随机抽样 也称概率抽样,这种抽样法最为常用,其特点是总体中每个总体单位都有同等的机会被抽取为样本单位,用这种方法抽得的样本叫随机样本。抽样的均等性常用抽签法或查随机数字表法来保证。

抽签法是将有限总体的每一个个体编上号,再作一套号签,对号签进行随机抽取,直至满足所需样本容量为止,抽中的号签所对应的个体即组成随机样本。

随机数字表(附表1)有10000个数字,它们是随机排列的,表内共有100行和100列,行号和列号各从00到99,使用随机数字表的方法是:假如某有限总体有1000个基本单位,予以编号为000,001,002,...,999,从其中抽取一个样本容量  $n=20$  的样本,可从表内任一点开始,沿横行依次取20个三位数,也可沿直行取20个三位数,遇有重复的数字应除去。本例假定在附表1中随机选择第15行第10列的交叉处作为开始点,沿横行取20个三位数,则这些三位数的号码是:

844,606,284,659,149,224,873,073,443,481,473,477,040,501,510,899,994,391,843,419。

在以上有限总体中,属于这些号码的20个基本单位,就是我们要抽取的20个单位,组成了样本。

当总体为无限时,随机数字表就无法使用了。事实上即使总体是有限的,但有时由于其容量相当大,也无法使用随机数字表,这时我们可以采用一些简单的抽样方法。但无论如何,从总体中抽取样本,都必须遵循随机的原则的。

(二) 试验 对于一些理论性的无限总体,一般是通过设计各种类型的试验来收集观测数据,这时通过试验获得的资料,便是随机样本的数据。有关试验方法和设计,将在五、六、

七章介绍。

## 第二节 资料的整理

获得试验数据后,可根据数据的多少,确定是否要通过一番整理。当数据不多时,可直接进行统计分析。当数据较多时,如有几百个数据,则直接进行统计分析将大大增加工作量,此时若把全部数据按数值的大小进行分组归类,制成次数分布表,则可大大减少统计分析的工作量,同时可以看出资料的集中和变化趋势,从而对资料的变化规律有一个初步的概念。次数分布表的制作方法因所研究的性状不同而有所不同。

### 一、试验资料的性质

试验资料一般可分为两大类:

(一)数量性状的资料 数量性状有计数和量测两种方式。

1.不连续性或间断性变数 指用计数方法获得的数据,每个数据必须是以整数来表示的,如小麦的基本苗数、分蘖数、穗数、每穗小穗数、每穗粒数等,在两个相邻的整数间,不容许有带小数的数据存在。由于在两个相邻整数间是不连续的,故称为不连续性或间断性变数。

2.连续性变数 指由称量、度量或测量等量测方法所得的数据,各个数据并不限于整数,在两个相邻的数值间可以有微量差异的其它数值存在,例如测定水稻的每穗粒重时,在2g至3g间,可以有2.3g或2.35g等数值存在,其小数位数的多少,因要求的精确度而异,这种变数称为连续性变数。

(二)质量性状的资料 指只能观察而不能量测的性状,即属性性状,如花药、子粒、颖壳等器官的颜色、芒的有无、茸毛的有无等。要从这类性状获得数量表征,可采用下列两种方法:

1.应用统计次数的方法 于一定样本内,统计其具有某种质量性状的个体数及不具某种质量性状的个体数,按两种类别计算其出现次数。例如在320株水稻植株中,有240株为紫色柱头的,80株为黄色柱头的。这类资料也称次数资料。

2.给予每类质量性状以相当数量的方法 例如小麦杂种的子粒颜色分离为白色与红色两种,可令白粒为0,红粒为1,则在粒色的群体中仅有0、1两种数量表示。

### 二、次数分布表

(一)间断性变数资料的整理 现以某小麦品种的每穗小穗数为例,在该品种中抽得随机样本,包含100个麦穗,计数其每穗小穗数,未加整理的资料见表1·1。

表1·1所述资料为间断性变数,每穗小穗数在15—20的范围内变动,把所有观察数据按每穗小穗数的多少加以排列,则共分为6组,组与组之间的相差为一个小穗,称为组距,这样可得表1·2形式的次数分布表。

从表1·2可以看到,一个原始资料经过初步整理后,便可了解资料的大致情况,如每穗小穗数以17个的为最多。另外,经过整理的资料也便于进一步统计分析。

表 1.1 100个麦穗的每穗小穗数

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 18 | 15 | 17 | 19 | 16 | 15 | 20 | 18 | 19 | 17 |
| 17 | 18 | 17 | 16 | 18 | 20 | 19 | 17 | 16 | 18 |
| 17 | 16 | 17 | 19 | 18 | 18 | 17 | 17 | 17 | 18 |
| 18 | 15 | 16 | 18 | 18 | 18 | 17 | 20 | 19 | 18 |
| 17 | 19 | 15 | 17 | 17 | 17 | 16 | 17 | 18 | 18 |
| 17 | 19 | 19 | 17 | 19 | 17 | 18 | 16 | 18 | 17 |
| 17 | 19 | 16 | 16 | 17 | 17 | 17 | 16 | 17 | 16 |
| 18 | 19 | 18 | 18 | 19 | 19 | 20 | 15 | 16 | 19 |
| 18 | 17 | 18 | 20 | 19 | 17 | 18 | 17 | 17 | 16 |
| 15 | 16 | 18 | 17 | 18 | 16 | 17 | 19 | 19 | 17 |

表 1.2 100个麦穗每穗小穗数的次数分布表

| 每穗小穗数 ( $x$ ) | 划 记 号 数       | 次 数 ( $f$ ) |
|---------------|---------------|-------------|
| 15            | 正 一           | 6           |
| 16            | 正 正 正         | 15          |
| 17            | 正 正 正 正 正 正 丁 | 32          |
| 18            | 正 正 正 正 正     | 25          |
| 19            | 正 正 正 丁       | 17          |
| 20            | 正             | 5           |
| 总次数 ( $n$ )   |               | 100         |

有些间断性变数资料，其观察数值的个数较多，且变异幅度又较大，因此不可能如上例那样按每一观察数值作一组的方法进行整理。例如研究某早稻品种的每穗粒数，共观察200个稻穗，每穗粒数的变异幅度为26—85粒，这样的资料如以每一观察数值为一组，则组数太多，资料的规律性就显示不出来，若每组包含若干粒数的幅度，例如以5粒为一组，则可使组数适当减少，本例经整理后分为12组，资料的规律性就较明显了，如表1·3。

**表 1.3 200个稻穗每穗粒数的  
次数分布表**

| 每穗粒数 (x) | 次数 (f) |
|----------|--------|
| 26       | 1      |
| 31       | 3      |
| 36       | 10     |
| 41       | 21     |
| 46       | 32     |
| 51       | 41     |
| 56       | 38     |
| 61       | 25     |
| 66       | 16     |
| 71       | 8      |
| 76       | 3      |
| 81       | 2      |
| 合 计      | 200    |

从表1·3可以看到，稻穗的每穗粒数以51—55粒这一组出现得最多。

(二) 连续性变数资料的整理 连续性变数不可能按间断性变数的归组方法来进行整理, 而必须先确定组数、组距、各组的上下限, 然后按观测值的大小来归组, 兹以表1·4的100株玉米穗位高度为例, 说明其整理方法。

1.求极差 从所有数据中找出最大观测值和最小观测值,再求其差数,即为极差,亦即整个样本的变异幅度。从表 1·4 中查到最大观测值为 158cm,最小观测值为 72cm,所以极差为  $158 - 72 = 86\text{cm}$ 。

2. 确定组数和组距 根据极差的大小把资料分为若干组, 每组的距离相等, 称为组



表 1·4 100株玉米的穗位高度

(单位: cm)

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 127 | 118 | 121 | 113 | 145 | 125 | 87  | 94  | 118 | 111 |
| 102 | 72  | 113 | 76  | 101 | 134 | 107 | 118 | 114 | 128 |
| 118 | 114 | 117 | 120 | 128 | 94  | 124 | 87  | 88  | 105 |
| 115 | 134 | 89  | 141 | 114 | 119 | 150 | 107 | 126 | 95  |
| 137 | 108 | 129 | 136 | 98  | 121 | 91  | 111 | 134 | 123 |
| 133 | 104 | 107 | 121 | 94  | 123 | 108 | 114 | 103 | 129 |
| 103 | 107 | 93  | 86  | 113 | 97  | 122 | 86  | 94  | 118 |
| 109 | 84  | 117 | 112 | 112 | 125 | 94  | 73  | 93  | 94  |
| 102 | 108 | 158 | 89  | 127 | 115 | 112 | 94  | 118 | 114 |
| 88  | 111 | 111 | 104 | 101 | 129 | 144 | 128 | 131 | 142 |

距。组数和组距是相互决定的, 决定组数时须考虑到资料整理的目的, 一方面在于简化资料, 以便于统计计算; 另一方面又须保持资料的真实面貌, 从中得出规律性的认识。当组数过少时, 则由次数分布表算出的统计数的精确度较差; 当组数过多时, 则又显得过于分散、看不出资料的集中情况, 且不利于以后的统计分析。为了使组数的多少和组距的大小比较适度, 应考虑: (1) 观测值个数的多少; (2) 极差的大小; (3) 便于计算; (4) 能反映出资料的真实面貌等方面。样本大小与组数多少的关系可参照表1·5来确定。

组数确定后, 还须决定组距, 组距 = 极差 / 组数。以表1·4的100株玉米穗位高度为例, 样本内观测值的个数为100, 查表1·5, 可分为8—16组, 假定分为9组, 则组距为  $86 \div 9 = 9.56\text{cm}$ , 为分组方便起见, 可以10cm作为组距。

3. 选定组限与组中点值 每组应有明确的界限, 才能使各个观测值划入一定的组内, 为此必须选定适当的组中点值及规定组

限。组值最好是整数, 或与观测值的位数相同, 以便于以后的计算。组限要明确, 最好比原始资料的数字多一位小数, 这样可使观测值归组时不致于含糊不清。首先选定第一组的中点值, 这一点选定后, 则该组组限确定, 其余各组的中点值和组限也可相应确定。第一组的中点值以接近于最小观测值为好, 这样可避免第一组内次数过多, 能较正确地反映资料的规律性。而最后一组的中点值以接近于最大观测值为好, 理由同前。

以100株玉米的穗位高度为例, 选定第一组的中点值为75cm, 则第二组的中点值为  $75 + 10 = 85\text{cm}$ , 余类推。每组有两个组限, 数值小的一个称为下限, 数值大的一个称为上限。在上述资料中, 第一组的下限为其中点值减去二分之一组距, 即  $75 - 1/2(10) = 70\text{cm}$ , 上限为其中点值加上二分之一组距, 即  $75 + 1/2(10) = 80\text{cm}$ , 为便于归组起见, 本组的上限可改为79.9cm, 故第一组的组限为70—79.9cm。其它各组的组限类推。

4. 把原始资料的各个观测值归组, 一般用“正”字来计数, 待全部观测值归组后, 即可求得各组的次数, 制成次数分布表(表1·6)。

(三) 属性变数资料的整理 属性变数的资料也可用类似次数分布的方法来整理。在整

表 1·5 样本大小与组数多少的关系

| 样本内观测值个数 | 分组时的组数 |
|----------|--------|
| 50       | 5—10   |
| 100      | 8—16   |
| 200      | 10—20  |
| 300      | 12—24  |
| 500      | 15—30  |
| 1000     | 20—40  |

表 1·6 100株玉米穗位高度的次数分布表

| 组限 (cm)     | 中 值 ( $x$ ) | 划 记 号 数     | 次数 ( $f$ ) |
|-------------|-------------|-------------|------------|
| 70—79.9     | 75          | 下           | 3          |
| 80—89.9     | 85          | 正 正         | 9          |
| 90—99.9     | 95          | 正 正 下       | 13         |
| 100—109.9   | 105         | 正 正 正 一     | 16         |
| 110—119.9   | 115         | 正 正 正 正 正 一 | 26         |
| 120—129.9   | 125         | 正 正 正 正     | 20         |
| 130—139.9   | 135         | 正 丁         | 7          |
| 140—149.9   | 145         | 正           | 4          |
| 150—159.9   | 155         | 丁           | 2          |
| 总次数 ( $n$ ) |             |             | 100        |

表 1·7 水稻杂种第二代植株米粒性状的分离情况

| 属性分组 ( $x$ ) | 次 数 ( $f$ ) |
|--------------|-------------|
| 红米非糯         | 96          |
| 红米糯稻         | 37          |
| 白米非糯         | 31          |
| 白米糯稻         | 15          |
| 合 计 ( $n$ )  | 179         |

理前,把资料按各种质量性状进行分类,分类数等于组数,然后根据各个体在质量属性上的表现,分别归入相应的组中,即可得到属性分布的规律性认识。例如某水稻杂种第二代植株米粒性状的分离情况归纳于表 1·7。

### 三、次数分布图

试验资料除用次数分布表来表示外,也可用图形来表示,次数分布图可以更形象地表明次数分布的情况。较普遍的图示形式有:方柱形图、多边形图和条形图。不论哪种形式的图示,都可画一直角坐标,用横坐标来表示各组的组距或中点值,用纵坐标来表示各组的次数。

(一) 方柱形图 方柱形图用于表示连续性变数的次数分布。现以表 1·6 100 株玉米穗位高度的次数分布表为例加以说明。该表有 9 组,所以在横轴上分为 9 个等分(因第一组的下限不是从 0 开始,故第一等分应离开原点一些,并在其前加折断号),并标上各组的组限,在纵轴上标定次数。横坐标与纵坐标的长度应有合适的比例(一般以 5 : 4 或 6 : 5 为好),绘成的图形才显得匀称。查表 1·6,第一组含有次数为 3,所以在第一组的上下限处绘两条纵线,其高度等于纵坐标上的 3 个单位,并用一横线连接两纵线的顶端,成方形,其余各组依次绘制,即可制成方柱形次数分布图(图 1·1)。

(二) 多边形图 多边形图也是表示连续性变数资料的一种方法。仍以 100 株玉米穗位高度的次数分布表为例,在图示时,以每组的中点值为代表,在横坐标上第一组中点值处垂直向上至纵坐标的 3 个单位处标记一点,其它各组均在其中点值处垂直向上按相应次数标记一点,最后把各点用直线连接起来,即成多边形次数分布图(图 1·2)。多边形图在左边最小组的组中点值外和右边最大组的组中点值外,应各伸出一个组距的距离,而使图形在此与横轴相交。

(三) 条形图 条形图适用于间断性变数和属性变数资料。一般在横轴上标出间断的中点值或分类性状,在纵轴上标出次数,现以表 1·7 水稻杂种第二代植株米粒性状分离情况

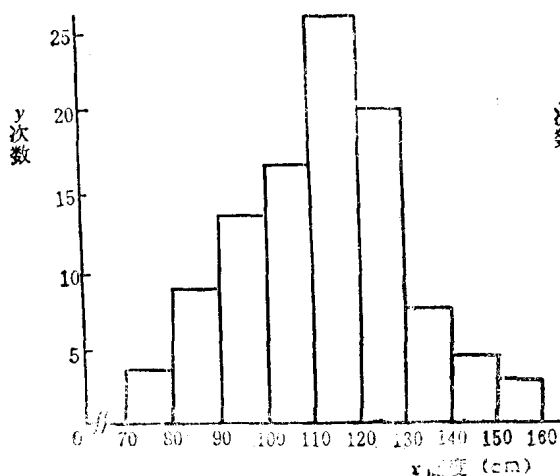


图 1.1 100株玉米穗位的次数分布方柱形图

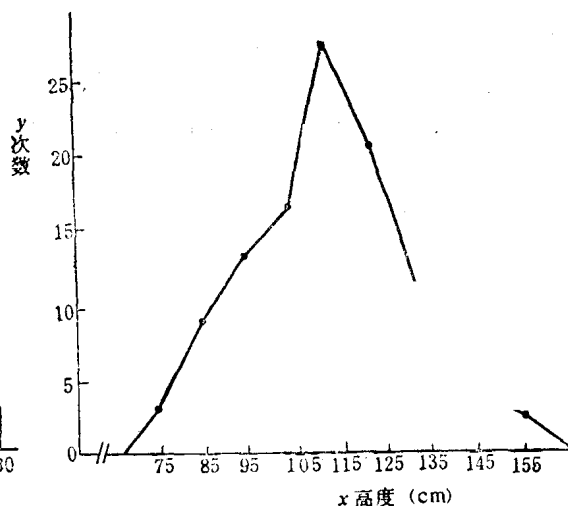


图 1.2 100株玉米穗位的次数分布多边形图

为例，在横轴上按等距离分别标出四种米粒性状，在纵轴上标定次数，在表1.7中，第一组为红米非糯稻，其次数为96，因此在红米非糯这一点上向上画一条形垂线，其高度为纵轴的96处，其余各组按同法绘制（图1.3）。

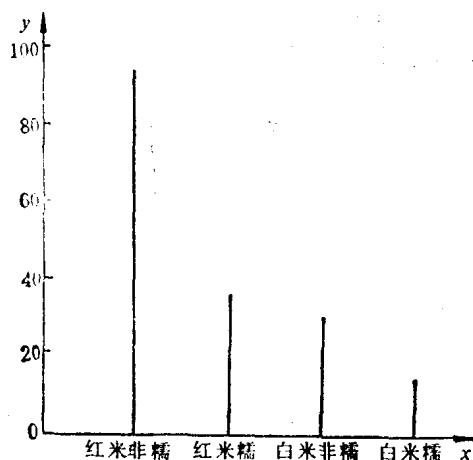


图 1.3 水稻 $F_2$ 代米粒性状分离条形图

### 第三节 资料的特征数

#### 一、平均数

（一）算术平均数 一个数量资料中各个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，记作 $\bar{x}$ 。算术平均数的应用非常广泛，其功用主要有：  
指出资料内变数的中心位置，用以衡量质量的一般水平。