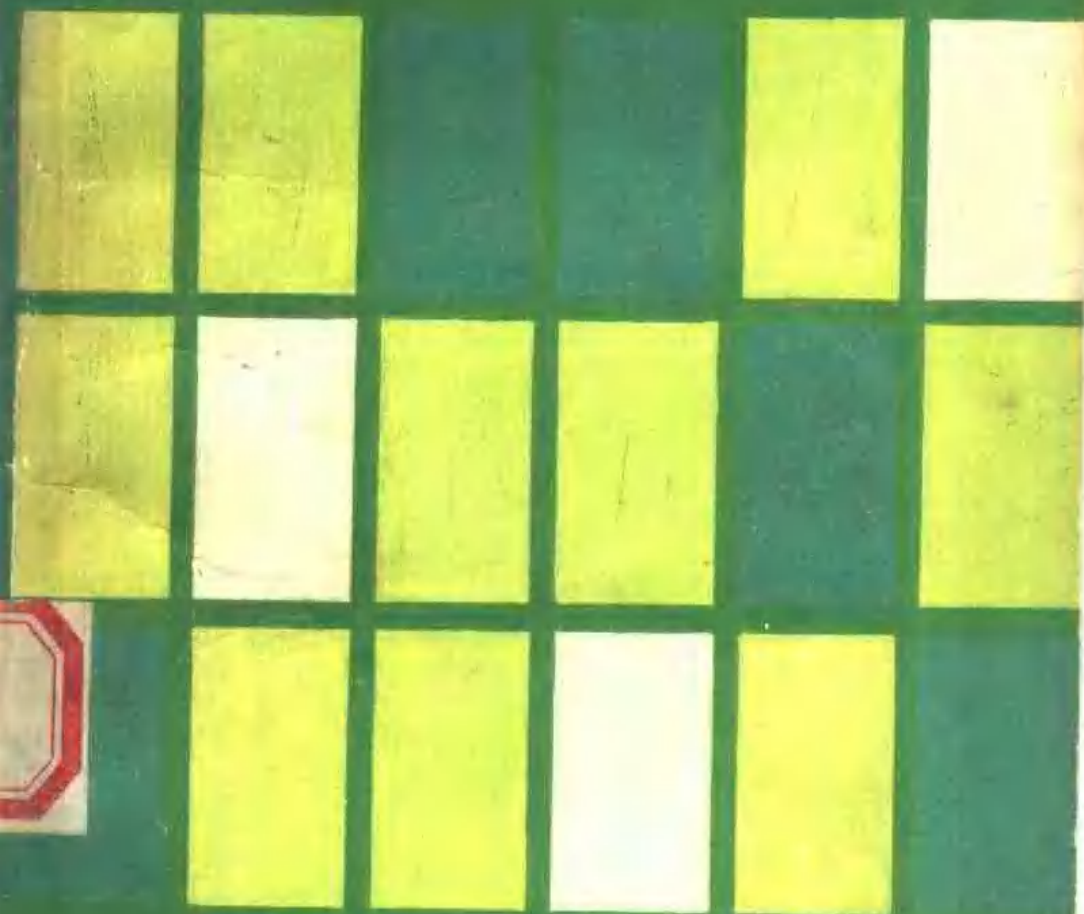


# 农业试验统计分析方法

胡秉民 张全德编写



# 农业试验统计分析方法

胡秉民 张全德

浙江科学技术出版社

责任编辑：沈虹

封面设计：王余

## 农业试验统计分析方法

胡秉民 张全德

·

浙江科学技术出版社出版

浙江印校印刷厂排版

浙江萧山印刷厂印刷

浙江省新华书店发行

开本850×1168 1/32 印张14.5 字数355,000

1985年2月第一版

1985年2月第一次印刷

印数1—10,200

统一书号：16221·107

定 价：2.40 元

## 前 言

当前，农业科学试验中运用统计分析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、便于自学的农业试验统计分析方法的图书。有鉴于此，我们编写了这本《农业试验统计分析方法》。

在编写这本书的时候我们考虑到两部分读者的需要。一部分是初学者，他们对农业试验统计模式和结果分析了解甚少，本书将作为一本入门基础书，通过对前八章的学习，掌握农业试验统计分析的基本技能和运用方法。对已经有些这方面知识的读者，我们同时编写了第九章经验曲线的配置，第十章多元回归，第十一章多点多年资料的综合分析和第十二章农业试验中的取样方法；并将概率与分布之二、某些统计分析中的重要概念的基本原理和袖珍式电子计算器的使用方法列于附录，以供农业科技人员根据自己的工作需要加以选择。

近年来，我们为农业科技人员举办了十余期生物统计培训班和讲座，本书就是在原有讲义的基础上，结合对大学本科生和研究生的教学实践，加以整理、充实和修改而成的。

全书蒙浙江农业大学高明尉老师审阅，谨此致谢。

限于作者水平，书中缺点和错误在所难免，恳请读者批评指正。

作 者

一九八三年六月于杭州

## 目 录

## 前言

|                             |        |
|-----------------------------|--------|
| 第一章 试验资料的抽样、整理与特征数 .....    | ( 1 )  |
| 第一节 抽样 .....                | ( 1 )  |
| 第二节 资料的整理 .....             | ( 3 )  |
| 第三节 资料的特征数 .....            | ( 7 )  |
| 思考与练习 .....                 | ( 17 ) |
| 第二章 概率与分布 .....             | ( 19 ) |
| 第一节 概率 .....                | ( 19 ) |
| 第二节 概率分布 .....              | ( 23 ) |
| 思考与练习 .....                 | ( 32 ) |
| 第三章 试验结果的显著性检验 .....        | ( 34 ) |
| 第一节 显著性检验的基本概念 .....        | ( 34 ) |
| 第二节 总体平均数的置信区间 .....        | ( 41 ) |
| 第三节 单个样本平均数的显著性检验 .....     | ( 44 ) |
| 第四节 两个样本平均数相比较的显著性检验 .....  | ( 48 ) |
| 第五节 方差齐性的检验 .....           | ( 57 ) |
| 思考与练习 .....                 | ( 58 ) |
| 第四章 计数资料的显著性检验 .....        | ( 61 ) |
| 第一节 卡平方( $\chi^2$ )检验 ..... | ( 61 ) |
| 第二节 适合性检验 .....             | ( 63 ) |
| 第三节 独立性检验 .....             | ( 65 ) |
| 思考与练习 .....                 | ( 70 ) |
| 第五章 方差分析 .....              | ( 72 ) |
| 第一节 单因素试验方差分析 .....         | ( 72 ) |

|                             |         |
|-----------------------------|---------|
| 第二节 两因素试验方差分析 .....         | ( 92 )  |
| 思考与练习 .....                 | ( 106 ) |
| 第六章 基本试验设计与结果分析 .....       | ( 109 ) |
| 第一节 田间试验的基本概念 .....         | ( 109 ) |
| 第二节 随机区组试验设计及其统计分析 .....    | ( 117 ) |
| 第三节 拉丁方试验设计及其统计分析 .....     | ( 128 ) |
| 第四节 对比法和间比法试验设计及其统计分析 ..... | ( 141 ) |
| 第五节 二因素随机区组试验结果的统计分析 .....  | ( 148 ) |
| 第六节 裂区试验设计及其统计分析 .....      | ( 157 ) |
| 第七节 条区试验设计及其统计分析 .....      | ( 171 ) |
| 思考与练习 .....                 | ( 185 ) |
| 第七章 正交试验设计及其统计分析 .....      | ( 191 ) |
| 第一节 正交表及其特点 .....           | ( 191 ) |
| 第二节 应用正交表设计试验 .....         | ( 194 ) |
| 第三节 正交试验结果分析 .....          | ( 196 ) |
| 第四节 有交互作用的正交试验设计与结果分析 ..... | ( 205 ) |
| 第五节 设置重复时的正交试验设计与结果分析 ..... | ( 213 ) |
| 第六节 正交试验设计与结果分析总结 .....     | ( 219 ) |
| 思考与练习 .....                 | ( 220 ) |
| 第八章 一元回归与相关 .....           | ( 222 ) |
| 第一节 直线回归方程的建立 .....         | ( 223 ) |
| 第二节 直线回归方程的显著性检验 .....      | ( 227 ) |
| 第三节 相关系数 .....              | ( 230 ) |
| 第四节 利用回归方程进行预测 .....        | ( 233 ) |
| 思考与练习 .....                 | ( 240 ) |
| 第九章 经验曲线的配置 .....           | ( 242 ) |
| 第一节 可线性化的曲线回归 .....         | ( 242 ) |
| 第二节 可线性化经验曲线方程的配置 .....     | ( 252 ) |
| 第三节 二次多项式经验曲线的配置 .....      | ( 262 ) |
| 思考与练习 .....                 | ( 267 ) |

|                                      |         |
|--------------------------------------|---------|
| 第十章 多元回归 .....                       | ( 269 ) |
| 第一节 多元线性回归方程的建立 .....                | ( 269 ) |
| 第二节 多元线性回归方程的显著性检验 .....             | ( 287 ) |
| 第三节 偏回归系数的显著性检验 .....                | ( 290 ) |
| 第四节 多元线性回归方程中自变量的剔除 .....            | ( 292 ) |
| 思考与练习 .....                          | ( 295 ) |
| 第十一章 多点多年试验资料的综合分析 .....             | ( 297 ) |
| 第一节 品种多点试验结果的统计分析 .....              | ( 297 ) |
| 第二节 品种多点多年试验结果的统计分析 .....            | ( 317 ) |
| 第十二章 农业试验中的取样方法 .....                | ( 335 ) |
| 第一节 农业试验中常用的取样方法 .....               | ( 335 ) |
| 第二节 试验小区产量的测定方法 .....                | ( 338 ) |
| 第三节 田间药效试验调查的取样方法 .....              | ( 340 ) |
| 附录与附表                                |         |
| 附录一 概率与分布之二 .....                    | ( 345 ) |
| 附录二 几个重要概念的基本原理 .....                | ( 359 ) |
| 附录三 袖珍式电子计算器的使用方法 .....              | ( 375 ) |
| 附表1 10000个随机数字 .....                 | ( 391 ) |
| 附表2 累积正态分布 $F_N(x)$ 值表 .....         | ( 397 ) |
| 附表3 正态离差 $u$ 值表 .....                | ( 399 ) |
| 附表4 $t$ 分布表 .....                    | ( 400 ) |
| 附表5 $F$ 分布表 .....                    | ( 401 ) |
| 附表6 Duncan's新复极差检验5%和1%SSR值表 .....   | ( 409 ) |
| 附表7 多重比较中的 $Q$ 表 .....               | ( 412 ) |
| 附表8 $\chi^2$ 分布表 .....               | ( 416 ) |
| 附表9 保证至少两个不同假设之一将被拒绝的样本容量<br>表 ..... | ( 418 ) |
| 附表10 常用正交表 .....                     | ( 419 ) |
| 附表11 相关系数显著性检验表 .....                | ( 432 ) |
| 附表12 泊松分布表 .....                     | ( 433 ) |

|                   |         |
|-------------------|---------|
| 附表13 常用对数表 .....  | ( 136 ) |
| 附表14 反对数表 .....   | ( 140 ) |
| 附表15 自然对数表 .....  | ( 144 ) |
| 附表16 希腊字母表 .....  | ( 149 ) |
| 附表17 常用统计符号 ..... | ( 150 ) |
| 参考书目 .....        | ( 152 ) |

# 第一章 试验资料的抽样、整理 与特征数

农业试验工作者在农业试验统计分析之前，首要关心的问题是，针对试验的全群，如何合理地去收集资料；又如何将收集到的资料加以整理，算出某些重要的特征数字，以作为进一步对统计资料进行分析，并据以作出科学结论的基础。

## 第一节 抽 样

在试验中，如果我们要对研究的全群进行全面检查，这在人力、物力各方面都是不允许的，因而存在着抽样检查的问题。下面说明抽样的几个基本概念。

### 一、总体与样本

研究对象的全体称为**总体**，组成总体的基本单元称为**个体**。总体往往是设想或抽象的，理论上包含无穷多的个体，例如水稻品种广陆矮4号的植株高度。但总体也可以是有限的，称为**有限总体**，例如研究在试验小区中广陆矮4号的植株高度。

总体是研究对象，但除较少的有限总体外，由于总体包含的个数太多，不可能全部测定。因而，一般只能从总体中抽取若干个个体来作出关于总体的推断，这些个体构成**样本**。而样本个体数称**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。农业试验中，常以30为分界线，即

$n \geq 30$ 为大样本,  $n < 30$ 为小样本。但这并不是一成不变的, 视具体问题而有差异。

由总体全部观察值算得的总体特征数, 如总体的平均数  $\mu$ , 称为参数, 用希腊字母表示。而从样本中各个体算得的特征数, 如样本平均数  $\bar{x}$ , 称为统计数(量)。它是总体相应参数的估计值, 一般用拉丁字母表示。

## 二、随机样本

为了考虑样本的代表性, 要求样本的各个体必须从总体中随机抽取, 这种从总体中随机抽取的样本称为随机样本。为了取得随机样本, 就要求在抽样方法上不依主观因素, 而使总体中的每一个个体被抽中是等可能的, 即每一个个体具有相等的被抽取的机会。

总体为有限时, 常用抽签法或随机数字表。抽签法是将有限总体的每一个个体编上号码, 再作一套号签, 对号签进行抽取, 直至抽取所需的容量构成一个样本。而随机数字表, 其表内数字的次序也是随机的, 任意抽取若干个数字, 就等于自己随机抽取了若干个签, 随机数字表见附表1。

随机数字表有10000个数字, 认为是以随机排列的, 其中 $5 \times 5$ 的方块仅仅为了读数的便利而分。随机数字表共有100行和100列, 编号各从00到99。使用随机数字表的方法是: 比如有限总体的1000个单元, 予以编号为000, 001, 002, 003, ..., 999。然后从表内任何一点开始, 沿横行依次取20个数, 也可以沿直行取20个数, 当然也可以隔一行或几行依次取20个数, 遇有重复的数字, 则应除去。表中的数字都认为是完全混和了的, 并没有什么特别的次序或类别, 任何一个序列都可以认为是随机的。

本例假定任意选择在附表1中15行10列之处开始, 沿横行选取20个三位数, 作为对应单元的编号, 则20个单元的编号是:

844, 606, 284, 659, 149, 224, 873, 073, 443, 481,  
673, 477, 040, 501, 510, 899, 994, 391, 843, 419.

以上20个单元就可以作为该有限总体的一个随机样本。

当总体为无限时,随机数字表就无法使用了。事实上,在许多情况下,即使总体是有限的,但由于容量相当大,也往往不使用随机数字表。但不论如何从总体中抽取样本,都必须遵循随机抽样的原则。

在抽样方法上,若把第一次抽得的样品仍放回去,第二次还可能抽到它,这种抽样叫重复抽样(或称返置抽样)。有时不可能进行重复抽样,特别是在大田中抽取样株后不可能放回去再作第二次抽样,这种情况叫非重复抽样(或称非返置抽样)。但是一般情况下,当总体所包含的个体数很大时,即使第一次抽出后不放回去而再作第二次抽样,其结果与重复抽样处理近似,故在总体很大情况下,一般都采用非返置抽样。

## 第二节 资料的整理

试验所得的数据,经检查与核对后,根据数据的多少确定是否分组。当数据不多时(小样本)不必分组,可直接进行统计分析。当数据较多时(大样本)有时需要将数据分成若干组,以便统计分析。经过分组归类后,制成较有规则的次数分布表,并根据次数分布表作出次数分布图。

### 一、次数分布表

根据表1-1的资料制作次数分布表1-3的步骤:

1. 计算全距(R)。全距又称极差,即数据中最大值减去最小值。在表1-1中,  $R = 158 - 72 = 86$  (厘米)。

表1-1 100株玉米穗位测定(单位:厘米)

|     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 127 | 118 | 121 | 113 | 145 | 125 | 87  | 94  |
| 118 | 111 | 102 | 72  | 113 | 76  | 101 | 134 |
| 107 | 118 | 114 | 128 | 119 | 111 | 117 | 120 |
| 125 | 94  | 124 | 97  | 88  | 105 | 115 | 134 |
| 89  | 141 | 114 | 119 | 150 | 107 | 126 | 95  |
| 137 | 108 | 129 | 136 | 98  | 121 | 91  | 111 |
| 134 | 123 | 138 | 104 | 107 | 121 | 94  | 126 |
| 108 | 114 | 103 | 129 | 103 | 127 | 93  | 86  |
| 113 | 97  | 122 | 86  | 94  | 118 | 109 | 84  |
| 117 | 121 | 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 |     |     |     |     |

2. 选择组数(K)。组数多少与样本容量有关,对大多数数据,7~20组就可以,即 $7 \leq K \leq 20$ 。决定组数时须考虑到资料整理的目的,即一方面在于简化资料,从中得到规律性的认识;另一方面又须保持资料的真实面貌。如组数过少,则由次数分布表算得的统计数的精确度较差,不能反映事物的真实面貌;如组数过多,则往往过于分散,看不到资料的集中情况,而且不便于以后的分析计算。表1-2可作为确定组数时的参考。

表1-2 样本容量与组数多少的关系

| 样 本 容 量 | 组 数   |
|---------|-------|
| 50      | 5—10  |
| 100     | 8—16  |
| 200     | 10—20 |
| 300     | 12—24 |
| 500     | 15—30 |
| 1000    | 20—40 |

本例一般不应小于8组，现假定取9组作次数分布表。

3. 选择合适的组距(C)。这是每一组上限与下限间的差数，以全距除以组数( $R/K$ )提供了组距的估计值，即：组距 = 全距/组数。但还要由研究人员根据具体情况作出最后的决定。如本例组距 =  $86/9 = 9.56$ ，根据本例的具体情况确定为10。组距以取整数为宜，力求简明，各组的组距一般要求相等，即采用等分分组。

4. 作出次数分布表。根据表1-1的资料，统计出落在各组内的个体数目，即组次数。今后在计算统计数字时，我们把落在各组中的观察值都视为相同，且它们等于组上限与组下限的平均值，称组中值，这样可得到次数分布表(表1-3)。

表1-3 表1-1资料的次数分布表

| 组 | 组限( $cm$ ) | 组中值 $\bar{x}$ | 划号记数  | 次数 $f$ | 累计次数 |
|---|------------|---------------|-------|--------|------|
| 1 | [70, 80)   | 75            | 下     | 3      | 3    |
| 2 | [80, 90)   | 85            | 正下    | 8      | 11   |
| 3 | [90, 100)  | 95            | 正正正   | 14     | 25   |
| 4 | [100, 110) | 105           | 正正正正  | 16     | 41   |
| 5 | [110, 120) | 115           | 正正正正正 | 25     | 66   |
| 6 | [120, 130) | 125           | 正正正正正 | 21     | 87   |
| 7 | [130, 140) | 135           | 正下    | 7      | 94   |
| 8 | [140, 150) | 145           | 正     | 4      | 98   |
| 9 | [150, 160) | 155           | 丁     | 2      | 100  |

上表如[70, 80)是表示一个区间，即代表 $70 \leq x < 80$ 这一范围，包括70而不包括80这段数。它比表1-1更清楚地表示玉米穗位高度的分布状况，分布在[110, 120)组为最多。

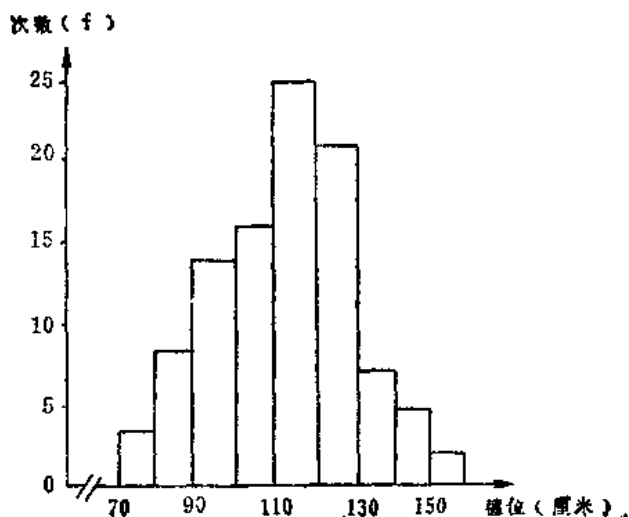
## 二、直方形图和多边形图

有时对于数量资料的描述，图示比次数分布表的数字更为有

效。常用的图示法有两种，直方形图和多边形图。不论哪一种形式的图示都是先画一直角坐标，用横坐标表示各组的组距或组中值，用纵坐标表示各组的次数。

直方形图是将分组数据用一系列毗连的长方形表示，如图1-1所示。这些长方形有相等的基底（组距），因此，它们的面积也与其次数成正比。多边形图是以组中点表示在横坐标，以次数表示在纵坐标，在直角坐标系中画上 $(\bar{x}_i, f_i)$ 各点，然后用直线将各点连接起来，即成多边形图。这个步骤与将相应的直方形图的长方形顶端中点用直线连接起来相同。表1-3数据的直方形图和多边形图示如图1-1。

次数分布表和图示可以使我们知道哪些组出现的次数较多，而且也可以约略看出资料的中心趋势和变异程度。



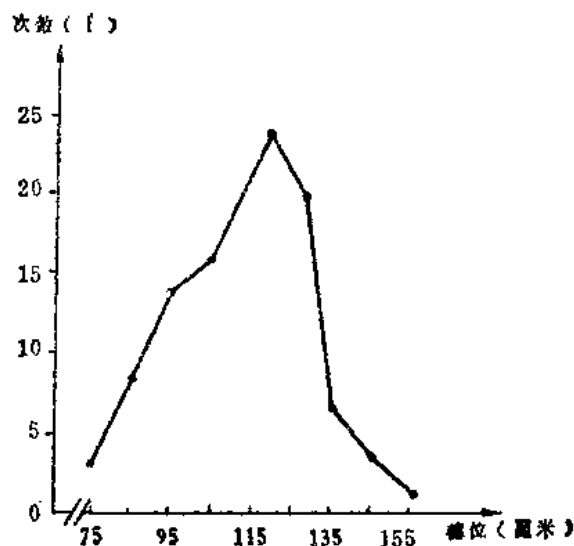


图1-1 表1-3数据的直方形图和多边形图

### 第三节 资料的特征数

为了进一步讨论样本资料的中心位置度量和变异程度，引进下述基本概念。

#### 一、平均数

平均数是表示资料内数据的中心位置。在处理样本数据时，经常要用样本平均数来代表某个总体的平均水平。

总体平均数用符号  $\mu$  表示，样本平均数分别用符号  $\bar{x}$ ,  $\bar{y}$ , ... 等表示。对于一个容量为  $n$  的样本： $x_1, x_2, \dots, x_n$ ，则有样本平均数：

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) \quad (1.1)$$

为书写方便, 记  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ , 并简记  $\bar{x} = \frac{1}{n} \Sigma x$ 。

计算表1-1玉米穗位的平均数

$$\begin{aligned} \bar{x} &= \frac{1}{100} \sum_{i=1}^{100} x_i = \frac{1}{100}(127 + 118 + \cdots + 142) \\ &= \frac{11236}{100} = 112.36(\text{厘米}) \end{aligned}$$

显然, 当样本容量  $n$  较大时, 计算颇为麻烦。对于资料已组成次数分布表的大样本计算平均数时是根据这一假设, 每组的中点值可用以代表组内的每一个观察值, 即认为中点发生的次数与组的次数相同。所以, 若从次数分布表计算平均数, 设  $n$  为样本总观察数,  $\bar{x}_i$  表示第  $i$  组样本的组中值,  $f_i$  为其组次数, 共分  $m$  组, 则有

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m x_i = \frac{1}{n} \sum_{i=1}^m f_i \bar{x}_i = \frac{1}{n}(f_1 \bar{x}_1 + f_2 \bar{x}_2 + \cdots + f_m \bar{x}_m) \quad (1.2)$$

这样的计算与真正的平均数是有误差的, 但组距不过大时, 误差相当小。

由表 1-3 利用公式 (1.2) 计算 100 株玉米穗位的平均数得:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^m f_i \bar{x}_i = \frac{1}{100} \sum_{i=1}^m f_i \bar{x}_i = \frac{1}{100}(3 \times 75 + 8 \times 85 + \cdots + 2 \times 155) \\ &= \frac{11250}{100} = 112.5(\text{厘米}) \end{aligned}$$

## 二、方差与标准差

一个总体中各数据间的变异是一个总体的重要特性，可以有多种方法使其数量化。

例如，有一对数分别为1和-1，其平均数为0。而另一对数分别是100和-100，其平均数同样是0。但很显然1和-1这对数相对于平均数0来说，其分散程度远小于100和-100这对数。前者集中而后者分散。

从上例可见，单靠平均数这一中心位置的度量来描述总体显然还不够，所以还要引进变异性的度量。变异性最常用的度量是方差与标准差。方差与标准差是衡量数据间的离散程度的一种指标。

总体方差以 $\sigma^2$ 表示，在实际问题中不易求得。通常用样本方差 $S^2$ 去估计总体方差 $\sigma^2$ ，样本方差 $S^2$ 的计算公式为

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (1.3)$$

对于此公式的来由我们作一些说明：从方差本身的意义来看，它是衡量数据间变异性的一个指标，即衡量数据间的离散程度。如有一个样本，其 $n$ 个观察值是 $x_1, x_2, \dots, x_n$ 。如何用定量来衡量这些数据的变异性呢？我们考虑到各数据离开平均数远，变异就大，反之，变异就小。这样，似乎可以用离均差之和

---


$$\text{*证: } \sum_{i=1}^n (x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x})$$

$$= (x_1 + x_2 + \dots + x_n) - n\bar{x} = \sum_{i=1}^n x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$$