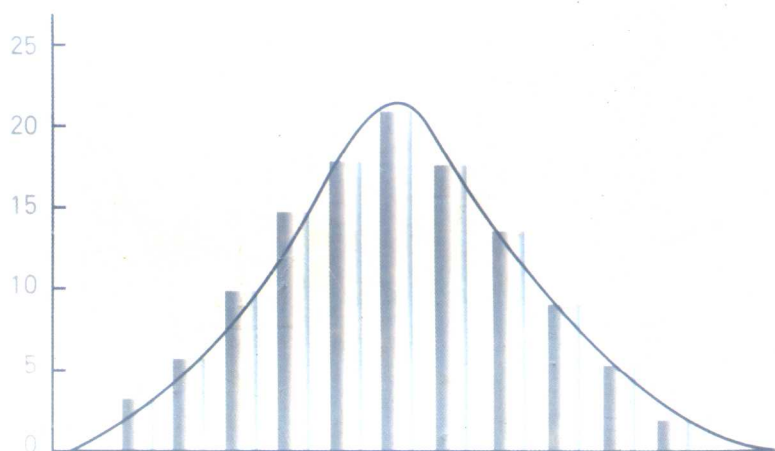




全国高等农林院校教材

试验方法及统计分析

白厚义 主编



中国林业出版社

全国高等农林院校教材

试验方法及统计分析

白厚义 主编

中国林业出版社

内 容 简 介

本书由四部分 10 章组成,较系统地介绍了生物统计的基本理论,试验研究的方案设计,常用的培养试验和田间试验的方法步骤,以及试验结果的统计分析,还介绍了多元分析中最常用的聚类分析和模糊聚类等。本书可作为高等农业院校农、林学类各专业以及生物技术各专业教学用书。也可作为其他专业和相关科技人员参考用书。

图书在版编目 (CIP) 数据

试验方法及统计分析/白厚义主编. —北京:中国林业出版社, 2005. 2

全国高等农林院校教材

ISBN 7-5038-3909-0

I. 试… II. 白… III. ①生物统计—试验设计(数学)—高等学校—教材 IV. Q—332

中国版本图书馆 CIP 数据核字(2004)第 142578 号

中国林业出版社·教材建设与出版管理中心

电话: 66170109 66181489 传真: 66170109

出版 中国林业出版社(100009 北京西城区德内大街刘海胡同 7 号)

E-mail: cfphz@public.bta.net.cn 电话: 66184477

发行 新华书店北京发行所

印刷 三河市富华印刷包装有限公司

版次 2005 年 2 月第 1 版

印次 2005 年 2 月第 1 次

开本 850mm×1168mm 1/16

印张 18.75

字数 400 千字

定价 25.00 元

阐述凡本书出现缺页、倒页、脱页等质量问题, 请向出版社发行部调换。

版权所有 侵权必究

全国高等农林院校“十五”规划教材

《试验方法及统计分析》编写人员

主 编：白厚义

副主编：袁 玲 张乃明 钱晓刚 荣湘民
田宵鸿

编 著 者：(按姓氏笔画排序)

王自林 (云南农业大学)

白厚义 (广西大学)

田宵鸿 (西北农林科技大学)

刘 强 (湖南农业大学)

张乃明 (云南农业大学)

张超兰 (广西大学)

陈佩琼 (广西大学)

陈秀虎 (广东清远职业技术学院)

周建斌 (西北农林科技大学)

周永祥 (西南农业大学)

杨 敏 (广东清远职业技术学院)

荣湘民 (湖南农业大学)

唐新莲 (广西大学)

袁 玲 (西南农业大学)

钱晓刚 (贵州大学)

前言

《试验方法及统计分析》是多年来多所院校在教学、科研和推广应用方面所获成果的基础上，利用国内外的最新理论，几经修改编著而成。本书不仅注意了内容的先进性，而且更着重于实用性和通俗性。对于基本原理，以深入浅出的方法作了较详尽的阐述，对其方法步骤，引用了大量富有成效的实际例子，做到了理论和实际的紧密结合。

本书共分四部分：

第一部分为生物统计的基本理论，包括总体、样本及误差，总体和样本的特征数，理论分布与抽样分布等。

第二部分为试验设计，包括试验研究概述，比较性试验设计，试验方法设计的基本原则等。

第三部分为试验方法，包括培养试验和田间试验等。

第四部分为数理统计，包括统计假设检验、方差分析、线性回归分析、非线性回归分析和聚类分析等。

本书适用于高等农业院校农、林学类本科各专业，以及生物技术专业，也可作为其他专业和相关科技人员的参考用书。全书内容较多，可根据各专业学时多少灵活安排。第一章至第八章可作为重点讲授内容，第九章和第十章可安排自学或作为参考内容。

本书编写分工：第一章（西南农业大学）、第二章（贵州大学、西北农林科技大学）、第三章（云南农业大学）、第四章（湖南农业大学）、第五章（广西大学、广东清远职业技术学院）、第六章（广西大学、贵州大学）、第七章（西南农业大学）、第八章（湖南农业大学）、第九章（湖南农业大学）、第十章（广西大学）。广西大学、广东清远职业技术学院编写附表。

本书引用了大量的资料，对提名和未提名的著者均表示衷心感谢！

书稿承蒙在植物营养研究法和生物统计等方面颇有造诣的中国农

业大学毛达如教授审阅和指正，特此表示深切感谢。

由于我们的水平有限，不妥之处在所难免，敬请读者批评指正。

编著者

2004. 10

目 录

前 言

第一章 生物统计的基本理论	(1)
第一节 总体、样本及误差	(1)
一、总体	(1)
二、样本	(1)
三、误差	(1)
第二节 总体和样本的特征数	(3)
一、表示数据集中趋势的特征数	(4)
二、表示数据分散性的特征数	(6)
第三节 理论分布与抽样分布	(10)
一、频率与概率	(10)
二、频数与频率分布	(11)
三、正态分布	(14)
四、抽样分布	(17)
五、置信限与置信概率	(21)
第二章 比较性试验设计	(23)
第一节 试验研究概述	(23)
一、试验研究的目的与任务	(23)
二、试验研究的种类	(23)
三、试验研究的一般程序	(24)
四、试验研究的基本要求	(26)
第二节 比较性试验方案设计	(27)
一、试验方案设计的基本概念	(27)
二、试验因素的效应和交互作用	(28)
三、试验方案设计的原则	(29)
四、比较性试验方案的设计	(30)

第三节 试验方法设计的基本原则	(37)
一、设置重复	(37)
二、随机排列	(38)
三、局部控制	(38)
第三章 培养试验法	(40)
第一节 概述	(40)
一、培养试验的性质、作用和特点	(40)
二、培养试验的分类	(41)
三、培养试验的发展概况	(41)
第二节 土培试验	(43)
一、土培试验的性质和任务	(43)
二、土培试验的技术	(43)
第三节 水培试验	(49)
一、水培试验的特点和任务	(49)
二、水培试验营养液的配制	(50)
三、水培试验的其它技术	(56)
第四节 砂培试验	(59)
一、砂培试验的特点和任务	(59)
二、砂培试验的技术	(60)
第五节 其它培养试验	(64)
一、隔离培养试验	(64)
二、流动培养试验	(68)
三、灭菌培养试验	(69)
四、渗滤水研究法	(70)
五、幼苗法	(71)
第六节 植物营养培养室的建立	(74)
一、培养室的组成	(74)
二、培养室的设计	(75)
三、培养室的管理	(76)
第四章 田间试验法	(78)
第一节 概述	(78)
一、田间试验的性质、作用和特点	(78)
二、田间试验的分类	(79)
三、田间试验方法的发展简介	(80)
第二节 田间试验的方法设计	(81)
一、小区的面积	(81)

二、小区的形状与方向	(83)
三、重复的次数	(84)
四、田间排列	(86)
第三节 田间试验的实施	(92)
一、试验田的选择和准备	(92)
二、试验的设置	(95)
三、试验的管理与观察	(98)
四、取样技术	(99)
五、收获与考种	(100)
第四节 特殊条件下的田间试验	(101)
一、微型小区试验	(102)
二、果树田间试验	(102)
第五节 示范推广试验	(103)
第五章 统计假设检验	(105)
第一节 统计假设检验的基本原理和方法	(105)
一、统计假设	(105)
二、统计假设检验的基本方法	(106)
三、统计假设检验的两尾检验与一尾检验	(109)
四、假设检验中的两类错误	(110)
第二节 单个平均数的假设检验	(112)
一、 u 检验	(112)
二、 t 检验	(113)
第三节 两个平均数相比的假设检验	(113)
一、两个正态总体相比较的方差比检验	(113)
二、成组数据相比较的假设检验	(114)
三、成对数据相比较的假设检验	(118)
第四节 百分数的假设检验	(120)
一、单个样本百分数的假设检验	(121)
二、两个样本百分数的假设检验	(122)
第五节 总体分布形式的假设检验	(124)
一、 χ^2 检验	(125)
二、柯氏检验	(128)
三、 W 检验	(130)
四、偏度、峰度检验	(132)
第六章 方差分析	(135)
第一节 方差分析的基本原理与方法	(135)

一、 F 分布与 F 检验	(135)
二、多重比较	(139)
第二节 常用比较性设计试验结果的统计分析	(143)
一、随机区组设计试验结果的统计分析	(143)
二、裂区设计试验结果的统计分析	(154)
三、正交设计试验结果的统计分析	(157)
第三节 方差分析的基本假定和数据转换	(161)
一、方差分析的基本假定	(161)
二、数据转换	(161)
 第七章 线性回归与相关	(165)
第一节 一元线性回归与相关	(165)
一、回归和相关的概念	(165)
二、一元线性回归分析	(167)
三、一元线性相关分析	(172)
第二节 多元线性回归与相关	(175)
一、多元线性回归	(175)
二、多元相关与偏相关	(182)
 第八章 曲线回归	(187)
第一节 可化为直线的曲线函数及最优方程的选择	(187)
一、可化为直线的曲线函数的主要类型	(187)
二、最优方程的选择	(189)
第二节 可化为直线的非线性回归分析的方法步骤	(191)
一、指数曲线回归分析的方法步骤	(191)
二、生长曲线回归分析的方法步骤	(193)
 第九章 多项式回归	(198)
第一节 一元多项式回归分析	(198)
一、一元多项式回归模型	(198)
二、一元多项式回归分析的一般步骤	(198)
三、一元多项式回归分析示例	(202)
第二节 多元多项式回归	(206)
一、多元多项式回归分析的一般步骤	(206)
二、多元多项式回归分析示例	(207)
第三节 正交多项式回归	(211)
一、一元正交多项式回归	(211)
二、多元正交多项式回归	(217)

第十章 聚类分析	(224)
第一节 聚类统计量	(224)
一、指标变量及其转换	(224)
二、距离	(226)
三、相似系数	(228)
第二节 系统聚类法	(231)
一、最短距离法	(232)
二、最长距离法	(234)
第三节 系统聚类法的应用	(235)
一、系统聚类在环境质量评价中的应用	(235)
二、在土壤分类上的应用	(239)
三、在遗传育种中的应用	(241)
第四节 模糊聚类	(244)
一、模糊聚类的基本原理和方法	(244)
二、模糊聚类在环境质量分类中的应用	(248)
三、模糊聚类在农业区划中的应用	(250)
参考文献	(256)
附录	(257)
附图 一些常见的可直线化的曲线函数图形	(257)
附表 1 正交表	(260)
附表 2 正态离差 u 值表 (两尾)	(263)
附表 3 学生氏 t 值表 (两尾)	(264)
附表 4 χ^2 值表 (一尾)	(265)
附表 5 柯尔莫哥洛夫检验临界值表	(266)
附表 6 正态性 W 检验的系数 ($a_{i,n}$)	(266)
附表 7 正态性 W 检验的临界值 (W_α) 表	(268)
附表 8 偏度、峰度检验的分位数表	(269)
附表 9 F 检验的临界值 (F_α) 表	(270)
附表 10 新复极差检验 SSR_α 值表	(276)
附表 11 q 值表 (双尾)	(277)
附表 12 百分数反正弦表 ($\sin^{-1}\sqrt{P}$) 转换表	(278)
附表 13 r_α 与 R_α 的 5% 和 1% 显著值	(281)
附表 14 正交多项式表 [$C=X(x)$ ($N=2\sim 13$)]	(282)
附表 15 10 000 个随机数字表	(284)

生物统计的基本理论

生物统计是资源环境学和生物学试验研究中广泛应用的一种数学方法。它利用概率论和数理统计学的原理和方法来处理资源环境科学和生物科学中的试验结果,推断试验结论,分析和解释有关现象,揭示内在规律。本章介绍生物统计的基本理论,包括总体、样本及误差的基本概念,总体与样本的特征数,理论分布与抽样分布,为进一步学习各种统计分析方法奠定基础。

第一节 总体、样本及误差

一、总 体

具有共同性质的个体所组成的集合体,称为总体,往往是我们研究对象的全体。总体一般是设想的或抽象的。例如,在研究环境污染状况时,各种污染物就可以看作一个总体。

总体按大小不同,可分为有限总体和无限总体。有限总体包含的个体数是可数的,数量是有限的。例如,在研究某地区不同水稻品种籽粒中的镉含量时,由于该地区水稻品种的数量是有限的,所以水稻品种可视为有限总体。无限总体包含的个体数量是不可数的,这种总体称为无限总体。例如,在研究不同条件下水稻籽粒的含镉量时,不同条件包括的内容很多,如不同的生态环境、不同的栽培方式、不同的水稻品种等,不同条件形成的个体是无限的,这些个体构成的总体称为无限总体。

二、样 本

研究任一事物,一般都在于揭示其总体的规律性。总体的性质决定了其中个体的性质。如果要对总体作出合乎实际的正确估计,最好是对总体中全部个体进行观察。但是由于一个总体包含的个体数往往很多,甚至无穷,以致在研究时不能对其逐一观察。因此,只能从总体中抽出一部分个体进行研究,这部分个体的总合称为样本或抽样总体。

如果从总体中抽取样本时,是采用随机的方法抽取的,所获得的样本称为随机样本。在一般的研究中,均要求随机样本。根据样本包含个体数目的多少,样本也有大小之分,包含个体数大于或等于 30 为大样本;小于 30 为小样本。

三、误 差

误差是指观测或测定结果与它对应的真值之差。真值是客观存在的,随着技

术的进步,测定值逐步接近于真值。在环境学研究中,一般是取已知标准样测定值作为真值,在没有标准样品的情况,常用已确认为标准方法的测定结果作为真值,或者取大样本的平均值作为真值的近似值。

由于观测、测定中往往受许多外来因素干扰,导致所取得的数据普遍存在误差,只是误差大小不同而已。产生误差的原因主要来源于如下三个方面:首先是技术方面的原因,如观测时所用的仪器、工具本身精度的限制。其次是测定方法、人员素质的影响。第三是外界环境条件带来的影响。因此,无论在理论上或应用上,人们对所收集的含有误差的统计数据,进行科学地分析处理,去伪存真就非常必要。要做到这一点,就必须弄清误差的基本概念及其计算方法。

(一) 误差的种类

误差按其性质不同可分为三种:

1. 偶然误差

观测、测定中由于偶然因素如微气流、微小的温度变化、仪器的轻微振动等所引起的误差称为偶然误差。通常所讲的误差,均是指这类偶然误差。

偶然误差具有偶然性,是不可预知的,所以这类误差是不可避免的,只能缩小,不能消除。偶然误差具有随机性,服从一定的概率分布,其发生受概率的大小所支配,也就是说,随机误差就其个体看是偶然的,而就其总体来说,却具有其必然的内在规律。根据研究,随机误差服从正态分布。随着重复次数的增加,随机误差由于正负相抵消,其平均值不断缩小,逐渐趋于零。因此,多次测定的平均值比单个测定值的随机误差小,这种性质称为抵偿性。

随机误差是在整个试验过程中形成的,试验的科学设计和正确的实施均能降低试验误差,提高试验精确度。

2. 系统误差

由于某个或某些固定的因素引起的有确定规律的误差,其误差的符号和数值在整个试验过程中是恒定不变的,或者是遵循着一定规律变化的,这类误差称为系统误差。例如温度升高或降低,仪器的量值偏高或偏低,以及化验人员的习惯和偏向等都能引起这类系统误差。系统误差的出现一般是有规律的,其产生的原因往往是可知的。因而这种误差可以根据其产生的原因加以校正和消除。系统误差表征着试验结果的准确度。一般地说,在试验过程中应尽可能设法预见到各系统误差的具体来源,并设法消除其影响。如果存在某种系统误差而未能预见和消除,就会增大试验误差。

3. 疏失误差

观测、测定中因某种不应有的错误造成所得数据与事实明显不符的误差称为疏失误差。例如记录错误、读数错误、称量错误、试剂加错、样本取错或仪器出现故障而未能发现所带来的错误等,均属疏失误差。这种误差实质上就是由于疏忽大意,操作不正确等主观原因所造成的,均无规律可循,在试验过程中应千方百计避免这种错误的发生。对于疏失误差,应作为异常值予以剔除。

(二) 误差的表示方法

误差的表示因性质、要求不同而有所区别，通常有如下几种表示方法：

1. 误差区间

表示误差区间的大小范围（误差范围），定义为一定条件下一组观察值的最高值与最低值之差，即

$$\max(x_k) - \min(x_k) \quad (k=1, 2, \dots, n) \quad (1.1.1)$$

2. 绝对误差

绝对误差指测定值与真值之差，即

$$\text{绝对误差} = \text{测定值} - \text{真值}$$

3. 相对误差

相对误差指误差与真值间的比值，即

$$\text{相对误差} = \text{误差} / \text{真值}$$

在应用上，一般取误差与测定值间的比值，即

$$\text{相对误差} = \text{误差} / \text{测定值}$$

4. 离差

离差指测定值 x_k 与测定值的平均数 $\bar{x}$ 之差，即

$$d_k = |x_k - \bar{x}| \quad (1.1.2)$$

5. 算术平均误差

算术平均误差指测定值的所有离均差绝对值之和除以测定次数，即

$$\bar{d} = \frac{\sum_{k=1}^n |x_k - \bar{x}|}{n} \quad (1.1.3)$$

此外，极差、方差、标准差、变异系数等均是误差的不同表示方法，这些将在后面作介绍。

第二节 总体和样本的特征数

任何事物的存在，总是与其环境条件密切相关，同一总体中的各个体不可能都处在绝对相同的条件之中，由于受偶然因素的影响，个体间的变异是必然存在的。例如同一稻田中，不同植株的高度、穗长、粒重等总有差异的。同一总体中个体间具有变异的每种性状或特征，在量方面可以表现为不同数值，对于这种因个体不同而变异的量，在统计学中称为变量，而不同个体在某一性状方面具体表现的数值称为观察值。如水稻株高就是一个变量，其中某株水稻株高 98cm 就是一个观察值。变量有连续和不连续之分，总体中相邻两个观察值之差可达无限小者为连续变量，相邻两个观察值之差最小为 1 者为不连续变量，或称非连续性变量。例如 1000 个不同土壤有机质的含量，它们构成连续变量；而不同地区年降雨次数则是非连续变量。

研究任何事物, 既应说明总体的集中性, 又应说明总体的变异性。凡能说明不同总体集中性和变异性特征的数值均称为总体特征数, 亦称参数。样本特征数是总体特征数的估计值, 称为统计值或统计数。统计数通常包括平均数、极差、方差、标准差和变异系数等。平均数是反映总体或样本集中性的特征, 而极差、样本平均差、平方和、方差、标准差、变异系数则为总体或样本变异性的特征数。

一、表示数据集中趋势的特征数

表示数据集中趋势的特征数包括平均数、中位数和众数。

(一) 平均数

平均数包括算术平均数、加权平均数、几何平均数、调和平均数和平方根平均数等。它们适用于性质相同的个体间求平均。在环境科学中, 前三者较常用。

1. 算术平均数

一个具有 N 个观察值的总体, 其观察值为 $x_1, x_2, \dots, x_N$, 则该总体的算术平均数 μ 定义为:

$$\mu = \frac{1}{N} (x_1 + x_2 + \dots + x_N) = \frac{1}{N} \sum_{i=1}^N x_i \quad (i=1, 2, \dots, N) \quad (1.2.1)$$

对含有 n 个观察值 $x_1, x_2, \dots, x_n$ 的样本, 其平均数 $\bar{x}$ 的定义为:

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i = \frac{\sum_{i=1}^n x_i}{n} \quad (i=1, 2, \dots, n) \quad (1.2.2)$$

因此, 一组数量资料, 各个观察值的总和除以观察值的个数所得的商, 称为算术平均数, 简称平均数或均数。算术平均数主要适用于具有对称分布的数据, 例如正态分布、二项式分布等。

2. 加权平均数

如果样本个体数据 $x_1, x_2, \dots, x_n$ 取值因频数不同或对总体的重要性有所差异, 则常采用加权平均方法。加权平均数是考虑了各数值所占比重后算出的平均数。加权平均数的定义式为:

$$\bar{x} = \sum_{i=1}^k f_i x_i / \sum f_i = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{f_1 + f_2 + \dots + f_k} = \frac{1}{\sum f_i} \sum_{i=1}^k f_i x_i \quad (1.2.3)$$

定义式中 f_i 为个体数据出现的频数, 或是因该个体对样本贡献不同而取的不同数值, 称为权重。

3. 几何平均数

以 G 表示, 定义式为:

$$G = \left(\prod_{k=1}^n x_k \right)^{1/n} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (1.2.4)$$

定义式中的 $\prod$ 为连乘符号。定义式取对数即成：

$$\lg G = \frac{\lg x_1 + \lg x_2 + \cdots + \lg x_n}{n}$$

计算几何平均数通常是求对数后，再用反对数的方法求出。几何平均数主要用于具有对数正态分布的数据。例如水文资料中河流的流量，一个地区多年降雨量往往符合对数正态分布。

4. 调和平均数

调和平均数以 H 表示，定义式为：

$$H = \frac{1}{\frac{1}{n} \sum_{k=1}^n \frac{1}{x_k}} = \frac{1}{\frac{1}{n} \left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n} \right)} \quad (1.2.5)$$

例如，从土壤表面下 30cm 处起，测定水分在土壤毛细管中上升的速度，第一个 10cm 的上升速度为 8cm/min，第二个 10cm 为 6cm/min，第三个 10cm 为 4cm/min，求其平均速度必须用调和平均数，即

$$H = \frac{1}{\frac{1}{3} \left(\frac{1}{8} + \frac{1}{6} + \frac{1}{4} \right)} = 5 \frac{7}{13} \text{ (cm/min)}$$

5. 平方根平均数

平方根平均数以 Q 表示，定义式为：

$$Q = \sqrt{\sum_{k=1}^n x_k^2 / n} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}} \quad (1.2.6)$$

(二) 中位数

中位数是指将样本数据从大到小，或从小到大依次排列，居中间位置的数。中位数以 M_c 表示，定义式为：

$$M_c = x_{(n+1)/2} \quad (n \text{ 为奇数时})$$

$$M_c = \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) \quad (n \text{ 为偶数时}) \quad (1.2.7)$$

在环境研究中所收集的数据，有时显得比较分散，个别的离群较远，难以判断去留。这种数据应用中位数表示其集中性较好。

(三) 众数

样本数据中频数出现最多的值称为众数。也就是频率最大的值，在一组数据中众数可以不止一个。众数以 M_o 表示。如果平均数以 $\bar{x}$ 表示，中位数以 M_c 表示，则众数 M_o 有如下经验公式：

$$M_o = 3M_c - 2\bar{x} \quad (1.2.8)$$

(四) 计算方法

[例 1.2.1] 某地排污明渠锌监测值 (mg/L) 0.114、0.300、0.126、

0.308、0.140、0.342、0.160、0.190、0.260、0.220。求算术平均数 $\bar{x}$ ，中位数 M_c ，众数 M_0 。

1. 算术平均数

$$\bar{x} = \sum_{k=1}^{10} x_k / n = \frac{x_1 + x_2 + \cdots + x_{10}}{10} = 0.216$$

2. 中位数

将例 1.2.1 数据排序，其计算为

$$M_c = \frac{1}{2} (x_{n/2} + x_{n/2+1}) = \frac{1}{2} (x_5 + x_6) = \frac{1}{2} (0.190 + 0.220) = 0.205$$

3. 众数

$$M_0 = 3M_c - 2\bar{x} = 3 \times 0.205 - 2 \times 0.216 = 0.183$$

二、表示数据分散性的特征数

每个样本有一批观察数据，以平均数作为样本的代表值。样本平均数代表性的优劣，依赖于样本内各个观察值的分散性，即变异程度。为了更全面地描述样本，只有平均数是不够的，还必须度量其变异度。表示变异度的统计数或参数较多，最常用的有极差、方差、标准差和变异系数。

(一) 极差

极差又称全距。资料中最大观察值与最小观察值的差值称为极差。极差以 R 表示，定义式为：

$$R = \max(x_1, x_2, \cdots, x_n) - \min(x_1, x_2, \cdots, x_n) \quad (1.2.9)$$

极差计算简便，当 $n < 10$ 时常采用。但极差作为度量资料的分散性不够理想，即不能较好地反映变异事物的变异程度。因为极差是由两个极端值决定，没有利用变异资料的全部信息，且受数据中极端值的影响。所以，当观察数量多，变异度大，特别是出现极端值时，不宜采用极差进行度量。

(二) 样本平均差

样本中各数据与其算术平均数之差的绝对值的平均数称为样本平均差，以 d 表示，定义式为：

$$d = \frac{1}{n} \sum_{k=1}^n |x_k - \bar{x}| \quad (1.2.10)$$

(三) 方差

1. 平方和

为了正确反映资料的变异度，较合理的方法是根据样本全部观察值来度量变异资料的变异度。这时要选定一个数值作为标准。平均数既为总体或样本的代表值，则以平均数作为标准最为合理。含有 n 个观察值的样本，其各观察值为 x_1 ，