

数据分析与试验优化设计

白新桂

清华大学出版社

9-N3/

G/1

数据分析与试验优化设计

白新桂 编著

清华大学出版社

内 容 简 介

本书介绍了如何使用电子计算机进行实验数据分析与试验优化设计。全书分三篇，共十一章，分别介绍了数据处理、回归分析及试验优化设计。书后附有习题与解答，线性方程组的计算机解法及常用数表。

本书可作为高等工科院校材料与机械工程类专业本科生、研究生教材，也可供有关研究单位或工厂科技人员参考。

数据分析与试验优化设计

白新桂 编著

清华大学出版社出版

(北京清华园)

北京京辉印刷厂印刷

新华书店北京发行所发行

开本：787×1092 1/16 印张：23 字数：569千字

1986年10月第1版 1989年4月第2次印刷

印数：8001~12000

ISBN 7-302-00211-8/TP·83

定价：4.55元

前 言

科学实验的数据分析与试验的优化设计，随着电子计算技术的迅速发展，不断赋予新的内容。

本书在引入有关概率论与数理统计基本概念的基础上，着重介绍了数据处理、回归分析及试验优化设计的计算方法与程序设计框图。并且附有典型问题的BASIC语言程序及其在机械与材料工程中的一些应用。

本书承蒙清华大学刘庄副教授、北京冶金设计院林连清高级工程师、冶金部有色金属研究院总院金其坚高级工程师、石力开、张葛丽工程师以及武庆兰、谢锋、夏为民等同志审阅并提出了宝贵的修改意见。在此，向他们表示衷心的感谢。

限于时间与编者的水平，错误与不当之处恳请读者批评指正。

编 者

一九八三年于北京

目 录

第一篇 数据处理

引言	1
第一章 有效数字与数据表示方法	2
§ 1—1 有效位数的意义	2
§ 1—2 古典“四舍五入”法则形成的舍入误差	2
§ 1—3 舍入法则	3
§ 1—4 有效数字的确定	4
§ 1—5 数据的表示方法	5
第二章 数字特征的计算与检验	7
§ 2—1 几个数理统计的基本概念	8
§ 2—2 数字特征的计算方法与程序设计	11
§ 2—3 示例	18
第三章 异常数据的删除与漏失数据的弥补	24
§ 3—1 拉依达 (Паїта) 准则	24
§ 3—2 肖维勒 (Chauvenl) 准则	27
§ 3—3 格拉布斯 (Grubbs) 准则	28
§ 3—4 狄克逊 (Dixon) 准则	34
§ 3—5 t 检验准则 (罗马诺夫斯基准则)	39
§ 3—6 漏失数据的弥补	44
第四章 误差分析	47
§ 4—1 真值与误差	47
§ 4—2 误差的表示方法	48
§ 4—3 实验最大可能误差的确定	50
§ 4—4 实验精度评定	56

第二篇 回归分析

第五章 简单回归分析及其程序设计	59
§ 5—1 回归分析简介	59
§ 5—2 一元线性回归分析	60
§ 5—3 回归方程的建立过程与程序设计	65
§ 5—4 Y 值的预报与控制——一元线性回归方程的应用	67
§ 5—5 二元线性回归分析	72
第六章 多元回归分析及其程序设计	84

§ 6—1	多元线性回归的数学模型	84
§ 6—2	多元回归系数的确定方法	85
§ 6—3	回归方程的显著性检验	87
§ 6—4	回归系数的显著性检验	89
§ 6—5	利用回归方程进行预测与控制	91
§ 6—6	三元线性回归的程序设计	92
第七章	逐步回归分析及程序设计	100
§ 7—1	选择“最优”回归方程途径	100
§ 7—2	逐步回归数学模型	101
§ 7—3	逐步回归分析的计算方法	104
§ 7—4	逐步回归计算过程	109
§ 7—5	逐步回归的程序设计	115
第八章	非线性回归分析及程序设计	124
§ 8—1	线性化回归	124
§ 8—2	多项式回归	128
§ 8—3	正交多项式回归	130
第三篇 试验优化设计		
第九章	简单试验设计	143
§ 9—1	试验设计的基本概念	143
§ 9—2	全面试验法	143
§ 9—3	简单对比法(孤立因素法)	144
§ 9—4	完全随机化试验法	144
§ 9—5	随机区试验法	146
§ 9—6	拉丁法与正交拉丁方试验法	148
第十章	正交试验设计	151
§ 10—1	正交试验设计	151
§ 10—2	正交试验的直观分析	153
§ 10—3	多指标试验设计的分析方法	160
§ 10—4	水平不同的正交试验设计	173
§ 10—5	有交互作用的正交试验设计	174
§ 10—6	正交试验设计的方差分析	179
第十一章	回归正交设计	188
§ 11—1	回归正交设计简介	188
§ 11—2	一次回归正交设计	188
附录		200
附录 1	习题与解答	200
附录 2	线性方程组的计算机解法	264
附表		291

附表 1	相关系数 R 表	291
附表 2	t 分布的双侧分位数 (t_{α})表	293
附表 3	F 检验的临界值 (F_{α})表	294
附表 4	正交多项式表 ($N = 2 \sim 30$)	300
附表 5	正交拉丁方表	309
附表 6	常用正交表	312
附表 7	常用回归正交表	347
参考书目	359

第一篇 数据 处理

引 言

材料工程的质量控制及材料科学的研究均以一定条件下得到的原始数据为前提。因此，人们经常需要加工处理所积累的大量技术数据，找出其中的变化规律和内在联系，用以指导科学与生产实践。

在近代的科学实验中，现象的随机性质是十分突出的，使过程的内在规律性往往被现象表面的偶然性所掩盖。恩格斯曾说：“凡表面上看去是偶然性在起作用的地方，其实这种偶然性本身始终是服从于内部的隐藏着规律的”。（《费尔巴哈与德国古典哲学的终结》，人民出版社，第38页。）

所谓实验数据处理，就是从带有偶然性的观测值中，运用适当的数学工具及以电子计算机为手段的“电算”方法，导出规律性结论的过程。只有科学地处理实验数据，方能充分有效地利用实验测试信息，得出正确结论，合理地组织科学实验。

总之，数据处理是测试信息的总结，是控制的先导。它的重要性与内容已超出了古典的实验误差处理的范围。在现代生产与科学实验中已经得到广泛的应用。

第一章 有效数字与数据表示方法

§1-1 有效位数的意义

人们在进行生产与科学实验时，需要把测试值记录下来并进行整理计算。究竟应该用几位数字表示所测得的数据和结果呢？有人误认为小数点后的位数越多越准确。其实不然，小数点后的位数不是准确性的标准，它只与所采用的单位大小有关。例如170.5cm与1.705m的准确性是完全一样的。另外，由于测量仪表的精度与实验观测者感官功能的限制，测量的准确度是有限的。因而，在计算过程中结果位数超出了测量所能达到的准确度的限度，数据的精度将是虚假的。相反，若写出数字的位数低于测得数据的位数时，将降低数据的精确度。

实验数据有效位数的确定的正确作法是所取位数除末一位数字为测量时的可疑数或估计数外，其余各位数字都是准确可靠的，通常末一位可疑数字上下可以有一个单位的误差。这样的数字称作有效数字。

例如，测长用的钢尺准确度为1mm，精度为0.5mm。当用钢尺测量某一物体长度为11.5mm时，前两位数字是准确的，而第三位是估计值。此时，该数称作三位有效数字。

当确定了有效数字的位数以后，在古典的数据处理中，通常采用的是“四舍五入”法则。

§1-2 古典“四舍五入”法则形成的舍入误差

一个有效数字的有效位数指的是最高位数字不为零的位数。一个数值如5.337，若保留两位有效数字，则通常是把第三位的数（这里是3），按古典的“四舍五入”法则进行处理，故舍去3，而得5.3。

设某一数值要保留 n 位有效位，则按古典的“四舍五入”法则对第 $n+1$ 位的数进行处理，就必然会带来舍入误差 ϵ 。第 $n+1$ 位数及其相应的舍入误差，如下表所示。

表 1-2-1

第 $n+1$ 位的数	0	1	2	3	4	5	6	7	8	9
舍入误差 ϵ	0	-1	-2	-3	-4	5	4	3	2	1

舍入误差 ϵ ，是一个离散型的随机变量。用大写字母 X 表示。

第 $n+1$ 位的数可能取值为0, 1, 2, 3, 4, 5, 6, 7, 8, 9。这十个数值可能出现的概率都是相等的。

$$P(X=0, -1, -2, -3, -4, 5, 4, 3, 2, 1) = \frac{1}{10}$$

一个离散型随机变量 X 取 x_i 值的概率为 P_i 。这意味着在极多的 N 次实践中,大体上有 NP_i 次的取值为 x_i ,从而极多次实践的取值总和大体为 $\sum_i NP_i x_i$,即极多次实践取值的平均值大体上为 $\frac{1}{N} \sum_i NP_i x_i = \sum_i P_i x_i$ 。我们把它叫做离散型随机变量 X 的数学期望值或均值。记作

$$\mu_x = E(X) = \sum_i P_i x_i$$

因此,古典的“四舍五入”法则的舍入误差(ε)的期望值 μ_ε 为:

$$\mu_\varepsilon = E(\varepsilon) = \sum_i P_i x_i = \frac{5}{10} = \frac{1}{2}$$

所以,古典的“四舍五入”法则的舍入误差的期望值等于第 $n+1$ 位单位的二分之一。

舍入误差是人为的引入误差。引入一种人为的误差总是希望它在多次实践中的均值基本上等于零才好。从上述的分析可以看出,古典的“四舍五入”法则并不具有所期望的性质。当第 $n+1$ 位出现了除5以外的其它数字时,其舍入误差在极多次实践中平均起来都可以抵消。只是在第 $n+1$ 位是5时,其舍入误差是5,无法相消。这就是古典“四舍五入”法则的症结所在。

如果规定:当第 n 位上的数是偶数(包括0)时,第 $n+1$ 位上的5予以舍去,此时舍入误差为-5,单位是第 $n+1$ 的单位;当第 n 位上的数是奇数时,第 $n+1$ 位上的5就进入,使第 n 位增加1单位,此时舍入误差为+5,单位是第 $n+1$ 的单位。由于第 n 位上出现偶数或奇数的概率各为二分之一,其相应舍入误差为-5或+5的概率也各为二分之一。从而极多次实践平均起来可以相消。这样,就修正了古典的“四舍五入”法则,使其具有所期望的性质。

§1-3 舍入法则

为了适应生产与科学技术工作的需要,我国科学技术委员会正式颁布了《数学修约规则》,通称为“四舍六入五单双”法则。概括说明:

四舍六入五考虑,五后非零必进一。五后皆零视奇偶,五前为偶应舍去,五前为奇则进一。

这一法则的具体运用,分述如下:

1. 若被舍弃的第一位数字小于5,则其前一位保持不变。如28.2345只取三位有效数字时,其被舍弃的第一位数字为3,小于5,则有效数字应为28.2。

2. 若被舍弃的第一位数字大于5,则其前一位数字加1。如28.2645只取三位有效数字时,其被舍弃的第一位数字为6,大于5,则有效数字应为28.3。

3. 若被舍弃的第一位数字等于5,而其后面数字全部为零,则视被保留的末位数字为奇数或偶数而定进或舍。如28.350及28.250,只取三位有效数字时,则分别应为28.4及28.2。

4. 若被舍弃的第一位数字为5,而其后面的数字并非全部为零,则进一。如28.2501,只取三位有效数字时,则进一,成为28.3。

5. 若被舍弃的数字包括几位数字时,不得对该数字进行连续的进位或舍弃,而应根据以上各条作一次处理。如2.154546,只取三位有效数字时,应为2.15,而不得按2.154546→2.15455→2.1546→2.155→2.16方法错误地连续进位至2.16。

6. 整数的修约也应遵照上述规则进行处理。如23438, 只取三位有效数字时, 则应为23400或 2.34×10^4 。

§1-4 有效数字的确定

在测量过程中, 对一次测量数据的有效数字位数, 应与所用仪器的精度相一致。然而, 在有效数字的运算过程中, 必须遵循“先进舍, 后运算”的原则。例如

1. 在加减计算中, 各数所保留的小数点后的位数, 应与所给各数中小数点后位数最少的相同。

如将23.62, 0.0083, 1.643三数相加时, 应根据上述取舍规则进行数据处理后再相加。应写成

$$23.62 + 0.01 + 1.64 = 25.27$$

2. 在乘除计算中, 应以有效数字最少的或百分误差最大的数字为准, 对其它各数值按上述取舍规则处理后, 再进行乘除运算; 所得积或商的精确度也不应大于相乘、除各数值中精确度最小的数值的精确度。

如将0.0121, 25.6432, 1.0578三数联乘时, 应先舍弃或进位处理后, 再进行联乘。写成

$$0.0121 \times 25.6 \times 1.06 = 0.328$$

3. 在对数计算中, 真数与对数的有效位数应相同。

如

$$n = \log_a b$$

即 对数 n 准确到几位, 真数 b 亦可准确到第几位。

4. 在计算均值时, 若为四个或多于四个数相平均, 则平均数的有效位数可增加一位。

5. 对常数 π , e 及 $\sqrt{2}$ 、 $\sqrt{3}$ 等的有效数字位数, 可以根据需要任意确定。

6. 界限数值不得修约。

通常表示数量界限的方法, 有以下几种:

1. 不大于 A , 表示从0— A 范围内均为合格(包括等于 A)。大于 A 者为不合格。如合金结构钢中, 残余Cr, Ni含量规定为各不大于0.35%。上述 A 或0.35%, 即为界限数值, 0—35%是非界限数值。

界限数值不得修约, 主要应视有效数字后的数值的位数而定。若在界限数值后的第一位数出现任何非零的数, 如0.351%, 则认为不符合标准。

但是, 若在界限数值后第二位数出现非零的数, 则不在此限。如0.3509%, 可视为0.35%。

2. 不小于 A , 表示从 A — ∞ 范围内均为合格(包括等于 A), 小于 A 者为不合格。

如35碳素结构钢的延伸率(δ_5)规定为不小于20%, 上述 A 或20%即为界限数值, 大于和等于20%为合格范围, 大于20%为非界限数, 可以修约。当出现19.8%(<20%)时, 应视为不合格, 不得修约成20%。

3. 在 A — B 范围内均为合格(包括等于 A 和 B), 小于 A 或大于 B 者为不合格。如普碳钢B2F钢的含碳量为0.09—0.15%, 上述 A , B 或0.09%, 0.15%为界限数值。同理, 若出现0.089%, 不得修约成0.09%, 0.151%也不得修约成0.15%。但是, 当出现0.1509%时, 按

前述原则应视其为0.15%。

4. $x \pm \frac{A}{2}$ 表示 $(x-B) - (x+A)$ 范围内均为合格, 小于 $(x-B)$, 大于 $(x+A)$ 者为不合格。如 $\phi 19\text{mm}$ 的热轧圆钢普通精确度的允许偏差为 ± 0.5 , 其界限数值则为 $\phi 19 \pm 0.5$, 即 $\phi 18.5 - \phi 19.5\text{mm}$ 。如果出现 $\phi 19.51\text{mm}$ 时, 应判为超差。但是, 如果出现 $\phi 19.509\text{mm}$ 时, 应视为 $\phi 19.5\text{mm}$ 。

§1-5 数据的表示方法

试验数据是实验信息与结果的记录, 要准确、简明、形象地表示出来, 通常有三种方法。即列表、作图和经验公式。

一、列表法

列表法的优点在于, 简单易作, 简明紧凑, 便于比较。

一般常用的有统计式、定性式、定量式及函数式。后两种用的较多。

根据试验数据列表时, 应注意以下几点:

1. 表的名称与项目要简明, 必要时可在表名下或表下加附注说明数据来源; 编列的表号应写在表名之前; 表中的项目应包括名称及单位, 一般应采用符号表示之, 表中主项代表自变量, 副项代表因变量。

2. 数字的写法应注意整齐统一, 正确。同一竖行的数字, 小数点要上下对齐; 数字为零时, 要注意有效数字位数, 如有效数为小数点后两位, 则应写为 0.00; 数字太大或太小时, 为避免对有效数字位数发生误解, 应乘数 10^n (n 为正或负整数) 表达小数点的位置。

3. 有效数字应取舍适当。在表达函数形式的表中, 一般均假定自变数没有误差, 所以应尽量取整。因变数的位数则取决于试验数据本身的精确度。

4. 列表时, 自变数应取整数或其它较方便的数值, 按递增或递减的顺序排列。自变数的间距一般应尽可能采用 1, 2, 或 5 乘 10^n 的数值。间距的数值要选得适当, 如过大, 使用时需要的插值多, 既不方便, 又不准确; 如太小, 又过于繁琐。

如试验数据的自变数与因变数都不够规则, 不便应用。可对数据先进行分度整理, 一般常用的方法是图解法: 先将未分度的原始数据描于坐标纸上, 划出通过各点的光滑曲线; 然后从曲线上读出自变数作等间距变化时的数值与相应的因变量值, 再以此数据列表表示之。

二、作图表示法

作图法优点在于, 形象简明, 便于直观比较。

作图法的坐标有直角坐标、单对数坐标、双对数坐标、三角坐标、极坐标及立体坐标等。最常用的是直角坐标。

以直角坐标作图时, 应注意以下几点:

1. 坐标分度

横轴 (x 轴) 代表自变数, 纵轴 (y 轴) 代表因变数。分度一般应采取 1, 2, 4 或 5 最方便, 应避免使用 3, 6, 7, 和 9。

坐标分度的起点不一定为零。通常应使图形占满整个坐标纸。

一般应使坐标纸的最小分格相应于试验数据的精确度。

可能时, 应尽量将变数加以变换, 使所得图形为直线或近似直线, 便于使用。

两坐标轴所代表的项目名称、符号和单位应分别标出，坐标轴的分度也应适当地标明。

2. 将试验数据在已作好分度的坐标纸上描点时，应能表示出试验数据的误差范围。

曲线① 各矩形边长分别代表自变数与因变数的误差范围，其中心则为测量数据的平均值。（见图1-5-1）。

曲线② 表明两变数的误差范围相同。描点为圆，其半径代表测量误差的范围，中心为测量数据的平均值。

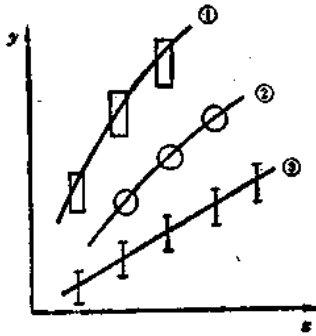


图 1-5-1 描点方法示意图

曲线③ 表明自变数无误差，或误差可以忽略不计，只有因变数有误差，则描点时以一垂直线段表示之，线段长度表示因变数的误差范围。

如同一坐标纸内，描绘几条曲线时，不同曲线的描点应用不同画法予以区别。

3. 根据各描点作曲线时，如数据过少不足以确定自变数和因变数间关系时，最好将各点用直线连接。如描点足够多，描出光滑连续曲线，不必通过所有各点，特别是两端的描点，但是应使曲线尽可能地与所有各点相接近，并使两边的点数及点与曲线间距离乘积的总和近于相等。

三、经验公式表示法

当作出表示数据的曲线后，根据曲线的形状，大多数的曲线可以进一步用一个方程式（或称为经验公式）来表示曲线或试验数据。以下介绍一些常见的经验公式及其选择，以及其中常数的求法等。

1. 经验公式的选择：一般没有简便的方法可获得一个理想的公式。通常是依据解析几何原理和经验来推测公式应有的形式。曲线类型参见本书§8-1。

2. 经验公式中常数的求法：经验公式中常数的求法有图解法、选点法、平均法、最小二乘法等很多种，最常用的是图解法和最小二乘法。前者简单易作，一般只适合于直线关系式；后者精确度最高，将在第二篇中详细介绍。以下简单介绍第一种方法。

直线图解法：凡经验公式为直线式或经变换后为直线式者均可用此法。直线式均可简化为 $y = a + bx$ 的形式。这样，如图1-5-2所示，在直角坐标纸上绘出的直线，其斜率 $\Delta y / \Delta x$ 或 $(y_2 - y_1) / (x_2 - x_1)$ 就等于所求常数 b ，而直线与纵轴（ y 轴）的截距就等于常数项 a 。用此法求得常数的精确度一般为 0.2%—0.5%，即其有效数字为三位，已可满足一般要求。

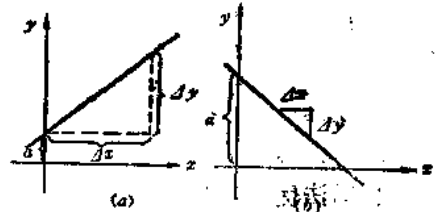


图 1-5-2 直线图解法示意图

第二章 数字特征的计算与检验

数据是实验的信息，是事物内在规律性的外部表现。研究、处理数据是为了认识事物的内在规律，以促进科学研究和指导生产。

事物的本质，现象的内在规律，往往隐藏在大量的数据之中。因此，对数据做科学的整理与分析，去粗取精，去伪存真，方能得到由表及里的科学认识。

例如，对某钢厂生产的25SiMn结构用钢的含硅量进行了检验，其含硅量数据如表2-1-1。

表 2-1-1

0.855	0.83	0.77	0.805	0.81	0.80	0.785	0.82	0.815	0.81
0.81	0.87	0.82	0.78	0.795	0.805	0.87	0.81	0.77	0.775
0.77	0.78	0.77	0.77	0.77	0.71	0.95	0.78	0.81	0.79
0.795	0.77	0.76	0.815	0.80	0.815	0.835	0.79	0.90	0.82
0.79	0.815	0.79	0.855	0.76	0.78	0.825	0.75	0.82	0.775
0.725	0.825	0.81	0.805	0.825	0.89	0.805	0.86	0.82	0.82
0.775	0.84	0.84	0.84	0.81	0.81	0.74	0.775	0.78	0.80
0.74	0.775	0.75	0.79	0.85	0.75	0.74	0.71	0.88	0.82
0.76	0.85	0.73	0.78	0.81	0.79	0.77	0.78	0.81	0.87
0.83	0.65	0.64	0.78	0.75	0.82	0.80	0.80	0.77	0.81
0.75	0.83	0.90	0.80	0.85	0.81	0.77	0.78	0.82	0.84
0.85	0.84	0.82	0.85	0.84	0.82	0.85	0.84	0.78	0.78

可以看出，这批数据有两方面特征：

1. 波动性

在同样条件下生产出来的25SiMn钢的含硅量并不完全一样，表现出一定的波动性。

2. 规律性

数据虽然有波动。但是并非杂乱无章。而是呈现出一定的规律性。可以发现，数据是在0.64—0.95之间波动，而在0.775到0.835的范围内，集中了较多的数据，越往两边数据越少。如在同样生产条件抽取若干批子样，仍出现这种规律性。如表2-1-2，图2-1-1所示。

表 2-1-2 25 SiMn 钢样本频数、频率分布表

分 组	0.635	0.655	0.675	0.695	0.715	0.735	0.755	0.775
	—0.655	—0.675	—0.695	—0.715	—0.735	—0.755	—0.775	—0.795
频 数	2	0	0	2	2	8	13	23
频 率(%)	1.6	0	0	1.6	1.6	6.6	10.9	19.3

分 组	0.795	0.815	0.835	0.855	0.875	0.895	0.915	0.935
	-0.815	-0.835	-0.855	-0.875	-0.895	-0.915	-0.935	-0.955
频 数	24	21	14	6	2	2	0	1
频 率(%)	20	17.5	11.9	5	1.6	1.6	0	0.8

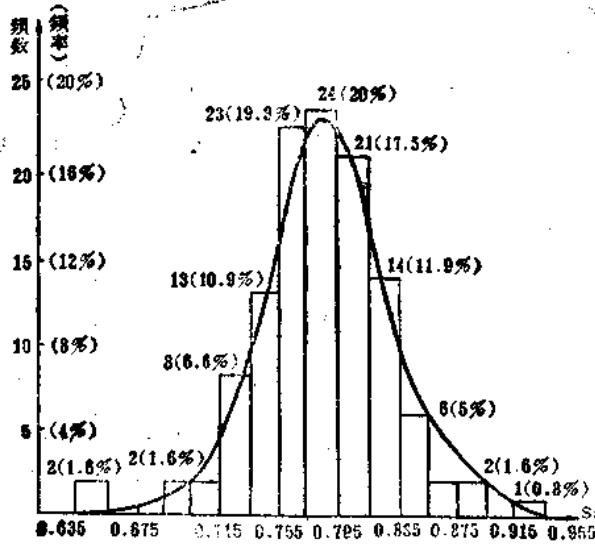


图 2-1-1 频数(频率)直方图

可以看出,中间有一峰值,两边逐渐减小。硅含量在[0.775, 0.835]内最多,约占56.8%,而硅含量大于0.915或小于0.675的只占2.4%

为了便于研究实验数据的数字特征,常常把积累的大批数据概括起来表示如下:

$$x_1, x_2, x_3, \dots, x_i, \dots, x_n$$

为了浓缩、简化实验数据中的信息,使其变得更加简明,易于处理。所以必须对数据的数字特征参数,诸如位置特征、散布特征、分布特征及相关特征的参数进行计算。

§2-1 几个数理统计的基本概念

一、总体、个体和子样

在概率统计计算中,把具有同一性质元素组成的集合体称为**总体**。组成总体的最小单位称为**个体**。例如,在实验中测量某一物理参量(强度、重量、阻值等)的全部测量值,组成所要研究的总体,而每次实验的观测数据 x_i 就是一个**个体**。

通过实验测试,所得到总体的部分数据,称为**子样**(又称**样本**)。子样是由个体所组成的。一个子样所包含个体的数目,称作子样容量,简称为**样本量**。

只要子样能够反映总体的特性,就可以根据抽取的子样,对总体的分布情况进行推断。也就是说,要通过样本来认识总体。因此,抽取子样要有两个要求,即有代表性和独立性。

二、事件与随机变量

通常人们把在一定条件下发生的现象、状态及测试结果称作**事件**。

在一定条件下必然发生的或不可能发生的事件分别称作**必然事件**或**不可能事件**。统称为**确定性事件**。

在一定条件下,一次实验中可能发生也可能不发生,而在大量重复试验中具有某种规律性的事件称为**随机事件**。

在自然界与社会中,到处存在着具有偶然性、不确定性,而内部确有其固有规律性的现象,称为**随机现象**。

随机变量是用来描述随机现象的。

为了研究随机事件的数量规律性而引入的特征量 X ，它在随机事件发生和不发生时取不同的数值，我们把 X 称为随机变量。

物理量的实际量测的可能值的全体，都是随机变量。对于某一次测量，随机变量取该次量测值是一个随机数。

随机变量分为两类：

1. 离散型随机变量——随机变量 X 的取值是不连续的，可排列为 x_1, x_2, \dots, x_i (x_i 可以是有限的或无限的)。例如产品的合格品与废品，材料的相组成，血液中的白血球数，……等都是离散型随机变量的例子。

2. 连续型随机变量——随机变量 X 的取值，可能是数轴上某个区间或是整个数轴。例如，测量误差，铸件的浇铸温度，分子运动速度等都是连续型随机变量的例子。

描述随机变量通常用概率分布和概率密度。

三、正态分布

大量的实验资料表明，测量的随机误差是服从正态分布的。例如，同一牌号的合金钢的成分或它的抗拉强度 (σ_s)，都是服从正态分布的。所谓正态分布就是指随机变量 X 的概率密度函数 $f(x)$ 为：

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right], \quad -\infty < x < +\infty$$

式中参数 μ 与 σ^2 分别是正态分布的数学期望与方差， e 是自然对数的底 ($e = 2.718$)

当 $\mu = 0$ 时， σ 取不同值及 $\sigma = 1$ 时， μ 取不同值时的正态分布密度曲线，参见图 2-1-2，2-1-3。

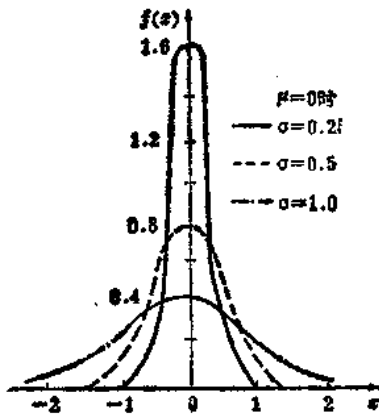


图 2-1-2

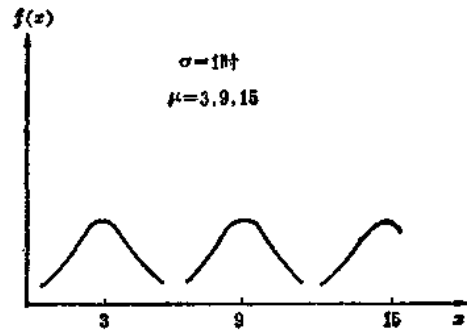


图 2-1-3

当 $\mu = 0$ ， $\sigma = 1$ 时得到的正态分布称为标准正态分布，其分布密度函数 $f(x)$ 与分布函数 $F(x)$ 表达式如下：

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$$

对于任何正态分布，只要进行变换，令 $\frac{x-\mu}{\sigma} = y$ ，就可以化成标准正态分布。标准正态分布 $F_0(x)$ 的函数值已制成表 2-1-3，为计算提供了方便。

例 为了制订某钢种的抗拉强度 (σ_b) 的合格标准, 统计了一批过去的生产数据。得出 σ_b 的均值 $\bar{x} = 60.5$, σ_b 的标准离差 $s = 3.74$, 欲使钢的废品不超过1%, σ_b 的标准值应订为多少?

设 σ_b 的标准数值为 x , 要使废品率 $\leq 1\%$, 就是要保证生产中实际 σ_b 值小于 x 的钢不超过总量的1%。

化成标准正态分布:

$$F_0(y) = F_0\left(\frac{x - 60.5}{3.74}\right) \leq 0.01$$

查表

$$F_0(2.33) = 0.9901$$

即
因而

$$F_0(-2.33) = 1 - 0.9901 = 0.0099 < 0.01$$

$$\frac{x - 60.5}{3.74} = -2.33$$

\therefore

$$x = 60.5 - 2.33 \times 3.74 = 51.79$$

就是说标准抗拉强度 σ_b 订为 51.79 kg/mm^2 , 可以使废品率控制在1%以内。

表 2-1-3 标准正态分布函数 $F_0(x)$ 的函数值

x	$F_0(x)$	x	$F_0(x)$	x	$F_0(x)$
0.00	0.500000	1.05	0.853141	2.10	0.982136
0.05	0.519939	1.10	0.864334	2.15	0.984222
0.10	0.539828	1.15	0.874928	2.20	0.986097
0.15	0.559618	1.20	0.884980	2.25	0.987776
0.20	0.579260	1.25	0.884320	2.30	0.989276
0.25	0.598706	1.30	0.903200	2.35	0.990613
0.30	0.617911	1.35	0.911492	2.40	0.991802
0.35	0.636831	1.40	0.919243	2.45	0.992957
0.40	0.655422	1.45	0.926471	2.50	0.993790
0.45	0.673645	1.50	0.933193	2.55	0.994614
0.50	0.691463	1.55	0.939429	2.60	0.995339
0.55	0.708840	1.60	0.940251	2.65	0.995975
0.60	0.725747	1.65	0.950528	2.70	0.996533
0.65	0.742154	1.70	0.955434	2.75	0.997020
0.70	0.758036	1.75	0.959941	2.80	0.997445
0.75	0.773373	1.80	0.964070	2.85	0.997814
0.80	0.788145	1.85	0.967843	2.90	0.998134
0.85	0.802338	1.90	0.971283	2.95	0.998411
0.90	0.815940	1.95	0.974412	3.00	0.998650
0.95	0.828944	2.00	0.977250		
1.00	0.841345	2.05	0.979818		