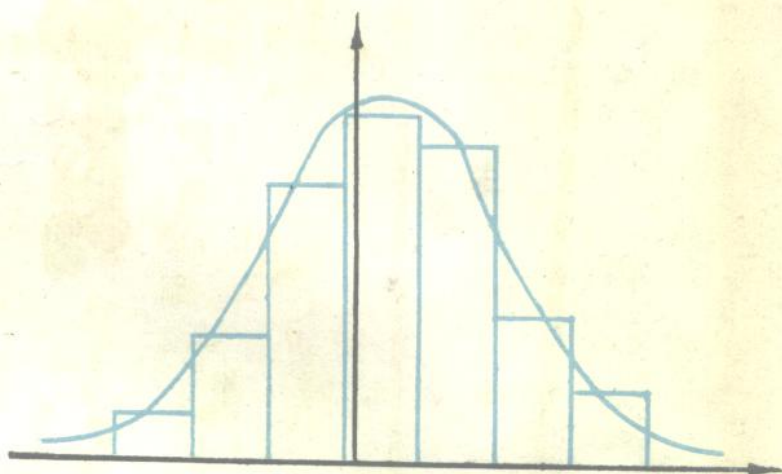


数据的统计处理和解释

—— 国家标准中若干统计方法

全国统计方法应用标准化技术委员会
标准化中统计方法应用分委员会 编



浙江大学出版社

359607

数据的统计处理和解释

——国家标准中的若干统计方法

全国统计方法应用标准化技术委员会
标准化中统计方法应用分委员会 编



浙江大学出版社

(浙)新登字10号

内 容 提 要

从生产过程中和科学试验中获得的数据, 如何处理? 其结果又作如何解释?

本书以十二个国家标准为基准, 给出了一套从简单到复杂的数据处理和解释的方法, 同时给出相应的理论根据和应用实例。

本书面向工矿企业、农、林、医等各条战线从事质量管理和标准化工作的工程技术人员和管理干部。由于本书给出的大部分内容在一般教科书上是不作介绍的, 因此本书也是大专院校、中等专业学校师生有益的教学参考书, 也可作为理工科大学的选修课教材。

DU98/06

数据的统计处理和解释

——国家标准中的若干统计方法

全国统计方法应用标准化技术委员会 编
标准化中统计方法应用分委员会

责任编辑 陈子饶

浙江大学出版社出版

浙江大学印刷厂印刷

浙江省新华书店发行

开本: 787×1092毫米 1/16 印张: 13.7 字数: 306千字
1991年8月第1版 1991年8月第1次印刷

印数: 1—3500

ISBN 7-308-00857-6

O·104 定价: 5.50元

编 者 的 话

随着我国经济建设的飞速发展,数理统计在工农业生产中得到了越来越广泛的应用。特别在以提高产品质量、增加经济效益为目的而开展的全面质量管理活动中,成为极重要的一部分。在这个新形势下,近几年国家技术监督局陆续颁发了一批统计方法应用的国家标准,以推进统计方法在工农业生产中的实际应用。为此,全国统计方法应用标准化技术委员会要求我们为更好地推广使用国标而编写一本带有指南性质的指导书。

本书介绍了有关数据处理和解释的十二个国家标准,并独立、完整地给出了一套从简单到复杂的数据处理和解释的方法,同时给出相应的理论根据和应用实例。旨在直接为工矿、企业、农、林、医等各条战线上从事质量管理和标准化工作的工程技术人员和管理干部掌握、使用国家标准服务。

本书给出的某些内容,如正态性检验,随机性检验,异常值的判断与处理;二项分布与泊松分布的参数估计、假设检验;检验的功效;数据的整理、数字修约规则等,这些在一般教科书上是不作介绍的,因此本书也是大专院校、中等专业学校师生有益的教学补充参考书。

本书文字力求深入浅出、通俗易懂。从应用角度,也能为具有高中以上文化程度,在产品全面质量管理(TQC)第一线工作的人员查用。

本书由全国统计方法应用标准化技术委员会委员、标准化中统计方法应用分委员会主任委员、武汉大学管理学院院长张尧庭教授审稿并作序。

在本书的编写和出版过程中,得到了全国统计方法应用标准化技术委员会和中国标准化与信息分类编码研究所的大力支持,对此我们表示衷心的感谢。

由于编者水平有限,书中难免有不妥或错误之处,欢迎读者批评指正。

编 者

1990年12月

序

标准化工作需要统计方法，从事标准化工作的同志经常要和大量的数据打交道，而对这些数据的整理分析和解释都离不开统计方法，因为不论来自哪门学科的数据都需要用统计方法去处理并加以解释。当今世界上的一些发达国家在科技和生产方面，对数理统计方法应用于标准化工作是极为重视的，他们有一批专业的数理统计工作者参与了标准化机构的工作，并且作出了显著的成绩。81年11月我国成立了全国统计方法应用标准化技术委员会，专门从事研究标准化工作中统计方法的应用，几年来配合国家技术监督局陆续制订了一批统计方法应用方面的国家标准，作出了一定成绩。

国家技术监督局和国际标准化组织 International Organization for Standardization (简称ISO)建立了经常联系，1980年国际标准化组织在ISO/PLACO中重申ISO/TC69 (ISO的第69技术委员会，即统计方法的应用技术委员会)的作用就是在国际标准化组织范围内作为统计方法的指导者和合作者，并建议制订一份总的指南帮助大家选择和修改统计方法的应用。后来，由于多种附加标准的发展，以及ISO/TC 69提供的咨询服务，使得统计方法的应用情况有了显著的改善。目前，国际标准化组织亦准备写一本统计方法在标准化中的应用指南，其中将详列ISO/TC 69的一般标准，并列举它们在各种工业部门中应用的例子，以便供与标准化和规范化有关的个人、技术委员会和设计机构参考。

我国的标准有些是从国际标准演变而来的，既结合我国的实际状况，又能促进我国的工业向国际先进水平靠拢。推广国标的使用对于提高产品质量、企业管理水平有指导、监督、调整…等作用。

本书是一本既介绍国标又给出相应理论根据和应用实例的通俗书。第一、二、三章由上海财经大学余澄扬撰写；第四、五章由浙江大学舒鹏撰写；第六、七章由安庆师院邓其封撰写。

作为一本指南性质的书，它的出版，相信能为促进数据处理统计方法国家标准的推广使用起到积极的作用。

张尧庭

1990年12月

目 录

第一章 数据的一般统计解释	1
§1-1 统计数据整理的意义	1
§1-2 统计方法研究的内容	1
§1-3 表达实验结果的规则	1
§1-4 统计资料的分组	2
一、频数分布 二、分组的规则	2
§1-5 分组的实例	3
§1-6 图示法	7
一、直方图 二、频数分布曲线 三、累积分布图	7
§1-7 数字修约规则	9
第二章 统计中的两个重要特征数	11
§2-1 集中趋势	11
一、算术平均数 二、中位数	11
三、众数 四、四分位数、十分位数和百分位数	15
§2-2 离散趋势	19
一、极差 二、平均差 三、四分位差	19
四、 n 个观察值的方差 五、 n 个观察值的标准差 六、变异系数	20
第三章 几种常用的统计分布	25
§3-1 常用的离散型随机变量及其分布	25
一、二点分布 二、二项分布	25
三、泊松分布 四、超几何分布	27
§3-2 常用的连续型随机变量及其分布	30
一、均匀分布 二、正态分布 三、 χ^2 分布	30
四、 t 分布 五、 F 分布 六、对数正态分布	36
七、指数分布 八、 Γ 分布 九、威布尔分布	38
十、分位数	40
第四章 常用统计分布的参数估计	42
§4-1 点估计	42
一、点估计中的常用术语	42
二、总体 X 的概率分布服从正态分布 $N(\mu, \sigma^2)$ 时, 总体均值 μ 的常用估计量	43
三、总体 X 的概率分布服从正态分布 $N(\mu, \sigma^2)$ 时, 总体方差 σ^2 的常用估计量	45
四、总体 X 的概率分布服从二项分布 $B(n, p)$ 时, 总体参数 p 的估计量	48
五、总体 X 的概率分布服从泊松分布 $\pi(\lambda)$ 时, 总体参数 λ 的估计量	49
六、极大似然估计	49
§4-2 区间估计	52
一、区间估计中的常用术语	52

二、单个正态总体均值 μ 与方差 σ^2 的置信区间	53
三、二个正态总体均值差的置信区间	55
四、二个正态总体方差之比的置信区间	56
五、总体为二项分布 $B(n, p)$ 时, 参数 p 的置信区间	57
六、总体为泊松分布 $\pi(\lambda)$ 时, 参数 λ 的置信区间	59
七、应用示例	59
八、样本大小 n 的确定	64
第五章 常用统计分布的参数检验	67
§5-1 假设检验中的常用术语	67
§5-2 检验公式	69
一、单个正态总体均值与方差的检验	69
二、两个正态总体均值差及方差之比的检验	71
三、二项分布参数 p 的检验	73
四、泊松分布参数 λ 的检验	80
§5-3 样本大小已事先确定时, 对假设进行检验的工作步骤	82
§5-4 应用示例	84
§5-5 区间估计与假设检验的关系	89
§5-6 检验的功效	90
第六章 正态性检验和随机性检验	106
§6-1 正态性检验	106
一、无方向检验	106
二、有方向检验	115
三、正态概率纸的应用	123
四、数据的变换	125
§6-2 随机性检验	125
第七章 正态样本异常值的判断与处理	130
§7-1 当 σ 已知时使用奈尔(Nair)检验法	130
一、单侧情形	130
二、双侧情形	133
§7-2 当 σ 未知时的检验方法	133
一、发现异常值的个数 ≤ 1 的情形	133
二、发现异常值的个数 > 1 的情形	139
附表	143
一、正态分布函数表	143
二、正态分布分位数表	144
三、 χ^2 分布分位数表	145
四、 t 分布分位数表	151
五、 F 分布分位数表	153
六、泊松分布函数表	207

第一章 数据的一般统计解释

§1-1 统计数据整理的意义

从事科学研究、经济管理和标准化工作的同志经常要和大量的数据打交道，而对这些数据的整理、分析和解释都离不开统计方法，因为不论哪门学科的数据处理和解释都需要一种统计方法。

统计是在面对不确定性的情况下作出决策的一种方法，它是研究数字资料的整理、分析和正确解释的一门学科。

人们各自从不同的来源取得各种数字资料，例如在日常生产活动中，各部门都有大量的原始资料，但这些资料通常都是杂乱无章的，必须经过整理和简缩才能加以利用，加以研究。在工矿企业、研究所等单位都拥有大量的试验数据，如果不经过整理和处理，它们无异是一堆废纸。但是，一旦使用完善的统计方法就可使数据整理、排列得有条有理，用图形或少量的几个重要参数，就可把一大堆数据的特征表达出来，同时既可避免不正确的解释，又可将获得满意数据的成本降低到最低限度，提高了经济效益，好处是十分明显的。

当这些数据经过整理和简缩而成一种适当的形式以后，就为进一步的统计分析工作奠定基础。

数据的一般统计解释包括方法和技术两个方面，内容十分广泛，是统计工作者研究的课题。

§1-2 统计方法研究的内容

统计方法的研究分三个部分，是有机统一体。它们是

1. 对总体的研究，以确定总体的聚集性质；
2. 对变异的研究，也就是除了要研究总体的集中趋势外，还要研究总体中各个个体之间的差异或离散程度；
3. 将一大批数据简缩为容易理解的形式。

在日常工作中，即使有可能给出全部的观测值，我们也宁可把它化为简缩的和容易理解的形式，给出主要的信息，以便讨论。在标准化工作中，要求用一种为大家所公认的方法来进行，以避免各搞各的一套，而且还要保证所采用的方法遵循在实践中被认为是最满意的方法。

§1-3 表达实验结果的规则

日常的科研工作中，在表达实验结果时，一般有以下一些规则：

1. 除非受到篇幅的限制, 一般实验的结果一定要详细报告, 如果观测的次数是大量的, 则可以把它放在一个附录里, 以避免妨碍分析和讨论。

2. 如果观测的次数大于30, 我们可以用频数分布来表达它的结果。关于将观测所得的数据进行分组和列表的方法将在后面讨论。

3. 在实验结果中, 一般要给出下列三个数字资料: (1)算术平均数; (2)标准差; (3)观测次数。不管实验结果中是否列出分布表, 上面这三个重要信息的数字资料必须列出。

4. 如果可能的话, 也要列出下列两个信息:

(1) 总体平均数的95%置信界限;

(2) 总体标准差的95%置信界限。

5. 最后应仔细检查一下, 观测值是否能安全地被修约, 如果这个数据以后仍有可能被用到, 并且要保持原始的状态被使用, 那么最好不要进行修约, 关于数字的修约规则见本章§1-7。

§1-4 统计资料的分组

一、频数分布

简缩数据的最简单方法就是分组。如果观测值的个数较多, 或观测值有重复时, 经常采用分组的办法来处理, 把个别差异不大的观测值放在同一个组里面。这样我们就可把全部数据用一张分布表来替代, 表格中显示出落在每个组中的原始观测值的个数。

“频数分布”在一般教科书上也称为“次数分布”, 英文称为 frequency distribution, 这里根据国标GB3358-82《统计学名词和术语》的规定而采用“频数分布”。

当我们把统计资料分组以后, 各组数值的变动范围就叫做组距(class interval), 组距的中间值称谓组中值(mid-value), 组距的终端数值称谓组限(class limit), 一个组里面的最小值称谓该组的下限(lower limit), 最大值称谓该组的上限(upper limit)。落在各组内的观测值个数称谓该组的频数(frequency)。

二、分组的规则

1. 组距大小的确定

组距的选择应该恰当, 通常是把全部观测值大致分成10至20个组, 以组距相等为好。有一个称为斯塔其(H.A. Sturge)的公式, 介绍分组组距(I)的计算, 可供参考, 即

$$I = \frac{\text{最大观测值} - \text{最小观测值}}{1 + 3.322 \lg N}$$

上式中, N 是全部观测值的个数, 即总的观测次数, $\lg N$ 是 N 的常用对数。

2. 组限的选择

组限的选择应该注意这样一个原则: 某个具体的观测值应归入哪一个组应该是十分明确, 不能模棱两可。这就是说, 组限不应是可能出现为观测值的数值。最满意的安排就是: 把两个相邻组之间的组限置于上一组的最高观测值和下一组的最低观测值的正中间。于是, 通常把组限表示成比记录观测值所用的数字多取一位小数, 例如表1-1中观测值带

一位小数，而表 1-2 中的组限值带两位小数。必须注意，需要进行分组的观测值事先不要进行修约，因为分组是修约的一种形式，而连续的修约容易导致系统误差。

§1-5 分组的实例

实例：对浇筑 7 天后的混凝土块抗压强度的 400 个观测值作统计数据分组。

表 1-1 浇筑 7 天后的混凝土块的抗压强度
(单位: N/mm^2)

24.0	30.4	25.3	29.7	27.0	28.5	36.7	24.5	29.3	30.4	28.7	26.2	23.0	27.7	30.4	26.2
23.8	27.9	21.0	27.7	25.7	23.4	25.5	30.8	28.7	25.8	27.0	28.8	30.7	28.1	28.3	29.0
28.8	31.1	28.8	26.8	27.2	22.3	33.0	43.6	25.9	25.3	25.8	23.2	27.7	26.6	29.7	31.7
28.8	27.7	23.6	28.8	28.5	24.3	26.6	33.0	39.1	30.3	31.0	30.1	29.6	25.3	28.1	23.6
23.2	27.4	29.1	24.7	22.8	26.6	24.0	30.6	30.8	31.2	22.1	18.3	34.8	18.7	25.9	25.8
32.0	27.2	26.4	31.2	27.9	35.4	24.7	25.2	27.9	29.2	27.9	26.0	26.5	28.1	24.7	23.2
27.9	22.4	27.6	27.9	28.3	27.0	24.1	28.4	37.3	27.4	25.0	26.8	24.2	27.0	28.8	27.5
27.6	26.8	31.1	30.3	26.8	27.7	27.4	40.3	37.2	27.2	27.2	24.0	28.9	26.8	33.2	30.5
28.6	24.6	34.3	25.8	27.2	22.5	27.9	24.0	27.9	27.4	28.7	24.9	27.0	29.0	25.8	28.1
28.5	28.3	26.3	27.2	25.9	24.5	29.0	17.0	27.6	28.1	29.0	29.4	30.0	28.3	32.8	25.1
28.1	24.9	28.0	25.5	26.5	30.1	28.6	25.1	28.8	22.3	26.4	26.0	19.1	34.1	30.1	32.2
27.0	23.6	26.5	22.1	27.9	24.5	30.1	25.1	35.5	24.4	27.4	26.3	26.9	19.1	23.9	26.4
27.0	29.1	27.4	30.2	25.6	28.5	31.3	23.0	32.8	26.8	25.5	27.0	24.5	24.5	24.8	21.9
32.4	29.8	26.8	24.1	30.5	29.0	28.9	29.4	30.9	27.4	21.9	34.9	24.7	30.1	24.1	25.3
31.3	24.4	25.5	24.7	22.8	27.4	26.0	26.8	31.5	29.3	22.8	34.2	22.1	22.1	23.5	24.7
28.8	22.5	26.8	26.9	28.3	28.5	27.8	31.2	30.1	29.3	23.4	25.6	30.1	31.7	27.1	23.8
31.5	25.3	27.6	32.1	25.1	29.6	31.6	25.1	26.4	27.2	20.8	30.5	21.6	19.3	27.4	24.9
30.3	25.9	29.4	31.4	23.4	25.1	27.9	31.4	29.6	23.4	32.5	29.2	23.2	29.6	23.4	24.9
29.0	29.2	29.2	26.4	27.2	25.1	25.4	28.8	27.0	27.9	28.8	31.8	22.1	27.0	31.6	23.9
32.6	27.0	23.8	25.6	25.9	37.3	23.4	24.5	30.1	30.2	32.5	32.0	26.6	23.4	28.0	26.9
34.6	28.1	33.9	22.5	26.6	25.8	30.4	24.7	27.6	25.9	27.6	25.1	25.3	20.4	23.3	31.8
26.1	29.4	27.7	23.9	25.3	25.3	19.8	21.2	35.1	32.6	27.9	26.4	28.3	27.0	28.8	30.5
24.3	31.5	27.4	26.0	25.7	23.2	23.2	24.5	28.8	29.4	23.9	24.9	28.5	26.7	26.3	30.0
30.7	32.6	26.8	24.8	27.4	25.1	28.5	28.3	27.0	26.3	27.6	23.6	21.6	26.7	19.8	29.2
30.9	28.5	28.7	29.6	25.9	29.2	23.5	33.3	25.5	22.1	25.4	26.0	21.4	25.4	18.3	31.9

注: N/mm^2 是压强单位; 其中 N 又叫牛顿, $1\text{N} = 1/9.81$ 公斤。

根据表 1-1 所给的数据进行分组时, 首先要决定分多少组, 组距多大。根据上面所介绍的斯塔其公式, 可求得组距应为 $I = \frac{43.6 - 17.0}{1 + 3.322 \lg 400} = 2.76 \text{N/mm}^2$, 故可分为 10 个组。但是为了实际应用和后面举例说明的方便起见, 选用 2N/mm^2 作为合适的共同组距, 这样就可分为 $(44.95 - 16.95)/2 = 14$ 组 (见表 1-2)。因为观测值都记录成带有一位小数 (单位 N/mm^2), 所以组限应表示成带有两位小数, 尾数为 0.05, 如表 1-2 第 1 栏所示。将每一个观测值在相应的组内做上记号, 为了记数方便, 记号可划成 “正” 字形, 每个 “正” 字表示 5 个观测值, 如表 1-2 第 2 栏所示。当每一个观测值都分配到各组以后, 数一下每个组有几个 “正” 字, 就得出各组的频数, 见表 1-2 第 3 栏。

表1-2 浇筑7天后的混凝土块的抗压强度
(观测值的分组法)

混凝土的抗压强度 (N/mm ²)	划 线 计 数	频 数
(1)	(2)	(3)
16.95—18.95	正	4
18.95—20.95	正丁	7
20.95—22.95	正正正正丁	22
22.95—24.95	正正正正正正正正正正正丁	62
24.95—26.95	正正正正正正正正正正正正正正正正丁	87
26.95—28.95	正正正正正正正正正正正正正正正正正正正正正	105
28.95—30.95	正正正正正正正正正正正正	60
30.95—32.95	正正正正正正丁	32
32.95—34.95	正正一	11
34.95—36.95	正	4
36.95—38.95	下	3
38.95—40.95	丁	2
40.95—42.95		0
42.95—44.95	一	1
合 计		400

每张表都应有一个标题，各类的标目应该清楚地用文字说明，为了便于参考，最好将各栏编号。同时，为便于进行工作，每张表都要注明制表日期，以便弄清是什么时间内的数据。

制表时应该明确地标明各组的下限和上限，任一个组的上限就是相邻的下一个组的下限，如果组距是按上述分组规则选择的话，就不会产生混淆，如表1-2所示。除了表1-2所介绍的方法以外，另外还有一种许多人愿意用的方法，这就是每个组可以用归入该组的最高观测值和最低观测值来确定，如表1-3所示。实际上表1-2和表1-3并没有大的区别，只是表达的方法不同而已。

表1-3是确定频数分布组的另一种方法。

表1-3 浇筑7天后的混凝土块的抗压强度

抗压强度 (N/mm ²)	观测次数
(1)	(2)
17.00—18.90	4
19.00—20.90	7
21.00—22.90	22
23.00—24.90	62
25.00—26.90	87
27.00—28.90	105
.....	...
43.00—44.90	1
合 计	400

在实际工作中, 频数可以用实际频数和相对频率 (通常表示为总计的百分数) 来表示, 如表1-4所示。其中

$$\text{相对频率} = \frac{\text{实际频数}}{\text{总频数}} \times 100\%$$

表1-4 表述频数分布的两种方法

(a) 实际频数分布		(b) 相对频率分布	
混凝土块的抗压强度 (N/mm ²)	各组的块数	混凝土块的抗压强度 (N/mm ²)	各组块数的百分数 (%)
(1)	(2)	(1)	(2)
16.95—18.95	4	16.95—18.95	1.00
18.95—20.95	7	18.95—20.95	1.75
20.95—22.95	22	20.95—22.95	5.50
22.95—24.95	62	22.95—24.95	15.50
24.95—26.95	87	24.95—26.95	21.75
26.95—28.95	105	26.95—28.95	26.25
28.95—30.95	60	28.95—30.95	15.00
30.95—32.95	32	30.95—32.95	8.00
32.95—34.95	11	32.95—34.95	2.75
34.95—36.95	4	34.95—36.95	1.00
36.95—38.95	3	36.95—38.95	0.75
38.95—40.95	2	38.95—40.95	0.50
40.95—42.95	0	40.95—42.95	0
42.95—44.95	1	42.95—44.95	0.25
合 计	400	合 计	100.00

为了某种目的, 对小于(或大于)某些特定数值的观测值的次数编制表格也是非常有用的, 用这种形式编制的频数分布表称为累积频数分布表。累积频数分布表是根据频数分布编制的, 其方法如下: 表中从上往下将第二组的频数加上第一组的频数, 记录合计数, 再将第三组的频数加上上述的合计数并记下结果。这样, 依次将每组的频数加上前面的合计数, 将每步的积累数记下来, 直到整个表的频数加完为止, 这种的累积方式, 叫做以下累积, 表示每组的上限以下的累积频数是多少, 如表1-5(I)所示。如果为了取得大于某些特定数值的观测值次数的积累表, 可仍采用累积的形式, 其过程同上, 所不同的只是从最后一组开始往上加, 如表1-5(II)所示, 这样的累积方式, 叫做以上累积, 表示每组下限以上的累积频数是多少。

表1-5 表述累积频数分布的两种方法

(I) 以下累积

(a) 实际累积分布			(b) 百分数累积分布		
抗压强度 (N/mm ²)	频数	小于某上限 的累积频数	抗压强度 (N/mm ²)	频率	小于某上限的累 积百分数 (%)
(1)	(2)	(3)	(1)	(2)	(3)
16.95—18.95	4	4	16.95—18.95	1.00	1.00
18.95—20.95	7	11	18.95—20.95	1.75	2.75
20.95—22.95	22	33	20.95—22.95	5.50	8.25
22.95—24.95	62	95	22.95—24.95	15.50	23.75
24.95—26.95	87	182	24.95—26.95	21.75	45.50
26.95—28.95	105	287	26.95—28.95	26.25	71.75
28.95—30.95	60	347	28.95—30.95	15.00	86.75
30.95—32.95	32	379	30.95—32.95	8.00	94.75
32.95—34.95	11	390	32.95—34.95	2.75	97.50
34.95—36.95	4	394	34.95—36.95	1.00	98.50
36.95—38.95	3	397	36.95—38.95	0.75	99.25
38.95—40.95	2	399	38.95—40.95	0.50	99.75
40.95—42.95	0	399	40.95—42.95	0	99.75
42.95—44.95	1	400	42.95—44.95	0.25	100.00
合 计	400	—	合 计	100.00	—

(II)以上累积(大于某些特定数值的观测值的次数累积)

(a)实际累积分布			(b)百分数累积分布		
抗压强度 (N/mm ²)	频数	大于某下限 的累积频数	抗压强度 (N/mm ²)	频率	大于某下限的累 积百分数 (%)
(1)	(2)	(3)	(1)	(2)	(3)
16.95—18.95	4	400	16.95—18.95	1.00	100.00
18.95—20.95	7	396	18.95—20.95	1.75	99.00
20.95—22.95	22	389	20.95—22.95	5.50	97.25
22.95—24.95	62	367	22.95—24.95	15.50	91.75
24.95—26.95	87	305	24.95—26.95	21.75	76.25
26.95—28.95	105	218	26.95—28.95	26.25	54.50
28.95—30.95	60	113	28.95—30.95	15.00	28.25
30.95—32.95	32	53	30.95—32.95	8.00	13.25
32.95—34.95	11	21	32.95—34.95	2.75	5.25
34.95—36.95	4	10	34.95—36.95	1.00	2.50
36.95—38.95	3	6	36.95—38.95	0.75	1.50
38.95—40.95	2	3	38.95—40.95	0.50	0.75
40.95—42.95	0	1	40.95—42.95	0	0.25
42.95—44.95	1	1	42.95—44.95	0.25	0.25
合 计	400	—	合 计	100.00	—

§1-6 图示法

一、直方图(histogram)

图形表示的资料比用一栏数字的形式所表示的资料更易领会和掌握,通过图形可直观地了解变量分布的变化特征。

频数分布常用直方图表示,这种频数分布图的纵坐标单位是个数,或叫次数,横坐标单位是变量的不同取值。沿着横坐标标出组限,在每个组限上画出一个直立的长方形,其面积与组内的频数成正比。这种图形有一个明显的优点,就是当各个组限间的频数有变动时,图形就会把这种变动反映出来。作图的过程比较简单,就是使长方形的高度与相应组中的频数成正比,表1-1中由于组距都相等,因此可以直接把频数用作直方图的纵坐标刻度。为了克服横坐标和纵坐标的尺度单位不同,避免引起误解,可在相应的长方形上标

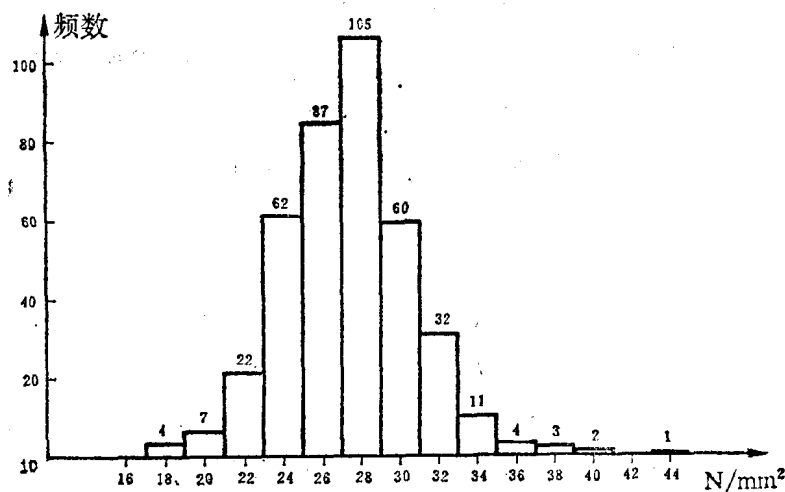


图1-1 400个混凝土块抗压强度的频数分布图

明各组的频数。根据表1-1所作的直方图的图形，如图1-1所示。

如果把上面直方图的纵坐标尺度（即频数）改换为频率，这样就得到相对频率的直方图。频率分布图的纵坐标单位是百分数（频率），横坐标和频数分布图相同，画出的图形和上面一样。

二、频数分布曲线(frequency distribution curve)

在绘制直方图时，是假定各观测数在组内是均匀分布的，虽然真实情况常常并不是这样，但实际上是可以接受的，因此，一般说来，在任何一个组里，最经常发生的倾向就是观测数趋向于组距中接近中心分布的那个点。因此，如果假定观测次数可以大大增加，就会发现分组的组数会增加（即组距缩小），而频数并没有变得不规则。这样，直方图的形状就会趋向于显示出一个比较平滑的轮廓，当组数分得无限多时，就会变成一条平滑的曲线。因此，通常假定任何一个直方图逼近一条曲线，并称这样一条曲线为频数（或频率）分布曲线，或经验分布曲线。

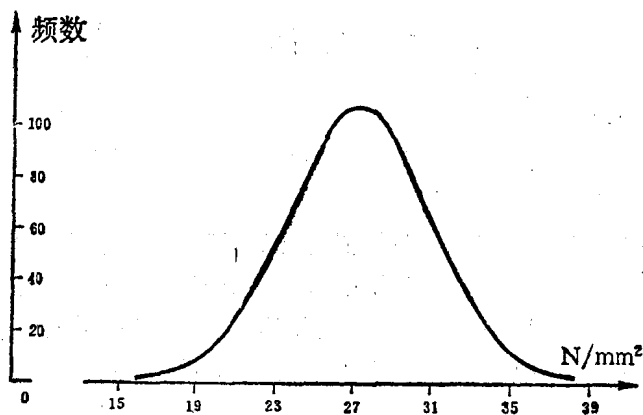


图1-2 400个混凝土块抗压强度的频数分布曲线

从导出频数分布曲线的方法中,可看到和直方图一样,在任何两个纵坐标之间的、曲线之下的面积与这两个纵坐标之间的频数成正比。如将表 1-1 中混凝土块抗压强度的资料画成频数分布曲线,其形状就如图 1-2 所示。分布曲线是数理统计的重要内容之一。

三、累积分布图(cumulative distribution diagram)

除了频数(或频率)分布图以外,还有另外一种分布图,它就是累积频数(或累积频率)分布图。如果要估计小于某个数的次数或百分数,这种图形就特别有用。纵坐标上的刻度表示频数或频率百分数,横坐标上的组限刻度仍与频数分布图一样。累积频数(或频率)是由各组上限的垂线来度量,把这些竖线的顶点联结起来,注意第 1 点连接刻度为 0 的最低频数(或频率),这样就画出了累积分布图。右边的图 1-3 显示了表 1-5(b)的百分数累积图。从图 1-3 中可以容易地看出低于某一观测值的百分数,或者看出有百分之几的观测数低于某值。累积分布图对于理论分布曲线的配制和研究很有用处。

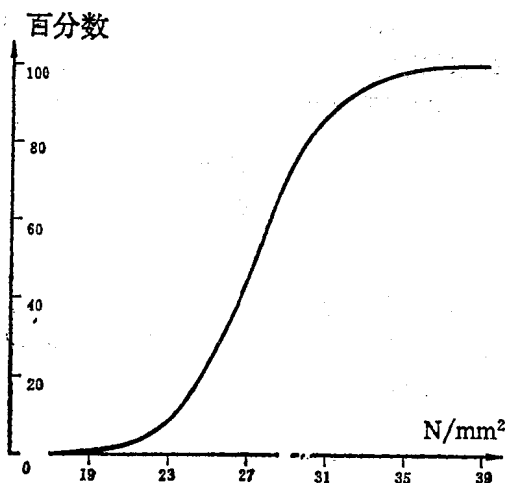


图1-3 400个混凝土块抗压强度的百分数累积分布图

§1-7 数字修约规则

在将数据进行整理和简缩时,往往会碰到需要对某些数字进行修约的问题,因为在各项标准中,以至在日常生产中,要规定各种界限数值,以便衡量技术指标或质量指标是否符合要求。在测量这些技术指标或质量指标时,由于测量、检验手段的限制,所测数据总会有一定的精确范围。此外,在进行数值计算时,也经常会带来误差。因此,对于实际数据应当如何处理,才能表达其精确程度,对于所规定精确程度范围以外的数字,又应如何处理,这些都是至关重要的。国家技术监督局对此作了专门规定,在1981年4月发布了中华人民共和国国家标准GB1.1—81《标准化工作导则 编号标准的一般规定》,其中第7、3、2款和附录 C,对在制订、修订标准中,各种测量、计算的数值需要修约时应遵守的《数字修约规则》作了下列明确的规定:

1. 在拟舍弃的数字中,若左边第一个数字小于 5 (不包括 5) 时,则舍去,即所拟保留的末位数字不变。

例如将 14.2432 修约到保留一位小数:

$$14.2432 \rightarrow 14.2$$

2. 在拟舍弃的数字中,若左边第一个数字大于 5 (不包括 5) 时,则进一,即所拟保留的末位数字加一。

例如将 26.4843 修约到只保留一位小数:

$$26.4843 \rightarrow 26.5$$

3. 在拟舍弃的数字中, 若左边第一个数字等于5, 且其右边的数字并非全部为0时, 则进一, 即所拟保留末位数字加一。

例如将1.0501修约到只保留一位小数:

$$1.0501 \longrightarrow 1.1$$

4. 在拟舍弃的数字中, 若左边第一个数字等于5, 且其右边的数字皆为0时, 所拟保留的末位数字若为奇数则进一, 若为偶数(包括“0”)则不进。

例如将下列数字修约到只保留一位小数:

$$0.3500 \longrightarrow 0.4$$

$$0.4500 \longrightarrow 0.4$$

$$1.0500 \longrightarrow 1.0$$

5. 所拟舍弃的数字, 若为二位以上数字时, 不得连续进行多次修约, 应根据所拟舍弃数字中左边第一个数字的大小按上述规定一次修约得出结果。

例如将15.4546修约成整数:

$$15.4546 \longrightarrow 15$$

不正确的做法是:

修约前	一次修约结果	二次修约结果	三次修约结果	四次修约结果
15.4546	15.455	15.46	15.5	16

上述修约规则, 习惯上称为“四舍六入五单双法”, 这种方法比起原先的“四舍五入法”更为合理。为了便于把这些规则记住, 有些同志编了一些口诀, 例如: “四舍六入五考虑, 五后非零则进一, 五后皆零视奇偶, 五前为偶应舍去, 五前为奇则进一”^{*}。

^{*} 参阅方开泰等著《数理统计与标准化》, 技术标准出版社, 1981年本第23页。