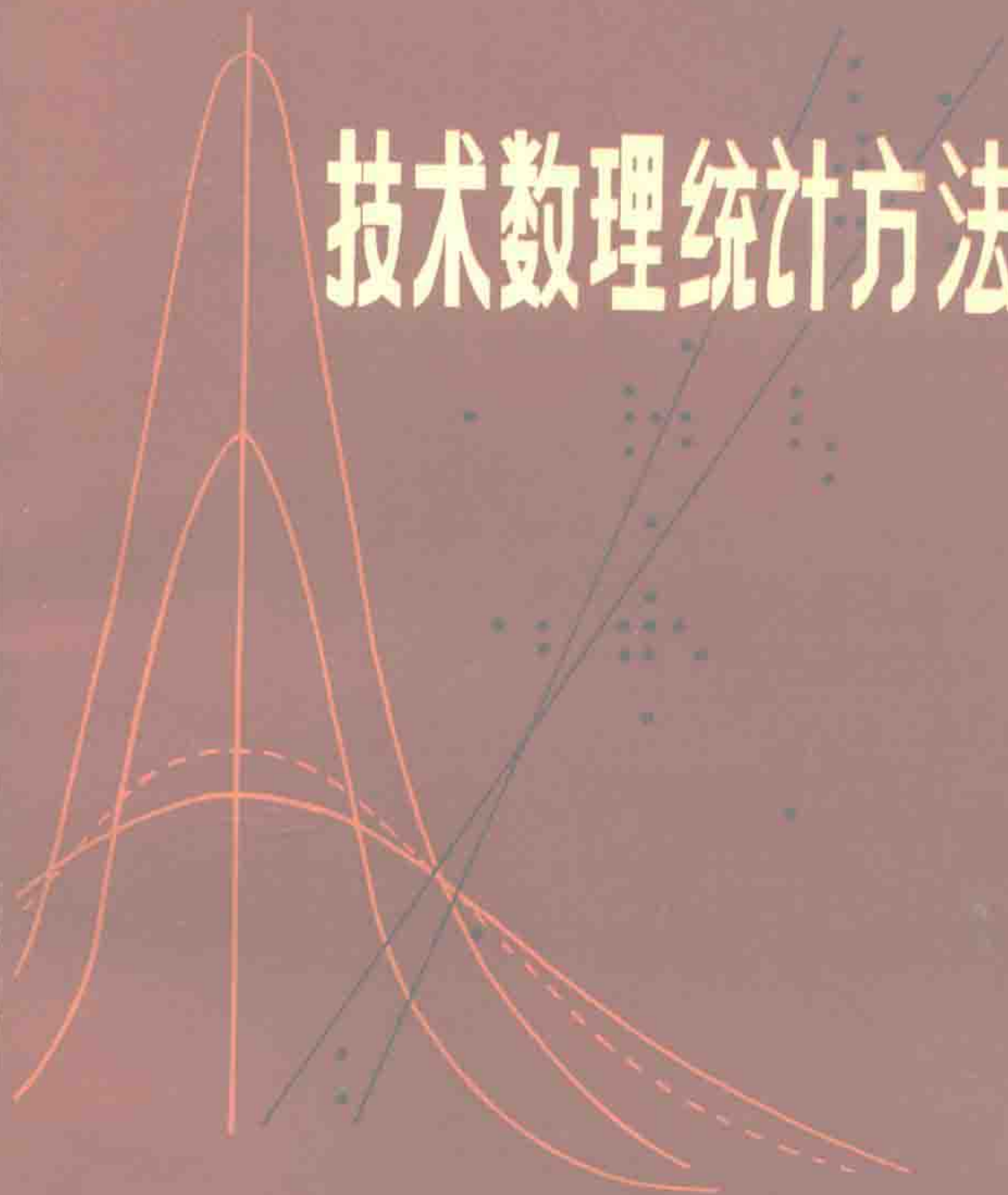


技术数理统计方法



曹 欣 总 编 著
安徽科学技术出版社

技术数理统计方法

(第 二 版)

曾秋成 编著

安 徽 科 技 出 版 社

责任编辑：孙述庆
封面设计：宋子龙

技术数理统计方法

曾秋成 编著

*

(合肥市跃进路1号)

安徽科学技术出版社出版

安徽省新华书店发行

六安新华印刷厂印刷

*

开本787×1092 1/16 印张28.25 字数696,000

1982年7月第1版 1983年11月第2版第3次印刷

印数：5,001—13,000

统一书号：15200·16 定价：2.90元

内 容 提 要

这是一本讲述数理统计方法实际应用的读物。全书共分二十章，第一章讲统计资料的整理，第二、三、四章讲概率统计的基本理论，第五、六、七、八、九章讲各种统计检验方法，第十、十一、十二、十三章讲试验设计，第十四、十五、十六、十七章讲各种回归分析与相关分析，第十八章讲经验分布曲线配线，第十九章讲质量评估与寿命评估方法，第二十章讲数理统计方法在电子计算机上应用的六个主要ALGOL语言程序。各章都附有较多的应用实例，书后还附有较为完全的数理统计用表。文字通俗，内容丰富、实用。便于结合生产与科研自学，也适合教学参考。

编 著 者 的 话

翻阅近代各科学领域的技术文献，我们都能看到许多有关数理统计方法应用的内容。它应用之广，发展之快，以致使最新式微型电子计算器的设计者，在计算器的有限键位上也专门设置了求“平均值”、“标准离差”等多种标准函数键，以满足计算中的实际需要，由此足见数理统计方法的重要性了。的确，近年来电子计算机技术的大发展，如同给数理统计方法插上了翅膀，使它在生产和科研中发挥的作用，越加显得突出。

编写一本便于技术人员学习并掌握数理统计方法的书，在内容的取舍，以至数学原理阐述的逻辑性与严谨性上都是不容易掌握的，因为它既不是一个知识性的科普读物，也不是一种数学专著，它应该成为那些需要掌握数理统计方法来解决自己工作中实际问题的同志们的手册性的工具书，这是我编著这本书的希望。

数理统计方法在生产和科研中的应用面，广如瀚海，因此本书在应用实例选择上，不可能面面俱到。我着重选用了冶金专业方面的实例，这不仅是因为笔者自己较为熟悉，而且在于这个专业领域有应用数理统计方法的巨大潜力，将使这本书多少具有一定的特色和侧重，以免零碎和分散。为使本书对从事生产、科研工作的技术人员和管理干部都能有所帮助，选收的实例专业性并不很强，使它带有相当的普遍性，这样广大读者就可以举一反三，联系自己的专业加以应用。

本书是在原马鞍山市科技局印行的资料《钢铁生产和科研的数理统计方法》基础上，进一步修改而成的。在编著过程中，得到了南京大学、安徽大学和中国科学技术大学三校数学系老师们的很多帮助，其中陈桂景、张守智、杜丹青等同志对初稿进行过详细地审阅，花费了许多精力。杨纪珂教授对我这方面工作一直十分关心，并给予许多具体指导与帮助。他的有关著作我在这本书中也有所引用。着手编著这本书的事，大致可以追溯到1973年初。现在这本书能够问世，是和许多同志与领导的鼓励、支持分不开的，谨此敬致衷心的感谢。

由于水平有限，书中的错误和缺点一定不少，热切希望广大读者批评指正，以便再版修改。

1981年1月

主要符号说明

ξ, η, ζ ——随机变量	C_e ——样本的峰度系数
x, y, z ——随机变量的取值	f ——频率
$\bar{\xi}$ ——随机变量的平均值	f° ——频数
$\bar{x}$ ——样本的平均值	$P(A)$ ——事件 A 出现的概率
$\bar{u}$ ——总体的平均值	$F(x)$ ——随机变量 ξ 的概率分布
$E(\xi)$ ——随机变量的数学期望	$p(x)$ ——连续随机变量 ξ 的概率密度
$D(\xi)$ ——随机变量的方差	$f(x)$ ——样本的频率函数
s^2 ——样本的方差	$F_N(x)$ ——样本的累积频率函数
σ^2 ——总体的方差	N, n ——样本容量
σ_{ξ}^2 ——作为随机变量看待的平均值的方差	ν ——自由度
μ'_r ——随机变量的 r 阶原点矩	α ——显著性水平
μ_r ——随机变量的 r 阶中心矩	λ ——舍弃区间的横坐标
ν_r ——随机变量的 r 阶定点矩	H_0 ——解消假设(零假设)
γ_r ——随机变量的 r 阶标准中心矩	H_A ——备择假设
m'_r ——样本的 r 阶原点矩	t ——标准化变量、司都顿函数
m_r ——样本的 r 阶中心矩	χ^2 —— χ^2 函数
C_v ——样本的变异系数	F —— F 函数
C_s ——样本的偏度系数	$\wedge$ ——估计值符号(用在相应符号的头上)

目 录

主要符号说明

第一章 统计资料的整理

§ 1	总体、个体与样本	1
§ 2	穿孔卡片	1
§ 3	统计资料的分组作图	5
§ 4	算术平均数	9
§ 5	中位数与众数	11
§ 6	极差、离差、平均差、方差与方根差	13
§ 7	矩	17
§ 8	变异系数、偏度、峰度	18

第二章 概率与随机变量

§ 1	随机事件与概率	22
§ 2	概率运算法则	25
§ 3	排列组合与二项式系数	28
§ 4	随机变量及其概率密度	30
§ 5	随机变量的数字特征	32

第三章 随机变量的分布

§ 1	二项分布	35
§ 2	波松 (Poisson) 分布	37
§ 3	正态分布	39
§ 4	负指数分布	43
§ 5	随机变量的函数的分布	46
§ 6	大数定律与中心极限定理	49

第四章 统计推断理论

§ 1	统计推断的含义	52
§ 2	差异显著性水平的选择	53
§ 3	两类错误问题	55
§ 4	点估计的协调性与无偏性	58
§ 5	区间估计与假设检验的关系	60

第五章 t 检验的应用

§ 1	大样本的统计假设检验	62
§ 2	用大样本对总体作区间估计	65
§ 3	检验小样本是否来自平均数已知的正态总体	67

§ 4	检验两个小样本是否来自平均数相等的正态总体	68
§ 5	t 检验应取多大的样本	71
§ 6	t 检验应用总结	73
第六章	属性资料的统计检验	
§ 1	属性统计引论	76
§ 2	理论频数与实测频数是否相符的检验	76
§ 3	两个属性样本的同质性检验	80
§ 4	计数小于 5 的四格列联表的精确检验	83
§ 5	$a \times b$ 列联表的 χ^2 检验	85
§ 6	χ^2 的可加性	88
§ 7	χ^2 检验的其他问题	90
第七章	其他统计检验方法	
§ 1	F 检验的应用	92
§ 2	符号检验	93
§ 3	秩和检验	95
§ 4	R 检验	97
§ 5	柯尔莫哥洛夫 (Колмогоров) 检验	101
§ 6	方差非齐性平均数的比较	103
第八章	方差分析初步	
§ 1	试验设计的几个问题	105
§ 2	试验单元相等的单因素方差分析	110
§ 3	试验单元不等的单因素方差分析	114
§ 4	几点说明	116
第九章	多因素方差分析	
§ 1	无交互作用的方差分析	120
§ 2	有交互作用的方差分析	125
§ 3	重复数不等的完全分析	132
§ 4	方差分析总结	138
第十章	考虑区组因素的试验设计	
§ 1	试验设计及其误差	141
§ 2	重复与随机化	143
§ 3	随机区组与拉丁区组	145
§ 4	正交拉丁区组	148
§ 5	平衡不完全区组	151
§ 6	尧敦区组 (Youden) 与平衡格子方	154
第十一章	用正交表做试验设计	
§ 1	正交表的概念	157
§ 2	表头设计	159
§ 3	不考虑交互作用的两水平试验	161

§ 4	考虑交互作用的两水平试验	165
§ 5	多水平试验	168
§ 6	缺失数据的估计	172
第十二章	正交表的改造	
§ 1	并列	178
§ 2	并列应用实例	180
§ 3	拟水平法	183
§ 4	拟因子法	188
§ 5	组合法	193
第十三章	正交表的灵活应用	
§ 1	直和法	198
§ 2	部份追加法	206
§ 3	直积法	209
§ 4	分割(裂区)法	213
§ 5	灵活应用正交设计一例	216
第十四章	一元回归分析	
§ 1	相关关系	223
§ 2	用最小二乘法配回归方程	225
§ 3	回归系数与相关系数	229
§ 4	加权回归(相关)实例	233
§ 5	回归方程的方差分析	237
§ 6	回归系数的统计检验与区间估计	238
§ 7	两条回归直线的比较和检验	240
§ 8	回归方程的预报与精度分析	242
第十五章	一元相关分析与一元非线性回归	
§ 1	二元正态分布、经验相关系数的分布	245
§ 2	相关分析问题	247
§ 3	常用的线性化曲线	251
§ 4	非线性回归实例	253
§ 5	相关指数	256
第十六章	多元回归分析与相关分析	
§ 1	最小二乘法的应用	261
§ 2	二元回归实例	264
§ 3	用逆矩阵求回归系数	265
§ 4	用逆矩阵解回归方程的实例	268
§ 5	相关的密切性	272
§ 6	偏回归差方和	274
§ 7	经验偏相关系数	276
§ 8	偏相关系数的差异显著性检验及其他	279

§ 9	一个完整的回归分析实例	280
第十七章	逐步回归与多项式回归	
§ 1	逐步回归分析	288
§ 2	多项式回归与多元非线性回归	295
§ 3	抛物线回归的合理幂次选择	298
§ 4	正交多项式回归	302
§ 5	两个正交多项式回归实例	304
第十八章	经验分布曲线的选配	
§ 1	概述	310
§ 2	皮尔逊 (Pearson) 曲线族及其 $\mathbf{I}$ 型分布	310
§ 3	对数正态分布与韦勃 (Weibull) 分布	315
§ 4	经验频率的计算	318
§ 5	概率坐标纸法	319
§ 6	峰度、偏度检验	326
§ 7	韦勃分布配线实例	328
§ 8	分布曲线选配的应用	330
第十九章	质量评估与寿命评估方法	
§ 1	质量评估图	334
§ 2	$\bar{X}$ 图和 R 图	335
§ 3	P 评估图和 m 评估图	339
§ 4	寿命的分布函数与故障率	343
§ 5	以完全寿命作评估	345
§ 6	以不完全寿命作评估	346
§ 7	以最低寿命作评估	350
第二十章	用计算机作统计运算	
§ 1	统计运算的根本出路	353
§ 2	单因素方差分析程序	354
§ 3	多因素方差分析程序	356
§ 4	多功能一元回归分析程序	361
§ 5	多项式逐步回归程序	365
§ 6	正态分布配线检验程序	370
§ 7	韦勃分布配线检验程序	372

附表目录

一、标准正态分布曲线的面积及纵坐标.....	1
二、 t 分布数值.....	5
三、 F 分布数值	6
四、 χ^2 分布数值	12
五、 $x = \sin^{-1} \sqrt{P}$ 变换	13
六、阶乘和阶乘的对数.....	15
七、一万个随机数字.....	17
八、符号检验表.....	21
九、秩和检验的显著界限.....	22
十、柯尔莫哥洛夫分布的概率.....	24
十一、 $WSD \sqrt{k/s^2}$ 表.....	25
十二、正交拉丁方.....	26
十三、平衡不完全区组 (BIB)	29
十四、尧敦方.....	33
十五、常用正交表.....	36
十六、 r 作 z 转换表.....	54
十七、 Γ 函数表.....	55
十八、正交多项式表 ($N=2 \sim 30$)	56
十九、负指数 $R=e^{-F}$ 值 ($F=\lambda t$).....	60

参 考 文 献

一、国内文献.....	65
二、翻译文献.....	65
三、国外文献.....	66

第一章 统计资料的整理

§ 1 总体、个体与样本

我们在生产和科研工作中，所统计和分析的资料，都是从研究对象中测取得来的。这个研究对象的全体，通称总体或母体；研究对象的一个单位，则称为个体。譬如，研究某耐火车间所生产的耐火砖，这里的全部耐火砖就组成一个总体，而每一块砖就是个体。

当人们要了解研究对象的总体情况，一般只能在这个总体中取出一定数量的个体来研究。这一定数量的个体的组成，又叫样本或子样。每个样本所包含的个体数，通常称为样本容量。正如要了解耐火车间所生产的耐火砖强度，而把所有耐火砖一一拿到压力试验机上去试验，无疑是不现实的，我们只能取子样来试验，比如每天取4块砖试验，则每天所取的这4块砖便组成一个样本，此样本容量即等于4。

随着研究对象的变化，总体、个体和样本所代表的对象也就随着转移。当我们的任务是研究一年中耐火车间生产的耐火砖强度，则这全年的耐火砖强度是一个总体。如果任务只是研究该年某一个月生产的耐火砖强度，则这一个月耐火砖强度便是一个总体。同理，如果任务是研究这一年耐火砖的密度，则全年耐火砖的密度，又是一个总体。

我们知道，这个耐火车间生产的耐火砖强度，不管取它的什么期间作研究对象，其总体所包含的个体数总是有限的，一年的砖也好，十年的砖也好，砖的总数量完全可以知道，而且一块砖也只能在压力机上试验一次，这种性质的总体，称为有限总体或有限母体。如果研究对象是耐火砖的密度，情况就有些不同。因为砖的总数量虽然有限，但每块砖密度的测定却可以做上任意多次。又如，研究对象为某轧钢厂加热炉的出钢温度，加热炉是连续工作的，出钢温度可在炉子出钢端的任何地方做任意次测量。这种总体的个体数，在理论上说是无限的，则称这种总体为无限总体或无限母体。

前面已经指出：就研究对象的特性而言，如果要完全了解其总体，对有限总体往往不能做到，对无限总体则根本不可能。所以，只能抽取有限个个体组成一定的样本来研究。样本容量应该多大，才能对总体有良好的代表性？怎样正确通过样本了解总体的特性？怎样能使抽取样本的工作最经济最合理？这一类的问题都是数理统计学研究的重要任务。

§ 2 穿孔卡片

在日常生产活动中，各部门都有大量的原始资料，但这些资料通常是零乱的，必须经过整理才能进行研究。如不加以整理，一般难以直接应用。

数据资料的整理，既重要又繁琐。现已出现了一些专门简化资料整理工作的机器和计算机。对大量的长期性质的原始数据，用穿孔卡片来记载，能大大简化统计时的整理工作，并

易于使整理工作机械化和计算机化。

穿孔卡片有很多种，一般常见的有适于手工和机械整理的纸片式穿孔卡片。图1—1即为某钢铁厂设计的一种纸片式穿孔卡片的形状。该厂生产某种中碳优质钢，用这种卡片可以记载各种质量指标的原始数据。这样一张卡片，它记载的数据，反映出一个炉罐的钢质量情况，其中包括钢的冶炼日期、炉罐编号、化学成份、机械性能、低倍组织和高倍组织等一系列检验结果。

这种穿孔卡片的记载方法，是把某一炉罐的有关检验结果，按卡片上每个孔所代表的数字，用破口机或剪刀对应地将孔剪缺成一个切口。由于1~9中的任意数值，可用1、2、4、7四个数中两个以内的数之和来表示，0(零)用无缺口来表示，故0~9中的任意数能表示成两个以内的缺口，不同数量级的数，用另一组1、2、4、7的数来表达。对于某些不会出现的数，可以减少表达数的孔数。特殊的首数或尾数，可用特殊的专用孔数码。属性指标，凡能以数码形式表达的，都将它变换成数码，如低倍组织的检验结果，是评为良好、轻微、中等、明显四个等级，我们则分别用1、2、3、4四个数码表示之。

【例1】某钢铁厂一号炉，1966年1月3日生产的第0007炉中碳优质钢甲罐的检验结果：

化学成份：	C	Si	Mn	P	S
	0.61	0.25	0.60	0.012	0.015
低倍组织：	一般疏松(中等) 液析(良好)				
	中心疏松(轻微) 夹杂(中等)				
高倍组织：	氧化物	第一试片1.5级			
		第二试片2.0级			
	硫化物	第一试片1.5级			
		第二试片2.0级			
	晶粒度	第一试片7~8级			
		第二试片6~7级			
机械性能：	强度	伸长率	收缩率		
	98.5kg/mm ²	14.0%	27.0%		
白点：	无				

记载于穿孔卡片上的情况，如图1—2所示。

如果日常就进行这样的记载，当需要研究某一时期内钢的质量，或研究某种质量指标条件下的其他指标情况，或查找某一炉钢的原始资料，或对某些质量指标进行分类研究时，只要用比孔径(3mm)小一些的钎子，在需要查找或分类的指标所列的数码相对应的孔中一穿，则不属于所需要的那些卡片就一定会被穿在钎子上，而需要的那些资料会全部掉落下来。这样查找或分类是很方便的，而且还可用专门分选机来进行。

卡片应以比较厚实的半硬纸片来做，孔径和孔距每张卡片都应一致。穿孔双排的目的，是在外孔使用破损后可用内孔继续工作，以提高卡片的使用寿命。人工查找，卡片应存放在专用的抽屉中，有顺序地整齐地竖立在抽屉内。

穿孔卡，在国外应用较多，有的已经专门生产(如图书分类穿孔卡片)统一编制发行。查找出的卡片，如果配合专门的计算分析机来进行整理，数据的各项统计结果能自动地计算出来。

[illegible]

图1—1 记载钢质量原始数据的穿孔卡片

钢 质 量 检 验 专 用 穿 孔 卡																							
强度(kg/mm ²)		收缩率(%)		伸长率(%)		液析(级)		中心疏松(级)		一般疏松(级)		白点		年		月		日		炉号		罐号	
10.0 1.0 0.1		10.0 1.0 0.1		10.0 1.0 0.1		1 2 4 7		1 2 4		1 2 4		有 无		十 个		十 个		十 个		千 百 十 个		十 个	
1 2 4 7 5		1 2 4 7 5		1 2 4 7 5		1 2 4 7		1 2 4		1 2 4		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7	
晶粒度(级)		硫化物(级)		氧化物(级)		段号		硫(%)		磷(%)		硅(%)		锰(%)		铁(%)		碳(%)		磷(%)		硫(%)	
1 2 5 1 2 5 1 2 5		1.0 0.1 0.1 1.0 0.1		1.0 0.1 0.1 1.0 0.1		1 2 4 7		0.001 0.01		0.001 0.01		0.01 0.1		0.01 0.1		0.01 0.1		0.01 0.1		0.01 0.1		0.01 0.1	
1 2 5 1 2 5 1 2 5		1 2 4 7 5		1 2 4 7 5		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7		1 2 4 7	

图1—2 【例一】数据在穿孔卡上的记载

§ 3 统计资料的分组作图

对于得到的一个样本,如果各个个体的观测数值没有重复,可按其出现的先后次序或按其数值大小顺序排列成表。

如果样本容量较大或个体数值有重复时,经常采用分组的办法来处理。数据分组本身,不可能提高它的信息质量,但正确的分组乃是减少信息损失、方便数据整理的重要条件。分组多少,组距多大,主要由样本的分布范围以及样本容量大小来决定。下面是推荐的分组组距间隔(I)的计算公式:

$$I = \frac{x_{\max} - x_{\min}}{1 + 3.322 \times \lg N} \dots\dots\dots (1-1)$$

式中 $x_{\max}$ 与 $x_{\min}$ —— 最大取值与最小取值
 N —— 样本容量

I 计算出来后,按舍零取整的原理对组距通常都取整数或整小数。一般各个组距宜取相等,这样计算和分析时就很方便。组距间的分界必须明确,不能列得含混,譬如10~20、20~30,这样的组距列法是不好的,属于20的个体到底列在哪一组就产生了混乱。应列为11~20, 21~30或10~19, 20~29,这样组距间的分界就明确了。如果是连续的变量,则应列为11~、20~的形式,这表示等于11到小于20的所有数均为一组。对于处于最大和最小值两端的组,如果它们端头数值在理论上说来还可以延伸出去时,这端头的组距通常列成半开区间的形式,记成某数以下(上)的形式。

【例2】试将下列20批优质钢样品的强度数据分组列表处理

炉号	强度(kg/mm^2)	炉号	强度(kg/mm^2)
61007	98.5	62002	98.0
61038	95.5	62052	97.5
61041	94.5	62078	98.0
61078	96.0	62091	90.5
61084	99.0	62092	99.5
61100	98.5	62125	96.5
61137	97.5	62134	95.5
61159	97.5	62154	98.0
61165	92.5	62164	89.5
61162	104.0	62102	94.5

【解】 $x_{\max} = 104.0$ $x_{\min} = 89.5$

$$I = \frac{104 - 89.5}{1 + 3.322 \times \lg 20} = \frac{14.5}{5.322} = 2.724$$

取组距 $I = 3 kg/mm^2$, 分组编排成表1-1

表1-1 20批优质钢样的强度分组

分 组	≤ 92	~ 95	~ 98	~ 101	~ 104
频 数	2	3	1	4	1

〔例2〕的表1—1，称为频数分布表。组距的中间值，称为组中值。分组后，一般把这一组所有变量的取值，都以组中值来代表，如〔例2〕分组以后，我们把强度在95.1~98.0 kg/mm^2 之间的10个个体（95.5二个、96一个、96.5一个、97.5三个、98.0三个），都看成强度为96.5 kg/mm^2 。每个组内的个体数，称为频数，用符号 f^* 表示。全部个体数（即样本容量 N ）与各组个体数之比的百分数，称为频率，用符号 f 来表示。

根据频数分布表，可以制成频数分布图或频率分布图。制成图形，其主要优点是能够直观地了解变量分布的变化特征。

频数分布图的纵座标单位是个体数（频数），横座标单位是变量的不同取值。频率分布图的纵座标单位是百分数（频率），横座标与频数分布图一样。

频数或频率分布图对于连续变量，常绘成直方形或多角形。

〔例2〕的分布表可以绘成频数和频率直方形图（图1—3）和多角形图（图1—4）。

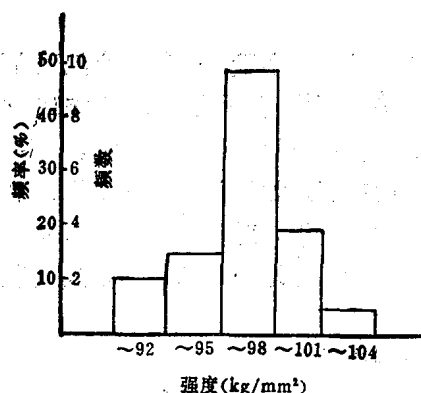


图1—3 20批钢样的强度分布直方图

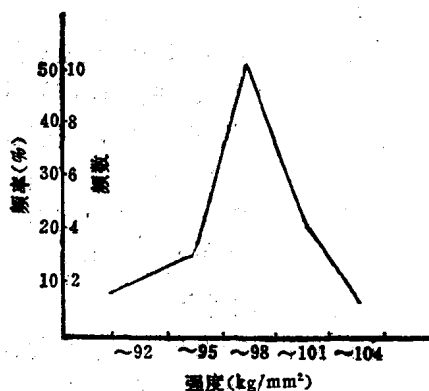


图1—4 20批钢样的强度分布多角图

可以知道：直方形图中所有矩形面积之和，在频数图中等于总的数量（ N ），在频率图中则等于总频率100%。这就是说，在频数和频率图中，每一个个体都占有相同的一个单位小面积。

多角形图没有直方形图上述严格的定量关系，它只把每个组的频数或频率的纵座标值简单地用直线连接起来而已。

比较直方形图和多角形图的特点，可以看出直方形图比多角形图更科学，今后应提倡使用直方形图。目前实际使用中，很多频数与频率分布图只简单地描绘变量的分布情况，要求并不严格。所以，多角形图仍然很广泛地被使用。

还有一种分布图——累积频数和累积频率分布图，对于理论分布曲线的配制和研究很有用处。累积频数和累积频率分布图，多采用直方形图，这样绘出的图形呈阶梯状，故有的又把它称为阶梯图。

所谓累积频数（累积频率），就是把各组的频数（频率）有序地累加起来所得的和。有序地累加，可按变量值由小到大的顺序，也可由大到小的顺序来累加。

譬如〔例2〕的累积频数和累积频率直方形图，如图1—5所示。

它们的累积频数和累积频率多角形图如图1—6所示。