# 质量管理学

## 第二分册

# 质量管理统计方法

中国质量管理协会培训教育部　编

企业管理出版社

# 质量管理统计方法

企业管理出版社

## 内 容 提 要

本册为《质量管理学》的第二分册，统计方法部分。它包括以下三方面的内容：

1．概率论与数理统计基础知识：从随机变量到中心极限定理，从总体、样本到统计推断，作者均作了扼要而较深入的叙述；

2．专门用于质量管理的两大类统计方法：统计质量控制和抽样检查，作者就其中的问题分别情况一一作了详细的介绍；

3．有广泛应用价值的一般统计方法：方差分析、正交设计、相关与回归等，作者强调了方法的实际背景和应用实效。

本册还编入了一些国内外较新发展的实用内容，如选控图、二次设计、回力设计等。

本册作为在质量管理人员中培训师资的教材而编写，也适合有志于质量管理的大专院校师生以及具有高中以上数学水平的工程技术人员阅读。

**质量管理学**

第二分册

**质量管理统计方法**

林少宫上编

企业管理出版社出版

煤炭工业出版社印刷厂印刷　中国质量管理协会发行

开本 787×1092 1/16　印张 22 1/2　字数553千字

1986年1月第1版　1986年1月第1次印刷

印数1-60,000册

统一书号：4207.060　定价3.00元

# 前　言

全面质量管理在我国试点、推行已经七年了。经过一大批企业的实践证明，推行全面质量管理是符合我国国情的，它确实是社会主义企业提高质量、降低消耗，改善企业素质，增强竞争能力，提高企业和社会经济效益的必由之路和重要措施。

随着全面质量管理的不断深入开展，广大质量管理工作者越来越感到需要进一步全面、系统地学习和掌握质量管理这门新的边缘学科。继续停留在推行初期的教育水平上，以学习国外一些经过归纳、总结后的几个观点和方法为主的教育内容已经不能满足指导实践的需要了。当前迫切需要在学习和借鉴国外质量管理理论和实践经验的基础上，编写一套结合我国实际、系统论述质量管理学科的教材。为此，我们组织部分学者，编写出一套“质量管理学”，献给经济战线的各级领导和广大质量管理工作者及企业中的技术人员。

这套教材曾在中国质协近年来举办的各期师资培训班上试用，并征集采纳了一些意见。全书根据内容分为两个分册，第一分册名为《质量管理原理与理论》，第二分册名为《质量管理统计方法》。

第一分册由中国人民大学工业经济系杨文士担任主编，整个分册分为五篇十八章及三个附录，其中第一至七章由杨文士负责编写；第八至十三章由北京航空学院管理工程系孙巩负责编写；第十四章至十七章由陕西机械学院北京研究生部廖永平负责编写；第十八章由中国质量管理协会张贵华、罗国英负责编写。第二分册由华中工学院数量经济研究所所长林少宫教授担任主编。整个分册分为七章，其中第一至三章由华中工学院数学系余明书负责编写；第四至五章由武汉建材学院数学教研室刘朝荣负责编写；第六至七章由内蒙古工学院管理工程系张绍镛负责编写。

本书可作为对在职技术人员和质量管理工作者的培训教材，并可供经济战线各级领导和从事科研、教育等工作的同志学习与研究使用，还可供大专院校开设质量管理课程参考。

由于我国幅员辽阔，行业繁多，开展全面质量管理的时间不长，企业的实践还十分有限，尚缺乏系统的总结与研究，要编写出一套各地区、各行业都能适用的教材，目前还有很多困难。加上我们水平有限，经验不足，恳切希望读者在使用本书的过程中提出宝贵意见，以便再版时进行修改和补充。尽量使这套“质量管理学”更能结合我国的国情，为指导我国的质量管理事业，提高企业现代化管理水平作出应有的贡献。

**中国质量管理协会培训教育部**

一九八五年十月

# 目　录


前言 …… 1

**第一章　数据的收集与整理** …… 1

第一节　引言 …… 1

第二节　数据收集 …… 1

第三节　数据整理 …… 8

**第二章　概率与概率分布** …… 22

第一节　引言 …… 22

第二节　概率及其计算 …… 23

第三节　离散型随机变量及其分布特征 …… 30

第四节　连续型随机变量及其分布特征 …… 52

第五节　中心极限定理 …… 60

**第三章　统计与统计推断** …… 66

第一节　引言 …… 66

第二节　抽样分布 …… 66

第三节　参数估计 …… 77

第四节　假设检验 …… 87

第五节　正态概率纸与$\chi^2$检验 …… 96

**第四章　统计质量控制** …… 103

第一节　常用统计方法 …… 103

第二节　工序能力分析 …… 116

第三节　工序统计控制 …… 133

**第五章　抽样检验方法** …… 171

第一节　衡量产品质量的方法 …… 172

第二节　计数抽样检验 …… 174

第三节　计量抽样检验 …… 197

**第六章　方差分析与正交试验设计** …… 209

第一节　方差分析 …… 209

第二节　正交试验设计 …… 224

第三节　可计算性产品的三次设计 …… 255

**第七章　相关与回归分析** …… 274

第一节　概述 …… 274

第二节　相关分析 …… 275

第三节　回归分析 …… 281

第四节　正交试验的回归设计 …… 308

**附录1**　随机数表 …… 318

**附录2**　$\sum_{d=0}^{c}\frac{\lambda^d}{d!}e^{-\lambda}$值表 …… 320

**附录3**　$\sum_{d=0}^{c}C_n^d p^d(1-p)^{n-d}$值表 …… 324

**附录4** 累积标准正态分布表——$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz$ …… 329

**附录5** $\chi^2$分布表——$P\{\chi^2(n)>\chi^2_\alpha(n)\}=\alpha$ …… 330

**附录6** t分布表——$P\{t(n)>t_\alpha(n)\}=\alpha$ …… 332

**附录7** F分布表——$P\{F(n_1,n_2)>F_\alpha(n_1,n_2)\}=\alpha$ …… 334

**附录8** 调整型抽样检验表（1）～(10) …… 337

**附录9** 正交表（1）～(5) …… 352

# 第一章　数据的收集与整理

## 第一节　引　　言

生产是为了满足社会日益增长的物质和文化生活的需要，为国家创造财富，同时也为进行生产的企业带来收益。任何企业，只要进行生产就必定存在产品质量问题。产品质量的优劣，不但直接关系到企业的信誉、经济收益和前途，并且也反映该企业的技术水平和管理水平。当今，质量的概念几乎渗透到各个领域。例如武器、弹药的质量关系到国防的防卫实力；药物、食品的质量关系到人们的身体健康；汽车、房屋、电器……的质量无不与人们的生活紧密相关。质量是人们对完善、适用、精美、真实、可靠等含意的综合表述。追求高质量，对用户负责，这是一种高尚的美的行为。

质量离不开数量。产品质量的提高，要用数量来表示；不合格品率的降低，也要用数量来表示；产品销售量、产品的规格、控制生产所用的管理标准、……，都要用数量来表示。现代工业生产的专业化方向势在必行。专业化的一个重要特点就是要求成批、大量地生产。企业生产量往往是数以万计、十万计、百万计、……。面对产量巨大的生产，如何才能保证它的质量呢？旧有的办法是用大量的检验人员，对产品进行全数检查，这显然是不经济的十分落后的办法。在产品生产过程中，能否及时发现不良征兆，发出警报以确保生产出来的产品都是合格品呢？如果影响产品质量的因素很多，能否从中找出主要因素并及时进行控制呢？根据生产实际所积累的数据，能否对设计参数、工艺参数进行优选，以保证达到预期的目的呢？

实践证明：质量管理统计方法能够提高产品质量、降低成本、优选设计参数及工艺参数。所谓统计质量管理（Statistical Quality Control），就是由数据来说话，用统计方法处理过的统计数字来表明管理结果。事实上，产品质量反映在数据的波动上，现代工业企业管理的技巧，主要在于数据的运用。

## 第二节　数　据　收　集

通过有目的地收集数据，运用数理统计的方法处理所得的原始数据，提炼出有关产品质量、生产过程的信息，再分析具体情况，作出决策，从而达到提高产品质量的目的。这就是通常所说的质量管理。

初期的质量管理，往往只是进行“事后把关”。全数检查即将出厂的产品，测试有关的质量特征指标，从中挑出不合格品，出厂合格品，见图1-1。这种方法，一方面不能预防生产过程中不合格品的产生；另一方面，如果要知道每一产品或工序被破坏时的指标，全数检查的方法显然是行不通的。

随着生产的发展和产量的大幅度提高，迫切地要求用经济、可靠的方法解决企业产品的质量检查问题，这就导致把全数检查发展成抽样检验。这种检验不仅要对最终产品进行抽样检验，以便对不合格品采取一定措施，减少可能引起的损失，而且要在产品生产过程

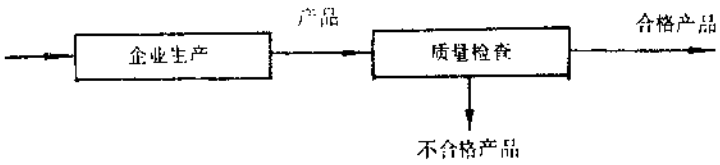


图 1-1

中做及时抽样检验，预防不合格品的出现，从而使生产过程保持稳定状态，图1-2。

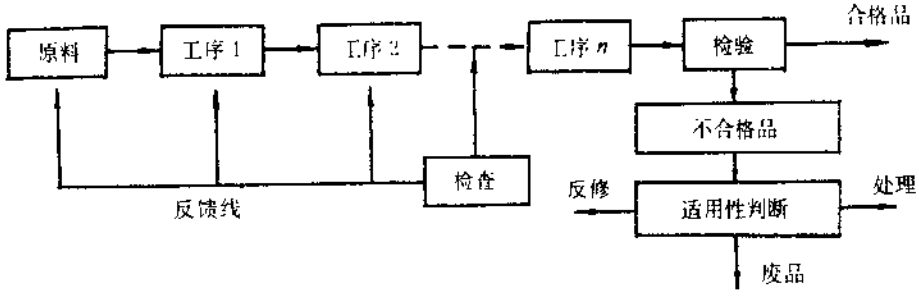


图 1-2

什么是抽样检验呢？简言之，抽样检验就是通过检验一部分产品而对一批产品的质量作出估计。只有合理地抽取样本，正确地运用统计方法处理数据，才能得到可靠的估计。下面我们简单地介绍抽样方法和一些需要注意的事项。关于抽样方法的细节，在以后各章节还要专门讲述。我们在这里仅仅借助于抽样问题说明数理统计的基本概念和一些运算。

（一）总体与样本

1. **总体** “总体”一词，是统计学中常用的一个术语。一批产品、一台设备或在某段时间内生产的同类产品的全体等，都可以叫做一个总体。例如，某灯泡厂三月份生产的全体灯泡，某纺织厂生产的布匹，或一道工序加工后的半成品，制造产品的原材料等等，都能成为一个总体。

构成总体的基本单位，叫做个体。这个基本单位又可叫做单位产品。单位产品有时可以很自然地划分出来，如果总体是一批电灯泡，那么其中的每只灯泡都可以看作一个个体。有些单位产品却不能自然地进行划分。比如，如果把一匹布作为一批产品，即叫做一个总体，那么这总体中的每个个体可以是一米布、十米布或二十米布，在这种情况下，个体的划分需要由具体问题而定。

所谓抽样检验，指的是从总体中抽取一部分个体，并测试被抽到的每个个体的有关质量特征的数量指标，得到一组数据，再对这些数据进行处理，然后对总体作出估计和判断。每个单位产品都有一个或一组表现其质量的数量指标与之对应。如电阻元件与阻值；电灯泡与其使用寿命；一米棉布与其上的疵点数。

总体中的每一个个体肯定要与某一个（组）数相对应，这是确定无疑的事实。这个数就是每个个体的质量表征，由于这个数的具体取值因个体的不同而异，因此通常称它为随机变量，并记作$X$。

综上所述，可以说总体是问题所涉及的全体对象，总体就是随机变量$X$的全体取值，

也可以说总体就是随机变量$X$。

总体可以是有限的，也可以是无限的。一个工厂生产的电子管的数目是有限的，但这些电子管的寿命作为一个总体来看就是无限的。

**2. 样本** 从一批产品中抽取一部分进行检验，被抽取的这一部分单位产品的全体，就叫做一个样本。换句话说，样本就是从总体中抽取的一部分个体的全体。

例如，我们要了解即将出厂的一批螺钉的长度是否合格，今从中抽取$n$个个体$X_1$，$X_2$，……，$X_n$。我们就称

$$\{X_1,\ X_2,\ \cdots\cdots,\ X_n\}$$

为总体的一个样本。这里的$n$通常称为样本容量或样本大小。

其中$X_1$，$X_2$，……，$X_n$分别表示$n$个螺钉的长度，$n$个螺钉被抽取后，客观上就有$n$个长度与其对应，当长度未被测定，则都用随机变量表示。待具体测定它们的长度后，所得值记作

$$\{x_1,\ x_2,\cdots\cdots,x_n\}$$

称它为样本值。

样本中的每一个个体叫做一个样品。今后，如果我们说抽取了样本大小为$n$的一个样本，实际上就是抽取了$n$个样品，也就是抽取了$n$个单位产品。如果在有限总体中，包含个体的总数为$N$，则常称$N$为总体的批量。

（二）数据的分类

抽取到样本以后，我们就要测试每个单位产品的质量特征指标，并把测得的原始数据记录在表格内。在统计质量管理中，总体样本、数据间的关系，可用框图1-3表示。

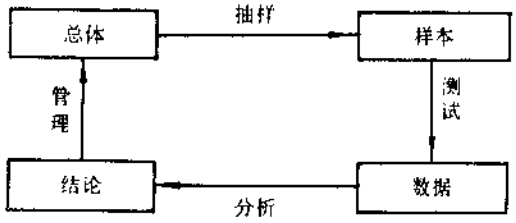


图 1-3

单位产品的质量可以用不同的方法来衡量。有的用“计数”方法来衡量，有的用“计量”的方法衡量。

**1. 计数值** 描述一米布或一铸件表面的质量，常用一米棉布上的疵点数、一铸件表面的气孔数或砂眼数表达。这就是所谓“计点”值描述产品质量的方法。

存在着一些单位产品的质量特征不能用定量的方法去度量。如洗衣机外表面的漆层有剥落，就可以认为其外观不合格。有许多产品，其外观不能用定量的方法来衡量，只能定性地分为好的与坏的，合格的与不合格的，这叫做用“计件”的方法表示产品的质量特征。

我们把“计件的”和“计点的”方法，统一称为计数的方法。计数的方法有一个重要的特点，即它的数据是不连续的，也就是说数据是以离散的状态出现的。

**2. 计量值** 许多产品的质量要用连续数量来表示。如棉纱的强力、钢的化学成分、电灯泡的寿命等，都是衡量产品质量的连续变量。使用连续变量来定量地度量产品质量的方法，称为计量方法。

计量值与计数值的不同，在于计量方法得出的数据是一组连续变量，而计数值是一组离散数据。

如何区分离散数据和连续型变量呢？所谓离散数据又称为间断数据。它的特点是在同

一变数之下，任何两个不同变数值之间不能插入无穷多个数值，而连续型变量则可以。

例如，记录机器每天发生故障的次数，属于计数值。记录得出来的数据是离散的。我们在3与7之间，插入4.56是无意义的。因为机器发生故障的次数不可能取4.56次。

在测试电灯泡寿命的一组数据里，取任意两个不同的数值，如1,230小时与2,000小时，在其中插入1,457.8小时是有意义的。因此，电灯泡的寿命是属于计量值。

（三）数据的修约

质量管理离不开数据的测定与计算。因此，也就出现了对超出规定精确度范围之外的数字，如何处理的问题。在数理统计中对数据的修约规则和传统的“四舍五入”法则，略有差别。具体规定如下。

（1）将某一数值修约为有效数字 $n$ 位，当第 $n+1$ 位的数字小于5时，舍去；当第 $n+1$ 位的数字大于5时，进1。

例如，若将下列左边的数值修约为4位有效数字，则可得出右边的结果。

$$3.224496 \longrightarrow 3.224$$

$$73.61601 \longrightarrow 73.62$$

（2）将某一数值修约为有效数字 $n$ 位，且第 $n+1$ 位数字等于5，则当第 $n+1$ 位以后的数字不全为0时，进1；当第 $n+1$ 位以后的数字全为0时，若第 $n$ 位数字为偶数（即2，4，6，8）或0时，则舍去；若第 $n$ 位数字为奇数（1，3，5，7，9），则进1。

例如，将下列左边数值修约为4位有效数字，则有

$$253.1501 \longrightarrow 253.2$$

$$4.8725 \longrightarrow 4.872$$

$$56.235 \longrightarrow 56.24$$

（3）若所拟舍去的数字不是单独一个数字，则不得对该数值连续进行修约，而应按所拟舍去的数字中最左边的第1个数字的大小，依上述规定（1）、（2）进行处理。

例如，将15.4546修约为两位有效数字，不应作如下处理：

$$15.4546 \rightarrow 15.455 \rightarrow 15.46 \rightarrow 15.6 \rightarrow 16。$$

应按下面方法处理：

$$15.4546 \longrightarrow 15$$

（四）抽样方法

必须研究抽样方法。因为合理地进行抽样，可以减少检验的数量。只有合理地进行抽样，才能使样品有代表性，才能得到精确可靠的推断，从而提高产品质量，确保生产处于稳定状态。怎样的抽样方法才是合理的呢？下面我们介绍几种常用的抽样方法。

**1. 随机抽样** 所谓随机抽样法，是指总体中的每个个体，均有同样被抽取机会的一种抽样方法。抽样时完全用偶然的方法抽取，事先不能考虑应抽取哪一个样品。拿产品来说，在随机抽样中，一批产品中的每一个产品都应有同样的机会被抽取，从而保证样品的代表性。目前较简便而又客观的随机抽取方法是利用随机数表（附表1）来进行抽样的。

例1-1 要从1000个产品中，随机抽取10个样品，组成样本大小为 $n=10$ 的一个样本。

利用随机抽样方法的步骤如下：

第一步，将1000个产品编号，每一个产品对应着一个号，编号的次序与方法不受任何限制，一般是从000编至999。

第二步，在随机数表中，任意指定一点，假设我们指定随机数表的第29行，第3列交叉处为起点，见附表1。从起点开始，由29行的第3、4、5列形成随机数字064，往右逢三个数字一读，得随机数字为：064，767，532，236，496，886，870，418，806，696。

以上十个编号，就是随机抽取的样本。以号对产品的编号，就得到10个待测试的样品。

注意，利用随机数表时，可以随意使用随机数表的任意一页。在该页上任意选定一个起始点后，可以由左向右，或由右向左，由上而下，或由下而上地取随机数。

如本例，当选29行的第3、4、5列形成的随机数字064为起始随机数，由此向下取数，则得：064，953，559，933，554，928，054，862，768，843等10个数字为入样数码。

例1-2　试从345个产品中，抽取15个样品，形成一个样本。

我们先将345个产品，由001～345编上相应的号，再在附表1随机数表上任选一行、三列形成三个数码的随机数。如选第84行，第9、10、11三列，由上向下取随机数，就得到：(655)，(520)，(662)，(380)，(473)，(405)，049，264，027，267，088，198，(457)，241，(675)，(753)，238。

凡数码在001～345之间的，都可以入样，其余的弃去不用。如此，仅有9个数码对应的单位产品入样，不足15个，需继续抽取。可以任意另选某行某列作起点，并自取确定三个数码的方法，凡抽到001～345之间的号，则入样，其余弃之，直至取满15个不同的数码为止。

用前述方法，可能要读许多数码才能选出15个在我们的编号以内的随机数。为加快取样速度，也常采用下述方法选取样品：

在任意选中了读数起点和顺序后，凡所读数码在001～345之间的，则该数码入样。当所读数码在401～745之间的，由该数码减去400后所得的数码入样。其余的000，346～400，746～999不要，且碰到重复的数码只算一次。用这样的方法，我们很快就可以选中例1-2中的15个样品，其编号可以是：225，120，262，073，005，049，264，027，267，088，198，057，241，275，238。

随机数表内，列出了100行、40列的0～9之间的共4000个数字，用查表的方法抽样十分方便。由于读数起点及读数顺序的任意性，随机抽样的答案有非唯一性，这正说明了总体中每个个体均有同样被抽取的机会。

**2. 分层抽样法**　分层抽样法是另一种常用的抽样方法，它用下述方法进行抽样：

先将含有$N$个个体的总体$S$，分成$K$组，分别记作$S_1$，$S_2$，……，$S_k$，称$S_i$为第$i$层，$i=1$，2，……，$K$。我们将每一层都看作一个小总体，一般情况下，每个小总体中含个体的数目可以不相等，如记$S_i$中含个体数为$N_i$，$i=1$，2，……，$K$，则有$N_1+N_2+\cdots+N_k=N$。这样，第$i$层可以表示为

$$S_i=\{w_1^{(i)},\ w_2^{(i)},\ \cdots,\ w_{N_i}^{(i)}\}\quad i=1,\ 2,\ \cdots,\ N_k \tag{1-1}$$

要对总体$S$抽取样本大小为$n$的样本，必须首先将样本大小$n$分成$n_1+n_2+\cdots+n_k$，当然

$$n = n_1 + n_2 + \cdots + n_k$$

再在第 $i$ 层 $S_i$ 中抽取一组大小为 $n_i$ 的样本，$i = 1, 2, \cdots, K$，这样就得到了 $K$ 组样本。最后把 $K$ 组样本合起来就形成为总体 $S$ 的分层样本。

分层抽样的结果，将是每层都有一些个体入样。因此，为了使样本有较好的代表性，分层时，一般将 $S$ 中某些有明显差异的个体分开，放在不同的层内，即将 $S$ 中相近的个体归为一层。

比如，要调查某种货物的销售额，抽查对象由百货商场至零售商店都应包括在内，为了使抽查合理，必须先将全部商场（店）依规模进行分层，再在各层进行抽取。

抽样时，应从每一层中抽多少个样品呢？一般可依比例进行分配。所谓按比例分配，是指每层都按同一比例抽取样品，含个体总数多的层分配样品数应多些，含个体总数少的层分配样品数按比例减少，具体个数可参考下面的方法进行计算。设总体 $S$ 中，含个体数为 $N$，现已将 $S$ 分为 $K$ 层：$S_1, S_2, \cdots, S_k$，且每一层所含个体数分别为 $N_1, N_2, \cdots, N_k$，$(N = N_1 + N_2 + \cdots + N_k)$。现要抽取样本大小为 $n$ 的一个样本，可用如下公式按比例分配各层样本大小，取

$$\frac{n_i}{N_i} = \frac{n}{N} \quad \text{或} \quad n_i = \frac{N_i}{N} n \qquad i = 1, 2, \cdots, K \tag{1-2}$$

显然应有，$n_1 + n_2 + \cdots + n_k = n$，但应注意这样算出来的 $n_i$ 不一定是整数，当 $n_i$ 为整数时，分层按比例抽样保证了总体中每个个体都有同等入样机会。

例1-3　有一批产品，共345件，试用分层抽样法抽取 $n = 15$ 个样品。

设依产品的每种特性差异分345件产品为5层，记作 $S_1, S_2, \cdots, S_5$，而且各层中产品的个数相应为45，55，65，95，85。由式（1-2），此时应有

$$N = 345,\ N_1 = 45,\ N_2 = 55,\ N_3 = 65,\ N_4 = 95,\ N_5 = 85,\ n = 15。$$

将以上数字代入式（1-2）得

$$n_1 = \frac{N_1}{N} n = \frac{45}{345} \times 15 \approx 1.96$$

$$n_2 = \frac{N_2}{N} n = \frac{55}{345} \times 15 \approx 2.39$$

$$n_3 = \frac{N_3}{N} n = \frac{65}{345} \times 15 \approx 2.83$$

$$n_4 = \frac{N_4}{N} n = \frac{95}{345} \times 15 \approx 4.13$$

$$n_5 = \frac{N_5}{N} n = \frac{85}{345} \times 15 \approx 3.7$$

由于产品个数不能取小数，故采取数据修约的方法，最后取

$$n_1 = 2,\ n_2 = 2,\ n_3 = 3,\ n_4 = 4,\ n_5 = 4$$

在各层中抽取样品数确定以后，具体在每一层中抽取哪一个样品，仍依前面讲的随机抽样的方法进行，这里就不重复了。因此，分层抽样法可描述为分层随机抽样，而前面讲的随机抽样也可描述为简单随机抽样，以示区别。

**3. 系统抽样法**　系统抽样法是比较方便的一种抽样方法，又叫做机械随机抽样。具

体做法是，先选定一正整数$K$，称$K$为抽取间隙。一般取$K$为尽可能接近$N/n$的整数。这里的$N$、$n$分别是总体和样本大小，然后将总体$S$中的$N$个个体依间隔排列，见表1-1。

表 1 1

| | | | |
|---|---|---|---|
| 1 | 2 | …, | K |
| K + 1, | K + 2, | …, | 2K |
| 2K +1, | 2K + 2, | …, | 3K |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 直至N为止 | | | |

以上的工作，事实上只是为抽样作准备，当准备妥当后，抽样是十分方便的。只要对号码1,2,……,$K$中作一次随机抽取，如果抽到第$i$号，则第$i$号入样，且称之为随机起始数。当第$i$号入样，则$K+i$,$2K+i$,……，都入样，直到抽满$n$(样本大小)个为止。即当起始数为$i$，则表1-1中第$i$列各号皆入样。不难理解，当$N/n$为整数时，每个个体都有同等的入样机会。

例1-4　试用系统抽样法，在批量$N=10000$的总体中，抽取样本大小$n=200$的一个样本。

我们依下列步骤完成系统抽样

第一步　决定抽取间隔$K$，令

$$K=\frac{N}{n}=\frac{10000}{200}=50$$

第二步　利用随机数表，从1到50中任取一号，作为起始数，如抽到13为起始数，对10000个产品编上相应的号，则

13, 63, 113, 163, 213, ……, 9963

就为所抽取的样本。

在实际工作中，具体实行系统抽样时，可以对某道工序中的半成品实行每隔$K$个抽取一个的方法，直到抽取所需的个数为止。大规模自动化生产线，可以采取每隔一段时间抽取一个样品的方式进行系统抽样。如要调查某天某道工序生产的情况，只需每隔十分钟抽一个样，在二十四小时内，就可得到由144个样品形成的系统样本。

注意，用$K=N/n$计算抽取间隔，常有K出现非整数的情况，这时仍然可以用近似取整数的办法来机械随机抽取。

如从编号为1～10的产品中，随机抽取三个样品，那么按系统抽样法

$$\frac{10}{3}=3.33\cdots$$

从1～10中随机抽取一个数，假设为4，以

$$\frac{4}{3}=1.33\longrightarrow 1\text{ 为起始抽号}$$

$$\frac{4}{3}+\frac{10}{3}=4.66\longrightarrow 5\text{ 为抽取的第二号}$$

$$\frac{4}{3}+2\times\frac{10}{3}=8\longrightarrow 8\text{ 为抽取的第三号}$$

以上简单地介绍了几种随机抽样的方法，在实际应用时采用哪一种方法，要由具体问

题而定。在选择抽样方法时，既要求进行抽样时迅速、经济，还要保持资料的精确度，更真实地反映总体的状况。

## 第三节　数据整理

抽样的目的，就是要通过局部（样本）反映整体（总体），即从一些样品判断一批产品的好坏。。怎样衡量一批产品的好坏呢？怎样才能知道生产过程是否稳定、正常呢？为此，首先必须对抽取的样本中的每个样品进行测试，取得原始数据，再整理加工这些数据，找出它们的特征，从而推测总体的变化规律和趋势，然后作出判断。下面先介绍整理加工数据的方法，对总体的分析和判断要留在以后的章节中再讲解。

先看下面的例子。

例1-5　于20天内，从维尼纶正常生产报表上看到的维尼纶纤度（表示纤维粗细程度的一个量）的情况，有如下100个数据，见表1-2。

表 1-2

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1.36 | 1.49 | 1.43 | 1.41 | 1.37 | 1.40 | 1.32 | 1.42 | 1.47 | 1.39 |
| 1.41 | 1.36 | 1.40 | 1.34 | 1.42 | 1.42 | 1.45 | 1.35 | 1.42 | 1.39 |
| 1.44 | 1.42 | 1.39 | 1.42 | 1.42 | 1.30 | 1.34 | 1.42 | 1.37 | 1.36 |
| 1.37 | 1.34 | 1.37 | 1.37 | 1.41 | 1.45 | 1.32 | 1.48 | 1.40 | 1.45 |
| 1.39 | 1.46 | 1.39 | 1.53 | 1.36 | 1.48 | 1.40 | 1.39 | 1.38 | 1.40 |
| 1.36 | 1.45 | 1.50 | 1.43 | 1.38 | 1.43 | 1.41 | 1.48 | 1.39 | 1.45 |
| 1.37 | 1.37 | 1.39 | 1.45 | 1.31 | 1.41 | 1.44 | 1.44 | 1.42 | 1.47 |
| 1.35 | 1.36 | 1.39 | 1.40 | 1.38 | 1.35 | 1.42 | 1.43 | 1.42 | 1.42 |
| 1.42 | 1.40 | 1.41 | 1.37 | 1.46 | 1.36 | 1.37 | 1.27 | 1.37 | 1.38 |
| 1.42 | 1.34 | 1.43 | 1.42 | 1.41 | 1.41 | 1.44 | 1.48 | 1.55 | 1.37 |

从上表100个数字中能看出什么呢？除了知道各个被抽取的维尼纶纤维粗细的不同以外，看不出其他任何规律性的东西。

必须对原始数据加以整理，在保留原数据中所包含的主要信息的前提下，对数据进行“压缩”，用简单明确的有代表性的数字说明总体的规律性。下面我们以例1-5的表1-2所列数字为例，介绍整理数据的方法。

（一）频数分布表

对原始数据进行分组，且求出频数分布表，是数据整理中不可少的一种方法。它的基本方法是将原来散布的数据，“压缩”到几个区域内，以便于我们分析总体的情况，其主要步骤如下：

第一步，分组。将数据的取值范围分为相互衔接的各个小区间，称为组。分组时，先要确定组数和组距。

（1）组数。通常用表1-3的标准决定分组数目。结合例1-5，由于表1-2中的数据总数$n=100$，故由表1-3可以分成$K=10$组。

（2）组距。组距是一组之宽度，对于等组距（即每组组距相等）的情形。组距h为

$$h=\frac{\text{全距}}{\text{组数}}=\frac{R}{K} \tag{1-3}$$

表 1-3

| 数据总数 $n$ | 适当的组数 $k$ |
|---|---|
| 50～100 | 6～10 |
| 100～250 | 7～12 |
| 250以上 | 10～25 |

式中 $R$ = 全距（又称为极差）

= 数据中的最大值 − 数据中的最小值

$$= \max_{1\le i\le n} x_i - \min_{1\le i\le n} x_i$$

$K$ = 组数。

一般将比值$R/K$选定为测量单位的整数倍，且以奇数倍较好。比如在例 1-5 中，

$$R = 1.55 - 1.27 = 0.028$$

故

$$h: 0.28/10 = 0.28 \approx 0.03$$

0.03是测量单位0.01的三倍，所以我们就取组距为0.03。在表1-2中，1.38，1.49，…数值，其小数点后两位数是精确的，又说精确至0.01，故在这里称0.01是测量单位。

第二步，决定组限。组的上、下界限值称为组限。由全数据的下端开始，每加一次组距就可以构成一个组限。在划分组限前，必须明确端点的归属，考虑到读数精度的限制（末位数字的舍入），故在决定组限时，只要比原始数据表中的有效数字的位数多取一位，则不存在端点数据的归属问题。

在例1-5中，原数据在1.27～1.55之间，现取为1.265～1.555之间。

分成10组，每组组距为0.03，故各组限为 1.265～1.295，1.295～1.325，…1.535～1.565。

第三步，作频数分布表。用唱票法，数出落在每组内观测值的个数，称为频数$f_i$，并计算频率$f_i/n$及累积频率$F_i$，列出频数分布表。

用上述步骤，得出例1-5的频数分布表，见表1-4。

表 1-4

| 组限 | 组中值$\tilde{x}_i$ | 频数计算 | 频数 $f_i$ | 频率$f_i/n$ | 累积频率$F_i$ |
|---|---|---|---|---|---|
| 1.265～1.295 | 1.28 | 正 | 1 | 0.01 | 0.01 |
| 1.295～1.325 | 1.31 | 正 | 4 | 0.04 | 0.05 |
| 1.325～1.355 | 1.34 | 正丅 | 7 | 0.07 | 0.12 |
| 1.355～1.385 | 1.37 | 正正正正下 | 22 | 0.22 | 0.34 |
| 1.385～1.415 | 1.40 | 正正正正下 | 23 | 0.23 | 0.57 |
| 1.415～1.445 | 1.43 | 正正正正正 | 25 | 0.25 | 0.82 |
| 1.445～1.475 | 1.46 | 正正 | 10 | 0.10 | 0.92 |
| 1.475～1.505 | 1.49 | 正 | 6 | 0.06 | 0.98 |
| 1.505～1.535 | 1.52 | 一 | 1 | 0.01 | 0.99 |
| 1.535～1.565 | 1.55 | 一 | 1 | 0.01 | 1.00 |

分组后的数据称为分组数据，每组数据所处的区间端点称为组限值。一组的上下限值的算术平均值，即区间的中点值称为组中值$\tilde{x}i$。如第四组的组中值为

$$\tilde{x}_4=\frac{1.355+1.385}{2}=1.37$$

对比表1-4和表1-2，我们明显地感到，经过加工后的分组数据，可以使人更清楚地看出数据的分布状况和规律性，并可以对总体作出一些判断。比如，不难作出结论，由20天内生产的维尼纶中抽取的100个样品，测其纤度，估计有70%的纤度在1.355～1.445之间。

（二）直方图

为了直观地表示数据分布，可以将数据中的频数、频率、累积频率分别用各种图形画出来。直方图能直观地形象地反映产品质量的分布情况，它是一种有效的现场管理工具。下面我们以表1-4列出的分组数据为例，介绍各种直方图的作法。对直方图的具体分析研究，留待第四章再进行讨论。

1. **频数直方图** 图1-4就是一个频数直方图。图中横轴表示变量的值，纵轴表示频数。各组所含频数的多少，以各组组距上的长方形的高度表示。频数直方图的具体作法如下。

第一步，选定直角坐标系。以横轴表示变量的值，刻上各组区间端点的值，以纵轴表示频数。

第二步，以各组频数为高度，各组组距为底边，对每一组划一长方形，频数直方图即告完成。见图1-4。

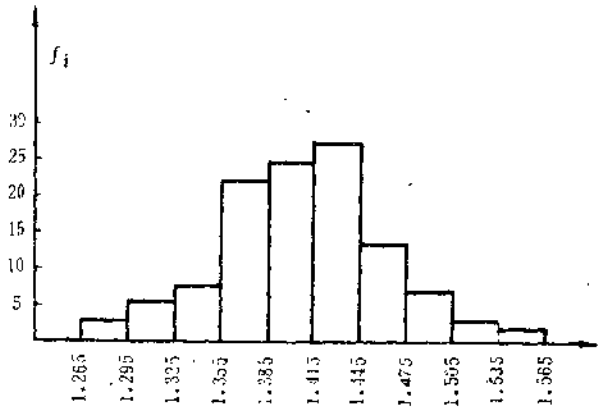


图 1-4

2. **频率直方图** 频率直方图以横轴表示变量的值，纵轴表示组距内的平均频率。如果令$\Delta x$表示组距（在例3-1中$\Delta x=0.03$），则纵坐标可取为

$$\frac{\text{频率}}{\text{组距}}=\frac{f_i/n}{\Delta x}=\frac{f_i}{n\Delta x} \tag{3-2}$$

选好坐标后，在每组上作一长方形，长方形的底边是组距$\Delta x$，其高为$fi/n\Delta x$。具体作法如下：

第一步，选定直角坐标系，以横坐标表示变量的值，刻以各组区间端点的值，纵坐标标以单位 频率/组距，不妨将纵坐标变量记作$y$，即令

$$y = \frac{fi}{n\Delta x}$$

第二步，以各组组距为底边，再以各组的频率除以组距为高，对每一组划一长方形，频率直方图即告完成，见图1-5。

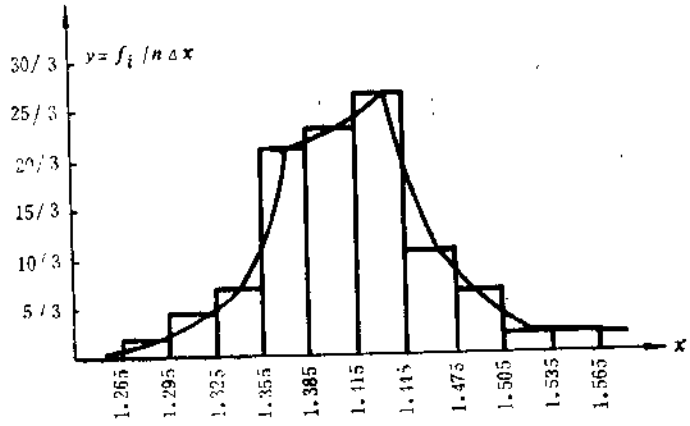


图 1-5

频率直方图与频数直方图的形象类似，差别仅在于纵坐标单位选取的不同。故在频率直方图中，不以长方形的高度表示频率，而以每块长方形的面积表示变量在该组上出现的频率数。

将频率直方图中所有的面积相加，其总和恒为 1，这是显然的。如果将图 1-5 中的频率直方图中每长方形的上边中点联接起来，就形成一条折线，此折线称为频率折线，或称为频率多边形。

**3. 累积频率直方图** 图1-6表示一个累积频率直方图。它以横坐标表示变量，其纵坐标直接用累积频率$Fi$表示。只要以各组组距为底边，以$Fi$为高，在每一组上划一长方形，则累积频率直方图即告完成，见图1-6。

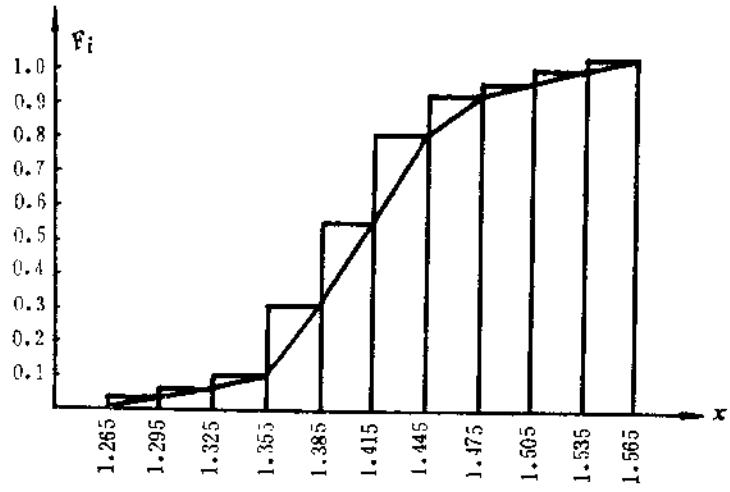


图 1-6