

质量管理的统计方法



科学技术文献出版社

质量管理的统计方法

科学技术文献出版社

内 容 简 介

本书译自日本标准协会出版的质量管理讲座“统计的方法”一书。内容包括产品质量管理中常用的统计思想、频数分布、计数值的二项式分布与泊松分布、母平均推测与正态分布以及分散分析与相关分析等。可供质量管理人员，标准化工作者，工厂工人、技术人员，以及大专院校师生参考。

质量管理的统计方法

(限国内发行)

编 辑 者：中国科学技术情报研究所

出 版 者：科学技术文献出版社

印 刷 者：北 京 印 刷 一 厂

新华书店北京发行所发行 各地新华书店经销

开本 $787 \times 1092 \cdot \frac{1}{32}$ 6.5 印张 137 千字

统一书号：15176·255 定价：0.70 元

1978 年 3 月出版

前 言

遵照伟大领袖毛主席关于“洋为中用”和“批判地吸收其中一切有益的东西”的教导，为适应我国产品质量管理现代化的需要，应广大质量管理人员和标准化工作者要求，本刊编辑部特翻译出版了这本日本标准协会质量管理基础教材。

统计质量管理，是一种各国普遍采用的科学质量管理方法。它是根据产品或半成品质量散差（即波动）的统计性规律，通过抽样检查预测和控制产品的质量，进而达到提高质量、减少次品的目的。

本书涉及到的一些统计数值表（如正态分布表、 t 表、 F 表等等），在这次编辑过程中也作了搜集，作为附录列于书后。此外，还译录了一些有关质量管理术语，以供参考。

由于水平有限，错误之处，请批评指正。

《标准化译丛》编辑部

目 录

前 言	(i)
第一章 统计的概念	(1)
1.1 统计的概念	(1)
1.2 数据的性质	(3)
1.3 散差的原因	(4)
第二章 频数分布	(6)
2.1 频数表	(6)
2.2 频数图	(10)
2.3 分层	(13)
2.4 巴氏曲线	(16)
2.5 易产生缺陷之处的调查方法	(16)
2.6 分散的加法性	(17)
第三章 计数值	(20)
3.1 几率(概率)	(20)
3.2 不合格个数的几率分布(二项式分布)	(31)
3.3 不合格率的推定	(36)
3.4 缺陷数的分布(泊松分布)	(38)
3.5 缺陷数的推定	(41)
3.6 百分率的推定(推算纸)	(42)
3.7 符号检定	(45)
3.8 不合格率之差的检定(分割表)	(48)
第四章 母平均的推测 (σ 已知)	(56)

4.1	正态分布	(56)
4.2	母平均 μ 的推定(σ 已知)	(67)
4.3	关于母平均 μ 的假设之检定(σ 已知)	(77)
4.4	平均差的推定与检定(σ 已知)	(83)
4.5	取样	(86)
第五章	分散的推测	(95)
5.1	不偏分散	(95)
5.2	母分散的检定	(102)
5.3	分散不同的推定与检定	(110)
5.4	用范围作的推定与检定	(114)
第六章	母平均的推测(σ 未知)	(122)
6.1	母平均 μ 的检定	(122)
6.2	母平均差的推定与检定	(129)
第七章	2 变量间的关系	(134)
7.1	2 元频数分布	(134)
7.2	相关系数	(140)
第八章	实验数据的分布	(149)
8.1	一元排列(母参数模型)	(149)
8.2	一元排列(变量模型)	(158)
8.3	二元排列(不重复)	(164)
8.4	二元排列(有重复)	(171)
8.5	分枝排列(多级排列)	(182)
附录:		(188)
χ^2 表		(189)
正态分布表		(190)
F 表(1)		(191)

F 表(2)	(192)
t 表	(193)
r 表	(194)
Z 变换表	(195)
术语与解释	(196)

第一章 统计的概念

1.1 统计的概念

现代质量管理有二个特点，一是以统计的思想和方法为中心，二是有组织的全面管理。统计质量管理一词是强调了第一点，目前广泛采用的有美国休哈特(W. A. Shewhart)提出的管理图法及其他各种统计方法。但是质量管理仅偏重于统计的方法而忽视其他方面，实际效果也不会太大；必须全厂上下协力有组织地进行质量管理。

在质量管理中统计的思想之所以重要，是由于客观“质量”经常在变动。不管用怎样精密的机械多么谨慎地操作，产品质量也一定存在散差(dispersion)。然而这种散差具有统计的规律性。掌握产品质量的统计性分布，再按统计推理进而预测和控制产品的质量，乃是“统计质量管理”的根本。

抓住统计规律，可预见全部产品的情况，用不着去注意单个产品，而且能从杂乱的变动中立刻找出规律性。这里须指出，我们推理的对象并不是被观测到的数据本身，观测的数据与统计推理对象有别。

现举二例说明。在金属丝验收检查时，要从许多金属绕线中取出几个试样测定其抗拉强度，然后根据测试结果决定这批金属线是否全部合格。在此，测得的每个数据是否满足规格要求？可不必关心；主要是根据这些数据推断整批产品是否合格。

再有在人造丝生产过程中，需定时测量一些人造丝的粗细度作 $\bar{x}-R$ 管理图。测丝的粗细，目的也不是要了解每一根纤维的粗细，而是要通过综合这些数据尽早发现其生产过程的异常情况。

注 1. 散差：表示测量值大小的不均一性即不整齐情况。散差的具体数值表示，一般采用标准偏差。

注 2. 批：指按某种目的将产品、半成品或原材料的单位体或单位量集合起来的整体。

为强调推理对象与数据间关系，称提供数据的原始集团——金属绕线的一批或人造丝的一个生产工序等为母集团 (Population)，从母集团中抽取出的部分称为试样 (Sample)。统计的推理就是以试样为依据，进行有关母集团情况的推测。整理这些关系，如图 1.1.1 所示。

注：抽样 (Sampling)；从母集团中抽取试样的活动。也称取样。

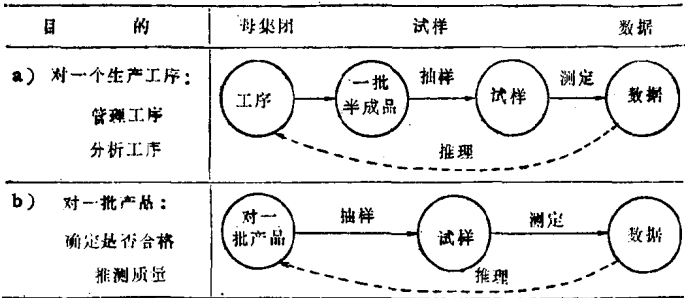


图 1.1.1 统计的推理

从杂乱的散差中找到规律性，常采用的有效方法是将母集团分层。例如，用 A、B 二台炉对钢制品淬火时，A 炉淬火的制品与 B 炉淬火的制品各为一层。测量一下每层制品硬度散差情况，可以获得这样资料：或是 A 层(或 B 层)散差小 B

层(或A层)散差大,或是每一层的制品散差都小,但由于二层间硬度差距大,使总体(A.B二层)散差大。再考虑到各层中的制品在炉中位置情况,就有可能改良加热法等技术。

1.2 数据的性质

进行统计推理必须有数据。而数据的选取要有一定的目的,统计推理则是按此目的进行。

我们测量的数据,分计量值与计数值两种。**计量值**系指作为连续量测得的质量特性值, **计数值**系指以个数数得的质量特性值,如次品件数或疵点数等。分别举例如下:

计量值例:药片1片的重量(克)、金属线的抗拉强度(公斤/平方厘米)、玻璃板的厚度(厘米)、电灯的光通量(流明)。

计数值例:每平方米棉布的疵点数、每个月的故事数、每张玻璃板的气泡数、涂饰不合格的零件数。

此外,圆棒的直径是计量值,但是也有用通止规检查圆棒直径是否在一定的规格限度内,以此划分圆棒合格还是次品的。这种情况下,本来为计量值(棒直径)结果作计数值(棒直径不合格即次品数)处理。在测定计量值费时费钱的场合,常采用这种简易测定方法。

测得的一组数据肯定有散差,表示散差分布的中心通常采用平均值 $\bar{x}$ 。表示分布的宽度即数据散差程度,采用范围 R (最大值减最小值)。

例如有一组数据为:

35, 32, 33, 37, 38

则平均值 $\bar{x}$ 为:

$$\bar{x} = \frac{35 + 32 + 33 + 37 + 38}{5} = 35$$

数据散差范围 R ;

$$R = \text{最大值} - \text{最小值} = 38 - 32 = 6.$$

但是, 从 5 个数据求出的 $\bar{x}$ 与从 100 个数据求出的 $\bar{x}$ 即便数值相同, 两者的可信程度也不同。对同一母集团而言, 可以想像观测数据越多, 数据散差分布宽度范围 R 就相应地越大。因而表示 $\bar{x}$ 与 R 时, 需要指出: 它们是从多少个数据中计算出的。

注: 表示数据的散差, 有时采用均方差的平方根

$$\hat{\sigma} = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}$$

鉴于选取数据是为了根据它进行有关母集团情况的推理, 所以必须让试样取得尽可能正确地反映母集团的状态。试样取的过少或者仅取合格品都没有作用。可是, 完全反映出母集团原有状态的取样, 一般是不可能的, 即使办得到, 也没有必要作这种取样观测。一般采用能使构成母集团的各个部分有相同的机会包含在试样中的抽取方法——即随机取样(random sampling). 采用随机抽取后, 就能够根据概率论进行客观的推理。

1.3 散差的原因

统计质量管理的目的, 也可以说在于减少质量的散差。影响产品质量的因素许许多多, 找到其中最重要的因素: 若是可管理者尽可能努力使它达到均等; 若是不能管理者则让它转变为其他因素, 最终让质量达到一定。

比如人造丝的粗细, 由纤维胶中纤维素浓度、纤维胶粘度(这些与原料有关)、纺纱机的喷出量(与制造工序有关)等决定。这些量一随时间、空间变动, 就要造成人造丝粗细的

散差，也就是说某一瞬间各个喷嘴喷出的人造丝粗细互不相同，同一喷嘴出来的人造丝粗细还随时间变动。因而，从几个喷嘴分几次取出人造丝测量粗细，可获得反映人造丝的粗细散差的数据。

对无数个变动原因皆能进行管理，当然理想，不过实际上办不到，而且也没有必要。细微的变动知其原因而放过也无妨，对于产生大变动的原因，则是必须捕捉住进行管理。对在工序过程中影响产品某一质量特性的几个变动原因间的关系，可用图表示出来，例如图 1.3.1，这种图示有助于我们找到哪些因素应作为管理的目标。

注：特性要因图，是一种表示质量特性与影响这种特性的主要因素之间关系的图。

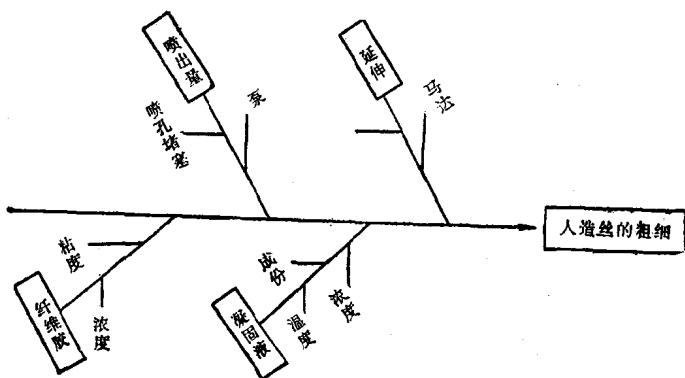


图 1.31 特性要因图

第二章 频数分布

要想找到产生质量散差的原因然后加以管理，首先需正确掌握散差变动的状态。为此，可以研究质量变动的频数分布(frequency distribution),即频数表和频数图。这些做起来很简单，但对理解统计方法却是重要的，而且它本身也有实际作用。

所谓频数表，是将数据中各个值反复出现的次数（即频数）进行排列而成的表；在数据的存在范围可分为若干个区间时，频数表则由属于各个区间的数据的出现次数(频数)排列而成。如以座标图表现频数就是频数图。

2.1 频数表

制造硫酸过程中，需使二氧化硫(SO_2) 气体转化为三氧化硫(SO_3)。现将记录的 100 个 SO_2 气体转化比率即 转化率数据按测定时间顺序排列，如表 2.1.1 所示。转化率大小决定着生产量。转化率散差受气体的浓度、温度、压力、速度等影响。作为公司标准，规定转化率从 91%至 94%。

从表 2.1.1 可知，转化率分布在从最小 90%至最大 94.8%之间；但仅观察这些数据还是不清楚分布状态如何。于是，把数据分布范围划分 7~20 个区间，然后统计各区间所包括的数据个数(见表 2.1.2)。

数据经这样整理而作成的表称为频数表。根据表 2.1.2 可以知道，转化率处在 92%~93%范围内时，频数最多(即

表 2.1.1 SO₂ 气体转化率(%)

No	1~10	11~20	21~30	31~40	41~50	51~60	61~70	71~80	81~90	91~100
1	94.3	92.8	92.7	92.6	93.3	92.9	91.8	92.4	93.2	91.4
2	93.4	92.2	93.0	92.9	92.2	92.4	92.2	92.8	94.8	90.0
3	92.4	93.9	93.5	93.6	93.0	93.4	91.4	94.2	93.8	90.6
4	92.8	93.2	92.2	92.5	93.6	93.9	92.4	91.8	92.8	91.6
5	93.8	93.6	92.1	92.0	92.6	92.0	91.8	90.8	91.2	93.7
6	94.0	92.8	92.8	91.2	93.0	93.1	93.0	92.6	92.8	94.7
7	94.2	93.2	93.4	92.0	92.8	93.4	92.0	92.4	92.7	92.4
8	92.9	91.8	92.1	94.6	91.4	91.0	92.1	92.2	93.4	92.2
9	92.8	92.8	92.0	93.8	91.2	93.2	91.9	93.0	92.8	91.9
10	93.1	92.0	92.4	93.5	92.0	92.8	92.4	92.4	93.2	94.0

数据出现次数最多)且都符合公司标准;低于标准下限(91%)者为 3 次,高于标准上限(94%)者为 8 次。

注: 频数表内还可列出有助于计算平均与标准偏差的项目 (参见下面表 2.1.3)。

作频数表时注意:

(a) 分组数 (划分的区间数) 7~20 个较为适宜。组数过少不能很好掌握分布状态;组数超过 20 以上, 则反倒难看清分布情况。

(b) 组的边界不要重叠。
上例中各组边界为: 90.0~90.4, 90.5~90.9, 91.0~91.4, ……。也可以将组边界

取在: 89.95, 90.45, 90.95, 91.45, ……。

(c) 有时要根据作出的频数表, 把现在的工序状态与规

表 2.1.2 频数表

区 间	记 号	频数
90.0~90.4	I	1
90.5~90.9	I	2
91.0~91.4	III	7
91.5~91.9	III	7
92.0~92.4	IIII IIIII IIIII	25
92.5~92.9	IIII IIIII IIIII I	21
93.0~93.4	IIII IIIII III	18
93.5~93.9	IIII III I	11
94.0~94.4	III	5
94.5~94.9	III	3

格值相对照；这种场合希望规格值尽可能接近边界值附近。

(d) 作频数表，原则上要取 100 个以上的数据，在不得已时也可取 50 个。

作出频数表能大体看清数据散差的样子，但要从数量上概括表现出来，得采用平均 m 与标准偏差(Standard deviation) σ 。计算方法如下：

步骤 1° 列出各区间数据的代表值(中心值) x_0 。

步骤 2° 利用公式 $u = \frac{x - x_0}{h}$ 将代表值 x 换算成简单数 u 。式中 x_0 为数据散差的大致中央代表值(此处 $x_0 = 92.7$)， h 为区间宽度(此处 $h = 0.5$)。可见 u 是一衡量各区间代表值 x 偏离中央代表值 x_0 有几个区间宽度 h 的数，显然 x_0 处对应的 $u = 0$ 。

步骤 3° u 乘以频数 f ，然后各个 uf 相加。对于此处的数据， uf 之和 $= -6$ 。

步骤 4° uf 再乘以 u 然后相加。此处 u^2f 之和 $= 328$ 。

经过以上步骤 1° 至 4°，可作出表 2.1.3。

步骤 5° uf 之和除以数据总个数 N ，所得商数为 m_u ，则

$$m_u = \frac{uf \text{ 之和}}{N} = \frac{-6}{100} = -0.06。$$

步骤 6° 从 u^2f 之和中减去 $(uf \text{ 之和})^2/N$ 然后除以 $(N-1)$ ，最后取其平方根。称该平方根为 σ_u ，则：

$$\begin{aligned}\sigma_u &= \sqrt{\frac{(u^2f \text{ 之和}) - (uf \text{ 之和})^2/N}{N-1}} \\ &= \sqrt{\frac{328 - (-6)^2/100}{99}} = 1.82\end{aligned}$$

步骤 7° x 的平均以及标准偏差 σ 可由 m_u 及 σ_u 求得：

表 2.1.3 利用频数表计算平均与标准偏差

区 间	代表值 x	$u = \frac{x - x_0}{h}$	记 号	频数 f	uf	u^2f
90.0~90.4	90.2	-5		1	-5	25
90.5~90.9	90.7	-4		2	-8	32
91.0~91.4	91.2	-3		7	-21	63
91.5~91.9	91.7	-2		7	-14	28
92.0~92.4	92.2	-1		25	-25	25
92.5~92.9	92.7	0		21	0	0
93.0~93.4	93.2	1		18	18	18
93.5~93.9	93.7	2		11	22	44
94.0~94.4	94.2	3		5	15	45
94.5~94.9	94.7	4		3	12	48
计	—	—	—	100	-6	328

$$m = x_0 + hm_u = 92.7 + 0.5 \times (-0.06) = 92.67$$

$$\sigma = h\sigma_u = 0.5 \times 1.82 = 0.91$$

平均 m 示出全体数据的中心位置,标准偏差 σ 则是表现数据散离平均的程度。

现在,假设在平均 m 的两侧各取一 σ 宽度(标准偏差的1倍),则得到区间 91.76~93.58[即 $(m - \sigma) \sim (m + \sigma)$]。在原始数据(表 2.1.1)中高于这区间上限的数据有 17 个:

94.3, 93.8, 94.0, 94.2, 93.9, 93.6, 93.6, 94.6, 93.8, 93.6, 93.9, 94.2, 94.8, 93.8, 93.7, 94.7, 94.0

低于这区间下限的数据有 11 个:

91.2, 91.4, 91.2, 91.0, 91.4, 90.8, 91.2, 91.4, 90.0, 90.6, 91.6

两者加在一起共 28 个,占全体的 28%。再在平均 m 的两侧各取 2σ 宽度(标准偏差的 2 倍),则得区间 90.85~94.49

[即 $(m-2\sigma)\sim(m+2\sigma)$], 原始数据中, 高于这区间上限的数据有 3 个:

94.6, 94.8, 94.7

低于区间下限的数据也有 3 个:

90.8, 90.0, 90.6

两者加起来有 6 个数据处于该区间之外, 占全体的 6%。进一步在平均 m 的两侧各取 3σ 宽度(标准偏差的 3 倍), 得区间 89.94~95.40。原始数据中超出该区间之外的数据一个也没有。

一般说来, 若在平均 m 上下各取一个 σ 宽度, 则处于该区间之外的数据约有 30%; 若取 2 个 σ 宽度, 处于该区间之外的数据约 5%; 取 3 个 σ 宽度, 区间之外的数据几乎 1 个也没有(1000 个数据中约有 3 个)。

注: 关于上述事实参照第 4 章的正态分布。

2.2 频数图

频数表对了解质量分布有很大作用, 而若将频数用图表示之, 质量分布的情况就更加一目了然。

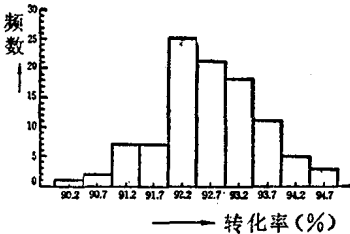


图 2.2.1 矩形频数图

此外, 如果以纵轴表示累积频数(Gumulative frequency)

以坐标横轴表示质量特性值, 纵轴表示频数, 通常称这样作出的图为频数图(或称矩形频数图 histogram)。如将频数图与规格值相对照, 频数图能帮助我们判断产品质量状态是否合格。表 2.1.2 作成频数图见图 2.2.1。