

■国家自然科学基金
■航空科学基金 资助出版

统计过程 控制

Statistical Process
Control

蒋家东 冯允成 著



 中国质检出版社

国家自然科学基金
航空科学基金 资助出版

统计过程控制

蒋家东 冯允成 著

中国质检出版社

北京

图书在版编目 (CIP) 数据

统计过程控制/蒋家东,冯允成著.—北京:中国质检出版社,2011.8
ISBN 978-7-5026-3400-1

I. ①统… II. ①蒋…②冯… III. ①企业管理-统计控制-过程控制
IV. ①F273.2

中国版本图书馆 CIP 数据核字 (2010) 第 258497 号

内 容 提 要

本书系统研究了统计过程控制的基本原理、统计方法、常用工具和关键技术,总结了当前国内外在统计过程控制领域的最新发展。主要内容包括:过程描述与图形表示,过程分布假设检验,过程能力分析,单变量统计控制图,多变量统计控制图,主成分统计过程控制,以及非线性与动态统计过程控制等。

本书既可以作为质量管理专业科研人员的研究参考书,又可以作为高等院校相关专业的本科生和研究生的学习参考书,也可以作为企业质量管理与工程技术人员开展统计过程控制的工具书和培训教材。

中国质检出版社出版发行

北京市朝阳区和平里西街甲 2 号 (100013)

北京市西城区复外三里河北街 16 号 (100045)

网址: www.spc.net.cn

电话: (010) 64275360 68523946

中国标准出版社秦皇岛印刷厂印刷

各地新华书店经销

*

开本 787×1092 1/16 印张 25.5 字数 465 千字

2011 年 8 月第一版 2011 年 8 月第一次印刷

*

定价: 58.00 元

如有印装差错 由本社发行中心调换

版权专有 侵权必究

举报电话: (010) 68510107

前 言

随着经济全球化的加速发展，国际国内市场深度融合，企业经营面临的外部环境更加复杂，市场竞争更加激烈，产品质量日益成为事关企业生存与发展的关键与核心问题。当前，全球经济正处在后金融危机时代增长乏力的困境之中，各种危机此起彼伏，围绕有限市场需求的争夺呈现白热化，打造产品在质量上的竞争力客观上成为企业吸引顾客，拓展市场，赢得竞争，做大做强的经营王道。为满足顾客不断升级的消费需求和与日俱增的质量期望，企业就必须全力加强质量管理，持续保证和提高产品质量。推广先进质量管理方法，实施统计过程控制，确保产品研制与生产过程的稳定受控，正是保证产品质量的重要手段。

发端于 20 世纪 20 年代的统计过程控制，已有近百年的发展，但仍然保持了鲜活的生命力。统计过程控制是基于过程波动的统计规律性，利用统计控制图对导致过程波动的影响因素进行分析和控制，并采取经济而适宜的工程措施保持过程在稳定状态下运行，使过程输出持续一致地满足顾客要求的科学方法。统计理论的不断发展和计算机应用的日益成熟，使得统计过程控制的理论基础更加完善，技术方法更加先进，适用范围更加广泛。经过 90 多年的发展，统计过程控制已经在机械、纺织、汽车、电子、航空等离散制造工业中得到了广泛应用，并逐渐拓展到造纸、炼油、化工、食品等连续、半连续制造工业，成为在线质量诊断与控制方面的重要技术方法。进入 21 世纪以来，由于传感技术、在线测量技术和计算机技术在产品研制与生产过程中的大量应用，对过程数据的实时采集变得非常容易，短时间内获取大量数据逐步成为现实，对这些过程数据的高效利用，成为在线质量诊断与控制方面一个十分迫切的需求。同时，随着回归分析、多变量分析、时间序列分析和非参数统计分析等多变量统计分析方法的不断成熟，统计过程控制已经从单变量情形深化发展到了多变量情形，在分析、诊断和监控研制与生产过程的质量绩效方面发挥了越来越大的作用。

作为质量管理领域的科研人员，我们对统计过程控制的关注、研究和应用已经走过了 15 个年头。自 20 世纪 90 年代中期以来，我们在航空工业不断推进统计过程控制的研究与应用，既取得了一定的成效，也遇到了很多的问题。尽管统计过程控制的思想方法、基本原理和技术工具是完全正确的，但其实际应用的效果却不尽如人意。总结统计过程控制在实际应用中遇到的困难，主要有四个方面：一是意识问题。与过去相比，我国企业的质量管理已经有了长足的进步，但长期形成的问题管理模式在许多企业仍居支配地位，其特点是重视对现有问题的处理，忽视对潜在问题的预防，致使基于预防思想的统计过程控制得不到企业管理层的真正重视。二是方法问题。统计过程控制是一个理论性较强的质量管理方法，需要较深的统计知识，只有正确使用才能产生正确的结果，但由于企业有关人员水平不高加之培训不足，不少企业对统计过程控制的应用比较粗糙，存在许多错误应用，导致难以产生积极的效果。三是工具问题。统计过程控制的有效应用离不开测量自动化、分析软件化和处理工程化，这对产品研制与生产过程的质量管理信息化要求很高。但质量管理信息化水平不高恰恰是我国企业经营管理中的薄弱环节，存在数据采集、处理和分析大量依靠人工，不仅误差大、时效性差，而且分析诊断能力弱，难以与产品研制和生产过程保持同步，导致统计过程控制与产品研制和生产过程两张皮。四是数据问题。统计过程控制用得好，必须有大量的历史数据。但我国企业的质量管理相对比较粗放，不太重视数据的积累和长期使用，导致数据量不足，难以形成科学的分析与预测。思考这些问题出现的根源，我们认为主要还是企业的质量管理与工程技术人员在统计过程控制方面的知识不足，没有形成基于数据决策的统计思维，为此我们决定写作本书，介绍和宣传统计过程控制的专业知识和最新发展。

与已有的有关统计过程控制的书相比，本书的写作力图体现三个特点：一是突出原创性。本书根据我们这些年来在统计过程控制方面的研究与应用实践，将一些最新的研究成果和应用案例纳入进来。本书中的一些案例来自于我们在企业的实际应用。二是突出系统性。本书从统计过程控制应用的内在规律入手，内容包括过程描述与图形表示、过程分布假设检验、过程能力分析、单变量统计控制图、多变量统计控制图、主成分统计过程控制以及非线性与动态统计过程控制等，基本上涵盖了统计过程控制从数据采集到统计控制图应用，从假设检验到能力分析，从单变量个别控制到多变量综合控制

等各个领域。三是突出前沿性。本书介绍了国内外在统计过程控制领域的最新研究进展，纳入了多变量过程能力分析 & 统计控制图设计方面的一些最新研究成果，部分研究属于探索性的研究应用。

本书的写作和出版得到了中国航空综合技术研究所、中国质检出版社、国家自然科学基金的大力支持。我们要特别感谢中国航空综合技术研究所的梁丽涛所长、吉世强副总工程师以及商广娟、郑立伟、汪邦军等同事们的支持和帮助，他们在相关领域的研究成果给我们很多借鉴和启发，使我们能够方便地了解 & 掌握统计过程控制领域的许多重要的研究信息和第一手资料。也要特别感谢中国质检出版社的黄洁编辑，她的努力和支持使得本书得以顺利出版。还要特别感谢我们的家人，他们的理解、支持和鼓励使我们能够全身心地投入到本书的写作之中。

受视野、信息、知识以及偏好所限，我们对统计过程控制的理解与研究一定还存在许多盲点与不足，甚至本书中的有些提法仍然值得商榷，一些研究也有待深化，对此我们真诚地希望得到每一位读者的谅解。我们热诚欢迎广大读者对本书中的不妥乃至错误之处提出质疑，并随时准备解释和讨论。我们希望本书的出版，能够引起更多的专家、学者和工程技术人员关注质量管理问题，更加重视统计过程控制的研究与应用，也希望本书的出版对我国企业实施统计过程控制有所帮助。

作者

2011年5月

目 录

第一章 过程描述与图形表示	(1)
1.1 数据收集	(1)
1.1.1 数据类型	(1)
1.1.2 数据抽取	(3)
1.2 一维随机样本的概念及表示	(5)
1.2.1 一维随机样本的数字特征	(6)
1.2.2 一维随机样本的图形表示	(6)
1.2.3 一维正态分布及其性质	(21)
1.2.4 中心极限定理	(21)
1.3 多维随机样本的概念及表示	(22)
1.3.1 多维随机向量及其性质	(23)
1.3.2 多维随机样本的定义与特点	(24)
1.3.3 多维随机样本的图形表示	(25)
1.3.4 多维随机样本的数字特征	(33)
1.4 多维正态分布	(34)
1.4.1 多维正态分布的定义	(35)
1.4.2 多维正态分布的性质	(36)
1.4.3 多维正态分布的参数估计	(36)
1.4.4 Wishart 分布及其基本性质	(37)
1.5 多维数据的正态性变换	(38)
1.5.1 一维边缘数据的正态性变换	(38)
1.5.2 多维数据的整体正态性变换	(39)
本章小结	(40)
第二章 过程分布假设检验	(41)
2.1 假设检验的原理与步骤	(41)

2.1.1	假设检验的基本原理	(42)
2.1.2	假设检验的步骤	(44)
2.2	一维总体及其参数的假设检验	(45)
2.2.1	正态总体及其参数的假设检验	(45)
2.2.2	二项分布参数的假设检验	(61)
2.2.3	泊松分布参数的假设检验	(65)
2.2.4	一维总体假设检验的模型选择	(68)
2.3	多维正态分布及其参数的假设检验	(68)
2.3.1	多维正态分布的假设检验	(69)
2.3.2	多维正态分布参数的假设检验	(73)
2.4	实例计算与分析	(79)
	本章小结	(85)
第三章	过程能力分析	(89)
3.1	单变量过程能力分析	(89)
3.1.1	单变量过程能力指数	(90)
3.1.2	过程能力指数与不合格品率之间的关系	(93)
3.1.3	给定置信水平下的过程能力指数	(95)
3.1.4	过程能力分析的作用与步骤	(96)
3.2	多变量过程能力分析	(97)
3.2.1	规范要求的表示	(98)
3.2.2	过程能力的表示	(101)
3.2.3	对规范要求的进一步修正	(102)
3.2.4	多变量过程能力指数	(105)
3.3	其他多变量过程能力指数	(109)
3.3.1	基于单变量分析的多变量过程能力指数	(109)
3.3.2	基于不合格品率的多变量过程能力指数	(110)
3.3.3	基于向量表示的多变量过程能力指数	(113)
3.3.4	基于体积比的多变量过程能力指数	(115)
3.3.5	基于椭球体规范的多变量过程能力指数	(118)
3.4	不同多变量过程能力指数算法的比较	(121)
3.5	对多变量过程能力指数问题的进一步讨论	(128)

本章小结	(131)
第四章 单变量统计控制图	(133)
4.1 统计过程控制技术的发展	(133)
4.2 统计过程控制的理论基础	(134)
4.3 统计过程控制的型式化应用框架	(136)
4.3.1 统计控制图的典型格式	(136)
4.3.2 统计控制图的两类错误	(138)
4.3.3 统计控制图的设计原理	(140)
4.3.4 判断过程正常的准则	(141)
4.3.5 判断过程异常的准则	(142)
4.3.6 基于波动理论的统计控制图监控原理	(149)
4.4 常用单变量统计控制图	(151)
4.4.1 均值—极差控制图	(153)
4.4.2 均值—标准差控制图	(167)
4.4.3 中位数—极差控制图	(168)
4.4.4 单值—移动极差控制图	(169)
4.4.5 不合格品率控制图	(170)
4.4.6 不合格品数控制图	(176)
4.4.7 缺陷数控制图	(177)
4.4.8 单位缺陷数控制图	(178)
4.5 统计控制图的使用	(179)
4.5.1 统计控制图的作用	(180)
4.5.2 应用统计控制图时需注意的问题	(180)
4.5.3 小批量问题的技术处理	(183)
4.5.4 面向输入参数的统计过程控制	(185)
本章小结	(186)
第五章 多变量统计控制图	(189)
5.1 单纯使用单变量统计控制图的缺陷	(189)
5.2 多变量统计控制图技术的发展	(191)
5.2.1 多变量统计控制图的研究现状	(191)
5.2.2 多变量过程的典型失控形态	(192)

5.2.3	多变量统计控制图的设计原理与适用范围	(195)
5.3	多变量均值控制图	(197)
5.3.1	多变量 χ^2 控制图	(198)
5.3.2	多变量 T^2 控制图	(198)
5.3.3	小样本多变量 T^2 控制图	(205)
5.3.4	多变量均值控制图中失控信号的诊断	(209)
5.3.5	多变量单值控制图	(211)
5.4	多变量离差控制图	(216)
5.4.1	多变量 $ S $ 控制图	(217)
5.4.2	多变量 W 控制图	(219)
5.4.3	多变量三向控制图	(221)
5.5	多变量累积和控制图	(227)
5.5.1	基于 T 统计量的 MCUSUM 控制图	(228)
5.5.2	向量累积和控制图	(228)
5.5.3	多变量累积和控制图的应用步骤	(229)
5.6	多变量指数加权移动平均控制图	(234)
5.6.1	基于 Z 统计量的 MEWMA 控制图	(234)
5.6.2	多变量 MEWMA 控制图的应用步骤	(235)
5.7	典型多变量控制图监控效果的比较	(237)
5.8	基于协方差阵特征结构的新型控制图研究	(245)
	本章小结	(257)
第六章	主成分统计过程控制	(258)
6.1	主成分分析的基本原理	(258)
6.1.1	主成分的几何意义	(259)
6.1.2	主成分分析的一般数学模型	(261)
6.2	主成分的求解与性质	(262)
6.3	样本主成分的导出	(264)
6.4	主成分分析在多变量过程控制中的应用	(267)
6.4.1	基于主成分模型的统计控制图	(268)
6.4.2	主成分多变量过程控制实例	(272)
6.4.3	对主成分多变量控制图诊断能力的分析	(279)

本章小结	(281)
第七章 非线性与动态统计过程控制	(282)
7.1 非线性多变量统计过程控制	(282)
7.1.1 滑动主成分分析模型	(283)
7.1.2 对主成分关系变化的监控	(288)
7.1.3 非相似性分析模型	(291)
7.2 动态多变量统计过程控制	(296)
7.2.1 多向主成分分析模型	(296)
7.2.2 多向偏最小二乘模型	(298)
7.2.3 动态主成分分析模型	(299)
7.2.4 动态独立成分分析模型	(302)
本章小结	(305)
附录 A 有关概率统计的知识	(307)
A.1 概率基础	(307)
A.2 概率的定义	(309)
A.3 随机变量及其分布	(311)
A.4 随机变量的数字特征	(315)
A.5 常用统计量及其分布	(316)
A.6 总体参数的估计	(319)
附录 B 有关矩阵的知识	(328)
B.1 正定矩阵及其性质	(328)
B.2 分块矩阵及其代数运算	(331)
附录 C 常用分布表	(338)
附录 D 常用系数表	(359)
附录 E 三维正态随机数据	(362)
参考文献	(372)

第一章

过程描述与图形表示

过程是自然界和人类社会生活中一切变化或活动的基础。企业的生产经营活动也是由一系列相互衔接的过程构成的。正是通过过程的转化作用，一组输入特性被转化为另一组输出特性。过程的状态决定了转化的效率与效果。研究过程的状态，本质上就是要研究过程输出特性的状态。由于各种偶然因素的影响，尽管人们对特定过程的具体功能有先验的认识或预先的设计，但过程输出特性的状态常常具有某种程度的不确定性或随机性。通过采集和分析过程数据，可以估计过程的总体分布及分布参数，评价过程参数的变化，识别导致过程状态发生变化的特殊因素。对过程进行统计描述和图形表示是研究过程状态的基本方法。对于单变量过程，过程数据通常是一维的，其统计描述与图形表示一般比较简单。对于多变量过程，过程数据通常是多维的，其统计描述与图形表示变得较为复杂。在高维情况下，多变量过程数据的图形表示更是非常困难。

1.1 数据收集

实施过程控制，自然必须获得反映过程状况的基础数据。为此，必须有目的地进行过程数据的收集和整理，为进一步进行统计分析创造条件。对于单变量过程，采集的过程数据往往是一维的。对于多变量过程，采集的过程数据往往是多维的。由于实施统计过程控制的过程输出特性可能是经过数学变换转化或统计降维处理的新特性或综合特性，过程实际的控制维度与收集数据的维度可能并不一致，取决于所采取的具体控制方法。

1.1.1 数据类型

由于过程的类型千差万别，能够实际收集到的数据类型可能有很大差异，在收集过程数据时必须明确过程数据的计量标度。按照计量的精确程度，可

将计量标度分为四种类型：

1. 定类标度

定类标度 (nominal scale) 也称名义标度或类别标度，是精度最低的计量标度。定类标度只能按照过程的某种属性进行平行的分类。例如，对某切削加工过程中使用的刀具按照材质分为铝合金、铁合金等。定类标度只是测度了事物之间的类别差异，其各类之间是平等的并列关系，无法区分优劣、大小或高低，各类别之间的顺序是可以改变的。由于定类标度只能区分事物是同类或不同类，因而只具有相等或不相等的数学特性。虽然定类标度的计量结果只表现为某种类别，但为了便于统计处理，可以对不同类别用不同的数字或编码来描述。例如，对于某发动机缸体的浇铸缺陷，可以用“1”表示砂眼、用“2”表示裂纹、用“3”表示尺寸误差等。这些数字只是给不同类别的一个代码，并不意味着这些数字可以进行数学运算。对定类标度的计量结果，通常是通过计算出每一类别中各元素或个体出现的频率或频数来进行统计分析。

2. 定序标度

定序标度 (ordinal scale) 也称顺序标度，是对事物之间等级或顺序差别的一种测度。定序标度不仅可以将事物分成不同的类别，而且还可以确定这些类别的优劣或顺序。例如，钢杆的硬度等级就是对其刚性高低的一种定序测度。显然，定序标度对过程的计量要比定类标度精确一些，但它也只是测度了不同类别之间的顺序，而未测量出类别之间的准确差值，其计量结果虽然能对事物分门别类和比较大小，但也不能进行常规的数学运算。

3. 定距标度

定距标度 (interval scale) 也称间隔标度，不仅能将事物区分为不同类型并进行排序，而且可以准确地指出不同类别之间的差距是多少。定距标度是对事物类别或次序之间间距的测度，通常使用自然或度量衡单位作为计量标度。例如，用摄氏度或华氏度来度量发动机燃烧过程中的温度变化。由于定序标度的每一间隔都是相等的，只要给出一个度量单位，就可以准确地指出两个计量数值之间的差值。由于定距标度的计量结果表现为数值，并可以计算差值，因而它不仅具有定类标度和定序标度的特性，其结果还可以进行加、减等数学运算。

4. 定比标度

定比标度 (ratio scale) 也称为比率标度, 它与定距标度属于同一层次, 一般不作区分, 其计量的结果也表示为数值。它除了具有上述三种计量标度的全部特性外, 还具有一个特性, 那就是可以计算两个测量值之间的比值。这就要求定比标度中必须有一个绝对的“零点”, 这也是它与定距标度的唯一差别。采用定比标度计量的结果通常不会出现“0”值。定比标度的计量结果可以进行各种数学运算。

显然, 采用不同的计量标度就会得到不同类型的过程数据。甚至对于同一过程, 由于涉及多个输出特性, 不同的输出特性可能需要使用不同的计量标度进行数据的测定。由此, 可以得到四种类型的过程数据: ①定类数据, 表现为类别, 但不区分顺序, 是由定类标度计量形成的; ②定序数据, 表现为类别, 但有顺序, 是由定序标度计量形成的; ③定距数据, 表现为数值, 可进行加、减运算, 是由定距标度计量形成的; ④定比数据, 表现为数值, 可进行加、减、乘、除运算, 是由定比标度计量形成的。

前两类数据反映的是过程的类别特征, 不能用数值表示, 其结果表现为类别, 称为计数型数据或离散型数据。后两类数据说明的是事物的数量特征, 能够用数值来表示, 因此称为计量型数据或连续型数据。

1.1.2 数据抽取

1. 总体与样本

总体也称母体, 是指研究对象的全体。组成总体的每一个单位称为个体。总体中所包含的个体数量称为总体容量, 用 N 表示。

在统计过程控制中, 研究的对象是过程状况, 而过程状况是通过对过程的输入、输出特性值的研究而得到的。由此, 从统计学的角度看, 总体是过程中每个个体数据的全体。例如, 为了研究分析某个生产过程是否统计受控, 由该过程形成的所有产品的全体就构成一个总体。总体可以是有限的, 也可以是无限制的。如果总体的个体数量是有限可数的, 称为有限总体; 如果总体的个体数量是不可数的, 称为无限总体。

样本也称子样, 是由从总体中抽取的部分个体组成的。组成样本的每个个体称为样品, 样本中所包含的样品数量称为样本容量或样本大小, 用 n 表示。一般用 X_1, X_2, \dots, X_n 表示总体的一个随机样本, 样本数据的取值记

为 x_1, x_2, \dots, x_n , 称为样本观测值。有时候, 为了表达方便, 对随机样本及其实际观测值不作区分, 均用 X_1, X_2, \dots, X_n 表示。

2. 随机抽样

在过程控制中, 通常不太可能为了掌握由某一过程输出的所有产品的信息而对整批产品全部进行检验, 特别是需要进行破坏性检验时。通常的做法是, 从全部产品中随机抽取一定数量的样品进行检验, 将样品检验结果组成样本数据, 通过对样本数据的分析来推断整批产品的状况。在统计过程控制中, 要通过对样本数据的研究来推断总体的状况, 这就要求样本的取得对总体具有代表性, 以便进行统计推断的误差尽量小。为了达到控制统计误差的目的, 通常必须进行随机抽样。

所谓随机抽样, 就是从总体中随机抽取一定数目的单位作为样本进行观察, 使每个总体单位都有一定的概率被选入样本, 从而使得根据样本所做出的结论对总体具有充分的代表性。随机抽样能有效地避免主观抽样带来的系统性误差, 确保样本数据能够有效地估计总体的数量特征, 识别过程的真实状态, 正确地选择和设置过程控制的方法, 并准确地进行故障诊断和异常原因识别和纠正。常用的随机抽样方法包括: 简单随机抽样、系统抽样、分层抽样、整群抽样等。

(1) 简单随机抽样。简单随机抽样是最基本的概率抽样方法, 指总体中的每一个个体被抽到的机会是相同的。在简单随机样本中, X_1, X_2, \dots, X_n 是相互独立的随机变量, 并且与总体有相同的概率分布。简单随机变量可以通过重复从总体中随机抽取得到。显然, 有放回地随机抽取得到的样本是简单随机抽样; 无放回地随机抽取得到的样本, 当样本容量 n 相对于总体容量 N 较小时, 或者总体容量 N 充分大时, 可以近似地看成简单随机样本。为了实现简单随机抽样, 可以采用抽签、查随机数表和掷随机数骰子等方法。

简单随机抽样的优点是, 方法简单直观, 计算抽样误差及对总体参数进行推断比较方便; 缺点是, 抽样程序比较复杂, 在实际工作中真正做到总体中每个个体被抽到的机会完全一样是不太容易的。

(2) 系统抽样。系统抽样也称机械抽样或等距抽样, 即将总体单位按某一标志排序, 然后按一定间隔来随机抽取样本单位。例如, 要从 100 件产品中抽取 10 件组成样本, 首先将 100 件产品随机按 1~100 顺序编号; 然后用抽签或查随机数表的方法确定 1~10 号中入选样本的产品编号, 假定为 4 号;

再按等距原则依次确定入选样本的产品编号为 14, 24, 34, 44, 54, 64, 74, 84 和 94; 最后由编号为 04, 14, 24, 34, 44, 54, 64, 74, 84 和 94 的 10 件产品组成样本。

系统抽样的优点是, 实施方便, 同时能够保证样本对总体的代表性。适合大量生产的流水线上产品的抽查; 缺点是, 若总体单位排序后呈现一定的规律性甚至周期性, 而抽样间隔的周期正好与之吻合, 依赖于这样排列的系统抽样就会产生系统性的偏差。在整体考虑了组内波动和组间波动的原则下, 系统抽样方法常应用于统计过程控制中的子组样本抽取中。

(3) 分层抽样。分层抽样也称为类型抽样, 即先将总体按某些重要的标志分成互不交叉重叠的若干层, 然后在各层中采用简单随机抽样或其他抽样方式抽取若干个样本单位, 由各层的样本单位组成一个样本。例如, 假设某企业 A、B、C 三个车间都生产同一种产品, 若 A 车间的生产量占全部产品产量的 30%, B 车间占 50%, C 车间占 20%, 需要抽取 100 件的样本。采用分层抽样可从 A 车间随机抽取 30 件、从 B 车间随机抽取 50 件, 从 C 车间随机抽取 20 件, 合起来共 100 件组成样本。

分层抽样的优点是, 样本代表性好, 抽样误差小; 缺点是, 抽样手续较为复杂。分层抽样方法常适用于产品质量的验收。

(4) 整群抽样。整群抽样也称集团抽样, 即在总体中, 不是抽取个别样品, 而是随机抽取整群的样品。这种方法是先将总体按某个标志分成若干群, 然后随机地抽取出部分群, 并由抽中的群中的所有个体组成样本。例如, 对某种产品抽取 5% 进行质量检查, 每隔 20 小时抽出 1 小时的产量组成样本, 然后对这些抽出来的样本进行检验并由此推断总体的质量状况。

整群随机抽样的优点是, 抽样实施比较方便; 缺点是, 由于样本只是来自个别几个群, 不能均匀地分布在总体中, 因而代表性差, 抽样误差大。

1.2 一维随机样本的概念及表示

在单变量过程的统计控制中, 往往需要研究一维随机样本的统计特征, 并根据这些统计特征判断单变量过程的状态变化。在随机样本的统计特征研究中, 最基本的就是要利用样本数据识别和推断过程的总体分布或分布参数, 进而针对不同的总体分布选择和建立相应的统计控制方法。其中最常用的就

是一维正态分布及其参数的估计和推断。

1.2.1 一维随机样本的数字特征

假设 X 为一个具有连续型数据的单变量过程, X_1, X_2, \dots, X_n 为来自与 X 相同总体的一个随机样本, 则称 X_1, X_2, \dots, X_n 为过程 X 的一个随机样本。如果过程 X 的概率密度函数为 $f(x)$, 且均值和方差存在, 则称 $E(X)$ 和 $D(X)$ 为过程 X 的总体均值和总体方差, 用 μ 和 σ^2 表示, 且有:

$$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx \quad (1-1)$$

$$\sigma^2 = D(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx \quad (1-2)$$

对于任意实常数 a 和 b , 过程 X 的总体均值 $E(X)$ 和总体方差 $D(X)$ 具有如下性质: ① $E(aX+b) = aE(X) + b$; ② $D(aX+b) = a^2 D(X)$ 。

对于总体 X 的一个随机样本 X_1, X_2, \dots, X_n , 称 \bar{X} 和 s^2 为样本均值和样本方差, 且有:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1-3)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (1-4)$$

样本均值 \bar{X} 和样本方差 s^2 分别具有如下性质: ① $E(\bar{X}) = E(X) = \mu$; ② $D(\bar{X}) = D(X)/n = \sigma^2/n$; ③ $E(s^2) = D(X) = \sigma^2$ 。由此, 将样本均值 \bar{X} 和样本方差 s^2 分别作为均值 μ 和方差 σ^2 的估计量, 即 $\hat{\mu} = \bar{X}$ 和 $\hat{\sigma}^2 = s^2$ 。

1.2.2 一维随机样本的图形表示

为了对单变量过程进行直观的研究, 通常需要借助于图形方法。目前, 常用的描述一维随机变量的样本数据的常用图形工具有直方图和运行图。前者有助于刻画一维随机样本的统计分布特征, 后者有助于刻画单变量过程随时间变化的趋势和演变模式。为了对单变量过程作进一步的分析, 还可以在收集过程相关影响因素的基础上, 绘制散布图和排列图。其中, 散布图可以展示成对的随机变量之间可能存在的相关关系, 排列图可以展示影响单变量