SHUINIQIYE TONGJISHOUCE

水泥企业统计手册

主　编　倪竹君　夏莉娜　张绍周
副主编　秦世景　姜胜平　徐觉慧

Shuini qiye tongji shouce

中国建材工业出版社

水泥企业统计手册

主　编　倪竹君　夏莉娜　张绍周
副主编　秦世景　姜胜平　徐觉慧

中国建材工业出版社

图书在版编目（CIP）数据

水泥企业统计手册／倪竹君，夏莉娜，张绍周编著.
北京：中国建材工业出版社，2005.1
ISBN 7－80159－790－7

Ⅰ. 水… Ⅱ. ①倪…②夏…③张… Ⅲ. ①水泥—质量控制—数理统计—手册 Ⅳ. TQ172－62

中国版本图书馆 CIP 数据核字（2005）第 001217 号

内 容 简 介

本书对水泥企业产品质量控制中常用的数理统计方法进行了系统介绍，对统计学中深奥的数学过程不做过多的推导和介绍，而是紧密结合水泥企业的生产实际，列举了大量的实例和图表，力求做到概念明确、步骤清晰、通俗易懂、实用性强，便于读者阅读和掌握。

本书是水泥企业开展质量管理工作必备的工具书，可作为水泥企业工程技术人员、化验室质量管理人员、检验人员、统计人员的培训教材，也可供水泥质检机构管理人员和建材专业学生学习数理统计课程时参阅。

水泥企业统计手册
主　编　倪竹君　夏莉娜　张绍周
副主编　秦世景　姜胜平　徐觉慧

出版发行：中国建材工业出版社
地　　址：北京市西城区车公庄大街 6 号
邮　　编：100044
经　　销：全国各地新华书店
印　　刷：北京鑫正大印刷有限公司
开　　本：787 mm×1092 mm　1/16
印　　张：18.75
字　　数：474 千字
版　　次：2005 年 1 月第 1 版
印　　次：2005 年 1 月第 1 次
定　　价：38.00 元

网上书店：www. ecool100. com
本书如出现印装质量问题，由我社发行部负责调换。联系电话：（010）88386904

前　言

水泥是事关国民经济建设和国防现代化建设的重要产品，各级建材行业主管部门历来十分重视对水泥产品质量的监督和管理。重大建设工程的用户对水泥产品质量的均匀性和稳定性的要求也逐渐提高。随着中国加入 WTO，市场经济体制在我国的建立和发展，水泥企业也面临着国内外激烈的竞争。形势要求水泥产品与其测试标准以及质量管理方法必须逐渐同国际标准接轨。ISO 9000 族质量管理体系认证的贯彻实施，已成为各企业提高竞争力、扩大市场份额的必经之路。

在质量管理体系的建设中，数理统计方法是提高企业产品质量和经济效益的一个非常重要而又难度较大的基本工具。在 ISO 9000 族标准中，对数理统计方法的应用提出了明确而严格的要求，强调“过程策划应能确保各过程按规定的方法和顺序在受控状态下进行”。ISO 9000 族 2000 年版将统计技术从质量管理的“一个要素”提升为“质量管理体系的基础”。我国自 2002 年 4 月 1 日起开始实施的《水泥企业质量管理规程》第八条化验室的职责中明确规定，“用正确、科学的数理统计方法，及时进行质量统计并做好分析总结和改进工作。”这些规定的目的是通过正确、科学的数理统计方法对水泥生产过程自始至终进行必要的分析和监测，及时调整生产工艺，以保证水泥产品的质量，提高企业的经济效益。

对我国大多数水泥企业而言，数理统计技术方面的应用尚是一个薄弱环节。为了提高水泥企业化验人员和质量管理人员的数理统计知识水平，通过科学的数理统计方法，从各个环节上抓好质量管理，促进企业的技术改造，编者根据培训工作的要求，参阅了有关文献和资料，结合水泥企业生产控制的实际情况，编写了本手册。本手册立足于水泥企业数理统计工作的实际需要，以便于理解和应用为主要目的，对统计学中繁杂的数学公式不作过多的介绍和推导，而将重点放在对水泥生产过程各环节质量控制指标的数理统计和分析上，帮助质量管理人员从中找出规律，对生产过程实施及时控制，以期提高产品质量，增加经济效益。

本手册第一章介绍数理统计的基本知识，第二章介绍分析检验中的数理统计方法，第三章介绍数理统计在水泥生产过程中的应用，第四章介绍过程质量控制中的数理统计知识，第五章介绍新老 14 种统计工具。以上五章是水泥企业分析检验人员和质量管理人员应该掌握的基本数理统计知识。为适应水泥企业技术改造和研制新产品、新工艺的需要，本手册又为质量管理人员、工艺技术人员收入了有关的数理统计方法，包括第六章的假设检验，第七章的方差分析与试验设计和第八章的回归分析。

本手册适用于水泥企业分析检验人员、统计人员和质量管理人员阅读使用,也可供各级水泥质量监督检验人员参考。

本手册在编写过程中参阅了有关文献资料。在此向这些文献资料的作者表示衷心的感谢。

由于编者水平所限,手册中难免有不妥之处,敬希读者不吝指正。

编　者

2004年8月

目 录

第一章 统计学基本知识

第二章 试验误差与数据处理

第三章 统计方法在水泥生产质量控制中的应用

第四章　过程控制中的统计技术

第五章　质量管理统计工具

第六章 假设检验

第七章 方差分析与试验设计

第八章 回归分析

附　录

第一章 统计学基本知识

第一节 统计的基本概念

统计技术是以概率理论为基础的应用数学的一个分支。统计技术是研究随机现象中确定的统计规律的学科。产品质量特性是一种随机现象,但这种随机现象在一定的范围内服从确定的统计规律——概率分布,其中最常见的是正态分布。按照实用型定义,统计技术是指与应用有关的统计方法,收集、整理、分析和解释统计数据,并对其所反映的问题的性质、程度和原因做出一定结论的科学技术。统计技术包括统计推断和统计控制两大内容。统计推断是指通过对样本数据的统计计算和分析,提供表示事物特征的数据,比较两个事物之间的差异,分析影响事物变化的原因,找出产品形成全过程中质量变化的规律,对总体质量水平进行推断,预测尚未发生的事件;统计控制是指通过对样本数据的统计计算和分析,采取措施消除过程中的异常因素,以保证产品质量特性的分布基本保持在设定值附近,使生产过程达到稳定受控状态。

应用统计方法要掌握分布的理论,要符合大数定律,即只有对大量数据取得的统计平均值才具有稳定性和代表性,才能得出比较准确的统计结论。因此,只有掌握基本的统计理论知识,才能较好地应用统计方法,发挥统计技术在质量控制中的作用。

一、统计数据

数据是统计的对象。习惯上把由数字组成的数字数据称为数据。

1. 数字数据

数字数据指由数字(0、1、2、3、4、5、6、7、8、9)和小数点组成的数据。数字数据是对可定量描述的特性的表达。可以通过抽样、测量、记录获得数字数据。任何数字数据又都可以形成(服从)一定的分布(统计规律)。

2. 数据的分类

(1)计量值数据。计量值数据是指可以连续取值,在有限的区间内可以无限取值的数据。长度、面积、体积、质量、密度、电压、电流、强度等,大部分质量特性的数值都属于计量值数据。

(2)计数值数据。计数值数据是只能间断取值,在有限的区间内只能取有限数值的数据。如到会的人数,今天生产的产品件数,产品表面的缺陷数等。所以计数值数据,是以正整数(自然数)的方式表现。计数值数据又分为计件值数据和计点值数据。

二、统计技术、统计方法和统计工具

统计技术中常使用三个名词:统计技术、统计方法和统计工具。这三种提法有其共性,即均是研究随机现象中确定的数字规律,但也有其各自的特点。

1. 统计技术

统计技术是一个大的概念，是就整个学科而言，指的是一门技术的总概括。

2. 统计方法

统计方法是指统计技术中的具体方法。如控制图，直方图，散布图等各是统计技术中的一种方法。原则上应称控制图、直方图、散布图等为统计方法。

3. 统计工具

统计工具指简化的统计方法。统计工具的开发是日本质量管理专家对质量管理工作的重要贡献。统计技术的理论基础是概率论，但对这一理论，初级技术人员难以掌握，因此妨碍了统计技术的推广应用。为此，针对基层工人和初级技术人员的特点，20 世纪 60 年代日本质量管理专家开发了因果图、排列图、调查表、直方图、散布图、控制图和分层法，称为质量管理七种工具。随着质量管理的不断深化，20 世纪 70 年代日本质量管理专家又开发出系统图、关联图、矩阵图、矢线图、KJ 法、PDPC 法和矩阵数据解析法，称之为质量管理新七种工具。所谓工具，指不讲统计方法的原理和设计，也不讲对统计结果的分析，只讲操作步骤。

第二节　总体和样本

一、总体

研究或统计分析的对象的全体元素组成的集合称为总体或母体。总体具有完整性的内涵，是由某一相同性质的许多个别单位（元素或个体）组成的集合体。当总体内所含个体个数有限时，称为有限总体；当总体内所含个体个数无限时，称为无限总体。在统计工作中，可以根据产品的质量管理规程或实际工作需要，选定总体的范围，如每个月的出厂水泥，某一批进厂煤或原材料，都可视为一个总体。

总体分布的特征值是指总体中单值 x 的分布特征值即分布中心 μ 及单值 x 的分散程度即标准偏差 σ。标准偏差 σ 的计算公式，如：

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}} \tag{1-1}$$

式中：$\sum$ ——代表加和；

x_i——总体中任意一个个体数据；

n——总体中的个体数，n 应趋向于无穷大（$n \to \infty$），至少要≥20。

总体的性质取决于其中各个个体的性质，要了解总体的性质，理论上必须对全部个体的性质进行测定，但在实际中往往是不可能的。一是在多数情况下总体中的个体数目特别多，可以说接近于无穷多，例如出厂水泥，即使按袋计数，也不可能对所有的袋进行测定；二是由无限个体组成的总体，例如对一种新分析方法的评价分析，每次测定结果即为一个个体，可以一直测定下去永无终止；三是有些产品质量的检测是破坏性的，不允许对其全部总体都进行检测。基于总体的这种种情况，在实际工作中只能从总体中抽取一定数量的、有代表性的个体组成样本，通过对样本的测量求出其分布中心和标准偏差，借助于数理统计手段，对总体的分布中心 μ 和标准偏差 σ 进行推断，从而掌握总体的性质。

二、样本

来自总体的部分个体的集合,称为样本或子样。从总体获得样本的过程称为抽样。样本中的每个个体称为样品。样本中所含样品的个数称为样本容量或样本大小。若样本容量适当地大,并且抽样的代表性强,则通过样本检测得到的分布特征值,就能很好地代表总体的分布特征值。

例如在水泥生料配制过程中,为控制生料的质量,每小时从生料生产线上采取一个样品,进行硅、铁、铝、钙的测定。每天共采取 24 个样品,构成了该日配制的生料总体的一个样本。对该样本中的 24 个样品的化学成分进行测定,可计算出该日配制的生料三率值的平均值。还可推广到整个生料库,将该生料库容纳的全部生料作为一个总体,其中每小时采取的样品之和作为样本,根据样本中所有样品的分析结果,计算该生料库中全部生料的三率值。又如,欲求得上月出厂水泥 28 天抗压强度的标准偏差,须以上月生产的全部水泥为一个总体,例如,上月生产20 000 t的水泥,按照《规程》规定,将每 400 t 作为一个编号,共分为 50 个编号,从每个编号的水泥中取得一个质量约为 6 kg 的样品(这 6 kg 样品应从 20 个不同部位中均匀抽取,混合均匀后作为一个样品),共取得 50 个样品,构成上月生产的出厂水泥总体的样本,样本容量为 50。对每个样品进行 28 天抗压强度的测定,按照公式计算得出上月出厂水泥 28 天抗压强度的标准偏差,作为确定当月出厂水泥 28 天抗压强度控制值下限的依据。

三、样本分布的特征值

由于总体的分布特征值一般是很难得到的,往往通过样本的分布特征值来进行推断。因此,在实际应用中,为了对总体情况有一个概括的全面了解,需要用几个数字表达出总体的情况。这少数几个数字在数理统计中称为特征值。因此,在进行统计推断前确定样本分布的特征值,具有重要的实用价值。

常用的样本分布特征值分为两类:一是位置特征值;二是离散特征值。

位置特征值一般是指平均值,它是分析计量数据的基本指标。在测量中所获得的检测数据都是分散的,必须通过平均值将它们集中起来,反映其共同趋向的平均水平,也就是说平均值表达了数据的集中位置,所以,对一组测定值而言,平均值具有代表性和典型性。位置特征值一般包括算术平均值、几何平均值、加权平均值、中位数、众数等。

离散特征值用以表示一组测定数据波动程度或离散性质,是表示一组测定值中各测定值相对于某一确定的数而言的偏差程度。一般是把各测定值相对于平均值的差异作为出发点进行分析。常用的离散特征值有平均差、极差、方差、标准偏差、变异系数等。

(一)表示样本分布位置的特征值(样本分布中心)

1. 算术平均值 $\overline{x}$

算术平均值的计算十分简单,应用也十分广泛。

将一组测定值相加之和,除以该组的样本的容量(测定所得到的测定数据的个数),所得的商即为算术平均值。设有一组测定数据,以 x_1、x_2、……、x_n 表示。这组数据共由 n 个数据组成,其算术平均值为:

$$\overline{x} = (x_1 + x_2 + \cdots\cdots + x_n)/n \tag{1-2}$$

或

$$\overline{x} = \sum_{i=1}^{n} x_i / n \qquad (1-3)$$

式中：n——样本的容量；

$\sum\limits_{i=1}^{n}$——在数理统计中，常用希腊字母 $\sum$ 表示加和。$\sum$ 下方的 $i=1$，表示从第一个数据开始加和，一直加和到 $\sum$ 上方所表示的第 n 个数据。

例如，对水泥中三氧化硫含量(%)的测定，得到 10 个数据：2.8、2.9、2.6、2.7、2.8、2.8、2.9、2.8、2.8、2.6。其算术平均值为：

$$\overline{x} = (2.8 + 2.9 + 2.6 + 2.7 + 2.8 + 2.8 + 2.9 + 2.8 + 2.8 + 2.6) \div 10 = 2.78$$

或记作：

$$\overline{x} = \sum_{i=1}^{n} x_i / n = (2.8 + \cdots\cdots + 2.6)/10 = 2.78$$

可以证明，样本的算术平均值是对总体分布中心的最佳推断。各测定值离开平均值的偏差之和等于零。各测定值距平均值之偏差的平方和小于距其他任何点的偏差的平方和。因此，常用平均值作为判断测定值大小的基准点。

2. 加权平均值

加权平均值是考虑了每个测量值的相应权的算术平均值。将各测量值乘以与其相应的权，将各乘积相加后，除以权数之和，即为加权平均值。其计算公式如下：

$$\overline{x}_W = \frac{W_1 x_1 + W_2 x_2 + \cdots\cdots + W_n x_n}{W_1 + W_2 + \cdots\cdots + W_n} = \frac{\sum W_i x_i}{\sum W_i} \qquad (1-4)$$

式中：$\overline{x}_W$——加权平均值；

x_1、x_2、……x_n——各测量值；

W_1、W_2、……W_n——各测量值相应的权；

$\sum W_i$——各相应权的总和；

$\sum W_i x_i$——各测量值与相应权乘积之和。

水泥企业计算某一时期内熟料的综合抗压强度时，应采用加权平均值。

【例 1－1】 某水泥企业有三台回转窑。1 号窑年产 20 万 t 熟料，平均抗压强度为 58.5 MPa；2 号窑年产 15 万 t 熟料，平均抗压强度为 57.8 MPa；3 号窑年产 12 万 t 熟料，平均抗压强度为 59.2 MPa。求全厂全年生产的熟料的综合抗压强度。

解：全厂全年生产的熟料的综合抗压强度为：

$$\overline{x}_W = \frac{20 \times 58.5 + 15 \times 57.8 + 12 \times 59.2}{20 + 15 + 12}$$

$$= \frac{274\,7.4}{47} = 58.46(\mathrm{MPa})$$

3. 中位数

中位数 $\tilde{x}$ 也是表示频率分布集中位置的一种特征值。其意义是将一批测量数据按大小顺序排列，居于中间位置的测量值，称为这批测量值的中位数。当测量值的个数 n 为奇数时，

第$\frac{1}{2}(n+1)$项为中位数；当测量值的个数 n 为偶数时，位居中央的两项之平均数即为中位数。

【例 1-2】 对出磨水泥每 2 小时测定一次三氧化硫含量，某日共得 12 个测量值：2.86、2.91、2.65、2.70、2.82、2.73、2.88、2.92、2.75、2.84、2.77、2.85。求这组测量值的中位数。

解：将 12 个测量值从小到大（或从大到小）依次排列为：

2.65、2.70、2.73、2.75、2.77、2.82、2.84、2.85、2.86、2.88、2.91、2.92

测量值个数 12 个为偶数，中位数是居于中间位置的二个测量值的算术平均值，故中位数为：

$$\tilde{x} = \frac{2.82 + 2.84}{2} = 2.83$$

中位数不受极端测量值的影响，计算方法比较简便，但准确度不高，多在数理统计和生产过程控制图中使用。

4. 众数

众数是指在一组测量数据中出现次数最多的测量值。例如某水泥企业控制出磨水泥的细度（筛余）范围为(7.0±1.0)%。每小时测定一次，某日早班的测量数据如下：7.4、7.1、7.8、7.4、7.5、7.4、7.6、7.5。

在这组数据中 7.4 共出现三次，多于其他任何数，故 7.4 即为这组测量数据的众数。

众数不受检测数据中所出现的极大值或极小值的影响，因此在检测值数列两端的数值不太明确时，宜于用众数表示检测结果的位置特征。但其缺点是当检测值未呈现明显的集中趋势时，其数列不一定存在有众数；众数没有明显的数学特征，一般不能用数学方法进行处理。

5. 均方根平均值

均方根平均值是各测量值平方之和除以测量值个数所得商值的平方根。计算公式如下：

$$u = \sqrt{\frac{x_1^2 + x_2^2 + \cdots\cdots + x_n^2}{n}} = \sqrt{\frac{\sum x_i^2}{n}} \tag{1-5}$$

式中：x_1、x_2、……x_n——各测量值；

n——测量值的个数；

$\sum x_i^2$——各测量值平方之和。

均方根平均值能较为灵敏地反映测量值的波动。

【例 1-3】 某班对出磨水泥细度的测量值（筛余）为：

7.2、7.3、7.4、8.8、7.9、7.6、7.4、7.5

求该班出磨水泥的平均细度。

解：用均方根平均值计算平均细度为：

$$u = \sqrt{\frac{7.2^2 + 7.3^2 + 7.4^2 + 8.8^2 + 7.9^2 + 7.6^2 + 7.4^2 + 7.5^2}{8}} = \sqrt{\frac{468.5}{8}}$$

$$= \sqrt{58.56} = 7.7(\%)$$

如用算术平均值计算平均细度为 7.6%，均方根平均值大于算术平均值，因该班测量值中出现了一个波动较大的值，即 8.8(%)。

（二）表示测量值离散性质的特征值

1. 极差 R

极差是最简单最易了解的表示测量值离散性质的一个特征值。极差又称全距,或范围误差,即在一组测量数据中最大值与最小值之差:

$$R = x_{max} - x_{min} \tag{1-6}$$

【例 1-4】 测得三块试体的抗压强度为 58.7、57.8、59.2、59.8、58.4、58.8(MPa),示此组试体抗压强度的极差。

解:极差为:

$$R = x_{max} - x_{min} = 59.8 - 57.8 = 2.0(\text{MPa})$$

极差是位置测定值,极易受到数列两端异常值的影响。测量次数 n 越大,其中出现异常值的可能性越大,极差就可能越大,因而极差对样本容量的大小具有敏感性。另外,极差只能表示数列两端的差异,不能反映数列内部频数的分布状况,不能充分利用数列内的所有数据。

尽管如此,极差在不少场合还是用来表示数列的离散程度。在正常情况下,只希望得知产品品质的波动情况,经常使用极差;在对称型分布中,使用极差表示数列的离散程度更为便捷,这时两极端的平均值非常接近于整个数列的平均值。

2. 平均绝对偏差

一组测量数据中各测量值与该组数据平均值之偏差的绝对值的平均数,称为平均绝对偏差。其计算公式如下:

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n} = \frac{\sum |d_i|}{n} \tag{1-7}$$

式中:$\bar{d}$——平均绝对偏差;

d_i——某一测量值与平均值 $\bar{x}$ 之差,$d_i = x_i - \bar{x}$。

【例 1-5】 以氟硅酸钾容量法测定某水泥熟料样品中二氧化硅的含量(%),所得结果为:21.50、21.53、21.48、21.57、21.52。计算该组测量结果的平均绝对偏差。

解:该组测量值的平均值:

$$\bar{x} = \frac{1}{5} \times (21.50 + 21.53 + 21.48 + 21.57 + 21.52) = 21.52$$

平均绝对偏差:

$$\bar{d} = \frac{1}{5} \times (0.02 + 0.01 + 0.04 + 0.05 + 0) = 0.024$$

平均绝对偏差是衡量数列离散程度大小的方法之一,比较适合于处理小样本,且不需精密分析的情况。与极差相比,比较充分地利用了数列提供的信息。但因其计算比较繁琐,在大样本中很少应用。与标准偏差相比,平均绝对偏差反映测量数据离散性的灵敏度不如前者高。

3. 方差

方差是指各测量值与平均值的偏差平方和除以测量值个数而得的结果。采用平方可以消除正负号对差值的影响。

如以 σ^2 代表总体方差,其计算公式为:

$$\sigma^2 = \sum (x_i - \mu)^2 / N \tag{1-8}$$

式中:x_i——每个测量值(变量);

μ——总体平均值;

N——总体所有变量的个数。

在实际工作中，往往用样本的方差来估计总体的方差。以样本的测量值估计总体方差的计算公式如下：

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n - 1} \tag{1-9}$$

式中：x_i ——样本中每个测量值(变量)；

$\overline{x}$ ——样本平均值；

n ——样本容量。

式中的 $n-1$ 称为自由度 f，$f=n-1$。所谓自由度可以理解为进行独立测量的次数减去处理这些测量值时所外加的限制条件的数目。此处独立测量的次数为 n，外加的限制条件是算术平均值 $\overline{x}$。如果已知 $(n-1)$ 个离差，则第 n 个离差也就可以确定下来，因此，在 n 个离差中相互独立的只有 $n-1$ 个。这就是自由度 $f=n-1$。这里是从自由度的物理意义出发进行理解，实际上从数学角度是可以得到严格证明的(详见参考文献 14)。

利用方差这一特征值可以比较平均值大致相同而离散度不同的几组测量值的离散情况。

【例 1-6】 某厂有两台水泥磨，在同一班里各自测定了出磨水泥的细度(筛余)，数据如下：

1 号磨：7.4、7.5、7.6、8.0、7.9、7.6、7.6、7.5

2 号磨：6.0、6.4、6.8、7.8、8.0、8.2、8.9、9.0

计算各自的平均值和方差。

解：

1 号磨：

平均值：$\overline{x}_1 = \frac{1}{8} \times (7.4 + 7.5 + 7.6 + 8.0 + 7.9 + 7.6 + 7.6 + 7.5) = 7.64$

各次测量值与平均值之差依次为：

-0.24、-0.14、-0.04、1.36、0.26、-0.04、-0.04、-0.14

方差：$s_1^2 = \frac{1}{8-1}(0.24^2 + 0.14^2 + 0.04^2 \times 3 + 1.36^2 + 0.26^2 + 0.14^2)$

$= \frac{1}{7} \times 2.02 = 0.29$

2 号磨：

平均值：$\overline{x}_2 = \frac{1}{8}(6.0 + 6.4 + 6.8 + 7.8 + 8.0 + 8.2 + 8.9 + 9.0) = 7.64$

各次测量值与平均值之差依次为：

-1.64、-1.24、-0.84、0.16、0.36、0.56、1.26、1.36

方差：$s_2^2 = \frac{1}{8-1} \times (1.64^2 + 1.24^2 + 0.84^2 + 0.16^2 + 0.36^2 + 0.56^2 + 1.26^2 + 1.36^2)$

$= \frac{1}{7} \times 8.84 = 1.26$

两台磨出磨水泥的细度平均值 $\overline{x}_1 = \overline{x}_2$，但方差却相差很大，$s_1^2 = 0.29$，$s_2^2 = 1.26$，显然 1 号磨出磨水泥的细度质量指标要优于 2 号磨。