SQC-5

统计质量控制

STATISTICAL QUALITY CONTROL

方差分析

Variance Analysis

陈国铭　主编
杨丽春　　编

中国石化出版社

SQC-5

统计质量控制

STATISTICAL QUALITY CONTROL

方差分析

Variance Analysis

陈国铭 主编
杨丽春 编

中国石化出版社

（京）新登字 048 字

内 容 提 要

本分册系统地向读者介绍了方差分析的概念、原理和方法，特别着重介绍了三因素以下交叉分组试验及系统分组试验的方差分析法，并针对容易忽略的方差分析前提、易混淆的几种检验分析方法和极差在方差分析中的应用进行了讨论。本书列举了大量的例题，并给予尽可能详细的题解，以助于读者掌握方差分析原理与方法。

本书可供企业质量管理人员、工程技术人员，有关院校师生阅读。

SQC-5

统计质量控制

STATISTICAL QUALITY CONTROL

·方差分析

Variance Analysis

陈国铭　主编

杨丽春　　编

*

中国石化出版社出版发行

（北京朝阳区太阳宫路甲 1 号　邮政编码：100029）

煤炭工业出版社印刷厂排版

中国纺织出版社印刷厂印刷

新华书店北京发行所经销

*

850×1168 毫米 大 32 开本 $5^1/_8$ 印张 137 千字 印 1—6400

1995 年 3 月北京第 1 版 1995 年 3 月北京第 1 次印刷

ISBN 7-80043-558-X/O·023　定价：6.00 元

中国石油化工总公司质量管理协会组织编写

生产技术顾问：张德义
统计技术审核：王经涛

主　编：陈国铭
副主编：张祖荫　郭耀曾
编委（按姓氏笔划）：李世英　陈国铭　杨丽春
张祖荫　饶上建　郭耀曾　崔廷铨
其他编辑校核人员：万　涛　刘秋萍　吕巧云
邱以玲　田从金

序　言

为了适应国际贸易往来和经济合作的要求，国际标准化组织经过十多年的努力，于1986和1987年相继正式发布ISO8402《质量——术语》标准和ISO9000质量管理和质量保证系列标准，将世界多年质量管理的经验进行了标准化。ISO9000系列标准的基本点是要求企业在生产过程中建立有效的质量保证体系，并对质量体系中相互关联、相互作用的若干要素进行有效的控制。在过程质量控制中，科学、有效方法之一就是数理统计方法。因此在ISO9000系列标准的各个模式中以及质量管理和质量体系要素指南中都要求在市场分析、产品设计、工序控制、性能评定、数据分析等方面广泛使用统计技术，其范围包括实验设计、方差分析、显著性检验、累积和控制图、抽样检验等技术。因此，研究学习统计质量控制技术对于贯彻ISO9000质量保证系列标准，提高科学管理水平是非常必要的。

回顾世界质量管理的发展史，可以看出，数理统计技术在质量管理中发挥了重要作用。从19世纪末到现在，质量管理在历史上经过了检验质量管理、统计质量控制和全面质量管理三个阶段。单纯检验质量管理的严重缺点：一是只能从产品中发现和挑出废品，事前预防功能不强；二是由于检验人员的差错，即使全数检验也可能漏检或错检；三是至关重要的破坏性试验不可能全数进行。产品是生产出来的，单靠检验是不能防止产生废品的。1924年美国贝尔研究所的休哈特（W. A. Shewhart）运用数理统计的原理提出了控制生产过程中的“6σ”方法，即后来发展的质量控制图和预防缺陷的概念。与此同时，同属贝尔研究所的道奇（H. E. Dodge）和罗米格（H. G. Romig）联合提出了在破坏性试验情况下采用的“抽样检验表”。二次大战初期，美国大批民用品转入军

试读结束：需要全本请在线购买： www.ertongbook.com

工生产，由于事先无法控制废品而不能满足交货期要求，又由于军工生产多属破坏性试验，全数检验不可能也不允许。美国国防部为了解决这一难题，邀集休哈特、道奇、罗米格以及美国材料与试验协会、美国标准协会、美国机械工程师协会等有关人员研究，于1941～1942年先后公布一系列“美国战时质量管理标准”，要求各公司普遍实行统计质量控制方法，结果半年内取得显著成效。后来统计质量控制取得了很大发展。

我国自从1978年从日本引进全面质量管理十多年取得了显著成效。纵观我国的质量管理发展历史，是由检验质量管理直跃全面质量管理，对数理统计方法的运用远不是像当年美国那样深入广泛，不少决策、设计、科研、生产、销售、服务部门在提出问题、解决问题、检查结果时有些人还不习惯于进行科学的数理解析。

为了普及数理统计基本知识并在生产实际中发挥作用，我们组织石化行业中具有实践经验的质量管理专家编写了这套《统计质量控制》系列丛书。本书共分十册，第一册是数据收集和整理，第二册概率和数理统计基础，第三册估计和检验，第四册控制图，第五册方差分析，第六册实验设计，第七册相关和回归分析，第八册抽样检验，第九册统计方法应用演示50例，第十册数表。

数理统计方法就是通过对生产实践中大量数据的收集、整理、解析，研究生产实际中的内在规律的数学方法。和目前国内其它有关数理统计的书籍相比，本系列丛书的显著特点：一是它不同于一般的数学教科书，特别突出了实际应用，因此在编写中尽量减少不必要的公式推导，是一本实用性较强的书籍；第二个特点是书中列举了大量社会和生产（特别是石油化工生产）实例，文章从实例引出理论，又从理论回到实例，便于读者理解和应用，适合于工业企业特别是石油化工等流程型行业设计、研究、生产、销售、辅助等系统技术人员和管理干部学习参考；第三是语言既通俗易懂，又有一定深度和广度，既可用于中等水平人员学习应用，又可适用于高等水平技术人员研究参考。

为了更好应用本书，建议学习中注意几点：一是随着计算机的高度发展，许多数理统计方法可完全不需用手工计算，即可很快得出结果，已经掌握了统计方法的人可直接借助计算机，但对于初学之人，还是先用手算为好，防止知其然而不知其所以然，不利于在实践中灵活运用；二是对于现场技术人员，不要去深究公式推导，只要求会实际灵活运用；三是统计方法只提供解决问题的手段，必须和固有技术相结合才能解决问题，因此要使读者学会用数学的思维考虑专门技术问题；四是质量管理所用的方法不限于数理统计方法，还包括许多其它方法，如价值分析（VA）、生产工学（IE）、操作研究（OR）、价值工程（VE）、可靠性工程（RE）等，本书这次没有列入，读者可根据需要深入研究，灵活运用。

张英之

1995年1月

目　录

1　方差分析的基本概念

1.1　什么是方差分析

所谓方差分析（*variance analysis*），就是利用试验观测值总偏差的可分解性，将不同条件所引起的偏差与试验误差分解开来；按照一定的规则进行比较，以确定条件偏差的影响程度及其相对大小。当已确认某几种因素对实验结果有影响时，可应用方差分析检验确定哪种因素对实验结果的影响最为显著，以及估计影响程度。

例 1-1　为了考察三种催化剂对某一化工产品收率的影响，在其它条件不变的情况下，每种催化剂重复试验 4 次，所得收率的数据见表 1-1。

表 1-1　收率数据表（%）

催化剂 \ 次数	1	2	3	4	平均
甲	35.2	33.1	35.5	36.4	35.05
乙	31.6	36.6	34.2	34.8	34.30
丙	34.9	36.8	36.3	35.8	35.95

从表中不难看出：不同催化剂的平均收率有差异，表明催化剂这个因素的变化，对产品收率有一定的影响；同一催化剂的 4 次试验结果之间也有差异，这是否表明这个产品收率不仅仅受催化剂改变的影响，还受到催化剂以外的某些随机因素的影响？现在的问题是如何从这组数据中确定：（1）不同催化剂的收率，究竟是否有显著地差异？（2）这种差异是催化剂这一因素的变化所引起的，还是试验中各种偶然因素的干扰所致？

例 1-2　某化合物的收率受催化剂品种与反应温度的影响，现选取三种催化剂（记做 A_1、A_2、A_3），四种不同的温度（记做 B_1、B_2、B_3、B_4），在催化剂与温度的 12 种不同的组合条件下，随机

地进行试验。若记催化剂为 A_i 和温度为 B_j 时的合成收率为 x_{ij}，得表 1-2 的试验数据。

表 1-2 数 据 表

温度 B / 催化剂 A	B_1	B_2	B_3	B_4
A_1	x_{11}	x_{12}	x_{13}	x_{14}
A_2	x_{21}	x_{22}	x_{23}	x_{24}
A_3	x_{31}	x_{32}	x_{33}	x_{34}

这时的问题是：催化剂、温度以及实验中的各种偶然因素的综合中，谁对实验结果产生了显著地影响？其中谁的影响最大？

方差分析法是由英国统计学家费雪（Fisher）于 1923 年首创的，现已成为数理统计方法中的一个很重要、也是最基本的方法。方差分析法已被广泛地应用于统计质量控制领域。

1.2 术语解说

1）因素（*factor*）

方差分析中所说的因素是指，从对实验产生影响的众多原因中挑选出的，可通过选取不同的条件以供考察的原因。这里，因素也称因子。

显然、方差分析中的因素，不应包括测不到数值的因素和实验中不可控制或难以控制的因素。

通常，我们用大写字母 A、B、C 等来表示因素。如例 1-1 中，催化剂是唯一的因素，可记它为 A。例 1-2 中有催化剂和温度两个因素，我们已分别记它们为 A 和 B。

2）水平（*level*）

因素在实验中所取的不同条件(或状态)、称为该因素的水平。水平也叫位级。若某因素记为 A，则因素 A 的 a 个不同水平可分别记做：A_1、A_2，……，A_a。A_i 称为因素 A 的第 i 种水平或第 i 种位级。

3）水平数（*level number*）

某因素选定的不同条件（或状态）个数，称为该因素的水平数。

试读结束：需要全本请在线购买： www.ertongbook.com

如例 1-1 中因素催化剂的水平数为 3。因为催化剂取 3 种不同状态。例 1-2 中因素 A（催化剂）的水平数为 3，因素 B（温度）的水平数为 4。

4）重复（*repetition*）

在相同的条件下进行 2 次及 2 次以上的实验，称重复实验或有重复实验。

如例 1-1 是有重复实验。

重复是解决实验的随机误差的基本手段。应在条件允许的情况下，尽可能地采取重复试验，使样本容量足够地大，以保证实验与分析的可靠性。

5）重复数（*repetition number*）

相同条件下重复试验的次数简称重复数。重复数常用小写字母 r、t、k 等来表示。

重复数相等的实验，指在实验的各因素、各水平下，均做同样多次重复试验。如例 1-1 是重复数相等且为 4 的试验。

重复数不等的实验，指在不同的试验条件下重复实验次数不完全相同的实验。实际中，应尽量避免采用不等重复的实验，因为在实验总次数一定时，等重复试验要比不等重复试验的精度高。

6）主效应（*main effect*）

某因素单独对实验结果所产生的影响或作用，称该因素的主效应。

7）交互效应（*interaction*）

在多因素实验中，两个及两个以上的因素相互作用，联合对试验结果产生的影响或作用，称为交互效应。交互效应也称交互作用。

如温度 A 与压力 B 相互作用，联合对转化率产生的影响，称为因素 A 与 B 的交互作用。把交互作用看做一个因素考察时，可记 A 与 B 的交互作用为 $A\times B$。这里符号“×”是“交互”的记法，不要理解为乘号。

8）条件误差（*conditional error*）

由试验条件不同所造成的差异，称为条件误差，也称条件变差。条件误差属于系统误差。

9）试验误差（*test error*）

试验中各种偶然的(随机的)原因对试验结果产生的影响，称为试验误差。在不致引起混淆的场合，尤其是在方差分析过程中，通常所说的误差，均是指试验误差。误差仍用字母 E 或 e 表示之。

误差给试验结果带来的影响也称为误差效应（*error effect*），它是方差分析中必定要考察的。

1.3 方差分析的类型

总体上讲，方差分析包括两大基本类型。第一种就是所谓纯方差分析，其特点是当试验中的因素均可以控制，各因素所确定的所有水平都可以通过努力而达到；第二种是协方差分析，它针对另一类的试验而言。这种试验中，具有少量的不可控制或难以控制的因素，试验是为了能确定这种因素的影响，并努力消除其影响。本书中仅介绍纯方差分析法，因此，以下所说的方差分析均是指纯方差分析。

从因素个数而言，方差分析可分为单因素方差分析、双因素方差分析和多因素方差分析。

从水平搭配的角度来看，试验可分为全面试验、非全面试验和系统分组试验。全面试验指对所有不同的试验条件均进行至少一次地试验，如试验选定两个因素 A 与 B，因素 A 取 3 个水平：A_1，A_2，A_3，因素 B 取 2 个水平：B_1，B_2，那么因素 A 与因素 B 共有 6 种不同的水平搭配（或称组合）方式，这 6 种方式便是 6 种不同的试验条件，可依次表示为：A_1B_1、A_1B_2，A_2B_1，A_2B_2，A_3B_1，A_3B_2。非全面试验如正交试验，它是从全面试验的全部试验条件中，取部分条件进行试验的一种方法。系统分组试验中，因素分有级别，二级因素在一级因素下分组，三级因素又在二级因素下分组。对于以上不同的试验，都可以进行方差分析，并分别称之为全面试验（或交叉分组试验）的方差分析、正交试验的方差分析和系统分组试验的方差分析。因正交试验有专门的介绍，故在

这里不再叙述。

例 1-3 试回答表 1-3 给出的是一个什么样的试验，并用适当的字母符号表示其中的因素、水平、重复数、条件数等。

表 1-3 收率数据表

收率（%） 反应时间（小时） 催化剂品种	2	4	6	8	10
VPO	58.15 58.0	61.25 63.30	62.28 67.04	63.54 61.50	59.15 59.48
M-VPO	69.81 70.16	72.15 70.60	69.85 70.26	68.24 69.31	68.02 65.44

解：

这是一个两因素重复数为 2 的全面试验。其中一个因素（催化剂品种）取 2 个水平，另一个因素（反应时间）取 5 个水平，共有 10 种试验条件。因此可记：

A＝催化剂品种

$A_1 = VPO$ ，　　$A_2 = M\text{-}VPO$

B＝反应时间

$B_1 = 2$（小时），$B_2 = 4$（小时），$B_3 = 6$（小时），

$B_4 = 8$（小时），$B_5 = 10$（小时）。

重复数　$r = 2$

试验条件数　$n = 10$

数据个数　$N = 20$

2 方差分析原理

2.1 偏差的构造

为了能既简单又明了地说明问题，我们以单因素等重复试验的数据为例，来说明各种偏差的构造。单因素等重复试验数据可记录成表 2-1 的形式，经过简单地计算，可得各水平下的平均，如水平 A_i 下的平均指表中第 i 行数据的平均，记为 $y_{i\cdot}$。总平均指全部数据的平均，用 $\bar{y}$ 表示。

表 2-1 单因素等重复试验数据表

水平＼次数	1	2	……	j	……	r	水平平均	总平均
A_1	y_{11}	y_{12}	……	y_{1j}	……	y_{1r}	$y_{1\cdot}=\frac{1}{r}\sum y_{1j}$	
A_2	y_{21}	y_{22}	……	y_{2j}	……	y_{2r}	$y_{2\cdot}=\frac{1}{r}\sum y_{2j}$	
⋮	⋮	⋮		⋮		⋮	⋮	$\bar{y}=\frac{1}{ar}\sum\sum y_{ij}$
A_i	y_{i1}	y_{i2}	……	y_{ij}	……	y_{ir}	$y_{i\cdot}=\frac{1}{r}\sum y_{ij}$	
⋮	⋮	⋮		⋮		⋮	⋮	
A_a	y_{a1}	y_{a2}	……	y_{aj}	……	y_{ar}	$y_{a\cdot}=\frac{1}{r}\sum y_{aj}$	

这里，有 3 种偏差，即总偏差、条件偏差和试验误差，它们应当如何表示呢？

直观地想法可能会是：用全体数据关于总平均形成的偏差之和，即 $\sum\sum|y_{ij}-\bar{y}|$ 来表示总偏差；用各条件下的均值 $y_{i\cdot}$ 代表该条件下的 r 个数据，则条件平均与总平均所形成的偏差之和

即$\sum\sum|y_i.-\bar{y}|$表示条件偏差；用同条件下的偏差之和，即条件A_i下各数据与A_i的平均$y_i.$所形成的偏差$\sum_j|y_{ij}-y_i.|$再关于各条件求和，也即$\sum\sum|y_{ij}-y_i.|$来表示试验误差。但是，事实上，方差分析中是用偏差平方和(*sum of squares*)来构造各偏差的。这是因为偏差平方和具有方差的一些良好特征，在下面的章节里，我们会逐渐看到这种构造的优点与道理所在。

用偏差平方和来构造各偏差，即：

(1) 总偏差用$S_{总}$或S_T表示，且

$$S_{总}=\sum\sum(y_{ij}-\bar{y})^2 \tag{2-1}$$

并称$S_{总}$为总偏差平方和，简称总平方和。

(2) 因素A偏差即系统偏差，用S_A表示，且

$$S_A=\sum\sum(y_i.-\bar{y})^2 \tag{2-2}$$

并称S_A为因素(A)偏差平方和，由于它是A的不同水平引起的、也称为组间平方和。

(3) 误差用$S_{误}$或S_e来表示，且

$$S_{误}=\sum\sum(y_{ij}-y_i.)^2 \tag{2-3}$$

并称$S_{误}$为误差平方和，由于$S_{误}$是由各水平内的偏差构造而成的，也称为组内平方和。

2.2 相对偏差平方和

由各偏差平方和的构造可以看出，在同样的波动程度下，测定数据越多，计算出的偏差平方和就越大。因此，仅用偏差平方和来反映数据的各种偏差显然是不够的，还应当考虑测定值个数对偏差平方和的贡献，这便是我们这里要说的相对偏差平方和(*relative sum of square*)。

相对偏差平方和定义为：偏差平方和除以平方和的自由度(*degree of freedom*)。

自由度又是怎么确定的呢？

通常，随机变数的自由度是由数据个数 n 及数据所受的线性约束方程个数 m 所决定的。即，当 n 个随机变数 $x_1, x_2, \cdots\cdots, x_n$，受到且仅受到（2-4）式 m 个独立方程的约束时，则这 n 个数据的平方和 $\sum x_i^2$ 的自由度为 $n-m$。

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots\cdots+a_{1n}x_n=0 \\ a_{21}x_1+a_{22}x_2+\cdots\cdots+a_{2n}x_n=0 \\ \cdots\cdots \\ a_{m1}x_1+a_{m2}x_2+\cdots\cdots+a_{mn}x_n=0 \end{cases} \tag{2-4}$$

方程组的系数矩阵秩为 m，且 $0 \leqslant m \leqslant n$

根据上述定义，我们来确定 2.1 节中给出的三个偏差平方和的自由度。首先，我们看一下平方和 $\sum_{i=1}^{n}(x_i-\bar{x})^2$ 的自由度 $\left(\text{其中 } \bar{x}=\dfrac{\sum x_i}{n}\right)$。为了简便起见，以下省去求和号“$\sum$”的变动域，即“$\sum$”总表示“$\sum_{i=1}^{n}$”。

令 $\quad y_i=x_i-\bar{x} \qquad (i=1, 2, \cdots\cdots, n)$

则 $\quad \sum(x_i-\bar{x})^2=\sum y_i^2$

当 x_i 满足关系：

$$\bar{x}=\frac{1}{n}\sum x_i \quad \text{也即} \quad n\bar{x}=\sum x_i$$

时，数据 $y_1, y_2, \cdots\cdots, y_n$ 满足且只需满足下面一个关系式：

$$y_1+y_2+\cdots\cdots+y_n=0 \tag{2-5}$$

这是因为：

$$\begin{aligned} & y_1+y_2+\cdots\cdots+y_n \\ =& (x_1-\bar{x})+(x_2-\bar{x})+\cdots\cdots+(x_n-\bar{x}) \\ =& \sum x_i-n\bar{x}=0 \end{aligned}$$

可见，平方和 $\sum y_i^2$ 仅受到（2-5）式的约束，所以平方和 $\sum y_i^2=\sum(x_i-\bar{x})^2$ 的自由度为数据 y_i 的个数 n 减去约束方程个数 1。

试读结束：需要全本请在线购买： www.ertongbook.com