

王永初 任秀珍 编著

# 数据资源与决策

1234567890

1234567890

1234567890

1234567890

1234567890



# 数据资源与决策

王永初 编著  
任秀珍

贵州人民出版社

责任编辑 刘世强  
封面设计 孙晓云  
技术设计 夏顺利

## 数据资源与决策

王永初 任秀珍 编著

---

贵州人民出版社出版发行

(贵阳市延安中路9号)

贵州新华印刷厂印刷 贵州省新华书店经销

787×1092毫米 32开本 15.875印张 340千字

印数1—(2,000)

1988年9月第1版 1988年9月第1次印刷

---

ISBN 7-221-00123-5/O.04 定价: 4.20元

## · 内 容 简 介 ·

数据是一种软资源。数据的利用关系到工业、农业、交通运输事业，是人们制定计划、拟订改革措施、寻找提高产品质量途径等的重要依据。数据资源的分析和决策涉及到自然科学与社会科学的众多领域。本书应用概率论、统计学、模型论、控制论、系统论、运筹学等方法，分析提高产品精度与可靠性的途径，介绍由数据演变成实验式和利用数据预测生长过程及科技进步的方法，以及数据在计划调度、仓库管理、诊断决策、人力物力开发等方面的应用问题。本书内容深入浅出，便于阅读，题例丰富，易于理解；实用性好，便于应用，是应用数据方面的一本专著。可供从事计划管理、分析化验、质量监督、科学技术实验等方面工作的有关人员参考。

HS42/10

## 前 言

实验或试验数据是一种资源，这个观点已被越来越多的人所承认。森林、矿藏是一种资源，这在人们的理解中不存在任何异义，因为这种资源看得见摸得着。至于数据是一种资源，可能有人难以理解，原因是它摸不着。本书以大量的实例说明实验数据这种无形的资源在国计民生、工农业生产中的作用。

实验数据资源是一种软资源。譬如在最优的工况下，同样的原材料可以生产出更多的产品，这些多出的产品是由实验数据产生的。通过可靠性分析，改善产品关键的结构，提高产品的使用期限，这同样是数据资源所起的作用。实验数据提供制作模型的依据，使得人们可以预测未来，下定决心作出决策。在我们从事科学研究的过程中深深感到这个问题的重要。

我们之所以决心写好这部著作是受到我们许多友好的鼓励与启发：浙江大学的吕勇哉教授、成都电讯工程学院的周立峰副教授、重庆大学的曹长修教授、沧州大学的陈思录校长，他们的许多精辟论述、丰富的实践，曾使我们受益非浅；感谢我们的长辈王文华医生，他在医学与医药方面给了我们大量的启示；感谢周春晖教授、周其鉴教授与薛定芳研

究员，他们在控制论、模型论与运筹学方面给我们许多有益的帮助。

我们认为数据资源的利用对所有的自然科学，甚至哲学与经济学都是有用的。实验人员、分析人员、质量监督人员、管理人员，乃至自然科学工作者、社会科学工作者，都会接触到大量的数据。应用好数据涉及的知识面较广，如概率论、统计学、可靠性理论、误差理论、模型论、系统学、决策论、排队论、库存论与运筹学。实验数据资源的利用，则是吸收了上述各种新生学科的精华，逐渐自成体系。在这部书中，我们以深入浅出的叙述和反复的举例说明，使读者可以理解各种方法的原理、应用步骤，以便选用各种方法去解决各自的问题。

我们都曾在西南工作了廿多年，西南是我们的第二故乡，在我们离开西南到新的工作岗位的时候，能有机会将此书奉献给西南人民，奉献给全国从事实验数据资源利用的同志们，应该感谢贵州人民出版社，为我们及时出版了这部著作，使得我们有机会表达自己的心愿。

华侨大学精密机械系陈希达主任、中国仪器仪表学会陆廷杰主任、中国自动化学会王洪宾主任在百忙中审阅了本书的原稿和修改稿，提出了许多宝贵的意见，藉本书出版的机会表示我们的谢意。

由于我们学识有限，知识面不够，书中难免有遗缺，欢迎读者批评指正。

编著者

1986年3月于华侨大学

# 目 录

<b>第一章 概述</b> .....	(1)
一、数据是一种资源.....	(1)
二、数据资源是可以利用的.....	(2)
三、合理利用数据的方法.....	(7)
<b>第二章 由实验数据分析产品质量</b> .....	(14)
一、问题的提出.....	(14)
二、误差数据处理的有关概念.....	(20)
(一) 测量误差的分布规律.....	(20)
(二) 测量数据误差的估计.....	(30)
三、静态精度分析.....	(36)
(一) 综合静态误差.....	(36)
(二) 误差分配.....	(38)
四、动态精度的分析.....	(41)
(一) 一个实例.....	(41)
(二) 动态误差的估计.....	(55)
五、质量评定与效果分析.....	(73)
(一) 子样大小的考虑.....	(75)
(二) 次品率的检验.....	(77)
(三) 革新效果的检验.....	(83)

(四) 比较检验 .....	(88)
(五) 试验方案优劣的评定 .....	(91)
六、应用举例 .....	(98)
<b>第三章 由实验数据制作实验式</b> .....	(111)
一、问题的提出 .....	(111)
二、由实验数据拟合实验式 .....	(117)
(一) 线性方程式拟合 .....	(118)
(二) 非线性方程式拟合 .....	(144)
(三) 拟合度的检验 .....	(150)
三、最少数据的最大利用 .....	(155)
(一) 正交多项式 .....	(155)
(二) 数据的正交变换 .....	(159)
(三) 正交实验结果的检验 .....	(179)
(四) 组合实验 .....	(184)
四、应用举例 .....	(212)
<b>第四章 由故障数据判断产品的可靠度</b> .....	(212)
一、问题的提出 .....	(212)
二、可靠性指标 .....	(216)
(一) 基本定义 .....	(216)
(二) 可靠性指标的概率联系 .....	(230)
(三) 无故障率具有威布尔分布的情况 .....	(236)
三、组合设备可靠度的基本运算关系 .....	(245)
(一) 串联的可靠度 .....	(246)
(二) 并联的可靠度 .....	(253)
(三) 冗余系统的可靠度 .....	(265)
四、整机系统的可靠性分析 .....	(276)



(一) 失效逻辑框图 .....	(276)
(二) 整机分析必要考虑的问题 .....	(290)
五、应用举例 .....	(296)
<b>第五章 由记录数据估计生长过程 .....</b>	<b>(305)</b>
一、问题的提出 .....	(305)
二、模型法及其应用 .....	(309)
(一) 系统模型的基本表示方法 .....	(310)
(二) 生长模型 .....	(316)
(三) 退化模型 .....	(336)
(四) 假设验证法 .....	(341)
三、经验模型法 .....	(352)
(一) 三种生长规律 .....	(352)
(二) 生长模型 .....	(353)
(三) 模型参数的确定 .....	(355)
(四) 基型方程的拟合 .....	(358)
(五) 应用举例 .....	(360)
四、选点拟合模型法 .....	(366)
(一) 基本方法 .....	(366)
(二) 应用举例 .....	(369)
<b>第六章 由综合数据形成决策 .....</b>	<b>(377)</b>
一、问题的提出 .....	(377)
二、由确定数据决定最优产销或指派方案 .....	(382)
(一) 动态规划方法与应用 .....	(383)
(二) 线性规划方法与应用 .....	(396)
三、服务部门的数据利用 .....	(419)
(一) 排队论及其应用 .....	(419)

(二) 库存论及其应用 .....	(452)
(三) 对策论及其应用 .....	(468)
<b>附录</b> .....	<b>(475)</b>
<b>参考文献</b> .....	<b>(495)</b>

# 第一章 概 述

## 一、数据是一种资源

有人认为信息、物质与能量是现代科学的三大支柱。信息是物质或能量存在与联系的表意。数据是信息的量化。譬如一个生产过程，产品的数量与质量是该生产过程属性与内外联系的必然的结果，是生产过程的信息。树木的年轮是树木生长的信息。中医诊断通过“望、闻、问、切”所掌握的是病家的信息。信息按时间排序所取出的具体值，就是数据。

信息是流动的，正象长江流水一样，此时此刻的信息同彼时彼刻的信息不一样，不一样的程度通过排序的数据反映出来。这里所谓时间排序就是按时间的先后为序的排列，可见信息赋存有时间的概念。信息存在是多样性的，譬如转炉脱炭过程，因为脱炭率高，钢水中的炭被燃烧多，发出的热量增加。脱炭率与炉温均可以作为炼钢过程的信息，这种多样性反映了信息可以被转换的特性。因此，我们在对某事物进行研究时，可以利用其直接的信息，也可以利用其转换后的信息。转换后的信息仍然是事物本质与内外联系属性的表现。对于一个具体的事物来说，信息流是无穷的、源远流长的，而数据总是在

有限时间内获得信息的一些散点值。我们要从这些片断、零散的数据，恢复信息的全貌，难免会引入许多误差。通过局部的信息对事物的本质作出判断，失误的可能性亦完全存在。正因为这样，研究从有限的信息尽可能地推断出全部信息，从局部的试验得到事物发展的全过程，有着许许多多的方法问题和技术问题必须解决。开发数据资源同开发一切矿藏一样，都有一些成熟和正在探讨的方法。用好数据资源这个课题，引起了许多自然科学工作者与社会科学工作者的重视。

一个有成就的科学家、工程师和实验科技人员，无不对自己工作中的数据感兴趣，他们往往从日积月累的数据中逐渐摸索到事物变化的规律，因而酝酿了定理、定律，导致了创造与发明。

虽然数据不是物质也不是能量，它不能象钢铁那样直接用来制造机床、汽车和拖拉机以及建设高楼大厦，也不能象石油那样可以用来提炼开动飞机的汽油、制造人造纤维的原料。但是它通过人们的观测与记录确实实是存在的，我们利用它，可以从石油中多提炼出许多的汽油，从一定量的钢铁中多生产许多的机床，这些多出产的汽油或机床是从实验数据的利用中产生的。正象从矿藏中“发掘”宝藏一样，我们藉助于实验数据，向生产、向大自然多要回了许多物质或能量，所以我们说数据是一种资源。

## 二、数据资源是可以利用的

数据虽然是局部的信息序列值，但它仍然是事物本质与

内外联系的反映。事物的本质与内外联系有其固有的规律，它所反映出的信息也必然有规律。马克思曾经说过，一切学科发展成一门真正的科学，都可以用数学的方法来解释。既然数据是信息的部分反映，也是某种规律的反映，其变化规律应该是可以用数学的方法来描述的。描述数据随时间变化的数学式，称为数学模型。

由数据演变成数学式的过程是一系列的数学推理与运算过程，我们称它为拟合，其意思是利用一个模型结构已知而参数待定的数学式来表征数据随时间的演变过程。参数的选择遵循这样一个规律：按照拟合模型计算出的数据同实测的数据，其总体误差的平方和最小，即模型参数的选择在误差平方和意义下达到最佳的拟合。若记拟合模型为  $y(t_j)$  为实

$$\hat{y}(t_j) = f\left(\sum_{i=1}^n \alpha_i, t_j\right) \quad (1-1)$$

测数据，则  $\alpha_i$  的选择应满足

$$\begin{aligned} J(t_j) &= \sum_{i=1}^m \varepsilon_i^2 = \sum_{i=1}^m [y(t_j) - \hat{y}(t_j)]^2 \\ &= \sum_{i=1}^m \left[ y(t_j) - f\left(\sum_{i=1}^n \alpha_i, t_j\right) \right]^2 \rightarrow \text{最小} \quad (1-2) \end{aligned}$$

显然， $\alpha_i$  由如下的  $n$  个方程式确定

$$\frac{\partial J(t_j)}{\partial \alpha_i} = 0, \quad i = 1, 2, \dots, n \quad (1-3)$$

将求出的模型系数的估值 $\hat{\alpha}_i$ ,  $i = 1, 2, \dots, n$ 代入拟合模型, 可以得到描述数据变化的计算函数

$$\hat{y}(t_j) = f\left(\sum_{i=1}^n \hat{\alpha}_i, t_j\right) \quad (1-4)$$

图 1-1 表示数据的散点图与拟合曲线, 其中拟合曲线是利用式 (1-4) 来描述的。虽然该曲线是基本上符合时间由 0 至  $t_j$  的数据变化规律, 但我们可以将该式的  $t_j$  当作变量处理, 延伸到超出  $t_j$  的时间范围, 这样, 式 (1-4) 可以用于预测。

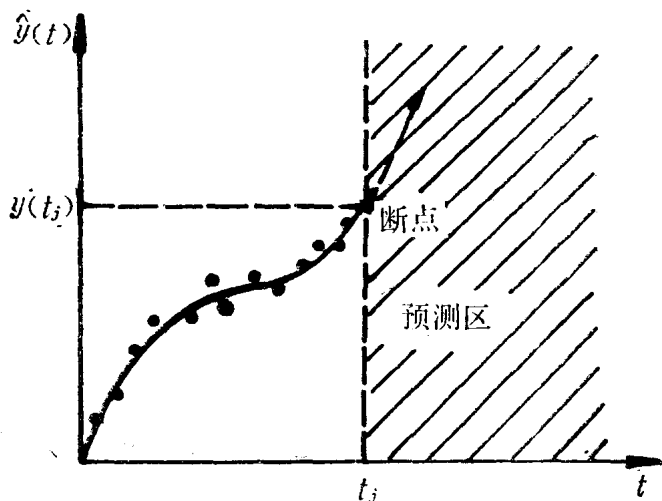


图1-1 数据散点图与拟合曲线

这种时间外延的函数用于估计生长过程就是生长曲线模型。生长模型在农牧业上有许多应用，譬如决定农作物或畜类什么时候收割（或屠宰），什么时候出售收益最大。生长模型在医学上应用也很多，根据实验得出的模型可以估计效应，估计服药的时间。

在我们进行实验或试验时，有时会出现偶然的误差，有些数据明显偏离大量数据的变化趋势，在制作模型时筛选需要非常慎重。对于短期使用的模型，可以将这些偶然的异常数据点剔除，但对于长期使用的模型，这些偶然的异常数据有时却是很重要的。譬如我们在估计某河流的水位变化时，异常高水位的数据往往很重要，据此，我们可以估计未来很长时间内的最高洪水水位，这对于建筑桥梁、修建堤坝，都有参考价值。从局部时间看，有些数据属于偶然的，但从长远看，这些偶然的数据常常是必然的结果。譬如海潮按一定的规律涨或落，其变化受到各种天文周期的影响，有时地壳的变化也会引起海平面的变化，捕捉这些异常的数据，可以用来估计地壳的变化，对于地震预报均有较大的作用。所以，当我们留心数据的变化时，不要轻易丢去一些认为无关紧要的数据。

实验测取数据是在较短时间内了解事物特性的重要手段。由实验测出的数据包含了两个方面的影响因素，一个是代表事物特性的内因，一个是代表事物与外界联系的因素——外因。描述事物内特性的函数是单位脉冲响应函数  $h(t)$ ，描述事物外因的函数是输入函数  $x(t)$ ，描述事物响应外因的函数是输出函数  $y(t)$ 。采用这样分解观察的方法，使我们可以及时根据当时当地的情况把握住事态的变化。但是  $y(t)$  并

不等于 $h(t)$ 乘以 $x(t)$ ，而是 $h(t)$ 与 $x(t)$ 的折积，即

$$\begin{aligned}y(t) &= h(t) * x(t) \\ &= \int_0^t h(t-\tau)x(\tau)d\tau\end{aligned}\quad (1-5)$$

这里 $*$ 代表折积符号。为了分析方便，我们较多地要借助于拉普拉斯变换。拉普拉斯变换是一种容易掌握与利用的函数变换，本书有不少地方介绍了这种变换的应用。根据我们的研究，得到一个十分有趣的结果，这就是 $h(t)$ 相当于外界条件恒定的自然生长函数，即生长曲线模型。正因为这样，有时我们可以不必进行很多试验，由生长的函数与预先规定的外界信号的变化函数，就可以决定全部的生长过程。

实验测取的数据是在条件控制的情况下取得的，因而数据比较规整，而自然积累的数据往往比较零乱。在自然变化条件下取得的大批数据，从本质上看是随机的，正因为数据变化的随机性，统计的方法尤其有用。采用统计的方法，我们可以从数据估计出观测变量的真值、分布规律、离差大小。这些统计的数据与统计的模型是判断产品精度，准确度和可靠性的基础，也使我们许多对许多随机变量有了量方面的认识。这些认识是十分宝贵的，它可以指导我们去改进产品的设计，去及时调整生产的工况。

实验数据也是帮助人们形成决心制定决策的依据。以市场供销问题为例，假如我们对某产品在市场的销售有了一定量的认识，又了解了其变化的趋势，我们就不会因盲目进货而造成商品的滞销或缺货。服务部门对服务对象往来的时间、数量及变化规律有了量的了解，就可以合理配备服务人



员与服务设施，获得最佳的经济收益。

### 三、合理利用数据的方法

实验数据利用得当可以使物质与能量的消耗减少，预见未来可能产生的事故，及时采取措施减少经济损失，保证设备与人身的安全，它起着与物质和能量一样的作用。要用好数据资源，有些科学的方法是必须的，譬如系统学的观点，可帮助人们建立全局的观点，有些事情从局部的角度看是不可取的，但从整体看是可取的，局部的需要就应该服从整体的需要。实验数据帮助人们建立总体与局部的模型、局部与局部的联系模型，有了这些模型，运筹学、决策论、排队论、库存论及预测学才能应用。因此说实验数据资源的利用是一门大学问，是实用科学与现代科学结合的结果。

不同的人对不同的事物有不同的了解，了解了事物常常会促使把事情办得好一些，不了解事物则常常导致把事情搞糟，但要求每一个人对事事都了如指掌，运筹于帷幄之中是不切合实际的。如果我们有了一些系统学、控制论、运筹学及概率论的基本知识，往往会触类旁通，将人的智能发挥到一个最好的程度，办好事情的可能性会大一些。正因为这样，本书介绍大量新的科学方法。由于照顾到读者面，我们的介绍也仅仅是初步，并且是通过例题引出概念，引出方法，这样处理的目的是便于应用。本书章节的内容虽然不同，但有些基本方法是共同的，学习时要掌握住这些基本的方法。