

统计方法应用标准化丛书

数据的 统计处理和解释

(第二版)

于振凡 张尧庭 于欣丽 陈玉忠 丁文兴 编著



 中国标准出版社

内 容 提 要

本书从实用的角度出发, 结合各行业数据统计处理与解释的实际需求及其相关的国家标准, 采用大量的质量管理方面的实例, 详细介绍了数据的统计处理和解释的理论、方法及原理, 力求通俗易懂, 深入浅出。读者通过本书可全面了解参数的统计检验、参数的估计以及异常值的判断与处理方面的基本知识和常用方法。

本书可作为各行各业数据统计处理和解释的培训教材使用, 适用于质量管理人、各类统计人员和工程技术人员阅读参考。

《统计方法应用标准化丛书》包括:

- 《数据的统计处理和解释》
- 《生产过程质量控制》
- 《产品质量抽样检验》
- 《可信性工程(可靠性、维修性、维修保障性)》

责任编辑: 王晓萍

封面设计: 徐东彦

版式设计: 李 玲

责任校对: 刘宝灵

责任印制: 邓成友

ISBN 7-5066-4220-4



9 787506 642200 >

ISBN 7-5066-4220-4/TB · 1650

定价: 15.00 元

统计方法应用标准化丛书

数据的统计处理和解释

(第二版)

于振凡 张尧庭 于欣丽 编著
陈玉忠 丁文兴

中国标准出版社

图书在版编目 (CIP) 数据

数据的统计处理和解释/于振凡等编著. —2 版.

北京: 中国标准出版社, 2006

(统计方法应用标准化丛书)

ISBN 7-5066-4220-4

I. 数… II. 于… III. 数理统计-数据处理-应用-国家标准-中国 IV. G307.0

中国版本图书馆 CIP 数据核字 (2006) 第 095873 号

中国标准出版社出版发行

北京复兴门外三里河北街 16 号

邮政编码: 100045

网址 www.bzcs.com

电话: 68523946 68517548

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 880×1230 1/32 印张 6 $\frac{1}{8}$ 字数 162 千字

2006 年 11 月第二版 2006 年 11 月第一次印刷

*

定价 15.00 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)68533533

丛书编委会

主 任 郑卫华

副主任 于欣丽 陈玉忠

委 员(按姓氏笔画)

丁文兴	于振凡	于 丹
马毅林	冯士雍	孙 静
何国伟	张尧庭	

丛书序

当今世界,我们面临着数字化、信息化的技术革命。高科技迅速发展及其成果的大规模商业化,使经济结构、产业结构和产品结构产生巨大变革,研究模式、管理模式也随之改变,这就需要现代统计技术应对这种变化。由于地区化、集团化经济的发展、贸易竞争日益激烈,产品质量的竞争已成为贸易竞争的最重要的因素。在这种形势下,各企业、公司、集团都深刻地感到不提高产品质量就没有出路,不能生存,产生了强烈的提高产品质量的紧迫感。提高产品质量,一要依靠技术进步,二要加强科学管理,产品质量是三分技术七分管理。质量工作、标准化工作都需要统计技术,统计技术是科学质量管理的重要手段,因为不论来自哪个学科的数据都需要用统计方法去处理并加以解释。目前世界上一些发达国家在科学技术和质量管理方面对统计技术的应用都极为重视,并且取得了显著的成绩。在这方面,我国相对比较薄弱。为了改善这种状况,更好地宣传、推广统计

技术,全国统计方法应用标准化技术委员会编写了这套《统计方法应用标准化丛书》。这套丛书第一版曾于1997年出版,取得了很好的效果,此次根据近几年国家标准的制修订情况、管理模式的发展及数理统计技术在管理中新的应用,又重新编写了第二版,包括下列四个分册:

《数据的统计处理和解释》

《生产过程质量控制》

《产品质量抽样检验》

《可信性工程(可靠性、维修性、维修保障性)》

这套丛书深入浅出地阐明了在科学技术研究与质量管理工作中,如何使用统计方法,并介绍了通过使用统计方法管理,提高产品质量、降低产品成本的有效途径。它的出版,对于推动统计方法的应用,建立科学的管理体系有着十分重要的指导意义。

这套丛书避免了高深的数学推导,以实用性为主,内容十分丰富,理论上既严谨又通俗易懂,具有可读性、可操作性,是广大科技人员、管理人员掌握数理统计技术的一套好书。衷心希望有更多的科技工作者、质量工作者通过阅读此书,成为通晓数理统计技术的高级专门人才,为我国现代化建设做出新贡献。

全国统计方法应用标准化技术委员会

2006年7月

前 言

随着科学技术和经济建设的飞速发展,数理统计在各行各业中得到了越来越广泛的应用,例如:质量管理、经济、金融、保险、环保等各方面都离不开数理统计,统计方法与技术是现代社会、经济研究及质量管理运用的基本方法。以概率论和数理统计为基础的各种统计技术的产生是理论与实践相结合的产物,这些方法已使深奥的纯理论数学走出了神秘的大门,变成了比较简单实用的工具。运用统计技术观察、分析问题比用直观的方法观察问题存在着本质的差别,它可以使人们透过现象看本质,透过现象发现原因,帮助人们识别直观所不易发现的问题和趋势。

自1969年设立诺贝尔经济学奖以来,已有40多位学者获奖。这些获奖者均精通现代统计方法,对统计方法的运用极为娴熟,在社会经济研究中取得了举世瞩目的成就。学习和运用统计方法已成为时代对我们的要求,在这个新形势下,近几年国家质量监督检验检疫总局又陆续颁布了一批统计方法应用的国家标准,以促进统计方法在各行各业中的应用。无论在工农业生产的质量管理还是实验室的质量管理中,都离不开统计方法和统计

技术。ISO 9000 族《质量管理体系标准》、ISO/IEC 17025《检测和校准试验室能力的通用要求》以及 ISO/TS 16949《质量管理体系 汽车生产件及相关维修零件应用ISO 9001 的特别要求》的颁布进一步推动了统计方法和统计技术在质量管理中的应用,为此,国际标准化组织还专门颁布了 ISO/TR 10017《统计技术应用指南》。严格地说,不使用统计方法和统计技术的质量管理不是科学的质量管理。

本书内容为数据的统计处理和解释,它是统计方法的重要组成部分,不仅适用于质量管理中的数据处理,也适用于各行各业的数据处理。

本书介绍了有关数据处理和解释的 19 项国家标准,并给出了一套从简单到复杂的数据处理和解释的方法。本书给出的某些内容,如正态性检验、异常值的判断和处理、二项分布与泊松分布的参数估计、假设检验、检验的功效等在一般教科书上不作介绍,因此本书也是大专院校师生有益的教学参考书。本书从应用的角度出发,从实际问题背景去理解统计学的概念和原理,不着重数学推理,文字力求深入浅出,通俗易懂。

编 者

2006 年 7 月

目 录

第 1 章 基本概念	1
1.1 引言	1
1.2 指标的分类	3
1.3 指标的分布	5
1.4 数据与取样	7
1.5 集中趋势	9
1.5.1 算术平均数	10
1.5.2 中位数	12
1.5.3 众数	13
1.6 离散性指标	17
1.7 分位数简介	20
1.8 其他	22
第 2 章 概率与分布	26
2.1 比例与概率	26
2.2 样本与统计量的分布	31
2.2.1 分布	31
2.2.2 随机变量与分布	34
2.2.3 期望、方差、矩	36

2.2.4	联合分布、条件分布、独立性	38
2.2.5	样本的联合分布	40
2.3	常见的离散型分布	41
2.3.1	两点分布	42
2.3.2	二项分布 $b(n, p)$	43
2.3.3	超几何分布	46
2.3.4	泊松分布	49
2.3.5	负二项分布	51
2.4	常见的连续型分布	52
2.4.1	正态分布	53
2.4.2	指数分布与威布尔分布	62
2.4.3	极值分布	68
2.4.4	位置—尺度参数族	69
2.4.5	Γ 分布	70
2.4.6	对数正态分布	71
2.5	统计量导出的分布	73
2.5.1	正态总体导出的统计量及其分布	74
2.5.2	样本的次序统计量	78
2.6	分位数	78
2.7	参数估计与假设检验简介	80
第3章	参数估计	83
3.1	点估计简介	83
3.2	点估计的优良性准则	84
3.3	点估计的方法	89
3.3.1	矩估计法	90
3.3.2	最大似然估计法	91
3.3.3	最小二乘法	94
3.3.4	贝叶斯估计	96

3.4	区间估计	98
3.5	二项分布参数的点估计与区间估计	103
3.5.1	经典估计	104
3.5.2	序贯估计	104
3.5.3	贝叶斯估计	104
3.5.4	极小极大估计	106
3.5.5	区间估计	107
3.6	泊松分布参数的点估计与区间估计	114
3.7	正态分布参数的点估计与区间估计	116
3.8	Γ 分布参数的点估计与区间估计	119
3.8.1	二参数 Γ 分布的参数估计	120
3.8.2	三参数 Γ 分布	122
第4章	假设检验	124
4.1	正态总体均值的检验	125
4.1.1	均值与给定值的比较	126
4.1.2	两个均值的比较	129
4.1.3	成对观测的两均值比较	131
4.2	方差分析	132
4.2.1	单因素的方差分析	132
4.2.2	两因素的方差分析	135
4.3	正态总体方差的检验	139
4.4	功效函数	141
4.4.1	U 检验的优良性质	142
4.4.2	假设检验的一些理论	145
4.4.3	OC 曲线的实际意义和应用	147
4.5	正态性检验	149
4.5.1	S-W 检验法(Shapiro-wilk)	151
4.5.2	D 检验法(D'Agostino)	151

4.5.3	偏度、峰度检验法	153
4.5.4	正态概率纸	155
4.6	二项分布参数的检验	156
4.6.1	单侧情形	156
4.6.2	双侧情形	158
4.7	泊松分布参数的假设检验	159
4.7.1	单侧情形	159
4.7.2	双侧情形	161
第 5 章	异常值的判断和处理	162
5.1	异常值处理的意义	162
5.2	正态总体异常值的判断和处理 (GB/T 4883)	165
5.2.1	奈尔(Nair)检验法	165
5.2.2	格拉布斯(Grubbs)检验法	166
5.2.3	狄克逊(Dixon)检验法	167
5.2.4	偏度-峰度检验法	168
5.2.5	一些说明	169
5.2.6	数据处理实例	170
5.3	指数分布异常值的判断和处理 (GB/T 8056)	178
5.3.1	全样观测时的方法	178
5.3.2	定数截尾时的检验方法	180
5.4	I 型极值分布异常值的判断与处理 (GB/T 6380)	180
5.4.1	狄克逊法	180
5.4.2	欧文(Irwin)法	181
	参考文献	183

第 1 章

基 本 概 念

1.1 引 言

从事科学研究、工农业制造以及管理工作的同志经常要和大量的数据打交道,而对这些数据的整理、分析和解释都离不开统计方法。统计是在面对不确定性的情况下做出决策的一种方法,它是研究数字资料的整理、分析和正确解释的一门学科。人们各自从不同的来源取得各种数字资料,例如在日常生产活动中,各部门都有大量的原始资料,但这些资料通常都是杂乱无章的,必须经过整理和简缩才能加以利用,加以研究。在各研究部门,各公司都拥有大量的试验数据,如果不经过整理和处理,他们无异是一堆废纸。但是,一旦使用完善的统计方法就可使数据整理、排列得有条有理,用图形或少量的几个重要参数,就可把一大堆数据的特征表达出来,同时既可避免不正确的解释,又可将获得满意数据的成本降到最低限度,提高了经济效益。

标准化是现代工厂公司管理上的一项重要工作,它对于保证产品质量、改善经营管理起着不可替代的作用。无论是制定标准还是推行标准,都要与大量的资料、数据打交道,怎样从大量的资料、数据中提取所要的信息,从而能作出比较正确的判断,这都离不开统计。

统计的功能不只是记录过去的状况,作一些简单的整理、归纳,让人们对于研究对象的总的轮廓有一个了解,更重要的是分析这些资料、数据为什么会不一样?应从哪些方面去探索形成这些差异的原因,揭示研究对象随时间的推移相应的变化规律,在这个基础上对未

来的状况作出估计和判断,帮助人们作出正确的判断和决策。所以统计不是只能描述、记录历史,而是分析历史、对现在和将来根据数据提供的信息来作判断和预测。它是一种认识客观事物规律的手段。

就以打靶的记录为例。如果我们观察了一个射手 100 次打靶的结果,其中 72 次中靶,28 次未中;中靶的 72 次弹着点的坐标记录下来是:

$$(x_1, y_1), (x_2, y_2), \dots, (x_{72}, y_{72})$$

很明显,一般说来,72 个中靶的位置不会是一样的。首先就要问;这些 $(x_i, y_i), i=1, 2, \dots, 72$ 为什么有差异? 如何从数据中分析出造成这些差异的原因? 粗略地说,有两个原因造成了这 100 次打靶结果的差异:

(i) 这个射击手的打靶技术。如果技术高,弹着点的密集程度高;如果技术差,弹着点就分散。

(ii) 这个射击手打靶时是否瞄准中心? 如果未瞄准,弹着点就会偏离中心,以致有的脱靶。

很容易看出,原因(i)需要长期训练后才能改进,而原因(ii)只要调整瞄准的中心位置就可改进,前者称为随机因素,也就是种种难以确定的因素造成的对射击手的影响,后者称为系统因素,这是比较明确的一个对结果有确定影响的因素。如果对这两者分析清楚了,就可以判断脱靶的原因大致是什么。怎样能从这 100 个打靶结果中分析出上述判断呢? 这就需要统计分析,统计分析的重要作用也就在这些方面。

从生产过程来看,每一道工序都在对产品进行加工,其中一道工序加工质量不好,就会影响后续的工序,甚至造成大量次品或废品。所以常常在生产时隔一定时间抽查若干件产品,检验这道工序加工质量是否合格,是否有变坏的趋势;应否调整,如何调整,这也要根据生产中历史资料以及现在的抽检结果作出判断,而统计分析也能提供较好的处理方法。

根据生产、管理中的实际需要,介绍一些统计的基本概念和常用的方法,有助于读者理解一些国家标准(或国际标准)制定时所依据的原理,掌握正确使用这些标准的方法,这就是本书的目的。我们不采用对统计方法方面的国家标准进行逐个介绍或逐个说明的方法,因为每个标准本身都已有这些内容。我们的目标是为使用这些标准的人提供一份便于自学、通俗易懂的教材,通过自学,理解和掌握有关的标准。这一章着重介绍一系列基本的概念,这些概念与每个统计标准都有联系,是理解各个统计标准必备的基础。

1.2 指标的分类

统计方法的研究分为三个部分,是有机统一体。它们是:

- (i) 对总体的研究,以确定总体的聚集性质;
- (ii) 对变异的研究,也就是除了要研究总体的集中趋势外,还要研究总体中各个个体之间的差异或离散程度;
- (iii) 将一大批数据简缩为容易理解的形式。

为了研究以上三个部分,我们首先谈谈指标的分类。我们把研究对象中观察、记录的内容称为指标,有些内容是描述属性的,有些内容是可以通过工具来测量的,大致上可以分为下面的四类:

(i) 计量的。如物体的长度、重量、室内的温度、湿度等等。这一类指标的特点是:原则上它的取值可以是实数某一区间内的任一个值,通常称这类指标是连续型的。这种指标的统计分析与具有密度的连续随机变量的分布相联,在通常的统计教科书中占有主要的地位。

(ii) 计数的。如一批产品中的不合格品个数、一个居民区内拥有的家用轿车辆数、固定资产在亿万元以上公司的个数等等。这一类指标的特点是:它的取值范围是整数或自然数,通常称这类指标是离散型的计数指标(或计件指标)。这种指标的统计分析与离散的随机变量的分布相联,在通常教科书中不占重要地位,偶尔论述到它。

(iii) 有序的。这是一种属性指标,可以反映程度深浅,例如颜色的深浅、味觉的鲜美程度、手感的丰满程度、弹性好不好等等,可以比较他们的顺序,但无法量化,这一类指标称为有序的指标。有时人们为了方便,对有序指标往往也用量化的符号与术语来表示,例如医生体检时,表示砂眼的程度用“+”号的个数来描述,工厂的产品质量用一等品、二等品、三等品来标记,这里的一、二、三没有确切量的意义,只是一个顺序的标记,所以这类指标不能认为是计数的指标,在使用统计方法时必须认识这种差异。

(iv) 名义的。这是单纯的属性指标,例如给产品编号,给公民以身份证号等等,这些号只是一个代码,它的值和大小顺序一般不表示任何意义。这一类指标称为名义指标。然而不要把名义值与名义指标混同起来。一个电阻的名义值是 10Ω (欧姆),但它的实际值可能是 10.04Ω ,名义值是产品规格上标定的值,这个值是有意义的,它不是代码。

这四大类指标可以归为两大类:

定量的:计量、计数;

定性的:有序、名义。

目前的标准大部分是与定量指标有关的,只有极少数似乎涉及定性指标,实际上也还是定量指标的问题,所以我们这本书也是以定量指标作为研究的主要对象。

指标的分类可以从不同的角度去考虑,例如生产中往往要考虑可控指标与不可控的指标,从心理学上往往要考虑是主观的指标还是客观的指标等等。我们上面介绍的分类主要是从统计分析处理的方法来考虑的,不同的指标往往处理的方法不一样,这些是必须注意的。在标准化工作中,要求用一种为大家所公认的方法来进行,以避免各搞各的一套,而且还要保证所采用的方法遵循在实践中被认为是最满意的方法。