

铁 矿 石 检 验 技 术 丛 书

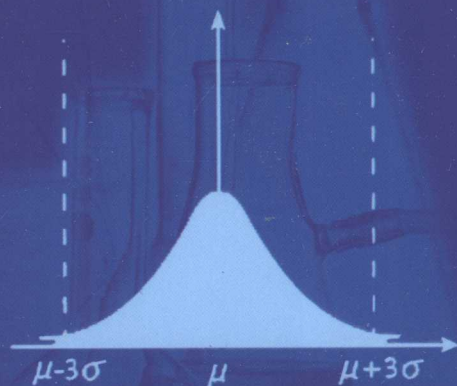
Series of Iron Ore Inspection Technology

铁矿石检验结果的 数据处理

Statistical Analysis of
Determination Results for Iron Ore

鲁国苗 程长征 应海松 主审

任春生 付冉冉 余 清 主编



冶金工业出版社
Metallurgical Industry Press

铁矿石检验技术丛书

铁矿石检验结果的数据处理

鲁国苗	程长征	应海松	主审
任春生	付冉冉	余 清	主编

北 京
冶金工业出版社
2009

内 容 提 要

铁矿石是我国经济发展所依赖的大宗战略性物资之一,一直受到国家质检部门和相关行业的高度重视,其品质分析结果直接涉及到铁矿石贸易双方的根本利益。铁矿石检测结果的数据分析与处理是铁矿石品质检测中的重要环节之一,本书综合介绍了铁矿石检测实验室检测过程中常用的数理统计分析方法。全书共分10章,主要介绍了统计检验、回归分析、方差分析、质量控制、分析方法评价、误差理论、分析结果表示、不确定度评定以及 Excel 在检测结果数据处理中的应用等,书中结合铁矿石检测实际工作中的处理个案对相关数据处理的方法进行了解析。

本书可供分析测试,特别是钢铁冶金相关企业分析测试领域的实验室工作人员、工程技术人员阅读,也可供相关专业的科研工作者和大专院校冶金及分析专业师生参考。

图书在版编目(CIP)数据

铁矿石检验结果的数据处理 / 任春生,付冉冉,余清
主编. —北京:冶金工业出版社,2009. 8

(铁矿石检验技术丛书)

ISBN 978-7-5024-5013-7

I. 铁… II. ①任… ②付… ③余… III. 铁矿物—
检验—数据处理 IV. TF521

中国版本图书馆 CIP 数据核字(2009)第 121855 号

出 版 人 曹胜利

地 址 北京北河沿大街嵩祝院北巷 39 号,邮编 100009

电 话 (010)64027926 电子信箱 postmaster@cnmip.com.cn

责任编辑 李 梅 张 卫 美术编辑 李 新 版式设计 张 青

责任校对 王永欣 责任印制 牛晓波

ISBN 978-7-5024-5013-7

北京兴华印刷厂印刷;冶金工业出版社发行;各地新华书店经销

2009 年 8 月第 1 版,2009 年 8 月第 1 次印刷

787 mm × 1092 mm 1/16;14 印张;335 千字;209 页;1-2000 册

42.00 元

冶金工业出版社发行部 电话:(010)64044283 传真:(010)64027893

冶金书店 地址:北京东四西大街 46 号(100711) 电话:(010)65289081

(本书如有印装质量问题,本社发行部负责退换)

《铁矿石检验技术丛书》

编写委员会

主 任 鲁国苗

副主任 王振新 康继韬 张忠义 应海松

委 员 (按姓氏笔画排列)

毛可辰 王志辉 王 艳 王振新 叶 卉

付冉冉 朱 波 任春生 孙锡丽 何光力

余 清 张忠义 陈健骅 陈育人 陈贺海

应海松 沈 逸 金进照 俞卫中 俞卫辉

贺存君 荣德福 郭大招 徐铭裕 陶惠君

康继韬 谢仁豪 鲁国苗 鲍仙彪 楼建元

廖海平

序

现代分析测试技术已经进入信息化时代,为有效利用分析数据,在最经济的条件下获得最理想的结果,数据处理手段已成为广大分析测试工作者的有效工具。在科技快速发展的今天,任何科学实验和工程实践所获得的数据信息,必须经过合理的数据处理并给出科学的评价,才具有实际价值。而实验室恰是检测、测量并产生数据信息的重要场所,也是生产和科研必不可少的重要基地,在各行各业的发展和进出口商品检验把关中发挥着极其重要的技术支撑作用。

在实验室的日常工作中,不可缺少的环节之一就是记录和整理大量检测数据。在这个过程中,实验室工作人员经常会遇到要判定测出数据的准确性、可靠性、评价分析方法、通过解析实验数据获得有限分析对象的最大限度的信息、利用标准物质进行质量控制、建立实验室(间)质量保证等问题。解决这些问题均需要技术人员具备相关数理统计分析知识和技巧,并根据这些知识和技巧,对实验进行全面的分析和评估,从中找出规律性的结论。在 ISO/IEC17025《检测和校准实验室能力的通用要求》中也非常重视统计评估处理方法的应用。

《铁矿石检验结果的数据处理》一书是宁波检验检疫局铁矿检测中心(国家级重点实验室)在多年的实践经验基础上,组织编写的一本实验室检测数据统计分析处理的技术参考书,该书综合了实验室常用的数理统计方法,可以为从事相关实验室检测工作的技术人员对实验室数据的分析与处理提供较全面的参考,有利于实验室,特别是铁矿石检测实验室的工作人员更加扎实有效地掌握检测数据的处理方法,从而确保检测结果的科学性和准确性。鉴于此,我特为本书作序,对其出版表示祝贺。



2009年4月15日

前 言

分析测试涉及各行各业,它是社会、经济和科技活动的技术基础,无论是政府组织,还是企业、团体或个人,都需要从分析测试的过程中获得相关信息。可以说,人类的现代生活与分析测试工作息息相关。但是,目前分析测试可靠性的现状一直令人担忧。一方面,错误结果导致错误判断的尴尬局面时有发生,重测、误测、误处理造成了巨大的经济损失和不良的社会影响;另一方面,分析测试的科技资源与快速发展的社会、经济和科技需求相比还有一定的差距,无法实现对各项测试投入足够的资源从而不断提高结果的准确度水平。因此,在考虑成本和效率的前提下,如何提高分析测量的有效性,使测量所获得的信息能够适合于应用目的,已成为数据使用者和广大分析工作者关注的焦点问题。

多年来,我国分析测试工作者在加强测试质量保证措施、不断提高分析测试结果的可比性和可靠性方面做出了积极的努力和探索,尤其是数理统计方法的引入和化学计量学的快速发展,使分析测试中获得的数据信息得到科学有效的处理,如在化学分析和仪器分析中,通过实验取得数据以后,需要合理地表示分析结果和对分析结果作出评价;判断应该舍去的无效测量数据;对分析方法作出评价;对影响分析结果的因素作出判断等。

铁矿石是我国经济发展所依赖的大宗战略性物资之一,一直受到国家质检部门和相关行业的高度重视,其品质分析结果直接涉及到贸易双方的根本利益。因此在分析测试中利用数理统计及化学计量学方法可以进一步提高所得数据分析的准确性和科学性。本书将铁矿检测实验室检测过程中常用的数理统计分析方法进行了综合介绍,内容包括统计检验、回归分析、方差分析、质量控制、分析方法评价、误差理论、分析结果表示、不确定度评定以及 Excel 在检测结果数据处理中的应用等,并结合实际工作中的处理个案对相关数据处理的方法进行了解析。希望本书能对冶金及相关行业即将步入分析测试领域和正在从事分析测试工作的技术人员以及大专院校相关专业师生有所帮助,起到抛砖引玉的作用。

本书共分 10 章,其中第 1、2、3 章和第 9 章中的第 9.1~9.3.7 节由任春生编写,第 4 章和第 6 章由付冉冉编写,第 5、7、8 章由余清编写,第 10 章由应

海松编写,第9章中第9.2.8节由廖海平编写、第9.2.9节由贺存君编写,全书由任春生统稿。于景欣、张爱珍承担了本书中表格的绘制,鲍惠君承担了部分书稿的校对工作,本书的第10章编写得到王增法的帮助;本书的出版得到了宁波检验检疫局铁矿检测中心(国家级重点实验室)的资助,也得到了冶金工业出版社的大力支持,对此表示衷心的感谢。

编者力图结合多年来的学习和实践经验,将本书编好,但由于我们的知识和水平有限,书中不妥之处,恳请广大读者批评指正。

编 者

2009年4月30日于宁波

目 录

1 概 论	1
1.1 数理统计的重要作用	1
1.2 数理统计的内容	2
1.3 数理统计的应用领域	3
1.3.1 在农业中的应用	3
1.3.2 在工业中的应用	4
1.3.3 在医、药学中的应用	4
1.3.4 在自然科学和技术科学中的应用	4
1.3.5 在社会、经济领域中的应用	5
1.3.6 在企业经营管理中的应用	5
1.4 数理统计方法中的基本概念	6
1.4.1 总体与样本	6
1.4.2 统计量	7
1.4.3 经验分布函数	7
1.4.4 正态总体统计量的分布	7
1.5 铁矿石检测中常见的数据处理基本方法	10
1.5.1 列表法	10
1.5.2 图示法	10
1.5.3 图解法	11
1.5.4 逐差法	12
1.5.5 最小二乘法	12
2 统计检验	14
2.1 概述	14
2.2 异常值检验	14
2.2.1 “ $4\bar{d}$ ”准则法	15
2.2.2 γ 检验法	16
2.2.3 格鲁布斯(Grubbs)检验法	17
2.2.4 Q 检验法	19
2.2.5 Dixon 检验法	19

·VI· 目 录	
2.2.6 几种检验方法的总结及处理可疑数据的实际考虑	20
2.3 平均值检验	21
2.3.1 平均值与标准值的比较	21
2.3.2 两个平均值的比较	23
2.3.3 F 检验	25
2.4 非参数检验	26
2.4.1 符号检验法	26
2.4.2 秩和检验法	28
2.4.3 游程检验	29
2.5 样本分布及拟合检验	31
2.5.1 概率图纸法	31
2.5.2 χ^2 拟合优度检验法	32
2.5.3 柯尔莫哥洛夫拟合检验—— D_n 检验	37
2.5.4 偏度-峰度检验法	39
3 回归分析	41
3.1 概述	41
3.2 相关分析	42
3.2.1 相关关系的类型	42
3.2.2 相关系数	44
3.2.3 相关系数的计算	44
3.2.4 相关系数的显著性检验	46
3.2.5 最小二乘法基本原理	46
3.3 回归分析	47
3.3.1 一元线性回归	47
3.3.2 多元线性回归	57
3.3.3 多项式回归	60
4 方差分析	61
4.1 方差分析原理	61
4.1.1 方差分析的概念	61
4.1.2 方差分析中常用的名词与概念	61
4.1.3 方差分析的线性模型	63
4.1.4 方差分析的基本思想	63
4.1.5 方差分析的应用条件	63
4.1.6 方差分析的步骤	64
4.2 单因素方差分析	64
4.3 双因素方差分析	67
4.3.1 无交互效应时的双因素方差分析	67

4.3.2 有交互效应时的双因素方差分析	70
4.3.3 双因素系统分组方差分析	72
4.4 多因素方差分析	75
4.4.1 三因素交叉分组全面试验的方差分析	76
4.4.2 三因素系统分组方差分析	77
4.5 协方差分析	80
5 质量控制与控制图	84
5.1 质量控制的意义	84
5.1.1 实验室质量控制途径	84
5.1.2 实验室质量控制的方法	85
5.2 质量控制图	85
5.2.1 控制图的基本构成	85
5.2.2 控制图的理论依据	86
5.2.3 控制图的分类	86
5.2.4 计量控制图的建立方法	87
5.2.5 质量控制图的判别准则	94
5.2.6 质量控制图的应用	96
6 分析方法评价	98
6.1 检出限与检测下限	98
6.1.1 检出限	98
6.1.2 检测下限	99
6.2 方法的精密度	102
6.2.1 一组测定值的精密度表示方法	102
6.2.2 标准偏差与浓度水平的关系	102
6.2.3 多组测定值的精密度表示方法	104
6.3 方法的准确度	105
6.4 方法的灵敏度	106
6.5 方法的适用范围	107
6.6 分析空白	107
6.6.1 分析空白在痕量分析中的地位	107
6.6.2 分析空白的主要来源及其特点	109
6.6.3 分析空白的监测和空白值的正确扣除	110
6.7 痕量分析报告数据的准则	111
7 误差理论	113
7.1 误差的分类	113
7.1.1 系统误差	113

·VIII·	目 录	
7.1.2	偶然误差	115
7.1.3	粗大误差	115
7.2	误差的理论基础与分布特性	115
7.2.1	误差的正态分布	116
7.2.2	正态分布误差的区间概率	116
7.2.3	误差的 t 分布	117
7.2.4	均匀分布	118
7.3	误差的传递	118
7.3.1	误差的传递规则	118
7.3.2	误差传递的基本公式	119
7.3.3	误差传递在铁矿石分析中常见的应用公式	119
8	分析结果表示	121
8.1	有效数字与修约规则	121
8.1.1	有效数字	121
8.1.2	有效数字的正确表示方法	121
8.1.3	有效数字的应用说明	121
8.1.4	“四舍六入五留双”修约准则	123
8.1.5	计算修约规则	123
8.1.6	有效数字运算规则	126
8.2	分析结果的表示	127
9	铁矿石检测中的测量不确定度评定	129
9.1	检测结果不确定度评定的意义	129
9.2	测量不确定度基本知识	129
9.2.1	测量不确定度的概念	130
9.2.2	测量不确定度与误差的关系	130
9.2.3	不确定度的来源	131
9.2.4	几个概念	131
9.2.5	测量不确定度的分类	131
9.2.6	测量不确定度的评定流程	132
9.2.7	测量不确定度的报告	135
9.3	铁矿石检测测量不确定度评定举例	137
9.3.1	重铬酸钾滴定法测定铁矿石中全铁含量的不确定度评定	137
9.3.2	火焰原子吸收光谱法测定铁矿石中铜含量的不确定度评定	140
9.3.3	X 射线荧光光谱法测定铁矿石中二氧化硅含量的不确定度评定	143
9.3.4	高频红外碳硫仪法测定铁矿石中硫含量的不确定度评定	147
9.3.5	自动电位滴定法测定铁矿石中铁含量的不确定度评定	150
9.3.6	电感耦合等离子光谱法测定铁矿石中铜含量的不确定度评定	153

9.3.7 紫外可见分光光度法测定铁矿石中铜含量的不确定度评定	157
9.3.8 球团矿抗压强度检测结果的不确定度评定	163
9.3.9 铁矿石水分测定的不确定度评定	165
10 Excel 在检测结果数据处理应用	169
10.1 变异系数、数值比较	169
10.1.1 变异系数	169
10.1.2 数值分析	174
10.2 线性相关与回归	175
10.2.1 线性相关	175
10.2.2 线性回归	177
10.3 方差分析	185
10.3.1 单因素方差分析	185
10.3.2 双因素方差分析	187
10.4 其他 Excel 数据处理方法	189
10.4.1 非参数统计的符号检验	189
10.4.2 重复性和再现性计算	190
10.4.3 铁矿石水平测试数据分析	196
10.4.4 测量不确定度 Excel 分析实例	203
10.4.5 常用 Excel 函数	205
参考文献	209



概 论

随着自然科学的不断发展,在各学科领域中,仅仅依靠定性描述方法已远不能满足研究需要,定量研究方法日趋重要。然而像生物科学、大气科学、地球科学和环境科学等学科,其研究对象往往是一些受到多重复杂自然因素和人为因素影响的随机现象,一般的数学方法在这些现象的研究方面不太理想。正因为如此,这些学科研究的定量水平始终低于物理学等传统学科,而数理统计方法的发展在提高此类学科研究的定量水平方面无疑具有举足轻重的地位。

数理统计方法是研究随机现象内在规律的科学。所谓随机现象是指那些在相同条件下,可能得到不同观测结果的现象。虽然应用数理统计方法在不同学科领域的应用过程中已发展成为不同的分支,如生物统计学、医药统计学、工程统计学和经济统计学等,但这些分支学科除了在具体研究对象方面有所不同以外,其方法学本身并没有根本区别。

1.1 数理统计的重要作用

数理统计是研究大量随机现象的统计规律性的一门数学科学。它以概率论为基础,研究如何用有效的方式收集、整理和分析受到随机性影响的数据,从而为随机现象选择相应的检验数学模型,并且在此基础上对随机现象的性质、特点和统计规律作出推断和预测,直至为决策提供依据和建议。

进口铁矿属法定检验商品,所有从国外进口的铁矿都需经我国检验检疫部门检验才能销售使用,我国口岸检验检疫局与国内钢铁企业相比应当有更多的机会了解国外不同产地铁矿的品质情况。但以往每批次铁矿检验品质资料的保留都是以台账形式人工记录,而铁矿品质、数量的检验数据及资料多而复杂,因此在品质分析或数据统计时非常不便。

从研究方法论的观点看,人们总是从研究对象的局部着手,进行深入的考察、实验、分析、判断与推理,进而达到对研究对象总体全面而深刻的认识。例如,要检验一批出口罐头是否合格,并要求就此做出判断,这里涉及的是破坏性检验,不允许对全部罐头实施开罐检验。其道理非常明显,如果对全部罐头实施开罐检验,经过开罐检验之后,全部罐头也就都变成了不能再出售的废品了,这时不管对这一批被检验罐头的质量做出什么样的结论也都毫无意义了。

在有些情况下,全面检验虽然在客观上是可能的,在技术上也是做得到的,但也未必这样去做。例如,要研究某工厂含汞废水对附近水库中鱼的污染程度,虽然从理论上讲,可以将水库中的鱼全部捕捞上来,逐条进行鱼体内汞含量的测定,但实际上,人们绝不会这样做。其原因很简单,因为逐条鱼进行汞含量测定,检验费用高,经济上也不合理。更何况这样的检验也如前面提到的罐头质量检验一样,是一种破坏性检验,不允许进行全检,而只能

从水库中随机地捕捞若干条鱼,测定鱼组织内的汞含量,从所测得的汞含量数据去推断水库中鱼受汞污染的程度。由于不同鱼组织内汞含量的差异和测试过程中的误差,汞测定值会出现波动,这样就产生了一个疑问:由若干条鱼的汞测定值去推断水库中鱼总的受汞污染的程度,其结论可靠和可信吗?如果使用数理统计方法正确地进行统计推断,其结论当然是可靠和可信的;否则,就无法保证推断的正确性。

实验室技术工作者既有研究分析化学理论和改进分析方法的任務,又有从事测定物质的化学组成的实际工作。在完成上述任务和工作的过程中,都必须进行实验,因此也都必须进行数据处理工作。在分析化学中,通过实验只能得到一些数据(例如,容量分析得到的是体积(如毫升数),重量分析法得到的是质量(如克数),吸光光度法和原子吸收光谱分析法得到的是吸光度 A 的数值等),应用这些数据进行适当的数学运算才能得到分析结果。为了获得准确可靠的分析结果,或要判断测定结果的可靠性,或者对分析方法进行考核和评价,就必须进行多次实验,然后应用这些实验的数据进行适当的数学运算才能得出切合实际的结果或结论。从实验数据出发,到获得分析结果或结论的数学运算过程称为“数据处理”。数据处理必需应用概率论与数理统计方法来完成。

然而,为什么试验(测量)数据必需应用数理统计方法来处理呢?这是因为分析化学实验所得到的数据都不是绝对准确的,都会有随机性质的偶然误差。例如,即使是同一个工作者,用同一种分析方法,在相同的条件下,对同一种样品进行多次重复测定,测定结果也不会是完全相同的。偶然误差是由于多种随机因素所引起的随机事件,其数值是不可测的。虽然偶然误差是随机性的、不确定的,但是它仍然是有规律性的,而概率论和数理统计方法正是研究和处理随机事件的属性和规律的科学方法,所以分析化学实验的数据,应该且必须运用数理统计方法来进行处理。只有这样,我们才能“透过”偶然误差,从实验数据中获得可靠的结果和结论,才能收到多方面的效果,具体包括:

- (1) 判断应舍弃的错误数据或结果;
- (2) 得到最精确的分析结果;
- (3) 判断分析结果的可靠程度;
- (4) 评价分析方法,分析工作者的质量;
- (5) 评价影响分析结果的各种因素的影响;
- (6) 从测定数据中得出最多的结论;
- (7) 对测定工作进行质量管理;
- (8) 使实验工作最经济。

但是也应该指出,数理统计方法是不能改变实验或测定工作本身的质量的,只能是在已有的测定数据的基础上,发挥上述作用。

由此可见,对一个分析化学家或分析化学工作者来说,如果没有掌握必要的概率论知识和数理统计方法,不会正确地进行数据处理,那么即使他的实验工作做得很出色,也不能得到全面、正确的结果和结论。相反,通过正确的数据处理工作,工作人员往往可以发现测定工作中的缺陷或不足之处,指出改进的途径。

1.2 数理统计的内容

数理统计学是“数学的一个分支学科,研究怎样去有效地收集、整理和分析带有随机性

的数据,以对所考察问题做出推断或预测,直至为采取一定的决策和行动提供依据和建议。”这一定义明确了数理统计学的研究对象与研究目的。

数理统计学是应用性很强的学科,它已被应用于各种专门领域(如物理、化学、工程、生物、经济、社会等),但只涉及其中有关带随机性的数据的分析问题,而不是以任何一种专门的知识领域为研究对象。但是,在应用数理统计方法分析带有随机性数据时,从统计模型的选择、实验方案的制订、统计方法的正确使用以致所得结论的恰当解释,都离不开所讨论问题的专门知识。

数理统计学内容庞杂,分支学科很多,难以做出一个周密的分类,这里仅做一般性分类。这些分支学科大体可以划分为以下三类:

第一类分支学科包括抽样调查和试验设计,主要涉及与数据收集有关的理论和方法问题。

第二类分支学科主要涉及统计推断的原理与方法,包括:

- (1) 与特定的统计推断形式有关的参数估计和假设检验;
- (2) 与特定的统计观点有关的贝叶斯统计和统计决策理论;
- (3) 与特定的理论模型或样本结构有关的非参数统计、多元统计分析、相关分析、回归分析、方差分析、序贯分析、时间序列分析和随机过程统计。

第三类分支学科是一些针对特殊的应用问题而发展起来的分支学科,如产品抽样检验、可靠性统计、统计质量管理等。这类分支学科都不涉及或很少涉及任何一种专门学科的知识,但一般需要考虑数据的收集和统计推断两方面的问题。例如,产品抽样检验的任务是制定从一批产品中做随机抽样的方案,并依据由此获得的样本去决定是否接受该批产品。这里,有抽样方案的统计问题,也有使用数据作统计假设检验的问题。

除以上三类分支学科之外,还有一类称之为边缘学科分支,如生物统计学、医学统计学、气象统计学、地质统计学、教育统计学、经济计量学、社会计量学、政治计量学、语言计量学、历史计量学等。这类分支学科所讨论的是统计方法在某一特定学科中的应用,涉及大量有关学科的专门知识,故不适宜当作数理统计学的分支。

1.3 数理统计的应用领域

由于随机性影响无所不在,因而数理统计的应用十分广泛,在自然科学、社会科学、工程技术、军事科学、医药卫生和工农业生产中都常常用到数理统计的理论与方法,尤其是近二十年来,电子计算机的发展与普及,各种使用方便的统计程序包(或程序库)的出现,使得各行各业中只要具有基本统计知识的人,都可以利用统计分析的各种工具软件,来为自己的研究课题服务。数理统计正在发挥着越来越大的作用,它的应用更加广泛深入。

从上面数理统计学的分支学科的划分,我们大致可以看出数理统计学的应用领域与范围。数理统计方法在工农业生产、自然科学和技术科学以及社会经济领域中都有广泛的应用。数理统计方法的恰当应用依赖于所讨论问题的专门知识和经验,以至良好的组织工作,故从本质上来说,数理统计方法基本上是一种辅助性的工具。

1.3.1 在农业中的应用

数理统计方法在农业中应用的一个主要方面,是对田间试验进行适当的设计和统

计分析。试验设计的基本思想和方法,就是从田间试验开始发展起来的。数理统计方法在农业中应用的另一领域是数量遗传学。例如,在培育高产优质农产品的研究中,其数据分析就需要使用多种统计方法,如应用很复杂的回归和方差分量分析的方法来计算遗传力。

1.3.2 在工业中的应用

数理统计方法在工业中的应用,主要有两方面:一是应用正交设计、回归设计、回归分析、方差分析、多元统计分析等统计方法来解决以下一系列问题:试制新产品和改进老产品,改革工艺流程,使用代用原材料和寻找适当的配方,判定影响产品质量的重要因素、次要因素,决定一组最优的生产条件;二是应用统计质量管理的统计方法,通过种种形式的质量控制图、抽样检验、可靠性统计分析以解决下面一些与工业中大批量及连续生产有关的问题:工序控制,制定成批产品的抽样验收方案,对大批生产的元件的可靠性及包含大量各种元件的系统的可靠性。

1.3.3 在医、药学中的应用

医、药学是较早使用数理统计方法的领域之一。在医学研究中,为防治一种疾病,统计方法常被作为重要的研究工具,用来发现和验证导致这种疾病的种种因素。例如,应用统计方法证实肺癌与吸烟的关系。数理统计方法在医学中的应用,就形成了医学统计学、生存分析等边缘分支学科,也构成了理论流行病学、临床流行病学的主要内容。在药学研究中,通过临床试验,应用正交设计、交叉设计、回归分析、方差分析、列联表分析等统计方法,来确定一种药物对治疗某种疾病是否有效,用处多大,以及比较几种药物或治疗方案的效力。

1.3.4 在自然科学和技术科学中的应用

在基础理论研究中,一方面,一种学说或假说是否正确,或在多大程度上正确,要诉诸大规模的实验验证,其中就有实验的设计和数据的统计分析问题;另一方面,统计分析也有助于发现某种规律性。一个著名的例子就是孟德尔遗传定律的发现。孟德尔在豌豆试验中发现了这一定律,以后由许多人通过进一步的试验,并用数理统计学的拟合优度检验法验证了该定律。在寻求该定律的理论解释的过程中,便诞生了“基因学说”。

在应用性研究中,由于对所研究现象的规律性认识不充分,人们不得不依靠对实验和观测数据进行统计分析(统计推断与统计预测),以提出解决问题的办法。例如,统计方法用于地震、气象和水文方面的预报,都有一定的效果;在地质勘探中,人们在一个地区的若干点(点的选择也有统计上的考虑)进行观察,对其结果用种种统计方法,如趋势面分析、对应分析等进行处理,以建立某种经验性质的规律,用以指导找矿。

数理统计方法在自然科学和技术科学中,已得到很广泛的应用,随即便产生了许多诸如统计物理学、计量化学、地质统计学、气象统计学、数量遗传学、生物统计学、水文统计学、技术计量学等边缘分支学科。一般而言,无论是自然科学还是技术科学,都离不开实验观察,都有处理数据的问题,因此也就有统计方法的用武之地。可以这样说,凡是有数据的地方,都是数理统计方法的应用去向。

1.3.5 在社会、经济领域中的应用

数理统计方法在社会、经济领域中,有着重要的应用。在西方发达国家,数理统计方法在这些领域中的应用,要比其在自然科学和技术领域中的应用更早且更广泛。例如,社会学中的抽样调查、列联表分析,人口学中的人口发展动态模型、随机过程统计、经济学中的经济计量模型等的应用促进了人们对社会现象的研究向定量化发展。在经济科学中,定量化的发展趋势要比其他人文社会科学部门更早且程度更深,如早在 20 世纪的二三十年代,时间序列分析方法就已经用于市场预测,继而一门边缘分支学科——经济计量学便应运而生,值得注意的是,自 1969 年设立诺贝尔经济学奖以来,其得主三分之二以上的都是经济计量专家,此外,教育统计学、人口统计学、经济计量学、社会计量学、政治计量学、语言计量学、历史计量学等边缘分支学科也正是数理统计方法在社会、经济领域中的应用结晶。

1.3.6 在企业经营管理中的应用

1.3.6.1 市场研究

在市场经济中开展市场研究尤其重要。市场研究是为某一特定的市场营销问题的决策而开发和提供其所需的信息的一种系统的、有目的的活动或过程。这里所说的信息,不仅是市场调查所得的数据资料,还包括市场研究人员对资料进行分析所得的结果(如结论、建议等)。市场研究的范围包括:产品研究、销售研究、购买行为研究、竞争分析、广告及促销研究、销售成本和利润分析、营销环境研究。这些研究又可归结为四类:

(1) 探测性研究。适用于需要研究的问题和范围不是很明确,无法确定调查内容,引起问题的原因不很清楚的情形。通常的做法是:收集与分析第二手资料,集中专家或相关人员的意见,进行小规模的重点调查、定性研究、实例分析。

(2) 描述性研究。回答事先计划好的问题,如“什么”、“何时”、“怎样”。通常的做法是:给出统计图;计算基本统计量(如相对指标、平均指标、变异指标等),进行相关分析,确定市场营销中有关问题的相关因素、相关关系及关联程度。

(3) 因果关系研究。用于寻找问题的原因,或确定变量之间有无因果关系。通常的做法是:先确定分析对象及其影响因素,然后建立具有单方程或联立方程形式的经济计量模型,应用回归分析方法进行测定。

(4) 预测性研究。用于对市场潜力、企业产品前景等进行预测,以此作为市场营销决策的依据。通常的方法包括:专家意见法,移动平均法,指数平滑法,经济计量模型法,投入产出模型法等。

在市场研究中,常用且有效的方法就是应用数理统计方法,比较基本的有列表分析、相关分析、方差分析;比较高级的有多元统计分析中的聚类分析、判别分析、主成分分析、因子分析、多特性模型等。下面主要介绍多元统计分析中的方法。

多元统计分析中聚类分析的目的在于辨别在某些特性上相似的事物,并按这些特性将样本划分成若干类(群),使在同一类内的事物具有高度的同质性,而不同类的事物则有高度的异质性。在市场研究中,聚类分析主要用于对消费者群进行市场细分,对产品进行分类,选择试验市场,确定分层抽样的层次,分析消费者的性格特征和行为形态等方面。