

温度计量测试丛书 (三)

温度计量中的 数据处理方法

WENDU JILIANG ZHONG DE
SHUJU CHULL FANGFA

戚盛勇 編著
戴永海 審

中国计量出版社

71.44
570

温度计量测试丛书 (三)

温度计量中 的数据处理方法

戚盛勇 编著 戴乐山 审

2K506/16

中国计量出版社

新登(京)字 024 号

内 容 提 要

本书系“温度计量测试丛书”之一。主要内容包括：数理统计和概率论的基本概念；随机变量性质和概率分布；总体和样本的概念；数据报道；假设检验和曲线拟合的方法；应用数理统计理论进行管理和控制产品的质量方法等。

大书可供从事温度计量、测试和检定人员以及广大工程技术人员阅读，亦可供大专院校有关专业的师生参考。

温度计量测试丛书 (三) 温度计量中的数据处理方法

戚盛勇 编著 戴乐山 审

责任编辑 窦绪昕

—*—

中国计量出版社出版

北京和平里西街甲 2 号

中国计量出版社印刷厂印刷

新华书店北京发行所发行

—*—

开本 787×1092/32 印张 10 字数 223 千字

1992 年 3 月第 1 版 1992 年 3 月第 1 次印刷

印数 1—5 000

ISBN 7-5026-0471-5/TB·368

定价 6.50 元

温度计量测试丛书编委会

主任委员 王良楣

副主任委员 凌善康

委员 (以姓氏笔划为序)

石质彦 朱国柱 师克宽

汪时雍 陈守仁 陈锡光

张立儒 周本濂 赵 琪

秦永烈 崔均哲 窦绪昕

戴乐山

前 言

本丛书是根据中国计量出版社关于按学科分类组编丛书的总体规划和统一安排，在中国计量测试学会的热情关怀和实际支持下，由温度计量测试丛书编辑委员会编写的。

计量是现代化建设中一项必不可少的技术基础，在计量测试科学领域中，温度的计量与测试又是一个很重要的方面。温度是一个基本的物理量，它与其它许多物理参数有着密切的关系，因而在工、农业生产，在科学研究和日常生产中都离不开温度的准确测量和精密控制。广泛传播温度及温度测量仪表的基本知识，介绍国内外测温技术的先进经验，交流各项成果，培养技术人才，促进各项工作，为实现四个现代化创造条件，这就是组编本丛书的宗旨。

应该看到，目前，在基层企业中，受过计量测试训练的技术人员严重不足，很多职工渴望增长专业知识和提高操作技能；尤其是近年来，大批青年技术人员参加工作，这是发展计量测试科学的一支新生力量，但是他们深感知识不足，迫切需要系统地学习一些计量基础知识，熟悉各类仪器仪表的原理、特性、检定和使用方法，以便更快地掌握专业技术，提高生产效率。这套丛书主要是针对这部分人员编写的，当然也可作为温度计量短训班的教材及有关学校师生、工程技术人员和科研工作者的参考书。

本丛书计划分成16分册，每一分册独立地、深入浅出地加以阐述，将陆续出版与读者见面。本丛书在组编过程中得到广大计量工作者和工矿企业技术人员的关心与支持，在此

42254

1

一并致谢。丛书编委会热忱地期望我国广大科学工作者共同促进本丛书的编辑出版工作。为我国早日实现四个现代化贡献力量。

限于我们的经验和水平，本丛书可能存在不少缺点和错误，恳请广大读者批评指正。

温度计量测试丛书编辑委员会

编 者 的 话

本书的目的是介绍数据处理的原则和方法以及在计量工作中的实际应用。由于厂矿企业的计量工作者不仅需要具备计量工作方面的业务知识，而且还要逐步具备技术监督的职能；所以本书也介绍了质量管理、抽样检查、试验设计等方面的知识。

本书着眼于方法的介绍，而不过分强调数学上的严格性，适宜于在厂矿企业中作技术工作或管理工作的同志阅读使用。为配合本书内容，我们编写了相应的计算机程序。有IBM系列微机的单位，配上该程序就能方便地作数据处理和抽样检查等方面的工作。

根据由浅入深和从具体到一般的原则，在第一章中提出了薄膜铂电阻温度计阻值的数据处理问题，作为一个实例，引出计量工作中数据处理的一些基本概念，在二、三、四章中应用这个例子作进一步深入的分析。

第二、三章讨论了随机变量性质和它的概率分布，重点介绍正态分布的性质和期望值以及方差的意义。第四章讨论了总体和样本的概念，并从薄膜铂电阻温度计的实例出发，引出了样本统计量所服从的 t 分布， χ^2 分布和 F 分布的概念，也简述了它们的性质。这样引出的样本分布显得直观、形象，便于读者理解。

前四章是从计量测试的实例出发，介绍了数理统计和概率论的基本概念，从第五章起将应用这些基本知识介绍计量测试中数据处理的几种方法和它们的应用。第五章着重介绍

了数据报道、假设检验和曲线拟合三种方法和应用，并以近年来国际上在温度计量科研工作中实际的数据处理过程为例，说明它们的一些基本概念和应用。

第六章介绍了抽样检查的基本概念，并具体介绍了1987年原国家标准局公布的有关计量抽样检查和计数抽样检查的国家标准，也介绍了有关的国际标准。第七章介绍了应用数理统计理论进行管理和控制产品质量的几种方法，具体介绍了质量控制图。第八章讨论了最优化的试验设计问题，具体地介绍了正交试验设计和方差分析。

第九章讨论了误差和不确定度的分类和合成，并以测量水、氩三相点温度和锌凝固点温度为例，具体说明不确定度的合成方法。

本书由以下同志合作编写：戚盛勇同志编写第一、四、六、七、八章，潘玉莲同志编写第二、三章，王志敏同志参加了“程序”的编写工作，李白云同志编写第九章。本书是根据我们在计量学和近代物理实验工作的基础上写成的，由于我们水平有限，书中难免存在不少缺点或错误，敬请读者批评指正。

编 者

目 录

编者的话

第一章 一个实例	(1)
第一节 电阻测量值的表示方法.....	(1)
第二节 电阻测量值的分布状况——直方图	(7)
第三节 测量值的分布曲线.....	(10)
第二章 随机变量及其分布	(14)
第一节 随机现象和随机变量.....	(14)
第二节 离散性 (型) 随机变量.....	(17)
第三节 连续性 (型) 随机变量.....	(22)
第四节 正态分布.....	(27)
第五节 其他常用分布.....	(37)
第三章 随机变量的数字特征量	(45)
第一节 离散性随机变量的期望值和方差.....	(45)
第二节 连续性随机变量的期望值和方差.....	(50)
第三节 相关系数.....	(52)
第四节 数字特征量的运算法则.....	(60)
第五节 标准误差的传递公式.....	(62)
第六节 铜-康铜热电偶的分度误差	(69)
第四章 总体和样本	(75)
第一节 总体与样本的联系和区别.....	(75)
第二节 样本平均值分布.....	(78)
第三节 样本方差分布—— χ^2 分布.....	(82)
第四节 t 分布	(91)
第五节 F 分布.....	(97)
第五章 实验数据的报道、检验和拟合	(103)

第一节	平均值报道 (一) —— σ^2 已知	(103)
第二节	平均值报道 (二) —— σ^2 未知, t 报道	(107)
第三节	银凝固点和铈三相点温度值的报道	(112)
第四节	样本方差报道 —— X^2 报道	(117)
第五节	假设检验的基本原理和方法	(120)
第六节	假设检验在温度计量中的应用	(124)
第七节	测试方法的精密度 —— 重复性、再现性及 GB 6379—86	(128)
第八节	最小二乘法	(143)
第九节	气体温度计 (一): 热力学温度与第二维 里系数的计算	(150)
第十节	多项式和多元函数的拟合	(152)
第十一节	气体温度计 (二): 第二维里系数与温 度的关系	(155)
第十二节	铂电阻温度计的电阻-温度关系	(157)
第六章	抽样检查	(160)
第一节	抽样检查	(160)
第二节	计数和计量抽样检查	(166)
第三节	计数标准型一次抽检方案	(168)
第四节	计数调整型抽检方案 (国家标准 GB 2828 —87)	(181)
第五节	计量标准型抽检方案 (国家标准 GB 8053 —87)	(195)
第六节	计量调整型抽检方案 (国际标准 ISO 3951 —81)	(210)
第七章	数理统计在计量管理中的应用	(220)
第一节	直方图的应用	(220)
第二节	计量工作中的控制图	(227)
第三节	计量值控制图	(232)
第四节	计数值控制图	(242)
第五节	控制图分析	(245)

第六节	控制图在计量工作中的应用·····	(250)
第八章	试验设计与方差分析·····	(258)
第一节	正交试验设计·····	(258)
第二节	方差分析·····	(281)
第三节	正交试验中的方差分析·····	(290)
第九章	不确定度的合成·····	(298)
第一节	误差的分类·····	(298)
第二节	不确定度的分类与合成·····	(299)
第三节	复现水三相点的不确定度·····	(300)
第四节	复现锡和锌凝固点的不确定度·····	(302)
第五节	复现氩三相点的不确定度·····	(305)

第一章 一个实例

在这一章中，我们将通过对薄膜铂电阻测量值的处理介绍数据处理中的一些基本术语，如样本平均值、极差、样本标准偏差、频数、直方图、分布曲线……等等。为以后几章学习提供一个形象的实例，作为学习数据处理的一个入门。

第一节 电阻测量值的表示方法

某工厂生产阻值为 1K 的薄膜铂电阻，使用温度范围从 0℃ 到 660℃。为了检验产品的质量，从产品中随机地抽取 400 只电阻进行测试，结果如表 1-1 所示。

怎样表示这 400 只电阻的测量结果及对这批产品的质量作出评价呢？当然最简单的办法是求它们的平均值。即

$$\begin{aligned}\bar{R} &= (R_1 + R_2 + \cdots + R_{400}) / 400 \\ &= \frac{1}{400} \sum_{i=1}^{400} R_i \\ &= 999.999_5 \Omega\end{aligned}$$

一般的，可表示为

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i \quad (1-1)$$

式中， n 是抽取的只数， R_i 是第 i 个电阻的阻值， $\bar{R}$ 称为样本平均值。因为这 400 个电阻是从一批电阻中随机抽取的，它是这批电阻的一个样本，所以它的平均值就称为样本平均值。

表 1-1

薄膜铂电阻阻值

1 000.396	999.976	1 000.608	1 000.712	999.836
999.864	999.828	1 000.084	1 000.244	999.808
999.780	1 000.180	1 000.120	999.972	999.592
1 000.236	1 000.080	1 000.080	1 000.096	1 000.200
999.884	1 000.296	1 000.564	1 000.424	1 000.160
1 000.068	1 000.692	1 000.732	999.676	1 000.400
1 000.460	999.784	999.760	1 000.036	999.848
1 000.500	999.192	999.980	1 000.525	998.860
1 000.140	999.680	1 000.236	1 000.472	1 000.820
1 000.288	999.784	1 000.044	1 000.148	1 000.544
1 000.428	1 000.208	999.872	999.752	999.884
1 000.060	1 000.612	1 000.056	998.984	999.688
1 000.012	1 000.480	998.980	1 000.308	1 000.388
1 000.048	1 000.136	1 000.028	999.948	999.904
999.884	1 000.488	999.880	999.928	999.820
999.512	1 000.088	999.992	1 000.600	999.732
1 000.628	999.964	1 000.108	1 000.104	999.564
999.728	999.688	1 000.004	999.924	1 000.424
1 000.308	1 000.936	999.916	999.524	999.968
1 000.012	1 000.440	999.872	999.776	1 000.148
999.488	1 000.776	999.720	999.452	1 000.460
999.424	999.064	1 000.276	1 000.200	1 000.072
999.352	1 000.436	999.824	1 000.476	1 000.276
999.880	1 000.096	999.656		

续表

1 000.432	999.752	999.736	1 000.304	1 000.316
1 000.744	1 000.780	999.864	999.752	1 000.168
1 000.000	999.928	999.924	999.588	1 000.212
1 000.020	1 000.168	1 000.296	999.976	1 000.604
999.980	1 000.612	1 001.116	1 000.760	999.684
1 000.120	999.224	999.884	999.900	1 000.260
1 000.220	1 000.508	1 000.116	1 000.332	999.844
999.608	1 000.380	1 000.216	1 000.476	999.392
1 000.320	999.780	999.840	1 000.372	999.844
999.328	1 000.032	1 000.224	999.708	999.864
1 000.044	999.604	1 000.108	1 000.412	1 000.280
1 000.016	999.820	999.712	1 000.012	1 000.548
999.604	999.908	1 000.352	1 000.160	1 000.216
999.596	1 000.436	999.788	999.868	1 000.208
998.964	1 000.220	1 000.028	999.492	999.740
999.848	1 000.176	999.732	999.768	1 000.456
999.804	1 000.236	1 000.312	1 000.716	999.944
999.460	999.220	999.296	1 000.656	999.824
999.868	999.400	999.012	1 000.220	999.840
999.412	1 000.164	1 000.228	1 000.096	999.988
1 000.336	999.868	1 000.340	999.372	1 000.716
999.864	999.980	999.632	1 000.216	999.524
1 000.276	999.720	999.320	999.852	999.504
1 000.576	1 000.192	999.168	1 000.340	999.452

我们除了关心这 400 只电阻的平均值大小外，还常常需要了解这 400 只电阻阻值的弥散程度。因为电阻的平均值虽然与标准值 1K 的差值仅为 0.0005Ω ，但是如果电阻的阻值分布范围很宽，例如分布在 $900\sim 1100\Omega$ 范围内，那末我们认为这批电阻的质量还是很差的；相反，如果分布范围为 $999\sim 1001\Omega$ ，则这批电阻的质量就要比前者好得多。所以数据的弥散程度是一个重要参数。

表示数据弥散程度的最简单的办法是用极差 R_x 。它是本组数据的最大值 $R_{\max}$ 减去最小值 $R_{\min}$ 。即

$$R_x = R_{\max} - R_{\min} \quad (1-2)$$

在上述 400 个数据中， $R_{\max} = 1001.116\Omega$ ， $R_{\min} = 998.860\Omega$ ，所以极差

$$R_x = 1001.116 - 998.860 = 2.256\Omega$$

用极差 R_x 表示一组数据的弥散程度十分简便，所以是工厂中常用的一种方法。但它丢掉了中间大量数据的信息，因此用它表示数据的弥散程度往往是不很确切的。

为了更确切地表示数据的弥散程度，最佳的方法是样本标准偏差 S_x 。它的定义是

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (1-3)$$

这 400 只铂电阻阻值的样本标准偏差为

$$S_x = \sqrt{\frac{\sum_{i=1}^{400} (x_i - 999.999)^2}{399}} = 0.40$$

其中 x_i 即是 400 只电阻中的第 i 只的测量值。

为什么样本标准偏差 S_x 能较好地反映数据的弥散程度呢？我们可以分析一下式 (1-3) 的构成。其中 $(x_i - \bar{x})$ 是每一个测量值与平均值的偏差，它反映了每一个数据偏离平均值的程度。能否把这 400 个 $(x_i - \bar{x})$ 值简单的加起来，用它来反映 400 个数据的偏离程度呢？不能！因为偏差 $(x_i - \bar{x})$ 有正有负，如把它们简单相加，则这些正负偏差可能刚好相互抵消，所以 $\sum_{i=1}^{400} (x_i - \bar{x})$ 的结果接近于零，不能反映它们的弥散程度。如把 $(x_i - \bar{x})$ 取平方后再相加，即 $\sum_{i=1}^{400} (x_i - \bar{x})^2$ ，这样 400 个 $(x_i - \bar{x})^2$ 的值都是正数，就不会相互抵消了。

但是平方和 $\sum_{i=1}^{400} (x_i - \bar{x})^2$ 还是不能确切地反映数据的弥散程度。因为，虽然数据的弥散程度相同，但如果被测电阻数目 n 增加，则有更多的 $(x_i - \bar{x})^2$ 项相加，其结果 $\sum_{i=1}^n (x_i - \bar{x})^2$ 必然增大。为了克服这一缺点，把平方和 $\sum_{i=1}^n (x_i - \bar{x})^2$

再除以自由度 $(n-1)$ 。

什么叫自由度呢？简单地说，自由度就是在平方和中能自由地安排其值的数据的个数。作一个通俗的比喻：如四个人分吃一只月饼，第一个人可以取月饼中任意大小的一部分，第二个人可以取剩下的月饼中任意大小的一部分，第三个人也如此，但第四个人却没有自由了，他只能吃前面三个人留下的这一小块，没有选择的余地，所以这四个人的自由度为 3。

再举一个数学的例子：如有 x_1, x_2, x_3, x_4 四个数，约定它们的和等于 a ，即

$$x_1 + x_2 + x_3 + x_4 = a \quad (1-4)$$

在式(1-4)的四个数中，有三个数的大小是可以任意改变的，但第四个数的值就不能任意改变了，它受到式(1-4)的约束，设 x_4 是第四个数，则 $x_4 = a - x_1 - x_2 - x_3$ ， x_4 不可能是其他值，所以这四个数的自由度是 3。

现在回来讨论平方和 $\sum_{i=1}^n (x_i - \bar{x})^2$ 的自由度。同样在这 n 个 x_i 中有一个 $\bar{x} = (x_1 + x_2 + \dots + x_n)/n$ 的约束条件，因此它的自由度为 $(n-1)$ 。如果测量值的个数 n 越大，平方和的值 $\sum_{i=1}^n (x_i - \bar{x})^2$ 越大，但自由度 $(n-1)$ 也越大，所以

$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ 能确切地反映一组数据的弥散程度。我们称

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (1-5)$$

为样本方差。为了使量纲与被测量的量纲一致，把样本方差 S_x^2 开方，得样本标准偏差 S_x 。

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

在 S_x^2 、 S_x 的名称上都加上“样本”二字，是为了表明样本方差 S_x^2 或样本标准偏差 S_x 都是从一个样本中求得的。在本实例中，它是从 400 只被测电阻的阻值中求得的，如果