误差分析方法

何 国 伟 编著

国防工业出版社

内容简介

本书简要地介绍了误差的基本概念，和分析、估计、评定、分配及综合误差的方法，其中包括了异常数据的取舍、正态性检验等一些较新的方法。着重点在这些方法的概念、物理意义及如何应用。可供工程技术人员阅读参考。

误差分析方法

何国伟 编著

*

国防工业出版社 出版

北京市书刊出版业营业許可証出字第074号

新华书店北京发行所发行　　各地新华书店经售

山西省原平县印刷厂印装

*

787×1092 1/32　印张 4 3/8　88千字

1978年11月第一版　1978年11月第一次印刷　印数：00,001—50,000册

统一书号：15034·1733　定价：0.37元

序　言

本书是以一般工程技术人员所能接受的观点及数学水平介绍误差的基本概念，分析、估计、评定、分配及综合误差的方法。因此，我们着重的是这些方法的概念、物理意义及如何应用，而不详细从概率统计理论加以推导证明。有兴趣的同志可以查阅本书引用的参考资料。

这里，力图纠正我们在工作中遇到的、在工程技术人员中较多见的一些误差概念上的误解。例如，把子样均值及子样方差当成母体的均值及方差来进行误差的评定及综合，于是作出的结论的置信度是不高的，差不多有一半的可能把实际误差估小了。

我们推荐用国内外较新研究得到的方法来取代某些成套资料及工程技术人员中还在使用着的较落后的方法。例如，我们推荐用格拉布斯方法（1952，1969）代替肖维纳方法（1876）来决定异常数据的取舍……等等。

我们不把国内外较新研究得到的方法一一列举，而只是介绍经过我们的分析、实践，认为最有效的一种方法。例如，我们推荐用夏皮罗-威尔克方法（1965）来进行正态性检验。尽管在此之后，还出现一些新的方法，例如想法很新颖的熵方法（1976），但是其效果还不如前者，所以不予介绍。

这里亦介绍了我们及国内近几年来的一些研究成果。

名词术语以《电子工业技术词典》第三十四章可靠性（国防工业出版社出版，1976年）为基础，参考了一些国外

标准。

复杂系统的误差综合中有时遇到动态参数误差综合问题，有时要用蒙特-卡罗方法来进行非线性误差综合，由于不是普遍遇到且涉及数学知识较多，故扼要地列入附录，不详细展开。

§10 的综合评定方法的推导列为附录 3。

希望读者批评指正，使这本小册子能作为引玉之砖，为社会主义建设尽一点力量。

何国伟

1977 年 8 月

目　录

符 号 表

$A, B, \cdots$	事件
$\overline{A}$	事件A的对立事件
$E[X]$	随机变量X的期望值（均值）
$f(x)$	随机变量的概率密度函数
$f_X(x)$	随机变量X的概率密度函数
$F(x)$	随机变量的累积分布函数（分布函数）
$F_X(x)$	随机变量X的累积分布函数（分布函数）
$N(\mu, \sigma)$	均值为μ、标准偏差为σ的正态分布
$N(0,1)$	均值为0，标准偏差为1的正态分布，即标准正态分布
n	子样大小(在附录2中，n表状态向量的维数)
$P(A)$	事件A出现的概率
r	子样相关系数
R	子样的范围
R_n	大小为n的子样的极差
s^2	子样方差
T	$T=\sum_{i=1}^{n} x_i$，为子样个体值的总和；在§6中表示Grubbs统计量。(附录1中表示寿命)
$X, Y, \cdots$	随机变量
X_0	标称值，真值

x	子样特性值
x_1, x_2, $\cdots$	子样个体特性值
$x_{(1)}, x_{(2)}, \cdots$	把子样个体特性值按从小到大排列后的第 i 个值，顺序统计量
$\bar{x}$	子样均值
$\in$	属于
$\bar{\in}$	不属于
ε	误差
μ	随机变量的均值
μ_X	随机变量X的均值
σ^2	随机变量的方差
σ_X^2	随机变量X的方差
ρ	相关系数
$\chi^2(\nu)$	自由度为ν的 χ^2 分布

§1　数据整理方法

一、母体、子样、误差

我们所研究的对象的某特性值的全体称为母体（又称作总体）。

〔**例1.1**〕　按某产品（例如某种精密电阻）的生产技术文件规定生产这种产品。如果我们要考察这种产品的某一参数（例如电阻值），则每一个产品有一个参数值。所有可能得到的产品某参数值的全体（例如所有可能得到的电阻值）就是一个母体。

〔**例1.2**〕　在规定的使用条件下，用某一测量设备反复测量某一对象的某参数（例如长度），则每一次测量得到一个测定值。所有可能得到的测定值的全体（例如用这测量设备对此对象的所有可能测得的长度）就是一个母体。

对象的一个单位体（例如一支三极管）或单位量（例如一公斤橡胶）称为个体。符合一定条件的若干个体的总体称为一“批”。一批中所包括的个体数目叫做批量大小。这里要注意的是：批要“符合一定条件”。有些单位把原材料差异很大、生产条件差异很大的若干组产品合起来算一“批”，这样做法是不符合批的定义的。

为一定目的取得的若干个个体的全体叫做子样[1]，子样所包括个体的数目叫做子样大小。

---

[1] 在本书中，“子样”的含义都是指“随机子样”，其意义见后述。

产品某参数的要求值 X_0 叫标称值。实际上按规定技术文件生产的产品的参数值为 X，定义

$$\varepsilon = X - X_0 \tag{1.1}$$

为产品某参数值的误差。按规定技术文件生产的产品参数值的误差构成一个母体。

〔例1.3〕 按规定的技术文件生产某种精密电阻，要求的电阻值为 2.000 欧。在产品中取一个，精密测定其阻值为 2.015 欧，则这一个产品的阻值误差为 0.015 欧。按规定的技术文件生产的这种精密电阻的所有可能的阻值误差全体构成一个母体。

测量对象的某一参数有一客观存在的值叫真值 X_0。例如测定某一物的长度，这物的长度的真值是客观存在的，只是除了少数例外场合，这真值实际上是只能逐步逼近，而是求不到的。

我们用某种测量设备在规定的使用条件下对此参数进行测量，得到的测定值为 X，定义

$$\varepsilon = X - X_0 \tag{1.1}$$

为用该测量设备测此参数的测量误差。按规定使用条件测定的测量误差构成一个母体。

根据上述的定义，按规定技术文件生产某种产品，所有可能得到的某参数值的全体构成一个母体；这即按技术文件规定生产无穷多产品的所有可能得到的某参数值的全体，实际上这是一个理想的概念。同样，用某种测量设备在规定使用条件下测定某一对象的某参数，所有可能得到的测定值的全体构成一个母体；即在规定条件下反复测定无穷多次，所有可能得到的测定值的全体，实际上也是一个理想的概念。

我们实际上只能生产有限个产品，并且逐个检测产品的参数值不一定现实。因为一则往往需要的工作量太大，二则有些参数（如寿命）的检测是破坏性的，所以只能抽样检测。同样我们只能用某一测量设备，对对象的某一参数进行不太多次的反复测定。所以实际上得到的是子样。但是只要产品的生产或测定是在严格的质量控制下进行的，则子样的性质就能在一定程度上反映母体的性质。但子样毕竟不同于母体，因此子样的性质毕竟只能在一定程度上反映母体的性质。这一点差异是非常重要的。产品设计对参数的要求是对参数母体的要求，而实际上评定验证是否达到设计指标却只能根据子样。部分工程技术人员在这一个概念上是混淆的。

二、子样数据记录表

为了便于误差分析，子样数据记录表一般应包括如下内容：

被测产品的名称、规格，被测产品特性，测试单位，测试所根据的技术文件编号，技术文件所规定的子样大小及选取子样的方法、容许公差的上下限值，测试设备的名称、编号，测试日期、时间，测试时的温度、湿度及其它必要的环境数据，测试或检验人员的姓名。

测试结果应记录在如下的典型表格中。

〔**例1.4**〕 按规定的技术条件生产某种精密电阻，要求的阻值为 1.000（单位，欧），今按随机抽样规定（其意义见例 2.5）抽取 25 个电阻，在恒温条件下用精密设备测得阻值见表 1.1。

每一行数据的右侧记下该行数据的最大值与最小值，为今后进一步分析使用。每行十个数据是分析计算的比较适中

表1.1　测试记录表

样本号	测试结果(单位，欧)					最大测试值	最小测试值
1～10	1.00141 1.00519	0.99247 0.98769	0.98663 1.00185	0.99005 1.00407	1.00439 1.00606	1.00606	0.98663
11～20	1.00156 0.98328	0.99375 1.01008	0.97372 0.98727	1.01057 0.98056	1.00446 1.01685	1.01685	0.97372
20～25	1.01558	0.99733	0.97883	0.98803	1.00567	1.01558	0.97883
						最大值	最小值
						1.01685	0.97372

的数目。

我们看到：这些数据不是一目了然的。所以需要作科学的整理和分析，从中提炼出所需的情报。

三、随机变量

在研究、设计、生产、使用中出现的一定现象、状态、试验测试结果叫做事件。在一定条件下，可能出现也可能不出现的事件叫做随机事件。随机事件的特点是：在这事件出现以前，我们不能完全确切地指出它会出现还是不会出现。在〔例1.4〕中，一个电阻的阻值在1.01000欧与1.02000欧之间是一个随机事件。在一定条件下必然发生的事件叫必然事件。它是随机事件的一个特殊情况。例如“上述精密电阻的阻值不是负的”，这是一个必然事件。在一定条件下不可能发生的事件叫做不可能事件。例如“精密电阻的阻值是负的”，这是不可能事件。

在研究、设计、生产、使用中，如果某一个量在一定条件下取某一值或取某一范围内的值是一个随机事件，则这样

的量叫随机变量，通常用英文大写字母表示。〔例1.4〕中，精密电阻的阻值就是一个随机变量。

设随机变量X的取值可排列为x_1，x_2，…，（这些x_i可以是有限多个，也可以是无限多个），则X叫离散型随机变量。

〔**例1.5**〕 对一批产品进行抽样检查，抽样个数为n。则只要产品的不合格率不是零，就有可能抽得不合格品。抽得的不合格品数X只能取0，1，2，…等值，所以X是一个离散型随机变量。

〔**例1.6**〕 在水流量标准装置中，换向器时间的测定存在一种固定大小的误差，例如为30毫秒。但方向不定，当换向器向右时为正值，当换向器向左时为负值。在实际测量时，换向器向右、向左都有可能。这种误差X可以取两个值：即+30毫秒及-30毫秒中的任一个，所以是一种离散型随机变量。

设随机变量X可取坐标轴上某一区间内的任一数值，则X叫连续型随机变量。

〔例1.4〕中的精密电阻的阻值X就是一种连续型随机变量。

四、频数及频数表

在多次实践中，某随机事件出现的次数叫做该事件的频数。在〔例1.4〕中，阻值在1.01000～1.02000欧之间的有4个，即“阻值在1.01000～1.02000欧之间”这一随机事件的频数为4。

把随机事件的特性值存在的范围分成若干个子区间。统计随机事件的特性值落在各子区间的频数，排列成的表叫频数表，也叫频数分布表。它是在分析随机事件规律性时对数

据处理的第一步。频数表的作法如下(以〔例1.4〕为例):

〔第一步〕 找出数据的最大值与最小值。本例中,根据数据表的最大值及最小值,得最大值为1.01685,最小值为0.97372。

〔第二步〕 选择包含数据最大值与最小值的区间。当子样大小较多时,通常等分成10～20个子区间;当子样大小少于50时,等分成5～6个子区间。本例中,选包含最大值与最小值的区间为〔0.96500～1.02500〕。今数据有25个,等分成6个子区间。

〔第三步〕 确定各子区间的边界值。为了避免某具体数据正好是子区间的边界值时,不好分辨它落入那一个子区间,故把边界值加上最小测试单位的一半。本例中,最小测试单位为0.00001欧,故取子区间边界值为:0.965005,0.975005,0.985005, 0.995005, 1.005005,1.015005, 1.025005。

〔第四步〕 确定各子区间的中心值 t_i。第 i 个子区间的中心值 t_i 即第 i 个子区间左右两边界值 l_i 及 r_i 的平均值(去掉所加的最小测试单位的 $\frac{1}{2}$)。

〔第五步〕 统计数据落入各区间的频数。记法用"正"字法,记在如下的频数表中。子区间的号码是从小到大排列的。在频数列之下,计算 Σn_i (n_i 是数据落入第 i 个子区间的频数)记在末一行,它应等于总数据数 n。

本例的频数表如表1.2。

表1.2 〔例1.4〕的频数表

子区间号码	子区间边界值		子区间中心值 t_i	频数记录	频数 n_i	累积频数 N_i	相对频数 $f_i(\%)$	累积相对频数 $F_i(\%)$
	左端点 l_i	右端点 r_i						
1	0.965005	0.975005	0.97000	一	1	1	4.00	4.00
2	0.975005	0.985005	0.98000	𠃊	4	5	16.00	20.00
3	0.985005	0.995005	0.99000	正	5	10	20.00	40.00
4	0.995005	1.005005	1.00000	正下	8	18	32.00	72.00
5	1.005005	1.015005	1.01000	正	5	23	20.00	92.00
6	1.015005	1.025005	1.02000	丅	2	25	8.00	100.00
					$n=\sum n_i$		$\sum f_i$	
					25		100	

五、累积频数及累积频数表

随机事件的某特性值比某一值小的频数，称为随机事件的某特性值对该值的累积频数。

对于频数表中第 i 个子区间的右端边界值 r_i 而言，数据的累积频数 N_i 即数据落在此区间及以前诸区间的频数和

$$N_i=\sum_{k=1}^{i} n_i \tag{1.2}$$

〔例1.4〕的累积频数如表 1.2 所示。

六、相对频数及累积相对频数

随机事件的特性值落入某区间的频数相对于总特性值个数 n 的比值叫做随机事件的特性值落入某区间的相对频数，有人称它为频率。设随机事件的特性值落入第 i 个区间的频数为 n_i，总特性值的个数为 $n=\sum\limits_{i} n_i$，则随机事件的特性值落入第 i 个区间的相对频数 f_i 为

$$f_i=\frac{n_i}{n} \tag{1.3}$$

通常用“%”表示。

〔例1.4〕的相对频数如表 1.2 所示。

随机事件的特性值比某一值小的相对频数，称为随机事件特性值对该值的累积相对频数。对于频数表中第 i 个子区间的右端边界值 r_i 而言，随机事件特性值的累积相对频数 F_i 即数据落在此区间及以前诸子区间的相对频数和

$$F_i=\sum_{k=1}^{i} f_i \tag{1.4}$$

〔例1.4〕的累积相对频数如表 1.2 所示。

七、频数直方图及相对频数直方图

为了能更清晰地看出数据波动规律，把频数表的数据按如下步骤构成直方图，叫频数直方图。

〔第一步〕 以横坐标记数据值。

〔第二步〕 以纵坐标记频数。

〔第三步〕 在横坐标轴上，按各子区间的边界值划分边界线。

〔第四步〕 记下各子区间的中心值 t_i。

〔第五步〕 以各子区间为底边，在其上作出高度为相应频数的直方柱。

〔例1.4〕的相应频数直方图如图 1.1 。

在某区间上，长方柱的总面积就表示数据落入该区间的频数。

为了能更清晰地看出相对频数波动的规律，把相对频数表的相对频数数据构成直方图，叫相对频数直方图。相对频数直方图的构成方法与频数直方图的构成方法完全类似，所不同的是以纵坐标记相对频数。如果已作出频数直方图，只