检定与测试中的数据处理

中国计量科学研究院情报室

检定与测试中的数据处理

中国计量科学研究院情报室

说 明

在国内外一派大好形势下，我国计量事业得到飞跃的发展，大批新生力量充实到各个计量部门，普及计量技术成为当前的迫切需要。本书就是根据《计量工作》杂志不少读者的这一要求，由我室黄福芸同志编写的。原先准备在《计量工作》上逐期连载，考虑到本书字数较多，刊物的篇幅有限，因此改用小册子的形式另行出版。由于我们没有编写这类材料的经验，缺点错误在所难免，希望读者指正。

本书在编写过程中，得到济南第二机床厂计量室和北京汽车制造厂计量室的许多宝贵意见，在此谨致谢意。

中国计量科学研究院　情报室

一九七六年五月

目　　录

第一章　前　　言

检定和测试是计量工作中最普遍和最大量的一项日常工作，每一次检定或测试所取得的一系列数据都要经过数学处理，才能得出最后的结果，因此数据处理是检定与测试工作的一项重要内容。

当我们在做一项精度较高的检定或测试时，常常发现这一次测量和那一次测量所测得的数值不完全相同；用这一台仪器和用那一台仪测得的结果不完全相同；用这个标准和用那个标准检定出来的结果不完全相同；在这个条件下和在那个条件下测得的结果不完全相同；甚至这个人和那个人测量的结果也不完全相同等等。为什么会有这个现象呢？这是因为在每一次的测量中都存在着误差，而这些误差的大小又不一定完全相等，所以虽然被测的量是同一个，而测得的结果就可能各不相同。那么怎样确定该量的大小呢？我们在多大程度上可以信赖这个数值呢？这便是我们讨论检定与测试数据处理的目的。

因为任何测量中都存在着误差，所以计量工作者的任务之一就是努力查明各次测量中误差的大小，并尽可能地把它们限制在我们所希望的微小范围之内，使检定或测试的结果能满足对计量工作所提出的要求。正确的数据处理方法就是实现这一任务的一个很好的手段。

熟练地掌握数据处理的原则，对于正确选用测量方法，鉴定所建立的计量基准或标准，检验产品及零、部件是否符合图纸要求等等都是很有用处的。

本文将从实用角度首先介绍如何从一系列的测得值中求出作为被测量测量结果的算术平均值L、测量列的均方误差σ、

和算术平均值L的均方误差σ_L；然后从测量结果的算术平均值L中修正所能修正的系统误差，并最后确定最终结果M的有效位数和计算最终结果的总误差σ_M，从而完成数据处理的全过程，取得最终结果为

$$M \pm \sigma_M。$$

然后简略介绍如何运用在数据处理过程中所述及的有关理论，解决实际检测工作中常常遇到的一些具体误差问题。

由于篇幅关系，本文对一些数学公式的推导过程作了较多节删，但并不影响对问题的理解，如果读者需要进一步了解这些推导，可以从任何一本研究误差理论的书籍中查到。

第二章 测量及其误差

将待测的量与作为测量单位的标准量进行比较就是测量。计量工作中的检定及测试都属于测量工作的范围。

由于受测量条件和测量方法的限制，我们所得到的测得值l与被测量的真值$L_{真}$之间存在着一个差值，这就是测量误差δ，即

$$\delta = l - L_{真} \tag{1}$$

因为测得值l可能大于、也可能小于被测量的真值$L_{真}$，所以上式可改写为

$$L_{真} = l \pm \delta \tag{2}$$

可见，测量的精确程度由测量误差的绝对值大小来决定；测量误差绝对值愈大，测量的精确度便愈低。为了提高测量的精确度，就必须采取措施，减小测量误差的产生。

事实上被测量的真值是无法得知的，这是因为我们无法消除全部的测量误差，而且也不能完全确定误差的确切数值，我

们只能确定误差数值大小的一定范围。所以通过测量，我们得到的只是近似于被测量真值的近似值L。

从一系列测得值中计算被测量近似值L的方法，根据测量类别的不同而不同，为了了解被测量近似值的计算方法，首先要知道测量分类的情况。

第一节 测量的类别

在测量工作中，按测量条件可以分为等精度测量和不等精度测量；按测量方法可以分为直接测量、间接测量和组合测量；按测得结果的情况，可以分为绝对测量和相对测量；按被测量的表征参数可以分为单项测量和综合测量等等。

在测量条件不变的情况下所进行的一系列测量，由于它们相互间的测量精度是相同的，这种测量称为等精度测量。例如由同一个人在同一仪器上对同一被测量进行的多次测量，便是等精度测量。这种测量由于简单易行，能在一定程度上和在一定范围内提高测量结果精度，所以在计量工作中应用最为普遍。

如果测量工作是在几种不同的测量条件下进行，或者采用几种不同的测量方法（包括不同的测量仪器）、不同的测量次数或由不同的测量者所进行的一系列测量所组成，由于这些测量的精度各不相同，因此称之为不等精度测量。在计量工作中，当有必要进行更为精密的检定或测试时，常常采用这一方法，因为它有可能减小一些误差，能较确切地得知测量误差的具体范围，因而能提高测量精度。当然，不等精度测量的工作量比起等精度测量的工作量（包括数据处理的工作量）要大得多。

测量结果可直接从测得值得出或计算出来的测量叫做直接

测量。例如用卡尺测量零件的尺寸；用温度计测量温度；用秒表测量时间等等。

通过对几个与被测量有一定函数（方程式）关系的量进行直接测量，将直接测量得到的测得值代入上述函数关系式计算出被测量大小的测量方法，叫做间接测量。例如长度计量中用正弦尺测量锥体角度φ（如图1），当被测锥体1的上表面平行于平板4的表面时，利用正弦定律，从 $\mathrm{Sin}\varphi=\frac{h}{l}$ 关系

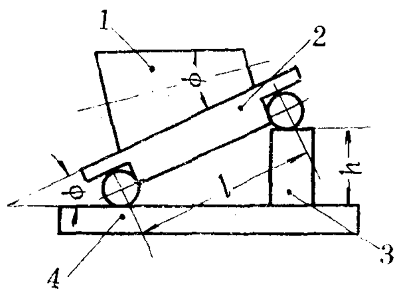

图 1

1—被测锥体，2—正弦尺，3—块规，4—平板

中，根据固定的l值和测量中所确定的h值而得出φ角的大小；又如圆柱形物体的密度D，可以在测出其质量m、直径d和高度h后，按关系式$D=\frac{m}{0.25\pi d^2h}$求得。

如果被测量的数目在一个以上，在对与这些被测量有一组已知函数关系的另一些量进行了一系列的测量以后，从这组方程式的解中求出各个被测量来，这样的测量就是组合测量。长度计量中封闭圆的分度测量和线纹尺的测量都是典型的例子。

直接测量、间接测量和组合测量都是计量工作中经常采用的方法。

利用作为标准单位的标准量比较被测量从而得出被测量大小的测量，称为绝对测量；测量结果仅只指出被测量与标准量两者的差值或比值关系时，这种测量叫做相对测量或比较测量。例如长度计量中用测微仪和块规测量圆柱直径，先用块规调好测微仪的零点，然后放上被测圆柱，所得到的结果为被测直径相对于标准长度的偏差。绝对测量和比较测量也是计量工作中经常采用的测量方法。

仅对被测量某一参数进行的测量称为单项测量；若某一被测对象不能用某一单项参数衡量而只能用其综合参数作总的衡量时，这种综合性的测量称为综合测量。

由于测量类别的不同，数据处理的方法也不相同，例如等精度测量、不等精度测量、直接测量、间接测量和组合测量就各有其不同的处理方法，因此在进行处理之前，必须搞清楚该测量所属的类别。

第二节　误差的分类及其产生的原因

在测量过程中所产生的误差，按其性质可以分为三大类，即：偶然误差、粗大误差和系统误差。

一、偶然误差

由许多暂时尚未被掌握的规律或一时不便于控制的微小因素所造成的误差，称为偶然误差。这种误差数值的大小和符号正负的出现，具有偶然的性质，不能事先知道，因而也就无法从测量过程中予以修正或把它们加以消除，但是根据偶然误差本身客观存在的规律，即它在多次重复测量中出现的规律，在一定的条件下，我们可以用增加测量次数的方法加以控制，从而减小其对测量结果的影响。

之所以要密切重视偶然误差，是因为它的存在决定了测量

工作的“精密度”，查明该测量中偶然误差的大小，就能确定该测量结果的可靠程度。

偶然误差通常由下面三个方面的原因造成：

1．由测量工具带来的误差

由于所使用的测量仪器结构不完善或零、部件制造不精密，会给测量结果带来偶然误差。例如由于轴与轴承间存在间隙，因而润滑油在一定的条件下所形成的油膜不均匀现象会给圆分度的测量带来偶然误差。

2．由测量方法带来的误差

测量方法的不完善或不合理会给测量带来偶然误差，例如使用目镜瞄准被测物，因瞄准的不精密所造成的读数误差等等。

3．由测量条件带来的误差

由于测量环境或条件不稳定等无法控制的因素（如温度的微小波动、湿度与气压的微量变化等等）会给测量带来偶然误差。例如在恒温室标准温度20℃的条件下，由于允许温度在±0.5℃范围内有微小的波动，便会给测量带来一定的偶然误差。

二、粗大误差

由于测量者在测量或计算时的粗心大意所造成的数值很大的误差，叫做粗大误差。是否存在粗大误差是衡量该测量结果是否“合格”“可用”的标志。没有从测得值中剔除粗大误差所取得的测量结果是不能采用的，因为这样的测量结果会导致错误的结论。因此，计量工作者必须以“三老”“四严”的科学精神，十分严肃认真地对待测量工作，力求不产生粗大误差。此外，计量工作者还必须学会从测得值中及时发现和剔除因疏忽而产生的粗大误差。

三、系统误差

在一定条件下，误差数值的大小和正负号或者固定不变，或者按一定规律变化的误差就是系统误差。这种误差可以通过实验或分析的方法，查明其变化的规律及其产生的原因，并在确定其数值后，可以从测量结果中予以修正，或在新的一次测量前，采用一定措施，改善测量条件或改进测量方法，从而使之减小或排除，但是系统误差不能依靠增加测量次数的办法使之减小或消除。

系统误差的存在决定了测量的“准确”程度，因为它的存在歪曲了测量结果的真实面目。

系统误差有以下几个来源：

1．工具误差

它由于所使用的测量工具结构上不完善或零部件制造时的缺陷与偏差所造成。例如块规两测量面的不平行性、微分螺丝付的死程、温度计分度的不均匀、天平两臂长的不等以及度盘的偏心等等。

2．调整误差

它由于测量前未能将仪器或待测件安装在正确位置（或状态）所造成。例如使用未经校准零位的千分尺测量零件，使用零点调整不准的电气仪表作检测工作等等。

3．习惯误差

它是由于测量者的习惯所造成的系统误差。例如用肉眼在刻度上估读时习惯地偏向一个方向；或者凭听觉鉴别时，在时间判断上习惯地提前或落后等。

4．条件误差

它由于测量过程中条件的改变所造成。例如测量工作开始与结束时的一些条件按一定规律发生变化（如温度、气压、湿

度、气流、振动等）后带来的系统误差。

5．方法误差

它由于所采用的测量方法或数学处理方法不完善而产生。例如在长度测量中采用了不符合阿贝原则的测量方法，或者在计算时采用了近似计算方法等等。

第三章　数据处理的依据

检定或测试中的数据处理原则基于偶然误差的正态分布理论。这种理论来源于科学实验的具体实践。当对同一量值进行大量的等精度重复测量得出一系列测得值（常称为测量列），设测量列中不含有能影响测量结果的系统误差和粗大误差，如果我们将从该测量列中求得的各个偶然误差 δ 和同一数值误差的数目 y 按图 2 的直角坐标描点，则各点的连线为一曲线，这一曲线称为偶然误差的正态分布曲线。*

偶然误差正态分布曲线告诉我们，在等精度测量条件下，

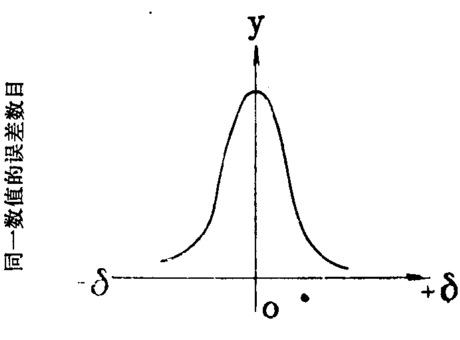

图　2

* 亦称为高斯正态分布曲线或高斯曲线

偶然误差有以下几个基本特点：

1．偶然误差中，数值为正的误差与数值为负的误差数目相等；

2．数值大小相等符号相反的偶然误差的数目相等；

3．绝对值小的误差比绝对值大的误差的数目多；

4．在一定的测量条件下，偶然误差的绝对值不会超过某一极限数值。

如所周知，在测量列中，如果数值小的偶然误差数目愈多，数值大的偶然误差数目愈少，则该测量的可靠性愈大。按偶然误差正态分布曲线来说，即曲线愈陡，偶然误差的极限值范围愈小，可靠性愈大；反之则可靠性愈低。由此可见，偶然误差正态分布曲线的陡与平，标志着该测量列的可靠性的大和小。

如果我们以 δ 代表偶然误差值、以 y 代表同一数值偶然误差的个数和可以由测量列中所有偶然误差 δ 算出的测量列均方误差 σ 来表示偶然误差正态分布曲线的数学式，则偶然误差正态分布曲线方程为：

$$y=f(\delta)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{\delta^2}{2\sigma^2}}$$

式中，e ——自然对数的底，e ＝2.71828；

π ——圆周率，$\pi=3.14159$。

由方程可知，当 σ 不同时，曲线的陡度也不相同。图3表示三种不同 σ 值的偶然误差正态分布曲线，其中误差为($^1/_2$) σ 的曲线最陡，偶然误差的极限（由＋ δ 到－ δ 之间的）范围最小，可靠性最大；而误差为 2σ 的那条曲线最平，偶然误差的极限范围最大，可靠性最差。可见 σ 愈大，测量的误差便愈大，因此测量列均方误差 σ 便成为表征测量精确度的重要参数之一。

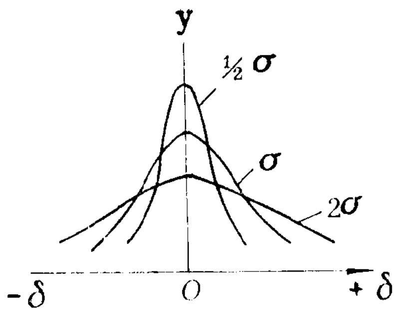

图　3

偶然误差正态分布曲线还表示误差出现的机会（通常称为概率或称为或然率），根据曲线的制备条件，我们可以知道，曲线下的全部面积相当于全部误差(正的和负的，大的和小的)出现的概率，即

$$\int_{-\infty}^{+\infty} y d\delta = 100\% = 1。$$

将上述偶然误差正态分布曲线方程的y值代入上式，得

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{\delta^2}{2\sigma^2}} d\delta = 1。$$

如果以新的变量$Z = \frac{\delta}{\sigma}$代入上式，则

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{Z^2}{2}} dZ = 1。$$

图4中曲线下阴影部分面积的量，相当于落在$0 \sim \delta_i$范围内误差出现的概率。此面积的积分式为：

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{0}^{\delta} e^{-\frac{\delta^2}{2\sigma^2}} d\delta$$

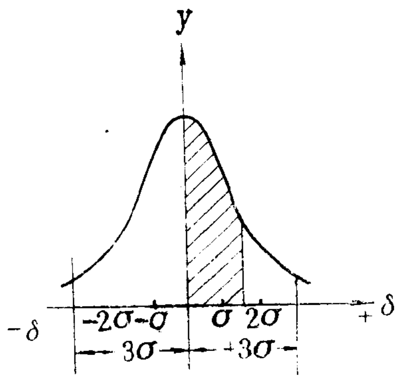

图 4

这一概率的函数式为

$$\phi\ (Z=\delta/\sigma)=\frac{1}{\sqrt{2\pi}}\int_{0}^{Z} e^{-\frac{Z^2}{2}}\,dZ。$$

这样，我们便可以知道，其绝对值不超过某偶然误差值的概率为上述面积的两倍，即，$2\phi(Z)$，或反过来说，超过 δ 的概率为 $1-2\phi(Z)$。

$\phi(Z)$值的大小可从 $\phi(Z)$积分值表中由 Z 值查得。

下表是假定在测量列中，绝对值超出 δ 的次数为 1 时，各种误差值 δ 与其应有测量次数 n 和超出或不超出的概率。

表 1 说明，随着 Z 的增大，超出 δ 的概率减小得很快，当 $Z=2$，即 $\delta=2\sigma$时，在 22 次测量中只有 1 次的误差绝对值超出 2σ 范围，而当 $Z=4$，即 $\delta=4\sigma$时，在 15625 次测量中才有 1 次误差的绝对值超过 4σ。由于一般测量中，测量的次数最多也只是几十次，因此可以认为，在任何情况下都不会出现绝对值大于 3σ 的偶然误差。通常把这最大可能出现的误差称为偶然误差的极限误差 Δ_{lim}，即

表1

Z	$\delta = Z\cdot\sigma$	不超过δ的概率 $2\phi(Z)$	超过δ的概率 $1-2\phi(Z)$	测量次数 n
0.67	0.67σ	0.4972	0.5028	2
1	1σ	0.6826	0.3174	3
2	2σ	0.9544	0.0456	22
3	3σ	0.9973	0.0027	370
4	4σ	0.9999	0.0001	15625

$\Delta_{lim} = \pm 3\sigma$。

综合以上所述，偶然误差正态分布理论不仅可用来确定均方误差，而且还可用来作为检验测量列中是否存在有粗大误差或系统误差的判断准则，因为粗大误差和系统误差是不服从偶然误差正态分布规律的。

第四章　数据处理的内容及方法

在检定和测试工作中的数据处理，按处理的先后顺序，大致有以下一些内容：

1. 确定被测量的值及测量列的精度；
2. 检验及剔除粗大误差；
3. 发现及修正系统误差；
4. 确定被测量的测量结果。
5. 确定被测量的最终结果。

第一节　确定被测量的值及测量列的精度

在经过一系列测量以后，计量人员的任务就是从所得到的测得值中确定被测量的值及测量列精度的大小，步骤是：

一、确定被测量的值L

假定我们对同一量进行一系列的等精度直接测量，得到一组测得值 l_1、l_2、$\cdots l_n$。设该量的真值为L真，各偶然误差（即各测得值与真值之差）为δ_1，δ_2，…，δ_n，则有：

$$L_{真}=l_1-\delta_1$$

$$L_{真}=l_2-\delta_2$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$L_{真}=l_n-\delta_n$$

将上式相加，得

$$n\cdot L_{真}=\Sigma l-\Sigma\delta$$

$$L_{真}=\frac{\Sigma l}{n}-\frac{\Sigma l}{n}$$

根据偶然误差正态分布理论当测量次数 n ⟶∞时，

$$\frac{\Sigma l}{n}=0$$

故上式改写为：

$$L_{真}=\frac{\Sigma\delta}{n}=\frac{l_1+l_2+\cdots+l_n}{n}。$$

上式告诉我们，当测量次数无限多时，被测量的算术平均值即为该量的真值。但是实际上我们不可能进行无穷多次的测量，所以也就不能得到真值，然而作为有限次数的测量，算术平均值则是最接近于真值的，因此以算术平均值* 作为最接近被测量真值的近似值L是最合理的，即

* 有的书籍称之为最或然值