

# 中华人民共和国国家标准

GB/T 6379.6—2009/ISO 5725-6:1994

---

## 测量方法与结果的准确度 (正确度与精密度) 第6部分:准确度值的实际应用

Accuracy (trueness and precision) of measurement methods and results—  
Part 6: Use in practice of accuracy values

(ISO 5725-6:1994, IDT)

2009-03-13 发布

2009-09-01 实施

---

中华人民共和国国家质量监督检验检疫总局 发布  
中国国家标准化管理委员会

## 目 次

前言 .....	I
引言 .....	II
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	2
4 限的确定 .....	2
4.1 重复性限和再现性限 .....	2
4.2 基于超过两个值的比较 .....	2
5 检查测试结果可接收性的方法及确定最终报告结果 .....	3
5.1 总则 .....	3
5.2 在重复性条件下所得测试结果可接收性的检查方法 .....	4
5.3 在再现性条件下所得测试结果可接收性的检查方法 .....	9
6 实验室内检查测试结果稳定性的方法 .....	10
6.1 背景 .....	10
6.2 检查稳定性的方法 .....	11
7 重复性标准差和再现性标准差在实验室评定中的应用 .....	22
7.1 评定方法 .....	22
7.2 由未经评定的实验室使用的测量方法的评估 .....	23
7.3 对已认可实验室的再评定 .....	25
8 与可替代测量方法的比较 .....	29
8.1 考虑可替代测量方法的原因 .....	29
8.2 比较测量方法的目的 .....	29
8.3 方法 B:作为候选的可替代标准方法(“标准化试验”未确立) .....	29
8.4 准确度试验 .....	30
8.5 方法 B 作为候选的常规方法 .....	35
附录 A (规范性附录) GB/T 6379 所用的符号与缩略语 .....	37

## 前 言

GB/T 6379《测量方法与结果的准确度(正确度与精密度)》分为以下六个部分,其结构及对应的国际标准为:

- 第 1 部分:总则与定义(ISO 5725-1:1994, IDT);
- 第 2 部分:确定标准测量方法的重复性和再现性的基本方法(ISO 5725-2:1994 IDT);
- 第 3 部分:标准测量方法精密度的中间度量(ISO 5725-3:1994, IDT);
- 第 4 部分:确定标准测量方法正确度的基本方法(ISO 5725-4:1994, IDT);
- 第 5 部分:确定标准测量方法精密度的可替代方法(ISO 5725-5:1998, IDT);
- 第 6 部分:准确度值的实际应用(ISO 5725-6:1994, IDT)。

本部分为 GB/T 6379 的第 6 部分。

本部分等同采用国际标准 ISO 5725-6:1994《测量方法与结果的准确度(正确度与精密度) 第 6 部分:准确度值的实际应用》及 ISO 于 2001-10-15 发布的对 1994 版 ISO 5725-6 的技术修改单。对 ISO 5725-6:1994 的错误作了如下更正:

- 根据计算结果,重新绘制了图 10;
- 原表 6 中  $x_1$  合计 16.84 一项计算错误,改为 16.74。

GB/T 6379 第 1 部分至第 6 部分作为一个整体代替 GB/T 6379—1986 及 GB/T 11792—1989。标准中将原精密度概念加以扩展,增加了正确度概念,统称为准确度;除重复性条件和再现性条件外,增加了中间精密度条件。

本部分的内容部分代替 GB/T 6379—1986 和 GB/T 11792—1989。

本部分的附录 A 为规范性附录。

本部分由广东出入境检验检疫局提出。

本部分由全国统计方法应用标准化技术委员会归口。

本部分起草单位:广东出入境检验检疫局、中国科学院数学与系统科学研究院、中国标准化研究院。

本部分主要起草人:李成明、冯士雍、张震坤、姜健、周崎、丁文兴、宋武元、于振凡、李政军、肖惠、刘建斌、陈玉忠。

本部分于 2009 年首次发布。

## 引 言

0.1 GB/T 6379 用两个术语“正确度”与“精密度”来描述一种测量方法的准确度。正确度指大量测试结果的(算术)平均值与真值或接受参照值之间的一致程度;而精密度指测试结果之间的一致程度。

0.2 考虑精密度的原因主要是因为假定在相同的条件下对同一或认为是同一的物料(物质/材料)进行测试,一般不会得到相同的结果。这主要是因为在每个测量程序中不可避免的会出现随机误差,而那些影响测量结果的因素并不能完全被控制。在对测量数据进行实际解释过程中,必须考虑这种变异。例如,测试结果与规定值之间的差可能在不可避免的随机误差范围内,在此情形,测试值与规定值之间的真实偏差是不能确定的。同样,比较两批物料的测试结果时,如果它们之间的差异是来自测量程序中的内在变化,则不能揭示两批物料间的本质差别。

0.3 GB/T 6379 的第1至第5部分讨论了精密度(用重复性标准差与再现性标准差表示)与正确度(用偏倚的诸分量表示)评定的背景,并给出了利用一种标准测量方法所得的测试结果对精密度和正确度进行评定的一些方法。然而,如果评定结果不能用于实际,这些评定方法就失去了意义。

0.4 测量方法的准确度一旦确定,GB/T 6379 本部分将利用这些知识为商业与贸易提供便利,同时用来监督并改进实验室的操作。

# 测量方法与结果的准确度 (正确度与精密度) 第6部分:准确度值的实际应用

## 1 范围

1.1 GB/T 6379 本部分的目的是说明能应用准确性数据的各种不同的实际情形:

- a) 给出计算重复性限、再现性限以及其他限的标准方法,所计算的这些限值将被用于检查使用标准测量方法所获得的测试结果;
- b) 提出对重复性或再现性条件下所获得的测试结果进行可接收性检查的方法;
- c) 描述如何评定一个实验室在某个时期内测试结果的稳定性,从而对实验室内操作提出“质量控制”方法;
- d) 描述如何评定某个特定的实验室是否具有正确使用给定的标准测量方法的能力;
- e) 描述如何比较可替代的测量方法。

1.2 本部分所涉及的测量方法特指对连续量进行测量,并且每次只取一个测量值作为测试结果的测量方法,尽管这个值可能是一组观测值的计算结果。

1.3 本部分假定测量方法的正确度及精密度的估计值已按 GB/T 6379 的第1部分至第5部分的方法获得。

1.4 有关使用场合的任何额外信息应在每次特定使用开始时给出。

## 2 规范性引用文件

下列文件中的条款通过 GB/T 6379 的本部分的引用而成为本部分的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版本均不适用于本部分,然而,鼓励根据本部分达成协议各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本部分。

GB/T 3358.1—1993 统计学术语 第一部分 一般统计术语

GB/T 4091—2001 常规控制图(ISO 8258:1991, IDT)

GB/T 6379.1—2004 测量方法与结果的准确度(正确度与精密度) 第1部分:总则与定义(ISO 5725-1:1994, IDT)

GB/T 6379.2—2004 测量方法与结果的准确度(正确度与精密度) 第2部分:确定标准测量方法的重复性和再现性的基本方法(ISO 5725-2:1994, IDT)

GB/T 6379.4—2006 测量方法与结果的准确度(正确度与精密度) 第4部分:确定标准测量方法正确度的基本方法(ISO 5725-4:1994, IDT)

GB/T 6379.5—2006 测量方法与结果的准确度(正确度与精密度) 第5部分:确定标准测量方法精密度的可替代方法(ISO 5725-5:1998, IDT)

GB/T 27025 检测和校准实验室能力的通用要求(GB/T 27025—2008, ISO/IEC 17025:2005, IDT)

ISO 3534-1:1993 统计学 词汇和符号 第1部分:概率和一般统计术语

ISO 5725-3:1994 测量方法与结果的准确度(正确度与精密度) 第3部分:标准测量方法精密度的中间度量

ISO 指南 33: 1989 有证标准物质(标准物质)的使用

ISO 指南 35: 1989 标准物质(标准物质)的定值 总则和统计原理

### 3 术语和定义

GB/T 3358.1 和 GB/T 6379.1 中给出的术语和定义在 GB/T 6379 本部分中仍适用。

GB/T 6379 使用的符号由附录 A 给出。

### 4 限的确定

#### 4.1 重复性限和再现性限

4.1.1 GB/T 6379.2 主要研究在重复性条件和再现性条件下所进行测量的各种标准差的估计。然而在通常的实验室工作中往往要求对两个(或多个)测试结果观测值的差进行检查,为此需确定一些类似临界差之类的度量,而不仅是标准差。

4.1.2 如果一个估计量是  $n$  个独立估计量的和或差,每个估计量的标准差均为  $\sigma$ ,则和或差的标准差为  $\sigma\sqrt{n}$ 。再现性限  $R$  和重复性限  $r$  均为两个测试结果之间的差,因而相应的标准差为  $\sigma\sqrt{2}$ 。在常规的统计工作中,为了检查两个测试结果之间的差异,往往用这个标准差的  $f$  倍作为临界差。临界差系数  $f$  的值依赖于与临界差相应的概率水平及测量结果所服从的分布。对重复性限和再现性限,概率水平规定为 95%。在 GB/T 6379 全部分析中,假定基本分布是近似正态的。对正态分布,95%的概率水平下, $f=1.96$ ,因此, $f\sqrt{2}=2.77$ 。GB/T 6379 本部分的目标是给出一些简单的经验法则供非统计专家使用,将  $f\sqrt{2}$  修约为 2.8。

4.1.3 如前所述,当标准差的真值未知时,估计精密度的结果给出了标准差的估计值。在统计实践中,标准差的估计用  $s$ ,而不用  $\sigma$  表示。按照 GB/T 6379.1 和 GB/T 6379.2 给出的程序,这些估计值是基于一定数量的测试结果,带给我们关于标准差的真值可能的最佳信息。因此,在以后的其他应用中,记  $s$  为基于相对有限次测试结果的标准差的估计值,以  $\sigma$  表示由完全精密度试验得到的值,并以该值作为标准差的真值,与其他估计值进行比较。

4.1.4 由 4.1.1 至 4.1.3 知,在对重复性条件或再现性条件下得到的两个单一测试结果进行检验时,应与重复性限  $r=2.8\sigma_r$  或再现性限  $R=2.8\sigma_R$  进行比较。

#### 4.2 基于超过两个值的比较

##### 4.2.1 一个实验室内两组测试结果的比较

在一个实验室内,如果在重复性条件下进行两组测量,第一组测试结果数为  $n_1$ ,其算术平均值为  $\bar{y}_1$ ;第二组测试结果数为  $n_2$ ,其算术平均值为  $\bar{y}_2$ ,则  $(\bar{y}_1 - \bar{y}_2)$  的标准差为:

$$\sigma = \sqrt{\sigma_r^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

在 95% 的概率水平下,  $|\bar{y}_1 - \bar{y}_2|$  的临界差为:

$$CD_{0.95} = 2.8\sigma_r \sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}$$

注 1: 如果  $n_1 = n_2 = 1$ ,上述临界差化简为  $r = 2.8\sigma_r$ 。

##### 4.2.2 两个实验室内两组测量结果的比较

如果在重复性条件下,第一个实验室测试结果数为  $n_1$ ,其算术平均值为  $\bar{y}_1$ ;第二个实验室测试结果数为  $n_2$ ,其算术平均值为  $\bar{y}_2$ ;则  $(\bar{y}_1 - \bar{y}_2)$  的标准差为:

$$\begin{aligned} \sigma &= \sqrt{\sigma_L^2 + \frac{1}{n_1}\sigma_r^2 + \sigma_L^2 + \frac{1}{n_2}\sigma_r^2} \\ &= \sqrt{2\sigma_L^2 + \sigma_r^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \end{aligned}$$

$$= \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2}\right)}$$

在 95% 的概率水平下,  $|\bar{y}_1 - \bar{y}_2|$  的临界差为:

$$CD_{0.95} = \sqrt{(2.8\sigma_R)^2 - (2.8\sigma_r)^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2}\right)}$$

注 2: 如果  $n_1 = n_2 = 1$ , 则上述临界差化简为  $R = 2.8\sigma_R$ 。

#### 4.2.3 一个实验室的测试结果与参照值的比较

如果在重复性条件下, 一个实验室得到了  $n$  个测试结果, 其算术平均值为  $\bar{y}$ , 那么就应将它与某个确定的参照值  $\mu_0$  进行比较, 在偏倚的实验室分量尚未确定的情况下,  $(\bar{y} - \mu_0)$  的标准差为:

$$\begin{aligned} \sigma &= \sqrt{\sigma_L^2 + \frac{1}{n}\sigma_r^2} \\ &= \frac{1}{\sqrt{2}} \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 \left(1 - \frac{1}{n}\right)} \\ &= \frac{1}{\sqrt{2}} \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 \left(\frac{n-1}{n}\right)} \end{aligned}$$

在 95% 的概率水平下,  $|\bar{y} - \mu_0|$  的临界差为:

$$CD_{0.95} = \frac{1}{\sqrt{2}} \sqrt{(2.8\sigma_R)^2 - (2.8\sigma_r)^2 \left(\frac{n-1}{n}\right)}$$

#### 4.2.4 多个实验室的测试结果与参照值的比较

如果有  $p$  个实验室, 分别在重复性条件下得到了  $n_i (i=1, 2, \dots, p)$  个测试结果, 每个实验室测试结果的算术平均值为  $\bar{y}_i$ , 所有实验室测试结果的总平均为:

$$\bar{y} = \frac{1}{p} \sum_{i=1}^p \bar{y}_i$$

将总平均与参照值  $\mu_0$  进行比较, 则  $(\bar{y} - \mu_0)$  的标准差为:

$$\begin{aligned} \sigma &= \sqrt{\frac{1}{p}\sigma_L^2 + \frac{1}{p^2}\sigma_r^2 \sum_{i=1}^p \frac{1}{n_i}} \\ &= \frac{1}{\sqrt{2p}} \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 + \frac{2\sigma_r^2}{p} \sum_{i=1}^p \frac{1}{n_i}} \\ &= \frac{1}{\sqrt{2p}} \sqrt{2(\sigma_L^2 + \sigma_r^2) - 2\sigma_r^2 \left(1 - \frac{1}{p} \sum_{i=1}^p \frac{1}{n_i}\right)} \end{aligned}$$

于是, 在 95% 的概率水平下,  $|\bar{y} - \mu_0|$  的临界差为:

$$CD_{0.95} = \frac{1}{\sqrt{2p}} \sqrt{(2.8\sigma_R)^2 - (2.8\sigma_r)^2 \left(1 - \frac{1}{p} \sum_{i=1}^p \frac{1}{n_i}\right)}$$

#### 4.2.5 报告比较结果

如果测试结果的绝对差超过以上条款给出的合理的限值, 应认为相应的绝对差可疑。此时, 用于计算此绝对差的所有测量都应被认为可疑, 需对它们作进一步审查。

### 5 检查测试结果可接收性的方法及确定最终报告结果

#### 5.1 总则

5.1.1 本章描述的检查方法仅适用于所使用的测量方法已标准化, 且重复性标准差  $\sigma_r$  和再现性标准差  $\sigma_R$  均已知的情形。因此, 当  $N$  个测试结果的极差超过第 4 章所给出的合理的限值时, 则认为  $N$  个测试

结果中的一个、两个、或者所有测试结果异常。建议从技术的角度查找发生异常的原因。然而,可能因为商业上的原因,有必要得到某个可接受值,此时应根据本章规定的方法对测试结果进行处理。

5.1.2 本章假定所有测试结果均在重复性条件或再现性条件下得到,计算中所用概率水平为 95%。若测试结果是在中间精密度条件(参见 ISO 5725-3:1994)下得到的,须用相应的中间精密度度量代替  $\sigma_r$ 。

5.1.3 在某些情况下,5.2 描述的方法将导致把测试结果的中位数作为最终报告结果,这些数据最好丢弃不用。

## 5.2 在重复性条件下所得测试结果可接收性的检查方法

注 3: 在 5.2.2.1 和 5.2.2.2 中,测量费用的低或高不仅指耗资多少,还包括测量是否复杂、执行是否困难以及是否费时。

### 5.2.1 单一测试结果

在商品检验中仅取一个测试结果的情形是不多见的。当仅有一个测试结果时,要立即对特定的重复性度量进行可接收性的统计检验是不可能的。若对测试结果的准确性有任何疑问,都应取得第二个测试结果。下面描述较常见的有两个测试结果的情形。

### 5.2.2 两个测试结果

两个测试结果都应在重复性条件下取得,测试结果之差的绝对值应与重复性限  $r=2.8\sigma_r$  比较。

#### 5.2.2.1 测试费用较低的情形

如果两个测试结果之差的绝对值不大于  $r$ ,可以接收这两个测试结果。最终报告结果为两测试结果的算术平均值。如果两测试结果之差的绝对值大于  $r$ ,实验室应再取两个测试结果。

此时,若 4 个测试结果的极差( $x_{\max} - x_{\min}$ )等于或小于  $n=4$  时概率水平为 95% 的临界极差  $CR_{0.95}(4)$ ,则取这 4 个测试结果的算术平均值作为最终报告结果。

临界极差按下式计算:

$$CR_{0.95}(n) = f(n)\sigma_r$$

式中  $f(n)$  称为临界极差系数。表 1 中列出了  $n$  从 2 至 40 以及从 45 到 100 的部分临界极差系数的值。

如果 4 个测试结果的极差大于重复性临界极差,则取 4 个测试结果的中位数作为最终报告结果。上述过程可以用图 1 的流程图表示。

#### 5.2.2.2 测试费用较高的情形

如果两个测试结果之差的绝对值不大于  $r$ ,这两个测试结果可以接收。最终报告结果为两个测试结果的算术平均值。如果两个测试结果之差的绝对值大于  $r$ ,实验室应再取一个测试结果。

此时,若 3 个测试结果的极差( $x_{\max} - x_{\min}$ )等于或小于  $n=3$  时概率水平为 95% 的临界极差  $CR_{0.95}(3)$ ,则取 3 个测试结果的平均值作为最终报告结果。

若 3 个测试结果的极差大于临界极差  $CR_{0.95}(3)$ ,则由下面两种情形之一来确定最终报告结果。

##### a) 不可能取得第 4 个测试结果的情形:

该实验室宜取中位数作为最终报告结果。

此过程可用图 2 的流程图表示。

##### b) 有可能取得第 4 个测试结果的情形:

该实验室应取第 4 个测试结果。如果 4 个测试结果的极差( $x_{\max} - x_{\min}$ )等于或小于临界极差  $CR_{0.95}(4)$ ,则取 4 个测试结果的算术平均值作为最终报告结果;如果极差大于  $CR_{0.95}(4)$ ,则取 4 个测试结果的中位数作为最终报告结果。

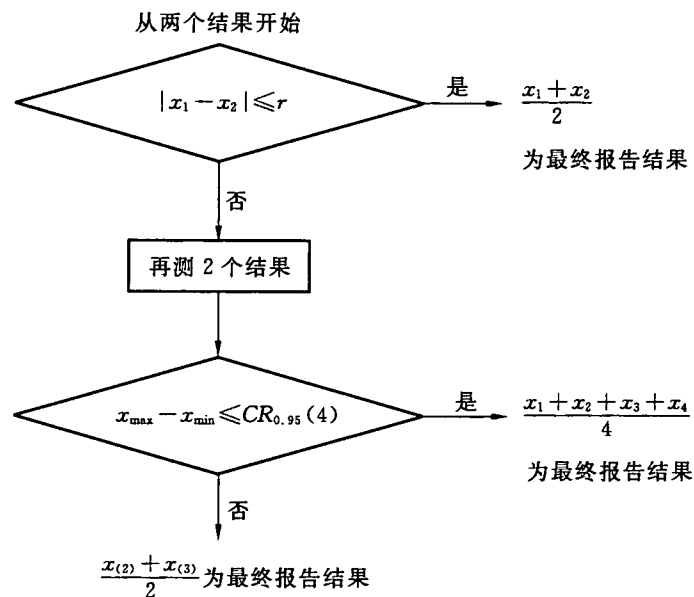
此过程可用图 3 的流程图表示。



表 1 临界极差系数  $f(n)$

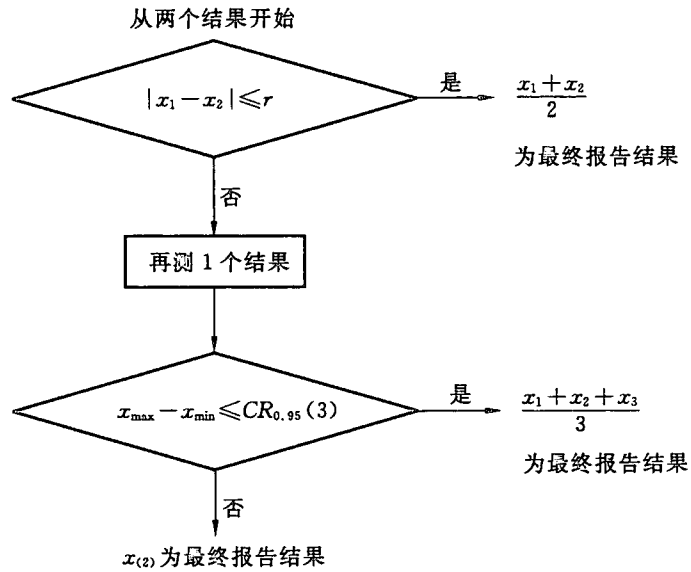
$n$	$f(n)$	$n$	$f(n)$
2	2.8	25	5.2
3	3.3	26	5.2
4	3.6	27	5.2
5	3.9	28	5.3
6	4.0	29	5.3
7	4.2	30	5.3
8	4.3	31	5.3
9	4.4	32	5.3
10	4.5	33	5.4
11	4.6	34	5.4
12	4.6	35	5.4
13	4.7	36	5.4
14	4.7	37	5.4
15	4.8	38	5.5
16	4.8	39	5.5
17	4.9	40	5.5
18	4.9	45	5.6
19	5.0	50	5.6
20	5.0	60	5.8
21	5.0	70	5.9
22	5.1	80	5.9
23	5.1	90	6.0
24	5.1	100	6.1

注：临界极差系数  $f(n)$  是  $(x_{\max} - x_{\min})/\sigma$  之分布的 95% 分位数，其中  $x_{\max}$  与  $x_{\min}$  分别为从标准差为  $\sigma$  的正态总体中抽取的样本量为  $n$  的样本的极大值与极小值。



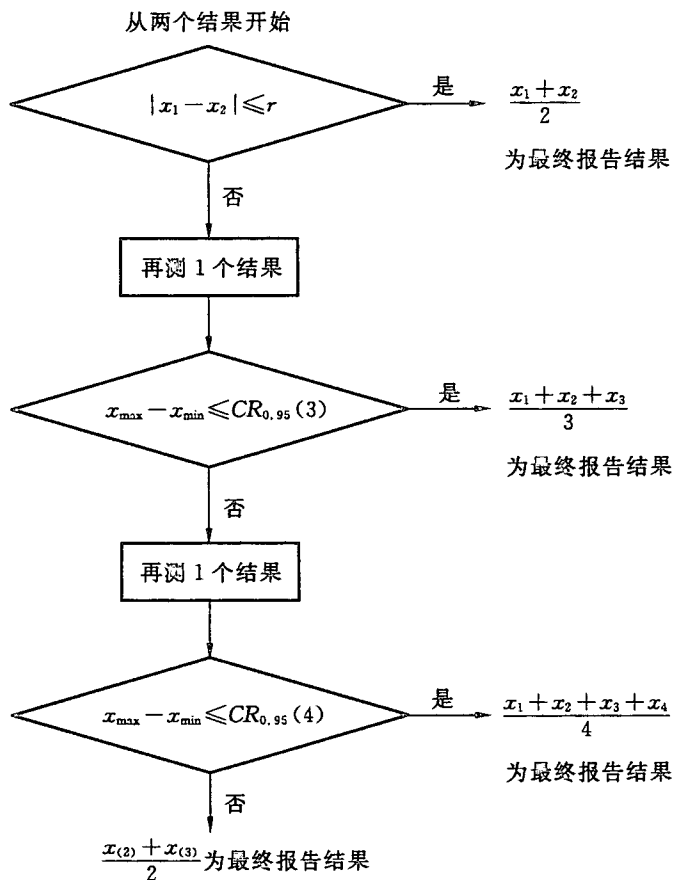
其中： $x_{(2)}$  为排序第二小的测试结果， $x_{(3)}$  为排序第三小的测试结果

图 1 在重复性条件下所得测试结果可接收性的检查方法  
(从两个测试结果开始且测试费用较低时：情形 5.2.2.1)



其中： $x_{(2)}$ 为排序第二小的测试结果

图 2 在重复性条件下所得测试结果可接收性的检查方法  
(从两个测试结果开始且测试费用较高时:情形 5.2.2.2a))



其中： $x_{(2)}$ 为排序第二小的测试结果， $x_{(3)}$ 为排序第三小的测试结果

图 3 在重复性条件下所得测试结果可接收性的检查方法  
(从两个测试结果开始且测试费用较高时:情形 5.2.2.2b))

### 5.2.3 从两个以上结果开始

实际工作中常有初始结果数大于2的情形。在重复性条件下， $n > 2$ 时确定最终报告结果的方法与 $n = 2$ 时的方法相类似。

将 $n$ 个结果的极差( $x_{\max} - x_{\min}$ )与按表1查得计算的临界极差 $CR_{0.95}(n)$ 比较:若极差等于或小于临界极差,则取 $n$ 个结果的算术平均值作为最终报告结果。

若极差大于临界极差,则由下面图4至图6表明的A,B,C三种情形之一来确定最终报告结果。

情形A与情形B分别对应于测试费用较低与较高的情形。情形C是可供选择的,推荐用于对 $n \geq 5$ 且测试费用较低,或 $n \geq 4$ 且测试费用较高的情形。

对测试费用较低的情形,情形A与情形C的区别在于:情形A需要追加 $n$ 次测量,而情形C只需要追加小于 $n/2$ 次的测量。按何种情形确定最终报告结果,最终依赖于 $n$ 的大小以及每次测量的难易程度。

对测试费用较高的情形,情形B与情形C的区别在于:情形C需要追加测量,而情形B不需要追加测量;当追加测量的费用非常高因而不可能时,情形B是唯一的选择。

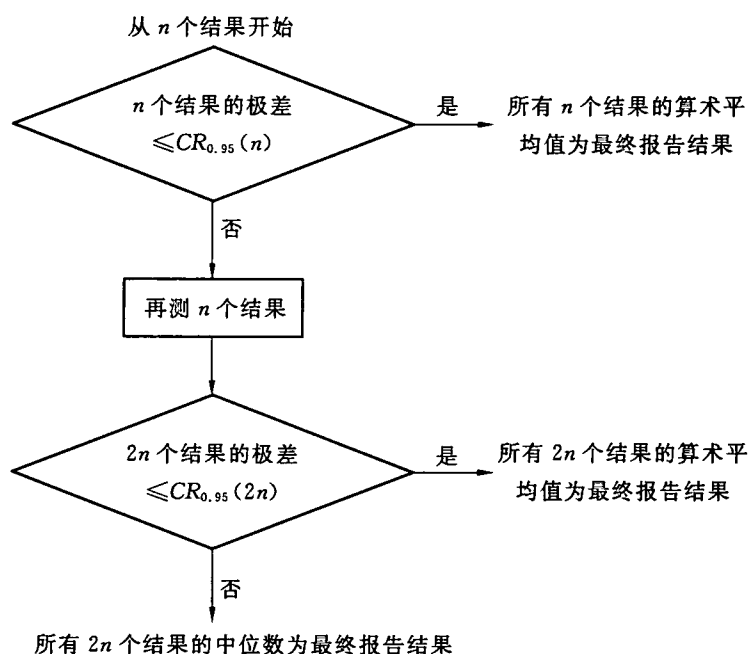


图4 在重复性条件下所得测试结果可接收性的检查方法  
(从 $n$ 个测试结果开始且测试费用较低时:情形A)

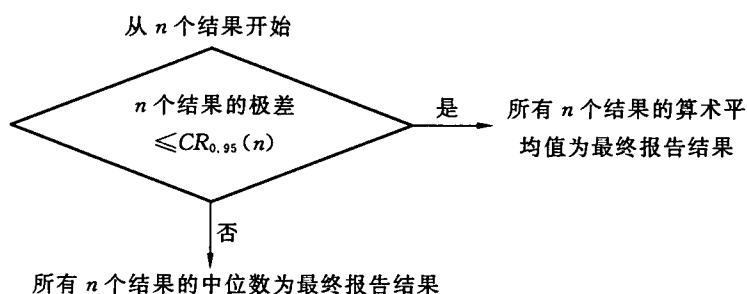
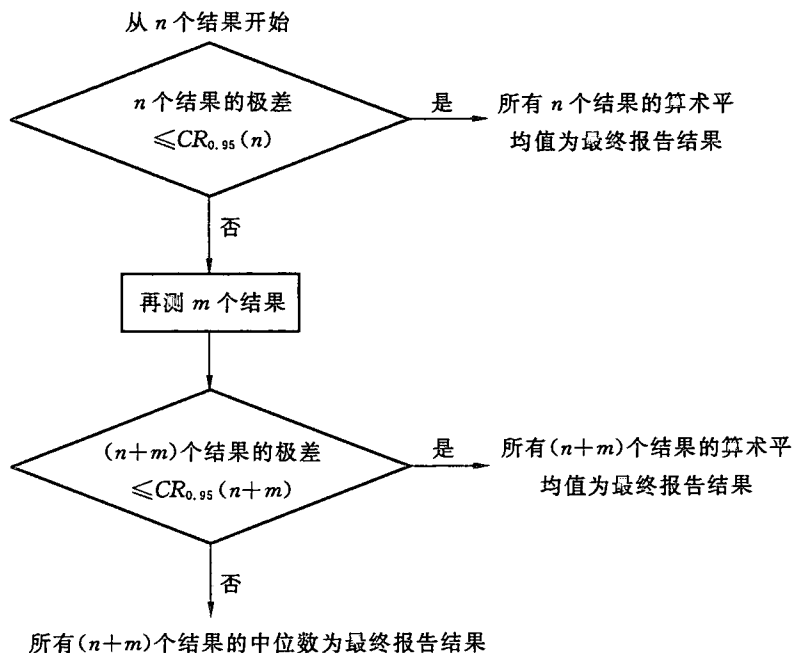


图5 在重复性条件下所得测试结果可接收性的检查方法  
(从 $n$ 个测试结果开始且测试费用较高时:情形B)



1)  $m$  应选为满足条件  $n/3 \leq m \leq n/2$  的整数

图 6 在重复性条件下所得测试结果可接收性的检查方法  
(从  $n$  个测试结果开始且测试费用较高时:情形 C)

#### 5.2.4 情形 B 的例:高费用的化学分析

在复杂而且耗时的化学分析中,费用昂贵的情形经常遇到,往往需要 2 天、3 天或者更多时间完成一次分析。如果在第一次分析中发现有技术上可疑的数据或者离群值,那么进行再分析将既费时又费钱。因此,通常在重复性条件一开始就获得 3 或 4 个测试结果,然后按照情形 B 的程序进行分析。参见图 5。

举个例子,用火试金分析法确定矿石中金、银的含量时,尽管有许多方法可用,但所有方法都要求有昂贵的特定设备、高度熟练的操作员和相当长的测量时间,通常需要 2 天左右。当矿石中含有铂族元素或其他共生的金属时,完成一次完整测量分析过程就需更多的时间。

以下是在重复性条件下精铜矿石中得到的金含量的 4 个测试结果,用情形 B 的方法对这些测试结果进行处理:

金的含量(g/t): 11.0 11.0 10.8 10.5

目前尚无测定矿石中金与银含量方法的国际标准,然而当给定  $\sigma_r = 0.12$  g/t 时,由表 1 查得  $f(4) = 3.6$ ,相应的临界极差为:

$$CR_{0.95}(4) = 3.6 \times 0.12 = 0.43 \text{ g/t}$$

因为上述 4 个测试结果的极差为:  $11.0 - 10.5 = 0.5$  g/t,该极差大于临界极差,所以最终报告结果为 4 个结果的中位数,即:

$$\frac{11.0 + 10.8}{2} = 10.9 \text{ g/t}$$

#### 5.2.5 关于精密度试验的说明

如按 5.2.2 或 5.2.3 所述的方法,结果频频超过临界值,则应对该实验室测量方法的精密度和(或)精密度试验进行调查。

#### 5.2.6 最终报告结果

如果仅需要最终测试结果时,应对如下两点进行说明:

——用于计算最终报告测试结果所用的测试结果数;

——最终报告测试结果用的是测试结果的算术平均值还是中位数。

### 5.3 在再现性条件下所得测试结果可接收性的检查方法

#### 5.3.1 总则

这些方法适用于有两个实验室且所得测试结果或结果的算术平均值有差异的情形。此时应当像重复性情形一样,用再现性标准差来作统计检验。

各种情况下均应保证有足够量的测试物料(物质或材料)用作测试,包括保存一部分备用样品以在必要时重新测试时使用。备用材料的多少取决于测量方法及其复杂程度。应妥善保存备用物料,防止损坏和变质。

两实验室的试样应是同一的,即两个实验室应使用样本制备阶段中的末级样本。

#### 5.3.2 两实验室测试结果一致性的统计检验

##### 5.3.2.1 每个实验室只有一个测试结果的情形

当每个实验室只有一个测试结果时,两实验室结果之差的绝对值应用再现性限  $R=2.8\sigma_R$  来检验。如果绝对差小于或等于  $R$ ,即认为结果一致,取其平均值作为最终报告结果;如果两个测试结果的绝对差大于  $R$ ,必须找出差异的原因是否是由于测试方法的精密度低和(或)测试样本有差别。两个实验室应遵循 5.2.2 的规定程序在重复性条件下对精密度进行检验。

##### 5.3.2.2 每个实验室有一个以上测试结果的情形

假定每个实验室都按 5.2 的规定程序取得了最终报告结果。因而,只要考虑两个最终报告结果的可接收性即可。为检验两个实验室的结果是否一致,应将两个结果的绝对差与临界差  $CD_{0.95}$  比较。不同情况的  $CD_{0.95}$  的表达式如下:

- a) 两个结果均为算术平均值(重复次数分别为  $n_1$  和  $n_2$ )时的临界差  $CD_{0.95}$  为:

$$CD_{0.95} = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2n_1} - \frac{1}{2n_2}\right)}$$

若  $n_1 = n_2 = 1$ ,上式化简为 5.3.2.1 给出的  $R$ ;若  $n_1 = n_2 = 2$ ,则化简为:

$$CD_{0.95} = \sqrt{R^2 - \frac{r^2}{2}}$$

- b) 两个结果之一为平均值(重复次数为  $n_1$ ),另一为中位数(重复次数为  $n_2$ )时的临界差  $CD_{0.95}$  为:

$$CD_{0.95} = \sqrt{R^2 - r^2 \left(1 - \frac{1}{2n_1} - \frac{\{c(n_2)\}^2}{2n_2}\right)}$$

其中  $c(n)$  为中位数的标准差与平均值的标准差之比,其值见表 2。

- c) 两个结果均为中位数(重复次数分别为  $n_1$  和  $n_2$ )时的临界差  $CD_{0.95}$  为:

$$CD_{0.95} = \sqrt{R^2 - r^2 \left(1 - \frac{\{c(n_1)\}^2}{2n_1} - \frac{\{c(n_2)\}^2}{2n_2}\right)}$$

$c(n)$  的数值见表 2。

如果两个结果差的绝对值不大于临界差,则两个实验室的最终报告结果均可接收,取两个结果的总平均作为最终报告结果。如果两结果之差的绝对值大于临界差,则应采用 5.3.3 给出的程序处理。

表 2  $c(n)$  的数值

测试结果数 $n$	$c(n)$
1	1.000
2	1.000
3	1.160
4	1.092
5	1.197

表 2 (续)

测试结果数 $n$	$c(n)$
6	1.135
7	1.214
8	1.160
9	1.223
10	1.176
11	1.228
12	1.187
13	1.232
14	1.196
15	1.235
16	1.202
17	1.237
18	1.207
19	1.239
20	1.212

### 5.3.3 解决两实验室的测试结果不一致的办法

引起两实验室的最终报告结果不一致的原因可能有：

- 两实验室之间的系统差异；
- 测试样本的差异；
- 确定  $\sigma_R$  和(或)  $\sigma_r$  过程中的误差。

若有可能交换所用的试样和(或)标准物料(标准物质/标准材料),每个实验室都应用另一实验室的试样进行测试,以判断系统误差存在与否及其程度。如无此种可能,每一实验室应使用一种共同的试样(最好是已知测量特性值的物料),这样做的优点是可以找出某个实验室或两个实验室的各自的系统误差。如果不能用这种方法来确定系统误差,两实验室应参照第三方参考实验室的结果来解决。

当不一致可能是因测试样本的差异引起时,两实验室应联合制作共同试样或委托第三方进行抽样。

### 5.3.4 仲裁

在合同签约或发生争议时,合同双方可通过仲裁解决。

## 6 实验室内检查测试结果稳定性的方法

### 6.1 背景

6.1.1 质量控制的第一步就是通过化学分析、物理测试和感官检验等手段进行量化。用上述量化方法获得的观测值总会有一些误差,依其来源可分为如下几类:

- 抽样误差;
- 样本制备误差;
- 测量误差,等。

本章中仅讨论测量误差,包括由于同一测试样本不同测试份样之间的不可分离的差异所引起的误差。

6.1.2 测量误差又可分为:

- 由随机因素引起的误差(对应于精密度);
- 由系统因素引起的误差(对应于正确度)。

6.1.3 考虑一个测量方法,自然希望它的精密度和正确度都能够满足要求。然而,当一个测量方法的精密度满足要求时,并不能保证它的正确度也满足要求。因此,当检查一个实验室内测试结果的稳定性时,有必要对测试结果的精密度和正确度都进行检查,将两种度量分别长期维持在所要求的水平。

6.1.4 对某些测量方法,真值可能不存在,或即使存在但由于没有标准物料(RM)而不能对测试结果的正确度进行检查。表3给出了若干例子。

如果没有标准物料(RM),要检查一个测试结果的正确度是困难的。然而,在许多实际情形,可以用如下的测试结果代替认定值,作为参照值使用。这个测试结果必须由技术熟练的操作员在设备精良的实验室内,严格、仔细、完全地按照标准测量方法(更确切的应称为“确定的”方法)进行测量获得。

6.1.5 为检查实验室内测试结果的稳定性,GB/T 6379 本部分需用到常规控制图(参见 GB/T 4091—2001)及累积和控制图(参见 GB/Z 4887—2006)。

当精密度或正确度有某种变化趋势或偏移时,使用累积和控制图来检查测试结果的稳定性较常规控制图效率更高;当正确度或精密度的变化是突然发生时,累积和控制图比常规控制图相比并没有明显的优势。

因为正确度的变化往往具有某种变化趋势或偏移,而精密度的变化往往是突然发生的,所以建议在检查正确度时使用累积和控制图,在检查精密度时使用常规控制图。

不过,对正确度和精密度的检查同时使用两种控制图方法也是值得的。

6.1.6 由于检查的过程可能要经历较长一段时间,在此期间,操作员、设备可能会发生变化,从而使真正的重复性条件不再适用。因此,检查也涉及到 ISO 5725-3 中所述的中间精密度度量。

## 6.2 检查稳定性的方法

### 6.2.1 总则

6.2.1.1 检查实验室内测试结果的稳定性时,需要考虑如下两种情况(测试物料特性分类见表3):

- a) 过程控制中使用的常规测试结果;
- b) 原材料和产品定价过程中使用的测试结果。

表3 为检查测试结果的准确度(正确度与精密度)  
根据真值和重要参数对测试物料特性的分类

分 类 <sup>1)</sup>	例		
	物料特性	是否有标准物料可用 <sup>2)</sup>	检查准确度的重要参数 <sup>3)</sup>
根据科学原理的一个理论值在实际中能作为真值	苯(甲)酸的化学成分	有标准物料 <sup>4)</sup>	$\Delta$ 和 $\sigma_w$
尽管真值在理论上存在,但在实际中用现有技术不能唯一确定;此时,可把一个由科学的或工程的协作组织在协同试验中确认的值作为约定真值	a) 矿石中铁含量的百分比	有标准物料	$\Delta$ 和 $\sigma_w$
	b) 黄铁矿中硫含量的百分比	无标准物料 <sup>5)</sup>	$\sigma_w$ 和 $\sigma_L$
依照某标准测试方法,由国际、国内或某个组织确定的一个指定值作为约定真值	a) 汽油中的辛烷值	有标准物料	$\Delta$ 和 $\sigma_w$
	b) 焦炭的强度	无标准物料 <sup>6)</sup>	$\sigma_M/\sigma_w, \sigma_L$ 和 $\sigma_w$
	c) 热塑性塑料的熔流率	无标准物料 <sup>7)</sup>	$\sigma_w$ 和 $\sigma_L$

1) 见 GB/T 3358.1。

2) 见 ISO 指南 35。

3)  $\Delta$  是实验室偏倚; $\sigma_w$  是实验室标准差; $\sigma_L$  是实验室间标准差; $\sigma_M$  是测试样本间的标准差。

4) 如果测试物料本身是纯的并且是稳定的,可用作标准物料。

5) 因为测试物料不稳定而不能确定标准物料。

6) 测试物料由大量易碎的,粒度、形状不同的固体颗粒组成,为测定其成分的测试又是破坏性的,从而不能确定标准物料。

7) 参照值由测量方法本身给定。

6.2.1.2 对于 a) 的情形,必须在一个特定的实验室长时间地分别对中间精密度条件下获得的测试结果进行检查(这里中间精密度条件包括一个因素不同、两个因素不同和三个因素不同等情况),以确保精密度度量维持在一个理想的水平(见 6.2.3 的例 2)。对此情形,大多数情况下只需检查精密度就足够了。因为即使测试结果是偏的,如果测试结果的变异相对于生产过程的变异足够小,只检查精密度就能反映过程变异。然而,如果把重复性标准差作为精密度的检查指标,由于过高的灵敏性会导致过程控制的过度反应;因此建议用适当的中间精密度标准差作为检查指标。

6.2.1.3 对于 b) 的情形,在检查精密度的同时,必须也对正确度(参见 6.2.4 中的例 3)进行检查,看两种度量是否分别维持在所需的水平。在此情形,需要一个认可的参照值。

6.2.1.4 下面给出四个例子:

——例 1 和例 2 说明如何用常规控制图方法检查重复性或中间精密度度量的稳定性;

——例 3 和例 4 说明如何用常规控制图或累积和技术检查正确度。

6.2.2 例 1 常规分析中重复性标准差的稳定性检查

6.2.2.1 背景

a) 测量方法:

用 ISO 6352:1985《镍铁 镍成分含量的测定 丁二酮肟重量法》测定镍的含量。

b) 资料来源:

某镍铁熔炼厂实验室 1985 年 9 月的常规报告。

c) 说明:

在镍铁熔炼厂的工作实验室里,每天都要进行化学分析来测定镍铁产品的化学成分,用实验室自己制备的标准物料进行镍含量的稳定性检查。

为检查镍铁成份的稳定性,每天在重复性条件下对实验室自制的标准物料的两部分进行测试分析,这里的重复性条件是指由同一个操作员、使用同样的设备,在同一时间进行测试。

以下是实验室自制标准物料的化学成分:

Ni	47.21%	Co	1.223%	Si	3.50%	Mn	0.015%
P	0.003%	S	0.001%	Cr	0.03%	Cu	0.038%

6.2.2.2 原始数据

表 5 给出了重复性条件下自制标准物料的常规分析测试结果,分别记为  $x_1$  和  $x_2$ ,以质量分数表示。

6.2.2.3 用常规控制图方法进行的稳定性检查

把常规控制图(R 图,即极差控制图,参见 GB/T 4091—2001)用于表 5 的测试结果,可以检查测试结果的稳定性,同时可以估计重复性标准差的大小。在计算中心线和控制限时,需用表 4 中列出的各种系数。

注 4: 此处 R 图是指 GB/T 4091 中的极差控制图,不要与本标准中使用的再现性符号 R 相混淆。

表 4 用于计算极差控制图的系数

计算中心线与行动限的系数 <sup>1)</sup>			用于计算警戒限的系数 <sup>2)</sup>		
子组中的观测数	用于中心线的系数 $d_2$	用于行动限的系数 $D_2$	用于警戒限的系数		
			$d_3$	$D_1(2)$	$D_2(2)$
2	1.128	3.686	0.853	—	2.834
3	1.693	4.358	0.888	—	3.469
4	2.059	4.698	0.880	0.299	3.819
5	2.326	4.918	0.864	0.598	4.054

1) 部分数值引自 GB/T 4091—2001 的表 2。

2) 用于计算警戒限的系数为:

$$D_1(2) = d_2 - 2d_3$$

$$D_2(2) = d_2 + 2d_3$$



表 5 6.2.2 例 1 控制图的数据表

1. 质量特性:自制标准物料中锂含量 2. 计量单位:%(m/m) 3. 分析方法:ISO 6352 4. 测试时间:1985年9月1日~9月30日 5. 实验室:某钢铁厂的实验室 A					
分析日期 (子组号)	观测值		极差 $w$	说 明	
	$x_1$	$x_2$			
1	47.379	47.333	0.046	位于警戒限之上	
2	47.261	47.148	0.113		
3	47.270	47.195	0.075		
4	47.370	47.287	0.083		
5	47.288	47.284	0.004		
6	47.254	47.247	0.007		
7	47.239	47.160	0.079		
8	47.239	47.193	0.046		
9	47.378	47.354	0.024		
10	47.331	47.267	0.064		
11	47.255	47.278	0.023		
12	47.313	47.255	0.058		
13	47.274	47.167	0.107		位于警戒限之上
14	47.313	47.205	0.108		位于警戒限之上
15	47.296	47.231	0.065		
16	47.264	47.247	0.017		
17	47.238	47.253	0.015		
18	47.181	47.255	0.074		
19	47.327	47.240	0.087		
20	47.358	47.308	0.050		
21	47.295	47.133	0.162		位于警戒限之上
22	47.310	47.244	0.066		
23	47.366	47.293	0.073		
24	47.209	47.185	0.024		
25	47.279	47.268	0.011		
26	47.178	47.200	0.030		
27	47.211	47.193	0.018		
28	47.195	47.216	0.021		
29	47.274	47.252	0.022		
30	47.300	47.212	0.088		
合计			1.660		
平均值			0.055 3	$\bar{w}/d_2=0.049 0$	
注: $\sigma_r=0.037 5$ a) 中心线= $d_2\sigma_r=1.128\times 0.037 5=0.042 3$ b) 行动限 $UCL=D_2\sigma_r=3.686\times 0.037 5=0.138 2$ $UCL=-$ c) 警戒限 $UCL=D_2(2)\sigma_r=2.834\times 0.037 5=0.106 2$ $UCL=-$					

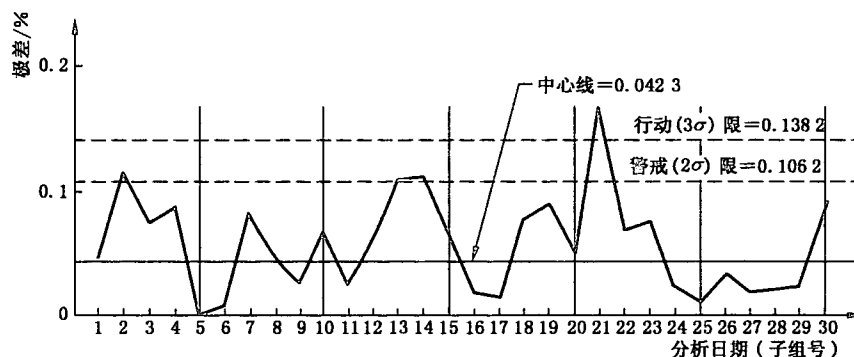


图 7 重复性条件下自制标准物料的镍含量(%)的极差控制图

本例中,把该年前一季度测试结果所得的重复性标准差  $\delta_r$  作为极差控制图的标准值,控制图的计算如下:

a) 中心线 =  $d_2\sigma_r = 1.128 \times 0.0375 = 0.0423$

b) 行动限

$UCL = D_2\sigma_r = 3.686 \times 0.0375 = 0.1382$

$LCL = -$

注:“—”表示无限制,不需标出,下同。

c) 警戒限

$UCL = D_2(2)\sigma_r = 2.834 \times 0.0375 = 0.1062$

$LCL = -$

计算重复性标准差的估计( $s_r$ )如下:

$$w = |x_1 - x_2|$$

$$s_r = \left( \sum_1^{30} w_i / 30 \right) / d_2 = \bar{w} / d = 0.0553 / 1.128 = 0.0490$$

共计算了 30 个子组的极差,每个子组中包含两个观测值。表 5 是计算过程中的工作表,图 7 则是标有控制限的数据点图。

在图 7 的控制图中,因为有一个点在行动限之上,且连续有两个点高于警戒限,这表明例中的测试结果不稳定。

### 6.2.3 例 2 常规分析中时间与操作员不同的中间精密度标准差的稳定性检查

#### 6.2.3.1 背景

a) 测量方法

用 ISO 351:1984《固体矿物燃料 总硫含量的测定 高温燃烧法》,测定高炉焦炭中的硫含量,测试结果以质量百分数表示。

b) 资料来源

1985 年 8 月某钢铁厂实验室的常规报告。

c) 说明

在生产高炉焦炭的炼焦炉中,每天三班轮换,在每一班生产的每一批产品中定期抽取焦炭样本,制作成实验室进行化学分析的测试样本,测定焦炭中硫的含量[%(m/m)]。

### 6.2.3.2 原始数据

表 6 是对焦炭中硫含量[% (m/m)]质量控制分析的测试结果,所测试的焦炭样本均在 1985 年 8 月取自 1 号焦炉。在测试过程中,把随机抽取的每个焦炭测试样本与其他测试样本分开,先由第一班的一名操作员对其分析,观测值记作  $x_1$ ; 然后由另一班的另一操作员在第二天对它再作分析,观测值记作  $x_2$ 。每天都对测试结果进行比较。

### 6.2.3.3 用常规控制图方法进行的稳定性检查

把常规控制图(R 图,参见 GB/T 4091)用于表 6 的测试结果,可以对测试结果的稳定性进行检查,同时可以估计时间与操作员不同的中间精密度标准差的大小。

有关计算中心线、行动限和警戒限(UCL 和 LCL)时所需的系数见 6.2.2 的例 1。把该年前一季度测试结果得到的时间与操作员不同的中间精密度标准差  $\sigma_{I(TO)}$  作为极差控制图的标准值。控制图的有关计算如下:

a) 中心线 =  $d_2\sigma_{I(TO)} = 1.128 \times 0.0133 = 0.0150$

b) 行动限

$$UCL = D_2\sigma_{I(TO)} = 3.686 \times 0.0133 = 0.0490$$

$$LCL = -$$

c) 警戒限

$$UCL = D_2(2)\sigma_r = 2.834 \times 0.0133 = 0.0378$$

$$LCL = -$$

时间与操作员不同的中间精密度标准差的估计  $s_{I(TO)}$  计算如下:

$$w = |x_1 - x_2|$$

$$s_{I(TO)} = \left( \sum_{i=1}^{31} w_i / 31 \right) / d_2 = \bar{w} / d_2 = 0.0142 / 1.128 = 0.0126$$

共计算了 31 个子组的极差,每个子组中包含两个观测值,数据见表 6,而图 8 则是标有控制限的这些数据的点图。

在图 8 的控制图中,没有证据表明测试结果不稳定。

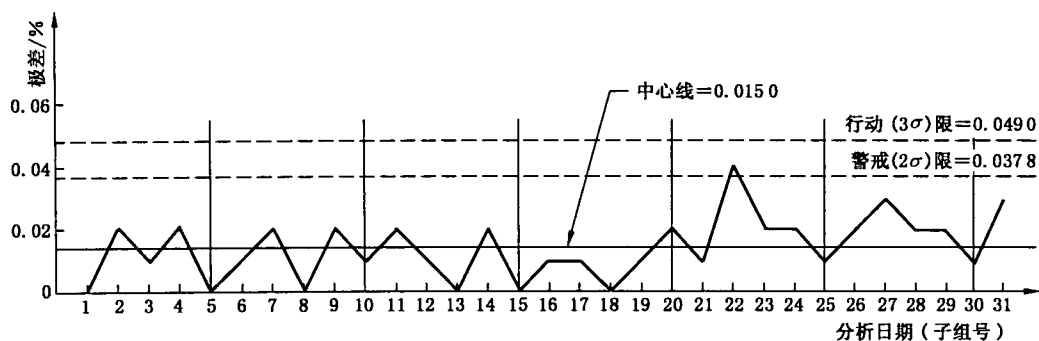


图 8 时间与操作员不同的中间精密度条件下所得的高炉焦炭硫含量(%)的极差控制图

表 6 6.2.3 例 2 控制图数据表

1. 质量特性:高炉焦炭中的硫含量 2. 计量单位:%(m/m) 3. 分析方法:ISO 351 4. 测试时间:1985年8月1日~8月31日 5. 实验室:某钢铁厂的实验室 B				
分析日期 (子组号)	观测值		极差 <i>w</i>	说 明
	$x_1$	$x_2$		
1	0.56	0.56	0.00	位于警戒限之上 <sup>1)</sup>
2	0.48	0.50	0.02	
3	0.57	0.58	0.01	
4	0.60	0.58	0.02	
5	0.58	0.58	0.00	
6	0.50	0.49	0.01	
7	0.56	0.58	0.02	
8	0.56	0.56	0.00	
9	0.48	0.46	0.02	
10	0.54	0.53	0.01	
11	0.55	0.57	0.02	
12	0.46	0.45	0.01	
13	0.58	0.58	0.00	
14	0.54	0.56	0.02	
15	0.56	0.56	0.00	
16	0.57	0.58	0.01	
17	0.46	0.45	0.01	
18	0.56	0.56	0.00	
19	0.56	0.57	0.01	
20	0.57	0.55	0.02	
21	0.44	0.45	0.01	
22	0.59	0.55	0.04	
23	0.55	0.57	0.02	
24	0.58	0.56	0.02	
25	0.46	0.45	0.01	
26	0.60	0.58	0.02	
27	0.59	0.56	0.03	
28	0.54	0.56	0.02	
29	0.47	0.49	0.02	
30	0.59	0.58	0.01	
31	0.49	0.52	0.03	
合计	16.74	16.72	0.44	
平均数			0.014 2	$\bar{w}/d_2=0.012 6$
注: $\sigma_{I(TO)}=0.013 3$ $x_1$ : 常规分析 $x_2$ : 由不同的操作员在第二天所作的第二次分析 a) 中心线 $=d_2\sigma_{I(TO)}=1.128\times 0.013 3=0.015 0$ b) 行动限 $UCL=D_2\sigma_{I(TO)}=3.686\times 0.013 3=0.049 0$ $UCL=-$ c) 警戒限 $UCL=D_2(2)\sigma_{I(TO)}=2.834\times 0.013 3=0.037 8$ $UCL=-$				
1) 为了获得 $x_2$ 的实际加热温度低于指定温度。				

## 6.2.4 例3 常规分析中正确度的稳定性检查

## 6.2.4.1 背景

## a) 测量方法

用 ISO 1171:1981《固体矿物燃料 灰分含量的确定》，测定煤炭中的灰分，测试结果以质量分数表示。

## b) 资料来源

1985年6月某钢铁厂实验室的常规报告。

## c) 说明

在钢铁厂中，按三班轮换方式用煤在炼焦炉提炼高炉焦碳。

为控制焦碳产品的质量，每一班对所用的原料煤按 ISO 1171:1981 给定的方法测定其中的灰分 $[\%(m/m)]$ ，并对常规分析时间与操作员不同的中间精密度标准差的稳定性进行检查，如 6.2.3 的例 2。

本例将说明对常规分析中正确度稳定性检查的方法，此间用到了自制的标准物料（灰分为 10.29%）。

## 6.2.4.2 原始数据

每天在三个班全体操作员中随机指定一人对标准物料进行分析，测试结果  $y$  列在表 7 中。

## 6.2.4.3 用常规控制图方法进行稳定性检查

将常规控制图用于表 7 的数据，对常规分析正确度的稳定性进行检查，同时估计偏倚的大小。

重复性标准差 ( $s_r$ ) 不能用于检查该特定实验室内偏倚，因为实验室内常规分析是在时间与操作员不同中间精密度条件下进行的，因而  $s_r$  不能代表该实验室所得测试结果的实际精密度。

为此采用较简单的移动极差控制图方法，而无需进行另外的试验以得到时间与操作员不同中间精密度标准差  $s_{I(TO)}$ 。

利用表 7 的注中给出的公式及事先确定的  $\mu$  和  $\sigma_{I(TO)}$  值，可绘制控制图。从图 9 中的控制图可以看出，有时偏倚和极差都很小，有时测试结果则不是很稳定，对此应进行调查。

## 6.2.4.4 用累积和控制图方法进行的稳定性检查

$\hat{\delta}$  的累积和控制图中 ( $H; K$ ) 的计算如下，其中  $(h; k) = (4.79; 0.5)$ ：

上侧：

$$H = h\sigma_{I(TO)} = 4.79 \times 0.06645 = 0.318$$

$$K_1 = \mu + k\sigma_{I(TO)} = 10.29 + 0.5 \times 0.06645 = 10.323$$

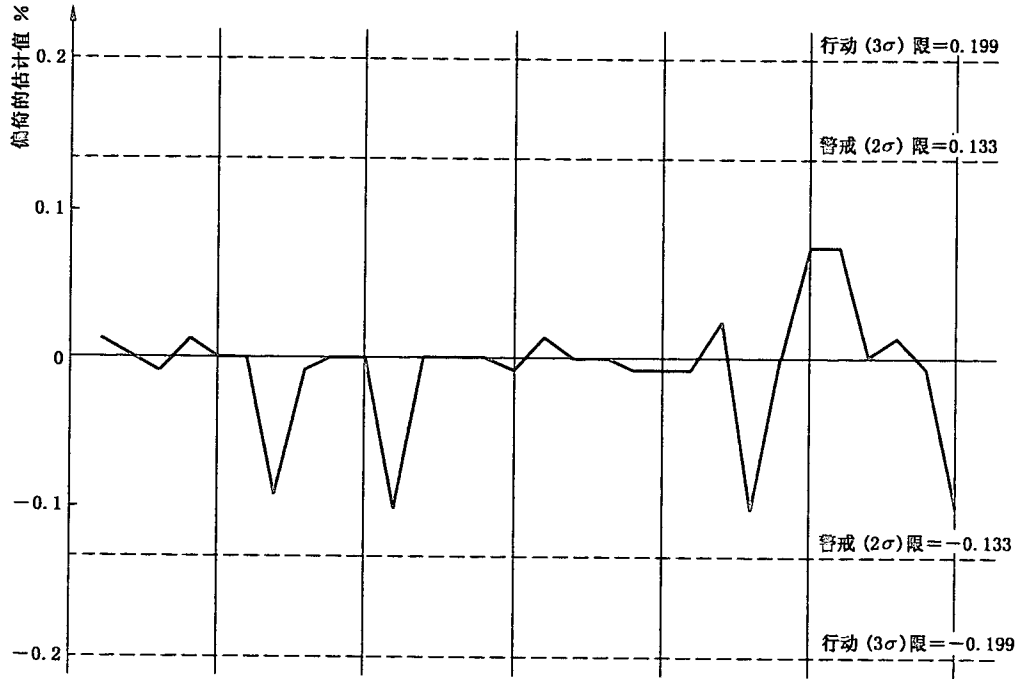
下侧：

$$-H = -0.318$$

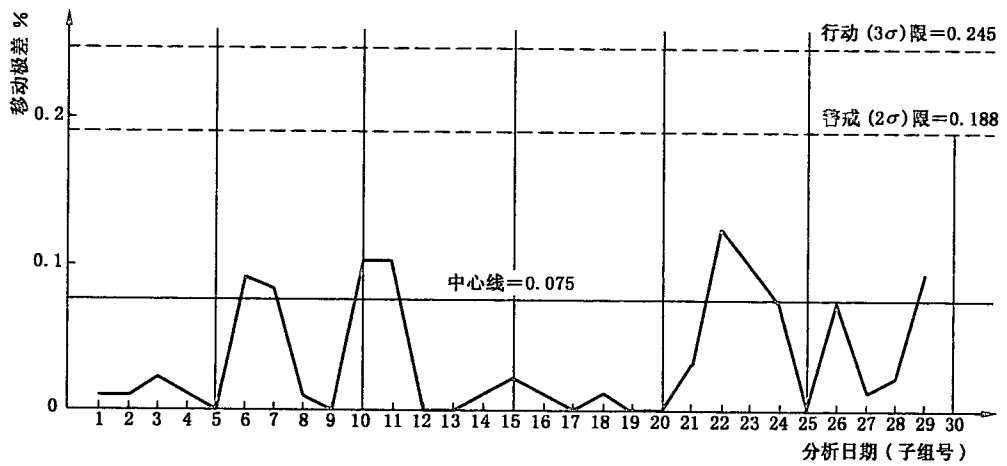
$$K_2 = \mu - k\sigma_{I(TO)} = 10.29 - 0.5 \times 0.06645 = 10.257$$

表 7 6.2.4 例 3 控制图数据表

1. 质量特性:自制标准物料中的灰分含量 2. 计量单位:%(m/m) 3. 分析方法:ISO 1171 4. 测试时间:1985年6月1日~6月30日 5. 实验室:某钢铁厂的实验室 C				
分析日期 (子组号)	测试结果 $y$	偏倚的估计值 $\hat{\delta}$	移动极差 $w$	说 明
1	10.30	0.01	0.01	
2	10.29	0.00	0.01	
3	10.28	-0.01	0.02	
4	10.30	0.01	0.01	
5	10.29	0.00	0.00	
6	10.29	0.00	0.09	
7	10.20	-0.09	0.08	
8	10.28	-0.01	0.01	
9	10.29	0.00	0.00	
10	10.29	0.00	0.10	
11	10.19	-0.10	0.10	
12	10.29	0.00	0.00	
13	10.29	0.00	0.00	
14	10.29	0.00	0.01	
15	10.28	-0.01	0.02	
16	10.30	0.01	0.01	
17	10.29	0.00	0.00	
18	10.29	0.00	0.01	
19	10.28	-0.01	0.00	
20	10.28	-0.01	0.00	
21	10.28	-0.01	0.03	
22	10.31	0.02	0.12	
23	10.19	-0.10	0.10	
24	10.29	0.00	0.07	
25	10.36	0.07	0.00	
26	10.36	0.07	0.07	
27	10.29	0.00	0.01	
28	10.30	0.01	0.02	
29	10.28	-0.01	0.09	
30	10.19	-0.10		
合计	308.44	-0.26	0.99	
平均值		-0.086 6	0.034 1	$\bar{w}/d_2=0.030 2$
注: 自制标准物料中的灰分含量: $\mu=10.29$ 上季度所得测试结果的标准差: $\sigma_{I(CTO)}=0.066 45$ 偏倚的估计值: $\hat{\delta}=y-\mu$ 移动极差: $w= \hat{\delta}_{i+1}-\hat{\delta}_i $ $\bar{x}$ -图: 中心线 =0 行动限 $UCL=+3\sigma_{I(CTO)}=0.199 4$ $UCL=-3\sigma_{I(CTO)}=-0.199 4$ 警戒限 $UCL=+2\sigma_{I(CTO)}=0.132 9$ $UCL=-2\sigma_{I(CTO)}=-0.132 9$ 移动极差图: 中心线 $=d_2\sigma_{I(CTO)}=1.128\times 0.066 45=0.074 96$ 行动限 $UCL=D_2(2)\sigma_{I(CTO)}=3.396\times 0.066 45=0.245$ $UCL=-$ 警戒限 $UCL=D_2(2)\sigma_{I(CTO)}=2.834\times 0.066 45=0.188 3$ $UCL=-$				



a)



b)

图 9 自制标准物料中灰分的偏倚估计  $\hat{\delta}$  [% (m/m)] 的常规控制图

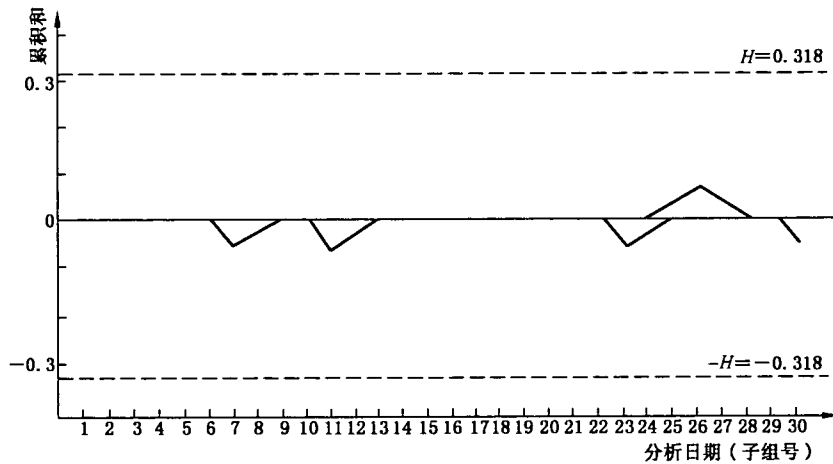


图 10 自制标准物料中灰分的偏倚估计  $\hat{\delta}[\%(m/m)]$  的累积和控制图

6.2.5 例 4 常规分析正确度稳定性检查的另一例

6.2.5.1 背景

a) 测量方法

用二乙基硫代氨基甲酸银-砷化三氢比色法确定氧化锌中的砷含量。

b) 资料来源

Kanzelmeyer J. H., “分析方法的质量控制”, 美国材料试验学会标准化通讯, 1977 年 10 月, 第 27 页, 图 2。

6.2.5.2 原始数据

见表 8。

6.2.5.3 用常规控制图方法进行的稳定性检查

利用表 8 注释中给出的公式以及事先确定的  $\mu$  和  $\sigma$  值, 可绘制  $\bar{x}$  的常规控制图(见图 11)。因有一个点高于行动限, 且有两个链长大于或等于 7 的链在中心线以下, 这表明例中的测试结果不稳定。

6.2.5.4 用累积和控制图法进行稳定性检查

对  $\bar{x}$  的累积和控制图(见图 12)中  $(H; K)$  的计算如下, 其中  $(h; k) = (4.79; 0.5)$  :

上侧:

$$H = h\sigma_r / \sqrt{n} = 4.79 \times 0.167 = 0.800$$

$$K_1 = \mu + k\sigma_r / \sqrt{n} = 3.800 + 0.5 \times 0.167 = 3.88$$

下侧:

$$-H = -0.800$$

$$K_2 = \mu - k\sigma_r / \sqrt{n} = 3.800 - 0.5 \times 0.167 = 3.72$$



表 8 6.2.5 例 4  $\bar{x}$  控制图数据表

1. 质量特性:自制标准物料中的砷含量				
2. 计量单位:质量 $\times 10^{-6}$				
3. 分析方法:二乙基硫代氨基甲酸银—砷化三氢比色法				
子组号	观测值		$\bar{x}$	说 明
	$x_1$	$x_2$		
1	3.70	3.80	3.75	位于行动限之上
2	3.76	3.86	3.81	
3	3.64	3.38	3.51	
4	4.01	3.62	3.82	
5	3.40	3.52	3.46	
6	3.65	3.53	3.59	
7	3.20	3.58	3.39	
8	4.19	4.65	4.42	
9	3.97	3.77	3.87	
10	2.95	3.69	3.32	
11	3.43	3.55	3.49	
12	3.85	3.53	3.69	
13	3.77	3.17	3.47	
14	3.19	3.60	3.40	
15	3.75	3.45	3.60	
16	3.55	3.25	3.40	
17	3.98	3.76	3.87	
18	3.56	3.78	3.67	
19	3.54	4.02	3.78	
20	3.35	3.55	3.45	
21	3.37	3.25	3.31	
22	3.42	3.42	3.42	
23	3.71	3.87	3.79	
24	3.77	3.62	3.70	
25	3.82	3.58	3.70	
26	3.73	3.02	3.38	
27	3.48	3.28	3.38	
28	4.01	4.19	4.10	
29	3.63	3.11	3.37	
30	3.51	3.23	3.37	
合计			108.28	
平均值			3.609	
注:				
自制标准物料中的砷含量:				
$\mu=3.80$				
以往的标准差:				
$\sigma_r=0.236$				
$\bar{x}$ -图				
中心线=3.80				
行动限				
$UCL=\mu+3\sigma_r/\sqrt{n}=4.300$				
$LCL=\mu-3\sigma_r/\sqrt{n}=3.299$				

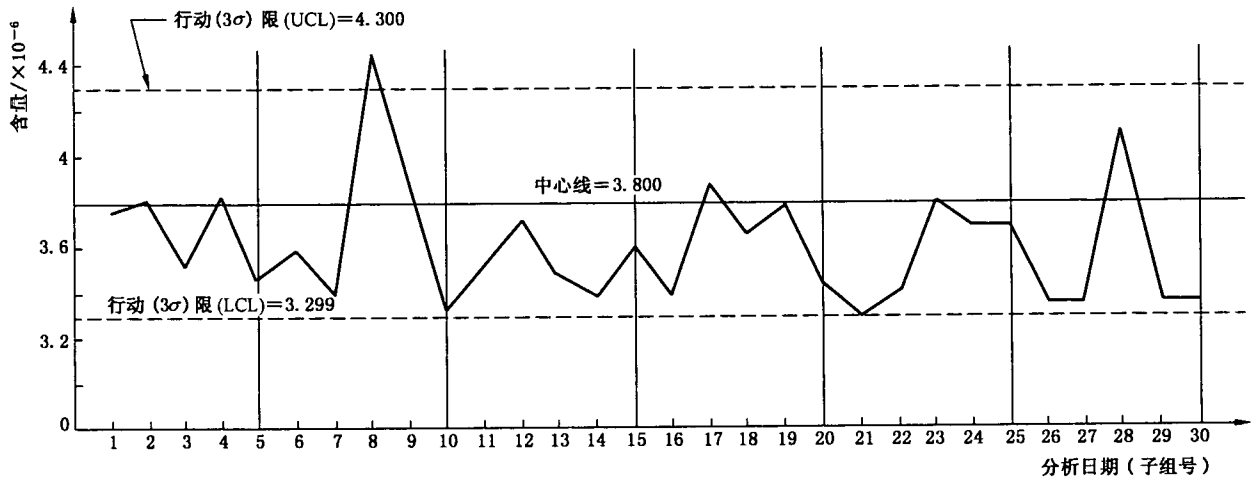


图 11 用二乙基硫代氨基甲酸银-砷化三氢比色法测定氧化锌中的砷含量的  $\bar{x}$  常规控制图

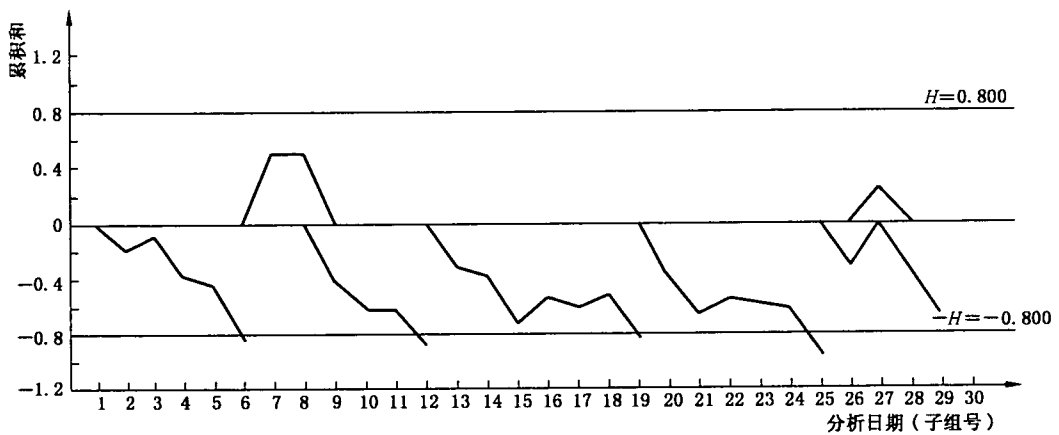


图 12 用二乙基硫代氨基甲酸银-砷化三氢比色法测定氧化锌中的砷含量的  $\bar{x}$  累积和控制图

## 7 重复性标准差和再现性标准差在实验室评定中的应用

### 7.1 评定方法

#### 7.1.1 总则

本章描述一种被许多实验室使用,且已标准化其测量方法的实验室的评定。因而该测量方法的精密度可以重复性标准差和再现性标准差的形式获得估计。假定这些值经精密度试验已事先确定。

根据测量方法是否存在标准物料和是否存在参照实验室的不同情况,实验室评定分为三种不同类型。当标准物料在一定数量水平上都存在时,只需待评实验室本身参与。对于不存在标准物料的测量方法,则不能按上述简单方法。待评实验室必须与一个被广泛认可的、可为评定提供基准的“高质量”实验室进行比较。对于实验室的连续再评定,常需对多个实验室同时进行,此时需进行一个协同评定试验。

进行协同评定试验的目的是把每一个实验室的测试结果与其他实验室的测试结果进行比较,以提高效能。

#### 7.1.2 协同评定试验定义的含义

测量方法的重复性标准差是对在同一实验室、一致条件下所得测试结果不确定性的度量,即 GB/T 6379.1 中定义的重复性条件下实验室内精密度的一种表示。

当所测量的真值存在且已知时,或者有标准物料时,实验室偏倚即可立刻确定。当所测量的真值未知时,实验室偏倚只能间接确定。一种方法是将该实验室与一个已知偏倚的实验室进行比较。然而这种方法过分依赖于所“参照”实验室的偏倚和精密度。

在协同评定试验的情形,再现性是表示不同实验室所得测试结果一致性的度量,因此可以用来估计每一实验室的偏倚。当评定试验的再现性已确定时,系统偏差大的实验室表现为离群实验室。

本章中,假定测量方法的精密度已经事先确定,这意味着重复性方差  $\sigma_r^2$ 、实验室间方差  $\sigma_L^2$  及再现性方差  $\sigma_R^2$  均已知。

第7章的方法主要用于检查实验室偏倚。当检查一个实验室内重复性或中间精密度时,第6章的方法更有效。

## 7.2 由未经评定的实验室使用的测量方法的评估

### 7.2.1 实验室运作的评估

关于实验室评估的一般准则,见 GB/T 27025。实验室应有良好的运作模式,且有令人满意的内部质量控制。内部质量控制的方法已在第6章中给出。

本部分控制仅基于对每一实验室日常工作状态的检查。这种控制可以直接进行,无需特定的测试物料,也不涉及别的实验室。

为了对实验室中测量方法的使用进行定量评估,必须进行对照试验。对照试验既可以在实验室内通过使用标准物料进行(参见 7.2.3),也可以通过与一个高质量的实验室进行比较来完成(参见 7.2.4)。

### 7.2.2 关于对照试验的一般考虑

在计划一个对照试验时,应考虑如下问题:

- a) 试验需在几个水平上进行? GB/T 6379.1—2004 的 6.3 中考虑了这个问题。
- b) 每一水平上应进行几次重复?

在协同评定试验情形,还要考虑:

- c) 需要多少实验室参与?

在安排对照试验时,需考虑 GB/T 6379.1—2004 的 6.1 及 GB/T 6379.2—2004 的第5章和第6章中的内容。

测试物料应匿名发送给实验室,其目的是确保实验室不会对测试物料以特殊对待,而应与平时操作保持一致。

### 7.2.3 有标准物料时的测量方法

#### 7.2.3.1 总则

7.2.3.1.1 当有标准物料时,可以对单个实验室进行评定。当测量方法的精密度已知时,已知的重复性标准差用来评估实验室内精密度,而偏倚则可将测试结果与参照值比较来确定。

有时需要引入一个可检出的实验室偏倚  $\Delta_m$ ,作为试验者希望从试验结果中能以高概率检测出实验室偏倚的最小值。

7.2.3.1.2 为评估实验室内部精密度,必须在实验室内进行重复测量。在对 7.2.2 所提的问题加以考虑后,按  $q$  个水平派送测试物料,每一水平进行  $n$  次重复测量。用 GB/T 6379.2—2004 第7章给出的方法评估测试结果。在评估实验室内部精密度时,将单元内标准差  $s_r$  与已知的重复性标准差  $\sigma_r$  进行比较。接收准则是:

$$s_r^2/\sigma_r^2 < \chi_{(1-\alpha)}^2(\nu)/\nu \quad \dots\dots\dots(1)$$

其中  $\chi_{(1-\alpha)}^2(\nu)$  是  $\chi^2$  分布的  $1-\alpha$  分位点,自由度  $\nu=n-1$ 。除非特别说明,显著性水平  $\alpha$  假定为 0.05。

$q$  个水平中大约有 95% 满足不等式(1)。由于通常  $q$  相当小,这意味着实验室中所有的  $q$  个水平都应满足准则(1)。

7.2.3.1.3 对偏倚进行评估时,要将每一水平的测试结果的平均值  $\bar{y}$  与相应的参照值  $\mu$  进行比较。因为

$$s_{(y)}^2 = s_L^2 + \frac{1}{n}s_r^2 = s_R^2 - s_r^2 \frac{(n-1)}{n} \dots\dots\dots(2)$$

故接收准则为：

$$|\bar{y} - \mu| < 2 \sqrt{\sigma_R^2 - \sigma_r^2 \frac{(n-1)}{n}} \dots\dots\dots(3)$$

实验室中所有的  $q$  个水平都应满足准则(3)。

当  $n=2$  时,接收准则(3)化简为：

$$|\bar{y} - \mu| < 2 \sqrt{\sigma_R^2 - \frac{\sigma_r^2}{2}} \dots\dots\dots(4)$$

当有可检出的实验室偏倚  $\Delta_m$  时,相应的接收准则为：

$$|\bar{y} - \mu| < \Delta_m/2 \dots\dots\dots(5)$$

7.2.3.2 例：混凝土中水泥含量的测定

7.2.3.2.1 背景

混凝土中的水泥含量是一个重要指标,因为它影响混凝土的耐久性。在混凝土规范中常标出水泥含量的最小值。水泥含量可以通过测量水泥(和凝料)样本及混凝土样品中的钙含量的而确定。为对一个实验室进行评定,可以事先准备已知水泥含量的混凝土样品。

为对 6 个实验室进行评定,制备了水泥含量为 425 kg/m<sup>3</sup> 的参照样品,每一个实验室对此做 2 次水泥含量的测定。

7.2.3.2.2 原始数据

见表 9。已知重复性标准差和再现性标准差的值为：

$$\sigma_r = 16, \sigma_R = 25$$

表 9 混凝土中水泥的含量

实验室 $i$	观测值	
	$y_{i1}$	$y_{i2}$
1	406	431
2	443	455
3	387	431
4	502	486
5	434	456
6	352	399

7.2.3.2.3 单元平均值和极差的计算

见表 10。

表 10 单元平均值和极差

实验室	单元平均值	极差
1	418.5	25
2	449	12
3	409	44
4	494	16
5	445	22
6	375.5	47

7.2.3.2.4 实验室内精密度的评定

用以下公式将表 10 中的极差与重复性标准差进行比较：

$$\frac{(y_{i1} - y_{i2})^2}{2\sigma_r^2} \leq \chi_{(1-\alpha)}^2(\nu)$$

当  $\alpha=0.05$  且  $\nu=1$  时,  $\chi_{0.95}^2(1)=3.841$

第 6 实验室的测试结果计算得到的检验值为:

$$\frac{(y_{61} - y_{62})^2}{2\sigma_r^2} = \frac{2209}{2 \times 16^2} = 4.31$$

因而该实验室为离群实验室。

7.2.3.2.5 偏倚的评定

式(4)的接收准则为:

$$|\bar{y} - 425| < 44.59$$

对第 4 实验室, 检验值为:

$$|\bar{y}_4 - 425| = 69$$

对第 6 实验室, 检验值为:

$$|\bar{y}_6 - 425| = 50.5$$

故上述两实验室的偏倚都不满足要求。

7.2.4 无标准物料时的测量方法

7.2.4.1 当没有标准物料可用时, 必须将待评实验室与一个高质量的实验室进行比较来评定。为了对待评实验室得出可靠的结论, 核心问题是找到一个精密度和偏倚都满意的实验室。

与标准物料存在的情形类似, 有时需要引入一个可检出的两个实验室偏倚之差值  $\lambda$ 。它是试验者希望能以高概率检出的两个实验室测试结果期望值之差的最小值。

7.2.4.2 将测试物料按 7.2.3.1.2 派送到两个实验室, 每个实验室内的精密度评定也按类似方法进行。两个实验室在每一水平的测试结果数  $n$  最好也相同。

7.2.4.3 在评定测量方法的偏倚  $\delta$  时, 要把两个实验室每一水平上所得测试结果的算术平均值进行比较。对一般情形, 记两个实验室所得测试结果数分别为  $n_1, n_2$ , 则有:

$$\begin{aligned} s_{(y(1)-y(2))}^2 &= 2\sigma_L^2 + \sigma_r^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \\ &= 2 \left[ \sigma_R^2 - \sigma_r^2 \left( 1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right) \right] \dots\dots\dots (6) \end{aligned}$$

接收准则为:

$$|\bar{y}_1 - \bar{y}_2| \leq 2\sqrt{2} \sqrt{\sigma_R^2 - \sigma_r^2 \left( 1 - \frac{1}{2n_1} - \frac{1}{2n_2} \right)} \dots\dots\dots (7)$$

$q$  个水平中的每一个都应满足接收准则(7)。

当  $n_1 = n_2 = 2$  时, 接收准则(7)化简为:

$$|\bar{y}_1 - \bar{y}_2| \leq 2\sqrt{2} \sqrt{\sigma_R^2 - \frac{\sigma_r^2}{2}} \dots\dots\dots (8)$$

7.3 对已认可实验室的再评定

7.3.1 连续对照试验的一般考虑

为了确保已通过评定的, 即已认可实验室的工作一直满足要求, 需要通过现场检查或参与评定试验来对实验室进行再评定。由于受到技术、经济和安全等诸多因素的影响, 实验室应间隔多长时间进行一次再评定, 没有简单硬性的规则可循, 负责部门应根据情况决定评定的频率。

再评定常会遇到需对多个实验室进行同时评定的情形。因此不建议用将每个实验室与一个高质量实验室进行比较的做法, 因为即使最好的实验室本身也必须进行检查。在此情况下, 有必要进行协同评定试验。

7.3.2 对实验室运作的评估

对实验室运作的评估, 按 7.2 所述的现场检查方式进行。

7.3.3 有标准物料时的测量方法

GB/T 6379.4 所描述的方法可相应地用于实验室的再评定。

7.3.4 无标准物料的测量方法

7.3.4.1 总则

7.3.4.1.1 在无标准物料可用的情形,每一个实验室的评定需通过由若干个实验室参与的协同评定试验进行。

安排一个评定试验与安排一个精密度试验非常相似,因此 GB/T 6379.1 和 GB/T 6379.2 提到的许多考虑对安排评定试验也适用。由于目标是对每一实验室进行评定,所以在每一水平上选择测试重复次数的方法与 7.2.2 中所述的一个实验室情形类似。

鉴于最终目的是对实验室进行评定,所以参与协同评定试验的实验室个数可少于精密度试验中的实验室数。例如,一种显而易见的方法是只有本国实验室参与试验。尤其重要的是参与试验的实验室数减少并不降低实验室间的系统偏差,因为系统偏差的降低会增加检出离群实验室的困难。

7.3.4.1.2 经过 7.2.2 提到的各种考虑,测试物料以  $q$  个水平派送到  $p$  个实验室,每一水平要求进行  $n$  次测量。用 GB/T 6379.2—2004 第 7 章描述的方法对测试结果进行估计。由于测试过程中可能产生测试结果的缺失或增加,各单元的结果数可能不同。

每个实验室内部精密度按第 6 章描述的方法进行评定。

7.3.4.1.3 为给出对偏倚的整体评定,对每一测试水平计算再现性方差(见 GB/T 6379.2—2004 的 7.5)。

$$s_R^2 = s_L^2 + s_p^2 \quad \dots\dots\dots(9)$$

其中

$$s_L^2 = \left\{ \frac{1}{p-1} \left[ \sum_{i=1}^p n_i (\bar{y}_i - \bar{y})^2 \right] - s_p^2 \right\} / \bar{n} \quad \dots\dots\dots(10)$$

而

$$\bar{n} = \frac{1}{p} \sum_{i=1}^p n_i \quad \dots\dots\dots(11)$$

将算得的实验室间方差的估计值  $s_L^2$  与已知的实验室间方差  $\sigma_L^2$  比较,按如下的接收准则进行检验:

$$\frac{\bar{n}s_L^2 + s_p^2}{\bar{n}\sigma_L^2 + \sigma_p^2} \leq \frac{\chi_{(1-\alpha)}^2(\nu)}{\nu} \quad \dots\dots\dots(12)$$

其中  $\chi_{(1-\alpha)}^2(\nu)$  是  $\chi^2$  分布的  $1-\alpha$  分位点,自由度  $\nu = p-1$ 。除非特别说明,显著性水平  $\alpha$  假定为 0.05。

如果接收准则(12)成立,则认为试验室间方差的估计值  $s_L^2$  是可接收的,并由此推断出所有实验室在所考虑的水平上获得了足够准确的测试结果。

如果接收准则(12)不成立,则通过计算格拉布斯检验统计量找出偏离最大的离群观测值,并剔除该实验室在这个水平上的所有测试结果,用剩下的  $(p-1)$  个实验室的测试结果重新估计实验室方差。如果这次的方差估计值满足接收准则(12),则这  $(p-1)$  个实验室通过评定,否则再次计算格拉布斯检验统计量。必要时可多次重复上述步骤。GB/T 6379.2 中曾提到,格拉布斯检验不宜重复应用。因此,当离群值较多时,应在所有水平上对所有数据进行检查。如果同样实验室在几个水平上偏离都大,则可以断定这些实验室的测量的偏倚过大。如果这些实验室仅在某一个水平上偏离较大,就有理由检查一下测试物料是否不均匀。如果偏离在多个实验室的多个水平上发生,可能是评定试验有缺陷。为了对此作出可能的解释,有必要对评定试验的每个环节进行严格检查。

如果某个实验室已被认为是离群实验室(无论是对实验室内精密度或偏倚),应将结果通知该实验室,并对实验室使用的方法进行检查,以改进该实验室的工作。

7.3.4.1.4 在进行再评定试验时,要采用不同的测试物料,以避免某些实验室对某些特定的测试物料有特别高的精密度,从而影响评定效果。此外,测试物料按 7.2.2 中的规定匿名地分发给各参与试验的实验室,以确保各实验室不对测试物料以特殊对待。

如果在一次评定试验产生的测试结果与以前试验得到的测试结果差异太大,重要的是分析可得到的所有信息,对这些意料之外的观测值找到可能的解释。

7.3.4.2 例:水的含碱量分析

7.3.4.2.1 背景

在对水质进行控制时,由多个实验室对水进行化学分析。这些实验室要得到认可,必须经过多次评估。本例中,用电位滴定法确定水的含碱总量,因无标准物料可用,所以评定必须通过评定试验来进行。

共有 18 个实验室参加评定试验,试验设两个水平,每个实验室在每一水平上进行两次测定。

7.3.4.2.2 原始数据

原始数据见表 11。

表 11 水的含碱量

实验室	水平		实验室	水平	
	1	2		1	2
1	2.040	5.250	10	2.170	5.520
	2.040	5.300		2.200	5.330
2	2.100	5.460	11	1.980	4.990
	2.110	5.460		1.940	5.020
3	2.070	5.240	12	2.120	5.340
	2.070	5.200		2.110	5.330
4	2.070	5.308	13	2.160	5.330
	2.090	5.292		2.150	5.420
5	2.740	5.850	14	2.050	5.330
	2.610	5.850		2.070	5.330
6	2.086	5.305	15	2.070	5.387
	2.182	5.325		2.056	5.335
7	2.128	5.296	16	2.010	5.210
	2.076	5.346		2.030	5.330
8	2.060	5.340	17	2.066	5.300
	2.080	5.340		2.070	5.280
9	2.060	5.310	18	2.060	5.300
	2.080	5.300		2.070	5.280

7.3.4.2.3 单元平均值和极差的计算

表 12 和表 13 分别给出了表 11 数据的单元平均值和单元极差。

表 12 表 11 数据的单元平均值

实验室	水平	
	1	2
1	2.040	5.275
2	2.105	5.460
3	2.070	5.220
4	2.080	5.300
5	2.675	5.850
6	2.134	5.315
7	2.102	5.321
8	2.070	5.340
9	2.070	5.305
10	2.185	5.425
11	1.960	5.005
12	2.115	5.335
13	2.155	5.375
14	2.060	5.330
15	2.063	5.361
16	2.020	5.270
17	2.068	5.290
18	2.065	5.290

表 13 表 11 数据的单元极差

实验室	水平	
	1	2
1	0.000	0.050
2	0.010	0.000
3	0.000	0.040
4	0.020	0.016
5	0.130	0.000
6	0.096	0.020
7	0.052	0.050
8	0.020	0.000
9	0.020	0.010
10	0.030	0.190
11	0.040	0.030
12	0.010	0.010
13	0.010	0.090
14	0.020	0.000
15	0.014	0.052
16	0.020	0.120
17	0.004	0.020
18	0.010	0.020

在所定的两个水平上重复性标准差和再现性标准差的事先确定值分别为：

$$\begin{aligned}\sigma_{r1} &= 0.023, \sigma_{r2} = 0.027 \\ \sigma_{R1} &= 0.045, \sigma_{R2} = 0.052\end{aligned}$$

#### 7.3.4.2.4 对实验室内精密度的评定

用如下公式将表 13 中所列的极差与重复性标准差进行比较：

$$w_{ij}^2 / 2\sigma_{rj}^2 < \chi_{(1-\alpha)}^2(\nu) / \nu$$

取  $\alpha=0.05, \nu=1, \chi_{0.95}^2(\nu) / \nu = 3.841$

运用上述准则，对水平 1，如下实验室由于检验值超出临界值(3.841)被判定为离群：

实验室 5:  $w^2=0.0169$ , 检验值=15.974

实验室 6:  $w^2=0.009216$ , 检验值=8.711

对水平 2，如下实验室由于检验值超出临界值被判定为离群：

实验室 10:  $w^2=0.0361$ , 检验值=24.76

实验室 13:  $w^2=0.0081$  检验值=5.55

实验室 16:  $w^2=0.0144$  检验值=9.88

#### 7.3.4.2.5 实验室偏倚的评定

根据表 12 的数据，按如下公式计算实验室间方差：

$$s^2 = \frac{1}{p-1} \sum_{i=1}^p n_i (\bar{y}_i - \bar{y})^2 = \bar{n}s_L^2 + s_r^2$$

对水平 1，求得：

$$n\sigma_L^2 + \sigma_r^2 = n\sigma_R^2 - (n-1)\sigma_r^2 = 0.003521$$

$$s^2 = 0.04436, \text{检验值} = 12.60$$

取  $\alpha=0.05, \nu=17, \chi_{0.95}^2(\nu) / \nu = 1.623$

发现实验室 5 的偏离最远。实验室 5 的格拉布斯检验值为：

$$G = (2.675 - 2.113) / 0.1489 = 3.77$$

用 GB/T 6379.2—2004 第 9 章的结果，在  $p=18$  时，5%水平的临界值为 2.651，将  $G=3.77$  与之相比较，判定实验室 5 为离群实验室。

用实验室 5 以外的其他实验室的测试结果计算得：

$$s^2 = 0.005357,$$

$$\text{检验值} = 1.521$$

取  $\alpha=0.05, \nu=16, \chi_{0.95}^2(\nu) / \nu = 1.644$ 。因此得出结论：除第 5 实验室以外的其他实验室在水平 1 下所得测试结果都有足够的准确度。

对水平 2，求得：

$$n\sigma_L^2 + \sigma_r^2 = n\sigma_R^2 - (n-1)\sigma_r^2 = 0.004679$$

$$s^2 = 0.05034, \text{检验值} = 10.758$$

其中  $\alpha=0.05, \nu=17, \chi_{0.95}^2(\nu) / \nu = 1.623$ 。

发现实验室 5 的偏离最远。实验室 5 的格拉布斯检验值为：

$$G = (5.85 - 5.337) / 0.1586 = 3.235$$

$p=18$  时，5%水平下的临界值为 2.651，将  $G=3.235$  与之相比较，判定实验室 5 为离群实验室。

用实验室 5 以外的其他实验室的测试结果计算得：

$$s^2 = 0.01867, \text{检验值} = 3.990$$

其中  $\alpha=0.05, \nu=16, \chi_{0.95}^2(\nu) / \nu = 1.644$ 。

发现实验室 11 的偏离最远。实验室 11 的格拉布斯检验值为：



$$G = (5.005 - 5.3069) / 0.09661 = -3.125$$

$p=17$  时, 5% 水平下的临界值为 2.620, 将  $G = -3.125$  与之相比较, 判定实验室 11 为离群实验室。

用实验室 11 以外的其他实验室的测试结果计算得:

$$s^2 = 0.00700, \text{检验值} = 1.496$$

取  $\alpha = 0.05, \nu = 15, \chi_{0.95}^2(\nu) / \nu = 1.666$ 。

因此, 除第 5 实验室和实验室 11 以外的其他实验室在水平 2 下所得测试结果有足够的准确度。

#### 7.3.4.2.6 结论

评定试验表明, 有多个实验室的内部精密度不好, 它们是第 5, 6, 10, 13 和 16 实验室; 而第 5 和 11 两个实验室在一个或两个水平上有较大的偏倚。应把评定结果告知相关实验室。

## 8 与可替代测量方法的比较

### 8.1 考虑可替代测量方法的原因

国际标准方法是指已经标准化的, 且能满足多方面要求的方法。这些要求包括:

- 测量方法应适用于测试特性的较大水平范围, 覆盖国际贸易中大多数物料。例如, 一种测定铁矿石中含铁总量的方法应当对尽可能多的国际贸易中的铁矿石适用;
- 测量所涉及的设备、试剂和人员在国际上能够找到;
- 测量费用在可接受的范围内;
- 测量方法的正确度和精密度应被结果的用户所认可。

标准测量方法可能因为过于繁琐而不适用于日常工作, 因此往往需要折中处理。某个特定实验室可能会找到一个比较简单但足够满足自身要求的方法。例如, 当测试物料来源相同, 其特性变异相对很小时, 使用一个简单花费较少的方法就够了。

由于历史原因, 某些地区可能更愿意采用某些测量方法。在此情形下, 也许希望有另外的测量方法来替代国际标准方法。

本章描述的比较是基于一个测试样本结果的。在比较两种测量方法的精密度和正确度时, 要求使用的测试样本应多于一个。比较所需的测试样本数依赖于多种因素, 如所关心的特性水平的范围及测量方法对样本成分变化的灵敏程度等。

### 8.2 比较测量方法的目的

8.2.1 8.2 描述比较两种测量方法的精密度和正确度的程序, 所比较的两种方法中的一种(称为方法 A)是国际标准方法或国际标准方法的首要候选方法。所作比较只提供两种方法的精密度和(或)正确度是否不同的证据, 不作其中一种方法比另一种方法更适宜于某种特定情形的建议。实际中使用哪种方法, 要结合费用及可利用的设备等多种因素而定。

8.2.2 8.2 内容主要是对以下应用设计的:

- 在一个国际标准方法的形成过程中, 技术委员会有时会面对这样问题: 在候选方法中, 哪一种最为适宜采用为国际标准方法? 方法的精密度和正确度是选择准则中的基础部分。
- 某些时候发现有必要开发可替代国际标准方法的另一种方法。候选方法应与国际标准方法一样准确。方法比较程序有助于确定候选方法是否满足要求。
- 对某些实验室而言, 大部分待测样本来源相同, 这些样本总体上有着非常一致的成分。在这种情况下, 若将国际标准方法作为常规方法采用会造成不必要的浪费。这时实验室可能需要有一个较简单的常规方法。用这种方法测得的结果应与现有国际标准方法具有相同正确度和精密度。

### 8.3 方法 B: 作为候选的可替代标准方法(“标准化试验”未确立)

对方法 A 和方法 B 的比较要依据精密度试验结果进行。如果方法 A 是一种已完全确立的标准方

法,它的精密度可作为比较的基础。如果方法 A 乃是一个尚在完善过程中的标准方法,它本身也应进行精密度试验。方法 A 和方法 B 的精密度试验均应根据 GB/T 6379.2 进行。

试验的目的如下:

- a) 确定方法 B 是否与方法 A 具有相同的精密度。试验结果应能检测出方法 B 和方法 A 的精密度度量之比是否大于某一特定值;
- b) 确定方法 B 是否与方法 A 具有相同的正确度。可通过精密度试验按以下方式确定:当分别使用两种方法对一样本进行测量时,所得测试结果的总平均值在统计上没有显著差异;或者当使用已定值的标准物料作为测试样本时,方法 B 在精密度试验所得的测试结果总平均值与标准物料的定值之间在统计上没有显著差异。

此外,试验应能检测出两种测量方法所得测试结果的期望值之差或每一方法所得测试结果的期望值与定值之差是否大于某一特定值。

#### 8.4 准确度试验

##### 8.4.1 一般要求

准确度试验应按 GB/T 6379.1 规定的总则进行。

应用文件形式明确两种测量方法的操作程序,对各个细节给予充分的说明,以免参与试验的实验室误解。在试验中对测量的程序不允许作任何更改。

参与试验的实验室应为测量方法所有潜在用户的一个代表性样本。

##### 8.4.2 测试样本

许多测量方法的精密度都要受测试样本的基体和测试特性水平的影响,对这些方法的精密度进行比较时,最好使用同一测试样本。另外,在对测量方法的正确度进行比较时也只能使用同一测试样本。鉴于此原因,应指定一位执行负责人,负责协调每种测量方法准确度试验的各工作组之间的事务和交流。

对测试样本的主要要求是均匀性,即每一实验室应该使用同一测试样本。如果怀疑单元内样本是非均匀的,应在文件中对选择试样的方法给予明确的说明。对一些测试样本,使用标准物料(RM)有其优点。在标准物料的均匀性已经得到保证的前提下,可以检查测试结果相对标准物料的定值的偏倚。使用标准物料的缺点是通常费用较高。在许多情况下,这可通过对标准物料单元进行再细分来克服。关于用某一标准物料作为测试样本的方法,参见 ISO 指南 33。

##### 8.4.3 测试样本的数量

试验中所使用的测试样本的数量,随所测特性的水平范围及准确度与测试水平之间依赖程度而变化。在许多情形,测试样本数量还受测量工作量以及在要求水平上能否找到测试样本等因素的制约。

##### 8.4.4 实验室数及测量次数

###### 8.4.4.1 总则

在实验室间试验中,两种测量方法所需的实验室数以及每个实验室所需测量次数依赖于:

- a) 两种测量方法的精密度;
- b) 两种测量方法精密度度量的可测出比  $\rho$  或者  $\phi$ ,即试验者希望能以高概率检测到从两种测量方法的测试结果所得的精密度度量之比的最小值。当精密度以重复性标准差表示时,该比值记为  $\rho$ ;当精密度以实验室间均方的平方根表示时,该比值记为  $\phi$ 。
- c) 两种测量方法偏倚的可检出的差  $\lambda$ ,即能检测到从两种测量方法所得测试结果期望值之差的最小值。

建议使用  $\alpha=0.05$  的显著水平来比较精密度的估计值。把不能测出选定的标准差之比最小值的风险,或偏倚之差的最小值的风险定为  $\beta=0.05$ 。

规定了  $\alpha$  与  $\beta$  的值,可用下面的公式计算可检出的差  $\lambda$ :

$$\lambda = 4 \sqrt{(\sigma_{LA}^2 + \sigma_{rA}^2/n_A)/p_A + (\sigma_{LB}^2 + \sigma_{rB}^2/n_B)/p_B} \dots\dots\dots(13)$$

其中下标 A 和 B 分别代表方法 A 和方法 B。

在大多数情形,方法 B 的精密度未知。对此种情形,可用方法 A 的精密度代替:

$$\lambda = 4 \sqrt{(\sigma_{LA}^2 + \sigma_{rA}^2/n_A)/p_A + (\sigma_{LA}^2 + \sigma_{rA}^2/n_B)/p_B} \dots\dots\dots(14)$$

试验者应把  $n_A, n_B, p_A$  和  $p_B$  的不同取值(从小到大)代入公式(13)或(14),直到找出能满足等式的  $n_A, n_B, p_A$  和  $p_B$ 。这些参数值是考虑安排一个适当的比较精密度估计的试验所需要的。

表 14 列出了在给定  $\alpha$  和  $\beta$  值的情况下,标准差之比作为自由度  $\nu_A$  和  $\nu_B$  的函数的最小值。(表 14 按 ISO 于 2001-10-15 发布的技术修改单改)

表 14  $\rho(\nu_A, \nu_B, \alpha, \beta)$  或  $\phi(\nu_A, \nu_B, \alpha, \beta)$  的值 ( $\alpha=0.05, \beta=0.05$ )

$\nu_B$	$\nu_A$																	
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	50	200
6	4.99	4.64	4.39	4.21	4.07	3.95	3.86	3.79	3.72	3.67	3.62	3.58	3.54	3.51	3.48	3.37	3.17	3.02
7	4.69	4.35	4.11	3.93	3.79	3.68	3.59	3.52	3.45	3.40	3.35	3.31	3.28	3.25	3.22	3.11	2.91	2.77
8	4.48	4.14	3.90	3.73	3.59	3.48	3.40	3.32	3.26	3.21	3.16	3.12	3.09	3.06	3.03	2.93	2.72	2.58
9	4.32	3.99	3.75	3.58	3.44	3.34	3.25	3.18	3.12	3.06	3.02	2.98	2.94	2.91	2.89	2.78	2.58	2.44
10	4.19	3.86	3.63	3.46	3.33	3.22	3.13	3.06	3.00	2.95	2.91	2.87	2.83	2.80	2.77	2.67	2.47	2.33
11	4.09	3.77	3.54	3.37	3.23	3.13	3.04	2.97	2.91	2.86	2.81	2.78	2.74	2.71	2.68	2.58	2.38	2.24
12	4.01	3.69	3.46	3.29	3.16	3.05	2.97	2.90	2.84	2.78	2.74	2.70	2.67	2.64	2.61	2.51	2.31	2.16
13	3.94	3.62	3.39	3.22	3.09	2.99	2.90	2.83	2.77	2.72	2.68	2.64	2.60	2.57	2.55	2.44	2.24	2.10
14	3.88	3.56	3.34	3.17	3.04	2.94	2.85	2.78	2.72	2.67	2.62	2.59	2.55	2.52	2.49	2.39	2.19	2.04
15	3.83	3.52	3.29	3.12	2.99	2.89	2.80	2.73	2.67	2.62	2.58	2.54	2.51	2.47	2.45	2.34	2.14	1.99
16	3.79	3.47	3.25	3.08	2.95	2.85	2.76	2.69	2.63	2.58	2.54	2.50	2.47	2.43	2.41	2.30	2.10	1.95
17	3.75	3.44	3.21	3.05	2.92	2.81	2.73	2.66	2.60	2.55	2.50	2.46	2.43	2.40	2.37	2.27	2.07	1.92
18	3.72	3.41	3.18	3.02	2.89	2.78	2.70	2.63	2.57	2.52	2.47	2.43	2.40	2.37	2.34	2.24	2.03	1.88
19	3.69	3.38	3.15	2.99	2.86	2.75	2.67	2.60	2.54	2.49	2.44	2.41	2.37	2.34	2.31	2.21	2.01	1.85
20	3.67	3.35	3.13	2.96	2.83	2.73	2.65	2.58	2.52	2.46	2.42	2.38	2.35	2.32	2.29	2.18	1.98	1.83
25	3.57	3.25	3.03	2.87	2.74	2.64	2.55	2.48	2.42	2.37	2.33	2.29	2.25	2.22	2.19	2.09	1.88	1.72
50	3.37	3.07	2.85	2.68	2.55	2.45	2.37	2.30	2.24	2.18	2.14	2.10	2.06	2.03	2.00	1.89	1.87	1.50
200	3.24	2.93	2.71	2.55	2.42	2.32	2.23	2.16	2.10	2.04	2.00	1.96	1.92	1.89	1.86	1.75	1.51	1.29

注:

- 1  $\rho = \frac{\sigma_{rB}}{\sigma_{rA}}; \nu_A = p_A(n_A - 1); \nu_B = p_B(n_B - 1)$
- 2  $\phi = \sqrt{\frac{n_B \sigma_{LB}^2 + \sigma_{rB}^2}{n_A \sigma_{LA}^2 + \sigma_{rA}^2}}; \nu_A = p_A - 1; \nu_B = p_B - 1$

当使用重复性标准差时,自由度为:

$$\nu_A = p_A(n_A - 1), \nu_B = p_B(n_B - 1)$$

当使用实验室间均方时,自由度为:

$$\nu_A = p_A - 1, \nu_B = p_B - 1$$

如果两种方法之一的精密度已完全确立,则用表 14 中自由度等于 200 的数值。

## 8.4.4.2 例:铁矿石中铁含量的测定

## 8.4.4.2.1 背景

考察了两种测定铁矿石中铁含量的分析方法,假定它们有相同的精密度:

$$\sigma_{rA} = \sigma_{rB} = 0.1\%$$

$$\sigma_{LA} = \sigma_{LB} = 0.2\%$$

## 8.4.4.2.2 要求

取  $\lambda = 0.4\%$ ,  $\rho = \phi = 4$ 。在计算所需最小实验室数时,假定两种方法取相等数目的实验室,每个试样分析两次,即:

$$p_A = p_B, n_A = n_B = 2$$

## a) 正确度试验要求:

$$0.4 = 4 \sqrt{(0.2^2 + 0.1^2/2)/p_A + (0.2^2 + 0.1^2/2)/p_B}$$

求得

$$p_A = p_B = 9$$

## b) 精密度试验要求:

从表 14 可以看出,当  $\nu_A = \nu_B = 9$  时,  $\rho$  (或  $\phi$ )  $\approx 4$ 。

比较重复性标准差时有:  $\nu_A = p_A, \nu_B = p_B$ , 故  $p_A = p_B = 9$ 。

比较实验室间均方时有:  $\nu_A = p_A - 1, \nu_B = p_B - 1$ , 故  $p_A = p_B = 10$ 。

## 8.4.4.2.3 结论

每个实验室间试验项目所需参与实验室数的最小值为 10。

## 8.4.5 测试样本的分发

实验室间试验项目的协调员应对获取、制备、分发测试样本担负最终责任。应确保参与实验室间试验的各实验室收到的测试样本状态良好,标记明确。应告知所有参与试验的实验室在相同条件下对样本进行分析。譬如,“在干燥条件下”表示样本称重前需在 105 °C 下干燥若干小时等。

## 8.4.6 参与试验的实验室

每一参与实验室间试验的实验室,应当安排一名成员负责组织实施试验协调员的指令。这名实验室成员应是合格的分析师。不应安排技术特别突出的人员(例如研究人员或“最佳”操作者)担任此职,以免得到一个低得不符合实际的标准差估计。被指定的实验室成员应在重复性条件下进行所要求次数的测量,而实验室负责在规定时间内向协调员报告测试结果。

## 8.4.7 测试结果的收集

每一方法试验项目的协调员要负责在合理的时间内收集到所有的测试结果。

协调员的职责是对测试结果进行仔细检查,以查出物理异常值,即查出那些由于可解释的物理原因而与其他测试结果不属于同一分布的测试结果。

## 8.4.8 测试结果的计算

测试结果应由合格的统计人员按 GB/T 6379.2 所描述的方法进行计算处理。对每一测试样本,要计算以下诸量:

$s_{rA}$  方法 A 的重复性标准差估计

$s_{rB}$  方法 B 的重复性标准差估计

$s_{RA}$  方法 A 的再现性标准差估计

$s_{RB}$  方法 B 的再现性标准差估计

$\bar{y}_A$  方法 A 所得测试结果的总平均值

$\bar{y}_B$  方法 B 所得测试结果的总平均值

## 8.4.9 方法 A 与方法 B 所得测试结果的比较

应在每一水平上对实验室间试验项目的测试结果进行比较。方法 B 在测试特性的较低水平上可

能准确度较高,而在测试特性的较高水平上准确度较低;或者正好相反。

#### 8.4.9.1 图形表示

可以将对每一水平的原始数据用图形来表示。有时通过图形,从精密度和(或)偏倚来看,两种方法的测试结果之间的差异非常明显,无须做进一步的统计评估。

所有水平的精密度与总平均值的图形表示也是需要的。

#### 8.4.9.2 精密度的比较

##### 8.4.9.2.1 方法 A 为已确立的标准方法情形

此时,方法 A 的精密度已经完全确定。

##### a) 实验室内精密度

若

$$\frac{s_{rB}^2}{\sigma_{rA}^2} \leq \frac{\chi_{(1-\alpha)}^2(\nu_{rB})}{\nu_{rB}}$$

则没有证据表明方法 B 的实验室内精密度不如方法 A;

若

$$\frac{s_{rB}^2}{\sigma_{rA}^2} > \frac{\chi_{(1-\alpha)}^2(\nu_{rB})}{\nu_{rB}}$$

则表明方法 B 的实验室内精密度比方法 A 的差。

这里  $\chi_{(1-\alpha)}^2(\nu_{rB})$  是自由度为  $\nu_{rB}$  的  $\chi^2$  分布的  $(1-\alpha)$  分位数,其

$$\nu_{rB} = p_B(n_B - 1)$$

##### b) 总精密度

若

$$\frac{s_{RB}^2 - (1 - 1/n_B)s_{rB}^2}{\sigma_{RA}^2 - (1 - 1/n_B)\sigma_{rA}^2} \leq \frac{\chi_{(1-\alpha)}^2(\nu_{LB})}{\nu_{LB}}$$

则没有证据表明方法 B 的均方不如方法 A;

若

$$\frac{s_{RB}^2 - (1 - 1/n_B)s_{rB}^2}{\sigma_{RA}^2 - (1 - 1/n_B)\sigma_{rA}^2} > \frac{\chi_{(1-\alpha)}^2(\nu_{LB})}{\nu_{LB}}$$

则表明方法 B 的均方不如方法 A。

这里  $\chi_{(1-\alpha)}^2(\nu_{LB})$  是自由度为  $\nu_{LB}$  的  $\chi^2$  分布的  $(1-\alpha)$  分位数,且

$$\nu_{LB} = p_B - 1$$

##### 8.4.9.2.2 两种方法均为新的候选标准方法情形

##### a) 实验室内精密度(按 ISO 于 2001-10-15 发布的技术修改单改)

令

$$F_r = \frac{s_{rB}^2}{s_{rA}^2}$$

若

$$F_{\alpha/2}(\nu_{rB}, \nu_{rA}) \leq F_r \leq F_{(1-\alpha/2)}(\nu_{rB}, \nu_{rA})$$

则没有证据表明两种方法的实验室内精密度有显著差异;

若

$$F_r < F_{\alpha/2}(\nu_{rB}, \nu_{rA})$$

则表明方法 B 的实验室内精密度优于方法 A;

若

$$F_r > F_{(1-\alpha/2)}(\nu_{rB}, \nu_{rA})$$

则表明方法 B 的实验室内精密度比方法 A 的差。

这里  $F_{\alpha/2}(\nu_{RB}, \nu_{RA})$  和  $F_{(1-\alpha/2)}(\nu_{RB}, \nu_{RA})$  分别是分子自由度为  $\nu_{RB}$ , 分母自由度为  $\nu_{RA}$  的  $F$  分布的  $\alpha/2$  和  $1-\alpha/2$  分位数, 且

$$\nu_{RB} = p_B(n_B - 1)$$

$$\nu_{RA} = p_A(n_A - 1)$$

b) 总精密度

令

$$F_R = \frac{s_{RB}^2 - (1 - 1/n_B)s_{rB}^2}{s_{RA}^2 - (1 - 1/n_A)s_{rA}^2}$$

若

$$F_{\alpha/2}(\nu_{RB}, \nu_{RA}) \leq F_R \leq F_{(1-\alpha/2)}(\nu_{RB}, \nu_{RA})$$

则没有证据表明两种方法的实验室间精密度有显著差异;

若

$$F_R < F_{\alpha/2}(\nu_{RB}, \nu_{RA})$$

则表明方法 B 的总精密度要好于方法 A;

若

$$F_R > F_{(1-\alpha/2)}(\nu_{RB}, \nu_{RA})$$

则表明方法 B 的总精密度比方法 A 差;

这里  $F_{\alpha/2}(\nu_{RB}, \nu_{RA})$  和  $F_{(1-\alpha/2)}(\nu_{RB}, \nu_{RA})$  分别是分子自由度为  $\nu_{RB}$ , 分母自由度为  $\nu_{RA}$  的  $F$  分布的  $\alpha/2$  和  $1-\alpha/2$  分位数, 且

$$\nu_{RB} = p_B - 1$$

$$\nu_{RA} = p_A - 1$$

注 5: 在许多  $F$  分布表中只列出了  $1-\alpha/2$  分位数, 此时可利用下面关系推出  $\alpha/2$  分位数:

$$F_{\alpha/2}(\nu_{RB}, \nu_{RA}) = 1/F_{(1-\alpha/2)}(\nu_{RA}, \nu_{RB})$$

$$F_{\alpha/2}(\nu_{RB}, \nu_{RA}) = 1/F_{(1-\alpha/2)}(\nu_{RA}, \nu_{RB})$$

### 8.4.9.3 正确度的比较

#### 8.4.9.3.1 平均值与标准物料定值的比较

当测试样本中的一个使用标准物料时, 使用如下的检验可将每种方法的总平均值与标准物料的定值进行比较:

a) 若

$$|\mu - \bar{y}| \leq 2 \sqrt{[s_{RB}^2 - (1 - 1/n_B)s_{rB}^2]/p_B}$$

则认为该方法的结果的总平均值与标准物料的定值的差异不显著;

b) 若

$$|\mu - \bar{y}| > 2 \sqrt{[s_{RB}^2 - (1 - 1/n_B)s_{rB}^2]/p_B}$$

则认为该方法测得结果的总平均值与标准物料的定值差异显著。

当二者差异显著时, 又有以下两种可能:

1) 若

$$|\mu - \bar{y}| \leq \delta_m/2$$

则没有证据表明该测量方法的偏倚是不可接受的;

2) 若

$$|\mu - \bar{y}| > \delta_m/2$$

则表明该测量方法的偏倚是不可接受的。

这里  $\delta_m$  是指试验者希望根据某一测量方法测试结果, 能以高概率检测出的测试结果期望

值与标准物料定值之差的最小值。

#### 8.4.9.3.2 方法 A 与方法 B 平均值的比较

a) 若

$$\left| \frac{\bar{y}_A - \bar{y}_B}{s} \right| \leq 2.0$$

则方法 A 与方法 B 平均值之间差异不显著；

b) 若

$$\left| \frac{\bar{y}_A - \bar{y}_B}{s} \right| > 2.0$$

则方法 A 与方法 B 平均值有显著差异；其中

$$s = \sqrt{s_A^2 + s_B^2}$$

$$s_A^2 = [s_{RA}^2 - (1 - 1/n_A)s_{rA}^2] / p_A$$

$$s_B^2 = [s_{RB}^2 - (1 - 1/n_B)s_{rB}^2] / p_B$$

当二者之间差异显著时，又有以下两种可能：

1) 若

$$|\bar{y}_A - \bar{y}_B| \leq \lambda/2$$

则没有证据表明两种测量方法偏倚的差是不可接受的；

2) 若

$$|\bar{y}_A - \bar{y}_B| > \lambda/2$$

则表明两种测量方法偏倚的差是不可接受的。

这里  $\lambda$  为两种测量方法偏倚之间的可检出的差。

### 8.5 方法 B 作为候选的常规方法

#### 8.5.1 参数

对一个常规的实验室测量方法，比较关注的参数有：长期均值  $\mu_t$ ，重复性标准差  $\sigma_r$  和时间不同的中间精密度标准差  $\sigma_{IT}$ 。

为估计上述参数，实验室应进行一种准实验室间试验，用“时间”因素代替参与的实验室（见 ISO 5725-3）。表示准实验室间试验所用的数学模型与一般实验室间试验的模型一致，只是将下标  $L$  用  $T$  代替（即用时间代替实验室）。此时，由时间不同引起的变异包含了实验室内通常发生的各种变化所产生的变异，比如设备校准、不同试剂、不同分析员及试验环境所引起的变异等。准实验室间试验的时间段在通常情况下应能覆盖这些变化。比较精密度的方法与 8.4.9.3 中描述的方法一致。

将一种方法应用于已定值的标准物料，即可确定该方法的偏倚。令  $\mu$  是标准物料在接受值（定值）。

#### 8.5.2 长期偏倚的检验

计算长期算术平均值：

$$\bar{y} = \sum_{i=1}^{p_{IB}} \sum_{j=1}^{n_B} \frac{y_{ij}}{n_B p_{IB}}$$

其中  $i$  和  $j$  分别表示长期（中间精密度）和短期（重复性）。

a) 若

$$|\bar{y} - \mu_t| \leq 2 \sqrt{\left( s_{IB}^2 + \frac{s_{rB}^2}{n_B} \right) / p_{IB}}$$

则认为测试结果的长期平均值与标准物料在接受值之间的差异不显著；

b) 若

$$|\bar{y} - \mu_t| > 2 \sqrt{\left( s_{IB}^2 + \frac{s_{rB}^2}{n_B} \right) / p_{IB}}$$

则认为测试结果的长期平均值与标准物料的接收值之间的差异显著。

当认为二者之间差异显著时,又有以下两种可能:

1) 若

$$|\bar{y} - \mu_r| \leq \delta_m / 2$$

则没有证据表明测量方法的长期偏倚是不可接收的;

2) 若

$$|\bar{y} - \mu_r| > \delta_m / 2$$

则表明测量方法的长期偏倚是不可接收的。

$\delta_m$ 是由试验者事先设定的长期可检测的差异。



附录 A  
(规范性附录)

## GB/T 6379 所用的符号与缩略语

- a* 关系式  $s=a+bm$  中的截距
- A* 用来计算估计值的不确定度系数
- b* 关系式  $s=a+bm$  中的斜率
- B* 表示一个实验室测试结果与总平均值的偏差分量(偏倚的实验室分量)
- $B_0$  表示在中间精密度条件下所有因素皆保持不变时 *B* 的分量
- $B_{(1)}, B_{(2)}, \dots$  表示在中间精密度条件下, 因素发生改变时 *B* 的分量
- c* 关系式  $\lg s=c+d\lg m$  中的截距
- $C, C', C''$  检验统计量
- $C_{\text{crit}}, C'_{\text{crit}}, C''_{\text{crit}}$  用于统计检验的临界值
- $CD_P$  概率 *P* 的临界差
- $CR_P$  概率 *P* 的临界极差
- d* 关系式  $\lg s=c+d\lg m$  中的斜率
- e* 发生在每次测试结果中随机误差分量
- f* 临界极差系数
- $F_p(\nu_1, \nu_2)$  自由度为  $\nu_1$  和  $\nu_2$  的 *F* 分布的 *p* 分位数
- G* 格拉布斯检验统计量
- h* 曼得尔实验室间一致性检验统计量
- k* 曼得尔实验室内一致性检验统计量
- LCL 控制下限(行动限或警戒限)
- m* 测试特性的总平均值; 水平
- M* 在中间精密度条件中考虑的因素数
- N* 交互作用数
- n* 一个实验室在一个水平(即一个单元中)上的测试结果数
- p* 参加实验室间试验的实验室数
- P* 概率
- q* 在实验室间试验中测试特性的水平数
- r* 重复性限
- R* 再现性限
- RM 标准物料(标准物质/标准材料)
- s* 标准差的估计值
- $\hat{s}$  标准差的预测值
- T* 总和
- t* 测试目标个数或组数
- UCL 控制上限(行动限或警戒限)
- W* 加权回归中的权数
- w* 一组测试结果的极差
- x* 用于格拉布斯检验的数据
- y* 测试结果

- $\bar{y}$  测试结果的算术平均值
- $\bar{y}$  测试结果的总平均值
- $\alpha$  显著性水平
- $\beta$  第二类错误概率
- $\gamma$  再现性标准差与重复性标准差的比值( $\sigma_R/\sigma_r$ )
- $\Delta$  实验室偏倚
- $\hat{\Delta}$   $\Delta$  的估计值
- $\delta$  测量方法偏倚
- $\hat{\delta}$   $\delta$  的估计值
- $\lambda$  两个实验室偏倚或两个测量方法偏倚之间的可检出的差
- $\mu$  测试特性的真值或接受参照值
- $\nu$  自由度
- $\rho$  方法 A 和方法 B 的重复性标准差之间的可检出的比
- $\sigma$  标准差的真值
- $\tau$  表示从上次校准始由时间变化引起的测试结果变异的分量
- $\phi$  方法 A 和方法 B 的实验室间均方的平方根可检出的比
- $\chi_p^2(\nu)$  自由度为  $\nu$  的  $\chi^2$  分布的  $p$  分位数

用作下标的符号

- C 校准-不同
  - E 设备-不同
  - i* 实验室标识
  - $I(\ )$  精密度的中间度量;括号内表示中间情形类型
  - j* 水平的标识(GB/T 6379.2);测试或因素的标识(ISO 5725-3)
  - k* 实验室 *i*, 水平为 *j* 的测试结果的标识
  - L 实验室间
  - m* 可检出偏倚的标识
  - M 试样间
  - O 操作员-不同
  - r* 重复性
  - R 再现性
  - T 时间-不同
  - W 实验室内
  - 1,2,3,... 测试结果按获得顺序的编号
  - (1),(2),(3),... 测试结果按数值大小递增顺序的编号
-