



中华人民共和国国家标准

GB/T 6379.1—2004/ISO 5725-1:1994

部分代替 GB/T 6379—1986

GB/T 11792—1989

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

Accuracy(trueness and precision)of measurement methods and results—
Part 1:General principles and definitions

(ISO 5725-1:1994, IDT)

2004-06-02 发布

2005-01-01 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	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 定义	1
4 准确度试验定义的实际含义	4
4.1 标准测量方法	4
4.2 准确度试验	5
4.3 同一测试对象	5
4.4 短暂的时间间隔	5
4.5 参与的实验室	5
4.6 观测条件	5
5 统计模型	6
5.1 基本模型	6
5.2 基本模型和精密度的关系	7
5.3 其他可供选择的模型	7
6 为估计准确度试验设计方面的考虑	7
6.1 准确度试验的计划	7
6.2 标准测量方法	7
6.3 准确度试验的实验室的选择	8
6.4 用于准确度试验物料的选择	10
7 准确度数据的应用	11
7.1 正确度和精密度数值的发布	11
7.2 正确度和精密度数值的实际应用	12
附录 A (规范性附录) GB/T 6379 所用的符号与缩略语	13
附录 B (规范性附录) 精密度度量不确定度的图示	15
附录 C (资料性附录) 参考文献	16

前 言

GB/T 6379《测量方法与结果的准确度(正确度与精密度)》分为六部分,其预期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(对应 ISO 5725-3:1994)
- 第4部分:确定标准测量方法正确度的基本方法(对应 ISO 5725-4:1994)
- 第5部分:确定标准测量方法精密度的可替代方法(对应 ISO 5725-5:1998)
- 第6部分:准确度值的实际应用(对应 ISO 5725-6:1994)

本部分为 GB/T 6379 的第1部分。

GB/T 6379 的本部分等同采用国际标准 ISO 5725-1:1994《测量方法与结果的准确度(正确度与精密度)》——第1部分:总则与定义及 ISO 于 1998-02-15 发布的对 1994 版 ISO 5725-1 的技术修改单。

GB/T 6379 第1部分至第6部分作为一个整体代替 GB/T 6379—1986 和 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。

本部分的附录 A 和附录 B 为规范性附录,附录 C 为资料性附录。

本部分由中国标准化研究院提出。

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:中国标准化研究院、中国科学院数学与系统科学研究院、辽宁出入境检验检疫局、广州出入境检验检疫局。

本部分主要起草人:于振凡、冯士雍、刘文、姜健、丁文兴、王斗文、肖惠、李成明。

本部分于 2004 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。类似的,当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,则不能表示这两批物料的本质差别。

0.3 很多不同的因素(除假定相同的样品之间的差异外)都能够引起测量方法的结果变异,它们包括:

- a) 操作员;
- b) 使用的设备;
- c) 设备的校准;
- d) 环境(温度、湿度、空气污染等);
- e) 不同测量的时间间隔。

由不同操作员所做的测量和在不同设备上进行的测量通常要比在短时间内由同一个操作员使用相同的设备进行测量产生的变异大。

0.4 描述重复测量结果之间的变异的一般术语是精密度。精密度的两个条件,即重复性和再现性条件对很多实际情形是必需的,对描述测量方法的变异是有用的。在重复性条件下,上面所列的因素 a) 到 e) 皆保持不变,不产生变异;而在再现性条件下,它们是变化的,能引起测试结果的变异。因此重复性和再现性是精密度的两个极端情况:重复性描述变异最小情况,而再现性则描述变异最大情形。当因素 a) 到 e) 的一个或多个允许变化时,位于精密度的上述两个条件的其他中间条件也是可以想象的,它们可用于某些特定的环境。精密度通常用标准差表示。

0.5 当已知或可以推测所测量特性的真值时,测量方法的正确度即为人们所关注。尽管对某些测量方法,真值可能不会确切知道,但有可能知道所测量特性的一个接受参照值。例如,可以使用适宜的标准物料或者通过参考另一种测量方法或准备一个已知的样本来确定该接受参照值。通过把接受参照值与测量方法给出的结果水平进行比较就可以对测量方法的正确度进行评定。正确度通常用偏倚来表示。例如,在化学分析中,如果所用的测量方法不能测出某种元素的全部,或者由于一种元素的存在而干扰了另一种元素的确定,就会产生偏倚。

0.6 ISO 5725 中使用的一般术语“准确度”,既包含正确度也包含精密度。

“准确度”这一术语在过去一段时间只用来表示现在称为正确度的部分。但是对很多人来说,它不仅包括测试结果对参照(标准)值的系统影响,也应包括随机的影响。

很长时间以来,术语“偏倚”一直被限制用于统计问题,由于它在某些领域中(如医学界和法律界)曾经引起过哲学上的异议,因此引进术语“正确度”似更强调其正面含义。

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

1 范围

1.1 GB/T 6379 系列标准的目的如下:

- a) 阐述评定测量方法与结果的准确度(正确度与精密度)的一般原理与应用,通过试验对不同测量进行实际的估计(GB/T 6379.1);
- b) 通过试验提供估计测量方法精密度的两个极端度量的基本的方法(GB/T 6379.2);
- c) 提供为获得精密度中间度量的程序,给出其适用的环境和估计方法(ISO 5725-3);
- d) 提供为确定测量方法正确度的基本方法(ISO 5725-4);
- e) 为某些特定的应用条件,提供有别于 GB/T 6379.2 和 ISO 5725-4 中基本方法的确定测量方法的正确度与精密度的其他可替代的方法(ISO 5725-5);
- f) 提供正确度和精密度上述度量的一些实际应用。

1.2 GB/T 6379 的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。

GB/T 6379 的本部分定量定义了一种测量方法给出正确结果的能力(正确度)与重复同样结果的能力(精密度)。这就意味着可用完全相同的方法来测量完全相同的事物,且测量过程是受控的。

GB/T 6379 的本部分适用于多种范围的物料(物质或材料),包括液体、粉状物和固体物料,这些物料可以是人工制造的,也可以是自然存在的,只要对物料的异质性进行适当考虑。

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的一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 3358.2—1993 统计学术语 第二部分:统计质量控制术语

GB/T 3358.3—1993 统计学术语 第三部分:试验设计术语

GB/T 6379.2—2004 测量方法与结果的准确度(正确度和精密度) 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 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(正确度和精密度) 第3部分:标准测量方法精密度的中间度量

ISO 5725-4:1994 测量方法与结果的准确度(正确度和精密度) 第4部分:确定标准测量方法的正确度的基本方法

3 定义

下列定义适用于 GB/T 6379 的本部分,其中部分定义引自 ISO 3534-1。

GB/T 6379 使用的符号由附录 A 给出。

3.1

观测值 (observed value)

作为一次观测结果而确定的特性值。

(引自 GB/T 3358.1—1993)

3.2

测试结果 (test result)

用规定的测试方法所确定的特性值。

(改自 GB/T 3358.1—1993)

注1: 测试方法宜指明观测是一个还是多个, 报告的测试结果是观测值的平均数还是它的其他函数(例如中位数或标准差)。它可以要求按适用的标准进行修正, 如气体容积按标准温度和压力进行修正。因此一个测试结果可以是几个观测值计算的结果。在最简单情形, 测试结果即为观测值本身。

3.3

精密度试验的测试水平 (level of the test in a precision experiment)

对某测试物料或试样, 所有实验室测试结果的总平均值。

3.4

精密度实验单元 (cell in a precision experiment)

由一个实验室在单一水平获得的测试结果。

3.5

接受参照值 (accepted reference value)

用作比较的经协商同意的标准值, 它来自于:

- a) 基于科学原理的理论值或确定值;
- b) 基于一些国家或国际组织的实验工作的指定值或认证值;
- c) 基于科学或工程组织赞助下合作实验工作中的同意值或认证值;
- d) 当 a)b)c) 不能获得时, 则用(可测)量的期望, 即规定测量总体的均值。

3.6

准确度 (accuracy)

测试结果与接受参照值间的一致程度。

(改自 GB/T 3358.1—1993)

注2: 术语准确度, 当用于一组测试结果时, 由随机误差分量和系统误差即偏倚分量组成。

3.7

正确度 (trueness)

由大量测试结果得到的平均数与接受参照值间的一致程度。

注3: 正确度的度量通常用术语偏倚表示。

注4: 准确度曾被称为“平均数的准确度”, 这种用法不被推荐。

3.8

偏倚 (bias)

测试结果的期望与接受参照值之差。

注5: 与随机误差相反, 偏倚是系统误差的总和。偏倚可能由一个或多个系统误差引起。系统误差与接受参照值之差越大, 偏倚就越大。

3.9

实验室偏倚 (laboratory bias)

一个特定的实验室的测试结果的期望与接受参照值之差。

3.10

测量方法偏倚 (bias of the measurement method)

所有采用该方法的实验室所得测试结果的期望与接受参照值之差。

注 6：实际操作中的例。如测量某化合物中硫的含量，由于测量方法不可能提尽所有的硫，因此该测量方法将有一个负的偏倚。对很多使用相同的方法的不同的实验室得到的测试结果求平均值，就可用来测定该测量方法的偏倚。测量方法的偏倚在不同水平下可以是不同的。

3.11

偏倚的实验室分量 (laboratory component of bias)

实验室偏倚与测量方法偏倚之差。

注 7：偏倚的实验室分量是针对特定实验室和实验室所具有的测量条件的，在不同的测试水平下也可以是不同的。

注 8：偏倚的实验室分量与测试结果的总平均值有关，而与其真值或标准值无关。

3.12

精密度 (precision)

在规定条件下，独立测试结果间的一致程度。

(改自 GB/T 3358.1—1993)

注 9：精密度仅仅依赖于随机误差的分布而与真值或规定值无关。

注 10：精密度的度量通常以不精密度表达，其量值用测试结果的标准差来表示，精密度越低，标准差越大。

注 11：“独立测试结果”指的是对相同或相似的测试对象所得的结果不受以前任何结果的影响。精密度的定量的测度严格依赖于规定的条件，重复性和再现性条件为其中两种极端情况。

3.13

重复性 (repeatability)

在重复性条件下的精密度。

(改自 GB/T 3358.1—1993)

3.14

重复性条件 (repeatability conditions)

在同一实验室，由同一操作员使用相同的设备，按相同的测试方法，在短时间内对同一被测对象相互独立进行的测试条件。

(引自 GB/T 3358.1—1993)

3.15

重复性标准差 (repeatability standard deviation)

在重复性条件下所得测试结果的标准差。

(引自 GB/T 3358.1—1993)

注 12：重复性标准差是重复性条件下测试结果分布的分散性的度量。

注 13：类似地可定义“重复性方差”与“重复性变异系数”，作为重复性条件下测试结果分散性的度量。

3.16

重复性限 (repeatability limit)

一个数值，在重复性条件下，两个测试结果的绝对差小于或等于此数的概率为 95%。

(引自 GB/T 3358.1—1993)

注 14：重复性限用 r 来表示。

3.17

再现性 (reproducibility)

在再现性条件下的精密度。

(改自 GB/T 3358.1—1993)

3.18

再现性条件 (reproducibility conditions)

在不同的实验室,由不同的操作员使用不同设备,按相同的测试方法,对同一被测对象相互独立进行的测试条件。

(改自 GB/T 3358.1—1993)

3.19

再现性标准差 (reproducibility standard deviation)

在再现性条件下所得测试结果的标准差。

(引自 GB/T 3358.1—1993)

注 15: 再现性标准差是再现性条件下测试结果分布的分散性的度量。

注 16: 类似地可定义“再现性方差”与“再现性变异系数”,作为再现性条件下测试结果分散性的度量。

3.20

再现性限 (reproducibility limit)

一个数值,在再现性条件下,两个测试结果的绝对差小于或等于此数的概率为 95%。

(引自 GB/T 3358.1—1993)

注 17: 再现性限用符号 R 表示。

3.21

离群值 (outlier)

样本中的一个或几个观测值,它们离开其他观测值较远,暗示它们可能来自不同的总体。

(引自 GB/T 3358.1—1993)

注 18: GB/T 6379.2 规定了在正确度和精密度试验中,用来识别离群值的统计检验和显著性水平。

3.22

协同评定试验 (collaborative assessment experiment)

一种实验室间的试验,在这样的试验中,用相同的标准测量方法对同一物料进行测试,以评定每个实验室的水准。

注 19: 在 3.16 和 3.20 中给出的定义适用于观测值为连续变化的情形。如果测试结果是离散的或经过修约的,那么前面所定义的重复性限和再现性限是各自满足以下条件值的最小值;两个测试结果差的绝对值小于或等于该值的概率不小于 95%。

注 20: 由 3.8 到 3.11、3.15、3.16、3.19 和 3.20 中所给的定义的诸量,指的都是实际中未知的理论值。实际确定重复性和再现性标准差及偏倚时用 GB/T 6379.2 和 ISO 5725-4 中所描述的试验,用统计语言说是这些理论值的估计值,因此会有误差。例如,与 r 和 R 相关的概率水平将不会正好等于 95%。当很多实验室参与一个精密度试验时,这些概率水平将近似等于 95%,但是当参与精密度试验的实验室数目少于 30 时,概率水平可能偏离 95% 较远。这是不可避免的,但是也不要过于低估它们的实际效用,因为设计它们的原意就是要作为一种工具,用来判断实验结果之间的差别是否是由于测量方法的随机不确定因素造成的。比重复性限 r 和再现性限 R 大的差值应该引起关注。

注 21: 符号 r 和 R 在其他地方有其它更一般的含义;例如在 ISO 3534-1 中, r 表示相关系数, R (或 W) 表示一组观测值的极差。如果有可能产生误解,特别是在标准中引用时,宜使用全称重复性限 r 或再现性限 R ,这样就不致引起混淆。

4 准确度试验定义的实际含义

4.1 标准测量方法

4.1.1 为使测量按同样的方法进行,测量方法应标准化。所有测量都应该根据规定的标准方法进行。这意味着必须要有一个书面的文件,规定有关如何进行测量的所有的细节,最好还要包括如何获得和准备试样的内容。

4.1.2 有关测量方法文件的存在意味着有一个负责研究测量方法机构的存在。

注 22: 在 6.2 中将更全面的讨论标准测量方法。

4.2 准确度试验

4.2.1 准确度(正确度和精密度的)度量宜由参加试验的实验室报告的系列测试结果确定。由为此目的而专门设立的专家组组织所有测试。

这样一个不同实验室间的试验称为“准确度试验”。准确度试验根据其限定目标也可称为“精密度的试验”或“正确度试验”。如果目标是确定正确度,那么应事先或同时进行精密度的试验。

通过这样试验得到的准确度的估计值,宜指明所用的标准测量方法,且结果仅在所用的方法下才有效。

4.2.2 准确度试验通常可以认为是一次标准测量方法是否适合的一个实际测试。标准化的主要目标之一就是尽可能的估计用户(实验室)之间的差异,由准确度试验提供的数据将会揭示出这个目标是如何有效取得的。实验室内方差(见第 7 章)或实验室均值之间的差异可能表明标准测量方法还不够详细,可以进一步改进。如果这样,宜将问题报告给标准化团体以便进一步调查。

4.3 同一测试对象

4.3.1 在一个准确度试验中,规定物料或规定产品的样本从一个中心点发往位于不同地点,不同国家,甚至不同洲的许多实验室。重复性条件的定义(3.14)指出在这些实验室中进行的测量应该对同一测试对象,并在实际同一时段内进行。为此应满足以下两个不同的条件:

- a) 分送各实验室的样本应该是相同的;
- b) 样本在运输过程 and 在实际测试前所耗费的时间须保持相同。

在组织精确度试验中,要仔细考察这两个条件是否得到满足。

注 23: 在 6.4 中将更全面讨论物料的选取。

4.4 短暂的时间间隔

4.4.1 根据重复性条件的定义(3.14),确定重复性的测量必须在恒定的操作条件下进行;即在整个测量时间段内,在 0.3 中所列的那些因素必须保持不变。特别,设备在两次测量之间不应重新校准,除非校准是单个测量中一个基本的组成部分。在实际中,在重复性条件下进行的试验宜在尽可能短的时间间隔内进行以便使那些不能总是保证不变的因素,比如环境因素的变化最小。

4.4.2 影响不同观测之间的时间间隔的另一因素是测试结果的独立性假定。为避免前面的测试结果可能会影响以下的测试结果(从而可能低估重复性方差),就有必要按以下方式提供样本:操作员根据样品编号不知道哪些样品是相同的。指示操作员按一定观测顺序操作,而顺序是随机的,以使所有的“同一”测试对象的测试不会一起进行。这也许意味着违背了重复测量应在一个短的时间段内完成的初衷,除非全部测量能在一个很短的时间间隔内完成。

4.5 参与的实验室

4.5.1 GB/T 6379 的本部分的一个基本假定是对一个标准测量方法而言,重复性对使用这个标准程序的每个实验室应该或至少是近似相同的,这样可以允许建立一个共同的平均重复性标准差,它适用于任何实验室。然而,每个实验室在重复性条件下进行一系列观测时,都能就该测量方法得到一个自己的重复性标准差的估计值,并可根据此与共同的标准差的值来校核该估计值。ISO 5725-6 详细地讨论了这种方法。

4.5.2 在 3.8 到 3.20 中定义的量,理论上适用于可能使用所述测量方法的所有实验室。但在实际上,它们是根据这个实验室总体的一个样本来确定的。选择这个样本的进一步细节将在 6.3 中讨论。当参加试验的实验室数及测量数都达到 6.3 中规定的数量时,所获得的正确度与精密度的估计值即可满足要求。然而,如果将来某一时间,有证据表明参加测试的实验室不能或不再能真正代表所有使用该标准测量方法的实验室,那么测量就将重新进行。

4.6 观测条件

4.6.1 在 0.3 中列出了能使在一个实验室内获得的观测值产生变异的所有因素。这些因素包括时间、

操作员与设备等。因为在不同时间进行测试时,由于环境条件的改变及设备的重新校准等都会使观测值受到影响。在重复性条件下,观测值是在所有这些因素不变的情况下取得的;在再现性条件下,观测值是在不同的实验室获得的,由于实验室的不同,不仅所有其他因素会发生改变,而且由于在两个实验室之间的管理和维护以及观测值的稳定性检查等诸多方面的差异也会对结果产生不同的影响。

4.6.2 有时也有必要考虑中间精密度条件,即观测值是在相同的实验室获得,但是允许时间、操作员或设备中的一个或几个因素发生改变。在确定测量方法的精密度时,很重要的一点就是要规定观测条件,即上述时间、操作员和设备这三个因素哪些不变,哪些改变。

此外,这三个因素所引起的差异的数值大小也与测量方法有关。例如,在化学分析中,“操作员”和“时间”是主要因素;微量分析中,“设备”和“环境”是主要因素;而在物理测试中,“设备”和“校准”是主要因素。

5 统计模型

5.1 基本模型

为估计测量方法的准确度(正确度和精密度),假定对给定的受试物料,每个测试结果 y 是三个分量的和:

$$y = m + B + e \quad \dots\dots\dots (1)$$

式中:

m ——总平均值(期望);

B ——重复性条件下偏倚的实验室分量;

e ——重复性条件下每次测量产生的随机误差。

5.1.1 总平均值 m

5.1.1.1 总平均值 m 是测试水平;一种化学品或物料的不同成分的样品(例如不同类型的钢材)对应着不同的水平。在很多技术场合,测试水平仅由测量方法确定,独立真值的概念并不适用。然而,在某些情况下,受试特性的真值 μ 的概念仍可使用,例如一种正在滴定溶液的真正浓度。总平均值 m 未必与真值 μ 相等。

5.1.1.2 在检查用相同测量方法获得的测试结果间的差异时,测量方法的偏倚不会对其产生影响,因此可以忽略。然而,当把测试结果和一个在合同中或标准中规定的值进行比较时,其中合同或标准中指的是真值 μ 而不是测试水平 m ,或者比较不同的测量方法得到的结果时,必须考虑测量方法的偏倚。如果存在一个真值,并且可以获得满意的参照物,那么就应用 ISO 5725-4 中的方法确定测量方法的偏倚。

5.1.2 分量 B

5.1.2.1 在重复性条件下进行的任何系列测试中,分量 B 可以认为是常数,但是在其他条件下进行的测试,分量 B 则会不同。当只对两个相同的实验室比较测试结果时,有必要确定它们相应的偏倚,通过准确度实验测定各自的偏倚,或通过它们在它们之间专门的试验确定。然而,若对不特别指定的两个实验室之间差异进行一般性的描述,或者对两个还没有确定其各自偏倚的实验室进行比较时,必须考虑偏倚的实验室分量的分布,这就是引进再现性概念的理由。在 GB/T 6379.2 中给出的程序,是在假定偏倚的实验室分量是近似正态分布情况下得到的,但在多数实际情形只须假定分布为单峰的即可。

5.1.2.2 B 的方差称为实验室间方差,用下式表示:

$$\text{var}(B) = \sigma_B^2 \quad \dots\dots\dots (2)$$

其中 σ_B^2 包含操作员间和设备间的变异。

在 GB/T 6379.2 中描述的基本精密度试验中,这些分量没有被分拆。在 ISO 5725-3 中给出了测量 B 的某些随机分量的大小的方法。

5.1.2.3 通常, B 可以看作是随机分量和系统分量之和。这里并不试图列出所有与 B 有关的因素,

这些因素包括不同的气候条件、制造者允许的设备变差,甚至包括由于操作员在不同地点接受培训所引起的技术上的差异等。

5.1.3 误差项 e

5.1.3.1 误差项表示每个测试结果都会发生的随机误差。在 GB/T 6379 本部分中,所有程序是在假定误差分布近似为正态分布的情况下得出的,但是在多数实际情形只须假定为单峰的即可。

5.1.3.2 在重复性条件下单个实验室内的方差称为实验室内方差,用下式表示:

$$\text{var}(e) = \sigma_w^2 \quad \dots\dots\dots (3)$$

5.1.3.3 由于诸如操作员的操作技巧等方面的差异,不同实验室的 σ_w^2 值可能不同,但 GB/T 6379 本部分中,假定对一般的标准化测量方法,实验室之间的这种差异是很小的,可以对所有使用该测量方法的实验室设定一个对每个实验室都相等的实验室内方差。该方差称为重复性方差,它可以通过实验室内方差的算术平均值来进行估计,表达式如下:

$$\sigma_e^2 = \overline{\text{var}(e)} = \overline{\sigma_w^2} \quad \dots\dots\dots (4)$$

上式中的算术平均值是在剔除了离群值后对所有参加准确度试验的实验室计算的。

5.2 基本模型和精密度的关系

5.2.1 当采用 5.1 中的基本模型时,重复性方差可以直接作为误差项 e 的方差,但再现性方差为重复性方差和 5.1.2.2 中实验室间方差之和。

5.2.2 作为精密度度量的两个量:

重复性标准差

$$\sigma_r = \sqrt{\text{var}(e)} \quad \dots\dots\dots (5)$$

再现性标准差

$$\sigma_R = \sqrt{\sigma_L^2 + \sigma_e^2} \quad \dots\dots\dots (6)$$

5.3 其他可供选择的模型

有时需要考虑基本模型的推广,它们在 GB/T 6379 其他相关部分中描述。

6 为估计准确度试验设计方面的考虑

6.1 准确度试验的计划

6.1.1 估计一个标准测量方法的精密度和(或)正确度试验的具体安排应是熟悉该测量方法及其应用的专家组的任务。专家组中至少应该有一个成员具有统计设计和试验分析方面的经验。

6.1.2 当计划一个试验时要考虑以下的问题:

- a) 该测量方法是否有一个令人满意的标准?
 - b) 宜征集多少实验室来协作进行试验?
 - c) 如何征集实验室? 这些实验室应满足什么要求?
 - d) 在实际中什么是水平的变化范围?
 - e) 在试验中宜使用多少个水平?
 - f) 什么样的物料才能表达这些水平? 如何准备受试物料?
 - g) 宜规定多少次重复?
 - h) 完成所有这些测量宜规定多长的时间范围?
 - i) 5.1 中的基本模型是否适宜? 是否需要考虑修改?
 - j) 需要什么特别的预防措施来确保同一物料在所有的实验室、在相同的状态下进行测量?
- 这些问题在 6.2 到 6.4 中予以考虑。

6.2 标准测量方法

如同在 4.1 中指出的那样,所考察的测量方法应是一个标准化的方法。这样一个方法应是稳健的,

即测量结果对测量过程中的微小变动,不会产生意外的大变动。若测量过程真有很大的变化,应有适当的预防措施或发出警告。在制定一个标准测量方法中,应该尽一切努力力求消除或减少偏倚。

也可以用一些相似的测试程序来对已经建立的测量方法和最新标准化的测量方法的正确度和精密度的测试。在后一种情况下,所得到的结果宜被看作是初始估计值,因为正确度和精密度随着实验室经验的积累而改变。

建立测量方法的文件应该是明确的和完整的。所有涉及该程序的环境、试剂和设备、设备的初始检查以及测试样本的准备的重要操作都应该包括在测量方法中,这些方法尽可能地参考其他的对操作员有用的书面说明。说明宜精确说明测试结果和计算方法以及应该报告的有效数字位数。

6.3 准确度试验的实验室的选择

6.3.1 实验室的选择

从统计的观点来看,那些参加估计准确度的实验室宜从所有使用该测量方法的实验室中进行随机选取。自愿参加的实验室可能不代表实验室的实际组成。然而,其他的一些考虑,比如要求参加的实验室应该分布在不同的洲或不同的气候地域等可能对代表性模式产生影响。

参加的实验室应该不宜仅由那些在对该测量方法进行标准化过程中已获得专门经验的实验室组成。也不宜由那些特别的“标准”实验室组成,这些“标准”实验室是专家用该方法来演示准确度的确定的。

需要征集参加协同实验室间测试的实验室个数,以及每个实验室在每个测试水平需要进行的测试结果个数是有关的。在 6.3.2 到 6.3.4 中给出了如何决定这些数目的导则。

6.3.2 估计精密度所需实验室数

6.3.2.1 在第 5 章中,(2)到(6)式中符号 σ 表示的诸量是未知的标准差真值,精密度试验的一个目标就是对它们进行估计。当可对标准差真值 σ 求得估计值 s 时,可以得到关于 σ 的范围的结论,即估计值 s 期望所在的范围。这是一个熟知的统计问题,可通过 χ^2 分布和 s 的估计值所基于的测试结果数目得到解决。通常使用的公式是:

$$P(-A < \frac{s-\sigma}{\sigma} < +A) = P \quad \dots\dots\dots (7)$$

以 A 表示标准值估计值不确定度的系数,常用百分数来表示。(7)式表示可以预期标准差的估计值 s 以概率 P 位于标准差真值(σ) A 倍的两侧。

6.3.2.2 对单一测试水平,重复性标准差的不确定度依赖于实验室数 p 和每个实验室内的测试结果数 n 。对再现性标准差,其估计程序较为复杂,因为再现性标准差是由两个标准差所确定(见(6)式)。此时需要另一个因子 γ ,它表示再现性标准差对重复性标准差的比:

$$\gamma = \sigma_R / \sigma_r \quad \dots\dots\dots (8)$$

6.3.2.3 下面给出计算概率水平为 95% 下 A 值的一个近似式。此式的目的是计算所需征集实验室数,并确定每个实验室在每个测试水平所需的测试结果数。这些等式没有给出置信限,因此在计算置信限的分析阶段不宜使用。 A 的近似公式如下:

对重复性,

$$A = A_r = 1.96 \sqrt{\frac{1}{2p(n-1)}} \quad \dots\dots\dots (9)$$

对再现性,

$$A = A_R = 1.96 \sqrt{\frac{p[1+n(\gamma^2-1)]^2 + (n-1)(p-1)}{2\gamma^4 n^2 (p-1)p}} \quad \dots\dots\dots (10)$$

注 24: 可以假定具有 ν 个自由度和期望值 σ^2 的样本方差近似服从正态分布,其方差为 $2\sigma^4/\nu$ 。(9)式和(10)式是在这个假定下得出的。通过精确的计算可检验上述近似公式。

6.3.2.4 γ 值是未知的,通常可利用在该测量方法标准化过程中获得的实验室内标准差和实验室内标准差得到它的初步估计。表 1 给出了实验室数为 p , 每个实验室的不同测试结果数为 n 时,重复性标准差和再现性标准差不确定度系数的精确值(以百分数表示),其图示见附录 B。

表 1 重复性标准差和再现性标准差估计值的不确定度系数

试验室数 p	A_r			A_R								
				$\gamma=1$			$\gamma=2$			$\gamma=3$		
	$n=2$	$n=3$	$n=4$	$n=2$	$n=3$	$n=4$	$n=2$	$n=3$	$n=4$	$n=2$	$n=3$	$n=4$
5	0.62	0.44	0.36	0.46	0.37	0.32	0.61	0.58	0.57	0.68	0.67	0.67
10	0.44	0.31	0.25	0.32	0.26	0.22	0.41	0.39	0.38	0.45	0.45	0.45
15	0.36	0.25	0.21	0.26	0.21	0.18	0.33	0.31	0.30	0.36	0.36	0.36
20	0.31	0.22	0.18	0.22	0.18	0.16	0.28	0.27	0.26	0.31	0.31	0.31
25	0.28	0.20	0.16	0.20	0.16	0.14	0.25	0.24	0.23	0.28	0.28	0.27
30	0.25	0.18	0.15	0.18	0.15	0.13	0.23	0.22	0.21	0.25	0.25	0.25
35	0.23	0.17	0.14	0.17	0.14	0.12	0.21	0.20	0.19	0.23	0.23	0.23
40	0.22	0.16	0.13	0.16	0.13	0.11	0.20	0.19	0.18	0.22	0.22	0.22

6.3.3 估计偏倚所需的实验室数

6.3.3.1 测量方法的偏倚 δ 可由式(11)估计:

$$\delta = \bar{y} - \mu \quad \dots\dots\dots (11)$$

式中:

$\bar{y}$ ——所有实验室对一特定的测试水平所得到的所有测试结果的总平均值;

μ ——可接受参照值。

该估计值的不确定度可由下式表达:

$$P(\delta - A\sigma_R < \delta < \delta + A\sigma_R) = 0.95 \quad \dots\dots\dots (12)$$

上式表示这个估计值以 0.95 的概率距测量方法偏倚的真值不超过 $A\sigma_R$ 。利用系数 γ (参见(8)式)可得:

$$A = 1.96 \sqrt{\frac{n(\gamma^2 - 1) + 1}{\gamma^2 p n}} \quad \dots\dots\dots (13)$$

A 的值由表 2 给出。

表 2 测量方法偏倚的估计值的不确定度系数 A

实验室数 p	A 值								
	$\gamma=1$			$\gamma=2$			$\gamma=5$		
	$n=2$	$n=3$	$n=4$	$n=2$	$n=3$	$n=4$	$n=2$	$n=3$	$n=4$
5	0.62	0.51	0.44	0.62	0.80	0.79	0.87	0.86	0.86
10	0.44	0.36	0.31	0.58	0.57	0.56	0.61	0.61	0.61
15	0.36	0.29	0.25	0.47	0.46	0.46	0.50	0.50	0.50
20	0.31	0.25	0.22	0.41	0.40	0.40	0.43	0.43	0.43
25	0.28	0.23	0.20	0.37	0.36	0.35	0.39	0.39	0.39
30	0.25	0.21	0.18	0.33	0.33	0.32	0.35	0.35	0.35
35	0.23	0.19	0.17	0.31	0.30	0.30	0.33	0.33	0.33
40	0.22	0.18	0.15	0.29	0.28	0.28	0.31	0.31	0.31

6.3.3.2 在试验期间,实验室偏倚 Δ 可由下式估计:

$$\hat{\Delta} = \bar{y} - \mu \quad \dots\dots\dots (14)$$

式中:

$\bar{y}$ ——所有实验室对特定测试水平所得到的所有测试结果的算术平均值;

μ ——可接受参照值。

该估计值的不确定度可由下式表达:

$$P(\Delta - A_w\sigma_r < \hat{\Delta} < \Delta + A_w\sigma_r) = 0.95 \quad \dots\dots\dots (15)$$

上式表示估计值以 0.95 的概率距实验室偏倚的真值不超过 $A_w\sigma_r$ 。实验室内不确定度系数为:

$$A_w = \frac{1.96}{\sqrt{n}} \quad \dots\dots\dots (16)$$

A_w 的值由表 3 给出。

表 3 实验室内偏倚的估计值的不确定度系数 A_w

测试结果数 n	A_w 值
5	0.88
10	0.62
15	0.51
20	0.44
25	0.39
30	0.36
35	0.33
40	0.31

6.3.4 实验室选择的影响

实验室数的选择是在可利用资源与将估计值的不确定度减少至一个满意的水平之间的一种折衷。根据附录 B 中的图 B.1 和图 B.2, 可以看到重复性标准差和再现性标准差当参加精密度试验的实验室数很小($p \approx 5$)时, 其值变化较为显著。而当 p 大于 20 时, 再增加 2 到 3 个只能使不确定度降低很少。一般取 p 为 8~15。当 σ_L 大于 σ_r (即 γ 大于 2) 时, 每个实验室在每个水平的测试结果数 $n > 2$ 时, 并不会获得比 $n=2$ 时太多的信息。

6.4 用于准确度试验物料的选择

6.4.1 在确定一个测量方法的准确度的测试中, 所使用的物料应该完全能代表该测量方法在正常的使用中的那些物料。作为一般规则, 使用 5 种不同的物料通常就能够满足较大的水平变化范围, 用这些水平完全能够确定所要求的准确度。当怀疑是否有必要修改最近开发的测量方法时, 在对该方法进行首次调研时, 只需要用较小水平数的物料, 在此基础上再进行进一步的准确度试验。

6.4.2 当观测值必须要在各个不随测量而改变的分离的物料上进行测量时, 至少在原则上, 这些观测值应该在不同的实验室使用一系列相同物料进行测试。然而, 这样就有必要将相同的物料给分布在各个国家或洲的不同地方的许多实验室, 在运输过程中同时伴随着许多损失和风险。如果在不同的实验室使用不同的物料, 那么就要按照这样的方式来选择物料, 即要确保这些物料是完全相同的。

6.4.3 在选择代表不同水平的物料时, 应考虑在将准备样本分送前, 是将物料进行专门的匀质化处理, 还是将不均匀物料的影响包括在准确度数值中。

6.4.4 在对不能匀质化的固体物料(如金属、橡胶、纺织品等)进行测量时, 或不能对相同试样重复测量时, 测试物料的非匀质性将成为该测量方法精密度的一个重要分量, 此时物料的同一性的概念也不再成立。虽然精密度试验仍可以进行, 但精密度的值仅仅对所用的物料有效, 也只有在这种情况下方可使用。

要使所确定的精密度值能更广泛地应用,只有在能证明其数值不因生产者不同或不同时间生产物料有较大差别时才可。这要比 ISO 5725 中所考虑的试验更加精心的安排的试验。

6.4.5 通常,当涉及破坏性实验时,由试样之间的差异所产生的测试结果的变异与测量方法本身的变异相比较或忽略不计,或应将它视为测量方法变异的一个固有的组成部分,从而成为精密度的一个真正分量。

6.4.6 当所测量的物料随着时间而改变时,应考虑完成全部试验时间的范围。在某些情况下,宜规定样本测量的时间,这一点很重要。

6.4.7 在上述论述中,在不同实验室的测量隐含着将试样运至实验室。尽管有些试样不存在运输问题,如储藏罐中的油。在这些情况下,不同实验室的测量指的是把不同的操作员连同他们所使用的设备送往测试现场。在其他一些情形,被测量可能是瞬时的或可变的,如江河中的流水,此时要注意尽可能取位置靠近,条件相同的样本进行测量。指导原则是以确定重复相同测量的能力为目的。

6.4.8 上述确定一种测量方法精密度的方法时,都假定了精密度或与所测试的物料无关或与物料有某种可预测的依赖关系。对某些测量方法,引用精密度值时必须说明是对哪一类或哪几类物料而言的。在其它应用场合,这些数值仅能作为粗略的估计。更为常见的情形是,精密度与测试水平密切相关,因此,建议在公布精密度时,同时明确精密度试验中所用的物料及物料的变化范围。

6.4.9 为评定正确度,至少一种所用的物料要有接受参照值。如果正确度似乎随水平改变,则需要有若干水平的物料具有接受参照值。

7 准确度数据的应用

7.1 正确度和精密度数值的发布

7.1.1 当精密度试验的目的是为了获得在定义 3.14 和 3.18 的条件下的重复性和再现性标准差的估计值时,应使用 5.1 中的基本模型。GB/T 6379.2 提供了估计这些标准差的方法,ISO 5725-5 提供了某些可替代的方法。当精密度试验的目的是为了获得精密度中间度量的估计值时,则应使用 ISO 5725-3 中的模型和方法。

7.1.2 一旦确定了测量方法的偏倚,其值宜与确定该偏倚时所参照的有关说明一起发布。当偏倚随测试水平改变时,宜以表格的形式对给定的水平及所确定的偏倚和所用的参考说明进行发布。

7.1.3 当以实验室间试验进行准确度和精密度的估计时,宜向每个参加测试的实验室报告各自相对总平均值的偏倚的实验室分量。这个信息对将来进行类似试验是有用的,但不宜用作校准的目的。

7.1.4 任何标准测量方法的重复性和再现性标准差都应用 GB/T 6379 的第 2 到第 4 部分规定的方法来确定,其结果宜在发布该标准测量方法时作为专门标记为“精密度”一节的内容。这一节也可以列出重复性限和再现性限(r 和 R)。当精密度不随测试水平变化时,可单独给出每种情况下的平均数。当精密度随着测试水平变化时,宜以表格的形式进行发布,如表 4,也可以用数学公式来表示。精密度的中间度量也宜用类似的形式来表达。

表 4 报告标准差的方法示例

范围或水平	重复性标准差 s_r	再现性标准差 s_R
从……到……		
从……到……		
从……到……		

7.1.5 在精密度条款中应给出重复性和再现性条件的定义(3.14 和 3.18)。当涉及精密度的中间度量时,宜说明时间、操作员和设备这些因素中哪些因素允许变化。当给定重复性和再现性限时,还应增加其它陈述,它们把重复性限和再现性限与两个测试结果之间的差和 95% 的概率水平联系起来。建议的措辞如下:

在通常正确的操作方法下,由同一个操作员使用同一仪器设备,在最短的可行的时间段内,对同一物料所做出的两个测试结果之间的差出现大于重复性限 r 的情况,平均在 20 次测试中不会超过一次。

在通常正确的操作方法下,由两个实验室报告的对同一物料进行测试的测试结果的差出现大于再现性限 R 的情形,平均在 20 次测试中不会超过一次。

通过引用进行测试所要遵守的标准测量方法的条款的编号,或以其他方式确保测试结果定义的清晰。

7.1.6 通常,在精密章节结束部分应该增加对准确度试验的简要说明。建议措辞如下:

准确度数据是依照 GB/T 6379(部分),在 $\times\times\times\times$ 年,对 p 个实验室和 q 个测试水平所组织和分析的试验而得到的。() 个实验室数据包括离群值,在计算重复性标准差和再现性标准差时不包括这些离群值。

应有关于在准确度试验中所使用的物料的描述,尤其当正确度和精密度依赖于测试物料时。

7.2 正确度和精密度数值的实际应用

在 ISO 5725-6 中详细论述了正确度和精密度的实际应用,以下是若干例子。

7.2.1 对测试结果接收性的检查

产品规范可有在重复性条件进行重复测量的要求。在这种情形,重复性标准差可以用对测试结果的接收性的检验,以及决定当测试结果不可接收时应该采取什么行动。当供需双方对相同的物料进行测量,而试验结果不同时,可以用重复性标准差和再现性标准差来决定差异是否是测量方法所能允许的。

7.2.2 在一个实验室内测试结果的稳定性

通过根据标准物料进行定期测试,实验室能够检查其结果的稳定性,从而得出该实验室有能力控制实验的偏倚和重复性的证据。

7.2.3 对实验室水准进行评估

对实验室的认认证日益普遍。无论采用标准物料还是进行实验室间试验,所获得的测量方法的正确度与精密度数值能用于对一个候选的实验室的偏倚与重复性进行评定。

7.2.4 比较可供选择的测量方法

为测量某一特性,若有两种测量方法可用,其中一种要比另一种简单而价廉,但是一般使用较少。可以根据正确度和精密度值用来对某些限定范围的物料判断这种廉价方法的使用。

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的系数
- 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
- $C_{crit}, C'_{crit}, C''_{crit}$ 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 ν_1 和 ν_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
- 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
- $\overline{\bar{y}}$ 测试结果的总平均值
- α 显著性水平
- β 第二类错误概率
- γ 再现性标准差与重复性标准差的比值(σ_R/σ_r)
- Δ 实验室偏倚
- $\hat{\Delta}$ Δ 的估计值
- δ 测量方法偏倚
- $\hat{\delta}$ δ 的估计值
- λ 两个实验室偏倚或两个测量方法偏倚之间的可检出的差
- μ 测试特性的真值或接受参照值
- ν 自由度
- ρ 方法 A 和方法 B 的重复性标准差之间的可检出的比
- σ 标准差的真值
- τ 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
- ϕ 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 ν 的 χ^2 分布的 p 分位数

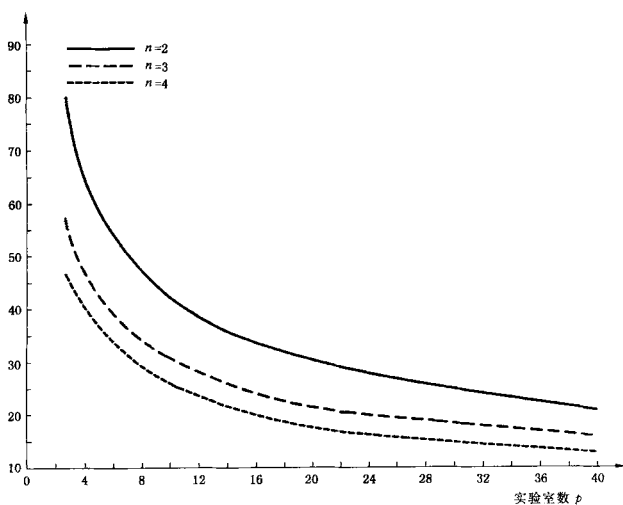
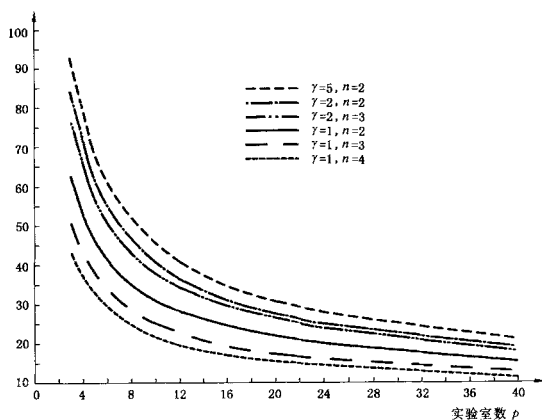
用作下标的符号

- C 校准差
- E 设备差
- 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 操作员差
- r 重复性
- R 再现性
- T 时间差
- 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

附录 B

(规范性附录)

精密度度量不确定度的图示

 s_r 的不确定度系数(%)图 B.1 以概率水平 95%，预期 s_r 的偏离不超过其真值的倍数 s_R 的不确定度系数(%)图 B.2 以概率水平 95%，预期 s_R 的偏离不超过其真值的倍数

附录 C
(资料性附录)

参考文献

- [1] ISO 3534-2:1993 统计学 词汇和符号 第2部分:统计质量控制
Statistics—Vocabulary and symbols—Part 2: Statistical quality control
- [2] ISO 3534-3:1985 统计学 词汇和符号 第3部分:实验设计
Statistics—Vocabulary and symbols—Part 3: Design of experiments
- [3] ISO 5725-5:1998 测量方法与结果的准确度(正确度与精确度) 第5部分:确定标准测量方法正确度的可替代方法
Accuracy (trueness and precision) of measurement methods and results—Part 5: Alternative methods for the determination of the precision of a standard measurement method
- [4] ISO 5725-6:1994 测量方法与结果的准确度(正确度与精确度) 第6部分:准确度值的实际应用
Accuracy (trueness and precision) of measurement methods results—Part 6: Use in practice of accuracy values
- [5] ISO Guide 33:1989, *Use of certified reference materials*
- [6] ISO Guide 35:1989, *Certification of reference materials—General and statistical principles*
-