



中华人民共和国认证认可行业标准

RB/T XXX—202X

定量类能力验证结果评价规范

Specification of evaluation for quantitative proficiency testing

草案稿

2021-12-19

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国国家认证认可监督管理委员会 发布

目 次

前 言..... III

引 言..... IV

1 范围..... 1

2 规范性引用文件..... 1

3 术语和定义..... 1

4 总则..... 3

 4.1 定量类能力验证的范围.....3

 4.2 定量类能力验证结果评价的一般原则.....3

5 能力验证的目的..... 4

 5.1 能力验证计划主要目的.....4

 5.2 不同机构组织的能力验证计划的主要目的.....4

6 能力验证计划的统计设计..... 5

 6.1 统计设计基础.....5

 6.2 统计假设.....5

 6.3 统计设计考虑事项.....5

7 物品..... 6

 7.1 物品的来源.....6

 7.2 物品的选择.....6

 7.3 物品的均匀性和稳定性.....6

8 指定值及其标准不确定度..... 7

 8.1 指定值的确定.....7

 8.2 指定值的标准不确定度的确定.....8

 8.3 指定值的验证.....9

9 能力评定准则的确定..... 10

 9.1 能力评定准则.....10

 9.2 确认评定准则的方法.....10

 9.3 对指定值不确定度的限定.....10

10 结果汇总与统计..... 11

 10.1 结果收集和初审..... 11

 10.2 结果的统计处理..... 11

11 能力评定.....12

 11.1 能力评定的通用原则..... 12

 11.2 能力统计量的计算..... 12

 11.3 评定结果分析..... 14

附录 A（规范性附录）符号..... 16

附录 B（规范性附录）样品均匀性和稳定性检验..... 18

附录 C（规范性附录）稳健分析..... 22

附录 D（资料性附录）评定结果分析（图示法）..... 31

附录 E（资料性附录）参加者结果统计计算示例..... 41

附录 F（资料性附录）R 语言代码..... 57

附录 G（资料性附录）定量类能力验证流程图.....67

参考文献..... 68

前 言

本标准按照 GB/T 1.1-2020 给出的规则起草。

本标准由国家认证认可监督管理委员会提出并归口。

本标准起草单位：华测检测认证集团股份有限公司、中国合格评定国家认可中心、中国检验检疫科学研究院、江苏省产品质量监督检验研究院、北京中实国金国际实验室能力验证研究有限公司。

本标准主要起草人： 。

引 言

为适应认证认可事业发展要求，充分发挥行业标准在认证认可重点工作领域中的重要作用，国家认证认可监督管理委员会组织制定能力验证方面的相关认证认可标准，以便认证认可工作实施时有所参考，本标准是能力验证认证认可标准的一部分，仅明确了定量类能力验证结果评价规范。

本标准主要对定量类能力验证结果的评价方法进行说明，对各种结果评价方法的选择进行了阐述，同时对定量类能力验证结果评价相关要素进行了规范，以便能力验证组织机构更好的进行参考。本标准按照 GB/T 1.1-2020 给出的规则起草。

定量类能力验证结果评价规范

1 范围

本标准规定了定量类能力验证结果评价所涉及的目的、统计设计、物品、指定值、结果分析和能力评定的要求和程序。

本标准适用于能力验证组织机构在开展定量类能力验证过程中所进行的结果评价，也可作为机构组织能力验证提供参考。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 27043 合格评定能力验证的通用要求（ISO/IEC 17043，IDT）

GB/T 28043 利用实验室间比对进行能力验证的统计方法（ISO 13528，IDT）

3 术语和定义

GB/T 28043 界定的以及下列术语和定义适用于本文件。为了便于使用，以下重复列出了GB/T28043 中的某些术语和定义。

3.1

实验室间比对 interlaboratory comparison

按照预先规定的条件，由两个或多个实验室在预定的条件下对同一或相似的检测对象进行测量或检测的组织、实施和评价活动。

[GB/T 28043-2019,3.1]

3.2

能力验证 proficiency testing

利用实验室间比对，按照预先设定的准则评价参加者的能力。

注：定量类能力验证是确定能力验证物品的一个或多个被测量的量。

[GB/T 28043-2019,3.2]

3.3

指定值 assigned value

对能力验证物品的某个特性赋予的值。

注：在定量类能力验证计划中，能力验证物品的特性是用量值来表示。

[GB/T 28043-2019,3.3]

3.4

能力评定标准差 standard deviation for proficiency assessment

用于能力评定的离散性度量。

注 1：该定义可以理解为由严格按照要求操作的实验室所组成的一个假设总体，其测试结果都总体标准差。

[GB/T 28043-2019,3.4]

3.5

测量误差 measurement error

测量结果减去参照值。

[GB/T 28043-2019,3.5]

3.6

最大允许误差 maximum permissible error

对规范或规程所允许的测量、测量仪器或测量系统，相对于已知参考量值的测量误差的极限值。

[GB/T 28043-2019,3.6]

3.7

z 值 z score

由能力验证的指定值和标准差计算的实验室偏倚的标准化度量。

[GB/T 28043-2019,3.7]

3.8

ζ 值 ζ score

实验室能力的标准化度量，由参加者结果、指定值、测量结果和指定值的合成标准不确定度得到。

[GB/T 28043-2019,3.8]

3.9

公议值 consensus value

由实验室间比对的一组结果得到的值。

[GB/T 28043-2019,3.11]

3.10

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。

[GB/T 28043-2019,3.12]

3.11

参加者 participant

接受能力验证物品并提交结果以供能力验证提供者评价的实验室、组织或个人。

[GB/T 28043-2019,3.13]

3.12

能力验证物品 proficiency test item

在能力验证中用于评定参加者能力的样品、产品、人工制品、标准物质/标准样品、设备部件、测量标准、数据组或其他信息。

[GB/T 28043-2019,3.14]

3. 13

能力验证提供者 proficiency testing provider

对能力验证计划建立和运作中所有任务承担责任的组织。

[GB/T 28043-2019,3.15]

3. 14

能力验证计划 proficiency testing scheme

在检测、测量、校准或检查的某个特定领域，设计和运作的一轮或多轮次能力验证。

[GB/T 28043-2019,3.16]

3. 15

标准物质/标准样品 reference material (RM)

具有一种或多种规定特性足够均匀且稳定的材料，已被确定其符合测量过程的预期用途。

[GB/T 28043-2019,3.17]

3. 16

有证标准物质/标准样品 certified reference material (CRM)

采用计量学上有效程序测定的一种或多种规定特性的标准物质/标准样品，并附有证书提供规定特性值及其不确定度和计量溯源性的陈述。

[GB/T 28043-2019,3.18]

4 总则

4.1 定量类能力验证的范围

该类能力验证物品被测量的结果通常是数值型的，多以区间或比例尺度来表示，并对结果偏离中心位置的程度进行评价。

注 1：有些能力验证尽管结果是用数值表示，但最终结果评价是按照定性方式的不在本范围内。

注 2：典型的定量类能力验证领域和测试项目有食品和药品领域的成分含量分析、建筑材料领域的物理性能（强度、刚度等）测试，水环境监测领域的理化性质（pH、电导率、溶解氧、温度等）检测等。

4.2 定量类能力验证结果评价的一般原则

4.2.1 统计设计应根据数据的特性、统计假设、误差的性质以及预期的结果数量，制定符合计划目标的统计设计。

4.2.2 只要统计学方法适合计划的目的且可靠，均可用于确定指定值及其不确定度，但需在计划文

件中对该方法有详细的描述，并详细告知参加者。无论使用何种方法确定指定值，均应对每轮能力验证计划指定值的有效性进行检验。同时，根据计划文件来考虑指定值的计量溯源性和测量不确定度的需求、关联性和可行性，以确保能力验证计划符合其目标。

4.2.3 数据分析及相关统计量的计算过程中，参加者结果可能出现各种统计分布情况，数据分析方法应与数据类型及其统计分布特性相适应。在能力验证计划报告中，能力验证提供者应为参加者提供计算方法的描述、结果的一般解释以及与该解释有关的任何限制性说明。

4.2.4 能力验证提供者应使用符合能力验证计划目标的有效评定方法。通过统计计算获得适合的能力统计量，对参加者反馈的结果值进行评价，必要时，还可对参加者反馈的过程信息和结果信息进行关联性分析，根据比较与分析的结果，综合评定参加者能否获得恰当测试结论的能力。

4.2.5 对参加者进行结果评价时，一般包括以下几个方面内容：

- a) 指定值和能力评定准则的确定；
- b) 结果的汇总和统计分析；
- c) 能力评定。

5 能力验证的目的

5.1 能力验证计划主要目的

- a) 评定实验室从事特定检测或测量的能力及监视实验室的持续能力；
- b) 识别实验室存在的问题并启动改进措施，这些问题可能与诸如不适当的检测或测量程序、人员培训和监督的有效性、设备校准等因素有关；
- c) 建立检测或测量方法的有效性和可比性；
- d) 增强实验室客户的信心；
- e) 识别实验室间的差异；
- f) 根据比对的结果，帮助参加实验室提高能力；
- g) 确认实验室声称的不确定度。

5.2 不同机构组织的能力验证计划的主要目的

5.2.1 政府或行业监管部门组织的具有强制性和监督性质的能力验证计划，一般主要以 a)， b) 为目的。

5.2.2 各能力验证提供者自行组织的能力验证计划，主要以 b)， e)， f) 为目的。

5.2.3 检测实验室自行组织的实验室间比对，主要以 c)， d)， f)， g) 为目的。

5.2.4 实验室自行参加的能力验证目的可包括以上全部 a)~g)。

6 能力验证计划的统计设计

6.1 统计设计基础

6.1.1 统计设计涵盖了能力验证计划的数据获取方案的策划、数据的收集、分析和报告等过程，统计设计通常基于能力验证计划的既定目标，比如确定已规定大小的误差，或确定带有规定了测量不确定度的指定值等。

6.1.2 数据分析方法可能从不做任何处理（如描述统计学）到使用复杂的统计变换（如利用带有概率假设或不同能力验证物品结果组合的模型）。

6.1.3 当能力验证计划根据委托人（如客户，或监管和能力认可机构等）给定规范设计时，可直接使用规范中的统计设计和数据分析方法。

6.1.4 在缺少统计设计所需的可靠信息时，可通过开展先期实验室间比对来获得。

6.2 统计假设

6.2.1 能力验证提供者应将用来确定指定值和评价参加者结果的统计设计和数据处理方法形成文件，并对选用的原因和假设进行说明，能力验证提供者应能证实统计假设合理并能保证统计分析按规定程序进行。

6.2.2 对数据的统计分析方法应与统计假设相一致。最常见的统计假设是，来自有能力参加者的结果服从近似正态分布（必要时可经转换）。能力验证提供者根据参加者结果的统计分布情况对统计假设进行确认，对于预期分布不够或不服从分布假设的情况，应考虑可能存在的原因以及选择相适应的数据分析方法。

6.3 统计设计考虑事项

6.3.1 能力验证中每个被测量或特性所要求或期望的准确度（正确度和精密度）以及测量不确定度。

6.3.2 达到统计设计目标所需的最少参加者数量；当参加者数量不足以达到目标或不能对结果进行有意义的统计分析时，应将评定参加者能力的替代方法的详细内容提供给参加者。

6.3.3 有效数字与所报告结果的相关性，包括小数位数。

6.3.4 需要检测或测量的能力验证物品数量，以及对每个能力验证物品或每项测定的检测、校准或测量的重复次数。

6.3.5 用于确定能力评定标准差或其它评定准则的程序。

6.3.6 用于识别和（或）处理离群值的程序。

6.3.7 只要适用，对统计分析中剔除值的评价程序。

6.3.8 只要适当，与设计相符的目标和能力验证轮次的频率。

7 物品

7.1 物品的来源

能力验证物品的来源主要可分为自制、外购和外包制备。在能力验证计划中，自制物品可直接取自天然样或人工配制，天然样和部分人工配制的物品需经过可靠的技术（多实验室协议定值、专家实验室定值等）确认其特性值，而部分经规范、可靠的配制过程完成的物品，其特性值可根据配方确定；外购的有证标准物质（CRM）通常具有标准值和不不确定度区间；外购的标准物质（RM）和外包制备物品则需要经过可靠的技术确认其特性值。

注：如使用上一轮计划物品作为本轮计划物品，则需对物品特性值进行验证。当物品特性值出现显著偏离时，需重新进行特性值确认。

7.2 物品的选择

7.2.1 定量类能力验证物品中被测量及干扰物的种类和水平、物品数量等，都应和能力验证计划的目的相符。

7.2.2 为了能够反映参加者的真实能力，物品的基质类型、干扰物种类和水平等，应与日常检测或校准的物品和材料的类型相同或相近。

7.2.3 物品的性质和制备方式不同亦会对能力验证结果造成影响。天然样往往更接近于实际检测的样品，而人工制备样多是通过基质中添加目标物来获得的，目标物的状态与实际样品可能存在差异，最终可能会影响测试的回收率和精密度。

7.2.4 定量类能力验证物品被测量水平还需要考虑检测方法定量限、行业限值及其他特殊因素（如地域、委托方要求等）。

7.3 物品的均匀性和稳定性

7.3.1 定量类能力验证计划中，物品的均匀性和稳定性，并非指物品从总体到个体毫无差异以及在规定的储存条件和时间间隔内毫无变化，往往是指这些差异和变化相对于被测量的不确定度或能力评定标准差可忽略不计，或是在规定的允许范围内，从而不影响对参加者的能力评价。

7.3.2 能力验证物品的不均匀性或不稳定性可能会造成参加者结果出现不可忽略的测量误差增大和精密度下降，进而影响参加者的能力评价。这种由制备、环境、时间和空间等因素变化造成的能力验证物品一致性差异，一定程度上影响了能力验证的客观性和公正性。

7.3.3 定量类能力验证物品应基于不均匀性和不稳定性对参加者能力评定可能产生的影响，建立合适的均匀性和稳定性评估准则。具体评估方法有：

a)附录 B 所述的实验研究方法，或能够提供同等的或更高均匀性和稳定性保证的替代实验方法；

b)根据先前轮次能力验证计划使用相似能力验证物品的经验，在本轮能力验证中使用相同的方法进行均匀性和稳定性检验；

c)评估本轮能力验证计划的参加者数据，确认其与以往轮次能力验证计划是否相一致，确认报告时间或生产程序是否变化，或是否存在由于不均匀性或不稳定性导致的任何意外波动。

7.3.4 定量类能力验证物品的均匀性检验和稳定性检验最常用到附录 B 所述的实验研究方法。如使用 F 或 t 检验，本质都是对物品总体的抽样检验过程，即从有限物品总体中随机抽样，再将样本测试值分析的各种统计量进行假设检验，最后对物品总体的分布函数进行统计推断；而使用 0.3σ 准则，则是直接判断物品的不均匀性或不稳定性量化在能力验证评定中是否可以忽略不计。

7.3.5 定量类能力验证物品进行均匀性检验和稳定性检验时，应根据计划目的、物品情况（制备数量、样品性质）、测试周期和精密度要求、检验成本等因素，选择物品的抽样方式、测试方法、测试方式、最小取样量、特性值以及数据分析评价方法。

7.3.6 通常应检查所有被测量的均匀性和稳定性。当选择部分被测量检查均匀性和稳定性时，所选的被测量应对样品制备、加工、存储、运输等过程的不均匀和不稳定来源反应敏感，如以下情形：

- a) 当被测量为比例时，小比例的被测量可能更难均匀化，因此，应选择小比例被测量；
- b) 若对物品进行热处理，则应选择对不均加热比较敏感的被测量；
- c) 若被测量在物品的制备过程中可能受凝结、沉淀或其他时间以来效应的影响，应依照顺序对该被测量进行检查。

7.3.7 定量类能力验证物品的稳定性检验应覆盖计划的全过程，尤其需对储存、包装和运输的条件进行监控和验证。如果在本轮计划结束后剩余物品要继续使用，则应对物品特性值进行定期稳定性监控。此外，一些特殊形式的能力验证计划（如顺序传递），物品需按预定的次序逐个传递给下一个参加者进行测试，整个过程时间较长，且涉及多次运输和转移，特性值往往可能出现变化，故需在计划前中后期对特性值进行稳定性检验。

7.3.8 当物品检验结果为不均匀时，首先应检查物品的来源、制备和检验过程是否可以改进，并对不均匀性的影响进行量化，并在评估中予以考虑，必要时重新设计和制备物品。

注：当不均匀性标准偏差 $S_S > 0.3\sigma$ ，应考虑不均匀性的影响，计算 $\sigma'_{pt} = \sqrt{\sigma_{pt}^2 + S_S^2}$ ，并使用 z' 值评价，可参考 11.2 和附录 B。

7.3.9 当物品检验结果为不稳定时，检查样品制备、储存和运输程序是否可以改进，并对不稳定性的影响进行量化，并在评估中予以考虑，必要时不评估参加者能力。

注：当稳定性检验均值差 $|\bar{x} - \bar{y}| > 0.3\sigma$ ，应考虑不稳定性的影响，计算 $\sigma'_{pt} = \sqrt{\sigma_{pt}^2 + |\bar{x} - \bar{y}|^2}$ ，并使用 z' 值评价，可参考 11.2 和附录 B。

8 指定值及其标准不确定度

8.1 指定值的确定

8.1.1 定量类能力验证的指定值常见的几种来源有：

- a) 已知值：根据特定能力验证物品配方（如制造或稀释）确定的结果作为指定值；
- b) 有证参考值：使用有证标准物品作为能力验证物品，以其有证特性值作为指定值；
- c) 单一实验室参考值：单一实验室使用参照方法（如基准测量方法）测定结果作为指定值；
- d) 由专家参加者确定的公议值：专家参加者（某些情况下可能是参考实验室）应当具有可证

实的测定被测量的能力，并使用已确认的、有较高准确度的方法，且该方法与常用方法有可比性；

e) 由参加者确定的公议值：能力验证计划设定的统计程序。将根据参加者报告结果所做的位置估计（稳健均值，中位值或算数平均值）作为指定值。如常用稳健统计方法计算稳健均值作为指定值。

8.1.2 指定值的确定应确保公平地评价参加者，并尽量使检测或测量方法间吻合一致。只要可能，应通过选择共同的比对小组以及使用共同的指定值达到这一目的。

8.1.3 指定值的确定方式往往与计划的目的、以及样品情况有关，下表为不同目的和样品条件下指定值常用确定方式：

表 1 不同情况指定值确定方式

计划目的	物品选择	指定值确定方式
a) 评定实验室从事特定检测或测量的能力及监视实验室的持续能力。	1. 国家标准物质/有证标准样品 (优先选择) 2. 自制样品	1. 由有证参考值确定 2. 由专家实验室公议值确定
b) 识别实验室存在的问题并启动改进措施，这些问题可能与诸如不适当的检测或测量程序、人员培训和监督的有效性、设备校准等因素有关； c) 建立检测或测量方法的有效性和可比性； d) 增强实验室客户的信心； e) 识别实验室间的差异； f) 根据比对的结果，帮助参加实验室提高能力。	1. 自制样品	1. 由专家实验室公议值确定 2. 由单一实验室参考值确定 3. 由参加者公议值确定
g) 确认实验室声称的不确定度。	1. 国家标准物质/有证标准样品 (优先选择) 2. 自制样品	1. 由有证参考值确定 2. 由专家实验室确定

8.2 指定值的标准不确定度的确定

8.2.1 当使用 8.1.1a)~c) 方法确定指定值时，应考虑测量过程、不均匀、不稳定和运输条件带来的不确定度分量，并根据如下指定值的标准不确定度 $u(x_{pt})$ 的计算模型获得：

$$u(x_{pt}) = \sqrt{u_{\text{char}}^2 + u_{\text{hom}}^2 + u_{\text{trans}}^2 + u_{\text{stab}}^2} \dots\dots\dots (1)$$

式中， u_{char} 为由定值引入的标准不确定度；

u_{hom} 为由于能力验证物品间差异而引入的标准不确定度；

u_{trans} 为由于能力验证物品的运输过程而引入的标准不确定度；

u_{stab} 为由于能力验证期间的不稳定性而引入的标准不确定度。

8.2.2 当使用专家实验室结果确定指定值时，可按照附录 E 中表 E9 的计算公式计算 $u(x_{pt})$ 。

8.2.3 当使用参加者结果确定指定值时，由于不均匀性、运输和不稳定性所带来的不确定度分量很大程度上反映在参加者测量结果的波动中，可按照下式直接计算 $u(x_{pt})$ ：

当以稳健平均值作为指定值时，

$$u(x_{pt}) = 1.25 \times \frac{S^*}{\sqrt{p}} \dots\dots\dots (2)$$

当以算术平均值作为指定值时，

$$u(x_{pt}) = \frac{S}{\sqrt{p}} \dots\dots\dots (3)$$

式中， S^* 为稳健标准差， S 为标准差。

8.3 指定值的验证

8.3.1 能力验证提供者在确定指定值时，应对该指定值的有效性进行验证。利用 7.1.1 [a-d] 方式确认的指定值，通常可与参加者结果获得的稳健均值或中位值比较来验证有效性，即当指定值与稳健均值或中位值的差值 x_{diff} 大于其标准不确定度 u_{diff} 的两倍时，指定值与稳健均值或中位值间存在显著差异，该计划指定值有效性存疑。同理，利用 7.1.1[e] 方式确认的指定值，亦可将其按照相同的准则与独立参考值（如均匀性检验数据的平均值）比较。

8.3.2 能力验证指定值有效性存疑时，应对样品的来源、制备和加工、测试方法、抽样方式、实验情况、指定值确定方法、独立参考值来源等因素进行评估，找出潜在的问题。

8.3.3 根据 x_{diff} 与 u_{diff} 间显著差异产生的原因，能力验证提供者应确定是否对结果进行评估，而且（对于连续能力验证计划）应确认是否修改后续能力验证计划的设计。如果差值较大，足以影响能力评定，或足以表明参加者使用的测量方法中存在重要偏倚，在本轮报告中应注明该差值。在这种情况下，未来能力验证计划的设计应考虑该差值。

8.3.4 对于定量类能力验证物品指定值可能出现的异议，应由专家组进行预先评估和后期的仲裁。

9 能力评定准则的确定

9.1 能力评定准则

9.1.1 能力验证是一种合格评定，它需要按照规定的准则，对参加者结果给予评价。

9.1.2 在定量类能力验证中，不同目标的能力评定基本方法是将参加者结果和指定值进行比较。将二者差值直接与设定的准则（如最大允许误差 δ_E ）进行比较，或将差值与测量误差允许值（如能力评定标准差 σ_{pt} ）进行比较，有时也可将差值与差值的不确定度进行比较。

9.2 确认评定准则的方法

9.2.1 与能力评价的目标和目的相符， δ_E 和 σ_{pt} 可由专家判定或法规规定（规定值）。

9.2.2 δ_E 和 σ_{pt} 可根据以前轮次的能力验证得到的估计值或由经验得到的预期值（经验值）。

9.2.3 σ_{pt} 可由统计模型得到的估计值（一般模型）。

9.2.4 σ_{pt} 可由标准方法的精密度试验得到的结果。

9.2.5 σ_{pt} 可由参加者结果分散性的公议值，如稳健标准差、标准化四分位距、传统标准差等得到（具体参见附录C）。

9.3 对指定值不确定度的限定

9.3.1 当指定值的标准不确定度 $u(x_{pt})$ 接近或大于能力评定标准差 σ_{pt} 时，某些参加者可能会因为指定值不准确而收到行动信号或警戒型号。因此，应确定 $u(x_{pt})$ ，并通报参加者。

如果符合以下准则，则可认为 $u(x_{pt})$ 可忽略不计，不需要在该轮能力验证结果中予以解释。

$$u(x_{pt}) \leq 0.3\sigma_{pt} \dots\dots\dots (4)$$

9.3.2 如果不符合该准则，则能力验证提供者应考虑采取以下措施，确保其符合能力验证计划的能力评定策略：

- a) 选择其他确定指定值的方法，使其不确定度满足式(4)；
- b) 在能力验证计划结果的解释中使用指定值的不确定度（如 z' 值）；
- c) 如果指定值来源于参加者结果，并且参加者中可识别子总体之间的差异导致较大的不确定度，应分别报告适用于每一个子总体的指定值和不确定度；
- d) 告知参加者，指定值的不确定度不可忽略，可能影响评定。

10 结果汇总与统计

10.1 结果收集和初审

10.1.1 定量类能力验证中结果的收集范围，应包括样品情况确认、结果值、测试记录、实验室情况调查等信息。

10.1.2 如果允许参加者自己选择测量方法，那么每一被测量或特性的单一指定值可能无法适合所有的参加者。在这种情况下，能力验证提供者可以针对每一种测量方法使用不同的指定值。

10.1.3 数据中明显错误的结果，如单位错误、小数点错误、计算错误或者错报为其他能力验证物品的结果，应从数据集中剔除，单独处理，以免对随后的统计处理造成损害。明显错误的结果应由专家进行识别和判断。

10.1.4 在做数据统计分析之前，应制作结果的直方图或核密度图来对结果进行直观检查，以确认结果的预期分布，识别异常值或意外变异的来源。

注：大部分定量测试结果是服从正态分布，亦有少量测试，如微粒计数服从泊松分布，一些校准项目结果服从指数或波形分布。

10.1.5 结果数量 p 会影响到离群值检验的有效性以及指定值确认方式和统计处理方法的选择：

a) 不同的离群值检验适用于不同规模的数据集。离群值检验需事先指定可能得离群值数量，当离群值较多时可能会失效，因此最适用于 $p > 10$ 的情景（依赖于可能的离群值比例）。

b) 如有可能，数量为 p 的结果数据的指定值应符合 9.3.1 给出的指定值不确定度准则。使用经典方法的均值作为指定值时，需满足剔除离群值后 $p \geq 12$ ；大多数均值的稳健估计在 $p \geq 12$ 时是可以接受的，但由于在 $p < 18$ 时不能符合 9.3.1 要求，故能力评定标准差需考虑指定值不确定度。

c) 当 p 较小时，结果的标准差估计可能存在很高的变异性，不宜使用如 nIQR 或 MADe($p < 30$) 等作为能力评定标准差，否则可能对评价结果产生影响。

10.2 结果的统计处理

10.2.1 根据第 8 章和第 9 章的内容可知，指定值和能力评定准则的确定可以独立于参加者结果，如使用 8.1.1[a-d] 方法来确定指定值以及使用 9.2.1-9.2.5 的方法来确定能力评定准则。

10.2.2 根据参加者结果来确定指定值和能力评定准则是能力验证中最常见的一种方式，即采用统计方法计算位置中心和离散性的估计值（公议值），进而通过能力统计量评定参加者结果。

10.2.3 参加者结果统计处理的方法很多，最常见的有经典统计法和稳健统计法（偶尔也有使用自助法等）。每一种统计方法都有其优点和适用性限制，因此，能力验证提供者应根据能力验证目的和数据特点，统计方法本身的崩溃点和效率，计算条件等因素选择合适的统计方法（稳健方法的计算原理参见附录 C，所有统计方法示例参见附录 E）。

10.2.4 稳健统计法最大的优势在于处理结果时，对结果的离群值和分布不对称性都具有良好的抵抗性。稳健统计法其稳健性能的高低，主要表现在稳健统计量本身的离散性大小和抵抗离群值及微小众数的能力高低。因此，常用效率、崩溃点和对微小众数的抵抗来衡量稳健统计法的稳健性

能，表 2 和表 3 列出了常见统计方法的效率、崩溃点和对微小众数的抵抗。

表 2 稳健统计估计量的相对效率

统计方法	统计估计	均值, $n=50$	均值, $n=500$	SD, $n=50$	SD, $n=500$
经典统计法	样本均值和标准偏差	100%	100%	100%	100%
稳健统计法	中位值和nIQR	66%	65%	38%	37%
	中位值和MADe	66%	65%	37%	37%
	算法A	97%	97%	74%	73%
	Q_n 和 $Q/Hampel$	96%	96%	73%	81%

表 3 统计估计量的崩溃点

统计方法	统计估计量	估计的总体参数	崩溃点	对小众模式的抵抗力
经典统计法	样本均值	均值	0	差
	样本标准偏差	标准偏差	0	差
稳健统计法	样本中位值	均值	50%	好
	nIQR	标准偏差	25%	中
	MADe	标准偏差	50%	中-好
	算法A	均值和标准偏差	25%	中
	Q_n 和 $Q/Hampel$	均值和标准偏差	50%	中 (对于距离 $6s^*$ 的小众模式效果好)

11 能力评定

11.1 能力评定的通用原则

- 11.1.1 用于能力评定的统计量应符合能力验证计划的目标。
- 11.1.2 不同被测量水平和不同轮次能力验证计划的能力评分应方便比较。
- 11.1.3 应对参加者结果进行复核，并确认与能力验证计划的设计所用假设一致，以获得有意义的
能力评定统计量。例如，没有证据表明能力验证样品变质，或参加者总体分布混杂，或有严重违
背数据性质的统计假设。
- 11.1.4 一般情况下，产生行动信号的数量不宜设置为固定比例。

11.2 能力统计量的计算

- 11.2.1 能力验证计划中有三种不同的能力评定方法，用来满足能力验证计划的不同目的。这些方
法包括：
- a)通过与预先设定准则比较进行能力评定，如根据专家、法规或相关标准确定的能力评定准
则；

注：某些情况下，当指定值是参加者结果的稳健均值时，能力评定基于事先规定的允许测量误差 δ_E 或能力评定标准差 σ_{pt} ，其中 $\sigma_{pt} = \delta_E/3$ 。

b)通过与其他参加者比较进行能力评定，如根据统计公议值确定指定值和能力评定标准差；

注：某些情况下，指定值可以是参照值，但 σ_{pt} 可以是参加者测量结果的稳健标准差。

c)通过与声称的测量不确定度比较进行能力评定。

注：使用测量不确定度来评定时，指定值通常是一个适当的参照值。

11.2.2 参加者结果和指定值的差值，或差值与能力评定标准差、最大允许误差、差值的不确定度、指定值等的比值，定义为能力统计量。能力统计量可按照根据能力验证计划的目标和要求设定的判定准则评定参加者的能力。

11.2.3 定量结果的常用能力评定统计量如下：

a)差值 D ：指定值与参加者结果的差值。

判定准则： $|D| < \delta_E$ 时，为可接受结果。

b)百分相对差 $D\%$ ：差值 D 与指定值比值乘以 100。

判定准则： $|D\%| < \delta_E/x_{pt} \times 100\%$ 时，为可接受结果。

注：①对于高度分散或者偏态的结果、顺序响应量、数量有限的不同响应量，百分位数是有效的，如根据允许误差来规定接受范围。但该方法仍应慎用。

c)允许偏差百分比 P_A ：差值 D 与最大允许误差 δ_E 比值乘以 100%。

判定准则： $|P_A| < 100\%$ 时，为可接受结果。

d) z 比分数：差值 D 与能力评定标准差 σ_{pt} 的比值。

判定准则： $|z| \leq 2.0$ 时，为可接受结果； $2.0 < |z| < 3.0$ 时，为给予警告信号； $3.0 \leq |z|$ 时，为不被接受的结果。

e) z' 比分数：差值 D 除以 σ_{pt} 与指定值标准不确定度 $u(x_{pt})$ 的方和根。

判定准则： $|z'| \leq 2.0$ 时，为可接受结果； $2.0 < |z'| < 3.0$ 时，为给予警告信号； $3.0 \leq |z'|$ 时，为不被接受的结果。

注 1：指定值用 8.1.1[a-d]方法确定时，由于未考虑指定值标准不确定度 $u(x_{pt})$ ，因此常用 z' 比分数评价。

注 2：指定值由参加者结果确定，且 $u(x_{pt}) > 0.3\sigma_{pt}$ 时，指定值标准不确定度 $u(x_{pt})$ 不可忽略，常用 z' 比分数评价。

f) ζ 比分数：差值 D 除以参加者自身结果标准不确定度 $u(x_i)$ 与指定值标准不确定度 $u(x_{pt})$ 的方和根。

判定准则： $|\zeta| \leq 2.0$ 时，为可接受结果； $2.0 < |\zeta| < 3.0$ 时，为给予警告信号； $3.0 \leq |\zeta|$ 时，为不被接受的结果。

g) E_n 值：差值 D 除以参加者自身结果扩展不确定度 $U(x_i)$ 与指定值扩展不确定度 $U(x_{pt})$ 的方和根。

判定准则： $|E_n| < 1.0$ 时，为可接受结果； $1.0 \leq |E_n|$ 时，为不被接受的结果。

注 1：用 ζ 比分数和 E_n 值评定时，指定值不可用参加者结果得到。

注 2：用 ζ 比分数和 E_n 值评定时，参加者需采用一致的方法评估不确定度。

注 3：扩展不确定度的包含因子 $k=2$ 。

h)其他的统计方法，可参见 GB/T 28043 和 IUPAC 国际协议等。

11.2.4 当对一个特定被测量使用了一个以上能力验证物品或有一组相关被测量时，可根据一轮能力验证计划中两个或两个以上的结果评定参加者的能力，这样可以对参加者能力进行全面评定。尽量不使用能力比分数的平均值，因为这将掩盖对一个或多个能力验证物品的较差的检测或测量能力，而这正是需要调查的。综合能力比分数常规应用，如统计全部或单个参加者可接受结果的数量（或百分比），以全面了解全部或单个参加者的能力情况。

11.2.5 能力验证计划中某些参加者的结果虽为不满意结果，但可能仍在相关标准或规范规定的允差范围之内，故在能力验证计划中，对参加者的结果进行评价时，通常不作“合格”与否的结论，而是使用“满意/不满意”或“离群”的概念。

11.2.6 能力评定中常规图示法的应用

a)能力评分直方图，可用于展示能力评分分布情况，便于观察分布中存在的异常情况。

b)顺序 z 比分数柱状图，可用于单一样品或样品对，是最常用的 z 比分数图示。该图示的优点是可以直观获得实验室的比对结果信息，并比较参加者在整个能力比对活动中所处的位置。

c)常规质量控制图，可以展示单一参加者每轮计划结果的能力统计量的过程趋势。通过发现图中异常和趋势变化，从而采取适当的措施来保证参加者质量体系的运行。

11.3 评定结果分析

11.3.1 分割水平设计和尧敦图

能力验证计划中，如向每个参加者发放两个基质和被测量水平都接近的样品，这样的样品称为分割水平样品对。分割水平设计中由于样品的相似性，可认为两个样品测试结果的系统误差相同，样品结果之和可以表示测试结果的系统误差，样品结果之差可以表示测试结果的随机误差。可通过两个样品结果的标准化差和标准化和的比分数来评价参加者测试水平和精密度（比分数判定准则与 11.2.3 相同）。

尧敦图是直接以参加者两个样品结果的比分数或测试结果做散点图，该方法可用于展示不同能力验证样品结果的相关性，并指导调查出现行动信号的原因。如尧敦图可通过单个参加者数据点在坐标图中的位置去分析单个参加者结果的误差来源，也可根据整体参加者数据点的分布形状来判定整体参加者结果的误差来源（具体参见附录 D）。

11.3.2 重复性标准差图

在一轮能力验证计划中，参加者获得 m 个重复测试结果时，利用这些结果的作重复性标准差图，可识别均值和标准差异异常的实验室。如给出参加者重复性标准差，则该图可用于识别偏倚异常大的参加者。若存在大量重复测量，该图也可用于识别重复性非常小的参加者（具体参见附录 D）。

11.3.3 曼德尔 h - K 统计量

在一轮能力验证计划中，参加者对每个水平样品重复条件下测试 m 次，根据实验室结果可以

计算曼德尔统计量 h 和 K 。 h 值是实验室间一致性的统计量，可以反映单个参加者的重复性偏差。因此，曼德尔统计量 h 和 K 可用来评价参加者间测试结果的准确度和参加者测试结果的精密度。

注：详细参见附录 D 和 ISO 5725-2。

附录 A

（规范性附录）

符号

d	测量样品的测量结果（平均值）与 CRM 样品测量结果（平均值）之差
$\bar{d}$	测量样品的测量结果（平均值）与 CRM 样品测量结果（平均值）之差的平均值
D	实验室偏倚
$D\%$	以百分数形式表示的实验室偏倚
δ_E	最大允许测量误差
E_n	含参加者结果和指定值不确定度的标准误差
m	每份样品的重复测量次数
p	一轮能力验证计划中参与的实验室数
P_A	参加者结果的“允差偏差百分比”（ D/δ_E ），可表示为百分数
s_r	重复性标准差的估计值
s_s	样品间标准差
s^*	稳健标准差（利用稳健算法计算的标准差）
σ_{pt}	能力评定标准差
σ_r	重复性标准差
σ_R	再现性标准差
u_{hom}	因能力验证物品的差异引起的标准不确定度
u_{stab}	因能力验证过程期间的不稳定性引起的标准不确定度
u_{trans}	因能力验证物品运输的不稳定性引起的标准不确定度
$u(x_i)$	参加者结果的标准不确定度
$u(x_{pt})$	指定值的标准不确定度
$u(x_{ref})$	参考值的标准不确定度
$U(x_i)$	参加者结果的扩展不确定度
$U(x_{pt})$	指定值的扩展不确定度
$U(x_{ref})$	参考值的扩展不确定度

w^*	稳健估计值（利用稳健算法计算的估计值）
x	测试结果
x_{char}	测定获得的特性值
x_{CRM}	CRM 的参考值
x_i	参加者 i 的能力验证结果
x_{pt}	能力验证物品的指定值
x_{ref}	参考值
x^*	稳健平均值（利用稳健算法计算的平均值）
$\bar{x}$	测试结果的算术平均值
z	用于能力评估的性能统计量值
z'	含指定值不确定度的修改 z 值
ζ	含参加者结果和指定值不确定度的修改 z 值

附录 B

(规范性附录)

样品均匀性和稳定性检验

B.1. 均匀性检验

B.1.1 均匀性检验的要求和方法

对能力验证计划所制备的每一个样品编号。从物品总体中随机抽取 10 个或 10 个以上的物品用于均匀性检验。若必要，也可以在特性量可能出现差异的部位按一定规律抽取相应数量的检验样品。

对抽取的每个物品，在重复条件下至少测试 2 次。重复测试的样品应分别单独取样。为了减小测量中定向变化的影响（漂移），物品的所有重复测试应按随机次序进行。

均匀性检验中所用的测试方法，其精密度和灵敏度不应低于能力验证计划预定测试方法的精密度和灵敏度。

特性量的均匀性与取样量有关。均匀性检验所用的取样量不应大于能力验证计划预定测试方法的取样量。

当检测物品有多个待测特性量时，可从中选择有代表性和对不均匀性敏感的特性量进行均匀性检验。

对检验中出现的异常值，在未查明原因之前，不应随意剔除。

可采用单因子方差分析法对检验中的结果进行统计处理。若物品之间无显著性差异，则表明样品是均匀的。

如果 σ 是某个能力验证计划中能力评价标准偏差的目标值， S_s 为样品之间不均匀性的标准偏差。若 $S_s \leq 0.3\sigma$ ，则使用的物品可认为在本能力验证计划中是均匀的。

B.1.2 单因子方差分析

为检验物品的均匀性，抽取 i 个样品($i=1, 2, \dots, m$)，每个样在重复条件下测试 j 次($j=1, 2, \dots, n$)。

每个物品的测试平均值：
$$\bar{x}_i = \sum_{j=1}^n x_{ij} / n_i$$

全部物品测试的总平均值：
$$\bar{\bar{x}} = \sum_{i=1}^m \bar{x}_i / m$$

测试总次数： $N = \sum_{i=1}^m n_i$

物品间平方和： $SS_1 = \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2$ 物品间均方： $MS_1 = \frac{SS_1}{f_1}$

物品内平方和： $SS_2 = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ 物品内均方： $MS_2 = \frac{SS_2}{f_2}$

自由度： $f_1 = m - 1$ $f_2 = N - m$

统计量： $F = \frac{MS_1}{MS_2}$

若 $F <$ 自由度为 (f_1, f_2) 及给定显著性水平 α (通常 $\alpha=0.05$) 的临界值 $F_\alpha(f_1, f_2)$, 则表明物品内和物品间无显著性差异, 物品是均匀的。

注： F 检验是基于均匀性检验数据呈正态分布进行的, 所以应检验数据的正态分布性, 以降低风险。

B.1.3 $S_s \leq 0.3\sigma$ 准则

从能力验证计划制备的物品中随机抽取 i 个样品 ($i=1, 2, \dots, m$), 每个样在重复条件下测试 j 次 ($j=1, 2, \dots, n$)。按 4.2 款计算均方 MS_1 、 MS_2 。

若每个物品的重复测试次数均为 n 次。按下式计算物品之间的不均匀性标准偏差 S_s :

$$S_s = \sqrt{(MS_1 - MS_2)/n}$$

式中, MS_1 物品间均方; MS_2 物品内均方; n 为测量次数。

若 $S_s \leq 0.3\sigma$, 则使用的物品可认为在本能力验证计划中是均匀的。式中 σ 是能力验证计划中能力评价标准偏差的目标值。

注： $S_s \leq 0.3\sigma$ 准则参考 9.3.1 要求, 直接判断物品的不均匀性在能力验证评定中是否可以忽略不计, 更符合均匀性检验的目的。

B.2 稳定性检验

B.2.1 稳定性检验的要求和方法

对于某些性质较不稳定的检测物品, 运输和时间对检测的特性量可能会产生影响。因此, 在物品发送给实验室之前, 需要进行有关条件的稳定性检验。

当检测物品有多个待测特性量时,应选择容易发生变化和有代表性的特性量进行稳定性检验。

稳定性检验的测试方法应是精密和灵敏的,并且具有很好的复现性。

稳定性检验的物品应从包装单元中随机抽取,抽取的物品数具有足够的代表性。在校准能力验证计划中,测量的物品需在参加实验室之间传递,作为被测特性量的监控,在计划运作的始末或期间应作稳定性检验。

稳定性检验的统计方法有 t 检验法、 $|\bar{x}-\bar{y}| \leq 0.3\sigma$ 准则法等。 t 检验法通常用于比较一个平均值与标准值/参考值之间或二个平均值之间是否存在显著性的差异。检验者可根据物品的性质和工作要求选用某一方法。

B.2.2 t 检验法

a) 一系列测量的平均值与标准值/参考值的比较:

按下式计算 t 值:
$$t = \frac{|\bar{x} - \mu| \sqrt{n}}{S}$$

式中, n 为测量次数; $\bar{x}$ 为 n 次测量的平均值; μ 为标准值/参考值; S 为 n 次测量结果的标准偏差。

注: 为了保证平均值和标准偏差的准确度, $n \geq 6$ 。

若 $t <$ 自由度为 $n-1$ 及给定显著性水平 α (通常 $\alpha=0.05$) 的临界值 $t_{\alpha}(n-1)$, 则表明测量的平均值与标准值/参考值之间无显著性差异。

b) 二个平均值之间的一致性:

按下式计算 t 值:
$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \cdot \frac{n_1+n_2}{n_1 \cdot n_2}}}$$

式中, $\bar{x}_1$ 为第一次检验测量数据的平均值;

$\bar{x}_2$ 为第二次检验测量数据的平均值;

s_1 为第一次检验测量数据的标准偏差;

s_2 为第二次检验测量数据的标准偏差;

n_1 为第一次检验测量的测量次数;

n_2 为第二次检验测量的测量次数。

注：为了保证平均值和标准偏差的准确度， n_1 和 n_2 均 ≥ 6 。

若 $t < \text{自由度为 } n-1 \text{ 及给定显著性水平 } \alpha \text{ (通常 } \alpha=0.05 \text{) 的临界值 } t_{\alpha}(n-1)$ ，则表明两个平均值之间无显著性差异。

注： t 检验是基于测试数据呈正态分布进行的，所以应检验数据的正态分布性，以降低风险。

B.2.3 $|\bar{x} - \bar{y}| \leq 0.3\sigma$ 准则

若 $|\bar{x} - \bar{y}| \leq 0.3\sigma$ 成立，则认为被检的样品是稳定的。

式中， σ 为该能力验证计划的能力评价标准偏差目标值；

$\bar{x}$ 为均匀性检验的总平均值；

$\bar{y}$ 为稳定性检验时，对随机抽出样品的测量平均值。

注：稳检抽样数 ≥ 3 ，对每个抽取的样品重复测试 2 次，每次分别单独取样。测量方法与均匀性检验用的测量方法相同。

附录 C

(规范性附录)

稳健分析

C.1.简单稳健分析方法

C.1.1 中位值

中位值是对称分布总体平均值的一种简单估计，该方法对离群值不敏感，可用 $\text{med}(x)$ 表示中位值，计算方法如下：

a) 将参加者提交的 p 个数据按递增顺序，表示为：

$$x_{\{1\}}, x_{\{2\}}, \cdots, x_{\{p\}}$$

b) 令：

$$\text{med}(x) = \begin{cases} x_{\{(p+1)/2\}} & p \text{ 为奇数} \\ \frac{x_{\{p/2\}} + x_{\{1+p/2\}}}{2} & p \text{ 为偶数} \end{cases} \cdots \cdots \text{(C.1)}$$

C.1.2 标准化四分位距 (nIQR)

经多项能力验证计划证实，标准化四分位距法是一种十分有效的稳健统计方法，该方法相对简单并使用广泛。可将参加者结果按递增顺利排列，通过计算 75%分位数和 25%分位数间的差值，然后乘以系数 0.7413 即可得到标准化四分位距。该统计量通常称为“标准化四分位距”(或 nIQR)，按(C.2)所示公式计算：

$$\text{nIQR}(x) = 0.7413[Q_3(x) - Q_1(x)] \cdots \cdots \text{(C.2)}$$

式中：

$Q_1(x)$ 表示 x_i 的 25%分位数 ($i = 1, 2, \dots, p$)

$Q_3(x)$ 表示 x_i 的 75%分位数 ($i = 1, 2, \dots, p$)

如 75%分位数与 25%分位数相同，则 nIQR 为零，需在剔除异常值后计算标准偏差作为总体标准差估计值或 C.5.2 中所述的程序计算稳健标准差。

C.1.3 尺度化中位数绝对差 (MADe)

能力验证中也可使用尺度化中位数绝对差 $\text{MADe}(x)$ 表示正态分布数据的总体标准偏差的估计值，该方法对离群值不敏感。按下式计算 $\text{MADe}(x)$ ：

a) 按下式计算绝对差 $d_i (i = 1, \dots, p)$ ：

$$d_i = |x_i - \text{med}(x)| \dots\dots\dots (\text{C.3})$$

b) 按下式计算 $\text{MADe}(x)$:

$$\text{MADe}(x) = 1.483\text{med}(d) \dots\dots\dots (\text{C.4})$$

C.2 稳健分析：算法 A

应用此算法计算可得到总体平均值和标准差的稳健值。

假设参加者提交的 p 个数据按递增顺序排列表示为：

$$x_{\{1\}}, x_{\{2\}}, \dots, x_{\{p\}}$$

这些数据的稳健平均值和稳健标准差分别用 x^* 和 s^* 表示。

按下式计算 x^* 和 s^* 的初始值：

$$x^* = \text{med}\{x_i\} (i=1,2,\dots,p) \dots\dots\dots (\text{C.5})$$

$$s^* = 1.483\text{med}\{|x_i - x^*|\} (i=1,2,\dots,p) \dots\dots\dots (\text{C.6})$$

注 1：本附录中的算法 A 和 S 来自 GB/T 28043。

注 2：在某些情况下，一半以上的结果 x_i 相同（例如，纺织品针数计数或者血清中电解质）。在这些情况下的 s^* 的初始值将是零，不能直接使用稳健方法。当初始 $s^*=0$ 时，可调查使样品标准差异常变大的粗大离群值后，使用初始 s^* 代替样品标准差。这种取代仅适用于初始 s^* ，替代后迭代算法可继续进行。

根据以下步骤更新 x^* 和 s^* 的值。计算：

$$\delta = 1.5s^* \dots\dots\dots (\text{C.7})$$

对每个 $x_i (i=1,2,\dots,p)$ ，计算：

$$x_i^* = \begin{cases} x^* - \delta & \text{若 } x_i < x^* - \delta \\ x^* + \delta & \text{若 } x_i > x^* + \delta \\ x_i & \text{其他} \end{cases} \dots\dots\dots (\text{C.8})$$

按下式计算 x^* 和 s^* 的新数值：

$$x^* = \sum_{i=1}^p x_i^* / p \dots\dots\dots (\text{C.9})$$

$$s^* = 1.134 \sqrt{\sum_{i=1}^p (x_i^* - x^*)^2 / (p-1)} \dots\dots\dots (\text{C.10})$$

稳健估计值 x^* 和 s^* 可由迭代计算得出，即根据 C.7 至 C.10 多次更新 x^* 和 s^* 的数值，直至

过程收敛。当稳健平均值和稳健标准差（ $\bar{x}^*$ 和 s^* ）的第三位有效数字在连续两次迭代中不再变化时，即可认为过程是收敛的。这是一种可用计算机编程实现的简单方法。收敛的替代准则也可根据能力验证结果的设计和报告要求确定。

C.3 稳健分析：算法 S

若参加者提交能力验证样品的某一被测变量的 m 次重复测量结果，或者 m 个相同的能力验证样品的测量结果，可以应用该算法计算标准差(或极差)。该算法可推出标准差或极差的稳健联合值。

将 p 个标准差或极差以递增顺序排列：

$$w_{\{1\}}, w_{\{2\}}, \dots, w_{\{p\}}$$

用 w^* 表示稳健估计值，与各 w_i 相关的自由度用 ν 表示。（ w_i 等于极差时， $\nu=1$ ；而 w_i 等于 m 个测试结果的标准差时， $\nu=m-1$ ）。从表 C.1 中查找算法所需的 ξ 和 η 值。

按下式计算初始 w^* ：

$$w^* = \text{med}\{w_i\} \quad (i = 1, 2, \dots, p) \dots\dots\dots \text{(C.11)}$$

注：如一半以上的 w_i 为 0，则初始 w^* 为 0，稳健程序将无法正确运行。如初始 w^* 为 0，则用排除可影响平均值的所有极端离群值后，再计算 w^* 替换算术联合平均标准差（或平均极差）。该等替代仅针对初始 w^* ，之后的程序应按说明继续运行。

按如下说明更新。计算：

$$\psi = \eta \times w^* \dots\dots\dots \text{(C.12)}$$

针对各 $w^*(i = 1, 2, \dots, p)$ ，按下式计算：

$$w_i^* = \begin{cases} \psi & \text{若 } w_i > \psi \\ w_i & \text{其他} \end{cases} \dots\dots\dots \text{(C.13)}$$

按下式计算新 w^* ：

$$w^* = \xi \sqrt{\sum_{i=1}^p (w_i^*)^2 / p} \dots\dots\dots \text{(C.14)}$$

稳健估计值 w^* 可由迭代算法得到，即多次更新 w^* ，直到过程收敛。当稳健估计值的第三位有效数字连续两次迭代后数值不再变化时，即可认为过程是收敛的。这是一种可利用计算机编程实现的简单方法。

注：当有一元正态分布的标准差时，算法 S 可用于计算总体标准差（因此，若满足 GB/T 6379.2 中的假设，可用于计算重复性标准差的估计值 s_r ）。

表 C.1 稳健分析所需的因子：算法 S

自由度 ν	限制因子 η	调整因子 ζ
1	1.645	1.097
2	1.517	1.054
3	1.444	1.039
4	1.395	1.032
5	1.359	1.027
6	1.332	1.024
7	1.310	1.021
8	1.292	1.019
9	1.277	1.018
10	1.264	1.017

注： ζ 和 η 的导出过程见ISO 5725-5:1998附录B。

C.4 计算密集性稳健估计量： Q_n 方法和 Hampel 估计量

C.4.1 利用 Q_n 方法决定稳健标准差

Q_n 是拥有高崩溃点、高效率的总体标准差估计量，是正态分布数据（假设无离群值）的无偏估计值。 Q_n 仅采用各参加者报告的某个结果（包括重复值的均值或中位数）。该计算使用数据集中的成对差值，因此与数据的均值或中位数估计值无关。本附录对计算过程进行了修正，以确保该估计量对所有实际数据集规模具有同等适用性。

计算由 p 个报告结果组成的数据集 $Q_n(x_1, x_2, \dots, x_p)$ 的 Q_n ：

a) 计算 $p(p-1)/2$ 绝对差值：

$$d_{ij} = |x_i - x_j| \quad \text{当 } i = 1, 2, \dots, p-1 \text{ 且 } j = i+1, i+2, \dots, p \dots\dots\dots (C.15)$$

b) 按以下顺序排列绝对差值 d_{ij} ：

$$d_{\{1\}}, d_{\{2\}} \dots d_{\{p(p-1)/2\}} \dots\dots\dots (C.16)$$

c) 计算：

$$k = \frac{h(h-1)}{2} \dots\dots\dots (C.17)$$

上式中， k 代表从 h 个对象中选择的数值对的数量，其中：

$$h = \begin{cases} p/2 & p \text{ 为偶数} \\ (p-1)/2 & p \text{ 为奇数} \end{cases} \dots\dots\dots (C.18)$$

d)按下式计算 Q_n ：

$$Q_n = 2.2219d_{\{k\}}b_p \dots\dots\dots (C.19)$$

其中，数据点的数量为 p 时， b_p 从表C.2中查找，或者 $p > 12$ 时，按下式计算：

$$b_p = \frac{1}{r_p + 1} \dots\dots\dots (C.20)$$

其中：

$$r_p = \begin{cases} \frac{1}{p} \left[1.6019 + \frac{1}{p} \left(-2.128 - \frac{5.172}{p} \right) \right] & p \text{ 为奇数} \\ \frac{1}{p} \left[3.6756 + \frac{1}{p} \left(1.965 + \frac{1}{p} \left(6.987 - \frac{77}{p} \right) \right) \right] & p \text{ 为偶数} \end{cases} \dots\dots\dots (C.21)$$

注1：2.2219为修正因子，用于在数据数量 p 较大时得出标准差的无偏估计值。数据数量 p 较小时使用的修正因子 b_p 从表C.2中查找； $p > 12$ 时， r_p 的计算如参考文件中所述，需利用大量模拟和回归分析。

表C.2修正因子 b_p ($2 \leq p \leq 12$)

p	2	3	4	5	6	7	8	9	10	11	12
b_p	0.993	0.993	0.513	0.844	0.612	0.858	0.669	0.873	0.720	0.889	0.757

C.4.2 利用 Q 方法决定稳健标准差

利用 Q 方法，可针对不同实验室报告的能力验证结果得出具有高失效点和高效率的标准差估计值。该 Q 方法不仅适用于拥有离群值的结果数据，还适用于（因不连续定量数据或舍入误差造成的）多数验证结果相同的情况。在这种情况下，多数成对差异为0，因此其它类似于 Q 方法的方法均无效。

各参加者仅报告一个结果（包括重复结果的均值或中位数）或报告重复结果时，均可使用该 Q 方法。在计算中直接使用重复结果可提高该方法的效率。

该计算使用数据集中的成对差异，因此与数据的均值或中位数估计值无关。与有限步骤算法一同用于C.5.3中描述的Hampel估计量时，该方法称为 Q /Hampel。

按实验室对所报告的测量结果进行分组：

$$\underbrace{y_{11}, \dots, y_{1n_1}}_{Lab1}, \underbrace{y_{21}, \dots, y_{2n_2}}_{Lab2}, \dots, \underbrace{y_{p1}, \dots, y_{pn_p}}_{Labp}$$

计算所有实验室间绝对差异的累积分布函数：

$$H_1(x) = \frac{2}{p(p-1)} \sum_{1 \leq i \leq j \leq p} \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{m=1}^{n_j} I\{|y_{ik} - y_{jm}| \leq x\} \quad \dots\dots\dots (C.22)$$

其中， $I\{|y_{ik} - y_{jm}| \leq x\} = \begin{cases} 1 & \text{若 } |y_{ik} - y_{jm}| \leq x \\ 0 & \text{其他} \end{cases}$ 表示示性函数。

按以下所示表示 $H_1(x)$ 的间断点：

$$x_1, \dots, x_r \text{ 其中 } x_1 < x_2 < \dots < x_r$$

针对所有正间断点 $x_1, \dots, x_r$ 运行以下公式：

$$G_1(x_i) = \begin{cases} 0.5 \cdot (H_1(x_i) + H_1(x_i - 1)) & \text{若 } i \geq 2 \\ 0.5 \cdot H_1(x_i) & \text{若 } i = 1; x_i > 0 \end{cases} \quad \dots\dots\dots (C.23)$$

并且设：

$$G_1(0) = 0$$

在不连续点 $0 \leq x_1 < x_2 < \dots < x_r$ 间进行线性插值，计算 $[0, x_r]$ 区间外所有 x 值的函数 $G_1(x)$ 。

计算各实验室验证结果的稳健标准差 s^* ：

$$s^* = \frac{G_1^{-1}(0.25 + 0.75 \cdot H_1(0))}{\sqrt{2} \Phi^{-1}(0.625 + 0.375 \cdot H_1(0))} \quad \dots\dots\dots (C.24)$$

其中， $H_1(0)$ 如(C.22)所示公式计算，在数据集中无准确联系的情况下等于 0， $G_1^{-1}(q)$ 为分位数 q 对应的 x 值， $\Phi^{-1}(q)$ 是标准正态分布的反函数。

C.4.3 利用 Hampel 估计量决定稳健均值

Hampel 估计量用于估计各实验室报告结果的整体均值，具有较大的稳健性与效率。当前无法得出 Hampel 估计量的准确公式，本段介绍了两种较为常用的算法。第一种算法计算起来较为简单，但不同的计算过程可导致不同的结果。第二种算法仅依据基础标准差，因此得出的结果相同。

C.4.3.1 下述计算为得出 Hampel 位置估计量的迭代重复加权计划：

- a) 将数据表示为 $x_1, x_2 \dots x_p$ ；
- b) 将 x^* 设为 $\text{med}(x)$ ；
- c) 将 s^* 设为合适的标准差文件估计值，如 MADe 、 Q_n 或利用 Q 方法得出的 s^* ；
- d) 针对各数据点 x_i ，按下式计算 q_i ：

$$q_i = \left| \frac{x_i - x^*}{s^*} \right|$$

e) 按下式计算权重 w_i :

$$w_i = \begin{cases} 0 & |q| > 4.5 \\ (4.5 - q)/q & 3 < |q| \leq 4.5 \\ 1.5/q & 1.5 < |q| \leq 3.0 \\ 1 & |q| \leq 1.5 \end{cases}$$

f) 按下式重新计算 x^* :

$$x^* = \frac{\sum_{i=1}^p w_i x_i}{\sum_{i=1}^p w_i}$$

重复第d-f步，直至 x^* 趋同。相邻两次迭代运算得出的 x^* 的差异小于 $0.01s^*/\sqrt{p}$ （相当于标准误差 x^* 的1%），假设趋同发生。如需确保准确度，也可使用趋同标准。

选择初始 x^* 和/或 s^* 时可能未包括数据集中的部分重要数据，因此无法保证Hampel估计量是特殊解决方法或可产生的最佳解决方法。因此，能力验证提供者应采取措施，检验所选位置是否有效或制定选择位置时需遵循的明确规则。最常用的规则是选择最靠近中位数的位置。通过审核结果确保数据集中的大部分数据在 $|q| > 4.5$ 范围内也有助于确定可行方法。

注 1：对于正态分布数据而言，Hampel 估计量的效率约为 96%。

注 2：该估计量的使用示例见附录 E。

注3：改变加权函数，可增大Hampel估计量的效率或微小众数耐受力。加权函数的一般形式如下：

$$w_i = \begin{cases} 0 & |q| > c \\ a(c-q)/[q(c-b)] & b < |q| \leq c \\ a/q & a < |q| \leq b \\ 1 & |q| \leq a \end{cases}$$

其中， a 、 b 和 c 为调整参数。在本附录中， $a = 1.5$ ， $b = 3.0$ ， $c = 4.5$ 。极差越大，效率越大；极差越小，对离群值或微小众数耐受力越强。

C.4.3.2 以下有限步骤算法产生 Hampel 位置估计值。

计算各实验室的算术均值 $y_1, y_2, \dots, y_p$

通过以下等式计算稳健均值 x^* :

$$\sum_{i=1}^p \psi \left[\frac{y_i - x^*}{s^*} \right] = 0 \quad \dots\dots\dots (C.25)$$

其中:

$$\psi(q) = \begin{cases} 0 & q \leq -4.5 \\ -4.5 - q & -4.5 < q \leq -3 \\ -1.5 & -3 < q \leq -1.5 \\ q & -1.5 < q \leq 1.5 \\ 1.5 & 1.5 < q \leq 3 \\ 4.5 - q & 3 < q \leq 4.5 \\ 0 & q > 4.5 \end{cases} \quad \dots\dots\dots (C.26)$$

s^* 是根据Q方法算出的稳健标准差。

准确结果可通过有限次运算得出,即无需迭代运算。 ψ 对于 x^* 而言具有部分线性,等式(C.25)

左侧的插值节点(此处视为 x^* 的一个函数)如下:

计算所有插值节点:

对于第一个数值 y_1 :

$$d_1 = y_1 - 4.5 \cdot s^*, d_2 = y_1 - 3 \cdot s^*, d_3 = y_1 - 1.5 \cdot s^*, d_4 = y_1 + 1.5 \cdot s^*, d_5 = y_1 + 3 \cdot s^*, d_6 = y_1 + 4.5 \cdot s^*$$

对于第二个数值 y_2 :

$$d_7 = y_2 - 4.5 \cdot s^*, d_8 = y_2 - 3 \cdot s^*, d_9 = y_2 - 1.5 \cdot s^*, d_{10} = y_2 + 1.5 \cdot s^*, d_{11} = y_2 + 3 \cdot s^*, d_{12} = y_2 + 4.5 \cdot s^*$$

以此类推至所有数值 $y_3, \dots, y_p$ 。

将这些数据 $d_1, d_2, d_3, \dots, d_{6 \cdot p}$ 按从小到大的顺序排列为:

$$d_{\{1\}}, d_{\{2\}}, d_{\{3\}}, \dots, d_{\{6 \cdot p\}}$$

将 m 设为 $1, \dots, (6 \cdot p - 1)$, 针对各数据运行以下等式:

$$p_m = \sum_{i=1}^p \psi \left[\frac{y_i - d_{\{m\}}}{s^*} \right]$$

检查是否符合以下条件:

若 $p_m = 0$ 。则 $d_{\{m\}}$ 为等式(C.25)的结果。

若 $p_{m+1} = 0$ 。则 $d_{\{m+1\}}$ 为等式(C.25)的结果。

若 $p_m \cdot p_{m+1} < 0$ 。则 $x_m = d_{\{m\}} - \frac{p_m}{\frac{p_{m+1} - p_m}{d_{\{m+1\}} - d_{\{m\}}}}$ 为等式(C.25)的解。

用 S 表示等式(C.25)所有结果组成的数据集。

最靠近中位数的结果 $x^* \in S$ 用作位置参数 x^* , 即:

$$|x^* - med(y_1, y_2, \dots, y_p)| = \min \{ |x - med(y_1, y_2, \dots, y_p)|; x \in S \}$$

可能存在多个结果。如同时有两个或没有数据最靠近中位数，则将中位数用作位置参数 x^* 。

附录 D
(资料性附录)
评定结果分析 (图示法)

例 1 尧敦图

该图是对每个参加者的测试结果 $x_{A,i}$ 和 $x_{B,i}$ 算出的平均值和标准差对应的能力统计量 ($z_{A,i}$ 和 $z_{B,i}$) 分别描点得到。平均值和标准差的计算可根据经典统计方法或稳健统计方法等进行。图中常用置信椭圆来表示散点的置信区间, 置信椭圆的表达式可以用霍特林统计量 T^2 、 $x_{A,i}$ 与 $x_{B,i}$ 的相关系数 $\hat{\rho}$ 来表示:

$$z_A^2 - 2\hat{\rho} z_A z_B + z_B^2 = (1 - \hat{\rho}^2) T^2 \dots\dots\dots (D.1)$$

式中, 霍特林统计量 $T^2=2\{(p-1)/(p-2)\}F_{(1-\alpha)}(2,p-1)$; $F_{(1-\alpha)}(2,p-1)$ 是自由度为 2 和 $(p-1)$ 的 F 分布的 $(1-\alpha)$ 分位数; p 为参加者数量。

表 D1 给出了检测两个相似能力验证物品的抗体浓度的数据, 列出了基于算法 A 计算出的稳健平均值和标准差对应的能力统计量 ($z_{A,i}$ 和 $z_{B,i}$) 。

表 D1 两个相似过敏源的抗体浓度的数据和计算

实验室 i	数据		z 值	
	过敏原 A $x_{A,i}$	过敏源 B $x_{B,i}$	过敏源 A $z_{A,i}$	过敏源 B $z_{B,i}$
1	12.95	9.15	0.427	0.515
2	6.47	6.42	-1.540	-0.428
3	11.40	6.60	-0.043	-0.366
4	8.32	4.93	-0.978	-0.942
5	18.88	13.52	2.228	2.023
6	15.14	8.22	1.092	0.194
7	10.12	7.26	-0.432	-0.138
8	17.94	9.89	1.942	0.770
9	11.68	4.17	0.042	-1.204
10	12.44	7.39	0.272	-0.093
11	6.93	7.78	-1.400	0.042
12	9.57	5.80	-0.599	-0.642
13	11.73	5.77	0.057	-0.652

14	12.29	6.97	0.227	-0.238
15	10.95	6.23	-0.180	-0.493
16	10.95	5.90	-0.180	-0.607
17	11.17	7.74	-0.113	0.028
18	11.20	8.63	-0.104	0.335
19	7.64	3.74	-1.185	-1.353
20	12.17	7.33	0.190	-0.114
21	10.71	5.70	-0.253	-0.676
22	7.84	6.07	-1.124	-0.549
23	20.47	15.66	2.710	2.762
24	12.60	11.76	0.321	1.415
25	11.37	4.91	-0.052	-0.949
26	11.36	13.51	-0.055	2.019
27	10.75	5.48	-0.241	-0.752
28	12.21	9.77	0.203	0.729
29	7.49	5.82	-1.230	-0.635
均值	11.54	7.66	0.00	0.00
标准差	3.29	2.90	1.00	1.00
相关系数	0.706		0.706	

注 1：表中数据为每升（L）样品中千单位（kU）抗体数，其中单位（U）是由国际标准物质的浓度定义。

注 2：表中 z 值的计算使用了未圆整的平均值和标准差数据，而不是使用本表底的舍入修约数值计算而得的。

尧敦图可以采用 Origin 软件进行作图：首先将表 D1 中的 $z_{A,i}$ 和 $z_{B,i}$ 作散点图；再在 Origin—Graph—Add Function Graph 中输入 z_A 关于 z_B 的置信椭圆方程以及 z_A 的取值范围，作置信椭圆曲线图。其中置信椭圆方程：

$$z_B = \hat{\rho} z_A \pm \sqrt{(1 - \hat{\rho}^2)(T^2 - z_A^2)} \quad -T \leq z_A \leq T \quad \dots\dots\dots (D.2)$$

尧敦图也可通过 R 语言程序进行作图，图 D1 的 R 语言代码见附录 F。

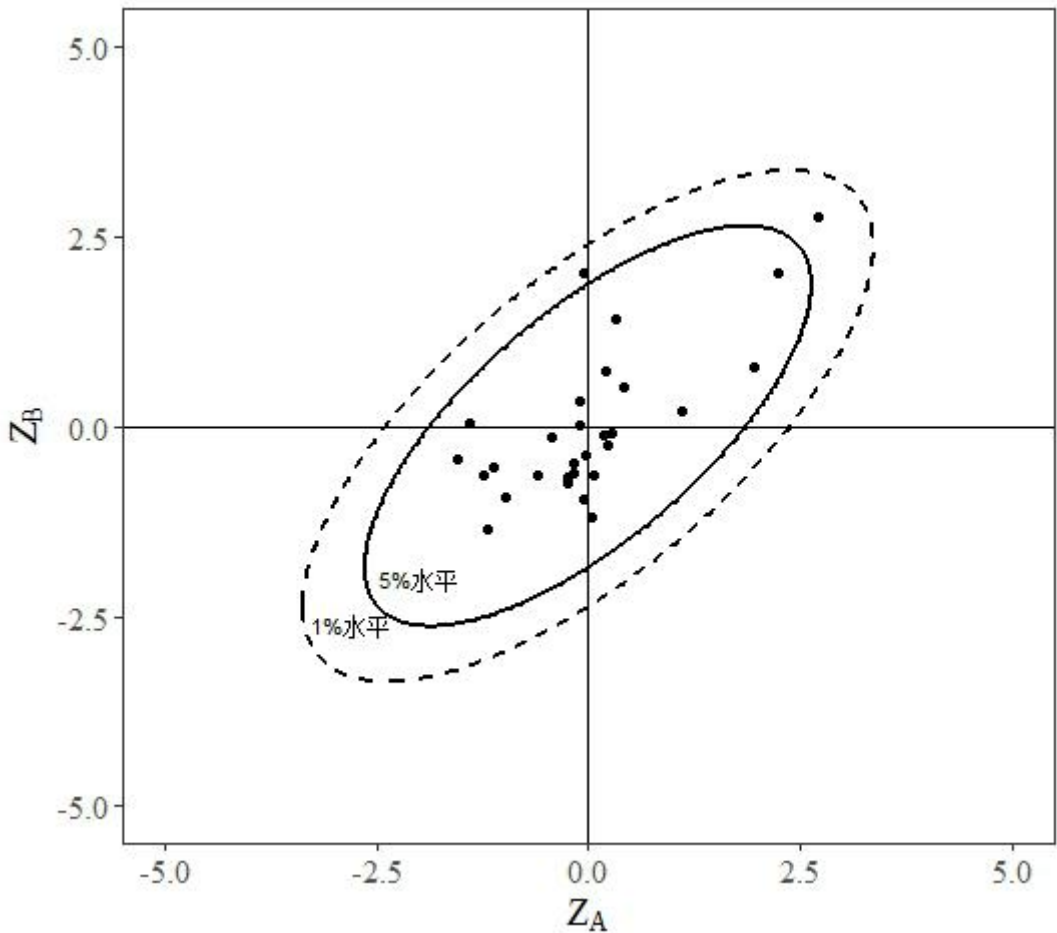


图 D1 两个相似过敏源的抗体浓度的 z 值尧敦图

图 D1 显示两个参加者（编号为 5 和 23）的 z 值在图中右上象限，因此有一致的正向偏 26 号参加者对能力验证物品 B 有一个高 z 值，对能力验证物品 A 得到的 z 值为-0.055，因此该实验室的重复性较差。

第 5、23 和 26 号实验室，应将其结果作为警戒信号处理，并在下一轮能力验证计划中检查这三个实验室的结果位于何处。图表和相关系数表明了 z 值的趋势（正向或负向），因此他们有机会通过更详细的指导来改善测量方法。

例 2 重复性标准差图

重复性标准差图是对每个参加者的实验室内标准差 s_i 和实验室均值 $\bar{x}_i$ 分别描点得到。或将标准差替换为重复测量结果的极差。令：

x^* 为用算法 A 计算的 $x_1, x_2, \dots, x_p$ 的稳健平均值。

w^* 为用算法 S 计算的 $s_1, s_2, \dots, s_p$ 的稳健估计值。

并假定数据服从正态分布。在原假设“各参加者的均值或实验室内标准差无差异”成立时，统计量：

$$\left(\sqrt{m} \frac{x_i - x^*}{w^*}\right)^2 + (\sqrt{2(m-1)} \ln(\frac{s_i}{w^*}))^2 \dots\dots\dots (D.3)$$

近似服从自由度为 2 的 χ^2 分布。因此，在以标准差 s 为纵轴、平均值 x 为横轴的坐标系中显著性水平为 1% 时的临界区域由下式确定：

$$s = w^* \exp\left\{\pm \frac{1}{\sqrt{2(m-1)}} \sqrt{\chi^2_{2;0.99} - \left(\sqrt{m} \frac{x - x^*}{w^*}\right)^2}\right\} \dots\dots\dots (D.4)$$

$$x^* - w^* \sqrt{\frac{\chi^2_{2;0.99}}{m}} \leq x \leq x^* + w^* \sqrt{\frac{\chi^2_{2;0.99}}{m}} \dots\dots\dots (D.5)$$

表D2是血浆能力验证物品中特定抗体浓度的测试结果。每个实验室在重复性条件下做四次重复测试。利用以上给出的公式可得图D2。该图表明一部分实验室收到了行动或警戒信号。

表 D2 血浆能力验证物品中特定抗体的浓度（每个实验室对一份样本进行 4 次重复测定）

实验室	平均值 kU/L	标准差 kU/L
1	2.15	0.13
2	1.85	0.21
3	1.80	0.08
4	1.80	0.24
5	1.90	0.36
6	1.90	0.32
7	1.90	0.14
8	2.05	0.26
9	2.35	0.39
10	2.03	0.53
11	2.08	0.25
12	1.25	0.24
13	1.13	0.72
14	1.00	0.26
15	1.08	0.17

16	1.20	0.32
17	1.35	0.4
18	1.23	0.36
19	1.23	0.33
20	0.90	0.43
21	1.48	0.40
22	1.20	0.55
23	1.73	0.39
24	1.43	0.30
25	1.28	0.22
稳健平均值	1.57	
稳健标准差		0.34
注：表中数据为每升（L）样品中千单位（kU）抗体数，其中单位 U 是由国际标准物质的浓度定义。		

重复性标准差图可以采用 Origin 软件进行作图：通过计算和查找获得 χ^2 、 x^* 、 w^* 、 m 的具体值，代入等式（D.4）和（D.5），得到 x 关于 s 的等式以及 x 的取值范围。首先将表 D2 中的平均值和标准差作散点图；再在 Origin—Graph—Add Function Graph 中输入关于 s 与 x 的等式以及 x 的取值范围，作函数曲线图。

重复性标准差图也可通过 R 语言程序进行作图，图 D2 的 R 语言代码见附录 F。

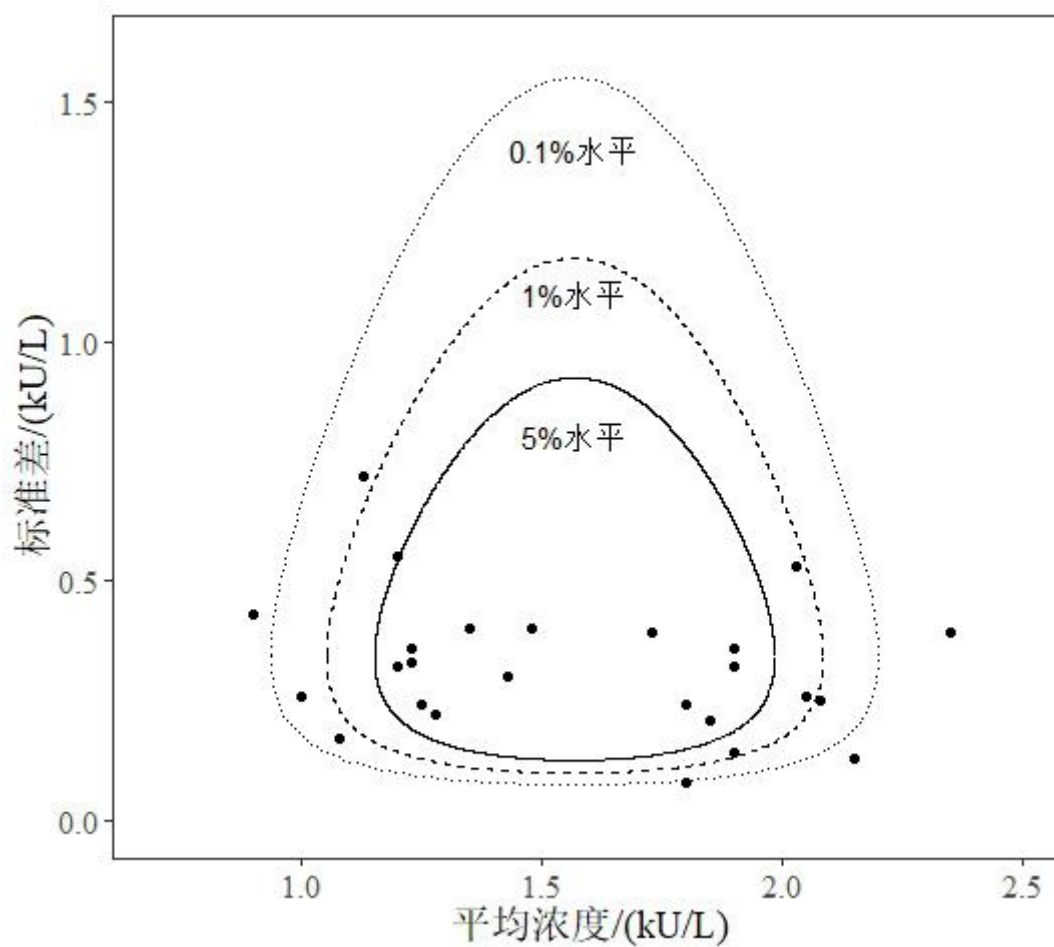


图 D2 重复性标准差图

例 3 曼德尔 h - K 统计量

在一轮能力验证计划中，参加者对每个水平样品重复条件下测试 m 次，根据实验室结果可以计算曼德尔统计量 h 和 K 。

h 值类似于 Z 比分数：

$$h_{ij} = \frac{\bar{Y}_{ij} - \bar{\bar{Y}}_j}{\sqrt{\frac{1}{(P_j - 1)} \sum_{i=1}^{P_j} (\bar{Y}_{ij} - \bar{\bar{Y}}_j)^2}}$$

式中： $\bar{Y}_{ij}$ 为 i 实验室在 j 水平测试结果的平均值；

$\bar{\bar{Y}}_j$ 为在 j 水平所有实验室测试结果的总平均值；

P_j 为实验室数量。

对于实验室测试重复性标准差用 K_{ij} 评价：

$$K_{ij} = \frac{S_{ij}}{\overline{S_{ij}}}$$

式中： S_{ij} 为 i 实验室在 j 水平的重复性标准差；

$\overline{S_{ij}}$ 为联合室内标准差，它等于在 j 水平所有实验室重复性标准差的平均值，即

$$\overline{S_{ij}} = \sqrt{\sum S_{ij}^2 / P_j}。$$

注： S_{ij} 近似服从正态分布时， $\overline{S_{ij}}$ 取算术平均值。对各实验室的结果，需用柯克伦检验剔除离群值后再计算平均值。在能力验证计划中可以用算法 S 计算标准差的稳健联合值 w^* 作为 S_{ij} 的公议值。

在 ISO 5725-2 中，根据 h_{ij} 和 K_{ij} 在显著性水平 5%和 1%临界域的位置，确定实验室的结果为正常值、歧离值或离群值。即：

- (1) 若统计量 h_{ij} 或 $K_{ij} \leq 5\%$ 显著性水平临界值，实验室结果为可接受的正常值；
- (2) 5% 显著性水平临界值 $\leq h_{ij}$ 或 $K_{ij} \leq 1\%$ 临界值，为歧离值；
- (3) h_{ij} 或 $K_{ij} \geq 1\%$ 临界值，为离群值。

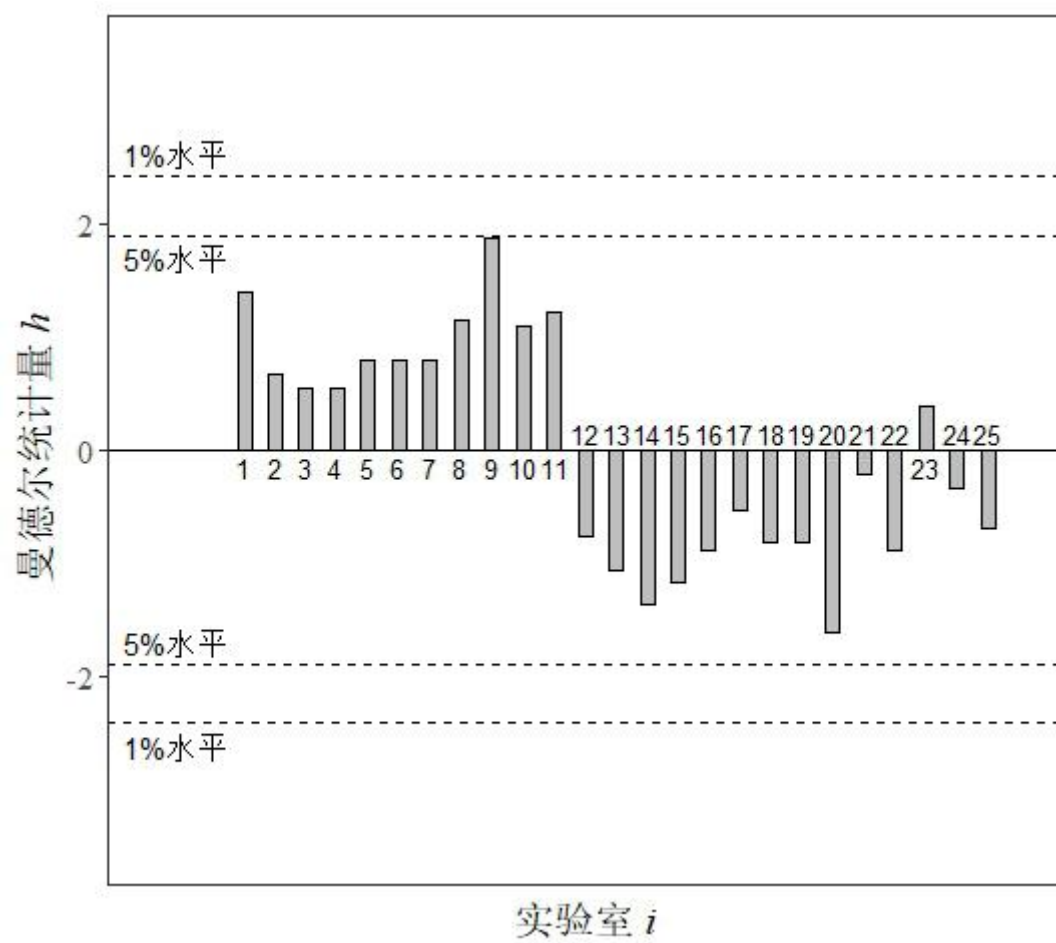
注：详细参见 ISO 5725-2。

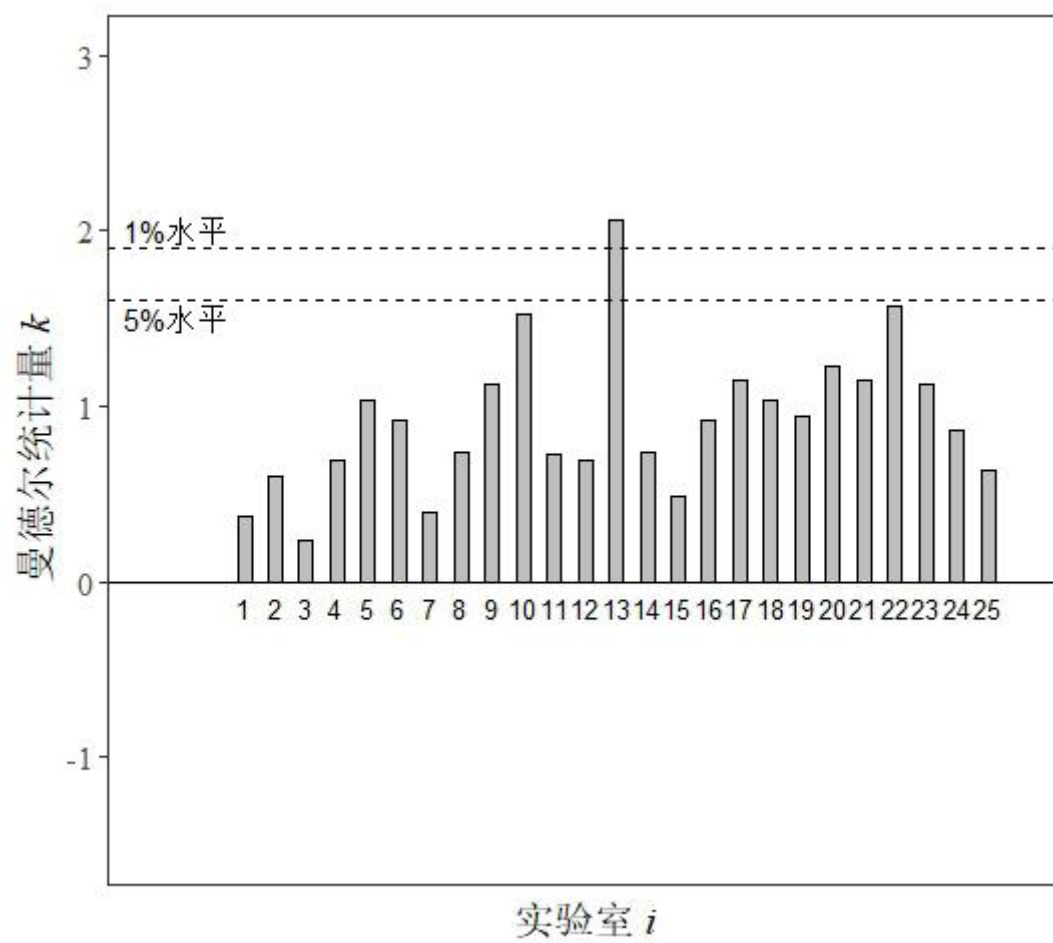
表 D3 是基于血浆能力验证物品中特定抗体浓度的测试结果的 h_{ij} 和 K_{ij} 计算。每个实验室在重复性条件下做四次重复测试。作出的曼德尔 h 图和 k 图。该图显示出各实验室准确度和精密度的具体情况。

表 D3 基于血浆能力验证物品中特定抗体的浓度测试结果的 h_{ij} 和 K_{ij} 计算

实验室	平均值 kU/L	标准差 kU/L	h_{ij}	K_{ij}
1	2.15	0.13	1.39	0.37
2	1.85	0.21	0.67	0.60
3	1.80	0.08	0.55	0.23
4	1.80	0.24	0.55	0.69
5	1.90	0.36	0.79	1.03

6	1.90	0.32	0.79	0.92	
7	1.90	0.14	0.79	0.40	
8	2.05	0.26	1.15	0.74	
9	2.35	0.39	1.87	1.12	
10	2.03	0.53	1.10	1.52	
11	2.08	0.25	1.22	0.72	
12	1.25	0.24	-0.77	0.69	
13	1.13	0.72	-1.06	2.06	
14	1.00	0.26	-1.37	0.74	
15	1.08	0.17	-1.18	0.49	
16	1.20	0.32	-0.89	0.92	
17	1.35	0.40	-0.53	1.15	
18	1.23	0.36	-0.82	1.03	
19	1.23	0.33	-0.82	0.94	
20	0.90	0.43	-1.61	1.23	
21	1.48	0.4	-0.22	1.15	
22	1.20	0.55	-0.89	1.57	
23	1.73	0.39	0.38	1.12	
24	1.43	0.30	-0.34	0.86	
25	1.28	0.22	-0.70	0.63	
平均值 $\overline{Y_j}$	1.57	h_{ij} (1%,p=25)	± 2.42	K_{ij} (1%,m=4,p=25)	1.90
标准差 $\overline{S_{ij}}$	0.35	h_{ij} (5%,p=25)	± 1.90	K_{ij} (5%,m=4,p=25)	1.60

图 D3 曼德尔统计量 h 图

图 D4 曼德尔统计量 k 图

附录 E

（资料性附录）

参加者结果统计计算示例

利用统计方法计算参加者提交结果的位置中心和离散性的估计值，是确定指定值和能力评定准则的重要方法。目前，能力验证常见的统计方法有经典统计方法、稳健统计方法等。能力验证提供者需根据能力验证的目的、自身条件和数据情况，如结果数、离群值数量、数据分布状况和对称性等，再结合方法本身崩溃点和效率，选择合适的统计方法。经典统计方法由于其崩溃点为零，统计计算时要求数据中无任何离群值，即需用格拉布斯（Grubbs）准则等统计方法剔除离群值后计算均值和标准差。而对于简单稳健统计方法（如中位值、NIQR、MADe）、迭代算法 A 和算法 S 等统计方法，均有着不错的抗离群值和抗微小众数的能力，因此适用于大多数回收数据服从（或近似服从）正态分布的能力验证计划。当数据中离群值的比例超过 20%，或者数据的分布中出现双峰或多峰时，前面提到的方法将不再适用，计算得到的估计值也不再可靠。对此，能力验证结果的处理可以采用复杂稳健统计方法（如 Q/Hampel、Qn/Hampel）或者自助法等。

例 1 经典算法应用

表 E1 给出一轮能力验证计划的 15 个结果，其中剔除 1 个明显错误的结果后，剩下 14 个结果的直方图呈近似正态分布，如图 E1 所示。采用经典统计方法来计算经格拉布斯检验剔除异常值后的结果的均值和标准差。

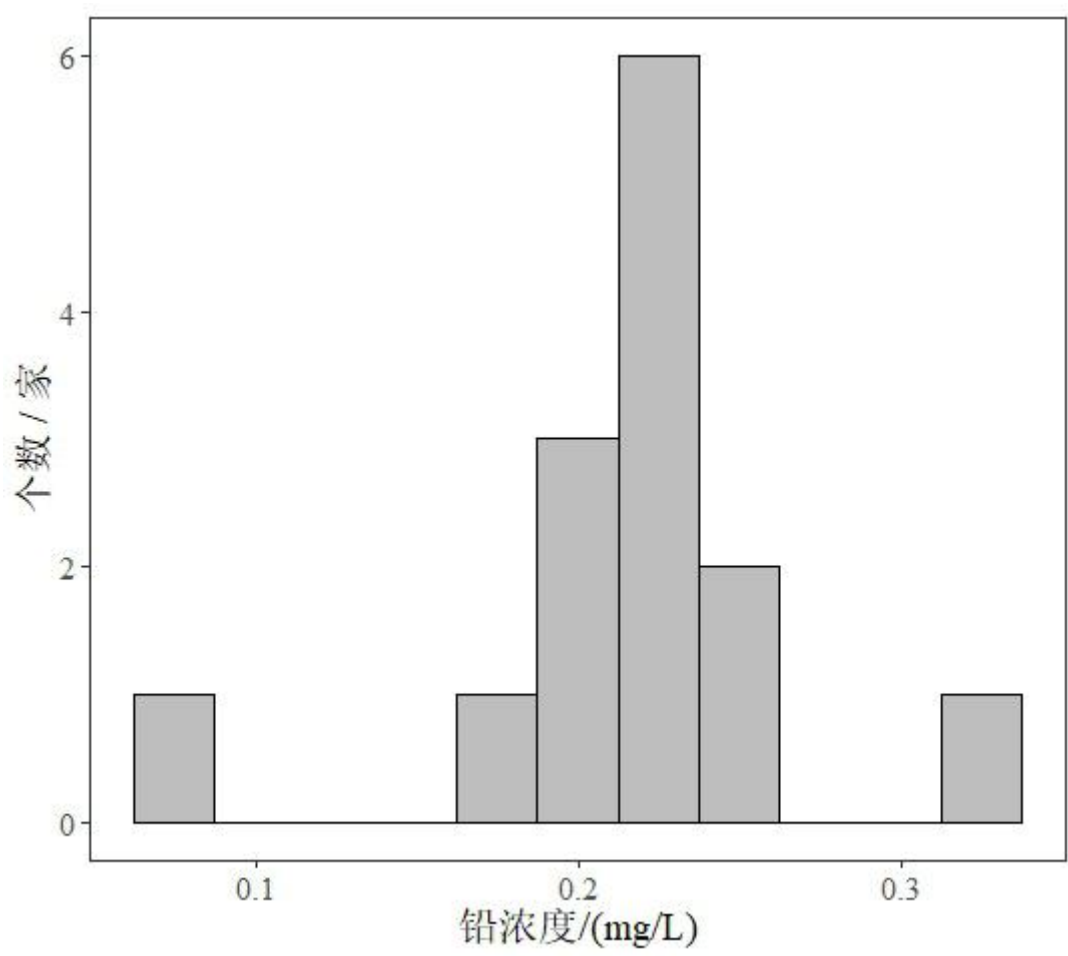


图 E1 参加者结果直方图

表 E1 水中铅含量测试结果数据的计算

实验室 代码	$x_i(\text{mg/L})$	实验室 代码	$x_i(\text{mg/L})$ (剔除明显错误后)	实验室 代码	$x_i(\text{mg/L})$ (格拉布斯检验剔除后)
001	0.206	001	0.206	001	0.206
002	0.321	002	0.321	002	0.321
003	0.070	003	0.070&	004	0.185
004	0.185	004	0.185	005	0.236
005	0.236	005	0.236	006	0.227
006	0.227	006	0.227	007	0.198
007	0.198	007	0.198	008	0.260
008	0.260	008	0.260	009	0.230
009	0.230	009	0.230	010	0.235

010	0.235	010	0.235	011	0.202
011	0.202	011	0.202	013	0.242
012	32.245*	013	0.242	014	0.228
013	0.242	014	0.228	015	0.231
014	0.228	015	0.231		
015	0.231				
				均值 $\bar{x}$	0.231
				标准差 S	0.034
				$\bar{x}$ 的标准不 确定度 u_x	0.009

注：“*”表示为有明显错误的结果；“&”表示为格拉布斯检验的异常值。格拉布斯检验方法具体参见 GB/T 4883。

例 2 中位值、nIQR(MADe)、算法 A 应用

饮用水中除草剂（莠去津）含量的测定能力验证计划共有 34 个参加者。为清晰起见，将参加者的原始数据按升序进行排列，如表 E2 所示。由于实验室家数大于 17 家，可以考虑使用稳健统计方法来计算结果的均值和标准差。用稳健统计方法做数据分析前应使用直方图或核密度图来对结果进行直观检查，以确认结果的预期分布，识别异常值或意外变异的来源。如图 E2 和图 E3 所示，测试结果呈单峰分布，且较为对称，可看做近似正态分布。

表 E2 生活饮用水中莠去津（Atrazine）含量测试结果数据

实验室代码	$x_i(\text{mg/L})$	实验室代码	$x_i(\text{mg/L})$
001	0.0400	017	0.2640
005	0.0550	018	0.2670
007	0.1780	022	0.2700
030	0.2020	012	0.2730
003	0.2060	019	0.2740
027	0.2270	029	0.2740
031	0.2280	013	0.2780
026	0.2300	014	0.2811
008	0.2300	034	0.2870

006	0.2350	033	0.2870
025	0.2360	021	0.2880
024	0.2370	004	0.2890
016	0.2430	020	0.2950
028	0.2440	002	0.2960
009	0.2450	011	0.3110
015	0.2555	032	0.3310
023	0.2600	010	0.4246

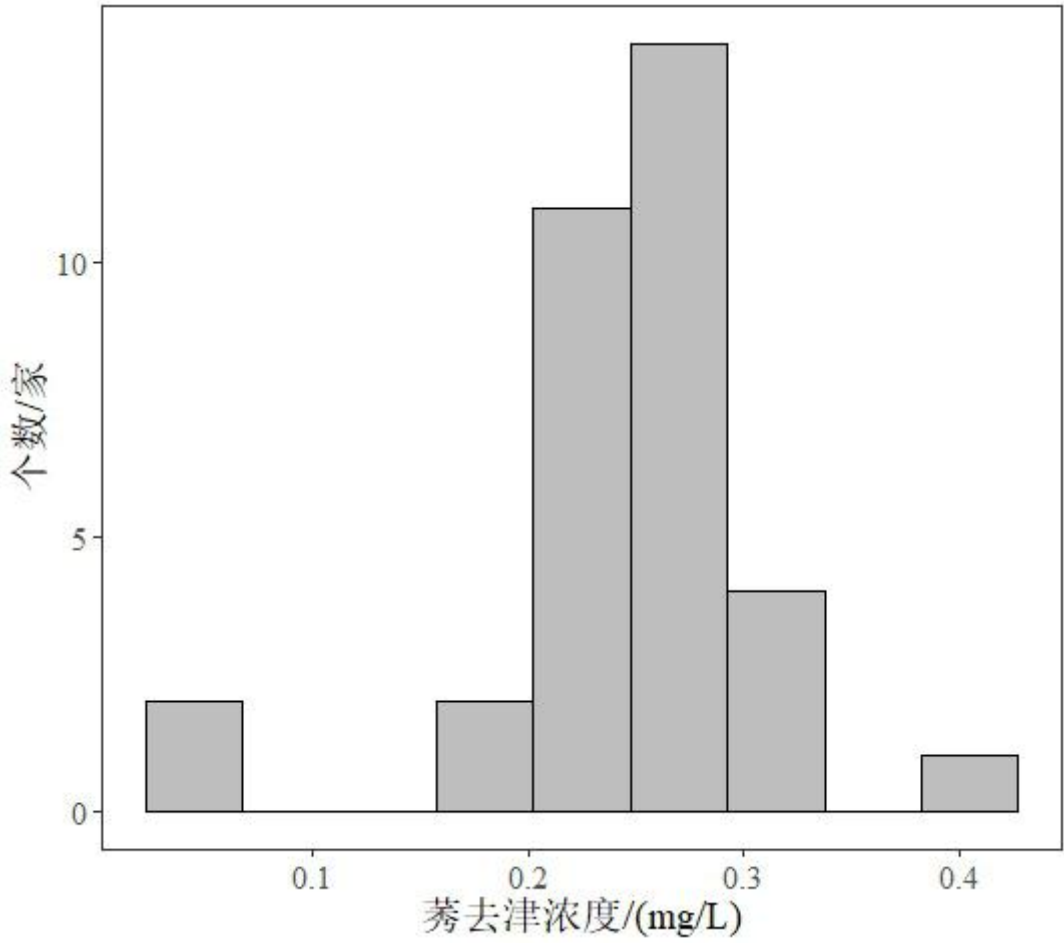


图 E2 参加者结果直方图

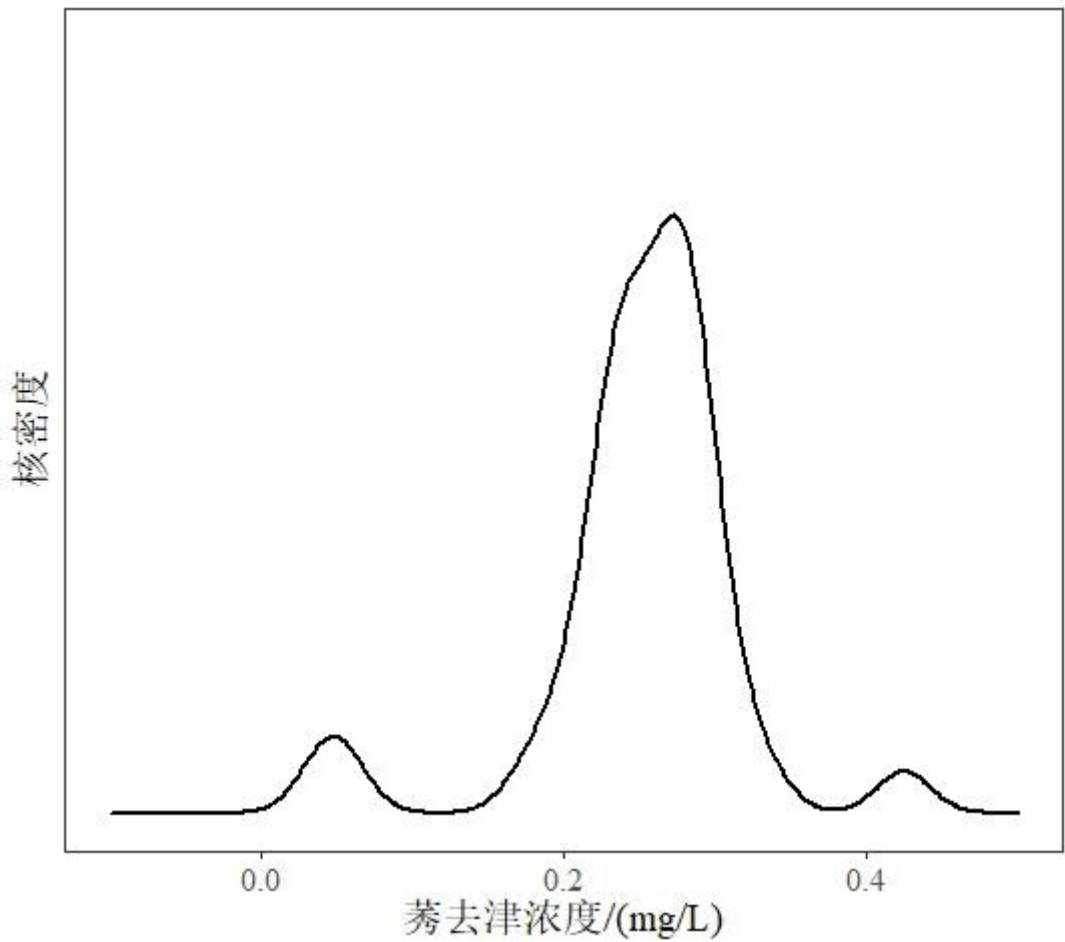


图 E3 参加者结果核密度图

如表 E3 所示，给出了用算法 A 计算得到的稳健平均值和稳健标准差。在经过 6 次迭代后，稳健平均值和稳健标准差的第三位有效数字不再发生变化。而表 E4 中给出了用四分位法计算得到的中位值和 nIQR(MADe)。

表 E3 饮用水中莠去津含量的稳健平均值和稳健标准差的计算

实验室代码	$x_i(\text{mg/L})$	第一次 迭代	第二次 迭代	第三次 迭代	第四次 迭代	第五次 迭代	第六次 迭代
$x^*-\delta$		0.204163	0.199732	0.198466	0.198037	0.197865	0.197790
$x^*+\delta$		0.319837	0.315969	0.315871	0.316065	0.316185	0.316243
001	0.0400	0.2042	0.1997	0.1985	0.1980	0.1979	0.1978
005	0.0550	0.2042	0.1997	0.1985	0.1980	0.1979	0.1978

007	0.1780	0.2042	0.1997	0.1985	0.1980	0.1979	0.1978
030	0.2020	0.2042	0.2020	0.2020	0.2020	0.2020	0.2020
003	0.2060	0.2060	0.2060	0.2060	0.2060	0.2060	0.2060
027	0.2270	0.2270	0.2270	0.2270	0.2270	0.2270	0.2270
031	0.2280	0.2280	0.2280	0.2280	0.2280	0.2280	0.2280
026	0.2300	0.2300	0.2300	0.2300	0.2300	0.2300	0.2300
008	0.2300	0.2300	0.2300	0.2300	0.2300	0.2300	0.2300
006	0.2350	0.2350	0.2350	0.2350	0.2350	0.2350	0.2350
025	0.2360	0.2360	0.2360	0.2360	0.2360	0.2360	0.2360
024	0.2370	0.2370	0.2370	0.2370	0.2370	0.2370	0.2370
016	0.2430	0.2430	0.2430	0.2430	0.2430	0.2430	0.2430
028	0.2440	0.2440	0.2440	0.2440	0.2440	0.2440	0.2440
009	0.2450	0.2450	0.2450	0.2450	0.2450	0.2450	0.2450
015	0.2555	0.2555	0.2555	0.2555	0.2555	0.2555	0.2555
023	0.2600	0.2600	0.2600	0.2600	0.2600	0.2600	0.2600
017	0.2640	0.2640	0.2640	0.2640	0.2640	0.2640	0.2640
018	0.2670	0.2670	0.2670	0.2670	0.2670	0.2670	0.2670
022	0.2700	0.2700	0.2700	0.2700	0.2700	0.2700	0.2700
012	0.2730	0.2730	0.2730	0.2730	0.2730	0.2730	0.2730
019	0.2740	0.2740	0.2740	0.2740	0.2740	0.2740	0.2740
029	0.2740	0.2740	0.2740	0.2740	0.2740	0.2740	0.2740
013	0.2780	0.2780	0.2780	0.2780	0.2780	0.2780	0.2780
014	0.2811	0.2811	0.2811	0.2811	0.2811	0.2811	0.2811
034	0.2870	0.2870	0.2870	0.2870	0.2870	0.2870	0.2870
033	0.2870	0.2870	0.2870	0.2870	0.2870	0.2870	0.2870
021	0.2880	0.2880	0.2880	0.2880	0.2880	0.2880	0.2880
004	0.2890	0.2890	0.2890	0.2890	0.2890	0.2890	0.2890
020	0.2950	0.2950	0.2950	0.2950	0.2950	0.2950	0.2950
002	0.2960	0.2960	0.2960	0.2960	0.2960	0.2960	0.2960
011	0.3110	0.3110	0.3110	0.3110	0.3110	0.3110	0.3110
032	0.3310	0.3198	0.3160	0.3159	0.3161	0.3162	0.3162
010	0.4246	0.3198	0.3160	0.3159	0.3161	0.3162	0.3162
平均值	0.2512	0.2579	0.2572	0.2571	0.2570	0.2570	0.2570
标准差 SD	0.0672	0.0342	0.0345	0.0347	0.0348	0.0348	0.0348
δ		0.0578	0.0581	0.0587	0.0590	0.0592	0.0592
新的 x^*	0.2620	0.2579	0.2572	0.2571	0.2570	0.2570	0.2570
新的 s^*	0.0386	0.0387	0.0391	0.0393	0.0394	0.0395	0.0395

x^* 的标准不确定度 u_x							0.0085
---------------------	--	--	--	--	--	--	--------

表 E4 饮用水中莠去津含量的中位值和 nIQR(MADe)的计算

实验室代码	$x_i(\text{mg/L})$	实验室代码	$x_i(\text{mg/L})$
001	0.0400	017	0.2640
005	0.0550	018	0.2670
007	0.1780	022	0.2700
030	0.2020	012	0.2730
003	0.2060	019	0.2740
027	0.2270	029	0.2740
031	0.2280	013	0.2780
026	0.2300	014	0.2811
008	0.2300	034	0.2870
006	0.2350	033	0.2870
025	0.2360	021	0.2880
024	0.2370	004	0.2890
016	0.2430	020	0.2950
028	0.2440	002	0.2960
009	0.2450	011	0.3110
015	0.2555	032	0.3310
023	0.2600	010	0.4246
中位值			0.2620
nIQR(MADe)			0.0402(0.0386)
中位值的标准不确定度 u_x			0.0086(0.0083)

例3 算法S应用

表E5是血浆能力验证物品中特定抗体浓度的测试结果。每个实验室在重复性条件下做四次重复测试。算法S可根据表中数据计算其标准差的稳健估计值，也称稳健联合值 w^* 。在经过2次迭代后，稳健联合值 w^* 的第三位有效数字不再发生变化。

表 E5 血浆能力验证物品中特定抗体的浓度（每个实验室对一份样本进行 4 次重复测定）

实验室	平均值 kU/L	标准差 kU/L	第一次迭代	第二次迭代	第三次迭代	第四次迭代
1	2.15	0.13	0.13	0.13	0.13	0.13

2	1.85	0.21	0.21	0.21	0.21	0.21
3	1.80	0.08	0.08	0.08	0.08	0.08
4	1.80	0.24	0.24	0.24	0.24	0.24
5	1.90	0.36	0.36	0.36	0.36	0.36
6	1.90	0.32	0.32	0.32	0.32	0.32
7	1.90	0.14	0.14	0.14	0.14	0.14
8	2.05	0.26	0.26	0.26	0.26	0.26
9	2.35	0.39	0.39	0.39	0.39	0.39
10	2.03	0.53	0.46	0.49	0.50	0.50
11	2.08	0.25	0.25	0.25	0.25	0.25
12	1.25	0.24	0.24	0.24	0.24	0.24
13	1.13	0.72	0.46	0.49	0.50	0.50
14	1.00	0.26	0.26	0.26	0.26	0.26
15	1.08	0.17	0.17	0.17	0.17	0.17
16	1.20	0.32	0.32	0.32	0.32	0.32
17	1.35	0.4	0.40	0.40	0.40	0.40
18	1.23	0.36	0.36	0.36	0.36	0.36
19	1.23	0.33	0.33	0.33	0.33	0.33
20	0.90	0.43	0.43	0.43	0.43	0.43
21	1.48	0.40	0.40	0.40	0.40	0.40
22	1.20	0.55	0.46	0.49	0.50	0.50
23	1.73	0.39	0.39	0.39	0.39	0.39
稳健平均值	1.57					
稳健联合值 w^*		0.32	0.3392	0.3446	0.3462	0.3467
界限值 ψ		0.4620	0.4898	0.4976	0.4999	0.5006
自由度 $\nu = n-1 = 3$; 限系数 $\eta = 1.444$; 修正系数 $\xi = 1.039$						

例 4 Q 或 Q_n /Hampel 算法应用

根据表 E2 中测试数据，使用 R 语言计算程序算得稳健均值和稳健标准差，见表 E6（R 代码见附录 F）。

表 E6Q 或 Q_n /Hampel 算法结果

方法	均值（mg/L）	标准差（mg/L）	$u(x_{pt})$ （mg/L）
稳健法：Q/Hampel 算法	0.2600	0.0426	0.0091
稳健法： Q_n /Hampel 算法	0.2592	0.0360	0.0077

注：Q 方法计算的稳健标准差是由文献[13]中程序计算得到的。

例 5 自助法计算指定值和指定值不确定度

35 个实验室参加了食品（牛奶）中大肠菌群的能力验证计划，该计划要求独立重复测试 5 次。指定值及其不确定度的评定是以每个实验室的 logCFU 的平均值作为依据（表 E7）。根据目标适用性，将“0.25 (log10 CFU/mL)”设为 σ_{pt} ，同时，核函数的标准差是 $0.75\sigma_{pt}$ （见 R 代码中的“bw”）。核密度图（图 E4）显示非对称性分布。采用自助法（1000 次重复抽样）估算该数据分布的核密度函数的众数和相应的标准误差，分别记作 x_{pt} 和 $u(x_{pt})$ 。导出结果如下：

$$x_{pt}=3.79(\log 10 \text{ CFU/mL}), u(x_{pt})=0.0888(\log 10 \text{ FU/mL})$$

注：因为 $u(x_{pt})>0.3\sigma_{pt}$ ，采用 z'评价实验室，即 $z'=\frac{x_i-x_{pt}}{\sqrt{\sigma_{pt}^2+u^2(x_{pt})}}$ 。

表 E7 食品（牛奶）中大肠菌群测试结果均值

实验室代码	$x_i(\log 10 \text{ FU/mL})$	实验室代码	$x_i(\log 10 \text{ FU/mL})$	实验室代码	$x_i(\log 10 \text{ FU/mL})$
001	3.80	013	2.83	025	3.25
002	3.90	014	2.75	026	3.45
003	3.07	015	2.06	027	3.64
004	3.64	016	3.75	028	2.86
005	4.06	017	3.73	029	3.17
006	3.40	018	3.82	030	3.19
007	3.59	019	3.86	031	3.17
008	3.39	020	3.88	032	4.22
009	3.47	021	3.97	033	3.82
010	3.47	022	3.96	034	3.82
011	3.77	023	3.80	035	3.95
012	3.53	024	3.88		

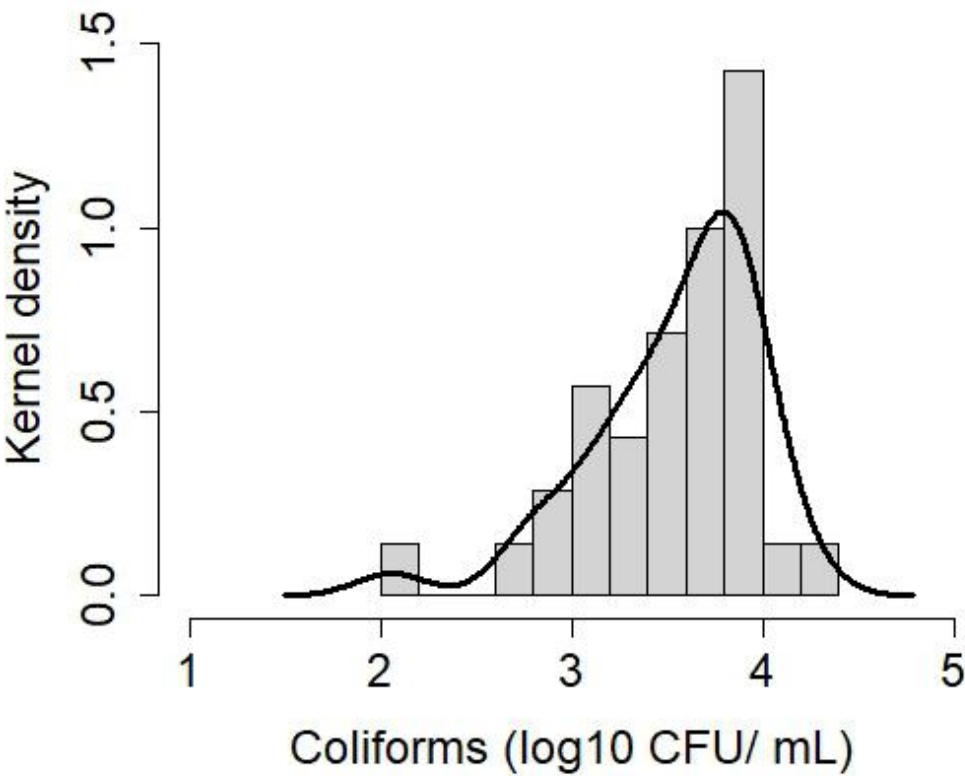


图 E4 参加者结果核密度图

例 6 单个实验室的参考值确定

表 E8 中给出的示例可能包括通过对一个能力验证物品（石料）和一个非常接近的 CRM 的一系列测试获得的数据（机械强度，LA），其中 CRM 的有证特征值是 $x_{\text{CRM}}=21.62$ LA，标准不确定度是 $u_{\text{CRM}}=0.26$ LA。该示例显示如何获得能力验证样品的参照值和不确定度。特别需要注意的是 CRM 的指定值的不确定度，包含了由不均匀性、运输和长期稳定性造成的不确定度。

能力验证物品的指定值： $x_{\text{pt}} = x_{\text{CRM}} + \bar{d}$ (E.1)

式中， $\bar{d}$ 为差值 d_i 的均值， d_i 为能力验证物品与 CRM 的测试均值之差。

能力验证物品的指定值的标准不确定度： $u_{\text{char}} = \sqrt{u_{\text{CRM}}^2 + u_{\bar{d}}^2}$ (E.2)

式中， $u_{\bar{d}}$ 为 $\bar{d}$ 的标准不确定度， d_i 的标准差为 S ，实验室家数为 n ， $u_{\bar{d}} = S/\sqrt{n}$ 。

表 E8 能力验证物品与 CRM 的测试的差值及差值的标准不确定度

样品编号	能力验证物品		CRM		能力验证物品与 CRM 的 测试均值之差 d_i LA
	测试 1 LA	测试 2 LA	测试 1 LA	测试 2 LA	
1	20.5	20.5	19.0	18.0	2.00
2	21.1	20.7	19.8	19.9	1.05
3	21.5	21.5	21.0	21.0	0.50
4	22.3	21.7	21.0	20.8	1.10
5	22.7	22.3	20.5	21.0	1.75
6	23.6	22.4	20.3	20.3	2.70
7	20.9	21.2	21.5	21.8	-0.60
8	21.4	21.5	21.9	21.7	-0.35
9	23.5	23.5	21.0	21.0	2.50
10	22.3	22.9	22.0	21.3	0.95
11	23.5	24.1	20.8	20.6	3.10
12	22.5	23.5	21.0	22.0	1.50
13	22.5	23.5	21.0	21.0	2.00
14	23.4	22.7	22.0	22.0	1.05
15	24.0	24.2	22.1	21.5	2.30
16	24.5	24.4	22.3	22.5	2.05
17	24.8	24.7	22.0	21.9	2.80
18	24.7	25.1	21.9	21.9	3.00
19	24.9	24.4	22.4	22.6	2.15
20	27.2	27.0	24.5	23.7	3.00
d_i 的平均值 $\bar{d}$					1.73
d_i 的标准差 S					1.07
$\bar{d}$ 的标准不确定度($S/\sqrt{20}$)					0.24
注：以上是在LA检测中，测量骨料机械强度获得的数据。					

采用上表数据，根据公式（E.1）和（E.2）计算能力验证物品的指定值和指定值标准不确定

度：

$$x_{pt}=21.62+1.73=23.35 \text{ (LA)}$$

$$u_{char}=\sqrt{0.26^2+0.24^2}=0.35 \text{ (LA)}$$

例 7 专家实验室公议值确定

选定的专家实验室需证明其对于被检测的特性量具有必要的能力。这可以提供实验室认可、资质认定、以前参加相关能力验证的结果、利用 CRM 的测量审核结果、所用检测方法的确认、不确定度评定方案与报告甚至在同行业测试中的信誉来证明。

在专家实验室确定后，按预先制定的规则，从制备的能力验证物品中随机抽取一定数量的样品分发给各专家实验室，并规定每个实验室应测试的样品数和重复测试的次数，在要求实验室提供结果的同时，还要求提供测量的不确定度及相关信息。

确定指定值的数据处理方式，类似于 CRM 中利用单个实验室定值的方式。指定值可以采用各个实验室结果的算术平均值或加权平均值指定值的不确定度也由相应的定值方式确定(见表 E9)。

表 E9 各种指定值及其不确定度计算方法

算术平均值	算术平均值 $\bar{x}$	$\bar{x}=\frac{1}{p}\sum_{i=1}^px_i$
	标准不确定度 $u(\bar{x})$	$S^2=\frac{1}{p-1}\sum_{i=1}^p(x_i-\bar{x})^2$ $u(\bar{x})=\frac{S}{\sqrt{p}}$
加权平均值	加权平均值 $\bar{x}$	$\bar{x}=\sum_{i=1}^pw_ix_i$ $w_i=\frac{\frac{1}{u^2(x_i)}}{\sum_{i=1}^p\frac{1}{u^2(x_i)}}$
	标准不确定度 $u(\bar{x})$	$u(\bar{x})=\sqrt{\sum_{i=1}^pw_i^2u^2(x_i)}$
中位值	中位值 $\text{Med}(x)$	$\text{Med}(x)=\begin{cases}x_{\frac{p+1}{2}} & p \text{ 为奇数} \\ \frac{1}{2}(x_{\frac{p}{2}}+x_{\frac{p}{2}+1}) & p \text{ 为偶数} \end{cases}$
	标准不确定度 $u(\text{Med}(x))$	$u(\text{Med}(x))=1.25\frac{\text{NIQR}}{\sqrt{p}}$

稳健平均值	稳检平均值 x^*	$x^* = \frac{\sum_{i=1}^p x_i^*}{p}$
	标准不确定度 $u(x^*)$	$u(x^*) = 1.25 \frac{S^*}{\sqrt{p}}$ $S^* = 1.134 \sqrt{\frac{\sum_{i=1}^p (x_i^* - x^*)^2}{p - 1}}$

如果数据的分布近似正态分布，而且各实验室结果的测量不确定也很接近，则在数据进行离群值检验和剔除后可以取算术平均值作为指定值。若采用加权平均值作指定值、则要求每个实验室所给的不确定度应当准确[14]。

若每个实验室不能提交测量不确定度，或提交的测量不确定度不充分可靠，则利用稳健统计法。根据公议值来确定指定值及其不确定度。

如表 E10 所示，从制备的能力验证物品中随机抽取 39 个样品分发给 13 家专家实验室，每个实验室测试 3 个样品，每个实验室提供三个样品的塑料维卡软化温度（VST）的测定结果，并计算算术平均值作为指定值。

表 E10 专家实验室塑料维卡软化温度（VST）的测定数据

实验室编号	样品 1 (°C)	样品 2 (°C)	样品 3 (°C)	均值 (°C)
1	97.8	98.1	97.6	97.8
2	96.8	96.5	95.9	96.4
3	94.8	95.2	94.9	95.0
4	96.0	97.8	97.6	97.1
5	95.6	95.9	96.5	96.0
6	94.7	95.2	95.5	95.1
7	95.6	95.4	95.9	95.6
8	94.5	95.8	94.9	95.1
9	95.5	95.8	96.1	95.8
10	95.3	95.8	95.7	95.6
11	96.0	96.5	96.8	96.4
12	95.8	95.9	96.5	96.1
13	97.5	97.1	96.8	97.1

算术平均值 $\bar{x}$			96.1
标准不确定度 $u(\bar{x})$			0.24

例 8 一般模型确认能力评定标准差（霍维茨方程）

霍维茨给出了化学分析方法再现性的一个一般模型，这个方法可得到以下再现性标准差的表达式： $\sigma_R = 0.02c^{0.8495}$

式中： c -待测化学成分的质量分数%。

如检测奶粉三聚氰胺含量的能力验证计划中使用两个能力验证样品 A 和 B，其参考水平为 A=1.195 mg/kg 和 B=2.565 mg/kg。计算再现性标准差能力验证样品 A(1.195 mg/kg)： $\sigma_R = 0.186$ mg/kg 或相对 $\sigma_R = 15.6\%$ ；能力验证样品 B(2.565 mg/kg)： $\sigma_R = 0.356$ mg/kg 或相对 $\sigma_R = 13.9\%$ 。

例 9 样品均匀性和稳定性检验

按照指定测试方法对含亚硝酸盐的奶粉样品进行均匀性测试。从样品总体中随机抽取 10 份样品进行检测，每份样品独立取样按顺序重复测试 2 次。首先，采用单因子方差分析法对样品的均匀性进行检验。从（ $\alpha = 0.05$ ）显著性的 F 值表上查得相应自由度的临界值为 $F_{\text{临界值}}$ ，若 $F_{\text{实验值}} \geq F_{\text{临界值}}$ ，说明各组间有系统差异，即样品之间存在差异，认为样品不均匀；反之，若 $F_{\text{实验值}} < F_{\text{临界值}}$ ，则认为组内与组间无明显差异，样品可视为均匀。

表 E11 均匀性测试数据

序号	测试结果/（mg/kg）	
	1	2
1	5.04	4.81
2	4.80	4.90
3	4.85	4.64
4	4.75	4.73
5	4.73	4.69
6	4.76	4.80
7	4.82	4.73
8	4.83	4.98

9	4.72	4.88
10	4.78	4.85
样品间标准差/（mg/kg）	0.026	
$F_{\text{实验值}}$	1.16	
$F_{0.05(9,10)}$	3.02	

为使用 $S_s \leq 0.3\sigma$ 准则进一步验证样品均匀性，计算样品中亚硝酸盐含量的不均匀性标准偏差 S_s 和 0.3σ （ σ 为能力评定标准差），计算结果见表 E12。样品中亚硝酸盐含量符合样品间标准差 $S_s \leq 0.3\sigma$ ，进一步表明样品是均匀的。

表E12 样品间标准差与 0.3σ 的比较

统计参数	测试结果/(mg/kg)
S_s	0.026
σ	0.39
0.3σ	0.117

为确保样品在整个能力验证过程中的稳定性，进行了样品稳定性检验。稳定性检验方法同均匀性检验。均匀性检验作为发样前稳定性检验。随机抽取 3 个最小包装单元，在 35℃ 下放置，15 天后取出，每个最小包装单元独立取样重复测试 2 次，作为发样后稳定性检验。

稳定性检验统计方法采用 t 检验法中“平均值一致性检验”方法，将稳定性检验结果与均匀性检验结果进行比较。若 t 小于显著水平 α （ $\alpha=0.05$ ）自由度为 n_1+n_2-2 的临界值 $t_{\alpha(n_1+n_2-2)}$ ，则发样前稳定性检验的平均值与发样后稳定性检验的平均值之间无显著性差异，判定样品稳定。表 E13 为稳定性评价数据汇总表。

表E13 稳定性测试数据

稳定性检验阶段	序号	测试结果/(mg/kg)	
		1	2
发样前	1	5.04	4.81
	2	4.80	4.90
	3	4.85	4.64

	4	4.75	4.73
	5	4.73	4.69
	6	4.76	4.80
	7	4.82	4.73
	8	4.83	4.98
	9	4.72	4.88
	10	4.78	4.85
发样后	1	4.64	4.76
	2	4.65	4.80
	3	4.85	4.68
检验结果 1,2		$t_{(1,2)} = 1.70$	
临界值		$t_{0.05(24)} = 2.06$	

比较发样前稳定性检验的均值 $\bar{x}$ 与发样后稳定性检验的均值 $\bar{y}$ 差的绝对值与对应的 0.3σ ，发现均值差的绝对值均小于对应的 0.3σ ，故样品在本轮次能力验证计划中稳定，见表 E14。

表E14发样前稳定性检验的均值与发样后稳定性检验的均值差与 0.3σ 的比较

统计参数	测试结果/(mg/kg)
$ \bar{x} - \bar{y} $	0.088
σ	0.39
0.3σ	0.117

附录 F
(资料性附录)
R 语言代码

图 D1 参加者结果直方图 (水中铅含量)

```
library(ggplot2)

xi<-c(0.206,0.321,0.070,0.185,0.236,0.227,0.198,0.260,0.230,0.235,0.202,0.242,0.228,0.231)

ggplot()+

  geom_histogram(aes(x=xi),stat="bin",position="dodge",fill="gray",binwidth=0.025,col="black")+

  theme_bw(base_family="serif") +

  theme(panel.grid =element_blank(),

  text = element_text(size=15)) +

  labs(x="铅浓度/(mg/L)",y="个数/家")
```

图 D2 参加者结果直方图 (生活饮用水中莠去津 (Atrazine) 含量)

```
library(ggplot2)

xi<-c(0.0400,0.0550,0.1780,0.2020,0.2060,0.2270,0.2280,0.2300,0.2300,0.2350,0.2360,0.2370,0.2430,0.2440,0.2450,0.2555,0.2600,0.2640,0.2670,0.2700,0.2730,0.2740,0.2740,0.2780,0.2811,0.2870,0.2870,0.2880,0.2890,0.2950,0.2960,0.3110,0.3310,0.4246)

ggplot()+

  geom_histogram(aes(x=xi),stat="bin",position="dodge",fill="gray",binwidth=0.045,col="black")+

  theme_bw(base_family="serif") +

  theme(panel.grid =element_blank(),

  text=element_text(size=15)) +

  labs(x="莠去津浓度/(mg/L)",y="个数/家")
```

图 D3 参加者结果核密度图（生活饮用水中莠去津（Atrazine）含量）

```
library(ggplot2)

xi<-c(0.0400,0.0550,0.1780,0.2020,0.2060,0.2270,0.2280,0.2300,0.2300,0.2350,0.2360,0.2370,0.2430,0.
2440,0.2450,0.2555,0.2600,0.2640,0.2670,0.2700,0.2730,0.2740,0.2740,0.2780,0.2811,0.2870,0.2870,0.
2880,0.2890,0.2950,0.2960,0.3110,0.3310,0.4246)

ggplot() +
  geom_density(aes(x=xi),size=1.0)+
  theme_bw(base_family="serif") +
  theme(panel.grid =element_blank()) +
  theme(text=element_text(size=15),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank()) +
  xlim(-0.1,0.5)+
  ylim(0,12)+
  labs(x="莠去津浓度/(mg/L)",y="核密度")
```

图 D4 参加者结果核密度图（食品（牛奶）中大肠菌群）

```
library(boot)

library(pastecs)

xi=c(3.80,3.90,3.07,3.64,4.06,3.40,3.59,3.39,3.47,3.47,3.77,3.53,2.83,2.75,2.06,3.75,3.73,3.82,3.86,3.88,
3.97,3.96,3.80,3.88,3.25,3.45,3.64,2.86,3.17,3.19,3.17,4.22,3.82,3.82,3.95)

options(digits=3)

stat.desc(xi)      #描述统计量

sigmat<-0.25      #适用性标准差

bw=0.75*sigmat     #核函数标准差

hist(xi,freq=F,main="",cex.axis=1.2,cex.lab=1.2,xlim=c(1,5),ylim=c(0,1.5),
      xlab="Coliforms (log10 CFU/ mL)",ylab="Kernel density",breaks=10)  #绘制直方图
```

```

lines( density(xi,kernel="gaussian",bw),col="black",lwd=3)           #绘制核密度图
theta<- function(y,i)                                              #统计量函数
{
dens<-density(y[i], kernel="gaussian",bw= bw)
mode<-dens$x[which.max(dens$y)]
}
set.seed(220)
boot.statistics<-boot(xi,theta,R=1000)
boot.statistics

```

图 E1 两个相似过敏源的抗体浓度的 z 值尧敦图

```

library(ggplot2)
ZAi<-c(0.427,-1.540,-0.043,-0.978,2.228,1.092,-0.432,1.942,0.042,0.272,-1.400,-0.599,0.057,0.227,-0.1
80,-0.180,-0.113,-0.104,-1.185,0.190,-0.253,-1.124,2.710,0.321,-0.052,-0.055,-0.241,0.203,-1.230)
ZBi<-c(0.515,-0.428,-0.366,-0.942,2.023,0.194,-0.138,0.770,-1.204,-0.093,0.042,-0.642,-0.652,-0.238,-0.
493,-0.607,0.028,0.335,-1.353,-0.114,-0.676,-0.549,2.762,1.415,-0.949,2.019,-0.752,0.729,-0.635)
T95=6.92821
T99=11.3098
r=0.706
ZA1<- seq(-sqrt(T95),sqrt(T95),0.00001)
ZB1=r*ZA1+sqrt((1-r^2)*(T95-ZA1^2))
ZB2=r*ZA1-sqrt((1-r^2)*(T95-ZA1^2))
ZA2<- seq(-sqrt(T99),sqrt(T99),0.00001)
ZB3=r*ZA2+sqrt((1-r^2)*(T99-ZA2^2))
ZB4=r*ZA2-sqrt((1-r^2)*(T99-ZA2^2))
ggplot()+
  geom_point(aes(x=ZAi,y=ZBi),cex=1.5)+
  geom_path(aes(ZA1, ZB1),lty =1,lwd=1)+
  geom_path(aes(ZA1, ZB2),lty =1,lwd=1)+

```

```

geom_path(aes(ZA2, ZB3),lty =2,lwd=1)+
geom_path(aes(ZA2, ZB4),lty =2,lwd=1)+
geom_text(aes(x=c(-2.8,-2.0),y=c(-2.6,-2.0),label=c("1%水平","5%水平")),
           position=position_dodge(width=0.5),cex=3)+
theme_bw(base_family="serif")+
theme(text=element_text(size=15),panel.grid =element_blank()+
geom_vline(aes(xintercept=0),colour="black")+
geom_hline(aes(yintercept=0),colour="black")+
xlim(-5,5)+
ylim(-5,5)+
labs(x=expression(Z["A"]),y=expression(Z["B"]))

```

图 E2 重复性标准差图

```

library(ggplot2)
Xi<-c(2.15,1.85,1.80,1.80,1.90,1.90,1.90,2.05,2.35,2.03,2.08,1.25,1.13,1.00,1.08,1.20,1.35,1.23,1.23,0.9
0,1.48,1.20,1.73,1.43,1.28)
Wi<-c(0.13,0.21,0.08,0.24,0.36,0.32,0.14,0.26,0.39,0.53,0.25,0.24,0.72,0.26,0.17,0.32,0.40,0.36,0.33,0.4
3,0.40,0.55,0.39,0.30,0.22)
 $\chi_{95}=5.99$ 
 $\chi_{99}=9.21$ 
 $\chi_{999}=13.8$ 
m=4
X=1.57
W=0.34
s1=W*sqrt( $\chi_{95}/m$ )
x1<- seq(X-s1,X+s1,0.00001)
S1=W*exp(sqrt( $\chi_{95}-m*((x1-X)/W)^2$ )/sqrt(2*(m-1)))
S2=W*exp(-sqrt( $\chi_{95}-m*((x1-X)/W)^2$ )/sqrt(2*(m-1)))
s2=W*sqrt( $\chi_{99}/m$ )

```

```

x2<- seq(X-s2,X+s2,0.00001)

S3=W*exp(sqrt( $\chi^2_{99-m*((x2-X)/W)^2}$ )/sqrt(2*(m-1)))
S4=W*exp(-sqrt( $\chi^2_{99-m*((x2-X)/W)^2}$ )/sqrt(2*(m-1)))
s3=W*sqrt( $\chi^2_{999}/m$ )

x3<- seq(X-s3,X+s3,0.00001)

S5=W*exp(sqrt( $\chi^2_{999-m*((x3-X)/W)^2}$ )/sqrt(2*(m-1)))
S6=W*exp(-sqrt( $\chi^2_{999-m*((x3-X)/W)^2}$ )/sqrt(2*(m-1)))

ggplot()+

  geom_point(aes(x=Xi,y=Wi),cex=1.5)+

  geom_path(aes(x1, S1),lty =1)+

  geom_path(aes(x1, S2),lty =1)+

  geom_path(aes(x2, S3),lty =2)+

  geom_path(aes(x2, S4),lty =2)+

  geom_path(aes(x3, S5),lty =3)+

  geom_path(aes(x3, S6),lty =3)+

  geom_text(aes(x = 1.57, y = c(0.8,1.1,1.4),label = c("5%水平","1%水平","0.1%水平")),

    position=position_dodge(width=0.5))+

  theme_bw(base_family="serif")+

  theme(text=element_text(size=15),panel.grid =element_blank())+

  xlim(0.7,2.5)+

  ylim(0,1.6)+

  labs(x="平均浓度/(kU/L)",y="标准差/(kU/L)")

```

图 E3 曼德尔统计量 h 图

```

library(ggplot2)

h<-c(1.39,0.67,0.55,0.55,0.79,0.79,0.79,1.15,1.87,1.10,1.22,-0.77,-1.06,-1.37,-1.18,-0.89,-0.53,-0.82,-0.8
2,-1.61,-0.22,-0.89,0.38,-0.34,-0.70)

i<-c(1:25)

ggplot()+

  geom_bar(aes(x=i,y=h),stat="identity",position="dodge",fill="gray",width=0.45,col= "black")+

  theme_bw(base_family="serif") +

  theme(panel.grid =element_blank()) +

  geom_hline(aes(yintercept=2.42), linetype="dashed")+

  geom_hline(aes(yintercept=-2.42), linetype="dashed")+

  geom_hline(aes(yintercept=1.9),linetype="dashed")+

  geom_hline(aes(yintercept=-1.9),linetype="dashed")+

  geom_hline(aes(yintercept=0),linetype="solid")+

  geom_text(aes(x=c(1:25),y=ifelse(h>=0,-0.15,0.15),label=c(1:25)),

            position=position_dodge(width=0.5),size=3.5)+

  geom_text(aes(x=-1.2,y=c(2.62,-2.62,1.7,-1.7),label=c("1% 水平","1% 水平","5% 水平","5% 水平
")),position=position_dodge(width=0.5),size=4)+

  theme(legend.position = "none")+

  theme(text=element_text(size=15),

        axis.text.x =element_blank(),

        axis.ticks.x = element_blank()) +

  ylim(-3.5,3.5)+

  xlim(-2,26)+

  labs(x=expression(实验室~italic(i)),y=expression(曼德尔统计量~italic(h)))

```

图 E4 曼德尔统计量 k 图

```

library(ggplot2)

k<-c(0.37,0.60,0.23,0.69,1.03,0.92,0.40,0.74,1.12,1.52,0.72,0.69,2.06,0.74,0.49,0.92,1.15,1.03,0.94,1.23,
1.15,1.57,1.12,0.86,0.63)

i<-c(1:25)

ggplot()+

  geom_bar(aes(x=i,y=k),stat="identity",position="dodge",fill="gray",width=0.45,col="black")+

  theme_bw(base_family="serif") +

  theme(panel.grid =element_blank()) +

  geom_hline(aes(yintercept=1.9), linetype="dashed")+

  geom_hline(aes(yintercept=1.6), linetype="dashed")+

  geom_hline(aes(yintercept=0),linetype="solid")+

  geom_text(aes(x=c(1:25),y=-0.15,label=c(1:25)),size=3.5,position=position_dodge(width=0.5))+

  geom_text(aes(x=-1.2,y=c(1.5,2.0),label=c("5% 水平 ","1% 水平 ")),size=4, position =
position_dodge(width=0.5))+

  theme(legend.position = "none")+

  theme(text=element_text(size=15),

        axis.text.x =element_blank(),

        axis.ticks.x = element_blank()) +

  ylim(-1.5,3)+

  xlim(-2,26)+

  labs(x=expression(实验室~italic(i)),y=expression(曼德尔统计量~italic(k)))

```

Qn 法计算代码（饮用水中莠去津）

```

M<-c(0.0400,0.0550,0.1780,0.2020,0.2060,0.2270,0.2280,0.2300,0.2300,0.2350,0.2360,0.2370,0.2430,0.
2440,0.2450,0.2555,0.2600,0.2640,0.2670,0.2700,0.2730,0.2740,0.2740,0.2780,0.2811,0.2870,0.2870,0.
2880,0.2890,0.2950,0.2960,0.3110,0.3310,0.4246)

p<-length(M)
N<-p*(p-1)/2
h<-floor(p/2)
k<-h*(h-1)/2
r<-ifelse(p%%2==0,(3.6756+(1.965+(6.987-77/p)/p)/p)/p,(1.6019-(2.128+5.172/p)/p)/p)
d<-numeric(N)
num=1
for(i in 1:(p-1))
{
  for(j in (i+1):p)
  {
    d[num] = abs(M[j]-M[i])
    num = num + 1
  }
}
ds<-sort(d)
dk<-ds[k]
Qn<-2.2219*dk/(1+r)
print(Qn)

```

Hampel 方法的代码（重复加权迭代算法）（饮用水中莠去津）

```
M<-c(0.0400,0.0550,0.1780,0.2020,0.2060,0.2270,0.2280,0.2300,0.2300,0.2350,0.2360,0.2370,0.2430,0.
2440,0.2450,0.2555,0.2600,0.2640,0.2670,0.2700,0.2730,0.2740,0.2740,0.2780,0.2811,0.2870,0.2870,0.
2880,0.2890,0.2950,0.2960,0.3110,0.3310,0.4246)

p<-length(M)

X<-median(M)

S<-0.0360

for(i in 1:100){
  q=abs((X[i]-M)/S)
  W=ifelse(abs(q)<=1.5,1,ifelse(abs(q)<=3,1.5/q,ifelse(abs(q)<=4.5,(4.5-q)/q,0)))
  X[i+1]=sum(M*W)/sum(W)
  if(abs(X[i]-X[i+1])<0.01*S*p^0.5){
    print(X[i+1])
    print(i+1)
  }
  break
}
}
```

Hampel 方法的代码（重复加权有限步算法）（饮用水中莠去津）

```
M<-c(0.0400,0.0550,0.1780,0.2020,0.2060,0.2270,0.2280,0.2300,0.2300,0.2350,0.2360,0.2370,0.2430,0.
2440,0.2450,0.2555,0.2600,0.2640,0.2670,0.2700,0.2730,0.2740,0.2740,0.2780,0.2811,0.2870,0.2870,0.
2880,0.2890,0.2950,0.2960,0.3110,0.3310,0.4246)

p<-length(M)

S<-0.0426

d=sort(c(M-4.5*S,M-3*S,M-1.5*S,M+1.5*S,M+3*S,M+4.5*S))

q=numeric()

P=numeric()

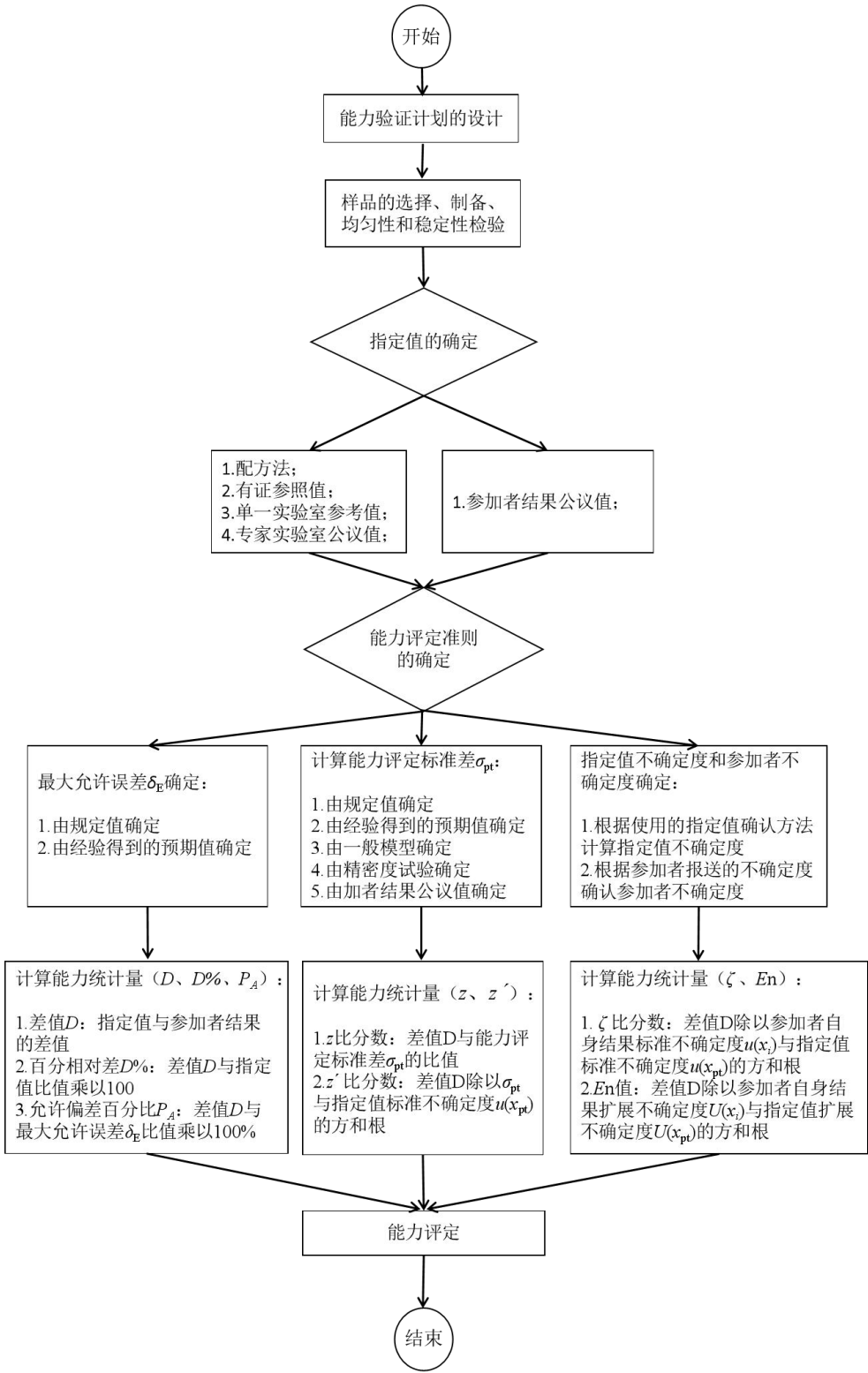
num=1
```

```

for(m in 1:(6*p)){
  for(i in 1:p){
    q[num]=(M[i]-d[m])/S
    num = num + 1
  }
}
Q=matrix(q,p,(6*p))
for(j in 1:(6*p)){
  Ψ<- ifelse((Q[,j]<=-4.5),0,
    ifelse((Q[,j]>-4.5&Q[,j]<=-3),-4.5-Q[,j],
    ifelse((Q[,j]>-3&Q[,j]<=-1.5),-1.5,
    ifelse((Q[,j]>-1.5&Q[,j]<=1.5),Q[,j],
    ifelse((Q[,j]>1.5&Q[,j]<=3),1.5,
    ifelse((Q[,j]>3&Q[,j]<=4.5),4.5-Q[,j],0))))))
  P[j]=sum(Ψ)
}
V=d[which(P[1:(6*p-1)]*P[2:(6*p)]<0)]-P[which(P[1:(6*p-1)]*P[2:(6*p)]<0)]*(d[which(P[1:(6*p-1)]*
P[2:(6*p)]<0)+1]-d[which(P[1:(6*p-1)]*P[2:(6*p)]<0)]/(P[which(P[1:(6*p-1)]*P[2:(6*p)]<0)+1]-P[whi
ch(P[1:(6*p-1)]*P[2:(6*p)]<0)]))
s=c(d[which(P==0)],V)
X=s[which.min(abs(s-median(M)))]
print(X)

```

附录 G
(资料性附录)
定量类能力验证流程图



参考文献

- [1] ISO 5725-2, 测量方法与结果的准确度(正确度与精密度) 第2部分: 确定标准测量方法重复性与再现性的基本方法.
- [2] ISO 5725-3, 测量方法与结果的准确度(正确度与精密度) 第3部分: 标准测量方法的精密度中间度量.
- [3] ISO 5725-4, 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.
- [4] ISO 5725-5, 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.
- [5] ISO 5725-6, 测量方法与结果的准确度(正确度与精密度) 第6部分: 准确度值得实际应用.
- [6] GB/T 28043-2019, 利用实验室间比对进行能力验证的统计方法.
- [7] GB/T 27043-2012, 合格评定 能力验证的通用要求.
- [8] ISO/IEC 17025, 检测和校准实验室认可准则.
- [9] GB/T 15000.3-2008, 标准样品工作导则(3) 标准样品 定值的一般原则和统计方法.
- [10] JJG 1006-1994, 国家计量技术规范一级标准物质.
- [11] JJF1059.1-2012, 测量不确定度评定与表示.
- [12] GB/T 27418-2017, 测量不确定度评定和表示.
- [13] [Http://quodata.de/en/web-services/QHampel.html](http://quodata.de/en/web-services/QHampel.html).
- [14] 马捷, 关淑敏, 茅祖兴. 能力验证及其结果处理与评价[M]. 北京: 中国质检出版社, 2016.
- [15] CNAS-GL002, 能力验证结果的统计处理和能力评价指南.
- [16] CNAS-GL003, 能力验证样品均匀性和稳定性评价指南.