



CNAS—GL02

**能力验证结果的统计处理和
能力评价指南**

**Guidance on Statistic Treatment of
Proficiency Testing Results and
Performance Evaluation**

中国合格评定国家认可委员会

二〇〇六年六月

能力验证结果的统计处理和评价指南

1. 目的与范围

1.1 本指南对 CNAS 能力验证结果的统计处理和根据统计处理结果对实验室进行能力评价的方法提供了指南。

1.2 本指南适用于 CNAS 的能力验证活动，以及将结果应用于 CNAS 认可的由其他方组织的能力验证活动。

2. 引用文件

下列文件中的条款通过引用而成为本文件的条款。以下引用的文件，注明日期的，仅引用的版本适用；未注明日期的，引用文件的最新版本（包括任何修订）适用。

ISO/IEC 指南 43-1: 1997 《利用实验室间比对的能力验证——能力验证计划的建立和运作》（GB/T15483.1, IDT）

APLAC PT001 《检测实验室间的比对》

APLAC PT002 《校准实验室间的比对》

《NATA 能力验证指南》

ISO 5725: 1994 1-6 《测量方法和结果的准确度（正确度和精密度）》

3. 统计处理

3.1 统计设计

3.1.1 能力验证的结果可以多种形式出现，并构成各种统计分布。分析数据的统计方法应与数据类型及其统计分布特性相适应。无论使用哪一种方法对参加者的结果进行评价，一般应包括以下几方面内容：

- a) 确定指定值；
- b) 计算能力统计量；
- c) 评价能力；
- d) 在某些情况下需预先确定被测样品的均匀性和稳定性。

3.1.2 在统计设计中应考虑下列事项及其相互影响：

- a) 所涉及测试的精密性和正确性;
- b) 在要求的置信水平下检出参加者之间的最小差异;
- c) 参加者的数量;
- d) 待检样品的数目和对每一被测样品进行重复检测/测量的次数。在校准能力验证计划中, 应考虑比对的周期;
- e) 估算指定值所使用的程序, 及识别离群值所使用的程序;
- f) 校准能力验证计划中, 参考实验室必须能够给出优于参加者的测量不确定度(应尽量选择拥有国家基标准的实验室)。

3.1.3 在缺乏 4.1.2 a)的可靠信息时, 可能有必要组织一次先导性实验室间比对(协同试验), 以获得该信息。

3.2 指定值及其不确定度的确定

3.2.1 确定指定值的方法有多种, 下面是最常用的几种。按不确定度增加的顺序(多数情况下如此)排列如下:

- a) 已知值——其结果由特定样品配制(如制备、稀释)时确定。
- b) 有证参考值——由定义法确定(用于定量检测)。
- c) 参考值——与一个可追溯到国家或国际标准的参考标准物质/标准样品或标准进行分析、测量或比对检测物品所确定的值。
- d) 由各专家实验室获得的公议值。专家实验室在对被测量的测定方面应具有可证实的能力, 其使用的方法已经过确认, 并且有较高的精密度和准确度, 与通常使用的方法具有可比性。在某些情况下, 这些实验室可以是参考实验室。
- e) 从参加实验室获得的公议值——利用 4.2.3 中的统计量, 并考虑到极端结果的影响。

3.2.2 为公正地评价参加实验室, 以及促进实验室之间和方法之间的协调一致, 应有确定的指定值。这一点通过参加共同的比对, 并使用共同的指定值就可以实现。

3.2.3 下述统计量适合于使用公议方法来确定指定值:

定性值——预先确定的多数百分率的公议值;

定量值——适当比对组的平均值, 如:

- a) 可以是加权或变换(如修剪平均或几何均值)的平均值;
- b) 中位值、众数或其它稳健度量。

3.2.4 应根据所开展项目的特定技术要求, 运用“测量不确定度表示指南”(由 BIPM、IEC、IFCC、ISO、IUPAC 和 OIML 等联合制定)中规定的程序确定指定值的不确定度。

3.2.5 极端结果

- a) 在使用参加实验室的数据确定指定值时，所用的统计方法应当使极端结果的影响降至最小，这可以通过使用稳健统计方法或在计算之前剔除离群值来实现（详见 ISO5725-2）。
- b) 如果参加者的结果作为离群值被剔除，那么该剔除应仅为了计算总计统计量，而在能力验证报告中仍需对这些结果进行评估，并且给出适当的能力评价。

3.2.6 其它需考虑的事项

- a) 按理想情况，如果用参考值或参加者的公认值来确定指定值，协调人应有一个程序来确定指定值的正确度以及检查数据的分布；
- b) 协调人必须有根据其不确定度判断指定值是否可接受的准则。

3.3 能力统计量的计算

3.3.1 单个检测项目的能力

3.3.1.1 能力验证结果常需转换成一个能力统计量以便于说明和衡量与指定值的偏差。

3.3.1.2 检测能力的评价对于能力验证的参加者应有意义。因此，对检测项目的能力评价应该和检测的要求相关，并能被理解或符合特定领域里的惯例。

3.3.1.3 变动性度量常用于计算能力统计量和能力验证计划的总结报告中。对一组比对的数据常用例子是：

- 标准偏差（ SD ）；
- 变异系数（ CV ）或相对标准偏差（ RSD ）；
- 百分数与中位值的绝对偏差或其它稳健度量。

3.3.1.4 定性结果通常不需要经过计算。定量结果常用的统计量如下：

- a) 偏差 $D = x - X$ ，这里 x 是参加者的结果值， X 为指定值
- b) 偏差百分比率 $\frac{D}{X} \times 100\%$
- c) 百分数或秩
- d) Z 比分数

$$Z = \frac{x - X}{s}$$

这里 S 是指变动性的适当的估计量/度量值。这种模式既适用于 X 和 S 由参加者结果推导出的情形，亦适用于 X 和 S 不是由全部参加者结果推导出的情形（例如，对指定值和变动性可作出明确规定时）。

利用四分位数稳健统计方法处理结果时, $Z = \frac{(x - X)}{0.7413IQR}$ 。式中 IQR 为四分位间距。

e) E_n 值 (该统计量通常用于测量比对计划和测量审核活动)。

$$E_n = \frac{x_{LAB} - x_{REF}}{\sqrt{U_{LAB}^2 + U_{REF}^2}}$$

式中 x_{LAB} 是实验室的测量结果, x_{REF} 是被测物品的参考值。 U_{LAB} 为参加者结果的不确定度; U_{REF} 是指定值的不确定度。

3.3.1.5 注意事项

- a) 参加者结果和指定值之间的简单差值可能足以确定能力, 且易被参加者所理解。数值 $(x - X)$ 在 ISO5725-4 中称为“实验室偏移的估计值”。
- b) 百分率差适用于浓度的变化, 参加者较易理解。
- c) 百分数或秩用于高度离散或偏态分布的结果和次序响应, 或不同的响应值有限时的情形。不要轻易使用该方法。
- d) 根据检测数据的性质须对结果实行变换。有时这种变换是必要的, 比如, 稀释的结果以几何尺度变化, 因而可以进行对数变换。
- e) 如果使用统计量作为评价标准 (如 Z 比分数), 变动性的估计必须可靠, 即基于足够的观察以减少极端值的影响和降低不确定度。

3.4 综合能力评估

3.4.1 在单独一次能力验证轮回中, 可以根据一个以上的结果对实验室能力进行评估。这种情况出现在一个特定测试物或一组相关的测试物有一个以上测试项目时, 将能提供更为全面的测试能力的评估方法。

某些图表, 例如尤登 (Youden) 图或曼德尔 (Mandel) h - 值图, 都是表示测试能力的有效手段。

综合评估的例子如下:

- a) 相同被测量的综合值
 - 满意结果的数目
 - z 比分数的平均值
 - 绝对偏差的平均值 (以单位或百分比表示)
 - 绝对偏差 (或平方偏差) 之和
- b) 不同被测量的综合值

- 满意结果的数目或百分比
- 绝对 Z 比分数的平均值
- 与评价极限相关的绝对偏差的平均值。

3.4.2 注意事项

- a) 数值可以根据需要进行变换,使它们都服从相同的假设分布(如 Z 比分数服从正态分布,偏差的平方服从 χ^2 分布)。
- b) 对严重影响综合能力评价的极端值应进行检查。

3.5 能力评价

3.5.1 在建立能力的评价标准前,应考虑能力的度量值是否具有下列特点:

- a) 专家公议:在这种情况下,顾问组或其它资深专家直接确定报告的数据是否符合要求,专家公议是评价定性检测结果的主要途径。
- b) 与目标的符合性:例如,应考虑方法的使用范围和参加者被认可的操作水平等。
- c) 数值的统计判定:这里的评价准则适用于各种结果值。一般将 Z 比分数分为:

$ Z \leq 2$	满意结果
$2 < Z < 3$	有问题
$ Z \geq 3$	不满意或离群的结果

将 E_n 值分为

$ E_n \leq 1$	满意结果
$ E_n > 1$	不满意结果

- d) 参加者的公议:由一定百分比的参加者或由某个参考标准组提供的比分数值或结果的范围。如:
 - 中心百分比(80%, 90%或 95%)满意,或
 - 单侧百分比(最低 90%)满意。

3.5.2 分割样品方案的设计,目的是识别不合适的校准或结果中严重的随机误差。对此,应依据足够数量的数据和较宽的浓度范围进行评估。

为识别和描述这些问题,可采用作图法,特别是采用平均值所作的图表明实验室间差异。结果用适当的参数或非参数技术与回归分析和残差分析进行比较。

3.5.3 只要可能,应使用图示法表示能力(如直方图、误差柱状图和 Z 比分数次序图)这些图示法可用来表示:

- 参加者结果的分布;
- 多个检测项目数据间的关系;
- 不同方法的分布比较。

有时,某些实验室出具的数据,在能力验证计划中为离群结果,但可能仍在其相关标准规定的允差范围之内,鉴于此,利用参加能力验证计划的结果来对实验室的能力进行判定时,通常不作出“合格”与否的结论,而是使用“满意/不满意”或“离群”的概念。

3.5.4 当利用测量审核对实验室的能力进行判定时,可利用 En 值或参照相关技术标准(包括统计技术方面的标准)进行判定,附件 C 给出了相应的统计方法信息。

附件 A:

检测实验室间能力验证计划的结果处理方法

A.1 引言

本文件概括了 CNAS 用于处理能力验证计划结果的程序。但仅适用于检测能力验证计划，而不适用于校准能力验证（测量比对）计划，校准能力验证计划的统计将在附录 B 中阐述。在检测验证计划中，结果的评价是建立在与给定值的比较之上，给定值通常是从所有参加者的结果中获得，即公议值。

本文件中所描述的统计程序能够适用于大部分的检测能力验证计划。只要可行，能力验证计划的结果分析应尽量采用这些程序，但在某些情况下，也可能需要采用其它更适合的统计技术。

对于所有的能力验证计划，统计分析只是评价其结果的一个方面。如果一个结果被认为是离群值，这意味着，从统计上看它明显地不同于本组的其他结果。然而，从所涉及的具体学科（如化学）看，结果可能没有“错”。这就是为什么规定结果的评价应由统计分析和技术专家共同参加的原因。

本文件的 A.4、A.5 和 A.6 列出了实际使用的统计分析（包括某些实例）——即能力验证计划最终报告中出现的统计、表格和图表。在 A.2 中将说明一些在能力验证计划筹划过程中需考虑的理论背景；在 A.3 中将描述开始统计分析前需进行的结果的收集、输入和检查。

A.2 统计设计

提供给实验室的检测物品之间的所有差别已降至最小，因此结果的变异性主要有两个来源：实验室间的变异（包括测量方法间的变动）和实验室内部的变异。我们将通过在这两种类型的变异上来评价实验室的结果和提供反馈。

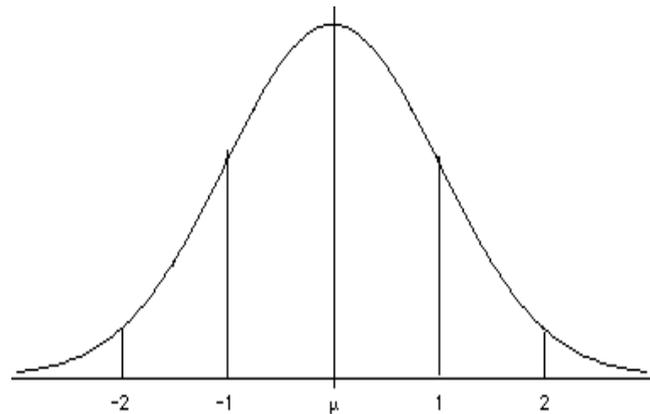
为了评定实验室间和实验室内这两种变异性，实验室必须进行多于一次（如两次）的相同检测。因此，只要可能，验证计划应设计成能够获得成对的相关结果。这可以用样品对来实现，如不可能，也可以通过对一个样品检测两次的结果来获得。

如果使用成对的样品，它们可以是等同的（即“均一对”）或者存在轻微的差别（分割水平对）。由样品对获得的结果分成两类：均一对，其结果预期是相同的（即两个样品完

全相同或同一样品检测两次)；分割水平对，其结果稍有差异。

均一对和分割水平对这两种类型结果的统计分析是相同的，但在解释上稍有不同（见 A.5）。有些验证计划不可能获得结果对，即只能获得单一样品的单个结果，在这种情况下，统计分析较为简单，但不能区分出两种不同类型的变异。

在能力验证计划设计过程中另一个重要的统计考虑是，假设分析的结果服从正态分布。这是统计分布中最常见的类型（见下图）。



正态分布

正态分布是一个连续的、对称的“钟形”曲线，并被定义为大约有 68% 的值位于平均值的一倍标准偏差内，95% 的值位于两倍标准偏差内，99% 的值位于三倍标准偏差内。因此，计划组织者（特别是技术专家）必须注意，所有获得的结果应近似于正态分布。

另外，对于具有连续数字的结果，技术专家必须仔细考虑给出所需的单位和小数位（或有效数字），否则数据可能出现大量的重复值。另一个应该避免的问题是当特性量是在非常低的水平上测试时，结果往往是不对称的（即偏向零）。

A.3 数据准备

在开始进行统计分析之前，应采取措施确保所采集的数据是正确、合理的。

必须确保正确地输入所有提交的结果。一旦收到了所有结果（或已超过上报结果的最后期限），必须仔细复查输入的数据。通过这个检查过程，一般可以识别出数据中的粗大误差和潜在问题。

在某些情况下，结果需经过转换，例如：微生物计数的统计分析通常按结果的对数计算，而不是按原始的数据计算。当所有结果已被输入并经过检查（必要时经过转换），然后制作显示结果分布的数据直方图，以检验正态性假设。

检查直方图可以看出结果是否连续和对称。如果不是，统计分析可能无效。还可能出现一个问题，即在直方图上出现两组有差异的结果（即双峰分布），这通常是由于使用了产生不同结果的两种检测方法。在这种情况下，应对两种方法的数据进行分离，然后对每一种方法的数据分别进行统计分析。

A.4 总计统计量

完成了数据准备，就可以用总计统计量来描述结果。至少应包含七种综合的统计量——结果数、中位值、标准四分位数间距（ IQR ）、稳健的变异系数（ CV ）、最小值、最大值和极差。

其中最重要的统计量是中位值和标准化 IQR ——它们是数据集中和分散的量度，与平均值和标准偏差相似。使用中位值和标准化 IQR 是因为它们是稳健的统计量，即它们不受数据中离群值的影响。

结果数是从一个特定检测中得到的结果总数，符号为 N 。

中位值是一组数据的中间值，即有一半的结果高于它，一半的结果低于它。如果 N 是奇数，那么中位值是一个单一的中心值，即 $X \left[\frac{(N+1)}{2} \right]$ 。如果 N 是偶数，那么中位值是两个中心值的平均，即是 $\frac{(X \left[\frac{N}{2} \right] + X \left[\left(\frac{N}{2} \right) + 1 \right])}{2}$ 。例如，如果 N 是 9，中位值是第 5 个值，如果 N 是 10，那么中位值是第 5 和第 6 个值的平均值。

标准化 IQR 是一个结果变异性的量度。它等于四分位间距（ IQR ）乘以因子 0.7413，其与一个标准偏差相类似。四分位间距是低四分位数值和高四分位数值的差值。低四分位数值（ Q_1 ）是低于结果的四分之一处的最近值，高四分位（ Q_3 ）是高于结果四分之三处的最近值。在大多数情况下 Q_1 和 Q_3 通过数据值之间的内插法获得。 $IQR = Q_3 - Q_1$ ，标准化 $IQR = IQR \times 0.7413$ 。

稳健 CV 是变异系数，稳健 $CV = \frac{\text{标准化}IQR}{\text{中位值}} \times 100\%$ 。

最小值是最低值（即 $X[1]$ ），最大值是最高值（即 $X[N]$ ），极差是它们之间的差值（即 $X[N] - X[1]$ ）。

计算了能力验证计划中的总计统计量后，为了及时地将信息反馈给实验室，可把中位值，标准化 IQR 以及实验室的结果列成表格，作为中期报告发至参加实验室。中期报告发

布之后，组织者不应对数据再做改动和添加（如迟到的结果）。

A.8 中给出了一个最终报告中的总计统计量的例子。例中使用了一个样品对，样品 A 和样品 B 是一对分割水平样品。

注：因子 0.7413 是从“标准”正态分布中导出。

A.5 稳健 Z 比分数和离群值

为了统计评价参加实验室的结果，可使用基于稳健总计统计量的 Z 比分数（中位值和标准化 IQR）。如果是样品对的结果（在大多数情况下），将计算两个 Z 比分数，即实验室间 Z 比分数（ZB）和实验室内 Z 比分数（ZW）。它们分别基于结果对的和与差值。

假设结果对是从 A 和 B 两个样品中获得的。把样品 A 所有结果的中位值和标准化 IQR 分别写为中位值（A）和标准化 IQR(A)，（样品 B 也类似）。仅对一个样品 A 的结果而言，简单的稳健 Z 比分数（用 Z 表示）为：

$$Z = \frac{A - \text{中位值}(A)}{\text{标准IQR}(A)}$$

当根据样品对的结果 A 和 B 计算 ZB 和 ZW 时，首先计算结果对的标准化和（用 S 表示）和标准化差值（Z），即：

$$S = (A + B) / \sqrt{2} \quad \text{和} \quad D = (A - B) / \sqrt{2} \quad (\text{保留 } D \text{ 的+或-号})$$

通过计算每个实验室的标准化和及标准化差值，可以得出所有的 S 和 D 的中位值和标准化 IQR，即中位值（Z），标准化 IQR(D) 等（这些总计统计量通常在报告表中列出，便于参加者自己计算 Z 比分数）。

随后计算实验室间 Z 比分数（ZB）和实验室内 Z 比分数（ZW），即

$$ZB = \frac{S - \text{中位值}(S)}{\text{标准IQR}(S)} \quad \text{和} \quad ZW = \frac{D - \text{中位值}(D)}{\text{标准IQR}(D)}$$

在报告中列表给出计算的 Z 比分数（见 A.8），并依据这些 Z 比分数来评定实验室的结果。把离群值定义为 Z（包括 ZB 和 ZW）绝对值大于等于 3 的结果或结果对，在表中，离群值在其 Z 比分数边上以（§）标出。

当实验室的 Z 比分数处在有问题的区间（即 $2 < |Z| < 3$ ）时，应鼓励实验室认真地检查

它们的结果偏差较大的原因。

对认为是离群的结果进行说明时，必须考虑 Z 比分数的符号和能力验证计划的设计。对于均一对和分割水平对，一个正的实验室间离群值（即 $ZB \geq 3$ ）表明该样品对的二个结果太高。而一个负的实验室间离群值（即 $ZB \leq -3$ ）表明其结果太低。

对于样品对，实验室内离群值（即 $|ZW| \geq 3$ ）表明其二个结果间的差值太大。

对于一个样品（ X ）的验证计划，一个简单的稳健 Z 比分数是离群值时， Z 比分数的符号可以表明结果太高（正）或太低（负），但不能确定离群是由于实验室间变动还是实验室内变动，或者是由二者所造成。

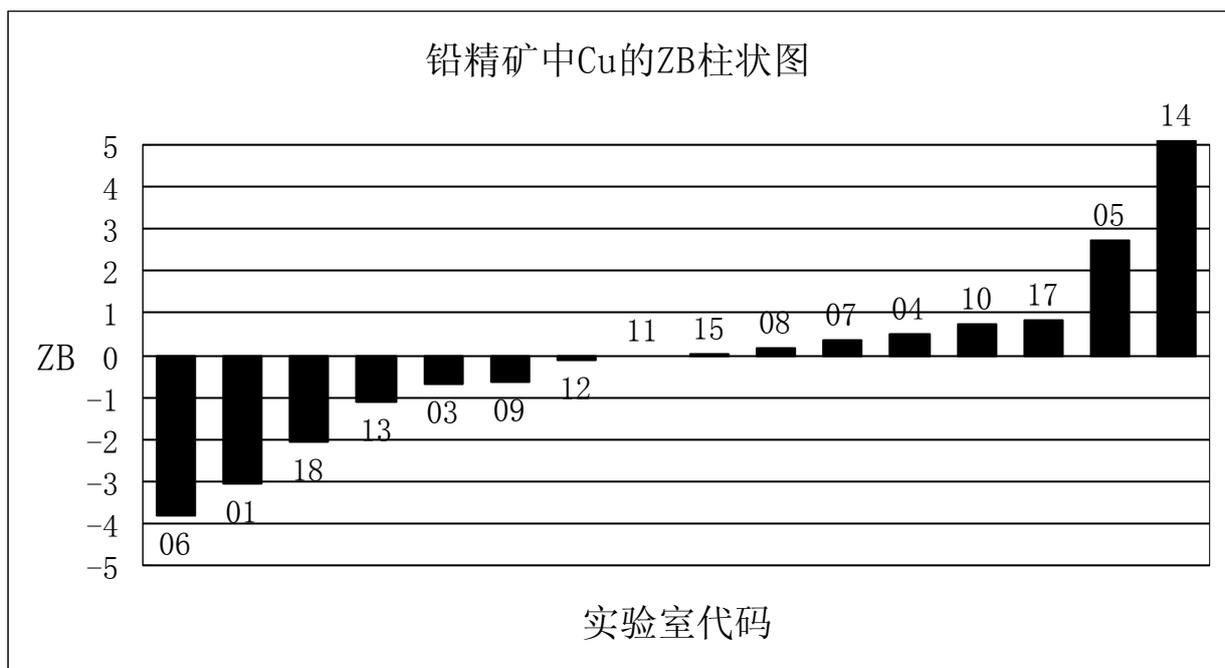
A.6 图形显示

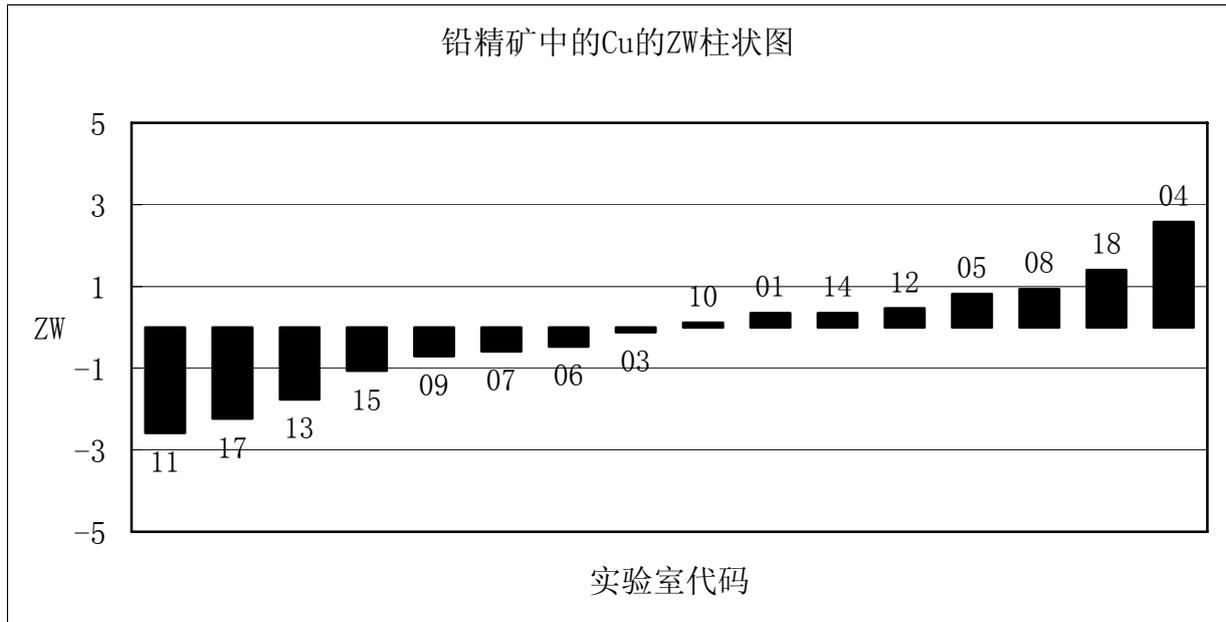
能力验证计划的报告除了包括结果、 Z 比分数表和总计统计量之外，通常还应包含一定数量的图表。两个最常使用的图形是 Z 比分数序列图和尤登图。

这些图能帮助组织者解释结果，而对于参加者也是非常有用的，特别是那些带有离群值的参加者，他们能够看到他们提交的结果与其他实验室结果的差异。

Z 比分数序列图

将 A.8 中所举的例子制作 Z 比分数序列图。图中按照大小的顺序显示出每个实验室的 Z 比分数（ ZB 和 ZW ），并标有实验室的编号，使每个实验室能够很容易地与其它实验室的结果进行比较。

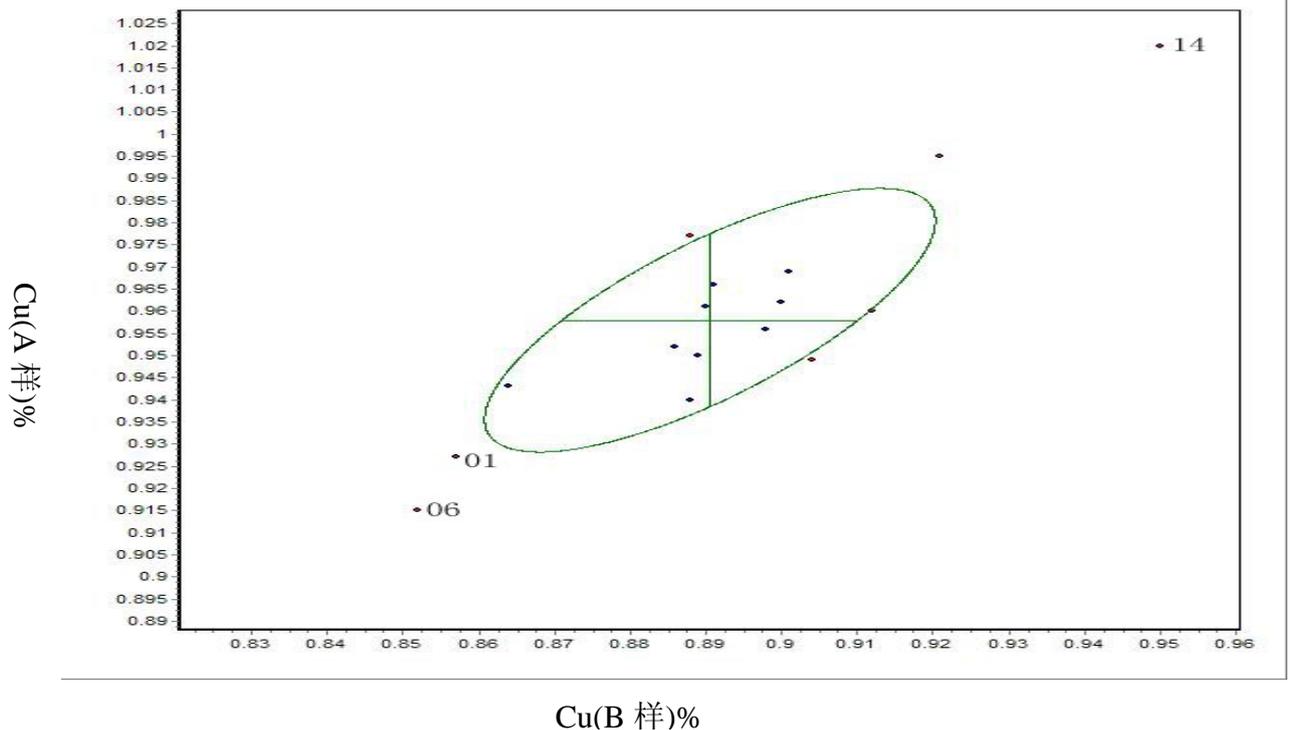




尤登图

尤登图是为二个样品的结果对而设计的。尤登图能显著地表示出实验室的系统偏差。是每个实验室的结果对，用黑点 • 表示。图中的椭圆表示约为 95% 概率的置信区域，椭圆的中心为二个样品中位值的交点。

铅精矿中 Cu 的分析尤登图



处于椭圆外的所有的点都标有相应的实验室编号。但要注意，这些点并不意味着都是离群。这是因为离群的标准 ($|Z| \geq 3$) 的置信水平约为 99%，而椭圆是约 95% 的置信水平。

这意味着，如果数据中没有离群值，期望大约有 5% 的结果将在椭圆外。然而因为能力验证的数据通常包含一些离群值，所以在大多数情况下将有超过 5% 的点在椭圆外。

尤登图中椭圆外的点，大体相当于那些 Z 比分数大于 2 或小于 -2 的值。因此，结果在椭圆之外但还不是离群值的实验室应当复查他们的结果。

尤登图的优点在于它们是真实数据的图示。在椭圆外的实验室能够看到它们的结果是怎样不同于其他的实验室。

从尤登图可以说明：

- (1) 含有明显系统误差的实验室（即实验室间变异）将在椭圆的右上象限或者在左下象限，即两个样品的结果异常地高或低；
- (2) 随机误差（即实验室内变异）明显高于其他参加者的实验室将处于椭圆外的左上或右下象限，即一个样品的结果过高，而另一个则过低。

然而应注意，尤登图只是用来说明数据，并不用来准确评定实验室的结果（结果的评定仍由 z 比分数确定）。

A.7 实验室综合表

在最终报告中，对检测样品中每个检测特性都应有一份综合表。该表包含了参加者的所有结果，同时带有相关的统计量和 Z 比分数（见 A.8 例）。如必要，还包含了对实验室的说明。

A.8 举例

铅精矿中 Cu 的测定结果和统计处理

实验室代码	铅精矿 A $w_{Cu}/\%$	铅精矿 B $w_{Cu}/\%$	S	ZB	D	ZW	方法代码
01	0.927	0.857	1.2615	-3.05§	0.0495	0.35	Cu-1
03	0.952	0.886	1.2997	-0.68	0.0467	-0.12	Cu-1
04	0.977	0.888	1.3188	0.51	0.0629	2.58*	Cu-1
05	0.995	0.921	1.3548	2.74*	0.0523	0.82	Cu-2
06	0.915	0.852	1.2495	-3.79§	0.0445	-0.47	Cu-1
07	0.962	0.900	1.3166	0.37	0.0438	-0.59	Cu-2

08	0.966	0.891	1.3131	0.15	0.0530	0.93	Cu-1
09	0.950	0.889	1.3004	-0.63	0.0431	-0.71	Cu-1
10	0.969	0.901	1.3223	0.73	0.0481	0.11	Cu-1
11	0.949	0.904	1.3103	-0.02	0.0318	-2.58*	Cu-1
12	0.961	0.890	1.3089	-0.11	0.0502	0.47	Cu-1
13	0.940	0.888	1.2926	-1.12	0.0368	-1.76	Cu-1
14	1.02	0.950	1.3930	5.11§	0.0495	0.35	Cu-1
15	0.956	0.898	1.3110	0.02	0.0410	-1.06	Cu-1
17	0.960	0.912	1.3237	0.81	0.0339	-2.23*	Cu-1
18	0.943	0.864	1.2777	-2.04*	0.0559	1.40	Cu-1
结果数	16	16	16		16		
中位值	0.958	0.891	1.3106		0.0474		
标准 <i>IQR</i>	0.0143	0.0106	0.01612		0.00603		
稳健 <i>CV</i> (%)	1.49	1.19	1.23		12.72		
最小值	0.915	0.852	1.2495		0.0318		
最大值	1.020	0.950	1.3930		0.0629		
极差	0.105	0.098	0.1435		0.0311		

注：加§号的数值为离群值，即 $|Z| \geq 3$ ；加*号的数值为可疑值，即 $2 < |Z| < 3$ 。

附录 B:

校准实验室间能力验证计划的结果处理方法

B.1 简介

本文件用于对校准实验室间能力验证计划（测量比对计划）结果的评定。

B.2 统计设计

对于校准实验室间的能力验证计划（也称为测量比对计划）结果的统计设计，采用国际上普遍接受的 E_n 值来进行评价。

B.3 E_n 比率

在测量比对计划中，使用 E_n 比率来评定某一实验室的每一个单独结果。按下式计算

E_n ：

$$E_n = \frac{x_{LAB} - x_{REF}}{\sqrt{U_{LAB}^2 + U_{REF}^2}}$$

式中： x_{LAB} -参加实验室的结果；

x_{REF} -参考实验室的结果；

U_{LAB} -参加实验室报告的不确定度（如是认可项目，即为认可的不确定度）；

U_{REF} -参考实验室报告的不确定度。

U_{LAB} 和 U_{REF} 的置信水平为 95%。

对于一个结果，可接受的一个 E_n 比率（也称 E_n 值）应在 -1 到 +1 之间，即 $|E_n| \leq 1$ （越接近零越好）。

在检测实验室间比对中， Z 比分数表明了实验室的测试接近给定值的程度。但在校准实验室间的比对中， E_n 值表明实验室是否是在参考值（给定值）的特定测量不确定度中。

E_n 值并不表明哪个实验室的结果最接近参考值，它只表明其测量结果是否符合对实验室要求的不确定度。因此，报告了小的不确定度的实验室，可能和在非常低水平（即较大的不确定度）上工作的实验室具有一个相似的 E_n 值。

在一系列相似的测量中， E_n 值期望呈正态分布。所以当考虑 $|E_n|$ 明显大于 1 的结果的时，应评价这个实验室出具的所有结果，观察是否存在一个系统偏离（例如 E_n 值始终是正值或负值）。

下面是 1 伏 直流电压标准实验室间比对的结果，实验室报出的不确定度和 E_n 比率列于表中。实验室 4 和 6 的结果被认为不满意。

1伏 DC

实验室代码	实验室结果-参考值 (μV)	U_{95} (μV)	E_n
参考	0	1	
1	-1	2	-0.45
2	2	2	0.89
3	-3	3	-0.95
4	2	1	1.41
5	0.5	1.5	0.28
6	-2.5	2	-1.12

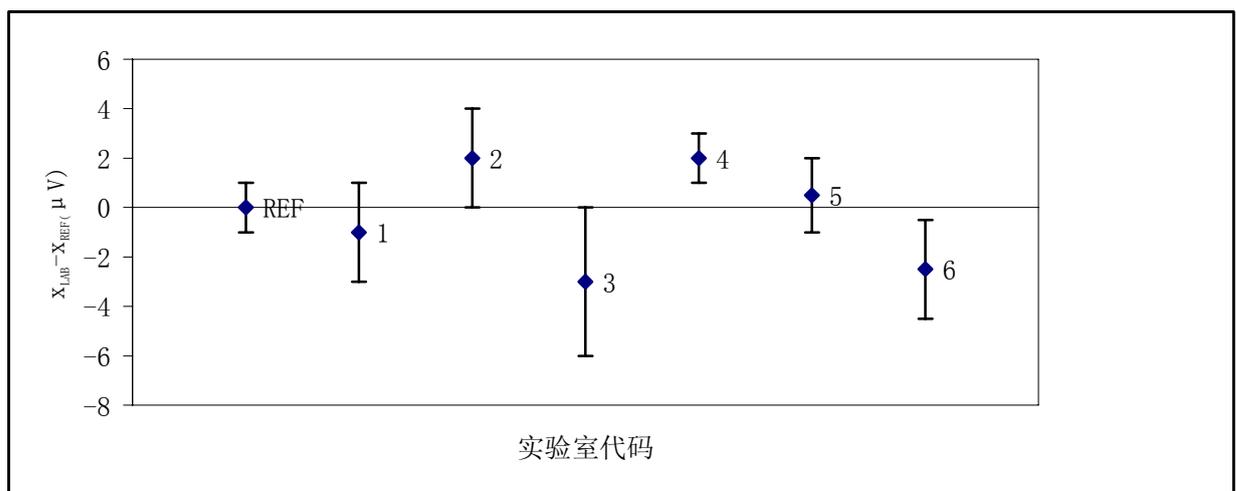
B.4 测量不确定度

在 E_n 值中使用了由实验室报告的测量不确定度。如果实验室报告的不确定度大于其被认可的不确定度，它应该作出解释。

B.5 图形显示

在最终报告中应有结果的图示及其相关的不确定度。以下面的图例显示了在 B.3 表中的结果。每个参加实验室的值和参考值用黑色菱形 ◆ 表示。实验室值向上和向下延伸的线段代表实验室报告的不确定度。

1Volt DC



但需要注意, 图形仅仅是一个数据的说明, 可以宏观地比较所有实验室的结果及其不确定度。它们不代表对一个结果的最终评定 (结果的评定由 E_n 值来确定)。

附录 C:

测量审核结果的评定

C.1 引言

测量审核是能力验证活动的形式之一。测量审核是将充分表征和校准过的物品送至一个实验室进行检测或校准,然后将实验室的结果与参考值进行比较,以确定实验室的能力。

本附录介绍几种测量审核结果的评定方式。对测量审核结果的评定,除了按本指南 4.3.1.4 中的 E_n 值进行判定外,使用者应根据实验室、测量方法及测量物品的具体情况,选用合适的方式对实验室能力进行评价。

C.2 测量审核结果的评定方式

1. 按 E_n 值评定

按下式计算 E_n 值:

$$E_n = \frac{x_{LAB} - x_{REF}}{\sqrt{U_{LAB}^2 + U_{REF}^2}} \quad (1)$$

式中: x_{LAB} -实验室的测量结果;

x_{REF} -被测物品的参考值;

U_{LAB} -实验室获得认可的能力的扩展不确定度;

U_{REF} -参考值的扩展不确定度。

U_{LAB} 和 U_{REF} 的置信水平为 95%。

若 $|E_n| \leq 1$, 则判定实验室的结果为满意, 否则判定为不满意。

利用 E_n 值评定测量结果是测量审核结果评定的基本方式, 但前提是必须正确评定该实验室对该项测量的不确定度。如果实验室不能正确评定其测量不确定度, 则无法使用该方法。

2. 按临界值 (CD 值) 评定

当实验室对检测和测量的不确定度缺乏正确的评定, 而用于该测量的标准方法提供有可靠的重复性标准偏差 σ_r 和复现性标准偏差 σ_R 时, 可采用本方法对测量审核结果进行判定。

根据 ISO5725-6, 按下式计算 CD 值:

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

实验室在重复条件下 n 次测量的算术平均值 \bar{y} 与参考值 μ_0 之差 $|\bar{y} - \mu_0|$ 小于临界值 CD 值, 则该实验室的测量结果可以接受, 实验时结果判定为满意结果, 否则判定为不满意结果。

3. 按专业标准方法规定评定

在 E_n 值及 CD 值不可获得时, 如相应专业标准规定了测试结果允许差, 可按标准规定评定。

按下式计算 Z 值:

$$Z = \frac{x_{LAB} - x_{REF}}{\Delta}$$

式中: x_{LAB} - 实验室的测量结果;

x_{REF} - 被测物品的参考值;

Δ - 标准中规定的允许差。

若 $|z| \leq 1$, 则判定实验室的结果为满意, 否则判定为不满意。