

国内外化学检测领域能力验证结果统计方法的探讨

伍 恒, 陈祝康*, 顾颂青

(上海市食品药品检验所, 上海 201203)

摘 要: 正确地选择和使用能力验证结果统计方法对于能力验证的成功实施具有至关重要的作用。本文对国内外化学检测领域能力验证的参考指南进行检索, 重点分析各指南或标准在指定值、能力评定标准差、能力统计量、图示法等能力验证结果统计关键环节的异同点。通过分析发现, 不同的认证认可组织对于能力验证结果的统计方法及能力评定既相似又有各自的特点, 这些组织制定的能力验证结果统计及能力评定准则大多以ISO/IEC17043、ISO 13528 为基本依据进行编制, 但又不完全相同, 有不同程度的细化、扩展和限制。本文提供的检索资源及提炼的观点可以为我所以能力验证提供者身份首次组织实施能力验证提供理论参考, 为能力验证统计人员正确选择和使用统计方法提供更多的信息和借鉴。

关键词: 能力验证; 结果统计方法; 能力验证提供者

Discussion on statistical methods used in analytical chemistry proficiency testing both in China and abroad

WU Heng, CHEN Zhu-Kang*, GU Song-Qing

(Shanghai Institute for Food and Drug Control, Shanghai 201203, China)

ABSTRACT: It is very important for the successful implementation of proficiency testing program to select and apply the statistical methods correctly. The domestic and overseas reference protocols for proficiency testing of analytical chemistry were searched, and the critical statistical techniques of each protocol or standard, such as the assigned value, the standard deviation for proficiency assessment, performance statistics and graphical methods were analyzed. It is found that different organizations had similarities as well as characteristics in statistical techniques. The proficiency testing statistical standards or protocols made by these organizations were mostly based on ISO/IEC17043, ISO 13528, but were not the same with varying elaboration, expansion or limitation. The retrieval resources and essential points of the study can provide theory references for organizing proficiency testing program as proficiency testing provider for the first time, and provide proficiency testing statistician more valuable information on the right selection and employment of statistical techniques.

KEY WORDS: proficiency testing; statistical techniques; proficiency testing provider

*通讯作者: 陈祝康, 主任药师, 主要研究方向为能力验证计划实施与质量管理。E-mail: great310g@sina.cn

*Corresponding author: CHEN Zhu-Kang, Senior Pharmacist, Shanghai Institute for Food and Drug Control, Shanghai 201203, China. E-mail: great310g@sina.cn

1 引言

能力验证对于评估一个组织机构的能力具有重要意义和辅助作用,其作为实验室内部质量控制的补充,提供了对其检测或测量能力的外部评价手段,是实验室将同一项目中的表现与其他实验室进行比较的有效工具。实验室参加能力验证有益于实验室长期能力的监控及整体风险管理,可促进实验室提高检测、校准能力,推动实验室开展检测人员教育、培训及能力监控。鉴于能力验证的重要意义,实验室认可国际和区域性合作组织均已把能力验证要求作为签署国际相互承认协议(mutual recognition arrangement, MRA)的基本条件。

本文归纳了国内外化学检测领域能力验证结果统计相关的重要指南及标准,从指定值、能力评定标准差及能力统计量等方面比较它们的异同点,以期为我所以能力验证提供者身份首次组织实施能力验证提供参考。

2 国内能力验证机构结果统计方法

国内有资格组织实施能力验证的机构包括:中国合格评定国家认可委员会(China National Accreditation Service for Conformity Assessment, CNAS)、国家认证认可监督管理委员会(Certification and Accreditation Administration of the People's Republic of China, CNCA)及获认可的能力验证提供者。这些机构组织实施能力验证一

般是依据 CNAS 制定的能力验证指南及部分国家推荐性标准。其中与能力验证结果的数据统计及评价相关的有 CNAS-CL03: 2010《能力验证提供者认可准则》^[1](附录 B 能力验证的统计方法,等同 GB/T-27043 附录 B)、CNAS-GL02: 2014《能力验证结果的统计处理和评价指南》^[2](根据 GB/T-27043 制定,同时参考 GB/T 28043)、GB/T 28043-2011《利用实验室间比对进行能力验证的统计方法》^[3](等同 ISO13528)以及 GB/T 27043-2012《合格评定能力验证的通用要求》^[4](等同 ISO/IEC 17043)。

由于以上提到的 4 个文件之间存在较大程度的相互引用,有较多重复内容,下面不再逐一介绍。重点从能力验证结果统计的关键点分析它们的异同,相同之处为:指定值的赋值方法,包括有证参考值、参照值、专家实验室公议值和参加者公议值;能力评定标准差的赋值方法,包括规定值、经验值、一般模型、精密度试验以及一轮能力验证计划所得数据确定;性能统计量,包括 D 值、 $D\%$ 、秩与秩百分数、 Z 值、 Z' 值、 En 值、 ζ 值及 EZ 值;图示法, CNAS-GL02、GB/T 27043 均指向 GB/T 28043 及国际理论和应用化学联合会(International Union of Pure and Applied Chemistry, IUPAC)报告的图示法。不同之处见表 1。在组织实施能力验证时,对于结果中离群值的处理及标准化四分位距计算稳健标准偏差可参考 CNAS-GL02;对于定性和半定量结果的处理可参考 GB/T27043;对于指定值的不确定度计算可参考 GB/T 28043。

表 1 国内能力验证结果统计及能力评价相关指南、标准的比较

Table 1 Comparison of relative guidance and standards on statistic treatment of proficiency testing results and performance evaluation in China

	CNAS-CL03 ^[1]	CNAS-GL02 ^[2]	GB/T28043 ^[3]	GB/T-27043 ^[4]
指定值	略	等同 GB/T-27043	有依据不确定度来判断指定值是否可接受的具体准则、指定值的比方法、测试结果有缺失时的处理方法。	定性实验推荐使用大多数参加者的结果作为指定值。
能力评定标准差	略	推荐两种迭代法和标准化四分位距法计算的稳健标准偏差,并提到一种简化的迭代算法 A。	推荐两种迭代法计算的稳健标准偏差,没有提到标准化四分位距法。	由参加者结果得到的传统标准差、稳健标准差均可。
性能统计量	略	/	/	定性结果和半定量结果的处理
离群值	略	与 GB/T-27043 相比,补充“明显错误的结果应由专家进行识别和判断”,删除“在较大的或常规的能力验证计划中,如存在有效的客观判据,则可自动剔除离群值。”	无离群值处理方法	提出离群值的处理方法

3 国外能力验证机构结果统计方法

目前国际上比较知名的为认证认可机构提供能力验证计划的相关组织包括: 亚太实验室认可合作组织(Asia Pacific Laboratory Accreditation Cooperation, APLAC)、澳大利亚国家检测机构协会(National Association of Testing Authorities, NATA)、美国分析化学家协会(Association of Official Agricultural Chemists, AOAC)、英国食品与环境研究院(The Food and Environment Research Agency, FERA)、国际认可合作组织(International Laboratory Accreditation Cooperation, ILAC)以及国际纯粹与应用化学联合会(International Union of Pure and Applied Chemistry, IUPAC)等, 他们制定或主要参考的能力验证结果统计相关指南如表 2 所示。

由于 APLAC、NATA 及 ILAC 能力验证结果评价所依据的指南为 ISO 13528: 2005 或 ISO/IEC 17043: 2010, 这两个指南已有国内转化标准, 在第 2 部分已有详细分析, 且在这 3 家机构的官网上没有查询到其制定的能力验证结果统计相关文件或应用范例, 下面主要介绍 AOAC、FERA 和 IUPAC 的结果统计方法。

3.1 美国分析化学家协会(Association of Official Agricultural Chemists, AOAC)

AOAC International 是经美国实验室认可协会认证的能力验证提供者。其组织的能力验证领域包括微生物和化学分析项目^[15-22]。AOAC 结果报告的特点在于, 不限制参

加者使用的检验方法, 并将等同的检验方法进行分组, 方便直观的比较每组之间结果、指定值和标准偏差的差异。

AOAC 官方网站上检索到的结果报告书中, 所有定量项目的数据统计方法指向 ISO 13528 的迭代法, 并报告指定值的标准不确定度, 结果评价采用 ISO 17043 的 Z 值, 见表 3。图示法采用 Z 值箱线图, 见图 1。

表 2 国际化学检测领域能力验证结果统计及能力评价依据的相关指南

Table 2 The international chemical testing guidance on statistic treatment of proficiency testing results and performance evaluation

机构	指南
APLAC	ISO 13528: 2005 ^[5] ISO/IEC 17043: 2010 ^[6] ISO 5725: 1994 1-6 ^[7] APLAC PT002 ^[8]
NATA	ISO/IEC 17043: 2010
AOAC	ISO 13528: 2005 ISO/IEC 17043: 2010
FERA	Protocol for proficiency testing scheme Part 1-Part 5 ^[9-13] IUPAC Technical Report ^[14]
ILAC	ISO/IEC 17043: 2010
IUPAC	IUPAC Technical Report ISO 13528: 2005 ISO Guide 43(已被 ISO/IEC 17043: 2010 代替)

表 3 国外能力验证机构结果统计及能力评价方法的比较

Table 3 Comparison of the method of international organizations on statistic treatment of proficiency testing results and performance evaluation

	AOAC	FERA	IUPAC
指定值的赋值	/	参加者结果的公议值、预设值及有证参考值	参考实验室的测量值、有证标准物质(CRM)中的认可值、通过 CRM 与能力验证测试材料的直接对比值、专家实验室的认同值、确定值、参加者共识值。
参加者公议值的计算	稳健均值(ISO 13528: 2005 迭代法)	除去异常值后, 结果分布为单峰, 且接近对称, 可用稳健均值、中位值、mode 作为指定值。	
指定值的不确定度	/	见 3.2.1	见表 4
能力评定标准差的赋值	参加者结果的稳健标准偏差 (ISO 13528: 2005 迭代法)	预测模型(Horwitz 方程或其变型)、测量方法信息、法规规定值、专家意见	推荐基于“fitness-for-purpose”的 σ_p (Horwitz 方程或其变型)或法规规定值, 不建议使用专家感知值、参加者结果的稳健标准偏差。
定量结果的评价	Z 值	Z 值, Q 值	可使用 Z 值、 Z_L 值、En 值、EZ 值, 不推荐 Z' 值、 ζ 值。
Z 值的判断标准	Z ≥ 2 满意, $2 < Z < 3$ 可疑, Z ≥ 3 不满意		

续表 3

	AOAC	FERA	IUPAC
定性结果的评价	将参加者结果与预期结果进行对照。	当指定值为预设值时, 结果评价为“满意”或“不满意”, 当指定值为参加者的公认值时, 结果评价为“一致”或“不一致”。	/
长期能力评价	/	/	组合 Z 值、图示法, 更推荐后者
图示	Z 值箱线图	一轮能力验证推荐 Z 比分数序列图(见图)	一轮能力验证的图示没有推荐的图示法, 长期能力评价推荐 Schewhart chart 和 J-chart

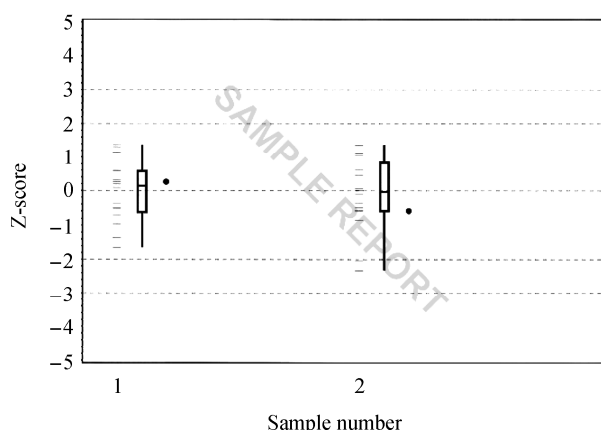


图 1 AOAC 肉类化学分析项目能力验证报告结果的 ZZ 值分布图
Fig. 1 ZZ-score distribution graph of proficiency testing results in the AOAC meat chemistry program

Z 值箱线图用以直观地观察某个实验室采用某种方法所得 Z 值在总体 Z 值中的位置。图中以黑色的圆点代表某实验室 Z 值, 以水平短线表示其他参加者 Z 值, 长方形的箱体代表所有实验室 Z 值中间的 50%, 箱体中的黑色水平线中位值, 须线代表合理拖尾, 须线的最高处代表最大值, 但不超过 1.5 倍箱高, 最低处类似。

3.2 英国环境与科学研究院(The Food and Environment Research Agency, FERA)

为了促进欧盟食品法规实施和执行的有效性, 英国农业部、渔业部、粮食部启动了评价分析实验室技术水平的计划。此计划由英国中心科学实验室(British Central Science Laboratory, CSL)的能力验证小组执行, 并被称为“食品分析能力评价体系”, 2009 年 CSL 正式更名为英国环境与科学研究院。

FERA 在“能力验证草案”(protocol for proficiency testing scheme)中对能力验证计划的组织、参与、能力评价等进行了详细介绍。“能力验证草案”包括 5 个部分, 第一

部分 common principles 为 FAPAS 组织的所有 PT 项目的共同原则; 第二部分到第五部分为对总则的补充, 分别为 FAPAS (Food analysis performance assessment scheme, 服务范围为食品、饲料和饮料的理化检测)、FEPAS(Food examination performance assessment scheme, 服务范围为食品、饲料的微生物检测)、GeMMA(Scheme the genetically modified materials analysis scheme, 服务范围为食品、饲料的转基因成分检测)、LEAP(Scheme the laboratory environmental analysis proficiency scheme, 服务范围为饮用水与水体环境检测)。

3.2.1 指定值的赋值及其不确定度

指定值的赋值及参加者结果公认值的计算方法见表 3。确定指定值之前要剔除离群值, 然后分析结果的分布, 若结果分布为近似正态分布, 则可用稳健法得到平均值, 若结果分布为不对称分布, 则先用核密度图识别多峰性。

针对指定值的 3 种统计方法, 不确定度的估计方法也有所不同。

表 4 FERA 能力验证草案中指定值不确定度计算方法
Table 4 The methods of determining uncertainty on the assigned value in FERA protocol for proficiency testing scheme

指定值 X_a	指定值的不确定度 v
稳健均值	$v = \hat{\sigma} / \sqrt{n}$
中位值	$v = sMAD(\text{绝对中位差}) / \sqrt{n}$
Mode	SEM(Mode 的标准偏差)

3.2.2 能力评定标准差的赋值

FERA 常用的能力评定标准差的赋值方法有 Horwitz 方程, 或根据测量方法提供的信息、法规规定、专家意见确定能力评定标准差(见表 3)。

3.2.3 能力验证性能统计量及结果评价

FERA 在第一部分总则中规定定量结果需转化为标准化统计量, 统计量的种类有 Z 值和 Q 值, Z 值是 FERA 能力

验证最常用的统计量。 Q 值则非常少用,只提示结果的相对误差,在结果评价中只作为参考方法,而不用做结果评价,适用于无法得到能力评定标准差的情况, Z 值和 Q 值的计算公式如下:

$$Z=(x-x_a)/\sigma_p, Q=(x-x_a)/x_a$$

x =参加者结果; x_a =指定值; Q =结果真值的最佳估计

σ_p =能力评定标准差

3.2.4 定量结果的图示法

FERA 定量能力验证计划采用 Z 比分数序列图(见图 2)。

3.3 国际纯粹与应用化学联合会(International Union of Pure and Applied Chemistry, IUPAC)

IUPAC 是各国化学学会的一个联合组织。化学分析实验室能力验证国际协调公约由国际标准化组织 AOAC、ISO 和 IUPAC 共同合作制定,目的是评价实验室对于已经建立的标准方法的检测能力。

IUPAC 对于能力验证结果的数据统计及评价有其独有的原则,即能力评定标准差和 Z 值的计算是基于“fitness-for-purpose”的原则,这个词在 IUPAC 报告中被反复强调,即综合考虑准确性、成本等因素选择的最优统计方法。

3.3.1 指定值的赋值及其不确定度的推算

指定值的赋值见表 3。其中用参加者公议值作为指定值时,除去结果中的异常值后,若结果分布为单峰,且接

近对称,可用稳健均值、中位值、mode 作为指定值。中位值更加适合数据频率分布严重倾斜的情况,而稳健均值(即 ISO13528^[5]附件 C 中的方法 A)更加适合数据分布为近似对称的情况,多数情况下比中位值有更小的标准误差,若结果分布为多峰或双峰分布,用 mode 作为指定值。由于参加者公议值不独立于参加者结果,导致无法合理识别整体偏差,若大部分结果是偏离的,正确的参加者将会被误判,并且当参加者数量少于 15 时,不确定度较高,由此产生了指定值不确定度的推算方法。本文将 IUPAC 技术报告中指定值不确定度的推算过程转化为直观的流程图,见图 3:

3.3.2 能力评定标准差的赋值方法

推荐 Horwitz 方程或法规规定值,不建议使用专家感知值、参加者结果的稳健标准偏差(见表 3)。

3.3.3 定量结果的评价

关于定量结果的评价,技术报告中允许使用的几种性能统计量有: Z_L 值、 Z' 值、 ζ 值、 En 值及 EZ 值。

Z_L 值适用于不基于“fitness for purpose”的能力验证项目,提供者没有参考任何外部实际需求,仅根据参加者结果计算评分,或者参加者发现提供者 IUPAC 基于“fitness for purpose”的方法并不适合其某项工作,在这种情况下,参加者可依据其自身的要求选择合适的评分方法,能力评定标准差采用 σ_{ffp} , Z_L 的计算公式为 $Z_L=(x-x_a)/\sigma_{ffp}$ 。

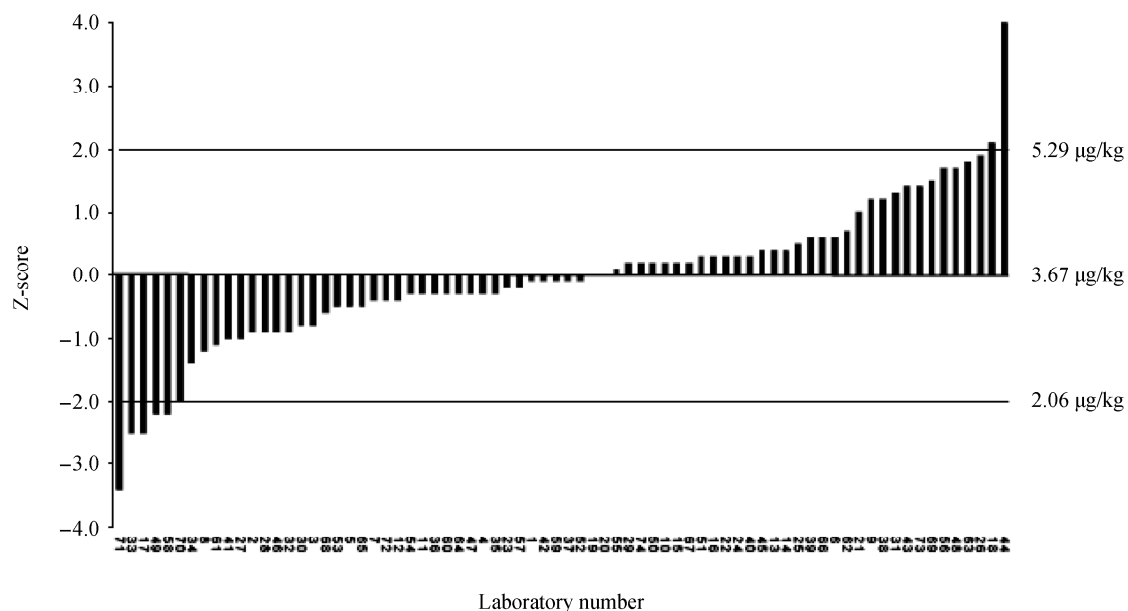


图 2 FAPAS 能力验证项目 17107 面粉中赭曲霉毒素 A 的 ZZ 值分布图

Fig. 2 ZZ-scores distribution diagram of ochratoxin A in 17107 wheat flour for FAPAS proficiency test

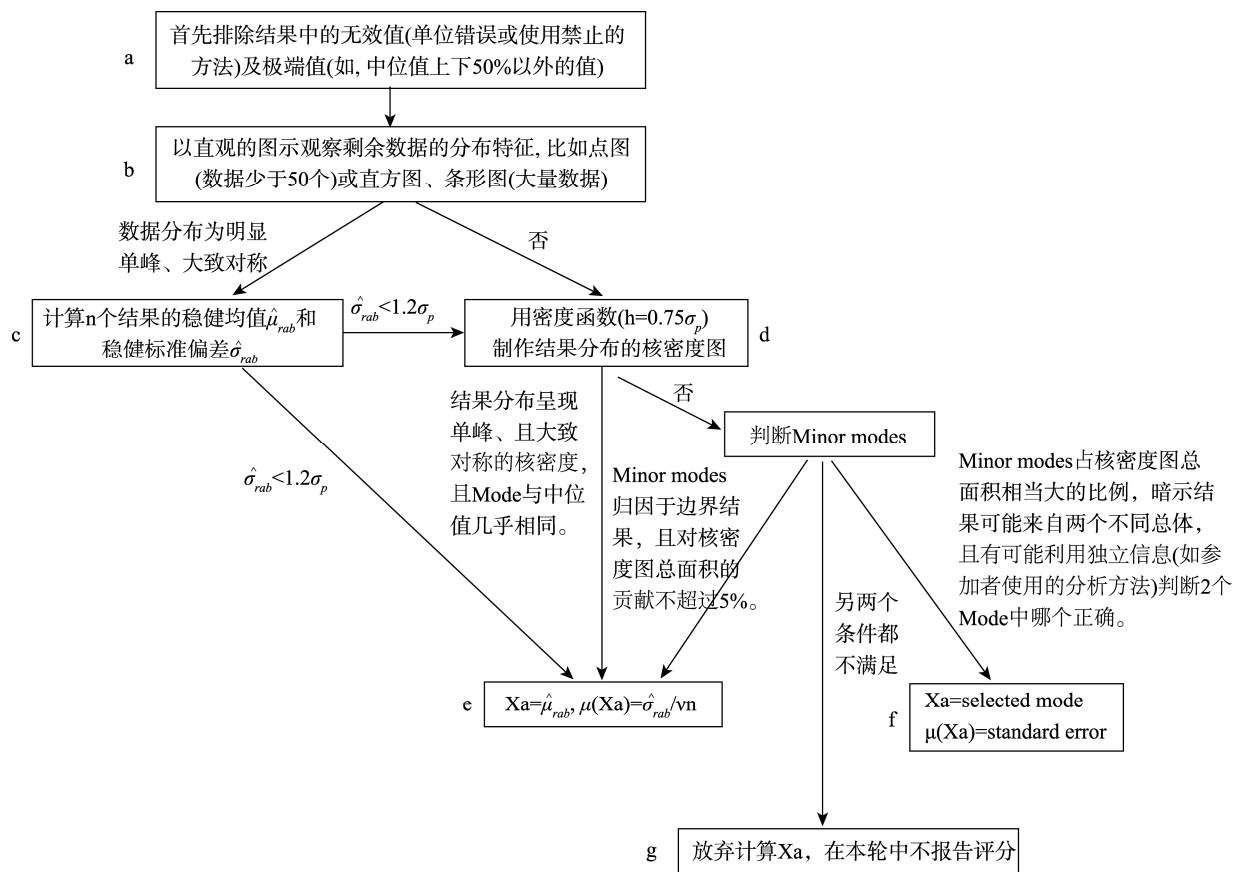


图 3 IUPAC 技术报告中推算指定值不确定度的流程图

Fig.3 The flow chart of determining the uncertainty on the assigned value in IUPAC technical report

ISO 13528 提出当指定值的不确定度不可忽略时(即 $\mu(x_a) > 0.3\sigma_p$) 需使用 Z' , 但在本准则中并不推荐使用 Z' , 原因是使用 Z' 可能会掩盖指定值不确定度过高的事实。同时, 本准则也不推荐使用 ζ 值, 除非参加者另有需求, 但并未说明具体原因。

3.3.4 长期能力评价

收集长期的 Z 值可以反应实验室长期的不确定性。可采用组合 Z 值及图示法评价。

(1) 组合 Z 值

组合 Z 值有以下两种形式:

$$S_{Z,rs} = \sum_i Z_i / \sqrt{n} \quad S_{ZZ} = \sum_i Z_i^2$$

使用组合 Z 值评价时需注意避免错误结论, 特别强调组合 Z 值对于不同分析成分的长期能力评价使用范围有限, 且存在潜在问题, 容易误用。

(2) 图示法

技术报告更加推荐用图示法表示实验室的长期能力, 此方法更加直观, 且不容易引起误解。常用的有 *Schewhart*

chart、*J-chart*, 见图 4、图 5。

图 4 中向下的箭头表示小于 0 的 Z 值, 向上箭头表示大于 0 的 Z 值; 小三角表示 $2 < |Z| < 3$, 大三角表示 $|Z| > 3$ 。

J-chart 结合了休哈特图和累积和图的特点, 可同时识别分析系统中较小偏差和突然出现的较大变异。作图前需要先将 Z 值转化成 J 值, 用 J 值作图。 Z 和 J 的转化关系如下, 当 J 的累积和达到 8 时需采取行动, 触发调查措施。

- If $Z \geq 3$ then $J=8$
- If $2 < Z < 3$ then $J=4$
- If $1 < Z < 2$ then $J=2$
- If $-1 < Z < 1$ then $J=0$
- If $-2 < Z < -1$ then $J=-2$
- If $-3 < Z < -2$ then $J=-4$
- If $Z \leq -3$ then $J=-8$

4 结 语

对于能力验证的组织实施方, 不仅要能对能力验证项目的进程进行宏观把控, 熟悉相关检验检测技术, 还要熟

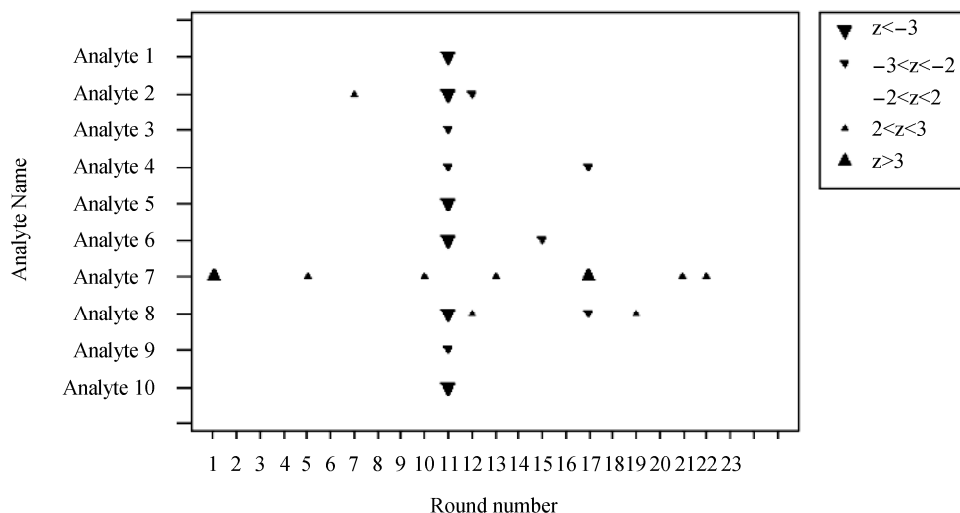


图 4 IUPAC 技术报告中的休哈特控制图
Fig.4 Schewhart chart in IUPAC technical report

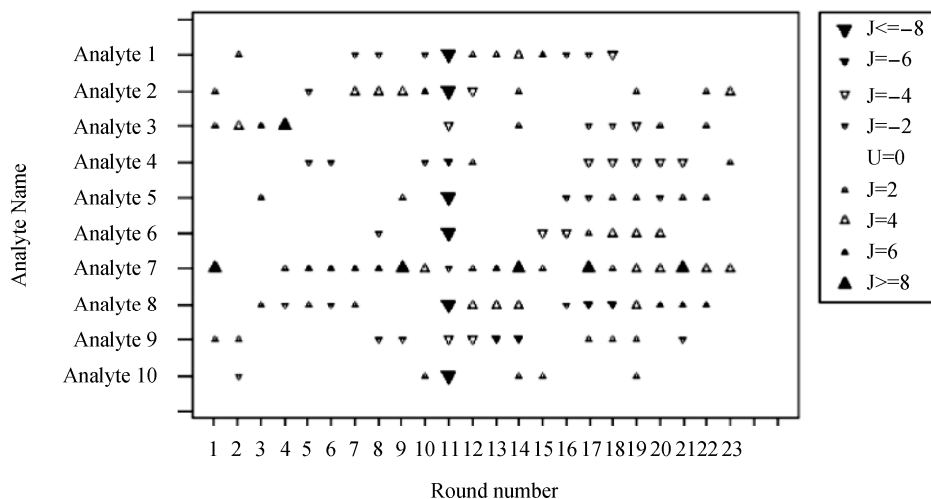


图 5 IUPAC 技术报告中的 J -chart
Fig 5. J -chart in IUPAC technical report

悉国内外相关指南的规定,掌握科学的统计学知识,才能为 P T 项目的顺利实施奠定坚实的基础。为认证认可机构提供能力考核计划的相关组织有很多,各个提供者对于能力验证结果的统计方法及能力评定既有相似又有各自的特点。本文通过对国内外相关指南进行分析,提炼出它们在指定值确定、能力评定标准差确定、图示法等结果统计关键环节的异同点,以期持续提升我所检验能力及其可靠性,为能力验证提供者组织实施能力验证提供理论参考。

参考文献

- [1] CNAS-CL03 能力验证提供者认可准则[S].
CNAS-CL03 Accreditation criteria for proficiency testing providers [S].
- [2] CNAS-GL02 能力验证结果的统计处理和评价指南[S].
CNAS-GL02 Guidance on statistic treatment of proficiency testing results and performance evaluation [S].
- [3] GB/T 28043-2011 利用实验室间比对进行能力验证的统计方法[S].
GB/T 28043-2011 Statistical methods for use in proficiency testing by interlaboratory comparisons [S].
- [4] GB/T 27043-2012 合格评定能力验证的通用要求[S].
GB/T 27043-2012 Conformity assessment-General requirements for proficiency testing [S].
- [5] BS ISO 13528:2005 Statistical methods for use in proficiency testing by interlaboratory comparisons [S].
- [6] Proficiency testing scheme providers ISO/IEC 17043 standard application document [Z].

- [7] ISO 5725: 1994 Accuracy (trueness and precision) of measurement methods and results [Z].
- [8] Testing Interlaboratory Comparisons-APLAC PT 002 [Z].
- [9] The Food and Environment Research Agency. Protocol for proficiency testing scheme, version 4, Part 1 Common principles [Z].
- [10] The Food and Environment Research Agency. Protocol for proficiency testing scheme, version 3, Part 2 FAPAS [Z].
- [11] The Food and Environment Research Agency. Protocol for proficiency testing scheme, version 3, Part 3 FEPAS [Z].
- [12] The Food and Environment Research Agency. Protocol for proficiency testing scheme, version 3, Part 4 GeMMA Scheme [Z].
- [13] The Food and Environment Research Agency. Protocol for proficiency testing scheme, version 4, Part 5 LEAP Scheme [Z].
- [14] Thompson M, Ellison SLR, Wood R. The international harmonized protocol for the proficiency testing of analytical chemistry laboratories [J]. Pure Appl Chem, 2006, 78(1): 45-196.
- [15] Report to participants in the AOAC laboratory proficiency testing program, Meat chemistry program [Z].
- [16] Report to participants in the AOAC laboratory proficiency testing program, Pathogen-free microbiology program [Z].
- [17] Report to participants in the AOAC laboratory proficiency testing program, Combination pathogen program in meat matrix, *Salmonella*, *E. coli* O157: H7 and *L. monocytogenes* [Z].
- [18] Report to participants in the AOAC laboratory proficiency testing program, Combination pathogen program in meat matrix, *Salmonella* and *L. monocytogenes* [Z].
- [19] Report to participants in the AOAC laboratory proficiency testing program, Standard microbiology program without *E. coli* O157:H7 & *Listeria* [Z].
- [20] Report to participants in the AOAC laboratory proficiency testing program, Standard microbiology program [Z].
- [21] Report to participants in the AOAC laboratory proficiency testing program, Pesticide residues in fruits & vegetables program, [Z].
- [22] Report to participants in the AOAC laboratory proficiency testing program, *Salmonella* in liquid egg program [Z].
- [23] FAPAS Proficiency Test Report 17107 Ochratoxin A in Wheat Flour [Z].

(责任编辑: 姚 菲)

作者简介



伍 恒, 硕士, 主要研究方向为食品安全与实验室质量管理。
E-mail: 369414949@qq.com



陈祝康, 硕士, 主任药师, 主要研究方向为能力验证计划实施与质量管理。
E-mail: great310g@sina.cn