

## • 质控与标规 •

# 四分位极差在血站 ELISA 质控的应用研究

苏武锦, 钟周琳, 邱昌文, 姜莹

(南宁中心血站, 广西南宁 530003)

**摘要:** 目的 使用四分位极差和中位数计算标准比值(稳健 Z-分数), 提高 Z-分数图识别质控离群值的能力。方法 基于四分位极差对参数  $\sigma$  进行估计具有稳健性的特点, 结合实例和模拟数据, 使用四分位极差和中位数计算稳健标准比值(稳健 Z-分数), 同时使用标准差和均值计算标准比值(Z-分数), 并比较两种分数图识别质控离群值的能力。结果 4 组质控数据的离群值, 稳健 Z-分数图检出离群值的例数与探索性统计一致, Z-分数图则仅能检出 1 例。结论 应用四分位极差和中位数指标构成的稳健 Z-分数图在非正态分布的质控数据中, 对离群值的检出能力优于应用标准差和均值指标构成的 Z-分数图。

**关键词:** 四分位极差; 质量控制图; 离群值

DOI:10.3969/j.issn.1673-4130.2014.06.035

文献标识码:A

文章编号:1673-4130(2014)06-0735-03

## Quartile range in the application of statistical quality control

Su Wujin, Zhong Zhoulin, Qiu Changwen, Jiang Ying

(Nanning Blood Center, Nanning, Guangxi 530003, China)

**Abstract: Objective** Using quartile range and median to calculate standard ratio(robust Z-score), improve the ability of labeling outliers. **Methods** Using quartile range estimation has the characteristics of robustness to parameter  $\sigma$ , combined with examples and simulation datas, using quartile range and median to calculate robust Z-score, and the use of standard deviation and mean to calculate Z-score, and comparison the ability of labeling outliers of the two score charts. **Results** Four groups of quality control of outlier data values, robust Z-score chart detected number and exploratory statistical agreement with the outlier, Z-score chart can only be detected in one case. **Conclusion** Application of quartile range and median of robust Z-score chart in quality control data of non-normal distribution, for outlier detection ability is superior to the application of the standard deviation and average value of Z-score chart.

**Key words:** quartile range; quality control chart; outliers

在统计学上, Z-score 称为标准比值, 也称 Z-分数, 或 Z-比值、Z 比分数等, 是单个观察值减去总体均值, 得到的差值除以总体标准获得的无因次度量。以 Z 分数作为离群值判断的方式, 其优点是方便快捷、原理简单易懂, 此法比较适合应用在符合正态分布的变量上。目前临床检验质量管理中, Z-分数图已经普遍用于质量控制图的绘制<sup>[1-3]</sup>。然而, 在质控实际检测中, 检测值不一定很符合正态分布要求, 存在右偏或左偏, 此时有些离群值不一定能被 Z-分数图检出。本文利用四分位极差对参数  $\sigma$  进行估计具有稳健性的特点<sup>[4]</sup>, 把它应用到标准比值的计算之中, 得到的 Z-分数称为“稳健 Z-分数”。

## 1 材料与方法

**1.1 样品** 北京康彻斯坦生物技术有限公司生产的弱阳性(0.2 IU/mL) HBsAg 质控品(批号: 201202001)及常规检测样品。

**1.2 仪器与试剂** STAR 全自动加样器、FAME 全自动酶免分析仪(瑞士 Hamilton 公司), 两者组合为全自动酶免检测系统。上海科华生物工程股份有限公司 HBsAg 诊断试剂盒(ELISA), 经国家食品药品检定研究院检验报告结论为合格的产品, 有效期内使用, 严格按照试剂盒说明和实验室规范进行操作。

## 1.3 方法

**1.3.1 数据收集** 每批样品使用全自动酶免检测系统

ELISA 检测, 按照试剂说明书要求平行检测试剂盒内阳性、阴性对照品和弱阳性质控品 1 次(孔); 实验结束后记录质控品的检测值, 共收集 20 例(A 组)。

**1.3.2 模拟数据** 为了说明方法的优劣, 本文从实例数据中衍生模拟出另外 3 种情况。情况 1(B 组): 由原始数据例 3 的检测值 3.30 设置为 1.35; 例 7 的检测值 2.43 设置为 1.43。情况 2(C 组): 由原始数据例 3 的检测值 3.30 设置为 2.30; 例 4 的检测值 2.17 设置为 4.17; 情况 3(D 组): 由原始数据例 3 的检测值 3.30 设置为 2.30; 例 4 的检测值 2.17 设置为 3.67; 例 16 的检测值 3.60 设置为 2.60; 例 20 的检测值 2.17 设置为 1.10。数据见表 1。

**1.3.3 质控图制作** (1) Z-分数图: 质控检测值与检测平均值之间的差, 除以质控品的标准差, 即  $Z = \frac{X_i - \bar{X}}{S}$ ,  $\bar{X}$  为质控品均值,  $X_i$  为单个检测值,  $S$  为质控品标准差。Z-分数图的刻度至少从 -4 到 +4, 平均数为 0,  $\pm 1$ 、 $\pm 2$ 、 $\pm 3$  为界限。Z-分数小于 -3 或者大于 3, 即认定为离群值。(2) 稳健 Z-分数图: 质控检测值与其检测值中位数之间的差, 除以质控品的标准差, 即  $Z = \frac{X_i - M}{QR} \times 1.349$ ,  $M$  为质控品检测值中位数,  $X_i$  为单个检测值,  $QR$  为四分位极差(第 3 个四分位数与第 1 个四分位数之差), 1.349 从标准正态分布上、下四分位数中导出。Z-分数图

的刻度至少从 -4 到 +4, 平均数为 0,  $\pm 1, \pm 2, \pm 3$  为界限。Z 分数小于 -3 或者大于 3, 即认定为离群值。

**1.3.4 离群值检出能力比较** 用两种计算方式对实例数据及模拟数据进行 Z-比值计算, 比值在同一质控图上描出, 同时使用 SPSS18.0 统计软件对质控数据进行探索性分析, MedCalc11 绘制箱须图。

**1.4 统计学处理** 使用 SPSS 18.0 统计学软件进行探索性统计、医学统计软件 MedCalc11.4.2.0 作箱式图以及 Excel 2003 进行 Z-分数图绘制。

## 2 结 果

**2.1 收集的原始数据(A组)及衍生模拟出的数据(B组、C组、D组),见表 1、2。**

**2.2 稳健 Z-分数图与 Z-分数图的比较** A 组计算获得的 20 个 Z-分数绘制成 Z-分数图, 未检出离群值, 稳健 Z-分数图则检出例 3、例 16 为离群值, 见图 1; B 组: 计算获得的 20 个 Z-分数绘制成 Z-分数图, 未检出离群值, 稳健 Z-分数图则检出例 3、例 7、例 16 为离群值, 见图 2; C 组: 计算获得的 20 个 Z-分数绘制成 Z-分数图, 检出例 4 为离群值, 稳健 Z-分数图则检出例 4、例 16 为离群值, 见图 3; D 组: 计算获得的 20 个 Z-分数绘制成 Z-分数图, 未检出离群值, 稳健 Z-分数图则检出例 4、例 20 为离群值, 见图 4。

**2.3 探索性统计及箱式图显示** A 组例 3 为离群值, 例 16 为极值; B 组例 3、例 7 离群值, 例 16 为极值; C 组例 4、例 16 为极值; D 组例 4、例 20 为极值。见图 5: 离群值, 使用“O”标记, 距箱上缘或下缘 1.5 倍至 3 倍四分位数间距的值为离群值; 极值, 使用“■”标记, 超出距箱上缘或下缘 3 倍四分位数间距的值。

**2.4 稳健 Z-分数图检出的离群值和极值与探索性统计一致。**

表 1 原始检测值及模拟数据(S/CO, n=20)

n	A组	模拟数据			n	模拟数据		
		B组	C组	D组		B组	C组	D组
1	2.18	2.18	2.18	2.18	11	2.36	2.36	2.36
2	2.60	2.60	2.60	2.60	12	2.09	2.09	2.09
3	3.30	1.35*	2.30*	2.30*	13	2.22	2.22	2.22
4	2.17	2.17	4.17*	3.67*	14	2.32	2.32	2.32
5	2.27	2.27	2.27	2.27	15	2.35	2.35	2.35
6	2.49	2.49	2.49	2.49	16	3.60	3.60	3.60*
7	2.43	1.43*	2.43	2.43	17	2.44	2.44	2.44
8	2.79	2.79	2.79	2.79	18	2.18	2.18	2.18
9	2.19	2.19	2.19	2.19	19	2.42	2.42	2.42
10	2.60	2.60	2.60	2.60	20	2.27	2.27	2.27
							1.10*	

\*: 模拟检测值。

表 2 各组数据的基本统计量

项目	A组	B组	C组	D组
均值( $\bar{x}$ )	2.46	2.32	2.52	2.38
标准差(s)	0.38	0.46	0.50	0.45
中位数(M)	2.36	2.30	2.36	2.36
QR	0.38	0.30	0.34	0.38

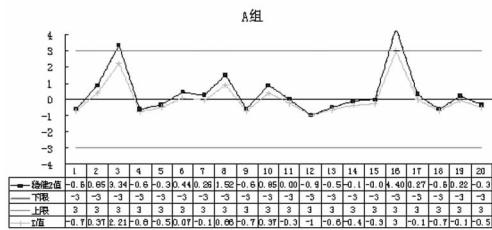


图 1 A 组 Z 分数与稳健 Z 分数的比较

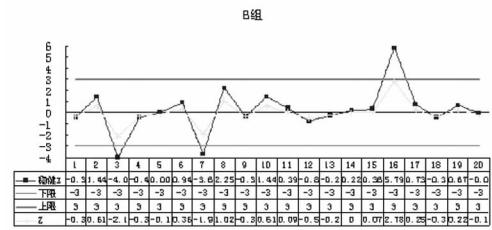


图 2 B 组 Z 分数与稳健 Z 分数的比较

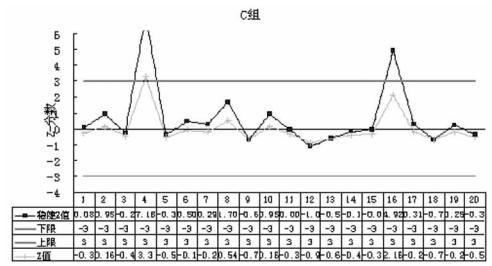


图 3 C 组 Z 分数与稳健 Z 分数的比较

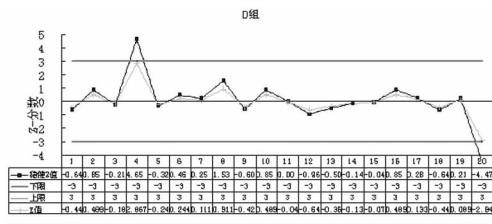


图 4 D 组 Z 分数与稳健 Z 分数的比较

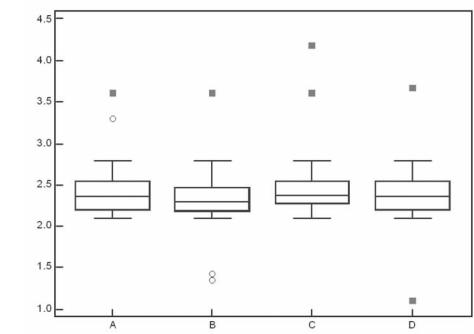


图 5 4 组数据的箱式图

## 3 讨 论

四分位极差, 即四分位数间距是指第 3 个四分位数(上四分位数, Q3)与第 1 个四分位数(下四分位数, Q1)相减而得, 记为  $QR = Q3 - Q1$ 。它表明除去两端各 25% 的数据后的极差, 是度量数据离散程度的指标, 与全距相比, 它排除了少数极端数值的影响。

正态分布根方差的估计假设总体为正态分布  $N(\mu, \sigma^2)$ , 则由于其上、下四分位数分别为  $\xi_0.75 = \mu + 0.6745\sigma$ ,  $\xi_0.25 =$

$\mu - 0.6745\sigma$ , 于是总体四分位极差为  $QR_1 = \xi 0.75 - \xi 0.25 = 1.349\sigma$ ; 若总体四分位极差  $QR^1$  用子样四分位极差  $QR$  替代, 则可得到正态分布  $N(\mu, \sigma^2)$  根方差的估计为  $= QR/1.349$ ; 利用四分位极差的方法对参数  $\sigma$  进行估计的最大优点就是具有稳健性<sup>[4]</sup>。目前在质量控制图中, 用于计算标准比值的公式为  $Z = \frac{X_i - \bar{X}}{S}$ 。本研究根据上述公式的思路和基于四分位极差对参数  $\sigma$  进行估计具有稳健性的特点, 使用四分位极差和中位数对稳健 Z 分数的计算进行了推导, 稳健  $Z = \frac{X_i - M}{\sigma} = \frac{X_i - M}{QR/1.349} = \frac{X_i - M}{QR} \times 1.349$ , 其中  $M$  为中位数,  $X_i$  为单个检测值,  $QR$  为四分位极差(第 3 个四分位数与第 1 个四分位数之差), 1.349 从标准正态分布上、下四分位数中导出。

本研究利用 Z 分数和稳健 Z 分数绘制质控图, 分别对 4 组数据进行了作图分析发现: 经探索性统计及箱式图确定的, A 组例 3 为离群值, 例 16 为极值; B 组例 3、例 7 离群值, 例 16 为极值; C 组例 4、例 16 为极值; D 组例 4、例 20 为极值。若使用 Z 分数分析时, 以上异常质控数据中只有 C 组例 4 数值为 4.17 被认为“失控”数据, 其余均为“在控”数据; 若使用稳健 Z 分数时, 则认为这些数据均为“失控”数据, 稳健 Z 分数图的离群值检出能力则与箱须图显示的完全相符。而本研究使用箱须图作为参照, 那是因为箱须图的箱子两端分别是上四分位数和下四分位数, 中间横线是中位数, 两端连线分别是除离群值外的最小值和最大值, 另外标记可能的离群值<sup>[5]</sup>, 它作为一种统计过程控制工具, 箱须图在离群值值检测上对于非正态分布的数据集中趋势相对于平均数更加合理<sup>[6-8]</sup>。

Z 分数图作为 Levey-Jennings 室内质控的一种图形, 其建立的理论依据之一就是数据正态性分布特性。然而, ELISA 实验受试剂、操作等因素影响, 质控实际检测中, 检测值不一定很符合正态分布要求, 存在右偏或左偏。因此, 当资料中存在有离群值时, 其标准差的值可能会很大, 所以相对地, Z 分数就会变小, 也就隐含在土 3 个 Z 分数范围以内, 当然就无法检出离群值了。再者, 很少有文献讨论到该方法对样本数的要求, 很容易被使用者误用。国外学者 Shiffler 在 1988 年就已经指出, 对任何一个变量来说, 其数据对应的 Z 分数可能的最大值为  $(n-1)/\sqrt{n}$ , 式中  $n$  为样本数<sup>[9]</sup>。按此计算, 当  $n=5$  时, 所有 Z 分数不可能大于 1.789, 当  $n=10$  时, 所有 Z 分数都小于 2.846, 也就是说当样本数小于 10 时, 不管离群值有多大, Z 分数图都无法检测出离群值。而四分位极差包括了居于中间位置 50% 的数据, 未考虑到两侧观察值的变异度, 故稳定性相对好<sup>[10-12]</sup>, 一般和中位数一起描述偏态分布资料。因为稳健 Z 分数图和箱须图的核心指标构成都是四分位极差和中位数, 所以稳健 Z 分数图在非正态分布的质控数据中, 对离群值的检

出能力优于应用标准差和均值等指标构成的 Z 分数图。当然, 稳健 Z 分数图对样本数是否有要求, 也是值得思考的。

需要说明的是: 稳健 Z 分数是当前实验室能力验证对结果进行评价的常用方法, 公式表达为  $Z = (\text{实验室结果}-\text{中位值})/\text{标准化 IQR}$ , 标准化  $IQR = IQR / (四分位数间距) \times 0.7413$ <sup>[13-15]</sup>, 事实上, 笔者文中的推导出公式与其是一样的, 即本文  $QR/1.349 = QR \times 1/1.349 = QR \times 0.7413 = IQR \times 0.7413$ 。

应用四分位极差和中位数指标构成的稳健 Z 分数图在非正态分布的质控数据中, 对离群值的检出能力优于应用标准差和均值指标构成的 Z 分数图。

## 参考文献

- [1] 丛玉隆, 冯仁丰, 陈晓东. 临床实验室管理学 [M]. 北京: 中国医药科技出版社, 2004: 140-148.
- [2] 中华人民共和国卫生部. 全国临床检验操作规程 [M]. 南京: 东南大学出版社, 2006: 84-85.
- [3] 冯仁丰. 临床检验质量管理技术基础 [M]. 2 版. 上海: 上海科学技术文献出版社, 2007: 238-240.
- [4] 胡宏昌, 占天易. 正态分布根方差的估计及其应用 [J]. 湖北师范学院学报, 2011, 31(1): 6-9.
- [5] 孙振球. 医学统计学 [M]. 3 版. 北京: 人民卫生出版社, 2010: 18-19.
- [6] David C. Hoaglin, Frederick Mosteller, John W. Tukey, 等. 探索性数据分析 [M]. 北京: 中国统计出版社, 1998: 62-101.
- [7] 官生平. SPC 统计过程管制 [M]. 厦门: 厦门大学出版社, 2004: 104-106.
- [8] 谢衷洁. 普通统计学 [M]. 北京: 北京大学出版社, 2004: 55-61.
- [9] Shiffler, R. E. Maximum Z score and outliers [J]. The American Statistician, 1988, 42(1), 79-80.
- [10] 陈平雁. SPSS 13.0 统计软件应用教程 [M]. 北京: 人民卫生出版社, 2005: 309-310.
- [11] 贺石林, 王健, 王净净. 中医科研设计与统计学 [M]. 长沙: 湖南科学技术出版社, 2011: 74-75.
- [12] 马斌荣. 医学统计学 [M]. 4 版. 北京: 人民卫生出版社, 2004: 147-148.
- [13] 王志国. 临床检验质量控制技术 [M]. 2 版. 北京: 人民卫生出版社, 2008: 308-309.
- [14] 张志清. 实验室 Z 比分数统计的辨析 [J]. 中国医疗器械信息, 2010, 16(1): 39-41.
- [15] 中国合格评定国家认可委员会. 能力验证结果的统计处理和能力评价指南 [S]. CNAS GL02, 2006.

(收稿日期: 2013-09-22)

(上接第 734 页)

- DP, et al. Apolipoprotein E genotype in matched men and women with coronary heart disease [J]. Ann Clin Lab Sci, 2005, 35(4): 391-396.
- [9] 马娟, 张鸿青, 王玮, 等. 载脂蛋白 B、载脂蛋白 E 及血管紧张素原基因多态性与冠心病的关系 [J]. 中国现代医生, 2011, 49(19): 4-6.
- [10] Wang CH, Zhou X. Meta analysis for relationship between ApoE gene polymorphism and coronary heart disease [J]. Zhonghua Yu

Fang Yi Xue Za Zhi, 2003, 37(5): 368-370.

- [11] Song Y, Stampfer MJ, Liu S, et al. Meta-analysis: Apolipoprotein E genotypes and risk for coronary heart disease [J]. Ann Intern Med, 2004, 141(2): 137-147.
- [12] Banares VG, Peterson G, Aguilar D, et al. Association between the ApoE4 allele and atherosclerosis is age dependent among Argentine [J]. Hum Biol, 2005, 77(2): 247-256.

(收稿日期: 2013-11-08)