

Bootstrap 方法建立生物参考区间的研究^{*}

李小佩, 王建新[△], 施秀英, 王惠民(南通大学附属医院检验科/南通大学公共卫生学院, 江苏南通 226000)

【摘要】 目的 探讨 Bootstrap 方法建立生物参考区间的可行性及最少样本量。**方法** 选取 120 例健康者检测血清尿素(BUN), 采用美国临床和实验室标准协会(CLSI) C28-A3 推荐参数、非参数方法及 Bootstrap 方法建立参考区间; 计算参考区间上下限值、均方根误差(RMSE)、精密度比值(I_d)等指标, 研究 Bootstrap 方法建立参考区间最少样本量。**结果** 非参数、参数、Bootstrap 百分位数、Bootstrap 标准方法建立参考区间分别为 3.00~7.39 mmol/L、3.05~7.33 mmol/L、3.12~7.18 mmol/L、3.10~7.33 mmol/L, 参考区间覆盖率大于 95%, Bootstrap 方法与 C28-A3 推荐方法结果一致; 进一步研究发现随着样本量减少, 参考区间增宽, RMSE 和 I_d 逐渐升高, 样本量减少到 60 时结果稳定。**结论** Bootstrap 方法可应用于各种数据分布类型生物参考区间的建立, 在小样本量时有一定优势, 为临床实验室制订参考区间提供了新思路。

【关键词】 Bootstrap 标准方法; Bootstrap 百分位数方法; 参考区间; 尿素

DOI: 10.3969/j.issn.1672-9455.2016.09.004 文献标志码: A 文章编号: 1672-9455(2016)09-1161-03

Estimation of reference intervals by bootstrap methods^{*} LI Xiao-pei, WANG Jian-xin[△], SHI Xiu-ying, WANG Hui-min (Department of Clinical Laboratory, Affiliated Hospital of Nantong University/School of Public Health, Nantong University, Nantong, Jiangsu 226000, China)

【Abstract】 Objective To research the bootstrap methods for reference intervals calculation are available and the minimum samples. **Methods** The concentrations of serum ureanitrogen (BUN) levels in 120 healthy reference subjects were determined and the reference intervals of BUN were calculated. Comparisons between results from parametric and nonparametric method which were recommended by Clinical and Laboratory Standards Institute (CLSI) C28-A3 and bootstrap methods. Reference interval limit and the values of RMSE, I_d were performed using RefValAdv to identify the minimum number of samples. **Results** The reference intervals calculated by nonparametric and parametric methods, percentile and standard bootstrap methods were 3.00–7.39 mmol/L, 3.05–7.33 mmol/L, 3.12–7.18 mmol/L, 3.10–7.33 mmol/L respectively, results were basically consistent. Overlap between reference intervals were greater than 95%. The values of RMSE and I_d increased, the reference intervals became wider with fewer samples. A minimum threshold of 60 could be considered a good limit to obtain reliable reference intervals. **Conclusion**

The bootstrap methods can be used to calculate reference intervals and provide new way of thinking for clinical laboratory, getting over the problem of the lack of sample size and the variable's probability distribution.

【Key words】 standard bootstrap method; percentile bootstrap method; reference intervals; BUN

参考区间是临床医生判断患者健康与否的标准, 是解释结果、分析检验信息、进行临床诊断的基本依据。国际标准化组织 15189 标准要求临床实验室对开展检验项目的生物参考区间进行定期评审。然而, 美国临床和实验室标准协会 (CLSI) C28-A3 推荐的参考方法需要较多样本 (大于 120 例), 实际操作耗时、费力。此外, 一些特殊项目 (脑脊液、精液检查等) 难以在较短时间内获得足够标本。如何科学、合理地制订参考区间是检验专家和统计学专家长期关注的问题。近年来, 国外学者将 Bootstrap 重抽样原理应用于参考区间的建立^[1-2], 其核心思想是对测量数据进行有放回地重复抽样, 解决了样本少、测量结果非正态分布等问题。目前, 国内未见相关报道, 本研究将 Bootstrap 方法应用于临床实验室尿素 (BUN) 参考区间的制订, 取得较好结果, 现报道如下。

1 资料与方法

1.1 一般资料 选取 2014 年 9 月, 南通大学附属医院健康体

检者 120 例, 男性 64 例, 女性 56 例; 年龄 18~60 岁。纳入标准: 60 岁以下成年男女 (女性为非孕期); 近 1 周内无大量蛋白 (肉食、海鲜) 摄入、无饮酒; B 超检查无肾结石、肾结核等; 尿、血生化检查肾功能指标正常; 无前列腺疾病、糖尿病、高血压; 心电图、X 线片、血常规及血生化检查无异常。

1.2 仪器与试剂 ADVIA2400 生化分析仪; BUN 试剂购自日本协和医药株式会社, 批号: 279ADD; 校准品为中生北控生物科技股份有限公司生产, 批号: 130321; 质控品购自伯乐生命医学产品 (上海) 有限公司, 批号: 15301。

1.3 方法 受检者安静状态下空腹采集静脉血 3 mL; 3 000 r/min 离心 10 min, 2 h 内分离血清, -70 °C 保存, 分析前将样本放置室温 30 min, 充分混匀, 上机检测。

1.4 统计学处理

1.4.1 离群值检测 应用 RefValAdv 软件中包含的 Dixon 方法进行离群值检验。

^{*} 基金项目: 国家质检总局公益性行业科研专项资助项目 (201210066); 南通市科技项目 (HS2012009)。

作者简介: 李小佩, 女, 检验技师, 硕士, 主要从事生物化学与临床酶学相关检验研究。 [△] 通讯作者, E-mail: wwbft@126.com。

1.4.2 正态性检验及正态性转化 绘制直方图并采用 RefValAdv 软件 Anderson-Darling (A-D) 检验判断数据的正态性, 正态性转化采用 Box-Cox 转化方法^[3]。

1.4.3 建立参考区间 原始数据选择非参数、Bootstrap 百分位数方法建立参考区间及其 90% 置信区间 (CI), 经 Box-Cox 方法转化的正态分布数据选择参数、Bootstrap 标准方法建立参考区间及其 90% CI^[4]。参数、非参数方法参照 CLSI C28-A3 文件。参考区间覆盖率指 2 种方法建立的参考区间最大值减去最小值的差值与重叠部分的比值^[5], 以此判断 2 种方法建立的参考区间是否存在差异。Bootstrap 标准方法是参数方法与有放回重抽样方法的结合, 该方法以计算机为基础从重抽样数据 $[x_i = (x_1, x_2 \cdots x_m)]$ 中有放回抽取 m 个样本 $[x_j = (x \times 1, x \times 2 \cdots x \times m)]$, $j = 1, 2 \cdots m$ (称为 Bootstrap 样本)], 重复此过程 B 次 ($B = 1\ 000$), 获得 B 个 Bootstrap 样本构成的经验分布, $\bar{x}_j$ 为 Bootstrap 样本的均值, $\bar{x}_{boot}$ 为经验分布的均值, 经验分布的标准误

$$s_{\bar{x}_{boot}} = \sqrt{\frac{\sum_{j=1}^B (\bar{x}_j - \bar{x}_{boot})^2}{B-1}}$$

经验分布的标准差: $s_{boot} = s_{\bar{x}_{boot}} \sqrt{n}$ 。

Bootstrap 标准方法计算参考区间 (RI): $RI_{boot} = \bar{x}_{boot} \pm 2s_{boot}$ 。

Bootstrap 百分位数方法是非参数方法与有放回重抽样方法的结合, 从重抽样数据 $[x_i = (x_1, x_2 \cdots x_m)]$ 中有放回抽取 m 个样本 $[x_j = (x \times 1, x \times 2 \cdots x \times m)]$, $j = 1, 2 \cdots m$ (称为 Bootstrap 样本)], 重复此过程 B 次 ($B = 1\ 000$), 获得 B 个 Bootstrap 样本构成的经验分布, $(\theta_{a/2}^*, \theta_{1-a/2}^*)$ 为 Bootstrap 样本百分位数, 1 000 个 Bootstrap 样本百分位数的均值即为经验分布 $100(1-a)\%$ 参考区间。Bootstrap 百分位数方法计算参考区间: $RI = [\sum_{i=1}^B \theta_{a/2}^*, \sum_{i=1}^B \theta_{1-a/2}^*]$ 。参数、非参数方法采用 RefValAdv 软件^[6], Bootstrap 方法采用 R 软件实现^[7]。

1.4.4 不同样本量参考区间上、下限值 从 120 例原始数据中进行无放回重抽样, 抽取样本量分别为 ($m = 110, 100, 90, 80, 70, 60, 50, 40$), 每组重抽样循环 p 次 ($p = 100$), 重抽样数据经转化采用 Bootstrap 标准方法, 原始数据采用 Bootstrap 百分位数方法建立 BUN 参考区间。如经无放回随机重抽样获得 $m = 70$ 的 100 个重抽样样本, 分别计算其上限及下限值, 采用 R 软件绘制上限及下限值箱式图。

1.4.5 参考区间上下限均方根误差 (RMSE) RMSE 用来评估 Bootstrap 方法的精密性^[8], 参考区间上下限均方根误差计算公式:

$$\text{参考区间下限 RMSE} = \{ \frac{1}{p} \sum_{i=1}^p [\hat{F}_{i(a/2)}^{-1} - F_{(a/2)}^{-1}]^2 \}^{1/2},$$

$$\text{参考区间上限 RMSE} = \{ \frac{1}{p} \sum_{i=1}^p [\hat{F}_{i(1-a/2)}^{-1} - F_{(1-a/2)}^{-1}]^2 \}^{1/2}.$$

$F_{a/2}^{-1}$ 为原始数据制订参考区间下限值, $\hat{F}_{i(a/2)}^{-1}$ 为重抽样数据制订参考区间下限值, $F_{(1-a/2)}^{-1}$ 为原始数据制订参考区间上限值, $\hat{F}_{i(1-a/2)}^{-1}$ 为重抽样数据制订参考区间上限值, 差异有统计学意义 ($P < 0.05$)。

1.4.6 不同样本量精密性比值 (I_d) I_d 表示不同样本量参考区间估计精密性比值^[9], I_d 值为重抽样数据标准误 $s_{\bar{x}_{boot}}^m$ 均值与原始数据标准误的比值, I_d 值计算公式: $s_{\bar{x}_{boot}}^m = \frac{1}{p} \sum_{i=1}^p s_{\bar{x}_{boot}}^m$, $I_d =$

$$\frac{s_{\bar{x}_{boot}}^m}{s_{120 \bar{x}_{boot}}}$$

2 结 果

2.1 离群值检测 原始数据及 Box-Cox 转化数据均未检测到离群值。

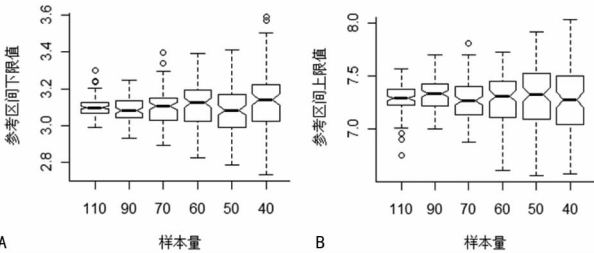
2.2 正态性检验及正态性转化 BUN 原始数据经 A-D 检验 $A = 0.930$, 不服从正态分布 ($P < 0.05$)。转化数据 $A = 0.344$, 服从正态分布 ($P > 0.05$)。转化数据经 Bootstrap 有放回重抽样 Bootstrap 样本均值 $\bar{x}_j$ 经 A-D 检验 $A = 0.245$, 服从正态分布 ($P > 0.05$)。

2.3 尿素参考区间的建立 BUN 原始数据为非正态分布, 采用非参数方法及 Bootstrap 百分位数方法建立参考区间及其 90% CI, 转化数据服从正态分布, 采用参数方法及 Bootstrap 标准方法建立参考区间及其 90% CI。样本量为 120 例时 4 种方法建立 BUN 参考区间基本一致, Bootstrap 百分位数方法区间最窄, Bootstrap 标准方法次之, 任意 2 种方法参考区间覆盖率均大于 95%, 表明 Bootstrap 方法可用于参考区间制订。见表 1。

表 1 BUN 参考区间 (mmol/L)

组别	下限	上限	下限 (90%CI)	上限 (90%CI)
非参数 (原始数据)	3.00	7.39	2.80~3.40	6.80~7.90
Bootstrap 百分位数 (原始数据)	3.12	7.18	2.81~3.44	6.61~7.76
参数	3.05	7.33	2.90~3.24	6.94~7.73
Bootstrap 标准 (Box-Cox 转化数据)	3.10	7.33	2.87~3.22	6.95~7.75

2.4 不同样本量参考区间上、下限值 样本量从 110 例减少至 60 例, Bootstrap 百分位数方法与 Bootstrap 标准方法建立 BUN 参考区间上限值与下限值覆盖范围增宽, 参考区间增宽, 样本量从 60 例减少至 40 例时上限值与下限值覆盖范围明显增宽。箱式图凹槽表示中位数的置信区间, 不同样本量 Bootstrap 标准方法中位数的置信区间均有重叠, Bootstrap 百分位数方法部分重叠, 表明 Bootstrap 标准方法比 Bootstrap 百分位数方法具有更好的稳健性。见图 1、2。

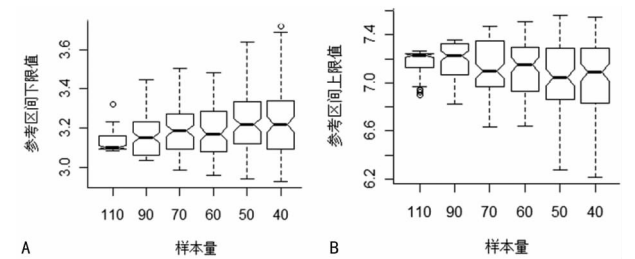


注: A 为 Bootstrap 标准方法下限值箱式图; B 为 Bootstrap 标准方法上限值箱式图。

图 1 Bootstrap 标准方法上、下限值箱式图

2.5 参考区间下、上限均方根误差 样本量从 110 例减少至 60 例, 2 种 Bootstrap 方法 RMSE 均增加, 精密性下降, Bootstrap 标准方法小于 Bootstrap 百分位数方法, 表明具有更好的稳健性, 继续减少样本量, RMSE 增加明显, 精密性变差。

2.6 不同样本量精密性比值 I_d 值量化样本量与参考区间估计精密性之间的关系, I_d 值临界值为 1.50, I_d 值大于 1.50 表明精密性较差。样本量从 110 例减少至 60 例, 2 种 Bootstrap 方法 I_d 值小于 1.50, 继续减少样本量, I_d 值大于 1.50。表明该法制订参考区间的最少样本量为 60 例。见表 2。



注:A 为 Bootstrap 百分位数方法下限值箱式图;B 为 Bootstrap 百分位数方法上限值箱式图。

图 2 Bootstrap 百分位数方法上、下限值箱式图

表 2 不同方法、样本量的 BUN I_d 值

样本量(n)	Bootstrap 标准方法 I _d	Bootstrap 百分位数方法 I _d
110	1.03	1.09
90	1.14	1.18
70	1.31	1.22
60	1.41	1.44
50	1.53	1.58
40	1.74	1.83

3 讨 论

国际临床化学和检验医学联合会(IFCC)建议临床实验室制订各自的参考区间。CLSI C28-A3 文件指出建立参考区间至少要 120 例健康个体。Linnet^[10]认为偏态数据需要更多样本,极度偏态数据至少需要 700 例。选择健康个体是 1 项费时、耗力的工作,临床实验室出于工作效率及成本考虑,很少能够按照文件建议建立参考区间,多使用《全国检验操作规程》、试剂说明书、仪器厂家提供的参考区间^[11],或对已有参考区间进行验证^[12]。在实际工作中,试剂厂家等提供的参考区间有时与临床诊断存在较大差异,出现参考区间验证无效的情况。因此,制订实验室合适的参考区间尤为重要。

参考区间的制订方法与数据分布类型、样本量有关。本研究表明,大样本量时上述 4 种方法制订参考区间结果一致,说明 Bootstrap 方法可用于参考区间建立。进一步研究发现 BUN 数据转化为正态分布,选用基于正态理论的参数方法与 Bootstrap 标准方法,后者区间较窄^[8]。试剂厂商提供的参考区间为 2.86~7.14 mmol/L, Coskun 等^[4]采用非参数方法建立参考区间为 2.43~6.83 mmol/L,参数方法为 2.75~7.10 mmol/L。本研究结果与厂商、国外学者报道存在一定差异,其原因可能与地区、人群不同有关。因此,直接引用厂商提供的参考区间可能影响临床诊疗工作。

美国兽医临床病例学会(ASVCP)指出,制订参考区间的健康个体小于 120 例时,Bootstrap 方法可用于参考区间上、下限 CI 的建立。Coskun 等^[4]研究表明样本量少时 Bootstrap 方法制订参考区间上、下限不确定度略微增加,认为该方法比 CLSI 推荐方法更为可靠。本研究结果随着样本量减少,参考区间增宽、RMSE 和 I_d 逐渐升高,当样本量为 60 例时出现明显改变。因此,建立可靠参考区间最少样本量为 60 例。本研究结果还表明 Bootstrap 标准方法比 Bootstrap 百分位数方法

更具稳健性,提示基于正态理论的统计方法效能较高^[10]。
总之,Bootstrap 方法可用来建立参考区间,灵活应用于各种数据分布类型,仅需传统方法一半的样本量,提高了实验室工作效率,值得临床实验室进行尝试。

参考文献

[1] Pavlov IY, Wilson AR, Delgado JC. Resampling approach for determination of the method for reference interval calculation in clinical laboratory practice [J]. Clinical and Vaccine Immunology, 2010, 17(8): 1217-1222.

[2] Pavlov IY, Wilson AR, Delgado JC. Reference interval computation; which method (not) to choose [J]. Clinica Chimica Acta, 2012, 413(13/14): 1107-1114.

[3] Horn PS, Feng L, Li Y, et al. Effect of outliers and non-healthy individuals on reference interval estimation [J]. Clinical Chemistry, 2001, 47(12): 2137-2145.

[4] Coskun A, Ceyhan E, Inal TC, et al. The comparison of parametric and nonparametric bootstrap methods for reference interval computation in small sample size groups [J]. Accreditation and Quality Assurance, 2013, 18(1): 51-60.

[5] Vojta A, Shek-Vugrovecki A, Radin L, et al. Hematological and biochemical reference intervals in Dalmatian pramenka sheep estimated from reduced sample size by bootstrap resampling [J]. Veterinarski Arhiv, 2011, 81(1): 25-33.

[6] Geffre A, Concordet D, Braun JP, et al. Reference value advisor: a new freeware set of macroinstructions to calculate reference intervals with Microsoft Excel [J]. Veterinary Clinical Pathology, 2011, 40(1): 107-112.

[7] R Development Core Team. R: a language and environment for statistical computing [EB/OL]. [2004-10-19]. <http://www.r-project.org>.

[8] Horn PS, Pesce AJ, Copeland BE, et al. A robust approach to reference interval estimation and evaluation [J]. Clinical Chemistry, 1998, 44(3): 622-631.

[9] Dimauroa C, Macciotta NPP, Rassu SPG, et al. A bootstrap approach to estimate reference intervals of biochemical variables in sheep using reduced sample sizes [J]. Small Ruminant Research, 2009, 83(1/2/3): 34-41.

[10] Linnet K. Two-stage transformation systems for normalization of reference distributions evaluated [J]. Clinical Chemistry, 1987, 33(3): 381-386.

[11] 钟堃, 王薇, 王治国, 等. 全国干化学室间质量评价项目参考区间现况调查 [J]. 检验医学, 2012, 27(8): 684-687.

[12] 邹麟, 张莉萍, 徐华健. Vitros 5.1 FS 自动生化分析仪生物参考区间验证 [J]. 重庆医学, 2013, 42(3): 332-333.