

层次结构定量数据组内相关系数的非参数 Bootstrap 估计

梁融 周舒冬 李丽霞 张俊国 郜艳晖

【导读】 探讨 Bootstrap 自抽样如何在层次结构数据中实现,为组内相关系数(ICC)可信区间的计算方法提供选择。文中利用混合效应模型估计重复测量数据和两阶段抽样数据的 ICC 及利用 Bootstrap 法估计 ICC 的可信区间,比较不同的自抽样模式下 ICC 可信区间结果。重复测量实例结果显示 Bootstrap 整群抽样估计的可信区间包含 ICC 真值,如忽视数据的层次结构特征,Bootstrap 方法得到无效的可信区间估计;两阶段抽样实例结果显示整群 Bootstrap 自抽样方法估计的 ICC 均数与原样本 ICC 偏差最小,可信区间宽泛。表明对层次结构数据进行 Bootstrap 自抽样,需考虑数据的产生机制,即高水平 Bootstrap 自抽样的统计量估计更接近原样本统计量。

【关键词】 层次结构数据;组内相关系数;可信区间;Bootstrap 自抽样

Non-parametric Bootstrap estimation on the intraclass correlation coefficient generated from quantitative hierarchical data LIANG Rong, ZHOU Shu-dong, LI Li-xia, ZHANG Jun-guo, GAO Yan-hui. Department of Epidemiology and Biostatistics, Guangdong Pharmaceutical University, Guangzhou 510310, China

Corresponding author: GAO Yan-hui, Email: gao_yanhui@163.com

This work was supported by a grant from the Natural Science Foundation of Guangdong Province (No. 10151022401000018).

【Introduction】 This paper aims to achieve Bootstrapping in hierarchical data and to provide a method for the estimation on confidence interval (CI) of intraclass correlation coefficient (ICC). First, we utilize the mixed-effects model to estimate data from ICC of repeated measurement and from the two-stage sampling. Then, we use Bootstrap method to estimate CI from related ICCs. Finally, the influences of different Bootstrapping strategies to ICC's CIs are compared. The repeated measurement instance show that the CI of cluster Bootstrapping containing the true ICC value. However, when ignoring the hierarchy characteristics of data, the random Bootstrapping method shows that it has the invalid CI. Result from the two-stage instance shows that bias observed between cluster Bootstrapping's ICC means while the ICC of the original sample is the smallest, but with wide CI. It is necessary to consider the structure of data as important, when hierarchical data is being resampled. Bootstrapping seems to be better on the higher than that on lower levels.

【Key words】 Hierarchical data; Intraclass correlation coefficient; Confidence interval; Bootstrapping

层次结构数据(hierarchical data)是医学研究领域一种常见数据形式,如对同一个体某变量进行重复测量,或流行病学研究中采用多阶段抽样调查,这时研究变量可能存在不独立的特征。如研究变量为连续型定量变量时,类内相关大小可用组内相关系数(intraclass correlation coefficient, ICC)描述,如重复测量数据中,ICC 值越大,表示个体内观测值间

相关性越强。单变量分析时,ICC 可通过估计类(个体)内方差和类(个体)间方差利用公式计算^[1];多变量背景时,对层次结构数据可采用混合效应模型(mixed effect model)考察影响因素对响应变量的作用,此时影响因素作为固定效应,类别效应作为随机效应,则 ICC 可由随机效应方差占随机效应方差与误差方差之和的比例来估计。目前常规统计软件如 SAS、SPSS 均可实现混合效应模型,估计随机效应方差和随机误差方差,由此研究者可自行计算 ICC 值,但是 ICC 的可信区间(CI)仍无法估计。

统计量分布不明确时,可应用 Bootstrap 方法估计其 CI。Bootstrap 方法通过对自身样本有放回重抽

DOI: 10.3760/cma.j.issn.0254-6450.2013.09.016

基金项目:广东省自然科学基金(10151022401000018)

作者单位:510310 广州,广东省分子流行病学重点实验室广东药学院公共卫生学院流行病与卫生统计学系

通信作者:郜艳晖, Email: gao_yanhui@163.com

样从而得到一系列自助样本,对每个自助样本估计统计量,以模拟原统计量的经验分布。由于该法对原统计量分布并无要求,因此自 Efron^[2]提出后已有广泛应用。当原始样本为独立数据时,自助样本可视为源自原始样本的随机样本,但当该方法应用于多阶段抽样或重复测量数据时,对样本进行有放回重抽样时应考虑数据自身固有的层次结构特征。本文探讨在层次结构数据中如何应用 Bootstrap 自抽样,并以 SAS 软件为例,说明利用混合效应模型估计 ICC 及利用 Bootstrap 法估计其 CI 的方法,比较不同自抽样模式对估计 ICC 及其 CI 的影响和提供选择。

基本原理

1. 混合效应模型和 ICC 的估计:以两水平的层次结构数据为例,可用混合效应模型拟合,即 $y = X\beta + Z\gamma + \varepsilon$,其中 y 为连续型响应变量, $X\beta$ 为固定效应项, $Z\gamma$ 为随机效应项,随机效应 $\gamma \sim N(0, \sigma_\gamma^2)$, ε 为随机误差且 $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ 。ICC 的计算公式为 $ICC = \sigma_\gamma^2 / (\sigma_\gamma^2 + \sigma_\varepsilon^2)$,表示高水平观测值之间的相关程度。在 SAS 软件中 Proc Mixed 可用于拟合混合效应模型,用 Random 语句指定随机效应,结果中输出随机效应方差及随机误差方差,并利用上述公式计算 ICC 值^[3]。

2. 估计 Bootstrap 的 CI:对独立数据,采用 Bootstrap 法计算 CI,即对原样本有放回自抽样,共自抽样 B 次,则得到 B 个自助样本。每个自助样本的大小 (N) 可以等于原样本量,也可以是原样本量的一定比例。对每个自助样本估计统计量,得到 B 个统计量,从而构造出待估统计量的经验分布^[4]。若经验分布服从正态分布则采用 t 分布法 $\bar{x} \pm t_{\alpha/2, v} \cdot s_x$ 估计 CI;若不满足正态性,则采用百分位数法计算;若考虑自助样本统计量与原样本统计量存在偏差,可能对 CI 的估计有影响,则根据偏差对百分位数法进行校正^[4,5]。此种 Bootstrap 抽样方法不依赖总体分布,也称为非参数 Bootstrap 方法。目前 SAS 软件中可用 Proc Surveyselect 实现以独立观测为抽样单位的 Bootstrap 自抽样。

3. 重复测量数据特征及 Bootstrap 抽样方法:对重复测量资料,设每个个体包含 t 个具有时间先后顺序的测量值,因此采用 Bootstrap 自抽样时不能以每个测量值为抽样单位,自助样本中个体内观测值的时间次序结构必须完整,即 Bootstrap 自抽样应以个体为抽样单位(本文称为 Bootstrap 整群抽样),抽中的个体内 t 次观

测值顺序应保持不变。实际应用时,在 SAS 软件中可调用 Proc Surveyselect 结合 Proc Mixed 实现 ICC 的 Bootstrap CI 估计。

需要注意的是, SAS 软件中用 Proc Mixed 实现混合效应模型要求的数据格式应以每个个体每次测量为一条观测(observation),即统计分析单位(图 1A, id 为 1 的个体有 3 次测量,以每次测量为分析单位,因此个体 1 有 3 行),而以个体为单位自抽样时, Proc Surveyselect 要求的数据格式应以每个个体为一条观测(图 1 B,个体 1 的 3 次测量在同一行),因此运用 SAS 软件对原样本和自助样本进行混合效应模型分析时,以及对原样本自抽样时,需将数据集格式在图 1A 和 B 间进行转换。

4. 多阶段抽样数据特征及 Bootstrap 抽样方法:以两阶段抽样数据为例,由于此类数据的层次结构并不像重复测量数据中具有不能随机变动的顺序,因此 Bootstrap 自抽样形式灵活多样。例如 Bootstrap 自抽样时可以高水平的层为抽样单位(本文称 Bootstrap 整群抽样),抽到的层内所有观测保持不变;另一种策略是固定高水平的层,以层内观测为抽样单位进行有放回重抽样(本文称 Bootstrap 分层抽样),这时自助样本中将包含所有的层,而层内观测为重复抽样结果。如不考虑层次特征,也可按独立数据方式直接以各观测为抽样单位。在 SAS 软件中,前一种自抽样方式类同重复测量数据,应用时需注意 Proc Surveyselect 与 Proc Mixed 对数据格式的要求(图 1);第二种自抽样方式可采用 Proc Surveyselect 的 Strata 语句定义高水平层次,使得有放回自抽样发生在各层内部。

实例分析

1. 重复测量数据:为分析语言教学方式对学生评判性思维(critical thinking skills, CTS)的影响,分别由 2 名教师各采用两种不同的教学方式对 4 个班学生授课。教学开始前进行语言水平测验,将学生

A: Proc Mixed 要求的数据结构				B: Proc Surveyselect 要求的数据结构						
id	Y	X ₁	X ₂	id	Y ₁	Y ₂	Y ₃	X ₁	X ₂	
1	a ₁	1	H	↔	1	a ₁	a ₂	a ₃	1	H
1	a ₂	1	H		2	b ₁	b ₂	b ₃	2	L
1	a ₃	1	H	
2	b ₁	2	L							
2	b ₂	2	L							
2	b ₃	2	L							
...							

注: id 为个体编号; Y 为个体重复测量的响应变量; X₁ 与 X₂ 为解释变量

图 1 重复测量数据 ICC 的 Bootstrap CI 估计过程中数据格式的转换

的语言水平分成高、中和低 3 个层次,并对学生分别进行 3 次 CTS 测试。

本例中 3 次 CTS 得分为重复测量的响应变量。采用混合效应模型分析教学方式、教师和学生教学前语言水平对响应变量的影响,估计随机效应方差为 2.7767,随机误差方差为 0.7761,计算 3 次 CTS 得分的 $ICC=0.7815$ 。为估计 ICC 的 CI,采用以个体为抽样单位的 Bootstrap 自抽样模式,得到 500 个自助样本的 ICC 值,由于不服从正态分布 ($P<0.0005$),故选择百分位数法和偏差校正百分位数法计算 ICC 的 95%CI (表 1)。自助样本 ICC 的均数 (m_ICC) 与原样本 ICC 的差值为 0.0335,偏差校正的 95%CI 较百分位数法窄。表 1 也列出忽略数据层次结构特征时完全随机抽样的结果,可见 m_ICC 的 95%CI 不包含原样本的 ICC 值,Bootstrap 法估计 CI 失效。

2. 两阶段抽样数据:采用一个两阶段抽样的模拟数据,包括 20 层,每层观测数为 10,共 200 个观测。定义两个连续型解释变量 X_1 和 X_2 ,均来自 $[0,1]$ 的均匀分布;层内的随机效应 $\gamma \sim N(0, 0.7^2)$,个体的随机误差 $\varepsilon \sim N(0, 0.2^2)$;连续型因变量定义为 $Y=-2+2X_1+3X_2+\gamma+\varepsilon$ 。先运用混合效应模型估计模拟数据的随机效应方差为 0.3597,误差方差为 0.0461,计算样本的 $ICC=0.8863$ 。

考虑数据的层次结构特征,分别采用 Bootstrap 整群抽样和 Bootstrap 分层抽样估计 ICC 的 CI (表 2)。结果显示,Bootstrap 整群抽样的 m_ICC 与原样本 ICC 偏差最小 (-0.0061),但 CI 范围较宽。如不考虑数据层次结构,Bootstrap 完全随机抽样则与 Bootstrap 分层抽样的结果相似。

讨 论

医学及公共卫生研究领域常见具有层次结构的数据,如整群随机对照临床试验、多阶段抽样、重复测量数据等,在统计分析时需要考虑这类数据的层次结构,估计和报告组内相关系数^[6,7]。而组内相关系数的 CI,可用 Bootstrap 法估计。本文采用非参数 Bootstrap 法是一种有放回自抽样、而不依赖总体分布计算统计量 CI 的方法,已广泛用于未知分布统计量的 CI 估计^[2,5,8,9]。但用于层次结构数据自抽样时,也应考虑数据的结构特征^[5-7,9-12]。

本文主要针对重复测量数据和两阶段抽样数据的层次结构特征,探讨不同 Bootstrap

自抽样方法估计 ICC 及其 CI 的策略。尽管两种数据均可看作高水平间独立和高水平内相关,并均可采用混合效应模型估计 ICC,但从数据产生的机制上两者有区别,因而在选择 Bootstrap 自抽样方式时应区别对待。重复测量数据中个体内的多次测量具有时间次序,而两阶段抽样数据中层内观测值则无次序特征,因此对重复测量数据进行 Bootstrap 自抽样时只能使用整群抽样方式,不能将个体内的多次测量值割裂开;而两阶段抽样数据层内观测无次序结构,因此均可采用整群和分层自抽样方式。本文重复测量实例显示,Bootstrap 自抽样时如忽视层次结构特征,得到无效的 CI 估计;两阶段抽样实例显示,几种方法均得到有效的 CI 估计。除本文采用的非参数 Bootstrap 外,还有参数 Bootstrap 法可用于再抽样。参数 Bootstrap 法假定原始数据来自某理论分布,通常将原始数据进行参数估计得到的统计量作为该理论分布的总体参数,再从理论分布中抽样得到一系列 Bootstrap 子样本。本文多阶段抽样数据,也可采用参数 Bootstrap 法,但用于重复测量数据时,在原始数据的分析模型中需考虑时间变量对结局的影响,否则同一个体各时间点上多次测量值的顺序难以参数化。

对多阶段抽样数据,本文探讨了 3 种非参数 Bootstrap 抽样方式,此外也可采用有放回自抽样与无放回随机抽样相结合的方式抽取自助样本^[10]。如 Ren 等^[12]在分析三层次结构数据时,分别采用从不同层次(第二层、第三层或第四层)有放回随机抽样,而抽到的层次内部无放回随机抽样(共做两次抽

表 1 不同层次特征 Bootstrap 法估计重复测量数据 ICC 的 95%CI

Bootstrap 抽样方式	ICC	m_ICC	偏差	95%CI	CI 宽度
整群抽样					
百分位数法	0.7815	0.8151	0.0335	0.7809 ~ 0.8421	0.0612
偏差校正百分位数法	0.7815	0.8151	0.0335	0.7662 ~ 0.7827	0.0165
完全随机抽样*	0.7815	0.8557	0.0742	0.8320 ~ 0.8766	0.0446

注:*百分位数法与偏差校正百分位数法结果相同

表 2 不同 Bootstrap 抽样方式估计两阶段抽样数据 ICC 的 95%CI

Bootstrap 抽样方式	ICC	m_ICC	偏差	95%CI	CI 宽度
整群抽样					
百分位数法	0.8863	0.8802	-0.0061	0.7826 ~ 0.9289	0.1463
偏差校正百分位数法	0.8863	0.8802	-0.0061	0.7760 ~ 0.9260	0.1500
分层抽样					
百分位数法	0.8863	0.8982	0.0119	0.8818 ~ 0.9155	0.0337
偏差校正百分位数法	0.8863	0.8982	0.0119	0.8704 ~ 0.8912	0.0208
完全随机抽样					
百分位数法	0.8863	0.8982	0.0119	0.8811 ~ 0.9150	0.0339
偏差校正百分位数法	0.8863	0.8982	0.0119	0.8670 ~ 0.8915	0.0245

样,虽然数据具有多于两层次的结构特征)的策略,比较从不同层次自抽样策略下自助样本统计量与原样本统计量的接近程度。结果显示,从最高层次自抽样样本的统计量最接近原样本统计量,从而认为高水平自抽样结果更优。本文只模拟了二水平结构数据,结果显示其有放回自抽样的 Bootstrap 样本统计量与原样本统计量偏差最小,但统计量的 95%CI 范围较宽,表明整群抽样误差较大。但在二阶段抽样数据的实例中,如忽略层次结构特征,也可得到 Bootstrap 的 95%CI,与低水平 Bootstrap 自抽样结果接近。由于本文的模拟数据仅为一次抽样的结果,更多系统模拟结果及其理论分析还有待进一步研究。此外,本文主要针对定量结局变量,当结局为定性变量时,应采用非线性混合效应模型计算 ICC^[11,12],但计算 CI 应与 Bootstrap 方法类似。

参 考 文 献

[1] Fang JQ. Medical Statistics and Computer Experiments with CD-ROM. 3rd ed. Shanghai: Shanghai Science and Technology Publishing House, 2006:430-442. (in Chinese)
方积乾. 医学统计学与电脑实验(附光盘). 3 版. 上海:上海科学技术出版社,2006:430-442.

[2] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat, 1979, 7(1):1-26.

[3] SAS Institute Inc. 2009. SAS/STAT® 9.2 User's Guide, 2nd ed. Cary, NC: SAS Institute Inc [EB/OL].

[4] Johnson RW. An introduction to the bootstrap. Teach Stat, 2001,

23(2):49-54.

[5] Ukoumunne OC, Davison AC, Gulliford MC, et al. Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. Stat Med, 2003, 22(24):3805-3821.

[6] Ip EH, Wasserman R, Barkin S. Comparison of intraclass correlation coefficient estimates and standard errors between using cross-sectional and repeated measurement data: the safety check cluster randomized trial. Contemp Clin Trials, 2011, 32 (2) : 225-232.

[7] Thompson DM, Fernald DH, Mold JW. Intraclass correlation coefficients typical of cluster-randomized studies: estimates from the Robert Wood Johnson Prescription for Health projects. Ann Fam Med, 2012, 10(3):235-240.

[8] Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci, 1986, 1(1):54-75.

[9] Xiao YH, Liu JW, Bhandary M. Resampling approaches for common intraclass correlation coefficients. J Stat Comput Simul, 2012, 82(9):1357-1366.

[10] Field CA, Welsh AH. Bootstrapping clustered data. J R Stat Soc Series B Stat Methodol, 2007, 69(3):369-390.

[11] Ren SQ, Yang SQ, Lai SH. Intraclass correlation coefficients and bootstrap methods of hierarchical binary outcomes. Stat Med, 2006, 25(20):3576-3588.

[12] Ren SQ, Lai H, Tong WJ, et al. Nonparametric bootstrapping for hierarchical data. J Appl Stat, 2010, 37(9):1487-1498.

(收稿日期:2013-05-13)

(本文编辑:张林东)

中华流行病学杂志第六届编辑委员会通讯编委名单

- | | | |
|---------------------|-------------------|------------------------|
| 陈 曦(湖南省疾病预防控制中心) | 窦丰满(成都市疾病预防控制中心) | 高 婷(北京市疾病预防控制中心) |
| 姜宝法(山东大学公共卫生学院) | 李 杰(北京大学医学部) | 李十月(武汉大学公共卫生学院) |
| 李秀央(浙江大学医学院公共卫生学院) | 廖苏苏(中国医学科学院基础医学院) | 林 玫(广西壮族自治区疾病预防控制中心) |
| 林 鹏(广东省疾病预防控制中心) | 刘爱忠(中南大学公共卫生学院) | 刘 刚(四川省疾病预防控制中心) |
| 刘 静(北京安贞医院) | 刘 莉(四川省疾病预防控制中心) | 刘 玮(军事医学科学院微生物流行病学研究所) |
| 鲁凤氏(北京大学医学部) | 欧剑鸣(福建省疾病预防控制中心) | 彭晓旻(北京市疾病预防控制中心) |
| 邱洪斌(佳木斯大学) | 赛晓勇(解放军总医院) | 苏 虹(安徽医科大学公共卫生学院) |
| 汤 哲(首都医科大学附属宣武医院) | 田庆宝(河北医科大学公共卫生学院) | 王 蓓(东南大学公共卫生学院) |
| 王素萍(山西医科大学公共卫生学院) | 王志萍(山东大学公共卫生学院) | 谢 娟(天津医科大学公共卫生学院) |
| 徐爱强(山东省疾病预防控制中心) | 徐慧芳(广州市疾病预防控制中心) | 严卫丽(新疆医科大学公共卫生学院) |
| 阎丽静(中国乔治中心) | 杨春霞(四川大学华西公共卫生学院) | 余运贤(浙江大学医学院公共卫生学院) |
| 曾哲淳(北京安贞医院) | 张 波(宁夏回族自治区卫生厅) | 张宏伟(第二军医大学) |
| 张茂俊(中国疾病预防控制中心传染病所) | 张卫东(郑州大学公共卫生学院) | 赵亚双(哈尔滨医科大学公共卫生学院) |
| 朱 谦(河南省疾病预防控制中心) | 祖荣强(江苏省疾病预防控制中心) | |