DOI:10.16644/j.cnki.cn33-1094/tp.2022.11.017

# 基于高校四级成绩数据的重采样方法研究*

夏 艳，张丽娟

(广州华商学院，广东 广州 511399)

**摘 要：**为了更好的评估高校各专业英语四级总体应试水平，本文突破传统估计均值的方法，在Bootstrap和Jackknife的基础上，从新的角度进行分析研究；采用Bootstrap-Jackknife估计方法对广州华商学院2017学年各专业四级成绩均值数据进行估计。从估计误差来看，Bootstrap-Jackknife估计误差要远远小于Bootstrap与Jackknife，这说明了Bootstrap-Jackknife在对估计总体样本均值问题上的优越性。

**关键词：**Bootstrap；Jackknife；Bootstrap-Jackknife估计；标准差估计；四级成绩

**中图分类号：**O212.7 **文献标识码：**A **文章编号：**1006-8228(2022)11-73-03

## Research on resampling method based on the score data of CET4

Xia Yan, Zhang Lijuan

(*Guangzhou Huashang College, Guangzhou, Guangdong 511399, China*)

**Abstract:** Breaking through the traditional method of estimating the mean value, on the basis of Bootstrap and Jackknife, the analysis and research are carried out from a new perspective to better evaluate the overall level of CET4. The Bootstrap-Jackknife estimation method is used to estimate the mean score of the CET4 of each major in Guangzhou Huashang University in the 2017 academic year. From the estimation error, Bootstrap-Jackknife method is much smaller than Bootstrap and Jackknife, which shows the superiority of Bootstrap-Jackknife in estimating the overall sample mean.

**Key words:** Bootstrap; Jackknife; Bootstrap-Jackknife estimate; standard deviation estimate; score of CET4

## 0 引言

大学英语等级考试一直以来都是高校学生参与度最高的全国性考试，其考试成绩不仅反映了学生的英语学习能力，其证书也是高校毕业生求职应聘中所必备的。

评估高校各专业整体英语应试水平，对于高校管理层在专业层面上提出相关英语教学改革措施极为重要。Bootstrap与Jackknife是抽样调查中常用的重采样方法，Jackknife是由Quenouille[1,2](1949/1956)作为减少系列相关系数估计量偏倚的一种方法提出的，后来逐渐成为复杂样本方差估计的一种重要方法。Bootstrap是由B.Efron[3](1979)在Jackknife的基础上提出的一种利用重抽样方法对总体参数进行估计的统计方法。吕萍[4](2017)指出在数据分析中，若忽视层、群等抽样设计的复杂性，直接利用调查数据按照传统数据分析方法，容易得出错误的结论，尤其是涉及标准误的估计。Bootstrap方法的优势在于对小样本进行评估时，可极大地降低评估样本不足对评估结果的影响[5]。该方法也在估计中存在些许不足，主要体现在重抽样都是在已知的样本观测数据中进行的，这使得自主样本与原样本的相似度较高，并且样本量越小，其相似度就越高，估计结果与真实分布的差异性也会越大[6]。Jackknife方法在方差分量估计和标准误估计上都较为准确，且其估计的准确性不随数据类型、研

**收稿日期：**2022-04-27

***基金项目：**2021年校级质量工程-数据科学与大数据技术(HS2021ZLGC02)

**作者简介：**夏艳(1996-)，女，湖北江陵县人，硕士，助教，主要研究方向：数据挖掘与大数据分析。

**通讯作者：**张丽娟(1994-)，女，湖南攸县人，硕士，助教，主要研究方向：数据挖掘与大数据分析。

究设计和方差分量的不同而产生波动，具有较强的稳健性[7]。Jackknife方法不足之处主要体现在：估计总体统计量时只利用了很少的信息，各采样样本之间的差异很小，每两个Jackknife样本中只有两个单一的观测值不同。本文在估计总体样本均值的过程中，考虑到Jackknife算法与Bootstrap算法存在的不足，提出Bootstrap–Jackknife算法，得到了更接近于总体样本均值的估计值。

## 1 数据与估计方法

### 1.1 数据来源与处理

本文采集广州华商学院各专业学生在2017学年的四级成绩数据，共计9860条有效数据，并对收集的数据进行对数化处理，数据对数化可以使得样本数据更加光滑，消除异方差,同时减小数据波动范围。

### 1.2 Normal

将采集得到的观测样本$x_1,\cdots,x_n$当做总体样本的近似，通过观测样本得到各样本统计量值以估计总体统计量，其中总体标准差的无偏估计如式(1)：

$$\widehat{se}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2},\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{1}$$

### 1.3 Bootstrap

Bootstrap是一种著名的方差估计方法，其思想是通过重复抽样来估计总体分布。具体来说就是将得到的样本$F_n(x)$当做总体$F(x)$的近似，$\hat{\theta}$是$\theta$的一个估计，通过从得到的样本中重复有放回抽样生成经验累积分布函数$F_n^*(x)$，对生成的$F_n^*(x)$样本进行相应计算得到$\hat{\theta}^*$，利用一系列$\hat{\theta}^*$实现$\hat{\theta}$的置信区间评定。具体步骤如下：

(1) 从观测样本$x_1,\cdots,x_n$中有放回地抽样生成样本$x^{*(b)}=(x_1^*,\cdots,x_n^*)$；

(2) 对第$b$个Bootstrap样本计算估计值$\hat{\theta}^{(b)}$，这里$b$的范围为1–2000，本文为了使全部的数据尽可能被采集，使得总体统计量的估计结果更为稳健，规定抽样次数$B=2000$；

(3) 对一个估计量$\hat{\theta}$的标准差进行Bootstrap估计就是将Bootstrap重复实验$\hat{\theta}^{(1)},\cdots,\hat{\theta}^{(B)}$的样本标准差作为估计值，如式(2)：

$$\widehat{se}\ (\hat{\theta}^*)=\sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{(b)}-\overline{\theta^{(*)}})^2},\overline{\hat{\theta}^{(*)}}=\frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}^{(b)}) \tag{2}$$

### 1.4 Jackknife

Jackknife可用于总体估计量的不确定估计，旨在减少估计的偏差。其思想为“去一”抽样，假设获取样本样本量为$n$，在第$i$次抽样中去除第$i$个样本数据$i=(1,2,...,n)$，用剩下的$(n-1)$个数据作为抽样样本计算$\hat{\theta}_{(i)}$，分别对生成的$n$个样本计算相应的样本统计量，如此得到$\hat{\theta}_{(1)},\cdots,\hat{\theta}_{(b)}$，从而实现总体统计量的置信区间估计。具体步骤如下：

(1) 从观测样本$x_1,\cdots,x_n$中做$i$次Jackknife抽样，生成第$i$个Jackknife样本$(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n)$；

(2) 对$n$个Jackknife样本计算估计值$\hat{\theta}^{(1)},\hat{\theta}^{(2)},\cdots,\hat{\theta}^{(n)}$；

(3) 当利用Jackknife对$\theta$进行标准差估计时，如式(3)：

$$\widehat{se}_{jack-knife}=\sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)}-\overline{\hat{\theta}_{(\cdot)}})^2} \tag{3}$$

其中$\frac{n-1}{n}$因子的原因是：

当$\hat{\theta}=\bar{x}$时，$\mathrm{var}(\bar{x})=\sqrt{\mathrm{var}(x)/n}$，因此$\frac{n-1}{n}$因子使得$\widehat{se}_{jack-knife}$成为无偏估计量。

### 1.5 Bootstrap–Jackknife

在实际应用中，Bootstrap对估计量的相关估计值具有随机性，即每一次运用Bootstrap算法抽样得到的估计值并不相同，而使用Jackknife对统计量进行估计时，各采样的样本之间的差异太小。本文考虑到Bootstrap与Jackknife的不足之处，结合两种算法，创新性地进行相关方差估计。采用Bootstrap选取多组样本，随后采用Jackknife对每组样本分别进行均值与标准差的估计，结合实际训练数据发现该方法得到的估计值稳健度更高。本文实现Bootstrap–Jackknife的具体步骤如下：

(1) 对于观测样本$x_1,\cdots,x_n$，进行$B=2000$次Bootstrap抽样，每次抽样$n$个样本；

(2) 假设$i=1{:}n$，每次选取上一步所有Bootstrap样本中不含有$x_i$的样本，并重新计算$\theta_{j(i)}$；

(3) 对第$j$个Bootstrap样本生成的所有$\theta_{j(i)}$计算相关估计值与标准差，标准差如式(4)：

$$\widehat{se}_j=\sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)}-\overline{\hat{\theta}_{(\cdot)}})^2} \tag{4}$$

(4) 最后对估计量$\hat{\theta}$的标准差进行Bootstrap估计，将$\widehat{se}\ (\hat{\theta}_1),\cdots,\widehat{se}\ (\hat{\theta}_j),\cdots,\widehat{se}\ (\hat{\theta}_B)$,的样本标准差作为估计值，得到$\widehat{se}_{Bootstrap-Jackknife}$如式(5)：

$$\widehat{se}_{Bootstrap-Jackknife}=\sqrt{\frac{1}{B-1}\sum_{j=1}^{B}(\widehat{se}\ (\hat{\theta}_j)-\overline{\widehat{se}\ (\hat{\theta}_{(*)})})^2},$$

$$\overline{\widehat{se}\ (\hat{\theta}_{(*)})}=\frac{1}{B}\sum_{j=1}^{B}(\widehat{se}\ (\hat{\theta}_j)) \tag{5}$$

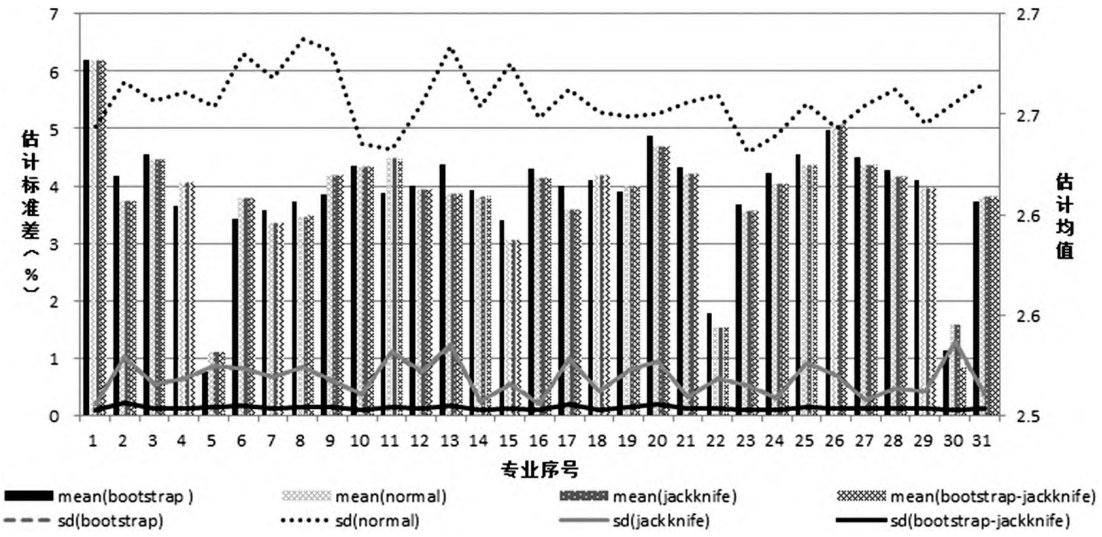


图1 四种方法估计在实际数据上的效果对比图

## 2 实例分析

分别采用Normal、Bootstrap、Jackknife、Bootstrap–Jackknife方法，对实际样本数据进行均值估计，实际训练样本为该校各专业学生在2017学年的四级成绩对数。估计结果对比情况如图1所示。

由图1数据可以看出：①对于Normal、Jackknife与Bootstrap–Jackknife这三种方法计算出的均值估计量仅有细微差异，而Bootstrap得到的均值估计值与其他三种方法得到的均值估计值相差较大；②对于标准差估计，Bootstrap–Jackknife估计得到的标准差要远远小于其他三种方法估计的标准差，这说明在对总体均值的估计中，Bootstrap–Jackknife的估计误差最小，即利用该方法得到的均值用来估计总体均值，其精度最高。另外Bootstrap与Jackknife的标准差估计值几乎重合为一条折线且远小于普通法的标准差估计值，这说明利用Bootstrap与Jackknife对估计量进行估计，其可信度要高于普通法得到的估计量值。

为了更明显的显示四种方法估计样本均值的差异，本文将四种方法得到的样本数据均值估计值进行排序，具体排序结果如表1所示(仅列举部分)。

表1 四种方法估计的均值排序对比

| | Bootstrap–Jackknife | Normal | Bootstrap | Jackknife |
|---|---|---|---|---|
| 英语 | 1 | 1 | 1 | 1 |
| 国际商务 | 2 | 2 | 2 | 23 |
| 会计学(ACCA班) | 3 | 3 | 3 | 5 |
| ... | ... | ... | ... | ... |
| 环境设计 | 29 | 30 | 29 | 17 |
| 视觉传达设计 | 30 | 31 | 31 | 29 |
| 产品设计 | 31 | 29 | 30 | 20 |

为比较Bootstrap–Jackknife方法与其他三种方法排序结果之间的差异，本文将各专业Bootstrap–Jackknife排序结果与其他三种方法得到的排序结果做差值处理，并进行绝对值运算，依据各差值结果绘制箱线图，如图2所示。

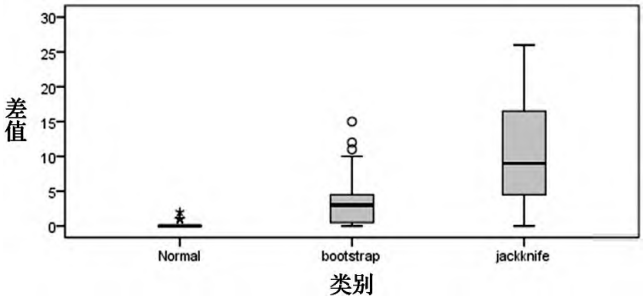


图2 各专业排序差绝对值箱线图

结合表1排序数据与图2箱线图可以看出：第一，Normal与Bootstrap–Jackknife在专业排序上的差异甚微，Bootstrap–Jackknife与Jackknife在专业排序上的差异最为显著，这说明就均值估计而言，Jackknife估计的稳定性并不高；第二，就排序数据上来看，该校英语四级应试能力前三的专业为英语、国际商务和会计学（ACCA班），而英语四级应试能力较差的专业为环境设计、视觉传达设计、产品设计这三个艺术专业。

## 3 结论

本文基于广州华商学院2017学年各专业学生四级成绩数据，运用Normal、Bootstrap、Jackknife和Bootstrap–Jackknife四种方差估计方法分别评估该校

(下转第80页)

均衡，是一种均匀的密集采样，导致训练困难。

## 4 结束语

本文选用的YOLOv3和SSD框架可实现四种阔叶材高效、准确辨识，YOLOv3框架辨识准确率更高，而SSD框架用时更短。总体而言，SSD对四种阔叶材做到了更高效自动辨识，可以在保证辨识的正确率前提下能够更快的处理样本，提高了阔叶材的识辨识效率。

本文识别准确率没有达到100%，综合分析与图像特点有关。本文只对四种木材样本进行研究，阔叶材种类相对单一，但是每种阔叶材采集的样本量较大，结果更具有适应性，下一步将从提高样本的多样性入手，增加不同阔叶材材种的训练集，从而提高模型的抗干扰和泛化能力，使其更适应于更多阔叶材材种的辨识。

**参考文献(References):**


[1] 赵子宇，杨霄霞，郭慧，等．基于卷积神经网络模型的木材宏、微观辨识方法[J]．林业科学，2021，57(6)：134-143

[2] 邵明伟，董军宇．基于深度学习的木材优选锯视觉检测算法[J]．林业科学，2020，56(12)：123-129

[3] 南玉龙，张慧春，郑加强，等．深度学习在林业中的应用[J]．世界林业研究，2021，34(5)：87-90

[4] 李若尘，朱悠翔，孙卫民，等．基于深度学习的木材缺陷图像的识别与定位[J]．数据采集与处理，2020，35(3)：494-505

[5] 林耀海，赵洪璐，杨泽灿，等．结合深度学习与Hough变换的等长原木材积检测系统[J]．林业工程学报，2021，6(1)：136-142

[6] 范佳楠，刘英，胡忠康，等．基于Faster R-CNN的实木板材缺陷检测识别系统[J]．林业工程学报，2019，4(3)：112-117

[7] 刘英，周晓林，胡忠康，等．基于优化卷积神经网络的木材缺陷检测[J]．林业工程学报，2019，4(1)：115-120

[8] 刘嘉政，王雪峰，王甜．基于深度学习的树种图像自动识别[J]．南京林业大学学报(自然科学版)，2020，44(1)：138-144

[9] KURDTHONGMEE W. A comparative study of the effectiveness of using popular DNN object detection algorithms for pith detection in cross-sectional images of parawood[J]. Heliyon，2020，6(2)：e03480

[10] HE T，LIU Y，YU Y，et al. Application of deep convolutional neural network on feature extraction and detection of wood defects[J]. Measurement，2020，152：107357

[11] HU J，SONG W，ZHANG W，et al. Deep learning for use in lumber classification tasks[J]. Wood Science and Technology，2019，53(2)：505-517

[12] 张晴晴，刘连忠，宁井铭，等．基于YOLOV3优化模型的复杂场景下茶树嫩芽识别[J]．浙江农业学报，2021，33(9)：1740-1747

[13] 吴天成，王晓荟，蔡艺军，等．基于特征融合的轻量级SSD目标检测方法[J]．液晶与显示，2021，36(10)：1437-1444

[14] 林相泽，张俊媛，徐啸，等．基于字典学习与SSD的不完整昆虫图像稻飞虱识别分类[J]．农业机械学报，2021，52(9)：165-171

[15] 李子茂，刘恋冬，夏梦，等．基于深度学习的月季多叶片病虫害检测研究[J]．中国农机化学报，2021，42(8)：169-176


(上接第75页)

各专业四级总体应试水平；对比估计结果发现：Bootstrap-Jackknife算法在估计总体均值方面上估计误差最低，在涉及排序问题上，Jackknife算法的排序稳定性最低。研究结果表明，Bootstrap-Jackknife算法可更精确、稳定的评估高校各专业总体英语应试水平，从而为高校在专业层面上制定科学的英语学习方式和可操作的实施办法[8]提供参考。

**参考文献(References):**


[1] Quenouille M H. Problems in plane sampling[J].The Annals of Mathematical Statistics，1949：355-375

[2] Quenouille M H. Notes on bias in estimation[J].Biometrika，1956，43(3/4)：353-360

[3] Efron B. Nonparametric estimates of standard error：the jackknife，the bootstrap and other methods[J].Biometrika，1981，68(3)：589-599

[4] 吕萍．抽样信息在复杂调查数据中的应用研究[J]．统计研究，2017，34(1)：108-118

[5] 宋一凡，贾云献，陈明华．基于Bootstrap方法的弹药最大射程评估[J]．计算机仿真，2020，37(6)：1-3

[6] 毛平．Bootstrap方法及其应用[D]．湘潭大学，2013

[7] 黎光明，黄梓龙．概化理论方差分量及其变异量估计：Jackknife方法与Traditional方法比较[J]．统计与决策，2020，36(6)：10-14

[8] 温静怡，崔盛．大学英语等级考试成绩对高校毕业生升学的影响——基于首都大学生成长追踪调查数据的实证研究[J]．教育经济评论，2020，5(6)：91-107