

# **IBM SPSS Bootstrapping 22**



**注释**

使用本信息及其支持的产品之前，请阅读 第 7 页的『声明』 中的信息。

**产品信息**

此版本适用于 IBM® SPSS® Statistics V22R0M0 及所有后续发行版和修改版，除非在新版本中另有说明。

---

## 目录

第 1 章 Bootstrap 简介 . . . . .	1	声明 . . . . .	7
第 2 章 Bootstrap . . . . .	3	商标 . . . . .	8
支持 Bootstrap 的过程 . . . . .	4	索引 . . . . .	11
BOOTSTRAP 命令附加功能 . . . . .	6		



---

## 第 1 章 Bootstrap 简介

在收集数据时，您通常对从中抽取样本的总体的属性感兴趣。您通过从样本计算得到的估计值来做出有关这些总体参数的推论。例如，如果随产品附带的 *Employee data.sav* 数据集为来自更大的职员总体的随机样本，则当前工资的样本平均值 \$34,419.57 为职员总体的当前平均工资的估计值。并且，此估计值对大小为 474 的样本具有标准误差 \$784.311，因此职员总体当前平均工资的 95% 置信区间为 \$32,878.40 至 \$35,960.73。但是这些估计量有多可靠呢？对于某些“已知”总体和表现良好的参数，我们非常了解样本估计值的属性，因此可以相信这些结果。Bootstrap 寻求发现有关“未知”总体和异常参数的估计量的更多属性信息。

### Bootstrap 的工作原理

考虑最简单的情况，对于样本大小为  $N$  的数据集，您可以通过放回方式从原始数据集中取得  $B$  个 bootstrap 样本（大小为  $N$ ），并为这  $B$  个 bootstrap 样本中的每个样本计算估计量。这  $B$  个 bootstrap 估计值为大小为  $B$  的样本，您可从中做出有关估计量的推论。例如，如果从 *Employee data.sav* 数据集取得 1000 个 bootstrap 样本，则当前工资样本平均值的 bootstrap 估计标准误差 \$778.76 可以替代估计值 \$784.311。

此外，Bootstrap 可进一步提供标准误差和中位数的置信区间，而参数估计则对此不适用。

### 产品对 Bootstrap 的支持

Bootstrap 作为子对话框包含在支持 bootstrap 的过程中。请参阅第 4 页的『支持 Bootstrap 的过程』以获取关于哪些过程支持 bootstrap 的信息。

当在对话框中请求 bootstrap 时，将在对话框生成的常规语法外粘贴新的单独 BOOTSTRAP 命令。BOOTSTRAP 命令按照您的指定创建 bootstrap 样本。产品内部对这些 bootstrap 样本的处理方式与拆分类似，尽管它们不会明确显示在“数据编辑器”中。这意味着，内部有效存在  $B*N$  个个案，因此在 bootstrap 期间处理数据时，状态栏上的个案计数器将从 1 计数到  $B*N$ 。输出管理系统 (OMS) 用于收集在每个“bootstrap 拆分”上运行分析的结果。这些 bootstrap 结果在汇聚后与过程生成的其余常规输出一起显示在“查看器”中。在某些个案中，您可能会看到对“bootstrap 拆分 0”的引用；这是原始数据集。



---

## 第 2 章 Bootstrap

Bootstrap 方法可以导出稳健的标准误差估计值，并能为诸如平均值、中位数、比例、几率比、相关系数或回归系数等估计值导出置信区间。它还可用于构建假设检验。当参数估计方法的假设存在疑问（例如，异方差残差拟合较小样本的回归模型），参数推论无法执行或需要非常复杂的标准误差计算公式（例如，为中位数、四分位数和其他百分位数计算置信区间）时，Bootstrap 是最好的替代选项。

**示例。** 一家电信公司每月大约会流失 27% 的客户。为了正确实施减少客户流失的举措，管理部门需要知道这一百分比在各个预定义客户组之间是否不同。使用 bootstrap 可以确定单一的客户流失率是否充分描述了四个主要的客户类型。

在查看员工记录时，管理部门对员工的以往工作经验比较感兴趣。工作经验向右偏斜，这使得平均值在作为员工“典型”以往工作经验的估计方面不如中位数理想。然而，在产品中无法获得中位数的参数置信区间。

管理部门还有兴趣使用线性模型拟合度当前和起始工资之间的差异，以确定哪些因素与员工工资上升存在关联。在 bootstrap 线性模型时，可以使用特殊的重新抽样方法（残差和狂野 bootstrap）以获得更准确的结果。

许多过程支持 bootstrap 抽样和对 bootstrap 样本分析结果的汇聚。指定 bootstrap 分析的控件作为公共子对话框直接集成在支持 bootstrap 的过程中。在 bootstrap 对话框上的设置会在不同过程中保留，因此如果您通过对话框使用 bootstrap 运行频率分析，则对支持此功能的其他过程而言，bootstrap 缺省打开。

获取 Bootstrap 分析

1. 从菜单中，选择一个支持 bootstrap 的过程，并单击 **Bootstrap**。
2. 选择执行 **bootstrap**。

此外，您还可以控制下列选项：

**样本数。** 对于生成的百分位数和 BCa 区间，建议使用至少 1000 个 bootstrap 样本。指定一个正整数。

**为 Mersenne 扭曲器设置种子。** 设置种子允许您复制分析。使用此控件类似于将“Mersenne 扭曲器”设为活动生成器并在“随机数生成器”对话框中指定固定起始点，两者的重大差别在于在此对话框中设置种子会保留随机数生成器的当前状态并在分析完成后恢复该状态。

**置信区间。** 指定一个大于 50 且小于 100 的置信度。百分位数区间简单地使用对应于置信区间百分位数的有序 bootstrap 值。例如，一个 95% 的百分位数置信区间使用 bootstrap 值的第 2.5 个和第 97.5 个百分位数作为区间的下限和上限（必要时插 bootstrap 值）。偏差修正加速 (BCa) 区间为调整区间，它更加准确，但代价是需要更长的计算时间。

**抽样。** 简单方法为通过放回方式从原始数据集进行个案重新取样。分层方法为通过放回方式从原始数据集进行个案重新取样，但在层次变量的交叉分类定义的层内。如果层中的单元格相对均一，且不同层间的单元格相差较大，则分层 bootstrap 抽样非常有用。

---

## 支持 Bootstrap 的过程

下列过程支持 bootstrap。

注意:

- Bootstrap 不能用于多重插补数据集。如果在数据集中存在 *Imputation\_* 变量, Bootstrap 对话框将被禁用。
- 如果存在非整数权重值, 那么引导过程不运行。
- Bootstrap 使用成列删除来确定个案基础; 也就是说, 在任何分析变量上具有缺失值的个案将从分析中删除, 因此当 bootstrap 生效时, 成列删除也处于生效, 即使分析过程指定了其他缺失值处理方式。

### Statistics Base 选项

**频率。** 以下功能受支持:

- 统计表支持平均值、标准差、方差、中位数、偏度、峰度和百分位数的 bootstrap 估计。
- 频率表支持百分比的 bootstrap 估计。

**描述。** 以下功能受支持:

- 描述统计表支持平均值、标准差、方差、偏度和峰度的 bootstrap 估计。

**浏览。** 以下功能受支持:

- 描述表支持平均值、5% 切尾平均值、标准差、方差、中位数、偏度、峰度和四分位距的 bootstrap 估计。
- M 估计表支持 Huber 的 M 估计、Tukey 的双权重、Hampel 的 M 估计和 Andrew 的 Wave 的 bootstrap 估计。
- 百分位数表支持百分位数的 bootstrap 估计。

**交叉表。** 以下功能受支持:

- 定向测量表支持 Lambda、Goodman 和 Kruskal Tau、不确定性系数和 Somers 的 d 的 bootstrap 估计。
- 对称测量表支持 Phi、Cramer V、列联系数、Kendall 的 tau-b、Kendall 的 tau-c、伽玛、Spearman 相关性和 Pearson 的 R 的 bootstrap 估计。
- 风险评估表支持几率比的 bootstrap 估计。
- Mantel-Haenszel 一般几率比表支持  $\ln(\text{Estimate})$  的 bootstrap 估计和显著性检验。

**平均值。** 以下功能受支持:

- 报告表支持平均值、中位数、组内中位数、标准差、方差、峰度、偏度、调和平均值和几何平均值的 bootstrap 估计。

**单样本 T 检验。** 以下功能受支持:

- 统计表支持平均值和标准差的 bootstrap 估计。
- 检验表支持平均值差值的 bootstrap 估计和显著性检验。

**独立样本 T 检验。** 以下功能受支持:

- 组统计表支持平均值和标准差的 bootstrap 估计。
- 检验表支持平均值差值的 bootstrap 估计和显著性检验。

**配对样本 T 检验。** 以下功能受支持:

- 统计表支持平均值和标准差的 bootstrap 估计。
- 相关表支持相关性的 bootstrap 估计。



- 检验表支持平均值的 bootstrap 估计。

**单因素 ANOVA。** 以下功能受支持:

- 描述统计表支持平均值和标准差的 bootstrap 估计。
- 多重比较表支持平均值差值的 bootstrap 估计。
- 对比检验表支持对比值的 bootstrap 估计和显著性检验。

**GLM 单变量。** 以下功能受支持:

- 描述统计表支持平均值和标准差的 bootstrap 估计。
- 参数估计值表支持系数、B 的 bootstrap 估计和显著性检验。
- 对比结果表支持差值的 bootstrap 估计和显著性检验。
- 估计的边际均值: 估计表支持平均值的 bootstrap 估计。
- 估计的边际均值: 成对比较表支持平均值差值的 bootstrap 估计。
- 事后检验: 多重比较表支持平均值差值的 bootstrap 估计。

**双变量相关性。** 以下功能受支持:

- 描述统计表支持平均值和标准差的 bootstrap 估计。
- 相关表支持相关性的 bootstrap 估计和显著性检验。

注意:

如果除 Pearson 相关性外, 还请求了非参数相关性 (Kendall 的 tau-b 或 Spearman), 则对话框将分别为其粘贴 CORRELATIONS 和 NONPAR CORR 命令, 以及单独的 BOOTSTRAP 命令。将使用相同的 bootstrap 样本计算全部相关性。

在汇聚之前, 将 Fisher Z 转换应用于相关性。在汇聚之后, 应用逆 Z 转换。

**偏相关。** 以下功能受支持:

- 描述统计表支持平均值和标准差的 bootstrap 估计。
- 相关表支持相关性的 bootstrap 估计。

**线性回归。** 以下功能受支持:

- 描述统计表支持平均值和标准差的 bootstrap 估计。
- 相关表支持相关性的 bootstrap 估计。
- 模型摘要表支持 Durbin-Watson 的 bootstrap 估计。
- 系数表支持系数、B 的 bootstrap 估计和显著性检验。
- 相关系数表支持相关性的 bootstrap 估计。
- 残差统计表支持平均值和标准差的 bootstrap 估计。

**序数回归。** 以下功能受支持:

- 参数估计值表支持系数、B 的 bootstrap 估计和显著性检验。

**判别分析。** 以下功能受支持:

- 标准化典则判别函数系数表支持标准化系数的 bootstrap 估计。
- 典则判别函数系数表支持非标准化系数的 bootstrap 估计。
- 分类函数系数表支持系数的 bootstrap 估计。

## Advanced Statistics 选项

**GLM 多变量。** 以下功能受支持:

- 参数估计值表支持系数、B 的 bootstrap 估计和显著性检验。

**线性混合模型。** 以下功能受支持:

- 固定效应估计值表支持估计值的 bootstrap 估计和显著性检验。
- 协方差参数估计值表支持估计值的 bootstrap 估计和显著性检验。

**广义线性模型。** 以下功能受支持:

- 参数估计值表支持系数、B 的 bootstrap 估计和显著性检验。

**Cox 回归。** 以下功能受支持:

- 方程中的变量表支持系数、B 的 bootstrap 估计和显著性检验。

## Regression 选项

**二元 Logistic 回归。** 以下功能受支持:

- 方程中的变量表支持系数、B 的 bootstrap 估计和显著性检验。

**多项 Logistic 回归。** 以下功能受支持:

- 参数估计值表支持系数、B 的 bootstrap 估计和显著性检验。

---

## BOOTSTRAP 命令附加功能

使用命令语法语言还可以:

- 执行残差和狂野 bootstrap 抽样 (SAMPLING 子命令)

请参阅命令语法参考以获取完整的语法信息。

---

## 声明

本信息是为在美国提供的产品和服务编写的。

IBM 可能在其他国家或地区不提供本文中讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。不过，用户应自行负责评估和验证任何非 IBM 产品、程序或服务的工作情况。

IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。您获得该文档并不意味着授予您任何这些专利许可。您可以将书面许可查询函件发送至：

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

有关双字节 (DBCS) 信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：

Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
1623-14, Shimotsuruma, Yamato-shi  
Kanagawa 242-8502 Japan

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。有些州/省不允许针对某些交易的明示或暗示免责条款，因此本声明可能不适用于您。

本信息可能含有技术误差或排版错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可以随时对本资料中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对任何非 IBM Web 站点的引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：(i) 允许在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及 (ii) 允许对已经交换的信息进行相互使用，请与下列地址联系：

IBM Software Group  
ATTN: Licensing

200 W. Madison St.  
Chicago, IL; 60606  
U.S.A.

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

本资料中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际软件许可协议或任何同等协议中的条款提供。

此处所含的性能数据均在受控环境下决定。因此，在其他操作环境中获得的结果可能差异较大。有些测量可能在开发级的系统中进行，不保证这些测量结果与常用系统上的测量结果相同。此外，有些测量结果可能通过推断来估计得出。实际结果可能有所差异。此文档的用户应针对其具体环境验证适用的数据。

非 IBM 产品的相关信息来自这些产品的供应商，及其发布的公告或其他公开来源。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

有关 IBM 未来方向或意向的所有声明均可能未经通知即变更或撤销，并且仅代表目标和目的。

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

凡这些实例程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明：

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

凡这些实例程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明：

©（贵公司的名称）（年）。此部分代码是根据 IBM Corp. 公司的样本程序衍生出来的。

© Copyright IBM Corp. \_（输入年份）\_. All rights reserved.

---

## 商标

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corp., 在全球许多管辖区域注册的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。最新的 IBM 商标列表可从 Web 站点 [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) 上的『版权和商标信息』部分获取。

Adobe、Adobe 徽标、PostScript 和 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国和其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国、其他国家或地区或两者的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国、其他国家或地区或两者的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

Java 和所有基于 Java 的商标和徽标是 Oracle 及/或其分支机构的商标和注册商标。



---

# 索引

## B

Bootstrap 3

支持的过程 4









Printed in China