

IBM SPSS Bootstrapping 22



お願い

本書および本書で紹介する製品をご使用になる前に、9ページの『特記事項』に記載されている情報をお読みください。

本書は、IBM® SPSS® Statistics バージョン 22 リリース 0 モディフィケーション 0 および新しい版で明記されない限り、以降のすべてのリリースおよびモディフィケーションに適用されます。

お客様の環境によっては、資料中の円記号がバックスラッシュと表示されたり、バックスラッシュが円記号と表示されたりする場合があります。

原典： IBM SPSS Bootstrapping 22

発行： 日本アイ・ビー・エム株式会社

担当： トランスレーション・サービス・センター

目次

第 1 章 ブートストラップの概要	1	特記事項.	9
第 2 章 ブートストラップ	3	商標.	10
ブートストラップをサポートするプロシージャ	4	索引	13
BOOTSTRAP コマンドの追加機能	6		

第 1 章 ブートストラップの概要

データを収集するときに、サンプルを収集した母集団のプロパティに関心を持つことがしばしばあります。サンプルから計算された推定値を使用して、これらの母集団パラメーターに関する推定を行います。例えば、製品に付属の *Employee data.sav* データ・セットが従業員の大規模な母集団からのランダム・サンプリングである場合、このサンプルの現在の給与の平均値である \$34,419.57 は、従業員の母集団についての現在の給与の平均値の推定値になります。さらに、この推定値には、サイズ 474 のサンプルに対し \$784.311 の標準誤差があるため、従業員の母集団での現在の給与の平均値の 95% の信頼区間は \$32,878.40 から \$35,960.73 になります。これらの推定値はどの程度信頼できるものなのでしょうか。特定の「既知の」母集団と正常に動作するパラメーターの場合、サンプル推定のプロパティについて熟知していて、これらの結果を信頼できます。ブートストラップは、「不明な」母集団と正常に動作しないパラメーターの推定値のプロパティについて、より多くの情報を明らかにしようとするものです。

ブートストラップの機能仕様

最も単純な動作は、サンプルのサイズ N のデータ・セットについて、元のデータ・セットからの置き換えにより、サイズ N の B 「ブートストラップ」サンプルを取り出し、これらの B ブートストラップ・サンプルそれぞれについて推定値を計算します。これらの B ブートストラップ推定値は、サイズ B のサンプルであり、これを元に推定値について推定を実行できます。例えば、*Employee data.sav* データ・セットから 1,000 ブートストラップ・サンプルを取り出すと、このサンプルの現在の給与の平均値に対するブートストラップ推定値の標準誤差 \$776.91 は、推定値 \$784.311 に対する代替値になります。

ブートストラップには、中央値に対する標準誤差および信頼区間も用意されています。これらは、パラメトリック推定では使用できません。

製品でのブートストラップのサポート

ブートストラップは、ブートストラップをサポートするプロシーチャーのサブダイアログとして組み込まれています。ブートストラップをサポートするプロシーチャーについては、4 ページの『ブートストラップをサポートするプロシーチャー』を参照してください。

ダイアログでブートストラップが要求されると、ダイアログによって生成された通常のシンタックスに加えて、新しい別個の `BOOTSTRAP` コマンドが貼り付けられます。この `BOOTSTRAP` コマンドにより、指定に従ってブートストラップ・サンプルが作成されます。製品は、内部的にはこれらのブートストラップ・サンプルを分割として処理します。ただし、データ・エディターには明示的に分割であることは示されません。つまり、内部的には有効なケースが $B*N$ 件あるため、ブートストラップ中にデータを処理するときに、ステータス・バーのケース・カウンターに 1 から $B*N$ の値が表示されます。各「ブートストラップ分割」に対して分析を実行した結果の収集には、出力管理システム (OMS) が使用されます。これらの結果はプールされ、プールされたブートストラップの結果は、プロシーチャーで生成されたその他の通常出力とともにビューアーに表示されます。一部のケースでは、「ブートストラップ分割 0」が参照される場合があります。これは元のデータ・セットです。

第 2 章 ブートストラップ

ブートストラップとは、平均値、中央値、比率、オッズ比、相関係数、回帰係数などの、推定に対する標準誤差および信頼区間の頑強な推定を派生させる方法です。これは、仮説検定の構築にも使用されることがあります。ブートストラップは、これらの方法の仮定が不確かな場合 (不均一残差の回帰モデルに適合するサンプル数が小さい場合など)、またはパラメトリック推定が不可能であるか、標準誤差を計算するには非常に複雑な数式を必要とする場合 (中央値、4 分位、その他のパーセンタイルの信頼区間を計算する場合など) に、パラメトリック推定の代替として最も役立ちます。

例。ある電話会社では、毎月 27% の顧客が解約しています。解約数の減少に向けたとりくみを適切に集中させるために、経営陣は、事前定義された顧客グループ間でこの解約率に違いがあるかどうかを知る必要があります。ブートストラップを使用して、単一の解約率が 4 つの主要な顧客タイプを適切に表しているかどうかを判別できます。

従業員レコードを見直して、経営陣は従業員の以前の業務経験に関心を持ちます。業務経験が右肩下がりであるため、従業員の「典型的な」以前の業務経験の推定値について、平均値の方が中央値よりも適切ではなくなります。しかし、この製品では、中央値に対してはパラメトリック信頼区間を使用できません。

経営陣は、線型モデルを現在の給与と初任給との差分に適用することで、従業員の昇給に関連付ける要素を決定することにも関心を持ちます。ブートストラップが線型モデルのときは、特別な再サンプリング方法 (残差ブートストラップおよびワイルド・ブートストラップ) を使用して、より正確な結果を取得できます。

多くのプロシージャーが、ブートストラップのサンプリング、およびブートストラップ・サンプルの分析から得られた結果のプールをサポートしています。ブートストラップ分析を指定するコントロールは、ブートストラップをサポートするプロシージャーの共通サブダイアログとして直接統合されています。ブートストラップ・ダイアログの設定は、プロシージャーをまたがって持続するため、ダイアログを使用してブートストラップにより度数分析を実行すると、ブートストラップをサポートする他のプロシージャーについて、ブートストラップがデフォルトで有効になります。

ブートストラップ分析を行うには、以下を実行します。

1. メニューからブートストラップをサポートするプロシージャーを選択し、「ブートストラップ」をクリックします。
2. 「ブートストラップの実行」を選択します。

オプションで、次のオプションを制御できます。

「**サンプル数**」。パーセンタイル区間および BCa 区間を生成する場合、少なくとも 1000 件のブートストラップ・サンプルを使用することをお勧めします。正の整数を指定してください。

「**Mersenne Twister のシードを設定**」。シードを設定すると、分析を複製できます。このコントロールを使用すると、アクティブ・ジェネレーターとして Mersenne Twister を設定し、「乱数ジェネレーター」ダイアログの固定開始ポイントを指定することと同様に機能しますが、このダイアログでシードを設定することによる大きな違いは、乱数ジェネレーターの現在の状態が保存され、分析の完了後にその状態が復元されることです。

「信頼区間」。50 より大きく、100 より小さい信頼度レベルを指定してください。パーセンタイル区間は、単に、信頼区間のパーセンタイルに応じて並べられたブートストラップ値を使用します。例えば、95%パーセンタイルの信頼区間は、ブートストラップ値の 2.5 番目と 97.5 番目のパーセンタイルを区間の下限および上限として使用します (必要に応じて、ブートストラップ値を補間します)。バイアス補正加速 (BCa) 区間は、区間をより正確に調整しますが、その代わり、より長い計算時間を必要とします。

「サンプリング」。Simple メソッドでは、元のデータ・セットからの置き換えにより、ケースを再サンプリングします。Stratified メソッドでは、ストラータ変数の交差分類によって定義されている階層内で、元のデータ・セットからの置き換えによりケースを再サンプリングします。層化ブートストラップ・サンプリングは、階層全体では単位が大きく異なるのに対し、階層内では単位が比較的同質である場合に役立ちます。

ブートストラップをサポートするプロシージャ

ブートストラップがサポートされるプロシージャは、次のとおりです。

注:

- ブートストラップは、多重代入データ・セットでは機能しません。データ・セットに *Imputation_* 変数がある場合、「ブートストラップ」ダイアログは無効になります。
- 整数でない重み値がある場合、ブートストラップは機能しません。
- ブートストラップは、リストごとの削除を使用してケース基準を決定します。つまり、分析変数のいずれかについて欠損値を持つケースは分析から削除されるため、ブートストラップが有効なときは、分析プロシージャで別の形式の欠損値処理を指定している場合でも、リストごとの削除が有効です。

Statistics Base のオプション

「度数」。サポートされる機能は次のとおりです。

- 統計表では、平均値、標準偏差、分散、中央値、歪度、尖度、およびパーセンタイルのブートストラップ推定がサポートされます。
- 度数分布表では、パーセントのブートストラップ推定がサポートされます。

「記述統計」。サポートされる機能は次のとおりです。

- 記述統計表では、平均値、標準偏差、分散、歪度、および尖度のブートストラップ推定がサポートされます。

「探索」。サポートされる機能は次のとおりです。

- 記述表では、平均値、5% トリム平均値、標準偏差、分散、中央値、歪度、尖度、および 4 分位範囲のブートストラップ推定がサポートされます。
- M 推定量表では、Huber の M 推定量、Tukey のパイウェイト、Hampel の M 推定量、および Andrew のウェーブのブートストラップ推定がサポートされます。
- パーセンタイル表では、パーセンタイルのブートストラップ推定がサポートされます。

「クロス集計表」。サポートされる機能は次のとおりです。

- 傾向性による類似度表では、ラムダ、Goodman と Kruskal のタウ、不確定性係数、および Somers の d のブートストラップ推定がサポートされます。
- 対称性による類似度表では、ファイ、Cramer の V、分割係数、Kendall のタウ b、Kendall のタウ c、ガンマ、Spearman の相関、および Pearson の R のブートストラップ推定がサポートされます。
- リスク推定表では、オッズ比のブートストラップ推定がサポートされます。

- Mantel-Haenszel の共通オッズ比表では、 $\ln(\text{推定値})$ のブートストラップ推定および有意度検定がサポートされます。

「平均」。サポートされる機能は次のとおりです。

- レポート表では、平均値、中央値、グループ中央値、標準偏差、分散、尖度、歪度、調和平均、および幾何平均のブートストラップ推定がサポートされます。

「1 サンプルの T 検定」。サポートされる機能は次のとおりです。

- 統計表では、平均値および標準偏差のブートストラップ推定がサポートされます。
- 検定表では、平均差のブートストラップ推定および有意度検定がサポートされます。

「独立したサンプルの T 検定」。サポートされる機能は次のとおりです。

- グループ統計表では、平均値および標準偏差のブートストラップ推定がサポートされます。
- 検定表では、平均差のブートストラップ推定および有意度検定がサポートされます。

「対応のあるサンプルの T 検定」。サポートされる機能は次のとおりです。

- 統計表では、平均値および標準偏差のブートストラップ推定がサポートされます。
- 相関表では、相関のブートストラップ推定がサポートされます。
- 検定表では、平均値のブートストラップ推定がサポートされます。

「一元配置分散分析」。サポートされる機能は次のとおりです。

- 記述統計表では、平均値および標準偏差のブートストラップ推定がサポートされます。
- 多重比較表では、平均差のブートストラップ推定がサポートされます。
- 対比の検定表では、対比の値のブートストラップ推定および有意度検定がサポートされます。

「GLM 1 変量」。サポートされる機能は次のとおりです。

- 記述統計表では、平均値および標準偏差のブートストラップ推定がサポートされます。
- パラメーター推定値表では、ブートストラップ推定と係数 B の有意度検定がサポートされます。
- 対比結果表では、差のブートストラップ推定および有意度検定がサポートされます。
- 推定周辺平均:推定表では、平均値のブートストラップ推定がサポートされます。
- 推定周辺平均:ペアワイズ比較表では、平均差のブートストラップ推定がサポートされます。
- その後の検定: 多重比較表では、平均差のブートストラップ推定がサポートされます。

「2 変量の相関分析」。サポートされる機能は次のとおりです。

- 記述統計表では、平均値および標準偏差のブートストラップ推定がサポートされます。
- 相関表では、相関のブートストラップ推定および有意差検定がサポートされます。

注:

Pearson の相関に加えてノンパラメトリック相関 (Kendall のタウ b または Spearman) が要求されると、ダイアログは、それぞれに個別の `BOOTSTRAP` コマンドとともに、`CORRELATIONS` コマンドおよび `NONPAR CORR` コマンドを貼り付けします。すべての相関の計算に、同じブートストラップ・サンプルが使用されます。

プールする前に、相関に Fisher Z 変換が適用されます。プールした後に、逆 Z 変換が適用されます。

「偏相関」。サポートされる機能は次のとおりです。

- 記述統計表では、平均値および標準偏差のブートストラップ推定がサポートされます。
- 相関表では、相関のブートストラップ推定がサポートされます。

「線型回帰」。サポートされる機能は次のとおりです。

- 記述統計表では、平均値および標準偏差のブートストラップ推定がサポートされます。
- 相関表では、相関のブートストラップ推定がサポートされます。
- モデルの要約表では、Durbin-Watson のブートストラップ推定がサポートされます。
- 係数表では、ブートストラップ推定と係数 B の有意度検定がサポートされます。
- 相関係数表では、相関のブートストラップ推定がサポートされます。
- 残差の統計量表では、平均値および標準偏差のブートストラップ推定がサポートされます。

「順序回帰」。サポートされる機能は次のとおりです。

- パラメーター推定値表では、ブートストラップ推定と係数 B の有意度検定がサポートされます。

「判別分析」。サポートされる機能は次のとおりです。

- 標準化された標準判別関数係数表では、標準化された係数のブートストラップ推定がサポートされます。
- 標準判別関数係数表では、標準化されていない係数のブートストラップ推定がサポートされます。
- 分類関数係数表では、係数のブートストラップ推定がサポートされます。

Advanced Statistics のオプション

「GLM 多変量」。サポートされる機能は次のとおりです。

- パラメーター推定値表では、ブートストラップ推定と係数 B の有意度検定がサポートされます。

「線型混合モデル」。サポートされる機能は次のとおりです。

- 固定効果の推定表では、推定のブートストラップ推定および有意度検定がサポートされます。
- 共分散パラメーターの推定表では、推定のブートストラップ推定および有意度検定がサポートされます。

「一般化線型モデル」。サポートされる機能は次のとおりです。

- パラメーター推定値表では、ブートストラップ推定と係数 B の有意度検定がサポートされます。

「Cox 回帰」。サポートされる機能は次のとおりです。

- 式中の変数表では、ブートストラップ推定と係数 B の有意度検定がサポートされます。

Regression のオプション

「2 項ロジスティック回帰」。サポートされる機能は次のとおりです。

- 式中の変数表では、ブートストラップ推定と係数 B の有意度検定がサポートされます。

「多項ロジスティック回帰」。サポートされる機能は次のとおりです。

- パラメーター推定値表では、ブートストラップ推定と係数 B の有意度検定がサポートされます。

BOOTSTRAP コマンドの追加機能

コマンド・シンタックス言語では、以下を行うこともできます。

- 残差ブートストラップ・サンプリングおよびワイルド・ブートストラップ・サンプリングの実行 (SAMPLINGサブコマンド)

シンタックス情報について詳しくは、『シンタックス参照コマンド』を参照してください。

特記事項

本書は米国 IBM が提供する製品およびサービスについて作成したものです。

本書に記載の製品、サービス、または機能が日本においては提供されていない場合があります。日本で利用可能な製品、サービス、および機能については、日本 IBM の営業担当員にお尋ねください。本書で IBM 製品、プログラム、またはサービスに言及していても、その IBM 製品、プログラム、またはサービスのみが使用可能であることを意味するものではありません。これらに代えて、IBM の知的所有権を侵害することのない、機能的に同等の製品、プログラム、またはサービスを使用することができます。ただし、IBM 以外の製品とプログラムの操作またはサービスの評価および検証は、お客様の責任で行っていただきます。

IBM は、本書に記載されている内容に関して特許権 (特許出願中のものを含む) を保有している場合があります。本書の提供は、お客様にこれらの特許権について実施権を許諾することを意味するものではありません。実施権についてのお問い合わせは、書面にて下記宛先にお送りください。

〒103-8510

東京都中央区日本橋箱崎町19番21号

日本アイ・ビー・エム株式会社

法務・知的財産

知的財産権ライセンス渉外

以下の保証は、国または地域の法律に沿わない場合は、適用されません。IBM およびその直接または間接の子会社は、本書を特定物として現存するままの状態を提供し、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。

この情報には、技術的に不適切な記述や誤植を含む場合があります。本書は定期的に見直され、必要な変更は本書の次版に組み込まれます。IBM は予告なしに、随時、この文書に記載されている製品またはプログラムに対して、改良または変更を行うことがあります。

本書において IBM 以外の Web サイトに言及している場合がありますが、便宜のため記載しただけであり、決してそれらの Web サイトを推奨するものではありません。それらの Web サイトにある資料は、この IBM 製品の資料の一部ではありません。それらの Web サイトは、お客様の責任でご使用ください。

IBM は、お客様が提供するいかなる情報も、お客様に対してなんら義務も負うことのない、自ら適切と信ずる方法で、使用もしくは配布することができるものとします。

本プログラムのライセンス保持者で、(i) 独自に作成したプログラムとその他のプログラム (本プログラムを含む) との間での情報交換、および (ii) 交換された情報の相互利用を可能にすることを目的として、本プログラムに関する情報を必要とする方は、下記に連絡してください。

IBM Software Group

ATTN: Licensing

200 W. Madison St.

Chicago, IL; 60606

U.S.A.

本プログラムに関する上記の情報は、適切な使用条件の下で使用することができますが、有償の場合もあります。

本書で説明されているライセンス・プログラムまたはその他のライセンス資料は、IBM 所定のプログラム契約の契約条項、IBM プログラムのご使用条件、またはそれと同等の条項に基づいて、IBM より提供されます。

この文書に含まれるいかなるパフォーマンス・データも、管理環境下で決定されたものです。そのため、他の操作環境で得られた結果は、異なる可能性があります。一部の測定が、開発レベルのシステムで行われた可能性があります。その測定値が、一般に利用可能なシステムのものと同じである保証はありません。さらに、一部の測定値が、推定値である可能性があります。実際の結果は、異なる可能性があります。お客様は、お客様の特定の環境に適したデータを確かめる必要があります。

IBM 以外の製品に関する情報は、その製品の供給者、出版物、もしくはその他の公に利用可能なソースから入手したものです。IBM は、それらの製品のテストは行っておりません。したがって、他社製品に関する実行性、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給者をお願いします。

IBM の将来の方向または意向に関する記述については、予告なしに変更または撤回される場合があります、単に目標を示しているものです。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する企業が実在しているとしても、それは偶然にすぎません。

それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生的創作物にも、次のように、著作権表示を入れていただく必要があります。

本書には、日常の業務処理で用いられるデータや報告書の例が含まれています。より具体性を与えるために、それらの例には、個人、企業、ブランド、あるいは製品などの名前が含まれている場合があります。これらの名称はすべて架空のものであり、名称や住所が類似する企業が実在しているとしても、それは偶然にすぎません。

それぞれの複製物、サンプル・プログラムのいかなる部分、またはすべての派生的創作物にも、次のように、著作権表示を入れていただく必要があります。

© (お客様の会社名) (西暦年). このコードの一部は、IBM Corp. のサンプル・プログラムから取られています。

© Copyright IBM Corp. _年を入れる_. All rights reserved.

商標

IBM、IBM ロゴおよび ibm.com は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、<http://www.ibm.com/legal/copytrade.shtml> をご覧ください。

Adobe、Adobe ロゴ、PostScript、PostScript ロゴは、Adobe Systems Incorporated の米国およびその他の国における登録商標または商標です。

インテル、Intel、Intel ロゴ、Intel Inside、Intel Inside ロゴ、Centrino、Intel Centrino ロゴ、Celeron、Xeon、Intel SpeedStep、Itanium、および Pentium は、Intel Corporation または子会社の米国およびその他の国における商標または登録商標です。

Linux は、Linus Torvalds の米国およびその他の国における登録商標です。

Microsoft、Windows、Windows NT および Windows ロゴは、Microsoft Corporation の米国およびその他の国における商標です。

UNIX は The Open Group の米国およびその他の国における登録商標です。

Java およびすべての Java 関連の商標およびロゴは Oracle やその関連会社の米国およびその他の国における商標または登録商標です。

索引

日本語, 数字, 英字, 特殊文字の順に配列されています。なお, 濁音と半濁音は清音と同等に扱われています。

[ハ行]

ブートストラップ 3

サポートされるプロシージャー 4



Printed in Japan