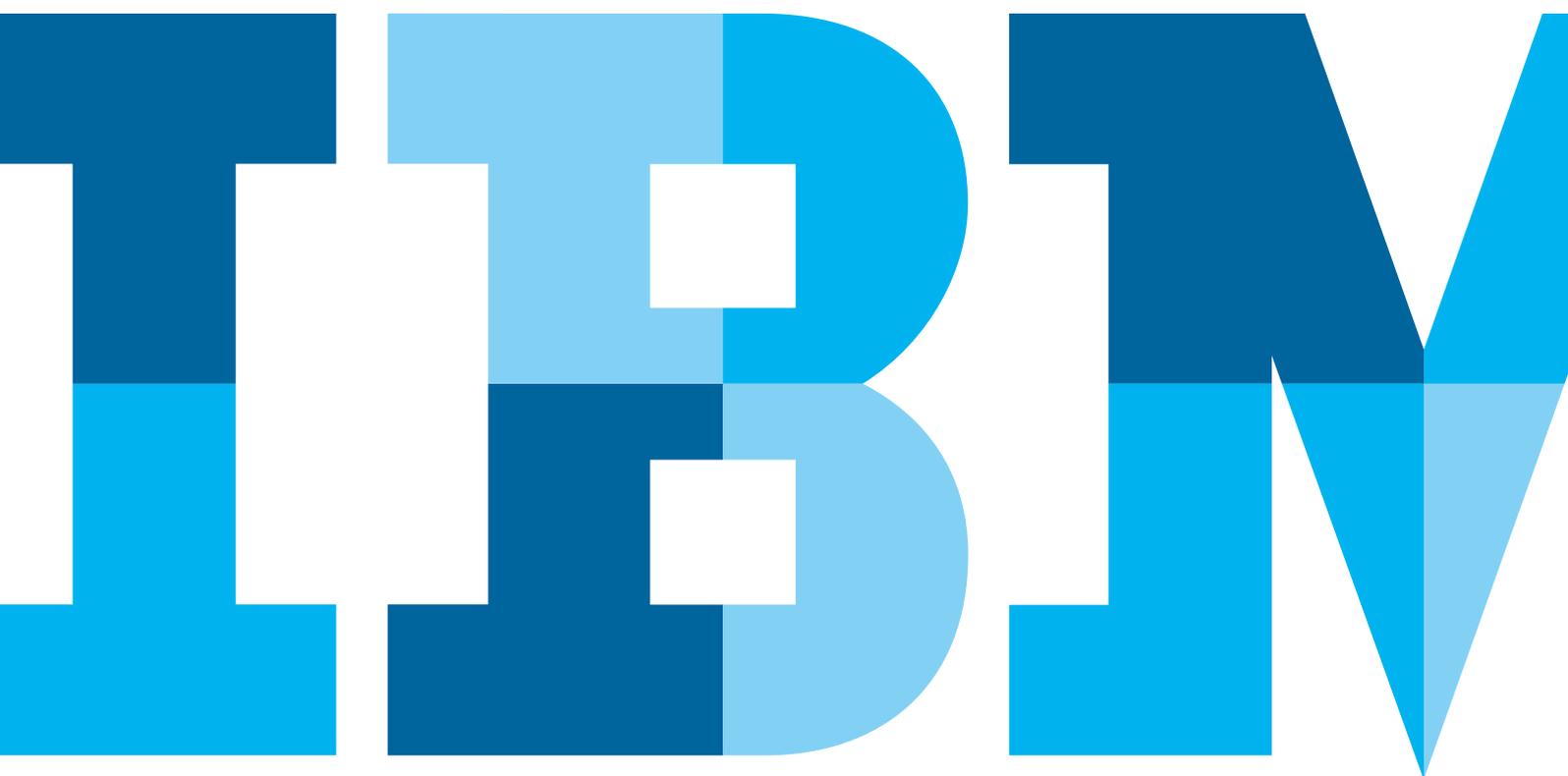


R と SPSS ソフトウェア: 誰もが勝者に



目次

- 2 概要
- 3 R とは
- 3 R が抱える制約
- 4 R と SPSS ソフトウェアの統合がもたらすベネフィット
- 7 結論
- 7 IBM アナリティクスについて
- 7 電話連絡を依頼する

概要

このホワイト・ペーパーの目的は、R プログラミング言語を単独で用いるのではなく、IBM® SPSS® Statistics および SPSS Modeler と併用した場合にもたらされるベネフィットをご紹介することです。SPSS と併用した場合、R のユーザーは、卓越したデータ管理、ポイント・アンド・クリック方式のインターフェース、プレゼンテーションに堪える品質のアウトプット、そしてスケーラビリティの向上を手にすることができます。SPSS ソフトウェアのユーザーは、絶えず拡充されている多彩な統計分析/グラフ作成ライブラリのコレクションにアクセスできます。これらのライブラリは、データからいっそう深い洞察を引き出す上で有用です。

SPSS と R を併せて利用することにより、両者が最大限に能力を発揮します。SPSS ソフトウェアは、SPSS インターフェースから R の構文を実行することができます。SPSS ソフトウェアに搭載されている強力なデータ操作、統計分析、予測のアルゴリズムを独自の R コードでさらに拡充し、機能と柔軟性を強化することが可能になります。カスタム分析を実行し、アウトプットを生成および操作し、新たな洞察を分析計画に統合することが可能です。また、SPSS カスタム・ダイアログ・ビルダーを使用すると、分析にプログラミングを用いることなく新たな分析オプションのベネフィットを活用したいユーザーと、R コードを共有し、再利用することができます。開発者がコーディングに専念できる一方で、ユーザーは分析に専念し、R を習得することなく R ベースの機能を利用できます。

R とは

R は、統計コンピューティングと統計グラフィックスに用いられるオープンソースのプログラミング言語およびソフトウェア環境です (www.r-project.org)。R 言語は、R 言語を用いて統計ソフトウェアを開発する統計学者やデータ・マイニング技術者の間では、ごく一般的な存在となっています。また、高度なデータ分析にも広く利用されています。さらに、線形/非線形モデリング、標準的な統計的検定、時系列分析、分類、クラスタリングなど、多種多様な統計手法とグラフ化機能を使用できます。R は、[Free Software Foundation](http://www.fsf.org) および [GNU General Public License](http://www.gnu.org/licenses/gpl.html) の条件の下で使用できます。

R では、4,800 種超のパッケージを使用することができます。これらは、計量経済学、データ・マイニング、空間分析、生命情報工学をはじめとする多くの要素に特化した、数多くのリポジトリに収められています。R は、各分野の著名な統計学者およびコンピューター科学者によって校閲されています。しかしながら、オープンソースであるために、リリースとパッケージについては体系的なプロセスが存在しません。

R が抱える制約

R は無償で使用できることから、一般的な認識では、商用の統計/モデリング・ソフトウェアの代替として使用することで、組織で大幅な節減を達成できるとされています。この認識は、若干の誤解の種となっています。R は数多くのメリットをもたらしますが、分析からビジネス価値を最大限に引き出す上で、すべての必要な要素を提供できるわけではありません。R ユーザーにとって関心事項となる制約には、以下のものがあります。

- 展開: R を使用して予測のアウトプットを運用環境に統合する作業は、困難なものになり得ます。
- インターフェース: R はモダンなグラフィカル・ユーザー・インターフェース (GUI) を備えていないため、R プログラマーではないユーザーが使用することは困難となっています。
- ラーニング・カーブ: R は、誰にとっても習得しやすいというものではありません。あらゆるユーザーがプログラマーであるわけではありません。
- データ: R は、簡単にはデータベースにネイティブに接続することができません。
- アウトプット: すぐに公開できるかたちでアウトプットを生成することは困難です。
- パフォーマンス: R は、使用可能なすべてのメモリーを瞬く間に使い果たすことがあります。
- コラボレーション: R は、特に R に関するチーム・メンバーの知識レベルが一様ではない場合、分析担当者チームの中で作業を分担することが困難です。
- 企業セキュリティ: ダウンロードするパッケージのセキュリティは保証されません。

また、パッケージおよびアウトプットを取りまとめるには、適切な R パッケージおよびコードを使用していることをユーザーが確認しなければなりません。これらの制約は、多くの場合、R 使用時のコスト増大につながります。

R と SPSS ソフトウェアの統合がもたらす ベネフィット

これらの制約に対処するには、R を SPSS ソフトウェアから使用することが効果的です。両者の強みを重ね合わせると、専門家の数を減らし、短期間で大量に学習することなく R のベネフィットを引き出したい組織のニーズに対応できます。SPSS ソフトウェアは利便性に優れたプラットフォームであり、R のユーザーは、大規模なデータ・セットを処理し、高品質のグラフをはじめとするさまざまな形式のアウトプットを生成できます。その他のベネフィットとしては、SPSS ソフトウェアの使いやすさに加え、統合済みの R パッケージを R に精通していない広範なユーザーに配布できることなどがあります。最も重要な点は、R を SPSS ソフトウェアから使用した場合、R が抱える制約に対処できることです。

展開

SPSS ソフトウェアは、R が抱える展開上の制約に対してコスト効率的に対処できます。たとえば、IBM SPSS Modeler を利用することで、ユーザーは、売上、顧客サービス、メンテナンスなどにビジネス・ルールと予測モデルがどのように作用するのかを最適化できます。SPSS Modeler は、R を組み込んだモデルの使用をサポートしています。SPSS Modeler は、IBM Cognos® TM1® との統合を通じてビジネス・インテリジェンスへのアクセスも合理化します。IBM は、予測インテリジェンスを使用するための多彩な手立てをユーザーの皆様にご提供できます。予測の洞察は、履歴データおよびリアルタイム・データのビューと併せて、ユーザーが対話型のモバイル・ダッシュボードに統合できます。追加のソフトウェアを作成あるいは購入する必要はありません。

インターフェース

SPSS Modeler および SPSS Statistics は、データの準備、統計分析、予測モデリング用のさまざまなアルゴリズムをサポートするシンプルな GUI を備えています。R のコードは、すでに提供されているその他すべての関数および機能とともに、同一のインターフェースで実行されます。また、ユーザーが GUI フロントエンドを R に追加することも可能です。プログラマーではないユーザーは、このフロントエンドを使用して変数名などのカスタム値を指定できます(図1参照)。これらのユーザーは、基盤となっているコードを理解することなく、高度な分析を活用できます。

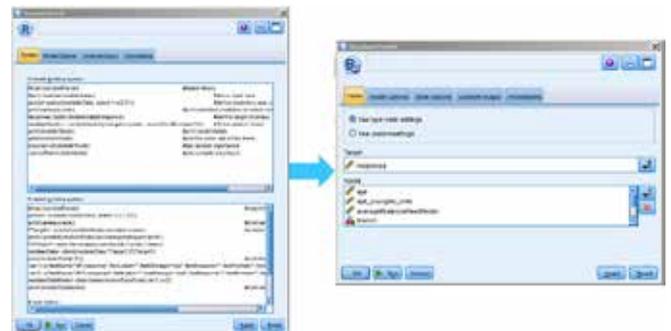


図 1: R のコードを使用してカスタム・ダイアログ・ボックスを作成することができるため、分析の実行中に、プログラマーではないユーザーがコードを活用しやすくなっています。

ラーニング・カーブ

R はプログラミング言語であるため、習得することが必要なものではありません。しかし、R の習得は簡単なタスクではありません。また、R の習得に時間を割いたとしても、R の能力のすべてを活用できるとは限りません。ここで、R を SPSS ソフトウェアと統合した場合、R がもたらすベネフィットをほぼすぐに引き出すことが可能になります。本当の意味でほかにない R ルーチンの習得に専念し、SPSS ソフトウェアにすでに用意されているデータ、統計分析、モデリングを使用することができます。

データ

R を使用する場合、分析に必要なデータにアクセスするには、膨大な時間と労力が必要です。大量のコードを記述し、パッケージを実装し、Java さえも駆使する必要があります。ここでデータ・アクセスの問題を解消するのが、SPSS Statistics および SPSS Modeler です。どちらの製品も、SQL、Oracle、SAP、IBM PureData for Analytics[®]、DB2[®]をはじめとする商用およびその他のデータベースと共に利用できます。SPSS Statistics および SPSS Modeler では、テキスト入力、スプレッドシート、SAS ファイルなどの形式を読み取ることが可能です。SPSS Modeler は、IBM Cognos Business Intelligence および Cognos TM11 からの直接読み取りに対応しています。事前ビルド済みのコネクタを用いるウィザードでデータにアクセスすることにより、データを分析前に抽出、変換、操作するという、多大な所要時間と負荷を伴う作業から解放されます。SPSS Statistics および SPSS Modeler では、ポイント・アンド・クリック方式のインターフェースに集約された、強力なデータ操作技術を利用することができます。ユーザーは、データを転置し、チェックして形式を変換することができます。SPSS Modeler では、データを予測モデリング用に 1 クリックで最適化する自動データ準備機能を実行することも可能です。

アウトプット

SPSS Statistics および SPSS Modeler には、プレゼンテーション対応のチャートとグラフを生成するための手段が数多く用意されています。SPSS Statistics ソフトウェアでは、R プログラマーが R の関数を SPSS ソフトウェアの構文にラップすることができます。結果として、プレゼンテーション品質のグラフ、ピボット・テーブル、その他の形式のアウトプットを生成することが可能になります(図 2)。また、ユーザーは、PDF、Word、PowerPoint、Excel などの一般的な形式で結果を簡単に出力できます。さらに、エンティティ分析、ソーシャル・メディア分析、テキスト分析をはじめとする SPSS Modeler の高度な機能を R に統合することも可能となっています。

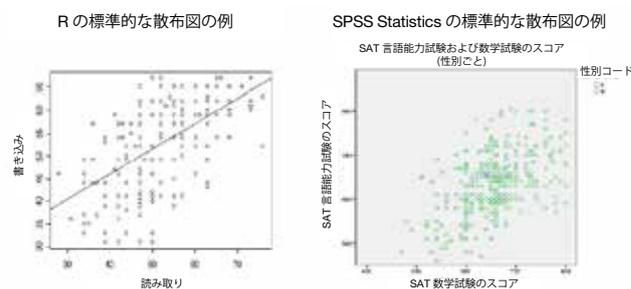


図 2: SPSS は、プレゼンテーション対応のチャートおよびグラフを生成することにより、ビジネス関係者が時間を取られることの多い、分析のアウトプットの準備と整形の時間を短縮します。

パフォーマンス

R のコマンドは、インメモリー・ワークスペースで運用されます。R のコマンドはメモリーを考慮した設計になっておらず、R セッションの進行中は、すべてのオブジェクトが一時的なワーキング・メモリーに保存されます。結果として、相当に早い段階で使用可能メモリーが使い果たされることがあります。R を SPSS Modeler と組み合わせて使用した場合、R に渡されるデータをパーティション分割または標本化することができます (図 3)。また、SPSS Modeler Server はメモリーを徹底的に活用するテクノロジーであり、分析をディスクに移転して、メモリーを使用可能な状態に維持できます。データのマイニングおよびモデリングの全体的なパフォーマンスに大きな影響を及ぼすことなく、R のコマンドを使用して、R のオブジェクトを作成することが可能です。さらに、SPSS Modeler では、たとえば IBM PureData for Analytics や SAP Hana の環境用に R をインデータベースでスケールリングできます。SPSS Analytic Server を使用する場合は、R を Hadoop でスケールリングすることも可能です。

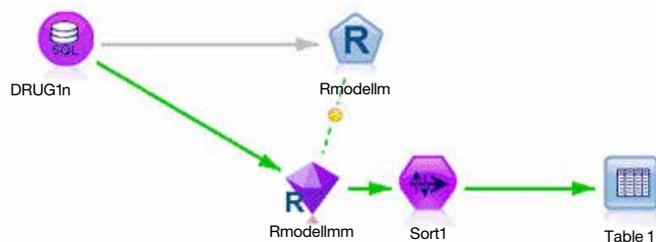


図 3: IBM SPSS Modeler は、レコードのスコアリングに使用される R モデルを用いて、SQL データベース内のデータにネイティブ・アクセスするために使用されます。紫色の部分は、分析ステップで SQL プッシュバックを使用していることを示します。プッシュバックによって、メモリー負荷の高いタスクをデータベース自体で実行することにより、パフォーマンスが向上します。

コラボレーション

R は 1 人で使用するものとして設計された言語です。その一方で、ほとんどの分析処理は、モデルや統計分析に関与する多数の人々による共同作業となっています。SPSS ソフトウェアを R と組み合わせて使用すると、単独での作業という側面を R から取り除いて、IBM が提供する世界屈指のコラボレーション機能を活かせるようになります。SPSS ソフトウェアでは、SPSS Statistics および SPSS Modeler で開発した分析資産を集約、セキュリティー保護、自動化するための網羅的なフレームワークが提供されます。したがって、開発されるモデルおよび統計分析をセキュリティーで保護した上で共有し、企業がバナンスを適用できることが保証されます。

企業セキュリティー

R には正式なリリース・プロセスが用意されていません。つまり、正式な品質保証 (QA) が組み込まれていません。QA を実施しない場合は、R を自らの責任で使用することになります。リリースまたはパッケージに実装されていると言明されている機能が提供されていることについて、確証を得ることはできません。また、R のパッケージは Web およびユーザー・コミュニティからダウンロードされるものであるため、パッケージのセキュリティーも問題になることがあります。事実、開発元について何も知らなければ、マルウェア、トロイの木馬ウイルス、データ盗聴メカニズムなどがダウンロードを通じて入り込む恐れがあります。一方、SPSS Modeler および Statistics は、IBM のソフトウェア QA プロセスの一環として厳格な検証が実施されています。SPSS ソフトウェアは開発元が IBM であるため、自社環境のセキュリティーを脅かすものになり得る、リスクを抱えた手順を適用する必要はなくなります。

結論

SPSS および R がそれぞれ誇る強みは、時間をかけて検証されてきたものであり、分析コミュニティで熱烈に支持されています。さらに、これらの強みが互いに補完し合うことで、分析コミュニティの全体にわたってベネフィットをもたらす、いっそう強力な数々の関数および機能が実現されています。R のユーザーは、より大規模なデータ・セットの処理を可能にする卓抜したデータ管理機能にアクセスできます。また、SPSS ソフトウェアを使用する R ユーザーは、一連の充実したグラフ/ピボット・テーブル出力オプションを駆使して、ユーザー体験の向上につなげることが可能です。つまり、SPSS ソフトウェアは、R のパッケージを広範なユーザーに配布する上で理想的な開発手段として機能します。

SPSS ソフトウェアのユーザーは、通常よりもはるかに多くの統計機能にアクセスできます。R をはじめとする複雑なプログラミング言語を習得するという煩わしさを味わうことなく、複雑な分析を実行することが可能です。R と SPSS ソフトウェアの併用を通じて得られるベネフィットについて、この機会にぜひご検討ください。

IBM アナリティクスについて

IBM アナリティクスのソフトウェアがもたらすデータ主体の洞察は、組織をよりスマートに運営し、競合事業者よりも優位に立つ上で有用なものとなります。このソフトウェアの包括的ポートフォリオには、ビジネス・インテリジェンス、予測アナリティクス、意思決定管理、パフォーマンス管理、リスク管理のソリューションが含まれています。IBM アナリティクスのソリューションをご利用いただくことで、業績に大きく影響し得る顧客アナリティクスなどの領域において、傾向とパターンを識別し、視覚化することが可能になります。これらのソリューションでは、シナリオを比較し、潜在的な脅威およびチャンスを予想し、リソースのプランニング、予算編成、予測を改善し、予想される成果とリスクのバランスを維持し、規制要件への対応を図ることができます。アナリティクスが広範にわたって利用可能となることで、ビジネス目標の達成に向けた短期および長期の意思決定に整合性を持たせることができます。詳細については、ibm.com/analytics/jp/ja/ をご覧ください。

電話連絡を依頼する

電話での連絡をご希望の場合、またはご質問がある場合は、ibm.com/contact/jp/ja/ をご参照ください。IBM 担当者が対応いたします。



© Copyright IBM Corporation 2016

日本アイ・ビー・エム株式会社
〒103-8510
東京都中央区日本橋箱崎町 19-21

Printed in Japan
July 2016

IBM、IBM ロゴ、ibm.com、Cognos、PureData、DB2、および SPSS 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、www.ibm.com/legal/copytrade.shtml をご覧ください。

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

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。

Java およびすべての Java 関連の商標とロゴは、Oracle やその関連会社の商標または登録商標です。

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。

本資料は発行日における最新情報を提供し、いかなる時点においても IBM によって変更される可能性があります。すべてのサービスがその他の国においても提供されるとは限りません。

本資料の情報は「現状のまま」で提供され、明示的にも黙示的にも、商品性の保証、特定目的への適合性の明示的保証、違反行為がないことを含むいかなる保証を行うものでもありません。IBM 製品に対しては、当該製品が準拠する契約書の契約条件に基づいて保証されます。



Please Recycle