

# KEYWORDS

特集 **1** オープンソース・ソフトウェアも活用できる  
**分析プラットフォームとしてのSPSS**  
 P.02 - P.04

特集 **2** 研究奨励賞復活!  
 データサイエンティストへの道  
**SPSS Datathon 2016**  
 P.05 - P.06



特集 **3** セミナーレポート  
**【SPSS♥ナイト】**  
**ビジネス成果を出すための  
 データ分析活用講座** P.07 - P.09

**教えて! データサイエンティスト** P.10 - P.11

SPSS Statisticsで統計解析をはじめよう P12

IBM SPSS研修スケジュール P13

SPSS研修コース対応表 P14

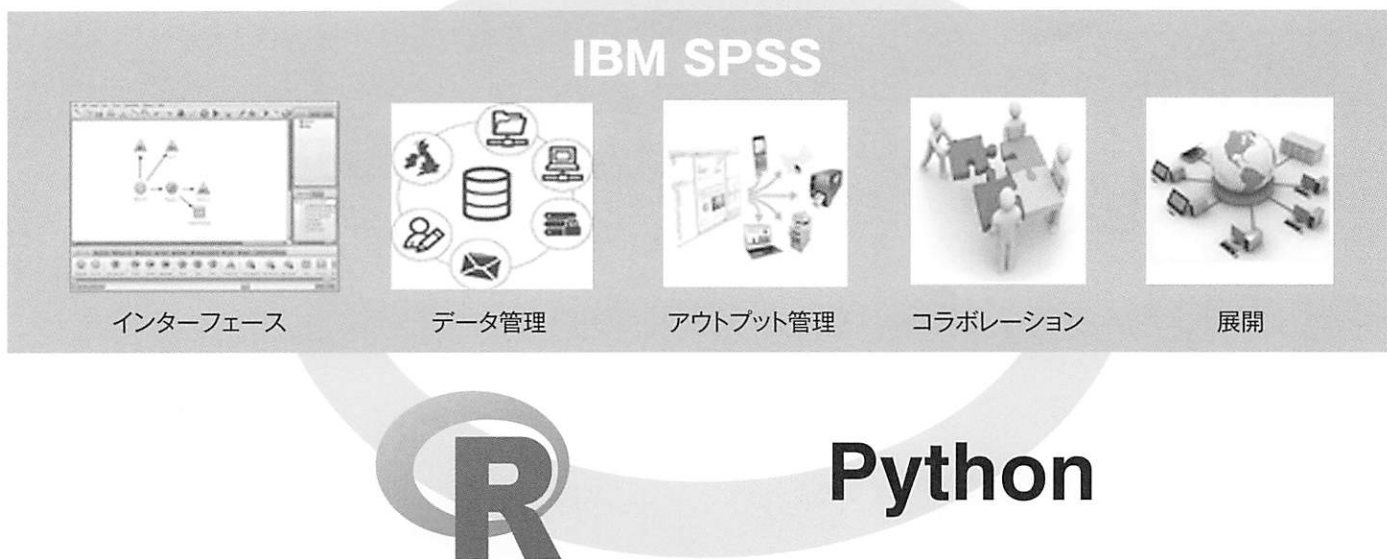
書籍情報 P15 - P.16



# 1 オープンソース・ソフトウェアも活用できる 分析プラットフォームとしてのSPSS

RやPythonといったオープンソース・ソフトウェアでは、次々と新しい分析手法が誕生しています。一方で、データ・ソースにアクセスしたり、データを加工して準備したり、試行錯誤の分析をしたりするには、SPSSのようなツールを利用することで、より効率的に作業を進めることができます。

2016年3月にリリースされたSPSS Statistics V24およびSPSS Modeler V18では、RやPythonの拡張機能をSPSSから呼び出せるようになったほか、RやPythonを使ったカスタム・ダイアログ・ボックスの作成もより簡単になりました。これにより、SPSSを分析プラットフォームと位置づけ、必要なところでRやPythonを呼び出し活用することができます。



RやPythonへの連携には、以下のような方法があります。

## IBM Predictive Analytics Gallery (拡張ハブ)

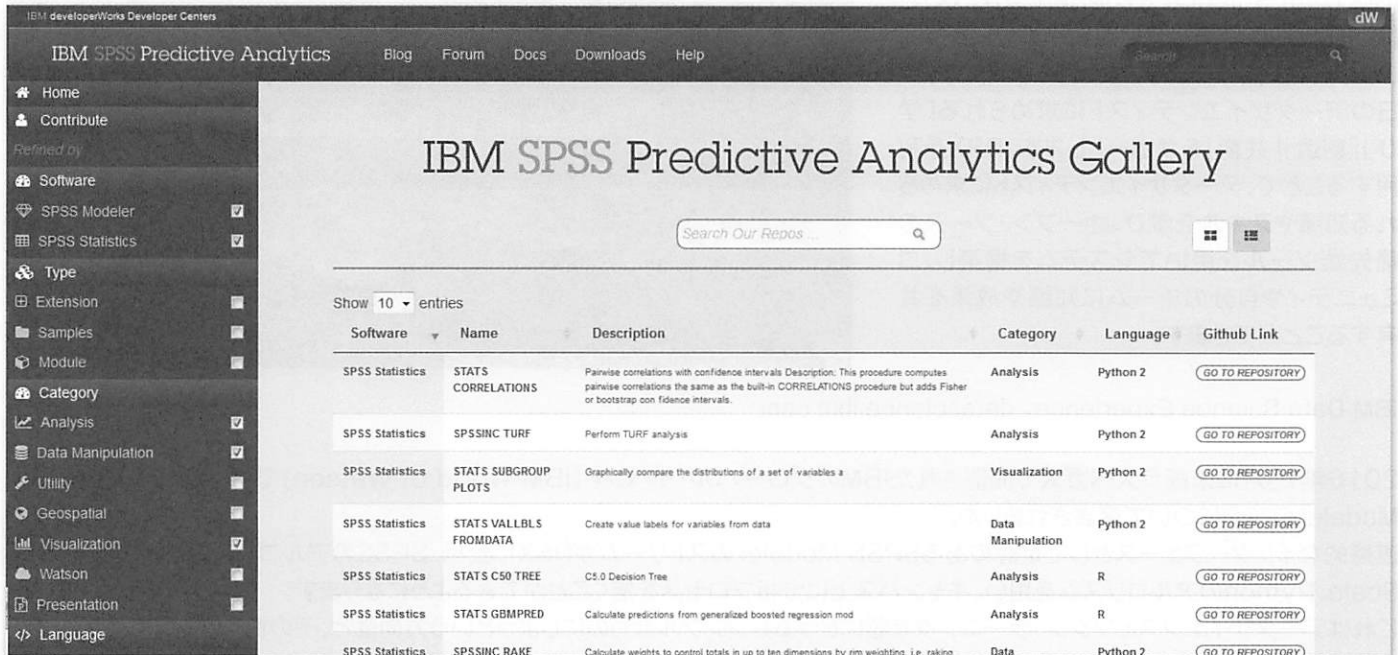
- IBM Predictive Analytics GitHubのGalleryには、SPSSを拡張するためのさまざまな拡張機能が掲載されています。
- SPSSから直接、GitHubに上がっている任意の拡張コマンドを呼び出したり、カテゴリ内の拡張コマンドを検索することができます。ブラウザからGitHubにアクセスし、ダウンロードする必要はありません。
- 見つけた拡張コマンドは、そのままSPSSにダウンロードし、インストールすることができます。拡張コマンドのファイルをダウンロードして任意のディレクトリに保存する必要はありません。
- RやPythonの分析手法やビジュアライゼーション、ユーティリティに加え、位置情報に関する分析手法やWatson APIへのリンク、JSONファイルのインポートなど、さまざまな拡張コマンドが利用可能です。





IBM Predictive Analytics Galleryには、さまざまな拡張機能が載っています

▶ <http://ibmpredictiveanalytics.github.io/>



## カスタム・ダイアログ・ビルダー

- RやPythonプログラムを用い、SPSSのカスタム・ダイアログボックスを作成できます。
- HadoopやAnalytics Serverを使うことなく、Pythonを実行するための拡張機能の作成も可能です。
- 拡張機能は、Sparkのマシン・ラーニング・ライブラリーMLlibや、NumPy、SciPy、scikit-learn、PandasといったPythonのライブラリーにアクセスすることができます。
- 作成した拡張機能を他のSPSSユーザーと共有することができるため、プログラミングをしない、あるいはプログラム・コードを理解しないユーザーも、オープンソース・ソフトウェアで作成されたこれらの機能を利用することができます。



RやPython for Sparkのスクリプトを実行するには、お使いの環境に、互換性のあるRやPythonのインストールが必要です。RのスクリプトをSPSSで実行するには、事前にプラグイン(無料)のインストールが必要です。下記よりダウンロードください。

▶ <http://www.ibm.com/developerworks/spssdevcentral>

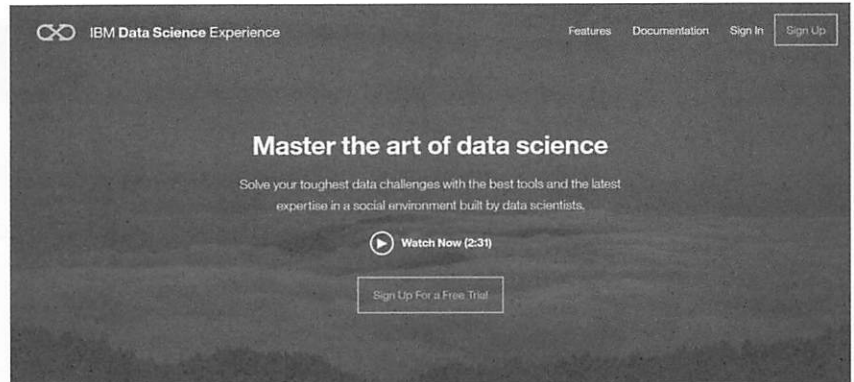
Python/SparkスクリプトをSPSS Modelerで実行するには、Python インストール済み環境を使用するように IBM SPSS Modeler サーバー (または IBM SPSS Modeler クライアントのローカル・サーバー) を構成する必要があります。詳しくは下記をご参照ください。

▶ [http://www.ibm.com/support/knowledgecenter/ja/SS3RA7\\_18.0.0/modeler\\_r\\_nodes\\_ddita/clementine/r\\_pyspark\\_api.html](http://www.ibm.com/support/knowledgecenter/ja/SS3RA7_18.0.0/modeler_r_nodes_ddita/clementine/r_pyspark_api.html)

# 1 オープンソース・ソフトウェアも活用できる 分析プラットフォームとしてのSPSS

## おまけ: SPSS Modeler と IBM Data Science Experience

IBM Data Science Experience (DSX) とは、IBMが2016年6月に発表したデータサイエンティスト向けのクラウド・サービスです。今日のデータサイエンティストに求められる「学び」「創造」「共創」をサポートします。DSXを利用することで、データサイエンティストに求められる知識やスキルを学び、オープン・ソースの最先端ツールを用いてシステムを構築し、コミュニティや自分のチームに知識や成果を共有することができます。

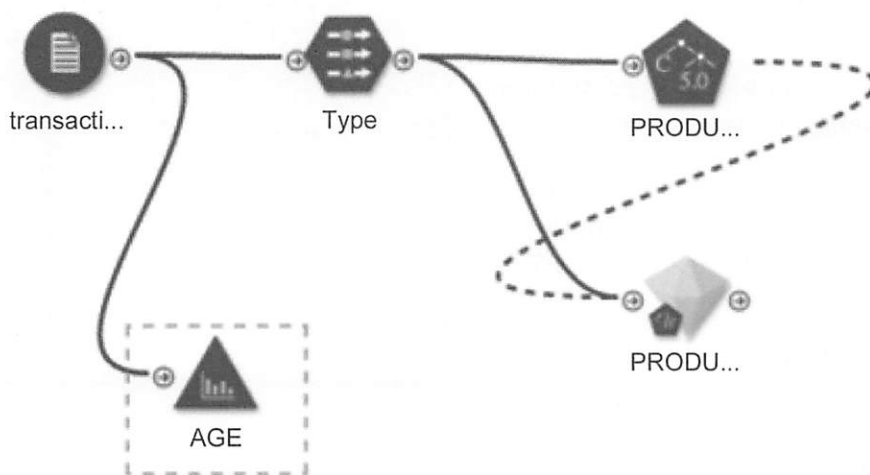


IBM Data Science Experience: [datascience.ibm.com](http://datascience.ibm.com)

2016年10月に米国ラスベガスで開催されたIBMのグローバル・イベント『IBM World of Watson』では、このDSXとSPSS Modelerの統合について発表されました。

直感的なインターフェースとして定評のあるSPSS ModelerのストリームがDSXに登場。SPSSのアルゴリズムだけでなく、RやScala、Pythonのアルゴリズムを用い、キャンバス上に分析プロセスを描くことができるようになります。

これはデータサイエンスとマシン・ラーニングを誰にとってもシンプルなものにしようという方向性と、その第一歩を示しています。GUIを使った分析プロセスの構築を可能にすることで、プログラミング・スキルのないビジネス・ユーザーにもオープンソース・ソフトウェアを活用した分析のケイバビリティを提供するだけでなく、プログラミングに長けたデータサイエンティストを含むすべてのユーザーが分析プロセスの設計やモデルの開発を効率化でき、データからパターンや洞察をいち早く見出せるようサポートします。



SPSS Modeler in Data Science Experienceのベータ版をご希望の方:  
以下のURLより、Waiting Listにご登録ください。

SPSS Modeler in Data Science Experience Create Machine Learning models without coding.  
[datasciencex.typeform.com/to/GVTA2n](http://datasciencex.typeform.com/to/GVTA2n)

IBM SPSSを使った統計コンテスト SPSS Datathon 2016。12月3日(土)の最終選考会にてフィナーレを迎えました。本コンテストは、教育機関及び医療機関の皆様を対象に、基礎的・応用的研究における更なる学術的發展を奨励する事を目的としています。SPSS Statistics Entry部門、SPSS Statistics 論文部門、SPSS Modeler Challenge部門の3つの部門に計48組のお申し込み、26組の成果物の提出をいただきました。

	SPSS Statistics Entry 部門	SPSS Statistics 論文部門	SPSS Modeler Challenge 部門
テーマ	問いません。 統計解析ソフトSPSS Statisticsを使い、論文やレポート、研究で、あなたの主張を支えるためにデータをどのように分析したのかをまとめ、ご応募ください。	問いません。 (学会などで発表する論文、自身で著作権を有する論文で過去の論文の引用や再分析したものも可)	『ID付POSデータに予測分析手法を用い、具体的な顧客販促施策の応用例を提案する』 ある小売店舗における一年間の購買データを分析し、例えばクラスタモデルを用いた適時レコメンデーションなど、データに基づいた販促施策を提案してください。
分析ツール	SPSS Statistics		SPSS Modeler
応募資格	学部生以下の学生 (高校生、高専生の応募可)	教育機関の在学学生、教員及び研究者(研究室等を含む) 医療機関の研究従事者(チームによる応募可)	

## SPSS Datathon 審査結果

### ■ SPSS Statistics Entry 部門

**グランプリ**



慶應義塾大学  
小野 梓さん、望月沙也佳さん、  
山中 昌玲さん、金城 博男さん

『ロングセラーブランドのリバイタリゼーション』

講評:企業がすぐに使えるような分析でした。発表のチームワークもよくわかりやすいプレゼンでした。

**準グランプリ**



同志社大学  
横田 貴仁さん、滑 健太さん、  
長壁 哲平さん、野田 憲冬さん

『臓器提供の意思表示における定性・定量分析』

講評:医療の世界で重要なテーマである行動変容、意思変容という重要なテーマへのチャレンジで興味深かったです。

**優秀賞**



青山学院大学  
吉田 真美さん、齋藤 珠希さん、  
新城 真由さん、小澤 拓実さん

『リーダーシップの機能的役割分化に関する定量的研究』

講評:テーマ、分析内容ともに難しいものに関心を持ってチャレンジしていました。

### 審査員 講評



**清水 聡**

慶應義塾大学商学部 教授

データを与えられて分析する場合でも、自分でデータを集めて分析する場合でも、仮説を立て、その仮説を説明するにはなにをすべきかを考えてから、分析に着手することがとても重要です。そのためには、グラフ化や記述統計でデータを理解したり、過去の研究を調べる必要があります。



**高橋 理**

聖路加国際病院 一般内科 医長  
聖路加国際大学 教授  
臨床疫学センター センター長

データを分析し適切に解釈するためには、データのソースや母集団、データの意味を、正しく理解しておくことが重要です。医療の世界でも大量のデータが貯まってきています。皆さんのスキルを医療の世界にも活かして欲しいと期待しています。



■ SPSS Statistics 論文 部門

**グランプリ**



金沢市立病院  
大坪 尚典さん

『当院地域包括ケア病棟の利用有無による影響調査』

講評: サンプル数が少ない中、結果を出そうとチャレンジされていました。今後、症例数を増やし、病気や怪我による違いをぜひ分析してみてください。

**準グランプリ**



北海道釧路市立音別小学校  
大野 昌宏さん

『ICT活用における教員の指導の在り方に関する実態研究』

講評: 緻密なアンケートをしっかりと回収した上で分析をされていることに感動しました。管理職の研究もされているとのこと、その分析も期待しています。

**優秀賞**




阿波市立土成中学校  
三橋 和博さん

『社会参画する力を育む生徒会活動』

講評: 変数をいろいろ入れ替えて因子分析に取り組んでいました。学生のモチベーションが上がっていく順番に沿っての分析となればさらに面白かったと思います。

■ SPSS Modeler Challenge 部門

**グランプリ**



筑波大学  
松本 健さん

『ID付POSデータを用いた優良顧客維持のためのレコメンデーションの作成』

講評: さまざまな手法を使い、最も高度でチャレンジングな分析を実施していました。

**準グランプリ**



福岡大学  
川原 優矢さん、萩尾 真広さん、茶木 綾音さん、  
桧枝 紫穂さん、白石 智紀さん、重松 勇輝さん

『PB商品に着目した分析 ~スイッチ分析からPB商品のあり方を考える~』

講評: 企業にとって重要なPB商品に着目した分析でした。PB商品でもカテゴリによって違うところが難しかったかもしれません。

**優秀賞**



日本英語検定協会  
夏目 泰秀さん、渋谷 佳樹さん

『顧客性別の推測による販促施策立案』

講評: お客様のクラスタリングは現場でも有効な分析です。最大のクラスターは最も重要で、これを深掘りできるとさらに面白かったです。

審査員 講評



松田 崇

株式会社マツモトキヨシホールディングス  
経営企画次長 兼 オンラインビジネスユニットマネージャー  
株式会社マツモトキヨシ 運営企画部部长

理論値と現実の施策の結果は大きく異なります。その差はなぜか、どうやって埋めるかが、学生の皆さんがビジネスの世界に進んだときの仕事となるでしょう。皆さんの実践の場での活躍を期待しています。



西澤 英子

日本アイ・ビー・エム株式会社  
アナリティクス事業部 ITスペシャリスト

大変興味深く拝見しました。時間的制約に加え、データや分析環境の準備などのご苦労もあったと思います。どのチームも分かりやすくまとめ、簡潔な発表でした。次の分析プランをお持ちの方がいらっしやるのも印象的でした。SPSS Datathonは更なる機会に繋がる第一歩をスタートしたのだと強く感じています。

自分が使っているデータや分析手法は最適なのか自信がない、他の分析者はどのような分析をしているのだろうか、といった疑問や不安を抱えている方、先進企業がどのようなデータソースを利用して、どのような分析を行い、分析結果からどのようなアクションに結びつけているのかをもっと知りたい、とお考えの方を対象にしたセミナーを実施しました。

## 第1回『ビッグデータ分析のための必須知識』

2016年11月18日(金)

### 成功するデータ分析 - はじめが肝心。

#### プロジェクトを成功に導くための、ビジネス理解とデータ理解

日本アイ・ビー・エム株式会社 アナリティクス事業部 SPSS ITスペシャリスト 西牧 洋一郎

CRISP-DM(P.9参照)は、古いかもしれませんが、基本的で重要な方法論です。いきなりモデル構築に着手してしまい、うまくいかないという例は多く聞きます。モデリングの前に、解決したいビジネス課題は何か、課題を解決するためにどんなデータが必要か、データをどう加工するかよく考えてみてください。業務とデータをよく理解することで、どんなモデルをどうやって作るかが見えてきます。

### Hot Topic - 「人工知能」時代における アナリティクスの役割

日本アイ・ビー・エム株式会社 アナリティクス事業部 SPSS ITスペシャリスト 西牧 洋一郎

「人工知能」によって、人によるデータ分析はますます重要なものになるでしょう。人工知能が得意とする処理がある一方、人ではないとできない分析があるからです。人工知能は質問を理解し、膨大な情報源から確信できる候補を自動で応答します。これに対し、人による分析では過去のデータを要約し、将来の推測値を求め、業務の自動化や最適化を実現していきます。これには、予測結果とビジネス・ルールを調和させたり、理論上の最適解とビジネス戦略を比べて採るべきアクションを決めたり、施策精度を上げるために取り込むべき新しいデータを検討したりと、人ならではの判断が必要になります。



データサイエンティスト・オブ・ザ・イヤー 2014  
渋谷 直正 氏

### SPSSユーザー講演

#### 「SPSS Modelerを使って自分で分析ができるマーケッターになろう」

日本航空株式会社 Web販売部 1to1マーケティンググループ アシスタントマネジャー 渋谷 直正 氏

マーケティングはお客さまの購入回数や訪問回数を増やしお得意さまになっていただくためのものです。データを分析するだけでなく、これを施策に活かさなくてはなりません。Webログデータからは、商品・サービスを購入したお客さまとそうでないお客さまの行動の違いが見え、ここから知見を得たり、施策を考えたりすることができます。施策につなげるには、業務知識やビジネス意識、企画力などがなにより重要です。統計や数学、高度な分析手法のすべてを身につけなくても、本当に必要ないくつかの分析手法を習得することで、お客さまをよりよく理解し、実行可能な施策へとつなげることができます。

渋谷氏は「教えて!データサイエンティスト」(P.11)にも登場します

### 参加者の声

- いつものビジネスセミナーとは雰囲気も客層も違い、新鮮でした。
- どこから手をつけたいか・・・という状態で来たので、分析にはまず仮説からというお話をはじめ、多くのヒントをもらうことができました。
- よくデータ分析の研修で言われているビジネスの理解が大事ということが、データサイエンティストから聞いてよかった。
- 純粋に分析の流れや事例が聞いて楽しかった。
- 実践的な内容で役立った。
- 業務理解があくまでベースになっているということに納得がきました。
- あらためて分析の重要な観点を認識することができました。
- SPSS ModelerのTipsを期待





## 第2回『スマートなデータ準備で分析のスピードとクオリティをアップする』

2016年12月6日(火)

## 成功するデータ分析 - 最も重要で、最も面倒なデータ準備を効率化する

日本アイ・ビー・エム株式会社 アナリティクス事業部 SPSS &amp; BI テクニカルセールス 佐藤 晴哉

データ準備には、分析全体の8割の時間がかかります。分析精度に大きく影響する重要なステップです。どんなデータや業種にも共通というものがあられるわけではなく、業務への実装を念頭に行う必要があります。モデルの精度を確保できるような入力を作ることは難しいポイントで、一度で高い精度を実現できるわけではありません。探索的にアプローチすることが成功の秘訣です。



## Hot Topic - Twitter分析からビジネスのために何が導き出せるか

日本アイ・ビー・エム株式会社 アナリティクス事業部 SPSS ITスペシャリスト 西牧 洋一郎

Twitterデータを無料で自由に分析できるとしたら、何をしますか。Tweetにはさまざまな価値ある情報が含まれています。ユーザーの嗜好や関心から360度プロフィールを作成したり、パーソナリティを分析することも、いまや可能です。拡散のメカニズムを分析し、炎上の予知もできるようになるかもしれません。Tweetに含まれる地名や、使われているWiFiスポットから、行動エリアを特定し、ユーザーの行動エリアと関心に基づき販促施策を展開することも、実現目前となっています。



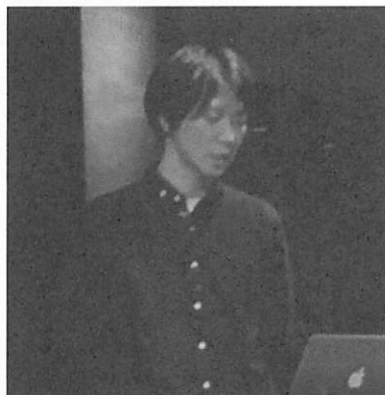
## SPSSユーザー講演

株式会社グラフ 代表取締役 原田 博植 氏

10月にリクルートライフスタイルから独立し起業しました。アジア展開を図り、フィリピンにも現地法人を開設。また、自身が主宰する、インハウスのデータサイエンティスト・コミュニティ「丸の内アナリティクス」を11月に社団法人化しました。分析業務の健全な内製化と標準化の促進、経済の発展への貢献を目的としています。

シンクタンク、外資系ITベンチャーなどを経て、最新テクノロジーやIT、データ分析といったさまざまな経験を積み、リクルートライフスタイルではEC事業等で改善施策をドライブしました。事業を理解した上で分析をビジネス施策に落とし込む、データサイエンスの「事業装着」が重要だと考えています。

赤の女王仮説「その場にとどまるためには、全力で走り続けなければならない」を信条に、自ら機会を作り出し機会によって自らを変えていくことを目指して、日々取り組んでいます。



データサイエンティスト・オブ・ザ・イヤー 2015 原田 博植 氏

## 参加者の声

- 普段自分がやっていることがたくさん出てきました。自分なりに整理することをやってみたいと思いました。
- SPSSの簡単な使い方を交えながら準備について具体的に話していただき参考になりました。
- 準備が8割という言葉は知りつつ、データの準備についてはあまり分かっていなかったのが、整理できて大変助かりました。
- ちょうどテキストマイニングに取り組んでいるところなので、少しでも今日のお話を活かしたいです。
- 多くの事例をご紹介いただき大変参考になった。
- 現場のアウトプットのイメージが出来て参考になりました。
- 全編通して、とても興味深い内容で勉強になりました。

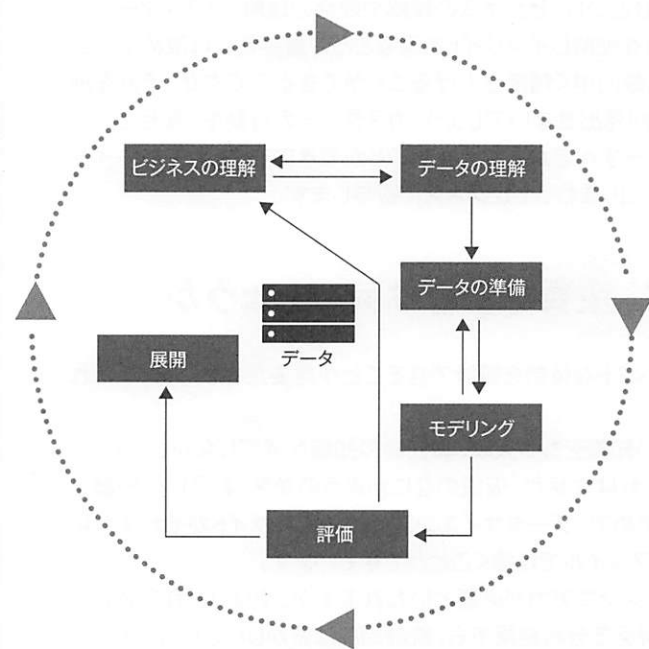




## おまけ：成功するデータ分析のためのメソッドロジー—CRISP-DM

CRISP-DM (CRoss-Industry Standard Process for Data Mining) は、企業向けにデータマイニング・プロセスを定義したもので、データマイニングを活用して優れた意思決定を行い、収益の増加やプロセスの効率化といった大きく具体的な成果を実現するのに役立ちます。

データマイニング・プロジェクトから期待どおりの結果を得るために、自社のビジネスの状況を的確に見極め、データマイニング・ツールがデータ・マイニング・プロセス全体をとおしてどのように機能するのか、このCRISP-DMを使って評価します。CRISP-DM ではデータマイニング・プロセスは6つのフェーズで構成されています。各フェーズの順序は厳密ではなく、異なるフェーズ間を常に行き来することが求められます。データマイニング・ツールを評価する際には、データマイニングのこの対話的な側面を念頭に置いておきます。企業は、プロジェクトをとおしてどのフェーズでも変更が可能な柔軟なツールを選択する必要があります。



### フェーズ①：ビジネスの理解

データマイニングで何を見つけだそうとしているのかをあらかじめ調査しておく必要があります。調査には、関係する人々をできるだけ多く集め、打ち合わせた内容をきちんと文書に記録しておくことをお勧めします。CRISP-DM フェーズの最終段階では、ここで集めた情報を元にどのようにプロジェクト計画を作成するかを議論します。

### フェーズ②：データの理解

データマイニングに利用できるデータを詳しく調査します。次のデータの準備フェーズは一般的に最も時間のかかる場所ですが、そこで予期しない問題が発生するのを避けるためにも、このステップでの作業が重要になります。

### フェーズ③：データの準備

データマイニングにおける最も重要な作業で、しばしば時間がかかります。実際、一般的なデータの準備作業は、プロジェクトに費やされる時間や労力の50~70%を占めると見積もられています。初期のビジネスの理解およびデータの理解フェーズに十分な労力を費やしていれば、このオーバーヘッドを最小限に抑えることができます。しかし、それでもデータの準備と事前処理にはかなりの時間を割く必要があります。

### フェーズ④：モデリング

分析ツールにデータを取り込み、そこから得られる結果から、ビジネスの理解フェーズで提起された問題を解決するための手がかりを得ます。

一般的にモデリングでは、繰り返し反復作業が行われます。データマイニングの担当者は、デフォルトのパラメータを使ってさまざまなモデルを実行し、次にパラメータを細かく調整したり、データの準備フェーズに戻って、選択したモデルに必要な操作を行います。単一のモデルや、1回モデルを実行しただけで満足する回答が得られることは滅多にありません。

### フェーズ⑤：評価

プロジェクトのはじめに作成したビジネス上の成功基準を使って、結果を評価する必要があります。これは、組織で結果を有効活用できるようにするためにも必要不可欠な作業です。

### フェーズ⑥：展開

展開は、新しい洞察を使って組織の改善を行うプロセスです。あるいは、データマイニングで得られた洞察を使って、組織の変更を実施することもあります。



# 教えて！ データサイエンティスト vol.2 vol.3

企業で活躍するデータサイエンティストに、気になるあのこと、このことについて聞いてみました



株式会社リクルートライフスタイル  
データサイエンティスト ネットビジネス本部  
プロダクトマネジメントユニット データサイエンスチーム **松本 健氏**

外資系医療機器メーカー、リサーチ会社を経て2007年リクルート入社。2012年10月、リクルートの分社化に伴い設立されたリクルートライフスタイルへ。データサイエンティストとして、じゃらんやホットペッパーグルメなどの日常消費領域のサービスに関する分析を提供している。

## Q1. データ分析はビジネスにどのように貢献するのでしょうか

データ分析の活用目的は大きく2つです。ひとつは直接的に売り上げに貢献すること。カスタマーへのレコメンドなどの施策の効果を1%改善するだけでも、大きなインパクトになります。もうひとつは、ビジネスの理解や競合の理解、カスタマーのインサイトから、そこにロジックを見出すことです。カスタマーの本質を理解しインサイトを得ることの重要性には改めて注目しています。マシンラーニングを使ったレコメンド・エンジンでは比較的早く精度を上げることができるのですが、それを施策に組み込んでしまうと、例えばうまく結果が出ないときに改善策が見出せないでしょう。カスタマーの行動や、なぜそのような意思決定をするのかを理解することで、改めて、どうアプローチすべきかを考え直すことができます。カスタマー・インサイトへの取り組みは、ビジネスに直結するものではありませんが、回りまわってビジネスに寄与します。

## Q2. データ分析を成功させるために、重要なことは何でしょうか

その組織にいる人材と組織の状態によって、試行錯誤をしながらベストな体制を設計できることが重要だと思います。これはすべての組織でできるわけではないかもしれません。

成功事例を振り返ると、データ分析者とビジネスのその分野に深い知識をもつ人が、お互いの知識を補完しながらうまくコラボレーションができていました。施策がうまくいかないときに、それはなぜか、仮説のなにが違うのかを、お互いの知識を動員し紐解いていき、少しずつ良くしていきました。片方だけではだめで、データサイエンスとビジネス・サイドがそれぞれに長けていて、この両輪がうまく回るという体制が、リクルートライフスタイルでは築くできています。

データサイエンティストには、データサイエンス力、ビジネス力、エンジニア力が必要といわれますが、やはりどれも必要です。最近ツールで簡単に分析できますが、一見うまくいくように見える分析結果でも、統計的にはおかしいということもあります。これをビジネスに取り込んでしまうのはリスクがあるでしょう。得意な領域を伸ばしつつ、広く深く知識を身につけていくべきだと思います。引き出しを多く持つことで、解けるか解けないかわからないような課題に対しても、アプローチを考え解いていくことができ、それがほかの人にはないバリューとなるでしょう。

## Q3. AI(人工知能)テクノロジーの適用が進む中、これからのデータサイエンティストには何が求められるのでしょうか。

モデルを作る部分はAIに任せ、データサイエンティストをはじめとする人間は、どのような変数をどう使うか、どのような施策を打つかに集中しようという世界を描く方たちもいますが、モデル作りをAIに任せっきりにするのではなく、人がなぜそのような行動をとるのか、どうしてそう意思決定するのかについては、人間がその本質を理解し、AIを補完していくのが良いのではないかと思います。





## 日本航空株式会社 Web販売部 1to1 マーケティンググループ 渋谷 直正 氏

2002年 日本航空に入社。2009年よりWeb販売部 1to1 マーケティンググループに所属。月間2億ページビューに上るWebページのログ解析や顧客情報の分析を担当。分析に基づき、航空券などのレコメンド施策の立案・企画・実施にあたる。お客さまの閲覧傾向に応じてコンテンツを変更する1to1マーケティングを実践し、購入率をアップするなどの成果を実現している。2014年に日経情報ストラテジーの「データサイエンティスト・オブ・ザ・イヤー 2014」を受賞。

### Q1. どのような分析業務をされているのですか

Webページのログなどの社内データはもちろんですが、最近では一部、社外データも組み合わせて分析するために、プライベートDMP\*の活用に取り組んでいます。

また、これまでは「ハワイ線の商品をもっと販売したい」といった商品志向の施策が中心でしたが、最近はお客さま志向の施策にシフトしています。お客さまを購買決定要因に応じてあらかじめ分け、そのお客さまが望んでいるものを提案するというアプローチです。データ分析をすれば「ハワイに行きそうな人」は予測できます。しかし、データ分析でターゲティングはできても人の心を変えることはできません。その人が何の目的でハワイに行くのか、その文脈を理解した上で購買意欲を高められないかを検討します。そのために外部データも使いながら施策を考えています。

\*Data Management Platform

### Q2. データ分析を成功させるための秘訣は何でしょうか

ビジネスの理解が最も重要だと考えています。いくら高い分析スキルがあっても、ビジネスを理解していなければ良い分析はできません。ビジネスの課題を理解し、共感したうえで、それを解決するためにどのデータを使えばよいかという、データの目利きになることがデータサイエンティストには求められると思います。データが食材、データサイエンティストが料理人だとすると、たくさんある食材から何を選んでどんな料理を作り出せるかが腕の見せどころです。

昨今、人工知能の進歩が目覚しく、データサイエンティストと共存できるのかといった議論も出ていますが、私は住み分けできると考えています。人工知能は世の中に多様にある非構造化データを分析できる形に変換してくれるため、人工知能が我々の食材を増やしてくれるとも言えます。ビジネスを理解し、課題を解決するために何をすべきかをコミュニケーションした上でデータの目利きをするというのは、人工知能にはまだ難しいため、そこはデータサイエンティストが担うべき役割だと考えます。

### Q3. データサイエンティストに必要なスキルについて教えてください

分析のスキルという意味では、数学から入ると挫折してしまうことが多いため、まずはツールを使って分析してみることをお勧めします。その際にはサンプルデータではなく、自分の持っているデータで分析する方が気づきを得られやすいです。また、分析結果をビジネスでどう利用するか、事前にイメージしておくことが重要です。例えば、メールの開封率が予想できたとして、それを具体的な施策にどのように活かすことができるのか、よく考えておく必要があります。

分析業務に取り組む際には大きく2つのハードルがあります。一つは統計解析、もう一つはデータを加工・集計・整形する作業です。統計解析については、必要な手法はそれほど多くないと考えています。通常の事業会社であれば、クロス集計、ロジスティック回帰分析、クラスター分析、決定木分析、アソシエーション分析の5つで十分です。

技術的に難しいのは実はローデータからの加工・集計です。ツールを使わない場合はSQLを書く必要がありますが、Excelしか使ったことがない人には相当ハードルが高いため、SPSSのようなツールを使って効率化するのがよいのではないかと思います。

# SPSS Statisticsで統計解析をはじめよう

## ■ IBM SPSS Statisticsモジュール製品

Statistics Baseにモジュール製品を加え、より高度でさまざまな分析を行うことができます。



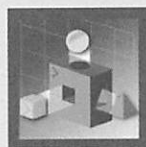
### IBM SPSS Advanced Statistics

複雑な関係性をより高い精度で分析  
一般線型モデル(GLM)、一般化線型モデル(GENLIN)、線型混合モデル、一般化推定方程式(GEE)、一般化線形混合モデル(GLMM)、生存分析(カプランマイヤー法やCox回帰)などの手法を搭載



### IBM SPSS Custom Tables

カスタム テーブルを即座に作成  
・ドラッグアンドドロップ インターフェイスを使用して、すばやく表を作成  
・プレビューしながら表を作成できるため、何度も表を作成し直す手間を省ける  
・結果を理解しやすいように表をカスタマイズ



### IBM SPSS Categories

カテゴリ型データの知覚マップから隠された関係性を明らかに  
・多変量カテゴリ型データを視覚化して探索  
・大規模な2次元、多次元表に含まれる情報を理解  
・分かりやすいバイプロットと知覚マップを使用してデータの関係を把握



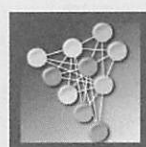
### IBM SPSS Missing Values

欠損データを推定し、高品質のデータを作成することでより精度の高いモデルを構築  
・欠損データの問題を克服  
・欠損データの範囲を即座に特定  
・欠損値を考慮する場合により統計的に有意な結果を得られる



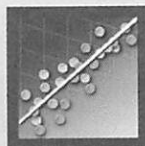
### IBM SPSS Bootstrapping

モデルの安定性を保証  
・推定量の標本分布を即座に推定  
・精度の低下につながる外れ値や異常値を排除  
・多数の分析手順をブートストラップする



### IBM SPSS Neural Networks

データの複雑な関連性を発見  
・より性能の高いモデルの構築  
・プログラミングが不要  
・MLPまたはRBFアルゴリズムから選択



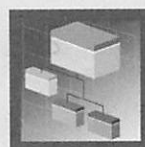
### IBM SPSS Regression

強力な回帰手続きで精度の高い予測結果を得る  
・多項ロジスティック回帰 (MLR) を使用し、カテゴリが3つ以上あるカテゴリ型の結果変数を予測  
・2項ロジスティック回帰を使用し、データを2つのグループに簡単に分類  
・モデルをより柔軟に調整可能



### IBM SPSS Exact Tests

サンプル数が少ない場合や稀な事象の場合でも正確な結果を得られる  
・サンプル数が少なく、従来の検定では正しい結果が得られないような場合でも使用可能  
・大きなデータセットの稀な事象を分析  
・30以上の正確率検定から選択



### IBM SPSS Decision Trees

グループをより的確に特定するための分類ツリーを作成  
・IBM SPSS Statistics内で直接ツリーを作成  
・データのパターン、区分、およびグループを特定  
・4つのツリー成長手法から選択



### IBM SPSS Conjoint

顧客の購入決定に作用する要素を発見  
・新規顧客にとって重要な製品機能を特定  
・「最も重要な」製品属性を特定  
・製品属性が顧客の購入決定にどのように影響するかを調査



### IBM SPSS Data Preparation

疑わしいケースや無効なケース、変数、データを簡単に特定  
・データの前処理を合理化  
・手間のかかる手動によるチェックが不要  
・より精度の高い結論を得られる



### IBM SPSS Complex Samples

複雑なサンプルの統計量を簡単に計算  
・母集団により統計的に有効な推定を得る  
・調査データからより精度の高い結果を得る  
・データの解析と結果の解釈が容易

※モジュール製品の利用には同一バージョンのIBM SPSS Statistics Baseが必要です。

## ■ 教育機関の方におすすめの製品構成

### 一番人気の構成

#### SPSS BARTモデル

社会学や心理学をはじめ各種研究にStatisticsを利用される方向け

基本機能を搭載したBaseに加えて、一般線型モデル(GENLIN)などに有用なAdvanced Statistics、ロジスティック回帰などの分析が可能なRegression、クロス集計を高度に行うCustom Tablesを組み合わせたモデルです。

IBM SPSS Statistics Base  
IBM SPSS Advanced Statistics  
IBM SPSS Regression  
IBM SPSS Custom Tables

### 教育機関向けスタンダード

#### SPSSアナリストモデル

アカデミック分野での高度な分析に

基本機能を搭載したBaseに、高度な統計手法を網羅したAdvanced Statistics、高度な回帰分析を行うオプションRegressionを組み合わせ。さらに高度な分析のためにはその他オプションを追加いただけます。

IBM SPSS Statistics Base  
IBM SPSS Advanced Statistics  
IBM SPSS Regression

### 医療統計

#### SPSSメディカルモデル

医療統計に必要な統計手法を網羅

これから医療統計・看護統計を始める医療機関の方向けのおすすめ構成です。

IBM SPSS Statistics Base  
IBM SPSS Advanced Statistics  
IBM SPSS Regression  
IBM SPSS Custom Tables  
IBM SPSS Exact Tests



# SPSS研修スケジュール 2017年 1月～3月

## ■ 操作コース SPSS製品の操作方法を学ぶトレーニングコースです。

コース名	時間	日数	マシン 使用	コースコード	1月	2月	3月
IBM SPSS Statistics 操作入門	9:30 - 17:00	2	○	0E5Y7	12-13	-	1-2
IBM SPSS Statistics データ加工	9:30 - 17:00	1	○	0E5Q7	-	-	3
IBM SPSS Statistics 基礎統計	9:30 - 17:00	2	○	0E5Z7	19-20	-	7-8
IBM SPSS Statistics 比較のための分散分析	9:30 - 17:00	1	○	0DA13	-	22	-
IBM SPSS Statistics 分散分析アドバンス	9:30 - 17:00	1	○	0E617	-	23	-
IBM SPSS Statistics 分類のための主成分・因子・クラスタ分析	9:30 - 17:00	1	○	0DA14	-	-	21
IBM SPSS Statistics 予測のための回帰・ロジスティック回帰分析	9:30 - 17:00	1	○	0DA15	-	-	27
IBM SPSS Statistics ディジジョンツリー	9:30 - 17:00	1	○	0E2C6	-	8	-
IBM SPSS statisticsによる傾向スコアの利用	9:30 - 17:00	1	○	0DA21	-	17	-
IBM SPSS Statistics 医療従事者のための分析1	9:30 - 17:00	1	○	0E657	18	-	-
IBM SPSS Statistics 医療従事者のための分析2	9:30 - 17:00	1	○	0E667	27	-	-
IBM SPSS Statistics 医療従事者のための分析3	9:30 - 17:00	1	○	0DA07	-	-	23
IBM SPSS Statistics によるノンパラメトリック検定:入門	9:30 - 15:30	1	○	0DA22	-	24	-
IBM SPSS Modeler 入門1	9:30 - 17:00	2	○	0D0P2	10-11	13-14	13-14
IBM SPSS Modeler 入門2 データ加工	9:30 - 17:00	1	○	0D0Q2	26	15	-
IBM SPSS Modeler 顧客分析【クラスタリング&アソシエーションモデル編】	9:30 - 17:00	1	○	0D0R2	-	27	-
IBM SPSS Modeler 顧客分析【予測モデル:カテゴリ編】	9:30 - 17:00	1	○	0DA18	-	28	-
IBM SPSS Modeler 顧客分析【予測モデル:スケール編】	9:30 - 17:00	1	○	0DA19	-	-	17
IBM SPSS Amos 操作入門	9:30 - 17:00	1	○	0E2D7	-	-	22
IBM SPSS Amos アドバンス	9:30 - 17:00	1	○	0E2E7	-	-	31
IBM SPSS Text Analytics for Surveys	9:30 - 17:00	1	○	0E320	-	-	9
IBM SPSS ロジスティック回帰分析入門	9:30 - 15:30	1	○	0E2I7	-	-	6

※一部コースの終了時刻が変更となりました。

## ■ アプリケーションコース アンケートや医療統計、データ加工など用途別に実践的な分析方法を学ぶコースです。

コース名	時間	日数	マシン 使用	コースコード	1月	2月	3月
IBM SPSS Statistics による アンケート調査:入門	9:30 - 15:30	1	○	0E6G7	-	-	15
IBM SPSS Statistics による アンケート調査:有意性検定の選択と実践	9:30 - 15:30	1	○	0E6H7	-	-	22
IBM SPSS Statistics による アンケート調査:多変量解析	9:30 - 15:30	1	○	0E6I7	-	-	27
IBM SPSS Statistics による アンケート調査:相関分析	9:30 - 15:30	1	○	0E6J7	-	-	24
IBM SPSS Statistics による テキストマイニングとデータ解析	13:00 - 17:00	0.5	○	0E5X7	-	3	-
IBM SPSS Statistics による 医療統計:入門	9:30 - 16:30	1	○	0E6L7	-	-	13
IBM SPSS Statistics による 医療統計:多変量解析	9:30 - 16:30	1	○	0E6M7	-	-	14
IBM SPSS 医療統計:論文の統計手法を読む	9:30 - 16:30	1	○	0D021	-	-	30

お客様のニーズに合わせた各種コースをご用意しております。  
 コースの詳細、最新スケジュールやお申込はWebサイトをご覧ください。

<http://www.i-learning.jp/service/it/spss.html>

●コース内容やスケジュールは止むを得ず一部変更になる場合がございます。予めご了承ください。 ●開催会場の詳細、地図についてはWebサイトにてご確認ください。  
 ※一部コース名が変更となっておりますのでご注意ください。

お申込み、お支払方法など、研修全般に関するお問い合わせ先:株式会社アイ・ラーニング ☎ 0120-623-629 (平日 9:00~17:00)

# SPSS研修コース対応表

この表は、製品群と研修コースの対応表です。どのコースにどの製品が必要か、ご確認頂けます。

コース名	日数	マシン 使用	難易度	Statistics Family											
				Statistics Base	Data Preparation	Advanced Statistics	Exact Test	Regression	Decision Trees	Forecasting	Categories	Text Analytics for Surveys	Amos	Modeler	SamplePower
操作コース															
IBM SPSS Statistics 操作入門	2	○	★	○											
IBM SPSS Statistics データ加工	1	○	★	○											
IBM SPSS Statistics 基礎統計	2	○	★★	○	○		○ (紹介のみ)								
IBM SPSS Statistics 比較のための分散分析	1	○	★★★	○		○									
IBM SPSS Statistics 分散分析アドバンス	1	○	★★★★	○		○									
IBM SPSS 分類のための Statistics 主成分・因子・クラスタ分析	1	○	★★★	○											
IBM SPSS 予測のための Statistics 回帰・ロジスティック回帰分析	1	○	★★★	○				○							
IBM SPSS Statistics ディンジョンツリー	1	○	★★★	○					○						
IBM SPSS Statistics 時系列分析	1	○	★★★★	○						○					
IBM SPSS 医療従事者のための分析1	1	○	★★	○											
IBM SPSS 医療従事者のための分析2	1	○	★★★	○		○									
IBM SPSS 医療従事者のための分析3	1	○	★★★	○		○	○	○							
IBM SPSS Modeler 入門	2	○	★★												○
IBM SPSS Modeler データ加工	1	○	★★												○
IBM SPSS Modeler 顧客分析 【クラスタリング&アソシエーションモデル編】	1	○	★★★												○
IBM SPSS Modeler 顧客分析 【予測モデル:カテゴリ編】	1	○	★★★												○
IBM SPSS Modeler 顧客分析 【予測モデル:スケール編】	1	○	★★★												○
IBM SPSS Amos操作入門	1	○	★★★												○
IBM SPSS Amos アドバンス	1	○	★★★★												○
IBM SPSS Text Analytics for Surveys	1	○	★										○		
IBM SPSS ロジスティック回帰分析入門	1	○	★★★	○					○						
理論コース															
IBM SPSS 入門はじめての分散分析と多重比較	1	×	★★	○		○									
IBM SPSS 入門初めての多変量解析	1	×	★★★	○											
アプリケーションコース															
IBM SPSS アンケート調査:入門	1	○	★	○											
IBM SPSS アンケート調査:有意性検定の選択と実践	1	○	★★	○			○								
IBM SPSS アンケート調査:多変量解析	1	○	★★	○				○				○			
IBM SPSS アンケート調査:コレスポンデンス分析	1	○	★★	○								○			
IBM SPSS テキストマイニングとデータの解析	0.5	○	★★	○								○	○		
IBM SPSS 医療統計:入門	1	○	★	○		○									
IBM SPSS 医療統計:多変量解析	1	○	★★	○		○		○							



# 関連書籍を上手に利用して分析力をアップ

操作方法を詳しく掲載している入門書から、中級以上の方向けの書籍まで多くの書籍が発刊されています。ぜひご活用ください。

## NEW SPSS完全活用法 データの入力と加工 (第4版)

酒井 麻衣子 著

- B5判変形(272頁) ●本体: 2,800円+税
- 出版社: 東京図書 ●発行: 2016年9月
- ISBN978-4-489-02247-0 C0040

◎知らないと損する裏ワザとテクニックが満載  
どんなに高度な手法を選んでも、データの下準備が不十分だったら正確な分析結果は得られない。本書では、目的にあったデータ収集や入力のコツ、データの加工法について、ウェブショップの顧客データや雑誌のアンケート事例などを用いて具体的に解説する。さらには、異なる手続きを一度で済ませたいとき、同じ手続きをコピーしたいとき、計算を手早く済ませたいときなどに使うシンタックスも多数紹介。知らないと損する裏ワザ&テクニックが満載。



## NEW SPSSによる多変量データ解析の手順 (第5版)

石村 貞夫・石村 光賢郎 著

- B5判変形(304頁) ●本体: 2,800円+税
- 出版社: 東京図書 ●発行: 2016年7月
- ISBN978-4-489-02242-5 C3040

◎入力の仕方から出力結果の見方まで、  
16の手法をやさしく解説  
重回帰分析をはじめとして、因子分析や主成分分析、階層型ニューラルネットワークや決定木、プロビット分析、パス解析、そして共分散構造分析などの16の分析手法とその手順をわかりやすく解説。とまどいがちなデータの入力から出力結果の読み取り方まで、画面を一つ一つ追っていけば、多変量解析がすぐに使えるよう書かれている。自慢の手順書、改訂版出来!



## NEW SPSSによるアンケート分析

寺島 拓幸・廣瀬 毅士 著

- B5判変形(292頁) ●本体: 2,500円+税
- 出版社: 東京図書 ●発行: 2016年6月
- ISBN978-4-489-02238-8 C0040

◎実際の場面を想定し回答形式ベースで  
まとめた

アンケートで得られる質的データを中心に、SPSSを使った分析方法とその理論を解いたアンケートの必携本。単一回答、複数回答、数値解答など、よく使われる典型的な回答形式ベースで分析方法を例示しているため、調査目的に適した分析を素早く探すことができる。本文は、①目的と考え方、②注意点、③前提条件、④SPSSの操作と出力、⑤補足、⑥論文・レポートでの提示方法をパターン化して仔細に解説。初学者でも無理なく読み進められるアンケートの新しい定本である。

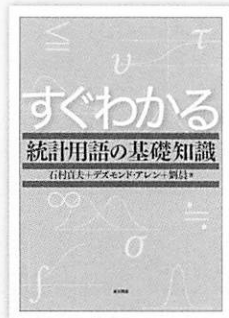


## NEW すぐわかる統計用語の基礎知識

石村 貞夫・D.アレン・劉 著

- A5判 (320頁) ●本体: 2,800円+税
- 出版社: 東京図書 ●発行: 2016年2月
- ISBN978-4-489-02233-3 C3040

◎コトバがわかれば統計はもっと面白くなる  
統計学を学ぶとき、データを集めて統計処理をするとき、コンピュータを使わない状況は考えられない。難しい計算に頭を悩ますことはなくなっても、理解するには基本が大切。言葉は似ていても全く違う、違うけれども同じ意味——厳密さよりも「わかりやすさ」を重視した簡潔な解説で、困ったときのあなたをサポートする。これから統計を学ぶ人にも、自分の知識の再確認にも必ず役に立つだろう。英語論文執筆のための基礎知識も付いている。



## SPSS超入門 — インストールからはじめるデータ分析

畠 慎一郎・田中 多恵子 著

- B5判変形(208頁) ●本体: 2,200円+税
- 出版社: 東京図書 ●発行: 2015年12月
- ISBN978-4-489-02228-9 C0040

データ分析をしなくてはならない。だけど何から始めればいいのか。統計が苦手でもパソコンが不得手でも大丈夫。インストールの方法から手取り足取り解説する。

統計解析の基礎のキソと、データ分析の手順、SPSSの基本操作をこの一冊にまとめた。さらに、知っていると便利な操作のコツや、もう一歩進むための参考書ガイドも収録。データが手元にあるのなら、まずは手を動かしてみよう。プロが教える超々初心者のためのデータ分析の入門書。



## 研究をブラッシュアップするSPSSとAmosによる心理・調査データ解析

小塩 真司 著

- B5判変形(272頁) ●本体: 2,800円+税
- 出版社: 東京図書 ●発行: 2015年7月
- ISBN978-4-489-02213-5 C0040

この分析とは違う分析方法をどうして発想できるのか、どうしてこの分析よりもあの分析方法のほうがより良いと考えられるのか——。データ分析はデータの収集前に立てる分析計画も重要だが、実際に試行錯誤しながら分析することも重要となる。本書は2人の大学生と心理学教授との会話を通して様々なデータを分析し、「誤りの発見→改善する→代案を示す」といった流れから、陥りがちな問題点とその解決策を学ぶ。データ分析の経験値を上げよう。



**NEW SPSSで学ぶ医療系データ解析 (第2版)**

対馬 栄輝 著  
 ●B5判変形(292頁) ●本体価格:3,200円+税  
 ●出版社:東京図書 ●発行:2016年12月  
 ●ISBN978-4-489-02258-6 C3047

◎分析内容の理解と手順解説、  
 バランスのとれた医療統計入門

理にかなった検定や解析を行う力を養うために、統計の専門的内容を省略したり、あまり簡単にはせず、かといって難しい理論まで深入りしない、ちょうど中間に位置する実用的テキスト。データ解析に対する知識のない初心者から、ある程度の知識を備えた方まで、幅広い読者層に支えられてきた初版をSPSSのバージョンアップに伴い改訂。リクエストの多かった「サンプルサイズの決定法」や「分析結果の読み解き」を強化した。



**SPSSによる医学・歯学・薬学のための統計解析 (第4版)**

石村 貞夫・謝 承泰・久保田 基夫・石村 友二郎 著  
 ●B5判変形(272頁) ●本体: 3,200円+税  
 ●出版社:東京図書 ●発行:2016年1月  
 ●ISBN978-4-489-02226-5 C3047

研究成果を学会や論文で発表するとき、どの統計手法を選んだらよいか、どうしたら客観的な評価をくだせるか? こうした悩みにこたえるために、医学系でよく使われる統計学の諸手法をわかりやすく解説。多重比較、ノンパラメトリック検定、2元配置の分散分析に始まって、コルモゴロフ・スミルノフ検定、マンテル・ヘンツェル検定、カプラン・マイヤー法、そしてコックス回帰分析と対数線型分析まで、画面で指示されたおりにクリックしていけば、誰でもすぐ使えるように書かれている。



**SPSSによる臨床心理・精神医学のための統計処理 (第2版)**

石村 貞夫・加藤 千恵子・石村 友二郎 著  
 ●B5判変形(280頁) ●本体: 2,800円+税  
 ●出版社:東京図書 ●発行:2015年11月  
 ●ISBN978-4-489-02225-8 C3040

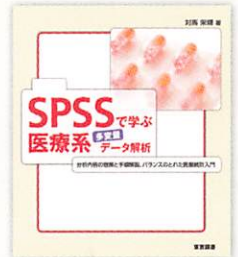
論文やレポートで研究成果を発表するときに必要な統計の上手な使い方と分析手法を解説する。アンケート調査の基本をはじめとして、基礎統計量の算出、グラフ表現、ノンパラメトリック検定や回帰分析、そして因子分析や主成分分析など、臨床心理や精神医学の研究に必要な分析手法を一通り網羅している。はじめての卒業論文、専門的な分析が必要な修士論文、プレゼンのためのデータ処理などに迷わず使えるよう、画面を追って説明。初めての人でも本書の手順通りにクリックすると……



**SPSSで学ぶ医療系多変量データ解析**

対馬 栄輝 著  
 ●B5判変形(256頁) ●本体: 3,200円+税  
 ●出版社:東京図書 ●発行:2008年10月  
 ●ISBN978-4-489-02041-4 C3047

医学に限らず、複雑化する現象から得られたデータを捉えるためには多変量解析の手法が必要となるが、それぞれの手法は高度な数学理論のうえに成り立っているために、統計初心者が実務的解析を行なうには優れた統計ソフトウェアの助けが不可欠である。本書では各解析手法の解説を「解析のしくみ」と「解析の実際」の2章構成とし、SPSSの操作手順とあわせて理論と実用の両面からデータ解析を学べるようになってきている。初心者にも、ある程度の経験を積んだ人にも役に立つ。



**Facebook、TwitterでSPSS最新情報を入手しよう!**

**IBM Analytics Japan Facebook**  
<https://www.facebook.com/BusinessAnalytics>



**IBM SPSS Twitter**  
<https://twitter.com/spssjapan>



**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番21号

© Copyright IBM Japan, Ltd. 2017 All Rights Reserved Printed in Japan

IBM、IBMロゴ、ibm.com およびSPSSは、世界の多くの国で登録されたInternational Business Machines Corp.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商標リストについては [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) をご覧ください。