

# SPSS

## データサイエンティストUniversity version 1.0 <概要>

日本アイ・ビー・エム株式会社  
2017/07/03



# 目次

- よくあるお客様の課題に対する提供価値
- データサイエンティストUniversityの目的
- ステップ1 基礎編
- ステップ2 応用編
- ステップ3 実践編
- データサイエンティストUniversity

## よくあるお客様の課題に対する提供価値

データサイエンティストUniversityでは、お客様の典型的な課題に対して、データサイエンティストとして必要なスキルを、体系化されたプロセスに基づいて段階的に学習し、アドバイスを受けつつデータ分析を繰り返すことで定着化させます。時間をかけて習熟した知識を活かし、業務課題の改善に繋げることができる人材を増やすことは企業にとって大きなメリットになります。

### 典型的な課題

データから現状の問題点を把握し  
解決できるような人材を育成したい

分析課題とデータに適した分析手法を  
判断できる人材を育成したい

ビジネス戦略に役立つ分析プロジェクトの企画や  
進め方を身につけた人材を育成したい

### 提供価値

#### 問題解決力の養成

データの特徴や関係性から仮説を立て、  
検証手続きを繰り返すことで、  
問題解決への道筋を描ける力を養います。

#### データ分析力の養成

様々な手法を的確に理解し、  
どの手法が妥当かの判断ができ、  
実務に応用できる力を養います。

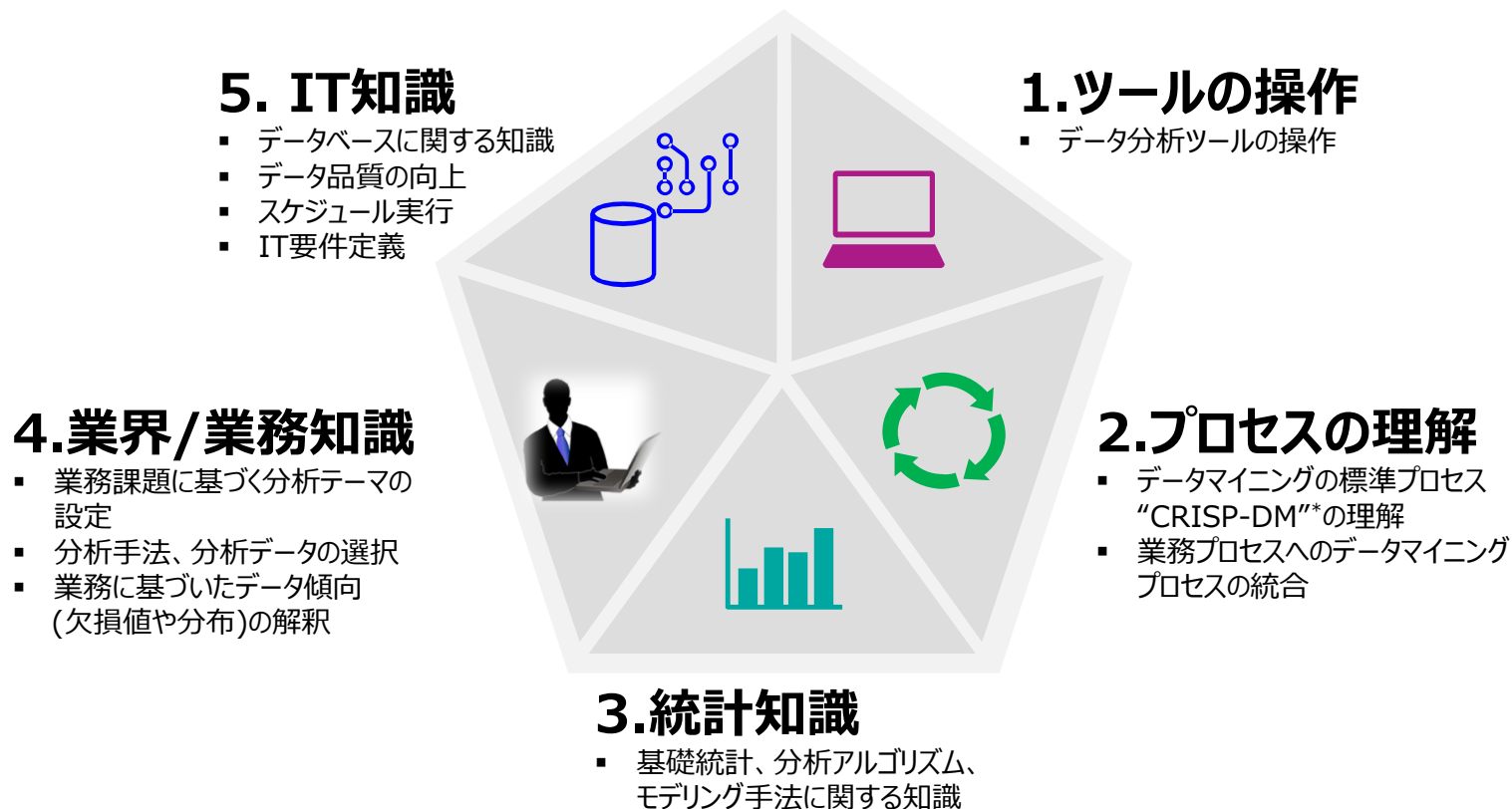
#### 自社ビジネスへの貢献力の養成

予測分析の結果をマネジメントの意思決定の  
強力な材料として活用できる力を養います。

# データサイエンティストUniversityの目的

データサイエンティストには、さまざまなスキルが必要とされます。要約すると以下の1.から5.が挙げられます。データサイエンティストUniversityでは、これらのスキルセットを伸ばすことを目的としています。主に、1.から3.のスキルにフォーカスしますが、業務に基づくデータの傾向やデータ分析に影響するデータの品質など、4.5.のスキルに関連する事柄にも触れます。このプログラムの修了時には、受講前よりも各スキルが向上していることを目指します。

## 「データサイエンティスト」のスキルセット



## 1.1 ステップ1 基礎編

<b>概要</b>	データ分析(データマイニング)のプロセスモデルであるCRISP-DM*を理解し、データ分析ツールであるIBM SPSS Modelerの一連の操作知識を習得して頂くことを目的としています。CRISP-DMには、ビジネスの理解、データの理解、データの準備、モデリング、評価、展開/共有の6つのフェーズがあります。まずは、これらのフェーズの意味合いと位置づけをご理解頂きます。つぎに、これらのフェーズのうち、データの理解から展開/共有のフェーズに関連するIBM SPSS Modelerの主な機能を網羅的に学習します。IBMが作成した公式テキストとサンプルデータ(顧客データ)を用いて、ハンズオン形式で実施します。
<b>メリット</b>	<ul style="list-style-type: none"><li>▪ 経験豊富な講師を派遣し、お客様先での実施が可能</li><li>▪ 体系的で説明が充実したテキストを使用するため、復習が可能</li></ul>
<b>期間</b>	3日間 10:00-17:00 (お客様のご都合にて調整が可能)
<b>お客様準備</b>	<ul style="list-style-type: none"><li>▪ ご受講者用PC、ホワイトボード、プロジェクターの用意</li><li>▪ ご受講者用PCと、その事前セットアップ</li></ul>

## 1.2 ステップ2 応用編

<b>概要</b>	<p>ステップ1 基礎編で習得した知識を基に、座学とハンズオン(演習)の2部構成で実施します。IBM SPSS Modelerで利用できる統計や機械学習の手法(次頁の一覧から選択)を習得して頂くこと、また、分析課題(お客様ご用意)を設け、貴社データを用いてデータの理解や準備、モデリング(データ分析)を実施することで、モデリングのプロセスと実務への適用イメージを習得して頂くことを目的としています。</p> <p>座学では分析手法を、ハンズオンでは貴社データでの分析プロセスを取り扱います。 ハンズオン資料は、貴社用に準備をさせていただきます。</p>
<b>メリット</b>	<ul style="list-style-type: none"><li>▪ 貴社データを利用した実践的なカリキュラムで実務への適用がスムーズ</li><li>▪ 貴社データでのIBM SPSS Modelerの操作、分析手法の適用、プロジェクトの進め方など、効果的な分析スキルの習得が可能</li></ul>
<b>期間</b>	<ul style="list-style-type: none"><li>▪ 1セッション：4時間(推奨13:00-17:00)</li><li>▪ 期間：5週間(セッションの間隔は、最短1週間)</li></ul>
<b>お客様準備</b>	<ul style="list-style-type: none"><li>▪ 分析課題(例:離反分析、異常検知分析など)を1つ決定</li><li>▪ 分析課題に該当するデータの用意(ハンズオン資料作成のため持ち出しが可能であること、ハンズオンで用いるためフィールド数とレコード数は適度に選別されていること)</li></ul>

## 応用編サンプルプログラム【マーケティング系, 顧客分析】

セッション	データ分析のフェーズ：目的	座学項目	所要時間
1	データの理解：基礎統計	■ 1 変量の要約(尺度・統計量/グラフ):データ検査	30分
		■ 母集団と標本・推定/検定	30分
		■ 2 変量の関係性(カイ 2 乗検定/ t 検定/無相関検定):フィールド選択	60分
2	モデリング：予測編①	■ 線型回帰	60分
		■ ロジスティック回帰	60分
3	モデリング：予測編②	■ デcisionツリー	90分
		■ ニューラルネット	30分
4	モデリング：分類編 連関編(アソシエーション)	■ クラスタ分析(K-means/Kohonen/Two Step)	60分
		■ アソシエーション分析(Apriori/Carma)	60分
5	モデリング：精度検討	■ バランス	30分
		■ 精度分析	30分
		■ アンサンブル学習	45分
		■ 全体を通しての質疑応答	15分

## 応用編サンプルプログラム【異常検知系, センサーデータ分析】

セッション	データ分析のフェーズ：目的	座学項目	所要時間
1	データの理解：基礎統計	■ 1 変量の要約(尺度・統計量/グラフ):データ検査	30分
		■ 母集団と標本・推定/検定	30分
		■ 2 変量の関係性(カイ 2 乗検定/ t 検定/無相関検定):フィールド選択	60分
2	データの準備：次元削減	■ フィールド選択	60分
		■ 因子(主成分)分析	60分
3	モデリング：外れ値検出	■ KNN	60分
		■ マハラビスの距離 R	60分
4	モデリング：変化点検出	■ 線型回帰	60分
		■ 時系列分析	60分
5	モデリング：異常行動検出 精度検討	■ バイズネット	60分
		■ 精度分析	30分
		■ AIC,BIC	15分
		■ 全体を通しての質疑応答	15分



## 各セッションの構成とサンプル【マーケティングの場合】

各セッションの枠組み	座学	一般的なデータ分析(モデリング)手法を理解する。 モデリング手法の知識を広げる。 Modelerを使用したモデリングの手順を知る。 など	第3回	座学	モデリング：予測編② ■ ディジションツリー (CHAID/C&R Tree/QUEST/C5.0) ■ ニューラルネット ゴール:ディジションツリーとニューラルネットワークの方法を習得
	演習	貴社データを用い、課題を設定し、 データ分析のプロセスモデル (CRISP-DM) に基づき、データ 分析を体験する。 ※進捗状況やディスカッションに基づき柔軟に対応		演習	Modelerを使用した予測分析 ● ディジションツリーノードを使用した分析 ● ニューラルネットノードを使用した機械学習での分析
第1回	座学	データの理解：基礎統計 ■ 1変量の要約 (尺度・統計量/グラフ) ■ 母集団と標本・推定/検定 ■ 2変量の関係性 (カイ2乗検定/ t 検定/無相関検定) ゴール:データを統計的に理解・把握する方法を習得	第4回	座学	モデリング：分類編/連関編 (アソシエーション) ■ クラスタ分析 (K-means/Kohonen/Two Step) ■ アソシエーション分析 (Apriori/Carma) ゴール:クラスタ分析とアソシエーション分析の方法を習得
	演習	IBM SPSS Modelerを使用した課題データの理解 ● データ検査ノードを使用した記述統計、欠損値の確認 ● フィールド選択ノードを使用したフィールド間の関係性の確認		演習	IBM SPSS Modelerを使用したクラスタ分析 ● クラスタリングノードを使用したクラスタ分析 ● アソシエーションノードを使用したアソシエーション分析
第2回	座学	モデリング：予測編①/精度分析 ■ 線型回帰 ■ ロジスティック回帰 ゴール:予測のための線型回帰とロジスティック回帰分析の方法を習得	第5回	座学	モデリング：精度検討 ■ 精度分析 精度分析の習得と全体のまとめ ゴール:精度分析の習得と全体を通しての復習と質疑応答
	演習	IBM SPSS Modelerを使用した予測分析 ● 線型回帰ノードを使用した分析 ● ロジスティック回帰ノードを使用した分析		演習	IBM SPSS Modelerを使用した精度の検討 ● 精度分析ノードを使用したモデルの精度の検討 全体を通しての復習と質疑応答 今後の分析課題に対するディスカッション

## 各セッションの構成とサンプル【異常検知の場合】

各セッションの枠組み	座学	一般的なデータ分析(モデリング)手法を理解する。 モデリング手法の知識を広げる。 IBM SPSS Modelerを使用したモデリングの手順を知る。など	第3回	座学	モデリング：外れ値検出 ■ マハラビスの距離 ■ KNN ゴール:外れ値を検出する方法を習得
	演習	貴社データを用い、課題を設定し、 データ分析のプロセスモデル(CRISP-DM)に基づき、データ分析を体験する。 ※進捗状況やディスカッションに基づき柔軟に対応		演習	IBM SPSS Modelerを使用した外れ値の検出 ● ヒストグラムノードを使用した外れ値の検出 ● KNNノードを使用した多変量での外れ値の検出
第1回	座学	データの理解：基礎統計 ■ 1変量の要約(尺度・統計量/グラフ) ■ 母集団と標本・推定/検定 ■ 2変量の関係性(カイ2乗検定/ t 検定/無相関検定) ゴール:データを統計的に理解・把握する方法を習得	第4回	座学	モデリング：変化点検出 ■ 線型回帰 ■ 時系列分析 ゴール:変化点検出のための回帰・時系列分析の方法を習得
	演習	IBM SPSS Modelerを使用した課題データの理解 ● データ検査ノードを使用した記述統計、欠損値の確認 ● フィールド選択ノードを使用したフィールド間の関係性の確認		演習	IBM SPSS Modelerを使用した外れ値の検出 ● 線形回帰ノードを使用した外れ値の検出 ● 時系列ノードを使用した外れ値の検出
第2回	座学	データの準備：次元削減 ■ フィールド選択 ■ 因子(主成分)分析 ゴール:ターゲットに関連のあるフィールドを絞り込む方法を習得	第5回	座学	モデリング：異常行動検出/精度検討 ■ ベイズネット ■ 精度分析(AIC,BIC含む) ゴール:異常行動検出のためのベイズネットと精度分析を習得
	演習	IBM SPSS Modelerを使用したフィールドの絞り込み ● フィールド選択ノードを使用した関連フィールドの選択 ● 因子分析ノードを使用したフィールドの削減		演習	IBM SPSS Modelerを使用した異常行動の検出 ● ベイズノードを使用した異常行動の検出 ● 精度分析ノードなどを使用したモデルの精度の検討

## 1.3 ステップ3 実践編

<p><b>概要</b></p>	<p>ステップ1,2で習得した知識を基に、アドバイスを受けながらお客様自身でデータ分析を実施して頂くことが目的です。</p> <ul style="list-style-type: none"> <li>▪ 分析の方針決めやゴール設定、分析結果の理解、モデルのブラッシュアップなどへのアドバイスの提供</li> <li>▪ データ分析プロセス中に発生する質問や疑問への回答</li> <li>▪ IBM SPSS Modelerの発展的な操作方法や応用編取扱データ分析手法一覧から追加選択した、分析手法の知識の習得 などをご支援します。</li> </ul> <p>オンサイトで、4時間/週~をご利用頂けます。          オンサイトの利用内容は、お客様のご要望に応じてアレンジが可能です(以下、オンサイトの利用例を参考)。</p>
<p><b>オンサイト 利用例</b></p>	<ul style="list-style-type: none"> <li>▪ 複雑なデータ加工やモデリングの結果の読み取りなどのQA対応</li> <li>▪ IBM SPSS Modelerの発展的な機能を習得するアドバンスドメニュー (次頁に詳細)の実施</li> <li>▪ 応用編取扱データ分析手法一覧の中から、さらに追加して座学の実施</li> <li>▪ 分析結果の報告会を開催し、お客様とIBMにて相互に育成プログラムの達成度を確認</li> </ul>
<p><b>メリット</b></p>	<ul style="list-style-type: none"> <li>▪ オンサイトでも、データ分析に関する疑問点のアドバイスを受けることが可能</li> <li>▪ ステップ1 基礎編から発展したパラメータの利用や繰り返し処理など、さらに便利なIBM SPSS Modelerの操作方法の習得が可能</li> <li>▪ ステップ2 応用編取扱データ分析手法一覧から追加で分析手法の習得が可能</li> <li>▪ 分析結果の報告会を実施することで、データサイエンティストとしての達成度の評価が可能</li> </ul>
<p><b>期間</b></p>	<p>1ヶ月～3ヶ月(4ヶ月以上の場合は要相談)</p>

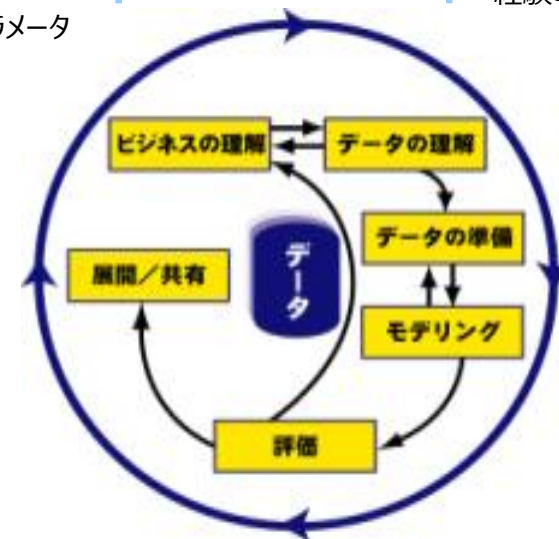
## 1.4 データサイエンティストUniversity

プログラム	内容		全期間																																							
	ステップ1 基礎編 ステップ2 応用編 5回開催	3か月	ステップ3 実践編 1か月 ステップ3 実践編 3か月	4か月 6か月																																						
特典	プログラム期間中、IBM SPSS Modelerに関するデータ分析の支援窓口にてメールでIBM SPSS Modelerの操作方法や分析手法に関連する問合せが可能（週8時間程度）																																									
前提	対象人数：12名様までを想定 実施場所：お客様オフィスを想定																																									
期間長期のメリット	<p>QAサービスの利用は、プログラムの前半よりも後半になるほど、質問や疑問が増えてくることが想定されます。たとえば、複雑なデータの加工方法や実務におけるモデルの結果の解釈など。QAサービスの時間配分を前半は少なめにし、後半を多めにしたり、また、複雑なデータ加工のフェーズは多めにするなど、期間が長いほど柔軟にご利用頂くことが可能になります。</p> <p>[QAサービスの時間配分の例]</p> <table border="1"> <thead> <tr> <th colspan="4">1ヶ月</th> <th colspan="4">2ヶ月</th> <th colspan="4">3ヶ月</th> <th>合計</th> </tr> <tr> <th>1w</th> <th>2w</th> <th>3w</th> <th>4w</th> <th>1w</th> <th>2w</th> <th>3w</th> <th>4w</th> <th>1w</th> <th>2w</th> <th>3w</th> <th>4w</th> <th>-</th> </tr> </thead> <tbody> <tr> <td>4h</td> <td>4h</td> <td>4h</td> <td>4h</td> <td>8h</td> <td>8h</td> <td>12h</td> <td>12h</td> <td>12h</td> <td>12h</td> <td>8h</td> <td>8h</td> <td>96h</td> </tr> </tbody> </table>			1ヶ月				2ヶ月				3ヶ月				合計	1w	2w	3w	4w	1w	2w	3w	4w	1w	2w	3w	4w	-	4h	4h	4h	4h	8h	8h	12h	12h	12h	12h	8h	8h	96h
1ヶ月				2ヶ月				3ヶ月				合計																														
1w	2w	3w	4w	1w	2w	3w	4w	1w	2w	3w	4w	-																														
4h	4h	4h	4h	8h	8h	12h	12h	12h	12h	8h	8h	96h																														
注意事項	<ul style="list-style-type: none"> <li>・ オフサイトのQAサービスは、期間中就業時間内(9:00-18:00)となります。</li> <li>・ 分析課題に対する結果の品質を保証するサービスではありません。</li> <li>・ スクリプトやRなどのプログラムの作成は本サービスの対象外です。</li> <li>・ 参考価格には、ソフトウェアライセンスおよびハードウェアの価格は含まれていません。</li> <li>・ 参考価格には、消費税が含まれていません。なお、本プログラムに関わる交通費等の経費は含まれています。ただし、都内近郊以外は、交通費・宿泊費を別途御見積もりとさせていただきます。</li> <li>・ 実施内容、期間、実施場所、対象人数などが変更される場合には、別途御見積もりとさせていただきます。</li> </ul>																																									

# 参考資料:CRISP-DMとは

- **C**ross-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining(CRISP-DM)は、データマイニングの標準的なプロセスモデルです。データサイエンティストUniversityでは、このCRISP-DMを利用しています。

ビジネスの理解	データの理解	データの準備	モデリング	評価	展開/共有
<p><b>ビジネス目標の決定</b></p> <ul style="list-style-type: none"> <li>▪ 背景情報</li> <li>▪ ビジネス目標</li> <li>▪ ビジネスの成功基準</li> </ul> <p><b>状況の評価</b></p> <ul style="list-style-type: none"> <li>▪ リソースの調査</li> <li>▪ 要件,仮定および制約</li> <li>▪ リスクと予想される事態</li> <li>▪ 用語集</li> <li>▪ コストと利益</li> </ul> <p><b>データマイニング目標の決定</b></p> <ul style="list-style-type: none"> <li>▪ データマイニングの目標</li> <li>▪ データマイニングの成功基準</li> <li>▪ プロジェクト計画作成</li> <li>▪ ツールと手法の初期評価</li> </ul>	<p><b>初期データの収集</b></p> <ul style="list-style-type: none"> <li>▪ 初期データ収集レポート</li> </ul> <p><b>データの記述</b></p> <ul style="list-style-type: none"> <li>▪ データ説明レポート</li> </ul> <p><b>データの調査</b></p> <ul style="list-style-type: none"> <li>▪ データ調査レポート</li> </ul> <p><b>データ品質の検証</b></p> <ul style="list-style-type: none"> <li>▪ データ品質レポート</li> </ul>	<p><b>データセット</b></p> <ul style="list-style-type: none"> <li>▪ データセットの説明</li> </ul> <p><b>データの選択</b></p> <ul style="list-style-type: none"> <li>▪ 選択または除外の基準</li> </ul> <p><b>データのクリーニング</b></p> <ul style="list-style-type: none"> <li>▪ データクリーニングレポート</li> </ul> <p><b>データの構築</b></p> <ul style="list-style-type: none"> <li>▪ 派生属性生成されたレコード</li> </ul> <p><b>データの統合</b></p> <ul style="list-style-type: none"> <li>▪ 結合されたデータ</li> </ul> <p><b>データのフォーマット</b></p> <ul style="list-style-type: none"> <li>▪ 再フォーマットされたデータ</li> </ul>	<p><b>モデリング手法の選択</b></p> <ul style="list-style-type: none"> <li>▪ モデリング手法</li> <li>▪ モデリングの仮定</li> </ul> <p><b>テスト設計の生成</b></p> <ul style="list-style-type: none"> <li>▪ テスト設計</li> </ul> <p><b>モデル作成</b></p> <ul style="list-style-type: none"> <li>▪ パラメータの設定</li> <li>▪ モデル</li> <li>▪ モデルの説明</li> </ul> <p><b>モデルの評価</b></p> <ul style="list-style-type: none"> <li>▪ モデルの評価</li> <li>▪ 改訂されたパラメータの設定</li> </ul>	<p><b>結果の評価</b></p> <ul style="list-style-type: none"> <li>▪ ビジネスの成功基準に基づくデータマイニングの結果の評価</li> <li>▪ 承認済みのモデル</li> </ul> <p><b>プロセスの見直し</b></p> <ul style="list-style-type: none"> <li>▪ プロセスの見直し</li> </ul> <p><b>次のステップの決定</b></p> <ul style="list-style-type: none"> <li>▪ 実行可能なアクションのリスト決定</li> </ul>	<p><b>展開の計画</b></p> <ul style="list-style-type: none"> <li>▪ 展開の計画</li> </ul> <p><b>モニタリングとメンテナンスの計画</b></p> <ul style="list-style-type: none"> <li>▪ モニタリングとメンテナンスの計画</li> </ul> <p><b>最終レポートの作成</b></p> <ul style="list-style-type: none"> <li>▪ 最終レポート</li> <li>▪ 最終プレゼンテーション</li> </ul> <p><b>プロジェクトの見直し</b></p> <ul style="list-style-type: none"> <li>▪ 経験の文書化</li> </ul>



## お問い合わせ先

お見積・お問い合わせはこちらから

**[ibm.biz/spssuniv-jp](https://ibm.biz/spssuniv-jp)**

(※お問い合わせ内容の箇所に、“SPSS Universityについて”と記載ください)

IBMアクセスセンター（一次受付）  
電話番号:0120-550-210  
(識別コード:SPSS)

