



文部科学省

MINISTRY OF EDUCATION,
CULTURE, SPORTS,
SCIENCE AND TECHNOLOGY-JAPAN



大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics

文部科学省委託事業
「データサイエンティスト育成ネットワークの形成」
平成 25 年度事業報告書

平成 26 年 3 月



情報・システム研究機構
統計数理研究所

【再委託先】

東京大学大学院
情報理工学系研究科

事業期間／平成 25 年 7 月 8 日～平成 28 年 3 月 31 日

文部科学省委託事業
「データサイエンティスト育成ネットワークの形成」
平成 25 年度事業報告書

目 次

1. エグゼクティブ・サマリ	01
2. 平成 25 年度の主要成果と知見	02
2.1. 事業の設計	
2.1.1. データサイエンティストとは何か	
2.1.2. データサイエンティストの不足	
2.1.3. 事業の基本理念	
2.2. 事業の概要	
2.3. 平成 25 年度の成果と知見	
2.3.1 認知度向上・啓発	
2.3.2 人材ローテーション	
2.3.3 ベスト・プラクティスの調査	
2.3.4 教材の開発	
2.3.5 海外連携・標準化	
2.4. 今後の方針	
3. 事業実施内容の詳細	08
3.1. 認知度向上・啓蒙	
3.2. 人材ローテーション	
3.3. ベストプラクティスの調査	
3.3.1. 統計検定合格者へのアンケート調査	
3.3.2. 企業・機関への聞き取り調査	
3.4. 教材の開発	
3.5. 海外連携・標準化	
3.6. その他	
引用・参考文献一覧	29
付録 A：依頼状と調査票	31
付録 B：聞き取り調査結果一覧表	36

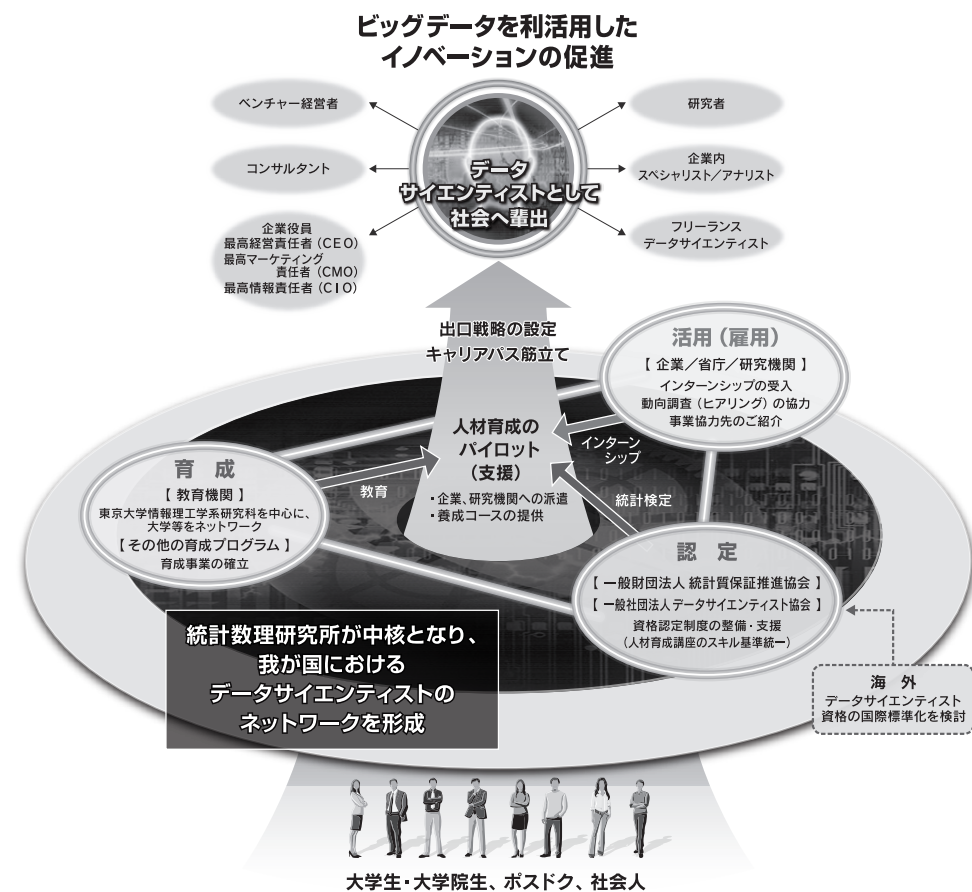
1. エグゼクティブ・サマリ

我が国におけるデータ分析専門家の持続的な育成と効果的な活用を目指して、我々は文部科学省委託事業「データサイエンティスト育成ネットワークの形成」を平成 25 年 7 月に開始した。本事業では、我が国におけるデータサイエンティストの「あるべき姿」を明らかにすると共に、データサイエンティスト育成に熱意を持つ教育機関と、データサイエンティストのスキルを利用したい企業・組織を広くネットワークし、それらの間で知識・経験を共有することで、多くのデータサイエンティストが育成され、有効に活用されることを狙う。この報告書は、2 年 9 ヶ月に渡る本事業の、最初の 9 ヶ月の成果に関するものである。

データ分析の専門家、いわゆる「データサイエンティスト」にはどのような類型があり、どのようなスキルを要求されているかについては、定まった合意があるわけではない。このため、今年度は特に、我が国におけるデータ分析専門家の現状を調査し、分析することに注力した。その結果、米国で一般に言われるデータサイエンティストとは異なった類型、特にキャリアパスとしてのデータサイエンティストがどうあるべきかという観点からの類型が明らかになり、我が国における今後の育成の一つの方向性が見えてきた。

また、データサイエンティスト育成のための試みとして、データ分析の基礎を短時間で学ぶためのオンライン教材「データサイエンティスト・クラッシュコース」の作成と、データ分析の現場を体験するためのインターンシップのトライアルを行った。

平成 26 年度以降は、本年度で得られた知見をもとに、「あるべき姿」の提案と浸透を図り、よりネットワークの活動の裾野拡大を狙う。



2. 平成 25 年度の主な成果と知見

2.1. 事業の設計

まず事業の設計の考え方について述べる。

2.1.1. データサイエンティストとは何か

事業の設計にあたり、まずデータサイエンティストとは何か、について文献等を調査した。そもそも、データサイエンティストという言葉には一定の定義がないようである。たとえば、Harvard Business Review 誌の記事は、かなりハードルの高い人材のことを述べている。新しいデータ分析アルゴリズムを開発する力を持ちながら、同時に経営に携わることのできるタイプの人材、敢えて言えば Google のラリー・ページのような人材を想定している。

それに近い人材育成の考え方としては、Insight Data Science Fellows Program がある^[14]。物理学、生物学などで Ph. D を持つ人材に対してデータ分析のツールやプログラミングを教えることで、短期間にデータサイエンティストとして育成しようというものである。これらの人材も、年収 9 万ドルから 13 万ドルで Google や Facebook に採用されるということなので、かなりトップレベルの人材と言えるだろう。

日経 ITPro は、今年の「データサイエンティストオブザイヤー」として、大阪ガスの河本氏を選出した。河本氏は、その著書「会社を変える分析の力」^[15]で、データサイエンティストには分析する力だけではなく、「問題を発見する力」と「分析結果を現場に使わせる力」が必要であると主張している。数理的なスキルとしては Ph. D までは必要ないかもしれないが、一方で、コンサルティングのようなビジネス能力が高く求められている。いわゆる「ビッグデータ」かどうかに関わらず、多くの企業では日々の業務の中で必要にかられてデータ分析をやっている方々も多いだろう。

このように、データサイエンティストには決まった類型がない。O' Reilly が、「データサイエンティスト」に近い人達 250 名に対して行った Web アンケートをクラスタリングして得られた結果^[6]によれば、データサイエンティストをビジネス型(“Data Businesspeople”）、研究者型(“Data Researcher”）、ハッカー型(“Data Creatives”）、デベロッパー型(“Data Developer”）のようにラフに 4 類型に分けられるという。一概にデータサイエンティストと言っても、ビジネス型の人材にソフトウェア開発を要求してもうまくいかないし、デベロッパー型の人材を採用してビジネスのイノベーションを期待するのには無理がある。

2.1.2. データサイエンティストの不足

データサイエンティストの育成が必要だとしても、我が国において必要な人材の概数はどのくらいだろうか。Mckinsey Global Institute のレポート^[17]では、“deep analytical talent”は 2018 年までに US で 14 万人から 19 万人不足する、としている。また、“data-savvy manager and analyst”については 150 万人不足する、としている。これらの数字は、すべての企業がフルに big data 分析を行うようになった場合にどうなるか、という仮定の元に割り出したデマンドであり、いわばこれが上限の数字であろう。

2012 年 10 月に東京で行われた Gartner Symposium/ITxpo 2012 では、ガートナー社のシニアバイスプレジデントのソnderガード氏が「日本ではビッグデータ関連の雇用が 36 万 5000 人分増える見込み」だ、しかし「実際 に雇用条件を満たせる人材は 11 万人程度しかいない」と述べた、と報告されている^[32]。

これは「いつまでに」36万5000人という数字か不明だが、McKinseyと同じ2018年くらいのタイムフレームを仮定しているのであれば、米国と日本の情報産業のサイズの比から言って、McKinseyと（オーダーとしては）変わらない数字、と考えることができる。

これを受けたのかどうか正確なところはわからないが、日経新聞の平成25年7月17日付けの記事^[29]には、「米調査会社ガートナーは将来的に国内ではデータサイエンティストが約25万人不足すると予測する。」という表現がある。36万人から11万人を引き算すれば、25万人になるので、日経がこのソルダーガード氏の講演に基づいて記事を書いた、ということも考えられるが、「ビッグデータ関連の雇用」と「データサイエンティスト」は同じものとは言えないだろう。ビッグデータ関連の雇用の中には、データセンターの運用に関わる人材、データ入力に関わる人材、データ処理のためのプログラム開発やプロジェクトマネジメントに関わる人材などが含まれていると思われるからである。

2013年1-3月期の法人企業統計^[33]によれば、資本金1億円以上の企業は国内におよそ31,000社である。これらの企業がそれぞれ少なくとも一人のデータサイエンティストを必要とするとすれば、数万のオーダーの人材が必要と言えるだろう。調査会社の数字より一桁小さいが、このあたりが我々にとってはリーズナブルな仮説なのではないかと考えた。

2.1.3. 事業の基本理念

上記McKinseyやO'Reillyの主張が、我が国におけるデータサイエンティストの実情に当てはまるのかどうかは、議論の余地がある。我々は、本事業を通して、「我が国におけるデータ利活用人材のあるべき姿」を明らかにし、それに基づいて戦略的に人材育成のエコシステムを創りだしていきたい。このため、本事業においては以下の3点を基本理念とした。

スケーラビリティ

最終的に数万人規模のデータサイエンティストの育成が必要であったとしても、一つの機関ないしはプロジェクトがこれらの人材を輩出することはできない。必要なのはスケールする仕組みである。このため、本プロジェクトではデータサイエンティスト育成に熱意を持つ教育機関と、データサイエンティストのスキルを利用したい企業・組織を広くネットワークし、それらの間で知識・経験を共有することで、多くのデータサイエンティストが育成され、有効に活用されることを狙う。

このため、本プロジェクトでは、東京大学を始めとする教育機関、データサイエンティストを活用する企業・研究機関、それにデータサイエンティストのコミュニティを形作るNPOなどを広くネットワークする。

出口戦略

人材育成の取り組みにおいて、もう一つ大事なことは出口戦略である。いくら大人数のデータサイエンティストが育成されたとしても、彼らの就職先が無ければ意味のないものになってしまう。データサイエンティストとしてのキャリアパスは何か、その青写真まで含めて、データサイエンティストのエコシステム全体を考えることが大切である。

我が国の実情に即した方策

米国と日本では、産業構造、労働市場、教育の仕組みや国民のメンタリティなどが異なる。米国の事情は参考にするが、単に米国発のトレンドの後追いをするのではなく、我が国の実情に即した「あるべき姿」を追求する。

2.2. 事業の概要

本事業は、下記の5つのワークストリームから構成する。

1. 認知度向上・啓発：人材育成ネットワークを円滑に形成するために、認知度の向上を試みるとともに、啓発活動を通じて、広くステークホルダーに周知する。
2. 人材ローテーション：データサイエンティストにとって、現場力が最も重要であり、分析の知識だけではなく、実際にデータに触れる経験が必要である。そのため、各企業・研究機関における人材のローテーションを積極的に行う必要がある。そこでこのプロジェクトでは、どのような人材ローテーションを行えば、データサイエンティストの育成に効果的であるかを、人材ローテーションのパイロットを実施して、その結果から有効性の高いプログラムを模索し、最終的には本事業の終了後も持続可能な仕組みを構築する。
3. ベストプラクティスの調査：データサイエンティストがどのように育成され、どのように活用されているかについて、各企業や研究機関へ人材育成に関するベスト・プラクティスの調査をおこなう。そして、これらのベスト・プラクティスが、ある程度一般化・類型化できるのであれば、データサイエンティスト育成のためのプログラムや教材等の開発が可能となる。同時に、データサイエンティストを活用する際の、スキルレベルの標準化にも寄与すると考えられる。
4. 教材の開発：データサイエンティスト育成のための教材については、各企業がこぞって開発をしているが、定式化された教材がみあたらない。というのも、企業それぞれの目的に合わせた育成教材が開発されているからである。そこで統計数理研究所では、研究所の人的リソースを最大限に活用した「モデル教材」を開発する。
5. 海外連携・標準化：3事業年以後、長期的な海外連携・標準化の活動を模索する。

2.3. 平成25年度の成果と知見

以下、上記5つのワークストリームについて、平成25年度の成果と知見について記す。プロジェクトの初年度である本年度では、特に上記3のベストプラクティスの調査に力を入れた。なお、各項目の詳細については、第3部「事業実施内容の詳細」を参照されたい。

2.3.1. 認知度向上・啓発

本年度は、5回のシンポジウム・ワークショップ等の主催・共催を行った。また、データサイエンティスト協会立ち上げへの協力を行い、その様子は、平成25年7月16日のNHK全国ニュースで放映された。また、事業代表の樋口、実施責任者の丸山などが、多くの講演・執筆活動を行い、データサイエンティスト育成と、その活用についての議論を深めることの重要性の周知に務めた。また、事業のパンフレット及びWebサイトを制作した。

2.3.2. 人材ローテーション

本年度は、協力機関である東京大学大学院理工学系研究科の支援を得て、5名の学生のインターンシップを行った。平成25年9月25日に東京大学で行った説明会には約40名の学生が参加し、学生の間におけるデータサイエンティストへの関心が高いことを伺わせた。

インターンシップ受入機関として、国立研究機関3研究所ならびに私企業4社が協力を申し出てくれたが、受入機関が求める勤務形態と、学生の学業のスケジュールの折り合いがつかないなど、条件が合わ

ない点が多々あり、結果として 5 名の学生が 1 企業 2 機関でのインターンシップをするにとどまった。

来年度以降は、予め定形の覚書などを用意し、学生の時間が比較的自由になる夏休みなどを狙って十分に準備期間をとって計画することが必要である。

2.3.3. ベスト・プラクティスの調査

本年度は、このベスト・プラクティスの調査に力を入れた。この調査は、統計検定合格者へのある程度の数量的規模のあるアンケート調査と、企業・機関への聞き取り調査の組み合わせで行った^[13, 27]。

アンケート調査の結果

アンケートは、2013 年 11 月に行われた統計検定試験の合格者のうち、企業に勤める方に任意で依頼する形で行った。有効回答数は 319 であり、うち約 87.5%が男性であった。回答者の年代別構成を図 2-1 に示す。また、質問内容は以下の通りである。

- Q 1- 3: デモグラフィー
- Q 4- 6: 業種・職種
- Q 7-10: 仕事上のデータ分析の頻度と内容
- Q11-18: 情報、統計、ビジネススキルそれぞれの必要性と、習得方法
- Q19-20: キャリアパス

これらの回答を分析した結果、業種とデータ分析の頻度との関係においては、製造業・金融業においてデータ活用が進んでいるが、サービス業は官公庁ではデータの利用はあまり進んでいないことがわかった。また、データ分析のスキルは、主に独学と仕事上の OJT で学んでいて、系統的な教育を受けていない様子が浮かび上がってきた。

キャリアについての回答結果を図 2-2 に示す。現在自身の能力が十分に発揮されていると思う回答者は約 50%だが、データ活用の専門家としてキャリアを積みたいと思う人は 60%を超えた。

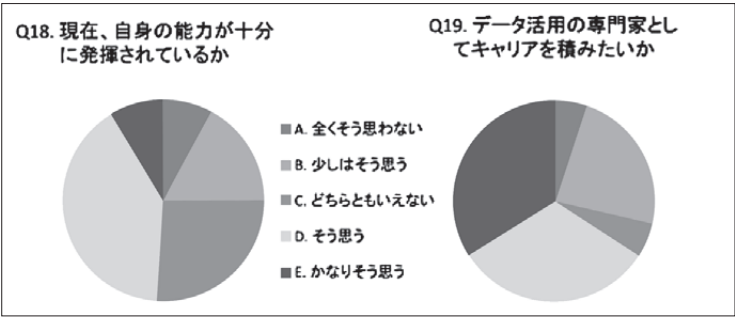


図 2-2 キャリアについての考え方

さらにキャリアを積みたい理由を問うたところ、図 2-3 のような回答を得た。

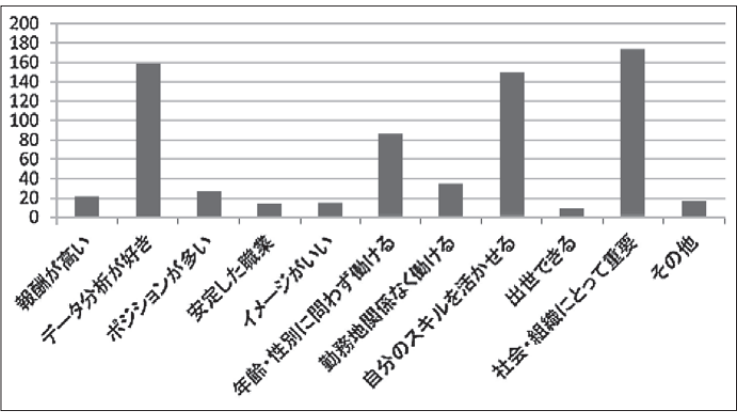


図 2-3 データ活用の専門家としてキャリアを積みたい理由

クラスタリングを用いてこれらの回答を 4 つのグループに分類した。それぞれに対して、特徴を捉えながら解釈すると以下の様に捉えることができる。

1. まだ実務経験は少ない若手だが、データサイエンティストになりたいという夢を持っているグループ。
2. 主にメーカーの製品開発・企画部門にいる中堅のエンジニアのグループ。自分の仕事の中で発生するデータを自分で分析する。社内では確実にデータの活用が進んでいる。キャリアパスも見えている。
3. 主に中小のサービス系の企業に勤めているが、比較的自由になる勤務形態として、データサイエンティストを望んでいるグループ。女性の比率が高いのが特徴。
4. IT サービス業等で、他者に対するデータ分析をプロとして長年実施してきているグループ。この仕事に誇りを持っている。

このクラスタリングの結果は、仕事の内容（自分のデータを分析するか、または他人のデータ分析を請け負うのか）と、目標とするキャリアパスに依存していて、2.1.1.節で述べた O'Reilly による米国でのデータサイエンティストの調査が、主に持っているスキルと経歴による類型になっているのと対照的である。調査の質問項目の違い、という理由もあるだろうが、短期的な即戦力を求める米国の労働市場と、長期的なキャリアパスを考えながら育成する日本の労働市場の違いを反映しているものとも考えられる。

聞き取り調査の結果

聞き取り調査は 20 件の企業・機関で「データサイエンティスト」と呼ばれる人々が、どのようなスキルを持ち、どのような役割で仕事をし、どのようなキャリアパスを考えているか、について伺った。対象者の業種は金融、製造、流通、公共、IT ベンダー、コンサルティング会社など多岐にわたり、会社の規模も大企業、ベンチャー企業、からフリーランスの方までをカバーした。その結果わかったことは、下記の 4 点である。

1. データサイエンティストのバックグラウンドは様々であること。現在、データサイエンティストとして活躍されている方々の出身は、理系も文系もあり、様々である。
2. データサイエンティストは全人的な能力であること。データサイエンティストの能力は、データ分析だけでなく、ビジネスの問題を発見してデータ分析の問題に落とし込む力、またデータ分析の結果を現場に適用してビジネスの成果につなげるためにコミュニケーション力も要求される。
3. データ分析は、個人の能力というよりは、組織の能力であること。「データサイエンティスト」という

- 個人がデータ分析のすべての局面を担当するのではなく、チームとしてそれぞれの得意分野を担当している。また、社員全員が基本的なデータ分析スキルを持つように教育している会社もある。
4. 発注側のリテラシーが大切なこと。どんなによいデータ分析結果を持っていても、経営者がそれに基づいて意思決定できなければ何にもならない。このため、発注側のリテラシーが重要である。

2.3.4. 教材の開発

データサイエンティスト育成には現場経験が必要だが、現場経験をする事前準備として最低限の知識を身につけて置かなければならない。このため、「データサイエンティスト育成クラッシュコース」と名づけたおよそ 2.5 時間のオンライン教材を準備した。この教材は 8 つのユニットからなり、データサイエンティストとは何か、から始まり、データサイエンティストが知っておくべき基本的な技法（統計、機械学習、最適化等）をひと通りカバーしている。受講者は、データ分析の現場で現れるキーワードをひと通り耳にすることができ、必要に応じて各ユニットの最後に提示される教科書や Web サイトなどの参考文献を参照できるように設計した。

本教材は Web 上で広く公開し、今後フィードバックを得て利用を拡大していただく予定である。

2.3.5. 海外連携・標準化

平成 26 年 3 月 6 日～7 日に、米国西海岸の数社を訪れ、シリコンバレーにおけるデータサイエンティストについての聞き取り調査を行った。その結果、米国におけるデータサイエンティストは、「データ・プロダクト」と呼ばれる、データ分析を使ったシステムを自身で構築できることが求められていて、そのためにプログラミングのスキルが必須であるということであった。この状況は、日本におけるデータサイエンティストが主にデータ分析のサービスを提供する職種であることと、対照的である。

日本において、「データ・プロダクト」を指向したデータサイエンティストが増えるのかどうか、あるいは我が国ではサービス指向のデータサイエンティストの育成に注力すべきであるかは、今後議論の必要がある。

2.4. 今後の方針

残りの 2 事業年において、今までの調査結果をもとに、我が国におけるデータ分析人材と、その育成のための「あるべき姿」を提言し、それに基づいて効果的かつ持続可能はデータサイエンティスト育成の仕組みを構築していく予定である。

3. 事業実施内容の詳細

3.1. 認知度向上・啓発

1. 広報用資料作成

- プロジェクトロゴの作成
(クラウドソーシングでフリーランサーを活用)
- リーフレットの作成 (2000 部)

2. ウェブサイトの構築

<http://datascientist.ism.ac.jp/>

3. シンポジウム

- イノベーションジャパン 2013 出展 (2013.8.29-30 東京ビッグサイト)
- 次世代リーディングシンポジウム 共催 (2014.1.27 一橋学術講堂・如水会館)
- Global Design Symposium : 知識創生 協賛 (2014.3.17 東京大学・福武ホール)

4. ワークショップ

- 「データサイエンティストネットワークの形成」事業紹介,サイエンティフィック・システム研究会 (2013.10.24 ホテルオークラ神戸)
- 「ビッグデータ利活用人材育成に向けたデータサイエンス教育の体系化：産学官連携の必要性」, 統計教育ワークショップ, 主催：統計教育大学間連携ネットワーク, 統計数理研究所, 2014.3.14
- 「文部科学省委託事業 データサイエンティスト育成ネットワークの形成」, ビッグデータ人材育成開発教材報告会, 慶應義塾大学, 2014.3.17

5. 学会発表等

- 丸山宏, 樋口知之, 竹村彰通『「データサイエンティスト育成ネットワークの形成」事業の概要』, 横幹連合シンポジウム, 香川大学, 2013.12.21-22
- 井川甲作, 丸山宏, 藤田茂, 藤原海世「データサイエンティストの現状と課題 ベストプラクティス調査」, 横幹連合シンポジウム, 香川大学, 2013.12.21-22

6. 講演

- 丸山宏「データに基づく意思決定ービッグデータブームの裏にあるものー」, SPES2013, 秋葉原 UDX, 2013.7.24
- Higuchi, T. Panel Discussion "Key Issues in Big Data". IEEE International Conference on Big Data 2013.
- 樋口知之「ビッグデータと統計学をつなぐもの」, 第 11 回地球シミュレーションシンポジウムーさらなる社会・未来・安心への貢献を目指してー, 秋葉原コンベンションホール, 2013.9.12
- 丸山宏「データに基づく意思決定ービッグデータブームの裏にあるものー」, 芝浦工業大学 MOT シンポジウム, 芝浦工業大学田町キャンパス, 2013.11.1
- 樋口知之「ビッグデータがせまる人への投資とサービスの高度化」, ビッグデータ&オープンデータ研究会設立キックオフイベント, 福岡 SRP, 2013.12.2
- 樋口知之「ビジネスにおけるビッグデータと統計学～受益者起点と予測モデル～」日経産業新聞フォーラム 2013, ベルサール八重洲, 2013.12.3



・丸山宏「文部科学省データサイエンティスト 育成ネットワーク形成事業」, データサイエンスサミット, ウェスティン東京, 2013.12.12

・丸山宏「データサイエンティスト育成ネットワークの形成 ～我が国におけるデータ分析人材の現状と展望～」, ビッグデータに関する研究開発シンポジウム ～ビッグデータによる新産業・イノベーションの創出に向けて～ (総務省・文部科学省・経済産業省 3 省合同シンポジウム), ベルサール飯田橋駅前<イベントホール>HALL「A」, 2014.3.26

7. 雑誌・記事等

- ・丸山宏『データ分析でできること、できないこと』, THINK!, pp26-33, 東洋経済新報社, 2013.8.1
- ・樋口知之『ビッグデータ 本当の破壊力』日経ビジネス 2013 年 9 月 30 日号
- ・丸山宏『ビッグデータの本質とは～データを活用する社会の実現に向けた課題と可能性～』
- ・NTT データ「DIGITAL GOVERNMENT」WEB 記事, 2013 年 10 月 3 日配信
http://e-public.nttdata.co.jp/topics_detail2/contents_type=2&id=968 (2014 年 3 月 1 日閲覧)
- ・樋口知之『ビジネスにおけるビッグデータと統計学～受益者起点と予測モデル～』 日経産業新聞フォーラム 2013. 2013.12.3
- ・樋口知之『ビッグデータ活用には「同床異夢」の解消が不可欠だ』日経 BP IT-Pro WEB 記事, 2013 年 12 月 13 日配信
<http://itpro.nikkeibp.co.jp/article/Interview/20131204/522593/> (2014 年 3 月 1 日閲覧)

3.2. 人材ローテーション

データサイエンティストにとって、現場力が最も重要であり、分析の知識だけではなく、実際にデータに触れる経験が必要である。そのため、各企業・研究機関における人材のローテーションを積極的に行う必要がある。そこで本プロジェクトでは、どのような人材ローテーションを行えば、データサイエンティストの育成に効果的であるかを、人材ローテーションのパイロットを実施して、その結果から有効性の高いプログラムを模索する。

人材のローテーション (再委託先：国立大学法人東京大学)

企業・研究機関へパイロットインターンするため、東京大学工学部の学部学生、情報理工学系研究科修士学生に協力頂いた。とくに研究機関へ派遣の際は、統計数理研究所 外来研究員の身分を付与し、修了者には認定証等の発行を検討中である。今後の予定としては、3 月下旬までにインターン完了となり、学生・受入担当者からヒアリングをおこない、5 月を目途にレポートをまとめる予定である。

3.3. ベスト・プラクティスの調査

データサイエンティストがどのように育成され、どのように活用されているかについてベスト・プラクティスを調査し、共有することは、このプロジェクトの主眼の一つである。これらのベスト・プラクティスがある程度一般化できるのであれば、データサイエンティスト育成のためのプログラム、あるいは教材の開発が可能なのはである。また、データサイエンティスト活用にあたって、そのスキルレベルを標準化することも必要となってくるだろう。

3.3.1. 統計検定合格者へのアンケート調査¹

平成 25 年 11 月 17 日に実施した「統計検定」合格者のうち社会人を対象として調査を行った。「統計検定」とは、統計に関する知識や活用力を評価する全国統一試験である。日本統計学会は、中高生・大学生・職業人を対象に、以下の各レベルに応じて体系的に国際通用性のある統計活用能力評価システムを研究開発し、統計検定として資格認定を行っている²。統計検定試験の種別には、以下の 7 種類がある。

- 1 級 大学専門課程の統計学の知識を専門分野で遂行する能力
- 2 級 大学基礎課程としての統計学の知識と問題解決能力
- 3 級 高校卒業段階までに求められる統計活用力
- 4 級 中学校卒業段階までに求められる統計活用力

統計調査士 統計調査実務に関連する基本的知識

専門統計調査士 統計調査全般に関わる高度な専門的知識

国際資格 (RSS/JSS) 英国王立統計学会 (Royal Statistical Society) との共同認定

調査方法

対象者 本調査対象者は、平成 25 年 11 月 17 日に実施された統計検定試験 1 級から 4 級, 統計調査士, 専門統計調査士の計 6 種類の合格者 780 名であった。

手続き 対象者は、Web サイト、あるいは紙ベースの調査票での回答を選択した。Web サイトで回答する場合、対象者は、<http://goo.gl/5Kbwdi> に各自アクセスして回答の入力をした。紙ベースの調査票で回答する場合、対象者は回答を記入後、郵便で調査票を返送した。回答期間は、平成 25 年 12 月 17 日から平成 26 年 1 月 21 日であった。Web サイトにおける回答は 160 名、紙ベースの調査票に対する回答は 159 名、計 319 名 (男性 279 名 (年齢内訳 20～29 歳：50 名、30～39 歳：84 名、40～49 歳：93 名、50～59 歳：46 名、60～69 歳：6 名、女性 38 名 (年齢内訳 20～29 歳：7 名、30～39 歳：20 名、40～49 歳：10 名、50～59 歳：0 名、60～69 歳：1 名、性別回答しない 2 名 (年齢内訳 20～29 歳：1 名、30～39 歳：1 名) から回答を得た。

調査項目は、以下のとおりであった。依頼状と調査票は付録 A に掲載した。

質問 1 受験番号

<基本属性 質問 2～質問 6 >

質問 2 年齢 10 歳階級 (7 区分)

質問 3 男女別 (3 区分 (「回答しない」を含む))

質問 4 所属組織の産業など (15 区分)

質問 5 【所属のある人】所属組織の職種など (13 区分)

質問 6 【所属のある人】所属組織の従業員規模 (7 区分)

<業務でのデータ活用 質問 7～質問 11>

質問 7 【所属のある人】データ活用の頻度 (5 区分)

質問 8 【所属のある人、かつデータ活用の機会がある人】

利用データの種類 (4 種類、複数回答)

¹ 統計検定センターから納品された調査報告書の内容を改編した (参考文献一覧を参照)。

² 統計検定試験の詳細や統計検定創設の経緯 <http://www.toukei-kentei.jp/about/index.html>

質問9 【所属のある人，かつデータ活用機会がある人】
データ活用の担当（7種類，複数回答）
質問10 【所属のある人，かつデータ活用機会がある人】
データ活用の重要度（5区分）
質問11 【所属のある人，かつデータ活用機会がある人】
データ活用のポジション（4種類，複数回答）
＜各種スキルと学習方法 質問12～質問17＞
質問12 利用可能な統計系スキル（9種類，複数回答）
質問13 統計系スキルの学習方法（8種類，複数回答）
質問14 利用可能な情報系スキル（8種類，複数回答）
質問15 情報系スキルの学習方法（8種類，複数回答）
質問16 利用可能なビジネス系スキル（9種類，複数回答）
質問17 ビジネス系スキルの学習方法（8種類，複数回答）
＜専門家としてのキャリア意向 質問18～質問20＞
質問18 自分の能力の発揮度（5区分）
質問19 専門家としてのキャリア意向（5区分）
質問20 【専門家としてのキャリア意向が少しでもある人】キャリア意向の理由（11区分）

調査結果
スキルの学習方法 データサイエンティストに必要と考えられる3種類のスキル（統計系スキル、情報系スキル、ビジネス系スキル）の学習方法は、全体的に見ると、どのスキルについても「独学」、次いで「OJT」という回答が多い傾向にあった。ただし、学習方法の組み合わせには、スキル毎に傾向が異なった。統計系スキル（表 3-1）では、独学（書籍）のみ78名（24.5%）、大学・大学院と独学（書籍）の組合せ25名（7.8%）、独学（WEB教材）と独学（書籍）の組合せ24名（7.5%）、研修とOJT、独学（書籍）の組合せの23名（7.2%）の順であった。したがって、このスキルの獲得には、教育機関が一定程度利用されていることが推測された。

表 3-1 統計系スキルの学習方法のパターン，統計検定種類別人数

質問13 統計系スキルの学習方法 選択パターン	総 数 ^(注1)		1級, 2級S/A, 専門統計調査士, 統計調査士		2級その他, 3級, 4級	
総 数	319		77		233	
独学(書籍)のみ 00000100	78	24.5%	15	19.5%	62	26.6%
大学・大学院, 独学(書籍) 10000100	25	7.8%	9	11.7%	14	6.0%
独学(WEB教材), 独学(書籍) 00001100	24	7.5%	4	5.2%	20	8.6%
研修, OJT, 独学(書籍) 00110100	23	7.2%	6	7.8%	17	7.3%
OJT, 独学(書籍) 00010100	20	6.3%	4	5.2%	16	6.9%
どのスキルも持っていない 00000001	17	5.3%	5	6.5%	11	4.7%
OJT, 独学(WEB教材), 独学(書籍) 00011100	11	3.4%	0	0.0%	11	4.7%
研修, 独学(書籍) 00100100	11	3.4%	2	2.6%	7	3.0%
大学・大学院, 研修, 独学(書籍) 10010100	10	3.1%	6	7.8%	4	1.7%
：						

注1 総数は受験級が不明の9名を含む。
注2 「1」は選択した状況, 「0」は選択していない状況を表す. 例えば「100000000」の場合, 「大学・大学院」のみを選択していることを表す.
注3 選択パターンは度数の大きいものから順に並べて, 10名以上の選択パターンのみを示す.

情報系スキル（表 3-2）では、「どのスキルも持っていない」を除けば、独学（書籍）のみ49名（15.4%）が最も多く、つづいて独学（書籍）と独学（WEB教材）の組合せ26名（8.2%）、OJTと独学（書籍）、独学（WEB教材）の組合せ22名（6.9%）の順であった。したがって、「どのスキルも持っていない」という回答が最も多かったことに注意する必要があるが、このスキルの獲得には、独学が主要な学習方法であることが推測された。

表 3-2 情報系スキルの学習方法のパターン，統計検定種類別人数

質問15 情報系スキルの学習方法 選択パターン	総 数 ^(注1)		1級, 2級S/A, 専門統計調査士, 統計調査士		2級その他, 3級, 4級	
総 数	319		77		233	
どのスキルも持っていない 00000001	56	17.6%	19	24.7%	36	15.5%
独学(書籍)のみ 00000100	49	15.4%	14	18.2%	35	15.0%
独学(WEB教材), 独学(書籍) 00001100	26	8.2%	2	2.6%	24	10.3%
OJT, 独学(WEB教材), 独学(書籍) 00011100	22	6.9%	6	7.8%	16	6.9%
OJT, 独学(書籍) 00010100	21	6.6%	5	6.5%	15	6.4%
研修, OJT, 独学(書籍) 00110100	13	4.1%	2	2.6%	10	4.3%
大学・大学院, 独学(書籍) 10000100	12	3.8%	2	2.6%	9	3.9%
研修, OJT 00110000	11	3.4%	2	2.6%	9	3.9%
研修, OJT, 独学(WEB教材), 独学(書籍) 00111100	11	3.4%	2	2.6%	9	3.9%
：						

注1 総数は受験級が不明の9名を含む。
注2 「1」は選択した状況, 「0」は選択していない状況を表す。例えば「1000000000」の場合, 「大学・大学院」のみを選択していることを表す。
注3 選択パターンは度数の大きいものから順に並べて, 10名以上の選択パターンのみを示す。

ビジネス系スキル（表 3-3）では、独学（書籍）のみ 42 名（13.2%）が最も多く、OJT と独学（書籍）の組合せ 31 名（9.7%）、どのスキルも持っていない 28 名（8.8%）、研修と OJT の組合せ 23 名（7.2%）、研修と OJT, 独学（書籍）の組合せ 23 名（7.2%）の順であった。したがって、このスキルの獲得には、独学と OJT が主要な学習方法であることが推測された。

表 3-3 ビジネス系スキルの学習方法のパターン，統計検定種類別人数

質問17 ビジネス系スキルの学習方法 選択パターン	総 数 ^(注1)		1級, 2級S/A, 専門統計調査士, 統計調査士		2級その他, 3級, 4級	
総 数	319		77		233	
独学(書籍)のみ 00000100	42	13.2%	5	6.5%	36	15.5%
OJT, 独学(書籍) 00010100	31	9.7%	10	13.0%	20	8.6%
どのスキルも持っていない 00000001	28	8.8%	9	11.7%	18	7.7%
研修, OJT 00110000	23	7.2%	3	3.9%	18	7.7%
研修, OJT, 独学(書籍) 00110100	23	7.2%	7	9.1%	16	6.9%
独学(WEB教材), 独学(書籍) 00001100	18	5.6%	4	5.2%	14	6.0%
OJTのみ 00010000	17	5.3%	7	9.1%	10	4.3%
OJT, 独学(WEB教材), 独学(書籍) 00011100	14	4.4%	4	5.2%	10	4.3%
大学・大学院, 独学(書籍) 10000100	11	3.4%	4	5.2%	6	2.6%
：						

注1 総数は受験級が不明の9名を含む。
注2 「1」は選択した状況, 「0」は選択していない状況を表す。例えば「1000000000」の場合, 「大学・大学院」のみを選択していることを表す。
注3 選択パターンは度数の大きいものから順に並べて, 10名以上の選択パターンのみを示す。

利用可能なスキル 「統計検定」合格者が実務で利用可能なスキルについて、3 種類のスキル毎（統計系スキル、情報系スキル、ビジネス系スキル）に回答を求めた。

実務で利用可能な統計系スキル（表 3-4）は、古典的統計 64.6%、時系列統計 60.2%、可視化 43.9%の順であった。

表 3-4 利用可能な統計系スキル，統計検定種類別人数

質問12 利用可能な統計系スキル	総 数 ^(注1)		1級, 2級S/A, 専門統計調査士, 統計調査士		2級その他, 3級, 4級	
総 数	319		77		233	
古典的統計(例:線形モデル, ANOVA等)	206	64.6%	51	66.2%	149	63.9%
機械学習(例:決定木, SVM, クラスタリング等)	83	26.0%	16	20.8%	65	27.9%
最適化(例:線形, コンベックス等)	88	27.6%	15	19.5%	70	30.0%
シミュレーション(例:離散, エージェントベース等)	79	24.8%	9	11.7%	67	28.8%
空間統計(例:GIS)	33	10.3%	9	11.7%	24	10.3%
時系列統計(例:予測, 時系列分析等)	192	60.2%	41	53.2%	144	61.8%
グラフィカルモデル(例:ソーシャルネットワーク等)	23	7.2%	4	5.2%	19	8.2%
ベイズ定理/モンテカルロシミュレーション(例:MCMC, BUGS等)	64	20.1%	16	20.8%	46	19.7%
可視化	140	43.9%	30	39.0%	106	45.5%

注1 総数は受験級が不明の9名を含む。
注2 割合(%)は総数を分母とし, 複数回答のため合計は100%にならない。

情報系スキル（表 3-5）については、統計的プログラミング 53.0%、市販統計パッケージ 48.9%、構造化データ 32.9%の順であった。

表 3-5 利用可能な情報系スキル，統計検定種類別人数

質問14 利用可能な情報系スキル	総 数 ^(注1)		1級, 2級S/A, 専門統計調査士, 統計調査士		2級その他, 3級, 4級	
総 数	319		77		233	
バックエンドプログラミング (例:Java,Rails,Objective C等)	74	23.2%	16	20.8%	55	23.6%
WEB系プログラミング (例:JavaScript,HTML,CSS等)	79	24.8%	13	16.9%	64	27.5%
ビッグデータ関連技術 (例:Hadoop, Map/Reduce等)	69	21.6%	12	15.6%	55	23.6%
構造化データ(例:SQL, JSON)	105	32.9%	17	22.1%	86	36.9%
非構造化データ(例:noSQL等)	32	10.0%	7	9.1%	25	10.7%
統計的プログラミング(例:R等)	169	53.0%	37	48.1%	128	54.9%
市販統計パッケージ (例:SAS, SPSS等)	156	48.9%	41	53.2%	109	46.8%
システム運用(例:ネットワークエンジニア, DBA, システム管理等)	62	19.4%	9	11.7%	51	21.9%

注1 総数は受験級が不明の9名を含む。
注2 割合(%)は総数を分母とし, 複数回答のため合計は100%にならない。

ビジネス系スキル（表 3-6）は、マーケティング・企画 57.1%、マネジメント 44.5%、研究・開発 39.2%の順であった。

表 3-6 ビジネス系スキル，統計検定種類別人数

質問16 利用可能なビジネス系スキル	総 数 ^(注1)		1級, 2級S/A, 専門統計調査士, 統計調査士		2級その他, 3級, 4級	
総 数	319		77	233		
マネジメント	142	44.5%	22	28.6%	114	48.9%
ファイナンス	77	24.1%	11	14.3%	63	27.0%
マーケティング・企画	182	57.1%	47	61.0%	130	55.8%
営業	73	22.9%	11	14.3%	60	25.8%
生産・製造管理	72	22.6%	10	13.0%	60	25.8%
研究・開発	125	39.2%	33	42.9%	89	38.2%
ITシステム設計・構築・運用	106	33.2%	19	24.7%	85	36.5%
コンサルティング	118	37.0%	27	35.1%	87	37.3%
その他	20	6.3%	6	7.8%	12	5.2%

注1 総数は受験級が不明の9名を含む。
注2 割合(%)は総数を分母とし、複数回答のため合計は100%にならない。

データ活用の担当とデータ活用の頻度、ならびに産業や職種とデータ活用の頻度との関係 「1 級，2 級 S/A，専門統計調査士，統計調査士」の場合、より高度なデータ活用を担当している人ほど「ほぼ日常的にデータ活用を行っている」人が多かった（表 3-7）。

表 3-7 統計検定種類，データ活用の担当，データ活用の頻度別人数

質問9、質問7 データ活用の担当， データ活用の頻度	総 数 ^(注1)		全く ない	月1日程度	週1日程度	週2・3日程 度	ほぼ日常 的に行って いる	無回答
総数	261	100%		81 31%	50 19%	31 12%	93 36%	6 2%
データ分析の目的を把握し，データでどのように 解決できるかを検討する	160	100%		45 28%	30 19%	19 12%	64 40%	2 1%
必要なデータの収集を行う	159	100%		45 28%	28 18%	22 14%	62 39%	2 1%
データを整理し，分析が可能なようにクレンジン グする	149	100%		37 25%	22 15%	23 15%	66 44%	1 1%
データを分析するモデルを作成する	97	100%		16 16%	16 16%	18 19%	46 47%	1 1%
モデルに従い，取得したデータを分析し結果を 抽出する	123	100%		23 19%	21 17%	19 15%	59 48%	1 1%
データ分析の結果をもとに，実際の業務への導 入を行う	104	100%		16 15%	21 20%	20 19%	46 44%	1 1%
導入した結果を検証する	78	100%		8 10%	16 21%	14 18%	38 49%	2 3%
1級, 2級S/A, 専門統計調査士，統計調査士	60	100%		19 32%	10 17%	4 7%	26 43%	1 2%
データ分析の目的を把握し，データでどのように 解決できるかを検討する	38	100%		13 34%	4 11%	2 5%	19 50%	0 0%
必要なデータの収集を行う	36	100%		11 31%	3 8%	4 11%	18 50%	0 0%
データを整理し，分析が可能なようにクレンジン グする	34	100%		10 29%	2 6%	3 9%	19 56%	0 0%
データを分析するモデルを作成する	24	100%		5 21%	1 4%	2 8%	16 67%	0 0%
モデルに従い，取得したデータを分析し結果を 抽出する	28	100%		5 18%	2 7%	1 4%	20 71%	0 0%
データ分析の結果をもとに，実際の業務への導 入を行う	16	100%		3 19%	3 19%	1 6%	9 56%	0 0%
導入した結果を検証する	14	100%		1 7%	3 21%	1 7%	9 64%	0 0%
2級その他, 3級, 4級	192	100%		59 31%	39 20%	26 14%	63 33%	5 3%
データ分析の目的を把握し，データでどのように 解決できるかを検討する	115	100%		29 25%	25 22%	16 14%	43 37%	2 2%
必要なデータの収集を行う	116	100%		32 28%	24 21%	17 15%	41 35%	2 2%
データを整理し，分析が可能なようにクレンジン グする	109	100%		26 24%	19 17%	19 17%	44 40%	1 1%
データを分析するモデルを作成する	70	100%		11 16%	14 20%	15 21%	29 41%	1 1%
モデルに従い，取得したデータを分析し結果を 抽出する	91	100%		18 20%	18 20%	17 19%	37 41%	1 1%
データ分析の結果をもとに，実際の業務への導 入を行う	85	100%		13 15%	18 21%	18 21%	35 41%	1 1%
導入した結果を検証する	60	100%		7 12%	13 22%	12 20%	26 43%	2 3%

注1 総数は受験級が不明の9名を含み，質問4で「所属なし」あるいは質問7で「全くない」の58名を除く。
注2 割合(%)は行和を100%として，行の度数10以上の場合に総数あるいは検定種類別の全体と15ポイント以上の差のあるものを太字表示にした。

業種別にみると、製造業や金融・保険業でのデータ活用の頻度は高かった（表 3-8）。一方、サービス業や公務員・官公庁でのデータ活用の頻度は低かった。

表 3-8 統計検定種類，所属組織の産業など，データ活用の頻度別人数

質問4，質問7 所属組織の産業など， データ活用の頻度	総 数 ^(注1)		全くない	月1日程度	週1日程度	週2・3日程 度	ほぼ日常 的に行って いる	無回答
総数	310	100%	49 16%	81 26%	50 16%	31 10%	93 30%	6 2%
製造業	63	100%	6 10%	11 17%	11 17%	8 13%	25 40%	2 3%
出版・印刷業	5	100%	0 0%	2 40%	0 0%	0 0%	3 60%	0 0%
金融・保険業	28	100%	6 21%	2 7%	5 18%	2 7%	12 43%	1 4%
卸売・小売業	6	100%	0 0%	3 50%	1 17%	1 17%	1 17%	0 0%
卸売業(商社)	4	100%	0 0%	2 50%	2 50%	0 0%	0 0%	0 0%
IT関連業	80	100%	9 11%	22 28%	15 19%	11 14%	22 28%	1 1%
サービス業	20	100%	8 40%	3 15%	0 0%	3 15%	6 30%	0 0%
公務員・官公庁	29	100%	7 24%	11 38%	7 24%	2 7%	2 7%	0 0%
医療・福祉関連業	24	100%	3 13%	9 38%	4 17%	2 8%	6 25%	0 0%
建設・不動産業	6	100%	1 17%	0 0%	1 17%	0 0%	3 50%	1 17%
教育・保育・学校法人	18	100%	5 28%	7 39%	2 11%	0 0%	4 22%	0 0%
マスコミ・メディア関連業	6	100%	1 17%	1 17%	0 0%	2 33%	2 33%	0 0%
フリーランス	4	100%	1 25%	2 50%	0 0%	0 0%	0 0%	1 25%
その他の組織	16	100%	2 13%	5 31%	2 13%	0 0%	7 44%	0 0%
所属なし	1	100%	0 0%	1 100%	0 0%	0 0%	0 0%	0 0%
1級, 2級S/A, 専門統計調査士，統計調査士	75	100%	15 20%	19 25%	10 13%	4 5%	26 35%	1 1%
製造業	10	100%	1 10%	2 20%	1 10%	1 10%	5 50%	0 0%
出版・印刷業	2	100%	0 0%	1 50%	0 0%	0 0%	1 50%	0 0%
金融・保険業	7	100%	0 0%	2 29%	2 29%	1 14%	2 29%	0 0%
卸売・小売業	1	100%	0 0%	0 0%	0 0%	0 0%	1 100%	0 0%
卸売業(商社)	1	100%	0 0%	0 0%	1 100%	0 0%	0 0%	0 0%
IT関連業	10	100%	2 20%	4 40%	0 0%	0 0%	3 30%	1 10%
サービス業	11	100%	6 55%	1 9%	0 0%	1 9%	3 27%	0 0%
公務員・官公庁	8	100%	1 13%	2 25%	4 50%	0 0%	1 13%	0 0%
医療・福祉関連業	5	100%	0 0%	2 40%	1 20%	1 20%	1 20%	0 0%
建設・不動産業	1	100%	0 0%	0 0%	0 0%	0 0%	1 100%	0 0%
教育・保育・学校法人	9	100%	4 44%	1 11%	1 11%	0 0%	3 33%	0 0%
マスコミ・メディア関連業	1	100%	0 0%	0 0%	0 0%	0 0%	1 100%	0 0%
フリーランス	0 -		0 -	0 -	0 -	0 -	0 -	0 -
その他の組織	9	100%	1 11%	4 44%	0 0%	0 0%	4 44%	0 0%
所属なし	0 -		0 -	0 -	0 -	0 -	0 -	0 -
2級その他, 3級, 4級	226	100%	34 15%	59 26%	39 17%	26 12%	63 28%	5 2%
製造業	51	100%	5 10%	9 18%	10 20%	7 14%	18 35%	2 4%
出版・印刷業	3	100%	0 0%	1 33%	0 0%	0 0%	2 67%	0 0%
金融・保険業	19	100%	6 32%	0 0%	3 16%	1 5%	8 42%	1 5%
卸売・小売業	4	100%	0 0%	3 75%	1 25%	0 0%	0 0%	0 0%
卸売業(商社)	3	100%	0 0%	2 67%	1 33%	0 0%	0 0%	0 0%
IT関連業	69	100%	7 10%	17 25%	15 22%	11 16%	19 28%	0 0%
サービス業	8	100%	2 25%	1 13%	0 0%	2 25%	3 38%	0 0%
公務員・官公庁	20	100%	6 30%	8 40%	3 15%	2 10%	1 5%	0 0%
医療・福祉関連業	19	100%	3 16%	7 37%	3 16%	1 5%	5 26%	0 0%
建設・不動産業	5	100%	1 20%	0 0%	1 20%	0 0%	2 40%	1 20%
教育・保育・学校法人	9	100%	1 11%	6 67%	1 11%	0 0%	1 11%	0 0%
マスコミ・メディア関連業	5	100%	1 20%	1 20%	0 0%	2 40%	1 20%	0 0%
フリーランス	4	100%	1 25%	2 50%	0 0%	0 0%	0 0%	1 25%
その他の組織	6	100%	1 17%	1 17%	1 17%	0 0%	3 50%	0 0%
所属なし	1	100%	0 0%	1 100%	0 0%	0 0%	0 0%	0 0%

注1 総数は受験級が不明の9名を含み，質問4で「所属なし」の9名を除く。なお，質問4で「所属なし」10名のうち1名は回答あり。

注2 割合(%)は行和を100%として，行の度数5以上の場合に最も値の高いものを太字表示にした。

職務別にみると、調査や研究・開発でのデータ活用の頻度は高かった。一方、営業や IT 関係、その他の職種でのデータ活用の頻度は低かった（表 3-9）。

表 3-9 統計検定種類、所属組織の職種など、データ活用の頻度別人数

質問5、質問7 所属組織の職種など、 データ活用の頻度	総 数 ^(注1)													
	全くない		月1日程度		週1日程度		週2・3日程 度		ほぼ日常的 に行っている		無回答			
総数	310	100%	49	16%	81	26%	50	16%	31	10%	93	30%	6	2%
管理	29	100%	4	14%	5	17%	8	28%	3	10%	8	28%	1	3%
研究・開発	51	100%	3	6%	8	16%	10	20%	8	16%	22	43%	0	0%
調査	38	100%	2	5%	7	18%	1	3%	4	11%	24	63%	0	0%
企画	33	100%	4	12%	8	24%	9	27%	5	15%	6	18%	1	3%
営業	21	100%	6	29%	6	29%	3	14%	2	10%	2	10%	2	10%
製造	7	100%	0	0%	2	29%	1	14%	1	14%	3	43%	0	0%
総務	5	100%	4	80%	0	0%	1	20%	0	0%	0	0%	0	0%
人事	5	100%	1	20%	1	20%	1	20%	2	40%	0	0%	0	0%
経理	3	100%	1	33%	1	33%	0	0%	0	0%	1	33%	0	0%
コンサルティング	15	100%	1	7%	5	33%	1	7%	2	13%	6	40%	0	0%
IT関係	51	100%	6	12%	17	33%	10	20%	4	8%	12	24%	2	4%
その他の職種	49	100%	16	33%	20	41%	5	10%	0	0%	8	16%	0	0%
無回答	3	100%	1	33%	1	33%	0	0%	0	0%	1	33%	0	0%
1級、2級S/A, 専門統計調査士、統計調査士	75	100%	15	20%	19	25%	10	13%	4	5%	26	35%	1	1%
管理	4	100%	1	25%	2	50%	0	0%	1	25%	0	0%	0	0%
研究・開発	19	100%	0	0%	3	16%	4	21%	2	11%	10	53%	0	0%
調査	18	100%	2	11%	3	17%	1	6%	1	6%	11	61%	0	0%
企画	6	100%	2	33%	0	0%	2	33%	0	0%	2	33%	0	0%
営業	1	100%	0	0%	0	0%	1	100%	0	0%	0	0%	0	0%
製造	1	100%	0	0%	1	100%	0	0%	0	0%	0	0%	0	0%
総務	3	100%	3	100%	0	0%	0	0%	0	0%	0	0%	0	0%
人事	0	-	0	-	0	-	0	-	0	-	0	-	0	-
経理	3	100%	1	33%	1	33%	0	0%	0	0%	1	33%	0	0%
コンサルティング	3	100%	0	0%	2	67%	1	33%	0	0%	0	0%	0	0%
IT関係	6	100%	1	17%	2	33%	0	0%	0	0%	2	33%	1	17%
その他の職種	10	100%	5	50%	4	40%	1	10%	0	0%	0	0%	0	0%
無回答	1	100%	0	0%	1	100%	0	0%	0	0%	0	0%	0	0%
2級その他、3級、4級	226	100%	34	15%	59	26%	39	17%	26	12%	63	28%	5	2%
管理	24	100%	3	13%	3	13%	8	33%	2	8%	7	29%	1	4%
研究・開発	31	100%	3	10%	5	16%	6	19%	6	19%	11	35%	0	0%
調査	20	100%	0	0%	4	20%	0	0%	3	15%	13	65%	0	0%
企画	26	100%	2	8%	8	31%	7	27%	5	19%	3	12%	1	4%
営業	19	100%	6	32%	6	32%	2	11%	1	5%	2	11%	2	11%
製造	6	100%	0	0%	1	17%	1	17%	1	17%	3	50%	0	0%
総務	2	100%	1	50%	0	0%	1	50%	0	0%	0	0%	0	0%
人事	4	100%	1	25%	1	25%	0	0%	2	50%	0	0%	0	0%
経理	0	-	0	-	0	-	0	-	0	-	0	-	0	-
コンサルティング	12	100%	1	8%	3	25%	0	0%	2	17%	6	50%	0	0%
IT関係	44	100%	5	11%	14	32%	10	23%	4	9%	10	23%	1	2%
その他の職種	36	100%	11	31%	14	39%	4	11%	0	0%	7	19%	0	0%
無回答	2	100%	1	50%	0	0%	0	0%	0	0%	1	50%	0	0%

注1 総数は受験級が不明の9名を含み、質問4で「所属なし」の9名を除く。なお、質問4で「所属なし」10名のうち1名は回答あり。

注2 割合(%)は行和を100%として、行の度数5以上の場合に最も値の高いものを**太字**表示にした。

データ活用の重要度とデータ活用の頻度、ならびにデータ活用のポジションとデータ活用の頻度 データ活用を重要であると捉えている組織であるほど、データ活用の頻度が高かった（表 3-10）。

表 3-10 統計検定種類、データ活用の重要性、データ活用の頻度別人数

質問10、質問7 データ活用の重要度、 データ活用の頻度	総 数 ^(注1)											
	全くない		月1日程度		週1日程度		週2・3日程度		ほぼ日常的に行っている		無回答	
総数	261	100%	81	31%	50	19%	31	12%	93	36%	6	2%
全く活用されていない	10	100%	7	70%	0	0%	1	10%	2	20%	0	0%
全く活用されていないが、活用しようとする動きがある	27	100%	18	67%	3	11%	2	7%	4	15%	0	0%
決算発表として利用する程度で、日々の業務では活用されていない	13	100%	3	23%	4	31%	1	8%	5	38%	0	0%
まれに、データの分析、提案が求められる	88	100%	29	33%	23	26%	15	17%	19	22%	2	2%
マネジメントにおいてデータを重視しており、定期的にデータの分析、提案が求められている	100	100%	17	17%	15	15%	10	10%	58	58%	0	0%
無回答	23	100%	7	30%	5	22%	2	9%	5	22%	4	17%
1級、2級S/A, 専門統計調査士、統計調査士	60	100%	19	32%	10	17%	4	7%	26	43%	1	2%
全く活用されていない	2	100%	1	50%	0	0%	0	0%	1	50%	0	0%
全く活用されていないが、活用しようとする動きがある	4	100%	2	50%	1	25%	1	25%	0	0%	0	0%
決算発表として利用する程度で、日々の業務では活用されていない	3	100%	1	33%	1	33%	0	0%	1	33%	0	0%
まれに、データの分析、提案が求められる	19	100%	8	42%	4	21%	1	5%	6	32%	0	0%
マネジメントにおいてデータを重視しており、定期的にデータの分析、提案が求められている	25	100%	6	24%	2	8%	2	8%	15	60%	0	0%
無回答	7	100%	1	14%	2	29%	0	0%	3	43%	1	14%
2級その他, 3級, 4級	192	100%	59	31%	39	20%	26	14%	63	33%	5	3%
全く活用されていない	6	100%	5	83%	0	0%	0	0%	1	17%	0	0%
全く活用されていないが、活用しようとする動きがある	21	100%	14	67%	2	10%	1	5%	4	19%	0	0%
決算発表として利用する程度で、日々の業務では活用されていない	10	100%	2	20%	3	30%	1	10%	4	40%	0	0%
まれに、データの分析、提案が求められる	68	100%	21	31%	18	26%	14	21%	13	19%	2	3%
マネジメントにおいてデータを重視しており、定期的にデータの分析、提案が求められている	71	100%	11	15%	13	18%	8	11%	39	55%	0	0%
無回答	16	100%	6	38%	3	19%	2	13%	2	13%	3	19%

注1 総数は受験級が不明の9名を含み、質問4で「所属なし」あるいは質問7で「全くない」の58名を除く。

注2 割合(%)は行和を100%として、行の度数10以上の場合に総数あるいは検定種類別の全体と15ポイント以上の差のあるものを**太字**表示にした。

また、データ活用を行うメンバーの管理者ポジションが存在する組織では、データ活用の頻度が高かった（表 3-11）。

表 3-11 統計検定種類，データ活用のポジションなど，データ活用の頻度別人数

質問11, 質問7 データ活用のポジション, データ活用の頻度	総 数 ^(注1)		全く ない	月1日程度		週1日程度		週2・3日程 度		ほぼ日常 的に行って いる		無回答	
総数	261	100%		81	31%	50	19%	31	12%	93	36%	6	2%
データ活用専門のポジションは存在せず、通常 のマネジメントポジションのみ	151	100%		61	40%	29	19%	21	14%	38	25%	2	1%
データ活用を行う上級のポジションが存在する	22	100%		4	18%	5	23%	3	14%	10	45%	2	9%
データ活用を行うメンバーの管理者となるポジ ションが存在する	52	100%		9	17%	4	8%	6	12%	33	63%	1	2%
データ活用専門ポジションではないが、専門家 ポジションがある	49	100%		10	20%	9	18%	6	12%	24	49%	1	2%
1級, 2級S/A, 専門統計調査士, 統計調査士	60	100%		19	32%	10	17%	4	7%	26	43%	1	2%
データ活用専門のポジションは存在せず、通常 のマネジメントポジションのみ	28	100%		12	43%	2	7%	3	11%	11	39%	0	0%
データ活用を行う上級のポジションが存在する	7	100%		1	14%	1	14%	0	0%	5	71%	0	0%
データ活用を行うメンバーの管理者となるポジ ションが存在する	13	100%		1	8%	2	15%	1	8%	9	69%	0	0%
データ活用専門ポジションではないが、専門家 ポジションがある	13	100%		4	31%	4	31%	0	0%	5	38%	0	0%
2級その他, 3級, 4級	192	100%		59	31%	39	20%	26	14%	63	33%	5	3%
データ活用専門のポジションは存在せず、通常 のマネジメントポジションのみ	116	100%		46	40%	26	22%	17	15%	25	22%	2	2%
データ活用を行う上級のポジションが存在する	14	100%		3	21%	4	29%	3	21%	4	29%	2	14%
データ活用を行うメンバーの管理者となるポジ ションが存在する	37	100%		8	22%	2	5%	5	14%	22	59%	1	3%
データ活用専門ポジションではないが、専門家 ポジションがある	36	100%		6	17%	5	14%	6	17%	19	53%	1	3%

注1 総数は受験級が不明の9名を含み、質問4で「所属なし」あるいは質問7で「全くない」の58名を除く。
注2 割合(%)は行和を100%として、行の度数10以上の場合に総数あるいは検定種類別の全体と15ポイント以上の差のあるものを**太字**表示にした。

専門家としてのキャリア意向 現在の業務において、自身の能力が十分に発揮されているか（自分の能力の発揮度）について、「少しはそう思う」40.4%が最も多い回答であった（表 3-12）。

表 3-12 自分の能力の発揮度，統計検定種類別人数

質問18 自分の能力の発揮度	総 数 ^(注1)		1級, 2級S/A, 専門統計調査士, 統計調査士		2級その他, 3級, 4級	
総 数	319	100.0%	77	100.0%	233	100.0%
全くそう思わない	25	7.8%	10	13.0%	15	6.4%
ほとんどそう思わない	52	16.3%	13	16.9%	38	16.3%
どちらとも言えない	86	27.0%	21	27.3%	62	26.6%
少しはそう思う	129	40.4%	27	35.1%	98	42.1%
かなりそう思う	26	8.2%	5	6.5%	20	8.6%
無回答	1	0.3%	1	1.3%	0	0.0%

注1 総数は受験級が不明の9名を含む。

したがって、データ活用の専門家としてキャリアを積みたいか（専門家としてのキャリア意向）について、「かなりそう思う」32.9%が最も多い回答になっていたと推測される（表 3-13）。

表 3-13 専門家としてのキャリア意向，統計検定種類別人数

質問19 専門家としてのキャリア意向	総 数 ^(注1)		1級, 2級S/A, 専門統計調査士, 統計調査士		2級その他, 3級, 4級	
総 数	319	100.0%	77	100.0%	233	100.0%
全くそう思わない	17	5.3%	2	2.6%	14	6.0%
少しはそう思う	77	24.1%	20	26.0%	56	24.0%
どちらともいえない	20	6.3%	2	2.6%	18	7.7%
そう思う	98	30.7%	21	27.3%	75	32.2%
かなりそう思う	105	32.9%	30	39.0%	70	30.0%
無回答	2	0.6%	2	2.6%	0	0.0%

注1 総数は受験級が不明の9名を含む。

3.3.2. 企業・機関への聞き取り調査

13 事業所に所属する 20 人のデータサイエンティストに対して、半構造化面接を行った。主に、どのようなスキルを持ち、どのような役割で仕事をし、どのようなキャリアパスを考えているか、について質問した。対象者の業種は金融、製造、流通、公共、IT ベンダー、コンサルティング会社など多岐にわたり、会社の規模も大企業、ベンチャー企業、からフリーランスの方までをカバーした。その結果わかったことは、以下の 5 項目に関わることであった。

- 1) データサイエンティストのバックグラウンド
- 2) データサイエンティストに求められる能力・スキル
- 3) 事業所内におけるデータサイエンティスト組織
- 4) 発注側のリテラシー
- 5) データサイエンティストの育成／キャリアパス

以下では、これらの点について詳述する。聞き取り調査結果一覧表は付録 B に掲載した。

データサイエンティストのバックグラウンド

データサイエンティストの学術的バックグラウンドは様々であった。確かに、データサイエンティストという名称から想像されるとおり、いわゆる理系の出身者は多かったが、理系分野の中で特定の分野に限定されることはなかった。

- ・大学学部では、生命系、実験系であった。大学院では機械学習を学び、システム生命学で学位を取得した（I 社 S 氏）。
- ・原子核理論物理学で学位を取得した後、総合エンジニアリング企業で就業した後、現在の会社に就職した（I 社 N 氏）。
- ・天文学で学位を取得した（K 機関 T 氏）。
- ・薬学系の学部を卒業したが、その後は分子生物学を専門としている（I 機関 F 氏）。
- ・学生時代は電子工学を学んでいた（Z 社 S 氏）。
- ・学生時代は応用物理学を学んでいた（T 機関 O 氏）。

- ・学生時代は地質学を学んでいた（T 機関 S 氏）。
- ・コンピュータグラフィックスの研究で学位を取得した（B 社 Y 氏）。
- ・学生時代は都市工学を学んでいた（M 社 O 氏）
- ・数学（作用素環論）で学位を取得した（M 氏）。
- ・大学では数理工学を学び、就職してからは製品開発に携わった（O 社 K 氏）。
- ・大学ではデータマイニングを学んでいた（P 社 H 氏）。

一方で、理系領域と社会科学系領域の融合分野や、社会科学系出身のデータサイエンティストも活躍していた。

- ・大学学部では経済学を学び、ソフトウェア工学で学位を取得した（T 機関 S 氏）。
- ・大学ではシステム創成学専攻に在籍し、技術経営が副専攻だった。（C 社 T 氏）
- ・大学学部では数理統計学を学び、大学院では公共政策を修めた（T 機関 T 氏）。
- ・大学学部では生命科学を学び、国内 MBA で階層ベイズモデルの研究をした（R 社 S 氏）。

データサイエンティストに求められる能力・スキル

データサイエンティストには、データを分析するだけではなく、ビジネス場面における意思決定に関わることができるだけの能力・スキルが求められていた。

- ・技術的な側面と人的な側面の両方が必要で、特に、顧客と接して折衝する能力は重要（I 社 S 氏）。
- ・現場で说得できるスキルが必要。伝えられない知識には意味がない（Z 社 S 氏）。
- ・分析自体ではなく、課題を見つけて、分析して、解決していく力が必要（T 機関 T 氏）。
- ・業種のエキスパティーズよりも、業務のエキスパティーズの方が重要（B 社 Y 氏）。
- ・スキルには3層あり、計算ができること、ビジネスと計算結果を繋ぐことができること、ビジネス情報課題や収益をあげる方法を検討できること（M 社 O 氏）。
- ・統計解析や分析の知識と情報系の技術が必要。その他にビジネスを理解することが必要。例えば、分析のプロセスを設計して、計画を立てられること、業務の段取り、データを取得し、どのように分析し、どのように伝えたらよいか理解すること。さらにビジネス側の業務改善を提案することもある（R 社 S 氏）。
- ・仕事の中の経験則が一番の成功機会。予想外のことが起こったときにそれがどうしてか原因を追究できること（M 氏）。

また、ビジネスに関わらない場面においては、統計解析のスキルにとどまらないデータの「深い」分析力が求められていた。

- ・大量のデータからゴールデンルールや外れ値をマイニングしたり、多次元空間で分析していくことになる（K 機関 T 氏）。
- ・統計の専門家のポストはないが、彼らがチームに加わると可能性が広がるだろう（J 機関 T 氏）。
- ・解析パッケージで仮定されている点について理論的な理解も必要。データそのものの意味を考えられる人、サンプリングの見極めが出来る人が重要。（T 機関 S 氏）
- ・調査の設計と集計の両方ができること。自発性、探求心があること。表面だけを見るのではなく、本質を見る必要がある（T 機関 M 氏）。

事業所内におけるデータサイエンティスト組織

例えば、CRISP-DM（SPSS、NCR、クライスラー、OHRA などにより確立されたデータ活用の標準プロセス）では、データサイエンティストのワークフローとして、1) ビジネス要件定義、2) データの理解、3) データの準備／加工、4) モデル作成、5) モデル評価、6) モデルの展開（改善方策の提案や効果検証など）が挙げられている。これらすべての局面を一個人が行うのではなく、チームとして（外部へのアウトソースも含む）1つのワークフローをこなしていることが多かった。このような事業所では、顧客の業務毎にワークフローを形成してデータ分析を行っていた。その理由は、業種が異なっても業務上の問題は似ているからと考えられる。

- ・エンジニアとマーケティングの橋渡しを担当している。（C 社 T 氏）
- ・グループでチームを組んでやっている。外部から SE が 2 人来ているが、彼らは情報処理や計算結果の表示、ルーティーンワークを行っている（I 機関 F 氏）。
- ・大きく 2 つの部署に分かれてデータ分析を担当している。部署 A では、スーパーバイズ、マーチャンダイジング、計算を各班に分かれて担当している。部署 B では、データを集める基幹システムを構築・整備している（Z 社 S 氏）。
- ・調査の設計をしている（N 氏）。製表・審査をしている（M 氏）。結果の取りまとめをしている（O 氏、S 氏）。結果表の集計が正しいか審査している（S 氏）。定義の見直しなどの規格の検討をしている（O 氏）。既に行った調査を基にサンプリングフレームを作成し、効率的な調査を行うための DB を企画・管理している（B 氏）。（以上 T 機関）
- ・「データサイエンティスト」と「モデラー」の 2 階層に分かれている。業種のエキスパティーズよりも、どちらかといえば業務のエキスパティーズのほうが重要（例：コールセンターの最適化については、製造も金融も問題が似ている）（B 社 Y 氏）。
- ・部署単独の案件と他部署との協業がある。協業の場合は、シンクタンク部門と連携している（M 社 O 氏）。
- ・事業ごとに担当窓口を設けて分析サービスを提供している。部署のトップには、マネジメントがいて、人員の管理を行っている（R 社 S 氏）。
- ・仕事は、チーム単位で分担して行う。（1 チーム 10 人くらい）統計部門ではデータ分析だけでなく、申請のドキュメンテーションや作業計画作成なども行う。データの正確性を担保するために、作業計画に基づいて二人以上のメンバーが同じことを行い、結果が一致するまで繰り返す。二人が違うプログラムを書くためは一発で一致することはまずない。何度も繰り返す。（M 氏前職）
- ・部署の担当分けは、ソリューション別になっている。定型化した分析と、高度な分析については外部にアウトソースしている（O 社 K 氏）。

一方で、社員全員が基本的なデータ分析スキルを持つように教育している会社もあった。

- ・分析の中身がわからない様なブラックボックスでやろうという考えは無理。高度な分析を行っても、現場のマネージャーが理解できなければ、行動してくれない。そのため、社員が背景を理解して説明できる必要がある（Z 社 S 氏）。

また、すべての局面を一個人が行う業務形態の事業所もあった。

- ・個人ですべてのデータサイエンティスト業務（ビジネス課題の抽出から分析まで）。工学系・数学系と医学系は話が合わないことが多いが、その間をつなぐ業務。（I 社 S 氏）
- ・企業からのマーケティングサイエンス的な分析業務の受託。市場調査やマーケ関連のデータ分析等。

- 顧客への提案、説明や、ビジネス上の課題抽出など。(I 社 N 氏)
 - 分析のアウトソーシングはしていない(予算の制約上、難しい。データの解析は個人にまかされている)。(K 機関 T 氏)
 - 予測モデル(融資モデル)を作成していた。仕事を行う上では、一応ワークグループがあるが、あまり関係ない。興味のあるプロジェクトに手を挙げて掛け持ちでやったりする。(M 氏前々職)
- フリーランスのデータサイエンティストの場合、クラウドソーシングを利用するので、事業所内での組織という概念を考えることが難しい。しかしながら、現状のクラウドソーシングの枠組みがあらゆる意味で有効と考えられていないようである。
- クラウドソーシングは、企業によって有効な手段であるが、日本ではあまり受け入れられないかもしれない。日本ではデータの管理が必要以上に厳しいこともあり、データをパブリックにすることはあまり好まれない。金融系などはまずデータを出さない。(M 氏)

発注者側のリテラシー

- 発注側、あるいはデータ分析結果を受け取って意思決定をする立場の人間がデータ分析結果を活用できなくては、どんなによいデータ分析であってもデータ分析の価値が損なわれてしまうだろう。また、高度なデータ分析に基づいた意思決定が活かされていないと推測される事例もあった (C 社 T 氏)。さらに、組織によっては、特定の専門家を育てるのではなく、組織全体のデータリテラシーを上げることが求めていると考えられる。したがって、発注者側のリテラシー向上は重要 (T 機関 T 氏) と考えられる。
- トップがデータ分析を理解しているかどうかが大事。会社全体の意識が違う。(Z 社 S 氏)
 - クライアントのリテラシーが最大のボトルネック。難しいことはわからない。分析の価値を認識しないと投資もしない。よくわからない理屈もある。クライアント側では分析ができると出世できない。トップダウンが必要であるが、現場との意識の差は大きい。データサイエンティストを育てるだけではダメ、企業側が客観的事実に基づいて意思決定を行う文化の育成が必要。声の大きい人が正しいのではなく、試行錯誤を認め、すぐに成果を求めない文化が必要。経営者がデータを理解しないことによるデメリットがでてくるのではないか。そうすれば経営者も気づくのではないか。(M 社 O 氏)
- また、データ分析を受注する際に、問題がはっきり定義されていないという回答も得られた。そのような場合、データサイエンティストは発注者側の問題設定から関わっている。
- 問題を明確に定義していないお客様(例:「このデータから何か見つけて」というお客様)に対しては、問題設定ワークショップを行う。この際にも、データ分析手法の知識は必要。なぜならば、「問題設定+手法」は常にペアで考えなければならないからである。(B 社 Y 氏)
 - 主に課題は、ビジネス的な悩みとしてあがってくることが多く、それを現場と一緒にになって具体的な課題に落とし込んでいく。(課題を“ひらく”と表現)。アイデアは現場が持っていることが多いのでそれを引き出していく。(R 社 S 氏)
 - 実績を積むと依頼の来るプライベートのコンペは、モデル構築の前段階から始まり、通常のコンペとは異なる。通常のコンペは、インプットも、アウトプットも決まっており、モデルを作るところが中心。プライベートでは、もっとあいまいな状況が多く、データベースをまるまる渡されて考えてほしいとか、データそのものを取得するところから始まることもある。(M 氏)
 - 分析課題は、ビジネス上の課題をもとに、現場と一緒に考えるケースが中心。「こういった分析を

- してほしい」というように、現場が分析課題を持ってくるケースは筋が悪い。「こういったことで悩んでいる」というのは、筋が良い。意思決定者は、分析課題にしてはダメ。意思決定の問題として考えるべき。(O 社 K 氏)
- データサイエンティストはクライアント側(企業側)には置くべきと考えている。その理由の第1は、企業固有のデータの意味を理解している人材が必要。例えば、機械学習を行うにしても、何をインプットとするのかその勘所が重要で、外部のコンサルタントではそこは理解できない。また、多重共線性とかの問題が発生した際に外部の人材では、数値の意味合いが理解できず判断できない可能性がある。第2に、企業内の利害調整の問題が外部人材では解決できないからである。第3に、データ分析の場合は、やってみないとわからないケースがあり、外部への発注が難しい。(P 社 H 氏)

キャリアパスについて

- 現状では、データサイエンティストには明確なキャリアパスがない。データ分析を担当していても、データサイエンティストとして昇進等ができない、あるいは担当部署の設置から今日までの日が浅く今後の見通しが未定といった回答が得られた。中には、継続的な雇用自体が保証できないという事業所もあった。
- データサイエンティストのキャリアパスは存在せず、基本 IT のパスしかない (I 社 S 氏前職)。
 - コンサルタントのキャリアでは、データ分析の専門家で進めるキャリアパスはない。キャリアを積むと営業・マネジメントになる。マネジメントになると基本営業であり売上責任が発生する。データ分析を進めたい人は別の会社に移ってしまうことも多い。(C 社 T 氏前職)
 - 報酬面で(データサイエンティスト的な)人材をとどめておくことは難しい。統計学の専門家はいない(データの解析は個人にまかされている)。(K 機関 T 氏)
 - 統計的な手法でモデルを考える人は必要だが、その後のキャリアパスは不明なので、ある分野をたててリーダーとしてやっていけるか(引っ張っていけるか) どうかが必要。マネジメント能力が重要。(J 機関 T 氏)
 - 最優先するのは現場の知識・経験。基本的に、データ分析専門の社員の採用は行っていない。(Z 社 S 氏)
 - 2007 年よりデータ分析を行う部署として立ち上がった。(M 社 O 氏)
 - 次のポジションがいまのところ明確でないのが問題。今現在は、マネージャー職としての権限を持つ、スーパーバイザー。会社でもこれまでなかった組織なので、ポジションもこれから考えることになる。(R 社 S 氏)
 - 上に行くと営業の仕事が多くなる(前々職)。上にいくと管理職(統計家としてのキャリアパスはない)(前職)。(M 氏)
 - キャリアパスとしては、社内異動が基本。適正がない→他部署に。能力はあるが、本人が実際にビジネスをやりたい→他部署に。能力もあり、本人も分析続けたい→残る。これまでなかった部署なので、キャリアパスとしては、今のポジションまで。その先は未定。この先、自分がどういうポジションになっていくは、今後の組織運営上重要。(O 社 K 氏)
 - 今後についてはどちらかという民間でやっていくつもりだが、このまま定年までというのはあまり想像できない。(P 社 H 氏)
- このような傾向を必ずしも反映しているとは限らないが、データサイエンティストの育成は何らかの教育

プログラムによらず、個々人の自己学習に委ねられていることが多い。

- ・自分で課題を見つけて技術を向上させる必要がある。(I 社 S 氏、N 氏)
- ・基本は自分で勉強していく方式(スーパーバイザーがついているのでその下で勉強)(K 機関 T 氏)
- ・グループに任せている。(J 機関 T 氏)
- ・基本的には OJT。その他に必要な知識は自主ゼミや自己学習による。(T 機関)
- ・スキルについては四則演算なので教育ではない。2 週間 Excel・Access くらいは教える。(全員 Access を使うことが出来る) 手法を教えることより、現場感を大事にしている。(Z 社 S 氏)
- ・基本的には独学(技術的に難しいことは業務上の機会がない)。コンサルティングスキルについては、仕事を通してでないと身につかない。(M 氏前々職)
- ・割と研究があり、業務に必要なことは教えてくれた。ただ、一番時間がかかるのはエラーハンドリングであり、その分野は教えられるのではない。(M 氏前職)
- ・最近の若い人は積極的に勉強しない。「面白そうなセミナーの連絡が来たので参加していいか?」という姿勢ではだめ。自分のやりたい領域、知りたいことについて積極さが必要。GDP の予想などをコンサルに頼むが、その根拠の説明を求めても最後よくわからなかった。自分で知りたいので、大学に入学した。(O 社 K 氏)
- ・基本は自分で勉強する。人が多くなってくると本を輪読したり、自分がやったことを発表したりしている。レクチャー形式で、できる人が教えていく。教育は「重要度は高いが緊急性が低い」仕事。学術系から本を選んでいる(国際会議の論文とか、論文は書いてはいないが読んではいる)。機械学習、分散処理にフォーカスしていて、そこから広げていっている。(P 社 H 氏)

一方で、事業所内でデータサイエンティストとしての一定のキャリアパスがあっても、育成に関してはデータサイエンティストに特化していないという回答も得られている。

- ・育成について、社内におけるスキル定義の中には、特にデータサイエンティストに特化したものは無い。通常のコンサルのスキルセットの中で、昇進が決まっていく。若い人を育てたい。これは、常にシニアメンバーとペアを組んでプロジェクトに当たらせることで育成する。なぜなら、「プロジェクトを通してしか人は育たない」から。(特に、これはコンサルタントのスキルについて語っているものと思われる)(B 社 Y 氏)
- ・ビジネスに近いところで仕事をさせることで、ビジネスサイドもわかる人材に育てようと考えている。社内にいるだけでは、集計家になってしまう。分析の意味をビジネスに近いところで感じる必要がある。クライアントが求めているのは、「右か左か」だけ。だから、高度な数学的素養を持つエンジニアは自分の存在意義を見出せない。誰もがビジネスサイドも分かるようになるわけではない。もっと高度な分析をしたいと思っているメンバーもいる。こういったメンバーには、高度なコミュニケーションは無理でも、業務知識をつけて 1 言えば、10 わかる人材に育てる。業務知識を知ることで、ビジネス上のボトルネックが分かるようになる。例えば、分析の数値から出た施策が、チャネルの限界でできないということもある。出向させた先では、分析をさせるのではなく、マネジメントをさせるようにしている。(M 社 O 氏)
- ・必要なスキル 3 つ(①統計解析等、分析の知識、②情報系の技術、③その他ビジネス的・PM 的スキル)の三角形をより大きくしていくイメージ。(R 社 S 氏)

現場では、事業所内での育成や教育機関での教育に対する懸念があることも分かった。

- ・データを活用しようとしている会社は沢山あるのでデータサイエンティストを育てる価値はあると

思う。WEB 系の企業はどのようなサービスでも育てる価値はあるが、そういった教育は日本ではあまりない。(M 氏)

- ・アカデミックの人の意識はややずれている。ある学会で「大学も企業の役に立つ研究を」そのためには、「企業の人にもっと頑張ってもらわなければ」という発想が主流だった。もっと企業側に立った発想が必要と考えられる。教授がビジネスを始めるなどの取り組みを進めるべきでは。(O 社 K 氏)

雇用に関しては、業務の専門性が非常に高く、統計学、あるいは情報学系のスキルよりも、事業所の業務に精通している人材を雇用するという回答もあった。

- ・可能な限り、事業に必要な専門領域に精通した人材を雇用。(I 機関 F 氏)

それとは逆に、データ分析やモデリングの部分はソフトウェア等でカバーできる部分があるので、個々人の専門性を重視する事業所もあった。

- ・文系も雇ってみたい。なぜなら、モデラーの部分は「道具がカバーしてくれる」から。理系の中に、文系を「まぶしていく」イメージ。(B 社 Y 氏)
- ・大卒(インターンから)からもいるが、中途の人が多い。だが、今は半々くらい。中途はビジネス寄りの人が中心。新卒は各分野で一流の人材を採用する。(P 社 H 氏)

今回の聞き取り調査では、フリーランスのデータサイエンティストの場合、今後もフリーランスを続けていくかについて否定的な回答(M 氏)が得られた。また、日本の事業所は、データサイエンティストを雇用する意思があるのか分からない(M 氏)という指摘もあった。例えば、データサイエンティストには高度なスキルが要求されるはずなのに、求人条件の中で「未経験者大歓迎」という記述をしている事業所は求職者に何を求めているのか不明と受け取ることもできる。

3.4. 教材の開発

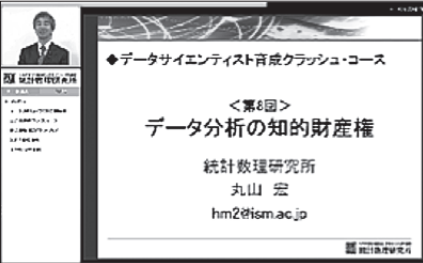
1. データサイエンティスト育成クラッシュコース

本事業では、インターンシップ学生向けに、データ分析の基礎を短時間で学ぶためのオンライン教材を開発した。構成は以下の通り。

- C0 コース概要 丸山宏(統計数理研究所 教授)
- C1 データサイエンティストとは:佐藤洋行(株式会社ブレインパッド)
- C2 データ解析基礎:馬場康雄(統計数理研究所 特命教授)
- C3 データ可視化とツール:中野純司(統計数理研究所 教授)
- C4 統計モデリングと機械学習:松井知子(統計数理研究所 教授)
- C5 統計的時系列モデリング:川崎能典(統計数理研究所 准教授)
- C6 最適化:伊藤聡(統計数理研究所 教授)
- C7 データ分析と意思決定:椿広計(統計数理研究所 教授)
- C8 データ分析の知的財産:丸山宏

2. 配信方法

MOOC として、クローズドなウェブ配信(インターンシップ学生のみ閲覧可能)を検討中である。また、今後の展開として DVD メディアでの配布をあわせて検討中である。



3.5. 海外連携・標準化

平成 26 年 3 月 6 日～7 日に、実施責任者の丸山宏がシリコンバレーを訪問し、Insight Data Science Fellows Program を始めとする数社で、米国西海岸におけるデータサイエンティストの育成と活用の状況に関してヒアリングを行った。訪問したのは、下記 4 社である。

1. Insight Data Science Fellows Program

日時： 3/6 10:30-11:30
訪問場所： Insight Data Science Fellows Program 本社 154 E Dana St, in Mountain View, CA
対応者： Dr. Kathy Copic 他 1 名

2. IBM Corporation

日時： 3/6 12:30-14:00
訪問場所： IBM Almaden Research Center 650 Harry Road, San Jose, CA
対応者： Dr. Jim Spohrer
Yassi Moghaddam (IISIP (The International Society of ServiceInnovation Professionals) Executive Director)

3. Facebook Inc.

日時： 3/7 14:00 - 15:00
訪問場所： Facebook Inc.本社 Willow Road, Menlo Park, CA
対応者： Ariel Evnine, Dr. Mike Develin

4. LinkedIn

日時： 3/7 16:00 - 17:00
訪問場所： LinkedIn 本社 Stierlin Court, Mountain View, CA
対応者： Andrew Clark

3.6. その他

1. 大学との連携

・「統計教育大学間連携ネットワーク」との連携
連携大学：東京大学，大阪大学，総合研究大学院大学，青山学院大学（代表校），多摩大学，立教大学，早稲田大学，同志社大学

・筑波大学・九州大学とのネットワーク強化、旧帝大ネットワーク(仮称)の構築

2. 一般社団法人データサイエンティスト協会、公益社団法人 企業情報化協会等との連携

3. 研究者一覧

研究代表者 樋口 知之（統計数理研究所 所長）
実施責任者 丸山 宏（統計数理研究所 副所長）
竹村 彰通（東京大学情報理工学系研究科 教授）※再委託先
研究協力者 清水 信夫（統計数理研究所 助教）
藤田 茂（統計数理研究所 特任研究員）
神谷 直樹（統計数理研究所 特任研究員）※2014.2.16～
井川 甲作（統計数理研究所 特別共同利用研究員／東京工業大学大学院）
藤原 海世（統計数理研究所 外来研究員／東京工業大学大学院）

4. 運営委員会 委員名簿

伊藤 聡	統計数理研究所 教授
井上 亮	株式会社アカリク 採用コンサルティング事業部
岡野原 大輔	株式会社 Preferred Infrastructure
柏木 宣久	統計数理研究所 教授
北村 浩三	日本アイ・ビー・エム株式会社 IBM 大学連携プログラム（当時）
木村 正一	統計検定センター
草野 隆史	株式会社ブレインパッド 代表取締役
佐藤 俊哉	京都大学医学研究科 社会健康医学系専攻 教授
鈴木 律夫	情報サービス産業協会 企画調査部
高田 唯史	自然科学研究機構 国立天文台 天文データセンター 准教授
竹村 彰通	東京大学情報理工学系研究科 教授
田村 義保	統計数理研究所 教授
坪井 誠司	独立行政法人海洋研究開発機構 地球情報センター 部長
中野 純司	統計数理研究所 教授
橋本 武彦	一般財団法人データサイエンティスト協会 事務局長補佐
藤山 秋佐夫	国立遺伝研究所 教授
堀 宗朗	東京大学地震研究所 教授
三國 純一	日本経済新聞社 クロスメディア営業局第一部

（敬称略・50 音順）
※運営委員会設置は平成 25 年度までとし、平成 26 年度以降は「メンバーシップ制度（仮称）」を設置し、より広くネットワークを構築する。

引用・参考文献一覧

[1] Bureau of Labor Statistics. "Occupational Outlook Handbook". Bureau of Labor Statistics. <http://www.bls.gov/ooh/Math/Statisticians.htm#tab-1>, (accessed 2013-03-11).

[2] Davenport, T.H.; Patil, D.J. Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review. October, 2012, p. 70-76. 日本版 データサイエンティストほど素敵な仕事はない. DIAMOND ハーバード・ビジネス・レビュー. 2013, 2月号, p. 84-95.

[3] Evans, J.R. Business Analytics: The Next Frontier for Decision Sciences. Decision Science. 2012, 43(2).

[4] 藤田 茂. データサイエンティスト育成ネットワークの形成. サイエнтиフィック・システム研究会, 2013.

[5] 濱中淳子. ポスドク就職難民問題～解決のための処方箋は何か～ . 大学と学生. 2008, no. 56, p. 14-20. http://www.jasso.go.jp/gakusei_plan/documents/daigaku530_06.pdf, (accessed 2013-03-11).

[6] Harris, H.; Sean M.; Marck V. Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work. O'Reilly, 2013, 32p.

[7] 樋口知之. ビッグデータと個人化技術. 統計. 2012, 9月号, p. 2-9.

[8] 樋口知之. ビッグデータが迫る研究開発の変革. RIETI BBL セミナー. 2012. <http://www.rieti.go.jp/jp/events/bbl/12051401.html>, (accessed 2013-03-11).

[9] 樋口知之. ビッグデータがせまる人への投資とサービスの高度化. ビックデータ&オープンデータ研究会. 2013.

[10] 樋口知之. ビジネスにおけるビッグデータと統計学～受益者起点と予測モデル～. 日経産業新聞フォーラム 2013. 2013.12.3

[11] 樋口知之. データ解析の真髄とは. DIAMOND ハーバード・ビジネス・レビュー. 2013, 2月号, p. 98-108.

[12] Higuchi, T. Panel Discussion "Key Issues in Big Data". IEEE International Conference on Big Data 2013. <http://cci.drexel.edu/bigdata/bigdata2013/programschedule.htm>, (accessed 2013-03-11).

[13] 井川甲作, 丸山 宏, 藤田 茂, 藤原海世. データサイエンティストの現状と課題 ベストプラクティス調査, 横幹連合コンファレンス, 2013.

[14] Insight Data Science Fellows Program White Paper. http://insightdatascience.com/Insight_White_Paper_2014.pdf, (accessed 2013-03-11).

[15] 河本 薫. 会社を変える分析の力. 講談社, 2013, 240p.

[16] 河本 薫. データ分析力と現場力を融合ー新たなイノベーションを起こす. 行政&情報システム. 2014, 2月号, p. 13-18.

[17] Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Byers, A.H. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 2011, 156p. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation, (accessed 2013-03-11).

[18] 丸山 宏. データに基づく意思決定ービッグデータブームの裏の本質. MATLAB Expo 2012 | JAPAN.

2012. <http://matlabexpo.com/jp/session/big-data.html>, (accessed 2013-03-11).

[19] 丸山 宏. ビッグデータ・ブームの本質. DIAMOND ハーバード・ビジネス・レビュー. 2013. <http://www.dhbr.net/category/d-bigdataboom>, (accessed 2013-03-11).

[20] 丸山 宏. データ分析でできること、できないこと. THINK. 2013, No. 46, p. 26-33.

[21] 丸山 宏. ビッグデータの本質とは～データを活用する社会の実現に向けた課題と可能性～. NTT Data DIGITAL GOVERNMENT & FINANCIAL TOPICS, 2013. http://e-public.nttdata.co.jp/topics_detail2/contents_type=2&id=968, (accessed 2013-03-11).

[22] 丸山 宏. データに基づく意思決定ービッグデータブームの裏にあるものー. 芝浦工業大学 MOT シンポジウム. 2013.

[23] 丸山 宏. データサイエンティスト育成ネットワークの形成. 第 3 回データサイエンスサミット. 2013.

[24] 丸山 宏. データ分析と意思決定. 行政&情報システム. 2014, 2月号, p. 7-12.

[25] 丸山 宏. データサイエンティスト育成ネットワークの形成. 次世代リーディングシンポジウム. 2014.

[26] 丸山 宏. データサイエンティスト育成ネットワークの形成 ～我が国におけるデータ分析人材の現状と展望～. ビッグデータに関する研究開発シンポジウム ～ビッグデータによる新産業・イノベーションの創出に向けて～（総務省・文部科学省・経済産業省 3省合同シンポジウム）. 2014.

[27] 丸山 宏, 樋口知之, 竹村彰通. 「データサイエンティスト育成ネットワークの形成」事業の概要, 横幹連合コンファレンス, 2013.

[28] 文部科学省委託事業 データサイエンティスト育成ネットワークの形成. <http://datascientist.ism.ac.jp/>, (accessed 2013-03-11).

[29] 日本経済新聞. ビッグデータ分析に人材の壁、25 万人不足見通しー I T 各社、育成へ本腰（2013/7/17 1:54）. 2013. <http://www.nikkei.com/article/DGXNZO57421630X10C13A7EA1000/>, (accessed 2013-03-11).

[30] 佐藤忠彦, 樋口知之. ビッグデータ時代のマーケティングーベイジアンモデリングの活用. 講談社, 2013, 215p.

[31] 統計検定センター. 統計検定合格者を対象とした調査報告書. 2014, 42p. http://datascientist.ism.ac.jp/H25_report/JSSC.pdf （2014 年 4 月公開予定）

[32] Sondergaard, P. Gartner Symposium/ITxpo 2012. <http://gartner-em.jp/symposium2012/index.html?o=>.

[33] 財務省. 法人企業統計調査結果（平成 2 5 年 1 ～ 3 月期）. <https://www.mof.go.jp/pri/reference/ssc/results/h25.1-3.pdf>, (accessed 2013-06-04).

付録 A：依頼状と調査票

社会人合格者の皆様へ アンケートのお願い

2013 年 12 月 17 日

一般財団法人 統計質保証推進協会 統計検定センター

社会人合格者の皆様へ

この度は、統計検定に合格されおめでとうございます。

統計検定センターでは、情報・システム研究機構 統計数理研究所の**文部科学省委託事業「データサイエンティスト育成ネットワークの形成」**事業の委託を受け、合格者の皆様にアンケートのご協力をお願いしております。ご回答いただきました内容は、適切な統計処理と分析を行い、「データサイエンティストのベストプラクティス調査」や教材開発等に活用することになっております。ご協力の程よろしくお願いいたします。

なお、このアンケートにはウェブサイトからも回答できます。ウェブサイトから回答いただける方は、統計検定センターホームページにあります「Web 合格発表」>「社会人合格者の皆様へ アンケートのお願い」よりお入りいただくか、下記のサイトをアクセスしてください。

http://goo.gl/5Kbwdi

本状と行き違いにご回答いただいております場合は、失礼のほどお許しくださいますようお願い申し上げます。

【対象者】

統計検定に合格された社会人の方をお願いしております。

【個人情報の使用目的について】

ご記入された受験番号を含む個人情報等は上記の目的にのみ使用します。また、データそのものは皆様の同意なく、委託者である統計数理研究所も含む第三者に提供いたしません。皆様の個人情報を保護します。

【実施機関】

一般財団法人 統計質保証推進協会 統計検定センター
情報・システム研究機構 統計数理研究所

統計検定合格者（社会人）アンケート

一般財団法人 統計質保証推進協会 統計検定センター

質問 1 【回答必須】 受験番号とイニシャルを記入してください。

<例> 受験番号：123456 氏名：統計あかり の場合 123456 TA

2 種類の統計検定に合格された方は、上位の試験または調査士の受験番号を記入してください。

質問 2 【回答必須】 年齢を選択してください。

- ☐16～19 歳
- ☐20～29 歳
- ☐30～39 歳
- ☐40～49 歳
- ☐50～59 歳
- ☐60～69 歳
- ☐70 歳以上

質問 3 【回答必須】 性別を選択してください。

- ☐男性
- ☐女性
- ☐回答しない

質問 4 【回答必須】 所属している組織について 1 つだけ選択してください。

※所属なしの方は、質問 12 まで飛んでください。

- ☐製造業
- ☐出版・印刷業
- ☐金融・保険業
- ☐卸売・小売業
- ☐卸売業（商社）
- ☐I T 関連業
- ☐サービス業
- ☐公務員・官公庁
- ☐医療・福祉関連業
- ☐建設・不動産業
- ☐教育・保育・学校法人
- ☐マスコミ・メディア関連業
- ☐フリーランス
- ☐その他の組織
- ☐所属なし

質問 5 【所属のある方】 所属している組織での職種について、1 つだけ選択してください。

- ☐管理
- ☐研究・開発
- ☐調査
- ☐企画
- ☐営業
- ☐製造
- ☐総務
- ☐人事
- ☐経理
- ☐コンサルティング
- ☐I T 関係
- ☐フリーランス
- ☐その他の職種

質問 6 【所属のある方】 組織の従業員数について、1 つ選択ください。

- ☐～10 名
- ☐11～50 名
- ☐51～100 名
- ☐101～500 名
- ☐501～1000 名
- ☐1001～5000 名
- ☐5001 名～

質問 16【回答必須】ビジネス系スキルについて、実際の業務で利用可能と思うものを**全て**選択してください。

☐マネジメント

☐営業

☐IT システム設計・構築・運用

☐ファイナンス

☐生産・製造管理

☐コンサルティング

☐マーケティング・企画

☐研究・開発

☐その他

質問 17【回答必須】上記で回答いただいたビジネス系スキルの学習方法を**全て**選択してください。

☐大学・大学院（MBA 含む）

☐外部スクール

☐仕事上での研修

☐仕事上での OJT

☐独学（WEB 教材）

☐独学（書籍等）

☐その他

☐どのスキルも持っていない

質問 18【回答必須】現在の業務において、ご自身の能力が十分に発揮されていると感じていますか。

☐全くそう思わない

☐ほとんどそう思わない

☐どちらとも言えない

☐少しはそう思う

☐かなりそう思う

質問 19【回答必須】データ活用の専門家としてキャリアを積みたいと思いますか？、
※「全くそう思わない」を選択されましたら、こちらでアンケート終了です。
ご協力ありがとうございました。

☐全くそう思わない

☐少しはそう思う

☐どちらともいえない

☐そう思う

☐かなりそう思う

質問 20【回答必須】質問 19 で、「全くそう思わない」を**選択しなかった方**へ質問です。
その理由で該当するもの**全て**を選んでください。

☐報酬が高そうだから

☐データ分析が好きだから

☐ポジションが多そうだから

☐安定した職業であるから

☐イメージがいいから

☐年齢・性別に問わず働けそうだから

☐勤務地関係なく働けそうだから

☐自分のスキルを活かせそうだから

☐出世できそうだから

☐社会にとって、組織にとって重要な仕事だと思うから

☐その他

理由を記述してください

ご協力ありがとうございました。

#お手数をお掛けしますが、同封の返信用封筒に封入してポストに投函してください。

付録 B：聞き取り調査結果一覧表

業種	事業所規模	育成	採用／キャリアパス	業務	バックグラウンド	求められる能力	発注者側のリテラシー
I社S氏			【SE前職】 データサイエンティストのキャリアパスは存在せず、基本ITのパスしかない。	個人ですべてのデータサイエンティスト業務（ビジネス課題の抽出から分析まで）。工学系・数力学と医学系は語が合わないことが多いが、その間をつなぐ業務。	大学学部では、生命系、実験系であった。大学院では機械学習を学び、システム生物学で学位を取得した。企業のソフトウェア部門で就業した後、現在の会社に就職した。	技術的な側面と人的な側面の両方が必要で、特に、顧客と接して折衝する能力は重要。	
I社N氏			【現職】 作業系は、個別技術が重要。コンサル系は、個人の資質（業務経験の有無、幅広い興味を持つ・持っているか、提案力、顧客開拓・獲得力）に依存する。	企業からのマーケティングサイエンス的な分析業務の受託。市場調査やマーケティングデータの分析等。顧客への提案、説明や、ビジネス上の課題抽出など。	原子核理論物理学で学位を取得し、総合エン지니어リング企業で就業した後、現在の会社で就業した。		
C社T氏		ベンチャー企業	IT関連	【前職】 コンサルタントのキャリアでは、データ分析の専門家とで進めるキャリアパスはない。キャリアを積むと営業・マネジメントになる。マネジメントになると基本営業であり売上責任が発生する。データ分析を進めた人は別の会社に移ってしまふことも多い。	エンジニアとマーケティングの橋渡しをしている。エンジニアとマーケティングの橋渡しをしている。	大学ではシステム創成専攻に在籍し、技術経営が副攻だった。	意思決定をしているのは社長。データがなくとも意思決定されるはず。
K機関T氏	研究	—	基本は自分で勉強していく方式（スーパーバイザーがついているのでその下で勉強）	分析のアウトソーシングはしていない（予算の制約上、難しい。データの解析は個人にまかされている）。	大量のデータからゴールデングルールや外れ値をマイニングしたり、多次元空間で分析していくことになる。		
J機関T氏	研究	—	グループに任せている。	モデルを仮定して計算して実際のデータと比較している。あるモデルに対して計算したカーネル（どここの構造に対して敏感なのかを表す関数）を使って観測データと計算データとの不一致を見つけている。	統計の専門家のポストはないが、彼らがチームに加わると可能性が広がるだろう。		
I機関F氏	研究	—	可能な限り、事業に必要な専門領域に精通した人材を雇用。	グループでチームを組んでやっている。外部からSEが2人来ているが、彼らは情報処理や計算結果の表示、ルーティーンワークを行っている。	薬学系の学部を卒業したが、その後は分子生物学を専門としている。		

データサイエンティスト育成ネットワークの形成 報告書 | 36

業種	事業所規模	育成	採用／キャリアパス	業務	バックグラウンド	求められる能力	発注者側のリテラシー
Z社SE	流通、製造	大企業	最優先するのは現場の知識・経験、基本的に、データ分析専門の社員の採用は行っていない。	大きく2つの部署に分かれてデータ分析を担当している。部署Aでは、スーパーバイズ、マーチャングダイジング、計算を各班に分かれて担当している。部署Bでは、データを集める基幹システムを構築・整備している。	学生時代は電子工学を学んでいた。	現場で読得できるスキルが必要。伝えられない知識は意味がない。	トップがデータ分析を理解しているかどうかが大事。会社全体の意識が違う。
T機関S氏			情報処理、データベース担当から、分析系の部署に移った。	調査の設計をしている（N氏）。製表・審査をしている（M氏）。結果の取りまとめをしている（O氏、S氏）。結果表の集計が正しいか審査している（ST氏）。定義の見直しなどの規格の検討をしている（O氏）。既にを行った調査を基にサンプリングフレームを作成し、効率的な調査を行うためのDBを企画・管理している（BK氏）。	大学学部では経済学を学び、ソフトウェア工学で学位を取得した。	解析パッケージで仮定されている点について理論的な理解も必要。データそのものの意味を考えられる人、サンプリングの見極めが出来る人が重要。	
T機関B氏			統理的な素養を必ずしも必要とされるわけではない。		大学学部では数理統計学を学び、大学院では公共政策を修めた。	分析自体ではなく、課題を見つけて、分析して、解決していく力が必要。	
T機関O氏			基本的にはOJT。その他に必要な知識は自主ゼミや自己学習による。		学生時代は応用物理学を学んでいた。		
T機関M氏		公共	昭和の時は調査票の集計を行っていた。			調査の設計と集計の両方ができること。自発性、探究心があること。表面だけを見るのではなく、本質を見る必要がある。	
T機関N氏			情報処理、プログラムがスタート。		学生時代は地質学を学んでいた。		
T機関S氏							
T機関T氏							発注者側のリテラシー向上は重要。
B社Y氏	IT関連	大企業	文系も雇ってみたい。なぜなら、モデラーの部分は「道具がカバールしてくれる」から。理系の中に、「文系を「まぶしていく」イメージ。	「データサイエンティスト」と「モデラー」の2階層に分かれている。業種のエキスパティエーズよりも、どちらかといえば業務のエキススパティエーズのほうが重要（例：コールセンターの最適化については、製造も金融も問題が似ている）。	コンピュータグラフィックスの研究で博士号取得した。5年前にコンサルティング部門へ転属。	問題を明確に定義していないお客様（例：「このデータから何か見つけて」というお客様）に対しては、問題設定ワークショップを行う。この際にも、データ分析手法の知識は必要。なぜならば、「問題設定＋手法」は常にベアで考えなければならぬからである。	

業種	事業所規模	育成	採用／キャリアパス	業務	バックグラウンド	求められる能力	発注者側のリテラシー
M社O氏	コンサル ティング	大企業	2007年よりデータ分析を行う部署として立ち上がった。	部署単独の案件と他部署との協業がある。協業の場合は、シンクタンク部門と連携している。	学生時代は都市工学を学んでいた。	スキルには3層あり、計算ができること、ビジネスと計算結果を繋ぐことができ、ビジネス情報課題や収益をあげる方法を検討できること。	クライアントのリテラシーが最大のボトルネック。難しいことはわからない。分析の価値を認識しないと投資もしない。よくわからない理屈もある。クライアント側では分析ができると出世できない。トップダウン世ではない。トップダウンが必要であるが、現場との意識の差は大きい。データサイエンティストを育てるだけではダメ、企業側が客観的事実に基づいて意思決定を行う文化の育成が必要。声の大きい人が正しいのではなく、誤行錯誤を認め、すぐに成果を求めない文化が必要。経営者がデータを理解しないことによるデメリットがでてくるのではないか。そうすれば経営者も気づくのではないか。
R社SE	コンサル ティング	大企業	次のボジションがいまのところで明確でないのが問題。今現在、マネージャー職としての権限を持つ、スーパーバイザー。会社でもこれまででなかった組織なので、ボジションもこれから考えることになる。	事業ごとに担当窓口を設けて分析サービスを提供している。部署のトップにはマネジメントがいて、人員の管理を行っている。	大学学部では生命科学を学び、国内MBAで階層ベイズモデルの研究をした。	統計解析や分析の知識と情報系の技術が必要。その他にビジネスを理解することが必要。例えば、分析のプロセスを設計して、計画を立てられること、業務の段取り、データを取得し、どのように分析し、どのような伝えたらよいか理解すること。さらにビジネス側の業務改善を提案することもある。	主に課題は、ビジネス的な悩みとしてあがってくることで多く、それを現場と一緒に becoming 具体的な課題に落として込んでいく。（課題を「ひらく」と表現）。アイデアは現場が持っていることが多いのでそれを引き出す。



文部科学省委託事業

**Data
Scientist
Training
Network**

データサイエンティスト育成ネットワークの形成

文部科学省委託事業
「データサイエンティスト育成ネットワークの形成」
平成 25 年度事業報告書

発行日／平成 26 年 3 月

情報・システム研究機構

統計数理研究所 データサイエンティスト育成ネットワークの形成

〒190-8562 東京都立川市緑町 10-3 Tel : 050-5533-8500(代表)

e-mail: info@datascientist.ism.ac.jp