

# はじめに

本書は、2006年に出版された『アカデミック・スキルズ—大学生のための知的技法入門』の姉妹編にあたります。同じく、慶應義塾大学出版会が刊行予定である『アカデミック・スキルズ—実地調査法入門』（長田進経済学部教授及び本書執筆の西山敏樹による）とも相関が深い書籍です。慶應義塾大学日吉キャンパスでは、自分で問題を発見し研究テーマを定め、データの収集・分析を経て研究テーマの動向を明らかにし、問題解決策を明示する一連の研究技法を「アカデミック・スキルズ」と命名して、少人数制の講義を開講しています。すでに、本シリーズの先輩の書として、『グループ学習入門—学びあう場づくりの技法』等があり、目下、「アカデミック・スキルズ」の少人数制の講義は一定の効果をあげて、そのノウハウを書籍に昇華させ社会に波及させるステージに突入しました。

筆者は、今まで社会調査法（データ収集法）や収集したデータの分析法について、慶應義塾大学教養研究センター設置科目「アカデミック・スキルズ」の講義を担当していました。また、慶應義塾大学の大学院である政策・メディア研究科や、現在の本務である大学院システムデザイン・マネジメント研究科でも、社会調査法や統計分析手法の講義や大学院生指導を行ってきました。そこで本書では、筆者の社会調査法やデータの分析を伴った学術論文の執筆、公的プロジェクト活動の経験も交えつつ、世の中に多く出回る当該分野の「理論書」ではなく「豊かな経験とノウハウ、つまり社会調査法（データ収集法）及びデータ分析法の真に活きたアカデミック・スキルズを後世に広く伝承する書」を目指して、本書をまとめました。

近年では、大学・大学院で社会調査やデータ分析を伴う実証的研究を行う学部生、大学院生が大変増えてきました。そこで問題なのは、データを集めてそれらを分析し、考察をまとめて何らかの提案を行えば学部卒業、さらには、大学院修了ができると思っている人々が、大変多いと

ということです。

データとその収集、分析の一連のプロセスは、社会的に見ても重要なことです。日常的に、データを見て、色々な人が意思決定し、何らかの行動をすることが大変多いからです。

データとは、人や組織の行動を決定づける大変重要な性格を持つものなのです。だからこそ、いい加減な姿勢でデータ分析を行うことはご法度です。人や組織が進むべき方向性を正しいものにするためにも、正しいデータ分析とはいかなるべきなのか、学問的お作法すなわち「アカデミック・スキルズ」として、社会調査（データ収集）やデータ分析の方法論を今、着実に身につけておきたいものです。

そこで本書では、社会調査（データ収集）やデータ分析を行うことの重要性や行うまでの心得から、定量・定性双方のデータ収集－分析－考察の事例と手法までを、わかりやすく解説することを目指しました。

導入書ではありますが、大学院へ社会から戻り入学した人たちが、あらためて学術的な研究を行う上で勘を取り戻す上でも効果的な書であると考えています。勘を取り戻して頂けるようにも、内容面で十分配慮致しました。もちろん、卒業論文を書く多くの学部生にも役立つ内容に仕立ててあります。さらには、これから社会調査（データ収集）やデータ分析の世界に入る多くの社会の方にも役立つ、汎用性の高い書であると確信しています。

筆者は、社会調査（データ収集）やデータ分析の基礎知識とそれに取り組む姿勢が、筆者の豊かな経験に根差す形で、みなさまの読後にしっかりと身につくよう工夫致しました。第1章で「データ分析とはどういうことか」、その世界観をわかりやすく解説し、第2章で「データの種類について」も説明しました。その基礎知識をもって、第3章の「データ分析を伴う研究の流れとポイント」を読んで頂けたら、力がぐんと伸びると思います。第3章ではデータを収集するためのノウハウや社会的なルールなどお作法としておさえておきたいことをすべてまとめてあります。そして第4章と第5章では定量的データの分析、定性的データの分析について実例にもとづき、わかりやすく説明しております。是非本

書を手に取られ、データ収集・分析の世界を味わってほしいと思います。そして学術的研究で、みなさまがしっかりとしたデータ収集・分析のアカデミック・スキルズを身につけ、実社会での活動や大学院での高等研究でそれを活かされることを期待致します。

なお、末筆ながら本書は極東証券株式会社の寄附講座である「アカデミック・スキルズ」が無ければ決して生まれなかつた知的資産です。同社の協力なくして、本書が刊行されることは無かつたと思います。ここに極東証券株式会社にあらためて感謝の意を表します。また、本書で用いたデータを提供して下さった慶應義塾大学総合政策学部の片岡正昭教授や学生諸君、看護医療分野の観点を中心に貴重なアドバイスを下さった目白大学人間学部の西山里利専任講師、編集過程で手厚いご指導を頂いた、慶應義塾大学出版会の佐藤聖様、浦山毅様、木下優佳様にも心より御礼申し上げます。

筆者は、本書の活きた社会調査（データ収集）とデータ分析のノウハウが、社会に広く貢献することを心より祈念しております。

2013年6月

筆者を代表して  
西山敏樹

## Contents

はじめに.....	3
<b>第1章 データ分析とはどういうことか .....</b>	<b>9</b>
1. データとは何か .....	10
2. データ分析の定義とデータ分析を行う者の心得 .....	11
3. データ分析の最低限の用意 .....	13
<b>第2章 データの種類について .....</b>	<b>15</b>
1. 「定量データ」と「定性データ」への大別 .....	16
2. 定量データの分類 .....	17
3. 定性データの分類 .....	18
<b>第3章 データ分析を伴う研究の流れとポイント .....</b>	<b>21</b>
1. 研究テーマの絞り込みと社会調査の計画 .....	22
2. 社会調査の準備 .....	39
3. 社会調査の実行 .....	65
4. 社会調査で得られたデータの分析・公開・社会への還元 .....	80
5.まとめ .....	90
<b>第4章 研究例（定量編） .....</b>	<b>91</b>
1.はじめに .....	92
2.用語確認 .....	93
3.平均値の差の検定 .....	99
4.クロス集計表 .....	102
5.回帰分析 .....	104
6.主成分分析と因子分析 .....	109

第5章 研究例（定性編）	115
1. 調査の実例：世界初の電動大型フルフラットバスを運転した ドライバーへの個別インタビュー	116
2. 具体的な研究活動例：日本語の日常会話のコーパス作成	133
会話参加者シート	140
会話録音（録画）情報シート	142
附録	145
社会調査質問票サンプル『交通環境のバリアフリー化に対する 価値観調査』	146
交通バリアフリー施策に対するグループインタビュー	167
社会調査原稿記録用紙	174
社会調査の倫理的配慮項目のリスト	175
参考文献リスト	178



## 第1章

---

データ分析とはどういうことか

---

## 1. データとは何か

---

今、読者のみなさんが、お店でジュースを買うとしましょう。店員が、「この新しいジュースはビタミンCがレモンよりたくさん含まれており、健康にもよく、よく売っていますよ」と言ってくるのと、「この新しいジュースはメーカーが分析したらビタミンCがレモンの**10倍**ほど入っており、**5種類**の病気に効果があり健康によく、最近では我々の店でも他のジュースの平均**3倍**は売っています」と言ってくるのと、どちらが、信頼度がより高く聞こえますか？

普通は、後者だと思います。メーカーまたはお店が独自に分析した数字、いわゆる「データ」があり、しかもそれが正確であればあるほど、新しいジュースへの買い手側の信頼度は増して、購買意欲も増すものです。

こうした我々の身近な生活シーンの例からも判るように、データは、物事の推論（この例では、買い手がお店で買うか買わないか、併せて買った後に、自分にとって、どの程度のプラス面とマイナス面があるかを市民が推定しようとしている段階ですね）の基礎となる事実であり、また意思決定の参考となる資料・情報と定義づけられます。

◎データとは◎

あらゆる物事の推論の土台となる事実であり、また、意思決定をするときの参考となる資料・情報のこと。

## 2. データ分析の定義とデータ分析を行う者の心得

ところで本書は、大学生でデータ分析を初めて行おうとする人や、社会人の大学院生で大学時代のデータ分析の勘を効果的、かつ効率的に取り戻したい人を対象にしています。要は、大学・大学院の研究・教育活動の様々な場面でデータを分析して、リポートや学術論文等にまとめる人たちに役立つよう、まとめています。

高等研究機関である大学または大学院で教育を受け、研究を行うみなさんに向けて、実社会で活躍する前に、正しくデータを集め、それらを分析し、あらゆる物事の推論の土台となる客観性の高い事実、あらゆる意思決定に役立つ資料・情報をまとめる手法をアカデミック・スキルズ（学問を行うまでの作法）の一つの重要な手法として、しっかり身につけてほしいと思っています。

分析とは、「ある物事を分解し、物事の側面やそれらを成立させている要素等を明らかにすること」です。要は、データ分析とは、次のプロセスです。

### ●データ分析●

あらゆる物事の推論の土台となる事実、また、意思決定をするときの参考となる資料・情報を客観的かつ公平に導き、整理する目的で、物事をできるだけ詳細に分解し、物事が持つ側面・性格や、成立させる要素等を明らかにする作業のこと。

上記の新しいジュースを例にとれば、「ビタミン C がレモンの 10 倍ほど入っている」、「5 種類の病気に効果があり健康によい」、「最近では我々の店でも他のジュースの平均 3 倍は売れている」というのは、新し

いジュースが持つ側面や性格を多くのデータから分析した大切な結果です。このデータを信じて新しいジュースを買って、健康を取り戻した人々がいれば、まさにデータを分析した意義が十分あったと言えます。

ただ一方では、データが人の行動を大きく左右するものである、ということを忘れないで欲しいと思います。新しいジュースを売りたいがために、メーカー側が「ビタミンCがレモンの5倍なのに10倍と言っている」、お店側が「本当は既存のジュースの1.5倍しか売れていないのに3倍売れていると言っている」という事実を知ったら、買い手はとてもがっかりし、裏切られた気分になります。

われわれは、この本をいわゆる「べからず集」にはしたくありません。ただ、残念なことにデータを改ざんしていた、という事例は、社会でもたくさん見受けられます。被験者から取得した薬の治験データを改ざんしていた、食品会社が摂取後の検査データを改ざんしていた、というような話は、マスコミでも時々取り上げられています。みなさんも、そういう話を一度は聞いたことがあるはずです。

分析したデータは、人の人生すら左右しかねない、ということをデータ分析者の心得として、抑えておいてほしいと思っています。なぜなら分析されたデータは、我々一人ひとりの意思決定に資する重要な情報となるからです。多数のデータに裏打ちされたという特効薬が、実は効果もなく、結局人々が購入代金をたくさんとられただけになったようなケースは、社会的に決して許されるものではありません。

要は、データ分析を行う時の三大心得は、次の通りになります。

### ◎データ分析を行う際の三大心得◎

- (1)客観的であれ—主觀を交えずに物事を冷静に見つめる姿勢
- (2)公平であれ—物事に偏見を含めず平等に見つめる姿勢
- (3)倫理的であれ—改ざんなどをせずに人としてのモラルを守る姿勢

みなさんは、大学や大学院で教育を受けて研究を行い、さらには実社会での業務で様々なデータ分析の機会を得ると思います。ゆえに、上記(1) – (3) を早い段階で身につけ、自分のデータ分析の結果で、人を誤った方向に導かないよう心がけが必要です。

---

### 3. データ分析の最低限の用意

---

まず、みなさんがたくさんの情報（データ）を集め、それらを分析する上では、データ分析用のコンピュータソフトウェアを用意することが必須となります。今日では、コンピュータの利用が標準的に行われるようになり、大学や大学院で研究を行う際に行政や企業等が提供してくれるたくさんのデータも、紙に印刷されたデータではなく、すぐにコンピュータ上で読み取れるデジタルデータの場合が多くなりました。

これから本書の解説では、データ分析の初心者、しかも数学的知識が無い人でも理解が進むように、特定のソフトウェアに偏ることなく記述しています。おそらく多くのみなさんのパソコンには、Microsoft Excel がすでにインストールされていることと想像します。もしくは、IBM SPSS Statistics (Excel のような初心者にはわかりにくい複雑、難解な関数の入力を省略でき、複雑な分析の手法もメニューの中からボタンを押すだけ) も大学でよく用いられており、利用可能な方もおられるでしょう。

是非みなさんには、こうしたデータ分析用のコンピュータソフトウェアをご用意頂きたいと思います。もちろん、ご自分が使いやすいものが良いので、われわれが特定のデータ分析用のソフトウェアを薦めることは致しません。本書では、以下で、様々なデータ分析の手法及びその事例をご覧頂きます。ご自分が用いるソフトウェアで、本書で紹介する各種分析手法を実践する際にどのような操作になるのか、また、どのよう

な操作を経れば事例で紹介する研究と同じプロセスを経ることが出来るのか、みなさんに一つひとつ確認しながら読み進めてほしいと思います。

本書では、以降で「データの種類」、「データ分析の進め方」についても、わかりやすく解説を行います。自分の研究テーマが決まっている人や分析したいデータがある人も多いと思います。本書を読み、どのような分析手法を用いて、どのようなコンピュータソフトウェアの操作を経れば分析が効果的かつ効率的か、想像しながら読むと良いと思います。

## コラム

### 「CSV形式とは」

データ分析の手法を大学や大学院で教えていると、「Excelに打ち込んだデータ類を他のソフトウェアで見たり、それらを分析したりすることは出来ないのでしょうか?」とよく質問されます。その際に用いるのが、「CSV形式」というものです。

CSV形式とは、データ類をカンマ(“,”)やタブで区切り並べたファイル形式のことです。カンマ区切りのものを「Comma-Separated Values」といい、タブ区切りのものを「Character-Separated Values」といいます。つまり、「CSV形式」は、こうした2種類のファイル形式の総称となります(2種類とも頭文字が「CSV」)。主にExcelやSPSS等のデータ分析用ソフトウェアでデータを保存する際に用い、異なる種類のデータ分析用ソフトウェア同士のデータ交換に使われることが多いです。例えば、Excelでデータの保存の際に形式をCSV形式として、他のデータ分析用のソフトウェアで開くと、そちらで閲覧・分析できる状態になる、というイメージです。

これなら、企業相互でデータ分析用のソフトウェアが異なっていても、作業を速やかに進めることができます。

## 第2章

---

### データの種類について

データにも種類があります。いわゆる「定量データ」と「定性データ」に大別されることをよく把握しましょう。本章ではデータの種類についてわかりやすく、簡潔にまとめます。

---

## 1. 「定量データ」と「定性データ」への大別

---

### ●定量データ

定量データは、「数値で計測・集計・分析が可能な情報」のことと言います。

例えば、駅の利用者数、病院の患者数、レストランの利用者数は、すべて数値としての計測・集計・分析が可能です。血圧値や心拍数などの数値で表される身体的データ、スーパーマーケットやコンビニエンスストアの基本的な属性（お客様の年齢・勤続年数等）、数値で表されるデータも定量データです。

こうして定量データなら、数値情報として計測・集計・分析の一連の作業が容易にできます。そのため、数値で明確に、わかりやすく人や組織の意思決定や判断に資する情報をまとめられる特徴があります。

一方で、ある商品を買うに至った気持ちの推移、ある野球のチームを応援するに至った具体的な経緯等、数値に表れない心の変化等を具体的に明確化することには向かず、それらは後述の定性データがカバーする範囲となります。

### ●定性データ

定性データは、「数値化できる定量データ以外の文章・画像・動画・音声等の形式をとる情報」のことと言います。要は、数値化できる情報を「定量データ」、それ以外のデータを「定性データ」と分類します。

例えば、新しいバス路線のバス停の間隔を検討するときに、許可を得

て高齢者の日常生活を動画にとらせてもらう場合があります。「どれくらい歩いて、立ち止まり、休むのか」を動画に記録しておけば、バス停の間隔を決めるのに役立ちます。こうした動画データそのものは、数値化して分析できるものではありません。ただし、高齢者が歩行中に立ち止まって休む時間のように一部のみをきりとり、数値で分析は可能です。

また、新商品の開発では、市民を7ないし8人程度を会場に集めて、開発する分野の多くの商品群から普段どれをどういう理由で買うのか、また新商品の試作品を体験してもらい買う理由、買わない理由等を詳しく調査することができます。こうした、グループインタビュー調査で記録した動画や音声そのものも、数値化して分析できるものではありません。

こうした定量データの行間にあるような人の心理や潜在意識、動画で無いとわからない事実等、数値化できない情報を定性データと呼びます。

データは大きく分けて「定量データ」と「定性データ」に分かれることを理解頂けたでしょうか。

その定量データもさらに2種類のデータに、定性データも同じくさらに2種類のデータに分けることができます。以下では、それを見ていきたいと思います。

---

## 2. 定量データの分類

---

定量データは、具体的に、次の2種類のデータに分類されます。

### ●比例尺度

比例尺度は、数値の差と数値の比の両方に意味がある尺度のことを言います。例えば、身長が<sup>g</sup>100cm から 120cm になれば数値の差は20cm、

数値の比としては20%伸びたことになります。同じく120cmから180cmになったときは、数値の差が60cm、数値の比としては50%伸びたことになります。こうして、数値の差と数値の比の両方に意味があることがわかります。

### ●間隔尺度

間隔尺度は、数値の差に意味を持つ尺度のことを言います。代表例は温度です。5度と10度の数値の差（5度分の差）には意味がありますが、5度と10度の比率（5度と10度、2種類の温度が50%ずつになっている）には意味がありません。こうして、数値の差のみに意味があることがわかります。

---

## 3. 定性データの分類

---

さらに定性データは、具体的に、次の2種類のデータに分類されます。

### ●順序尺度

順序尺度は、順序にのみ意味がある尺度のことを言います。間隔や比率には意味がありません。例えば、好きなゲームの順を1位将棋、2位オセロ、3位チェス、4位囲碁……というデータは、順序にのみ意味があり、好きな度合いの比較は順序から判断することは可能ですが、間隔や比率には意味がない点がわかります。

### ●名義尺度

データを分類するために、単に整理番号として数値を割り当てるものを名義尺度と言います。代表的な例は血液型で、A型を1番、B型を2

番、O型を3番、AB型を4番としてふった整理番号が名義尺度となります。要は、データを区分する目的でふる單なる番号でしかなく、数値の平均値等には全く意味がないことがわかります。

まとめると、「名義尺度<順序尺度<間隔尺度<比例尺度」の順に精度が高くなります。情報量が多い尺度は、情報量が少ない尺度を兼ねることができます。一方、その逆は当然ながらできません。例えば、年齢を32歳、36歳、45歳……というように具体的に数値で書いてもらえば、30歳代10人、40歳代6人というようにまとめられますが、30歳代、40歳代などに○をつけてもらった場合に、具体的な年齢は二度と読みとれなくなります。また、定性データも事前に調整すれば数値で定量データのように扱うことができます。

今後、データを扱う上で上記の分類は最低限覚えておきましょう。

## コラム

### ビッグデータ

最近、ビッグデータという言葉が新聞やテレビで見聞きされる様になりました。ビッグデータは、「インターネットの普及やコンピュータの処理速度の向上等に伴って生成される大容量のデジタルデータ」（朝日新聞『知恵蔵2013』より）のことを言います。近年、ブログやYoutube等の動画サイト、Facebookやmixi、Twitter等のSNS（ソーシャル・ネットワーキング・サービス）、Suica等の電子マネーの利用が進み、大量の文字・音声・写真・動画等のデジタルデータがインターネット上を飛び交っております。これらのデータは、まさしく大規模な社会調査が自動で行われていることと同意です。ビッグデータを分析すれば、人の移動や購買の状況、嗜好等を即時に把握する事が出来、マーケティングに役立てる事も可能となります。ただし、ビッグデータの売買が近年問題になっております。人の移動の状況や購買活動、嗜好等は個人情報そのものであり、集まったビッグデータを必要とする企業に販売する行為が社

会問題化しています。ビッグデータは誰のものか、という議論もあります。ビッグデータの使用方法については国民のコンセンサスも必要になっています。

.....

### 第3章

---

データ分析を伴う研究の流れとポイント

本章では、データ分析を伴う研究の標準的な流れについて説明します。ただやみくもにデータを集めのではなく、しっかりと以下のように研究計画を立て進めましょう。

まず本書では、データ取得に際して学問的に極めて重要な社会調査の計画から、取得したデータの分析・考察、さらには、成果のプレゼンテーションまでの標準的な流れを示します。データ分析を伴う研究は、データの収集方法を確かなものにし、かつ信頼のあるデータを集めることが第一義です。信頼性の低いデータを分析しても意味がないわけですから、データの収集方法から分析、考察、まとめの方法までをしっかりと本章で学んでください。

---

## 1. 研究テーマの絞り込みと社会調査の計画

---

ここではまず、研究テーマの絞り込みとデータ獲得のための社会調査の計画について述べます。

自分が解明したいテーマをじっくり見据え、いわゆる自分が行う社会調査の5W1H（Who、Which、What、Where、When、How）を決めていきます。そして、その計画が調査に答える人々のこととも考えた倫理的配慮（調査対象者のリスク回避と人としてのルールに則った調査か）がなされたものであるかをしっかりと点検するノウハウまでを述べます。

一部、社会調査を実践する過程で、次の節の「社会調査の準備」とクロスオーバーするところがありますので、ぜひ読み通してください。

①社会調査で解明したいことの絞り込み—自分の研究のオリジナリティを—  
ここでもう一度、社会調査とは何か、原点に立ち戻りたいと思います。辞書をいくつか引くと、「社会調査」 = 「物事の実態や動向、問題点等