

“データ分析の進め方を理解したい人に”

この

Book Review 3冊!

問題意識

前回のこのコーナーでは、分析の方法論について学ぶための3冊を紹介しました。ただし、あくまでも方法論ですので「では、具体的にどうするの?」についての答えは得られません。そこで、今回は、具体的にデータ分析を進めるには、どのような視点で、どのように進めるのかを学ぶための3冊を紹介したいと思います。

これまでのマーケティング・リサーチの教科書では、集計やデータ分析のステップについて具体的に学ぶための本はあまりなかったと感じています。統計や多変量解析で分析の基本を学び、あとは実務上でデータ取扱いの経験を積むことがデータ分析のための勉強だったのではないのでしょうか。ところが、ビッグデータに注目が集まることで、データ分析の進め方そのものに焦点を当てた本が多く出版されるようになり、データ分析の進め方について具体的なノウハウを学ぶことができるようになりました。

しかし、「なんだ、ビッグデータ分析か」と思わないでください。ここに書かれていることはビッグデータ分析にのみ当てはまることではなく、アンケート調査データや質的データの分析においても重要なポイントです。

3冊に共通するデータ分析のポイント

今回紹介した3冊に共通するデータ分析のポイントがあります。多くの人が指摘するということは、それぞれがデータ分析ステップの本質部分になると思います。特に重要だと思う3つのポイントを以下で整理していきます。

なかでも高橋氏(『データ活用実践教室』)のPPDACというデータ分析のマネジメントサイクルモデルがデータ分析の進め方を端的に示していると思いますので、以下に紹介しておきます。

Problem(問題の明示化と課題設定)
Plan(どのようなデータを集め、分析するか)
Data(データを集め、整備する)
Analysis(データを加工・分析する)
Conclusion(結論を出す)

①分析の課題を明確にすること

当たり前なことだと思われるかもしれませんが、多くの人がこのことを指摘しているということは、課題が明確ではない分析が多いからではないでしょうか。

分析の課題が明らかでなければ、どのようなデータを集めるべきか、どのような手法を使うべきか、どのようなアウトプットをするべきかがわかりません。まさに、問題の明確化と課題の設定が分析を進めるうえで欠かせないポイントになります(このことは、マーケティング・リサーチの重要なポイントでもあります)。

そして、あんちべ氏(『データ解析の実務プロセス入門』)の指摘するMust(必須な分析)、Should(できる限り行うべき分析)、Could(可能なら行う分析)、Won't(行わない分析)というブレイクダウンも分析の指針として重要な示唆を与えてくれます。

②データの背景を明らかにすること

ビッグデータをベースにデータ分析をする際には、データの背景(素性や性質とも)を明らかにすることも重要になり、このことを強調する方も少なくありません。具体的には、そのデータがどのような背景で集められたデータなのか、さらに今回の課題を明らかにするために適切なデータなのかを検討することです。

しかし、これはビッグデータに限らず、アンケート調査でも大切な視点だと思います。以前は母集団定義→適切な抽出フレーム→ランダムサンプリング→高い回収率という過程でデータが集められたことで、集まっ



1

トップデータサイエンティストが教える データ活用実践教室

高橋威知郎他著(2015)
日経BP社 2,000円(税別)

常に「達人たちが語る“極意”をつかめ」とあるように、データサイエンティストとして著名な方々(7名)がデータ分析とは何かについて整理している本です。ですから、具体的な内容というよりも分析のプロセスであり、勘どころであり、スタンスが紹介されています。しかし、最初からノウハウ中心の分析本に手を出すよりも、まずは本書で分析の本質や極意を理解することのほうが重要だと思います。これらを理解することが効率的で有効なアウトプット(アウトカムといったほうがよいでしょうか)につながるはずです。



2

データ解析の 実務プロセス入門

あんちべ著(2015)
森北出版 2,400円(税別)

データ解析について書かれた本の多くはデータが揃った段階から始まるものが多いのですが、実は「よきデータ(分析にふさわしいデータ)」を作り、「どんなデータなのか」を確認することがとても重要です。本書はこのデータ分析の最初の段階であるデータの整備(よきデータとは何か、そのための作業)や最初の分析(探索的データ解析の進め方や注意点)についてページを割いているところがおすすめポイントです。入門書ですので、すでに知っていることも多いかもしれませんが、データについて改めて理解し、具体的に何をするのかについて理解するには最適の本だと思います。



3

統計思考入門 —プロの分析スキルで「ひらめき」をつかむ

水越 孝著(2014)
プレジデント社 1,600円(税別)

著者は、本書について、「物事を考えるプロセスにおいて、統計の技術をどう活用するか」をテーマとした本だとしています。従って、内容も、具体的なビジネス課題を提示したうえで、実際のデータを使って分析のポイントやプロセスを解説したものになっています。データ分析(多変量解析も含めて)を、どう使うのかを理解する最初の取っかかりとなる本だと思います。著者は矢野経済研究所の方なので、データの素性の部分も母集団と標本という視点で書かれているなど、前2冊の本と比べて狭義でのリサーチ視点(スモールデータ視点)で書かれている本です。

たデータが分析対象として適切だという前提がありました。しかし、今ではこのサンプリング理論を実行すること自体がむずかしく、集まったデータがどのようなデータなのかを確認し、理解することが、分析に際しては重要なポイントになると思います。

の重要性についての認識が低くなり、またこの過程を体験する機会が減少したようにも思います。それゆえ、いま一度、データ整備の意味とノウハウを理解する必要があるでしょう。

③データを整備すること

分析に適するようにデータを整備(クリーニング、クレンジング)することも、多くの人が指摘するポイントです。

例えば、安宅氏(『データ活用実践教室』)は、「歪み」「汚れ」「欠落」という視点でデータを整備する必要性を述べています。

あんちべ氏(『データ解析の実務プロセス入門』)も「GIGO」という言葉を使い、データ整備の重要性を説いています。GIGOとは、Garbage In Garbage Out のことで、「ゴミを入れてもゴミが出てくる」という、データ分析でよく使われる戒めの言葉です。

データ整備は、ネットリサーチとなる前はアンケート調査でも重要なステップだったのですが、見かけ上、きれいなデータが得られるために、データクリーニング

データ分析ステップの大切さ

今回紹介している本では、データ分析の視点や具体的な解析ステップについても紹介しています。大きく整理すると、大きさをみる、比較する(差異と類似点)、分類する、となりそうです。しかし、これらの分析視点や解析は、正しいデータがあるから意味があります。だからこそ、今回のテーマであるデータ分析の進め方を理解することが、とても重要になるのです。

分析の課題を明確にすること、データの背景を明らかにすること、データを適切に整備すること。これらはデータを扱ううえで常に念頭に置いておきたいポイントですし、今回の3冊を通じて、その意味するところを理解しておきたいものです。

りんく考房

鈴木 敦詞