

データサイエンス基礎講座 超初級 第3限

フューチャーブリッジパートナーズ株式会社

長橋 賢吾

第3時限 データってどうまとめるの？

▶ ドクター：第2限では、第1限から一歩進んで、回帰分析、主成分分析、因子分析を取り扱いました。



▶ あゆみ：主成分分析、因子分析が難しかったです。



▶ ドクター：はい、これは誰でも戸惑う地点です。いろいろなデータについて、主成分分析、因子分析を使ってみて、体感する、これが重要です。



アソシエーション分析とは？

- ▶ ドクター：今回、最初に取り上げるのは、アソシエーション分析です
- ▶ あゆみ：アソシエーション分析??
- ▶ ドクター：はい、そうです。
- ▶ あゆみ：アソシエーションって何ですか？
- ▶ ドクター：アソシエーションとは、日本語に訳すと、関連です。
- ▶ あゆみ：えーと、関連って？



アソシエーション分析とは？

▶ ドクター：関連とは、たとえば、ECサイトで、この本を買った人は、これも買っています、というおススメが表示されているのを見たことありませんか？



▶ あゆみ：はい、あります。それが関連ですか？

▶ ドクター：その通りです。ある製品とある製品の関連性を分析、ルールを見つけて、そこからおススメを表示すれば、おススメ商品を買う率は上がりますよね？



▶ あゆみ：たしかに、そうですね～

重回帰分析とは？

- ▶ ドクター：一つの説明変数から複数の説明変数になると何が変わるのでしょうか？
- ▶ まゆみ：う～ん、変数が変わるくらいしかわかりません。
- ▶ ドクター：一つの説明変数であれば、一つの説明変数がどれだけ変数を説明できるかにフォーカスすればよいのですが、複数の場合、それぞれの説明変数がどれだけ説明できるのか、そして、説明変数間にどんな関係があるのか、理解する必要があります。
- ▶ まゆみ：ちょっと、わかりません～



コラム3 データサイエンティスト

筆者は、長年にわたってIT企業を内から、外から、いろいろな会社をみる機会がありました。そのなかで、感じたこと、それは、“デキる人”の存在です。

“デキる人”には、様々な意味があります。仕事ができる人、容量が良い人、ここ一番で活躍する人などです。筆者が思う、IT企業でのデキる人、それは、応用が利く人です。

ITのすべてとは言いませんが、エンジニアはたいてい、システムを作るために必要なプログラム言語を学習します。そして、たとえば、Javaを学習したとして、1. Javaしかできないプログラマー、2. Javaのエッセンスを活用して、Rubyなどほかのプログラミング言語にも精通するプログラマーです。企業の採用の観点から言えば、言うまでもなく、欲しいのは後者の人材です。

翻って、筆者が思うに昨今、流行っているデータサイエンティストもこのプログラマーの例に近いと言えます。すなわち、R, SPSS, Excelなどある特定の分野に関して対応できる人、もしくは、あるシチュエーション、分類をもとに、次のシチュエーション、分類を踏まえること、これが重要といえるのではないのでしょうか。