データサイエンス基礎講座 超初級　第２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１時限　どれで分析すればいいの？データはあるけど

- ドクター：第1限では、統計とは、ルールを見つけること、そこから、Ｒでグラフを描く、平均、相関について取り上げました。

- あゆみ：統計でルールを見つけるってことはわかったけど、具体的にどうすればいいか、まだ、わからないです。。。

- ドクター：場数をこなすことが重要ですね。そして、ルールを発見するためのツール（統計手法）の理解も合わせてやっていきましょう。

回帰分析とは？(1)

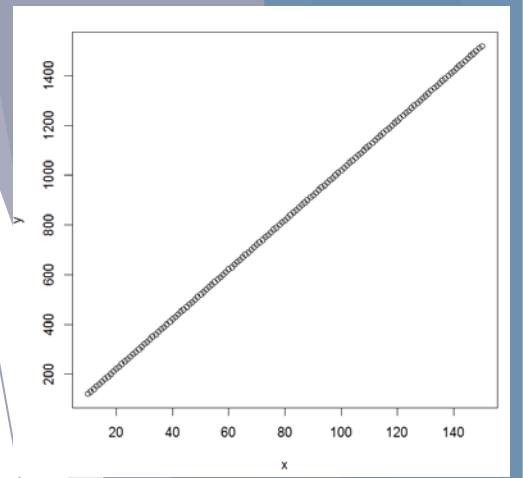
- ドクター：あゆみさん、y= ax + b っておぼえていますか？

- あゆみ：たしか、中学でやったような。
- ドクター：そうですね。

- あゆみ：それと統計とどう関係があるんですか？
- ドクター：たとえば、aが10, bが20であれば、 y = 10x + 20,xが10であれば,yは120になりますよね

- あゆみ：はい

回帰分析とは？(2)



- ドクター：y=10x+20をグラフにするとこうなります。

- あゆみ：はい、でも、統計とどう関係あるんですか？
- ドクター：いい質問です。これって、ある意味、ルールですよね。x にどんな値を入れても y はルール通りに決まります。

- あゆみ：たしかに、そうですね～

コラム２　ワインの方程式

「その数学が戦略を決める」（イアン・エアーズ、文春文庫、2010年）は、人間の下手な“先入観”より、コンピュータによる“数学”の方が、より有効な意思決定ができることを示唆しています。

そのなかでの、エピソードはワインの質。ボルドーワインは、毎年、気温などによってその質（クオリティ）は変わりますが、何が影響を与えるのか。長年、ワイン仲買人が自身の舌でその質を決めていましたが、そうした“アナログ”な状況に一石を投じたのが、統計学者アッシェンフィルターです。

彼の長年の観測によれば、ボルドーワインの質＝12.465 + 0.00117 ×　冬の降雨量 + 0.0614 ×　育成期平均気温 - 0.00386　×　収穫期降雨量、であると指摘します。これは言うまでもなく、今回取り上げた回帰分析の結果です、正確には、変数が2つあるので、重回帰分析です。

この方程式をどうとらえるか、これはその人次第です。ワインの質はこんなに完結に表現できるのかという指摘もあれば、結局のところ、最も重みのある係数(0.0614)は育成期平均気温であり、育成期平均気温が高ければよいワインが育つ、そんな指摘もできるでしょう。

筆者が思うに、この回帰分析は、たしかに、育成期平均気温に依存しているかもしれません。ただし、それを2014年、2015年、2016年と当てはめることによって、より、強固な説明力のあるモデルになりうるということです。という点において、所詮は数字かもしれません、その数字を上手く使うこと、それも重要と思うのです。