

傾向線

「傾向線」についてのトレーニングによろこそ。

付属の練習ワークブックをダウンロードして、実際に Tableau を操作してみてください。

傾向線の追加

傾向をつかむことは、データ分析に重要なインサイトをもたらします。例えばここでは、「風速が上がると、電力の出力はどのくらい増加するか?」という問いに答えることを考えてみましょう。風速が上がると電力の出力も増加することが確認できるため、そこには関係があることが分かります。しかし、正確には、どのような関係があるのでしょうか。

簡単な操作で、傾向線をビューに追加できます。

- アナリティクスペインを開き、[傾向線] を使いたいモデルタイプにドラッグします。
- 傾向線は削除するのも簡単です。ビューの外にドラッグするだけです。ここでは、傾向線が必要なので、操作を元に戻します。

傾向線のオプション

既定では、傾向線はペインおよび色ごとに設定されます。

- データタブに戻り、[場所] などのディメンションをビューに追加する場合には、ペインまたは散布図ごとに傾向線が設定されます。
- 同じように、[場所] を [色] に追加する場合、傾向線は 3 つに分かれます。
- [場所] を [色] で表示し、全体を 1 つの傾向線で表したい場合は、傾向線を変更します。

傾向線を編集するには、右クリックして [傾向線] を選択し、[傾向線の編集] をクリックします。

- [色ごとの傾向線を許可] のチェックを外すと、全体で 1 つの傾向線に戻ります。
- ビューをシンプルにするには、[信頼区間の表示] のチェックを外すこともできます。

このダイアログボックスには、他にもオプションがあります。

- 最初は [モデルタイプ] です。
 - 最初にアナリティクスペインから傾向線を設定するときと同じオプションがあります。
 - このオプションでは、1 つまたは両方の変数に対する変換に基づいて、線形回帰モデルを構築します。
 - ここでいう線形とは、係数 (けいすう) についてであり、変数の関係が線形ということではありません。
 - モデルタイプと変換の詳細については、[傾向線のモデルタイプ](#) についてのオンラインヘルプを参照してください。
- 信頼区間は、そのモデルにおける 95% の信頼区間を示します。
- また、y 切片を強制的に 0 にするオプションもあります。

傾向線の有意性

傾向線が意味のある情報を示しているかどうか、評価する必要があります。

- 傾向線にカーソルを置くと、傾向線の式と P 値、決定係数 R^2 を示すツールヒントが表示されます。
- 統計学において、P 値とは有意性の概念を示す数値です。
 - P 値が、通常 0.05 のカットオフ値より小さい場合、その結果は有意と解釈されます。

- 0 から 1 のスケールで P 値が大きい場合、データから見られる傾向が偶発的なものである可能性があります。
- この例では傾向線の P 値が非常に小さいので、信頼できます。

しかし、モデルの適性を正確に評価するには、P 値だけでは十分ではありません。

- R-2 乗を確認すると、モデルがデータに適しているかどうかを効果的に判断できます。R-2 乗値の範囲は 0 から 1 で、値が高いほど適正度が高いと判断されます。
 - この例では、R-2 乗値が非常に高く、0.956 です。
 - つまり、このモデルがデータに非常に適していることを示しています。R-2 乗が 1 ならば、それは完璧に適したものであるといえますが、たとえば R-2 乗が 0.999 など、ありえないほど高い場合、そのモデルに問題がある可能性があるので注意してください。一般的には、R-2 乗値が不自然に高いかどうかは、自由度が低いことから判断できます。あるいは、単にデータポイントが多すぎるのが考えられます。

傾向線の残差

P 値が小さく R-2 乗が大きいだけでは、そのデータを表す傾向線の適切さを判定するには不十分です。

データポイントのすべてが予測される傾向線に収まるわけではないからです。特定のポイントから予測値までの距離が、誤差もしくは残差です。

正しいモデルでは、説明変数に対してこの残差をプロットするとゼロ付近の無作為正規分布になります。この残差プロットが正規分布していない場合、データが予測値と合わない傾向、つまりそのモデルは最善ではないことを示しています。

傾向線を使用してビューの残差値を求めるには、次の手順を実行します。

- [ワークシート] > [エクスポート] > [データ] を選択します。
- ファイルを保存するよう求められるので、「傾向線の残差」という名前で保存します。現在サポートされているのは Microsoft Access 形式のみです。
- [エクスポート後に接続] を選択します。
- このデータソースには、散布図からの元データと傾向線から得られた予測値、残差が含まれています。

[風速] を [列] にドラッグして説明変数を水平軸に、[残差] を [行] にドラッグして垂直軸におき、残差プロットを構築します。[風車] を [詳細レベル] にドロップします。

良好なモデルであれば、ゼロ付近で正規分布が見られるはずですが。

この例のモデルでは、風速値に基づいて発電力を正しく予測できていないことは明らかです。この傾向線では P 値と R-2 乗値が良くても、残差プロットはひどい状態です。

まとめ

Tableau における傾向線の概要と、関連する統計の解釈方法について説明しました。ご視聴ありがとうございました。引き続き無料トレーニングビデオをご覧ください。