

Excelで行う相関分析

データ分析を行う場合、データの関係性を知ることが基本的かつ重要なことです。データの関係性を明らかにする分析手法に相関分析があります。相関分析では、大量のデータの中からデータ間の関係性を見つけることができ、関係性の有無を客観的に判断することができます。統計データ分析コンペティションの応募論文（高校生の部）中でも多く用いられる分析手法です。

相関分析とは

相関分析では、複数のデータ（変数）間にどのような関係性、すなわち「相関」があるかを見ます。変数間の相関を見る方法にはいくつかありますが、ここでは最も一般的な次の方法を行います。

- ・ 散布図を描く
- ・ 相関係数を算出する

相関分析は様々なツールで行うことができますが、ここでは最も一般的で手軽な Excel を使った方法を紹介します。

統計データ分析コンペティション 2022 の受賞論文（高校生の部：統計活用奨励賞）「都市部と地方の教育格差の要因と課題～日本の教育現場において～」

(<https://www.nstac.go.jp/sys/files/static/statcompe/files/2022/2022H4-katsuyo.pdf>)
を例に見てみましょう。

受賞論文では、都道府県間で教育格差の生じる要因を所得や教育費など様々なデータで相関分析を行うことで明らかにしています。

論文中の図4は、都道府県別の「一人当たり県民所得」と「教育費」（一世帯当たりの月間消費支出額）の関係性について相関分析を行った結果です。

これを実際に Excel で行ってみましょう。

分析で用いるデータは、「一人当たり県民所得」が SSDSE-基本素材 (SSDSE-E) の項目コード C121101 を、「教育費」が SSDSE-県別推移 (SSDSE-B) の項目コード L322108 です。

データは、SSDSE のウェブサイトからダウンロードできます。

(<https://www.nstac.go.jp/use/literacy/ssdse/>)

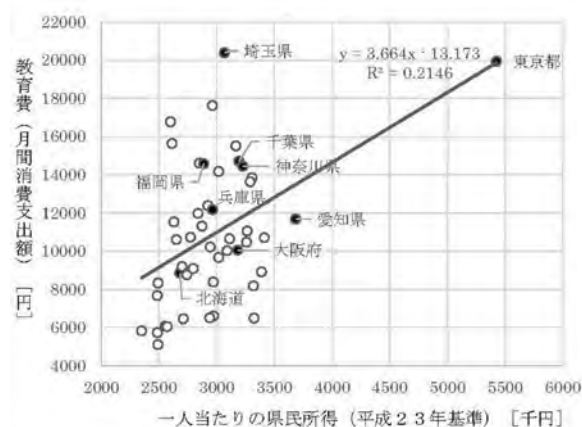


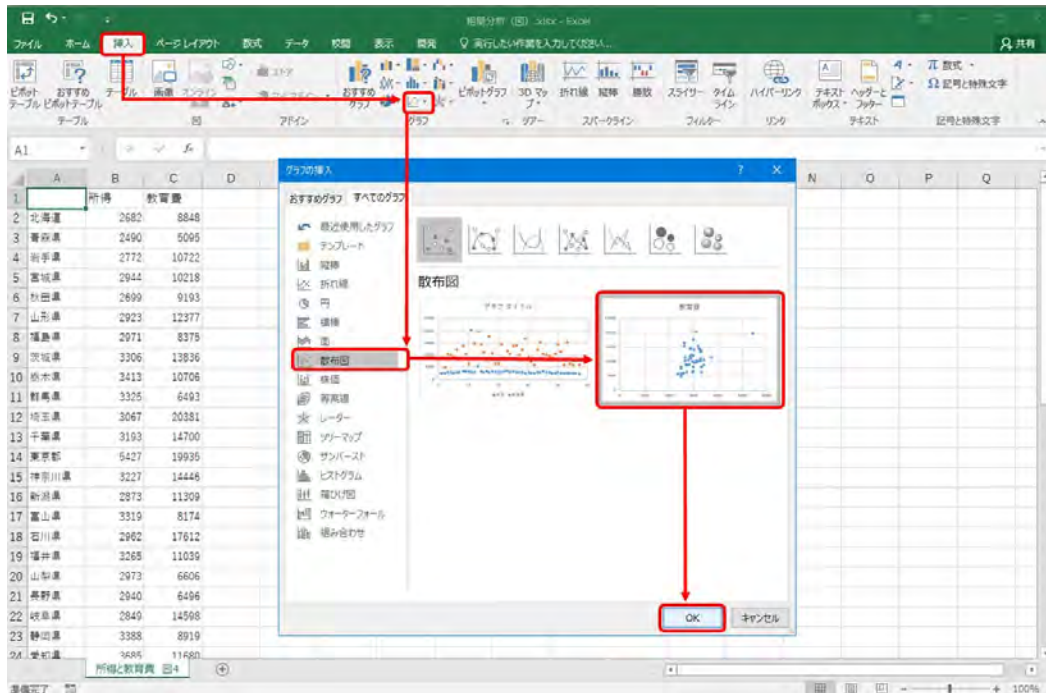
図4 一人当たりの県民所得と教育費の関係

Excel で行う相関分析

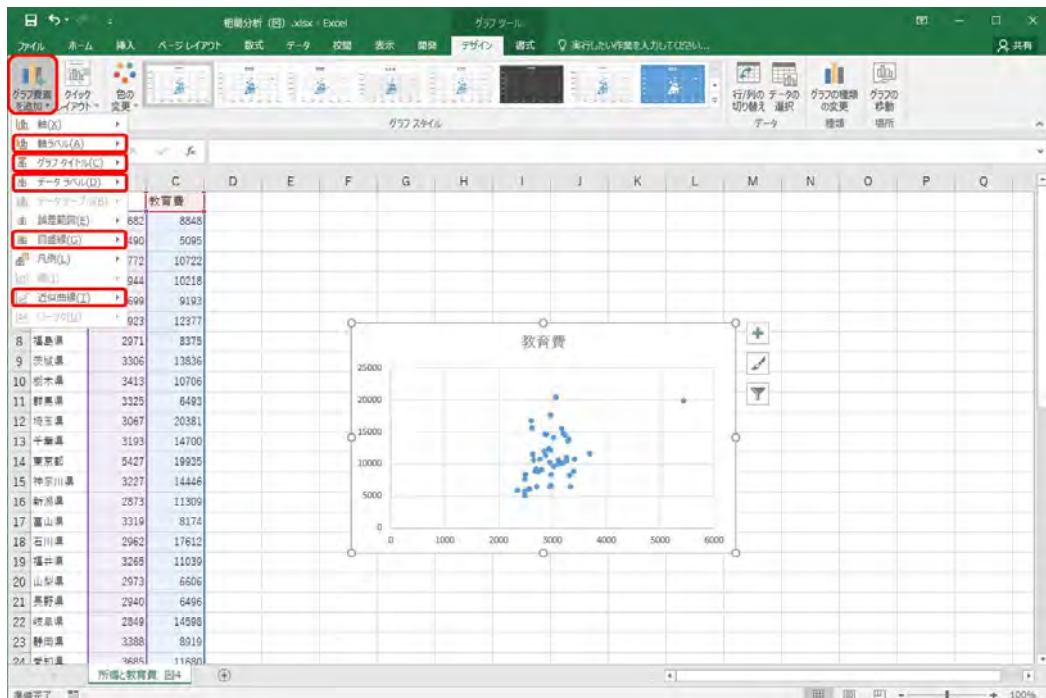
散布図を描く

散布図は2変数を縦軸と横軸に設定し、データを点でプロットしたグラフのことです。

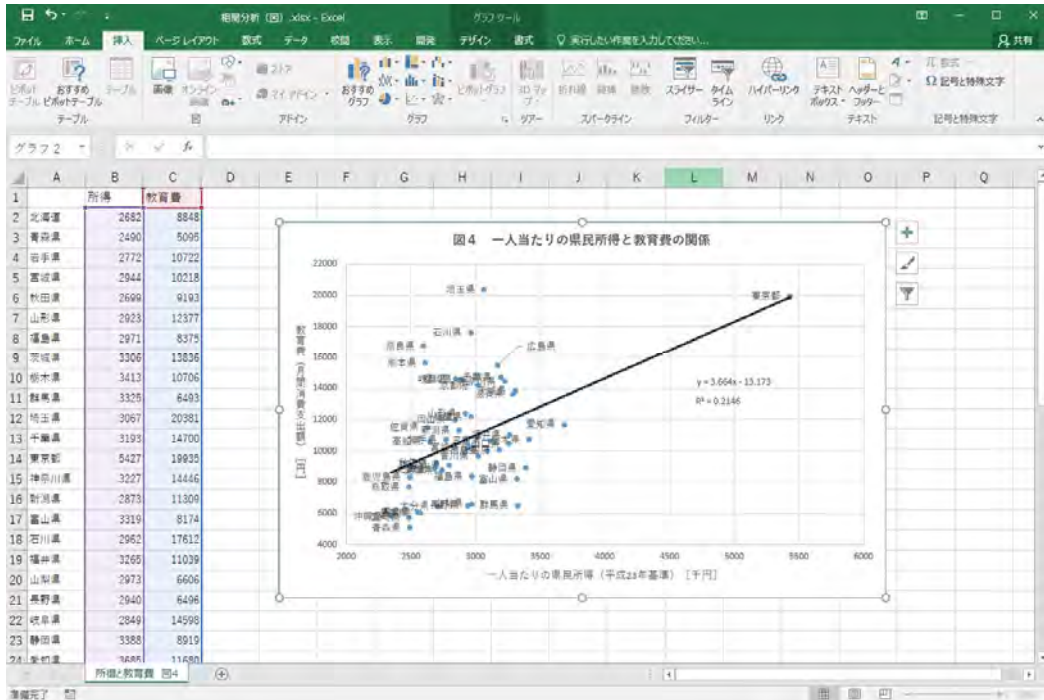
散布図を描くことで、データにどのような関係性があるか視覚的に確認することができます。データを入力した Excel シート内の関係性を調べたい2変数を選択し、「挿入」タブ内の「グラフ」メニューから「散布図 (X,Y) またはバブルチャートの挿入」⇒「散布図」を選択します。



シート内にグラフ（散布図）が挿入されます。



挿入されたグラフはデフォルトの状態ですので、「グラフ要素を追加」メニューから、「グラフタイトル」、「データラベル」、「軸ラベル」、「近似曲線」（回帰直線）等を挿入するなどしてグラフを見やすく整えます。



相関係数を算出する

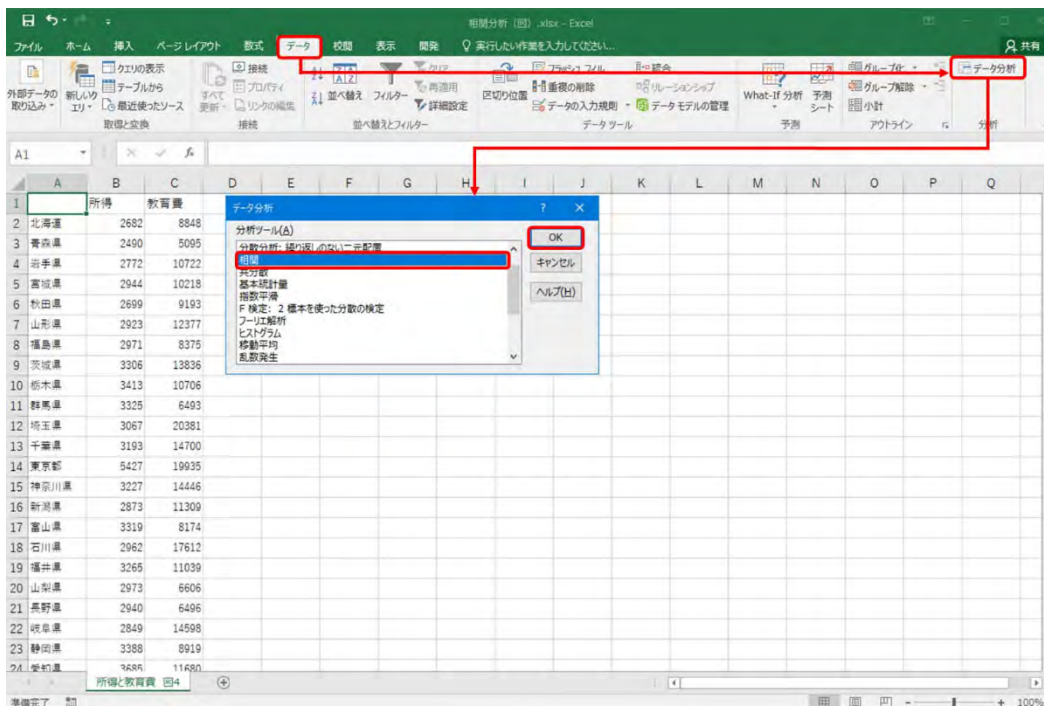
散布図を見ることでデータの関係性を視覚的に確認することができますが、関係性の解釈は見る人によって異なるかもしれません。そこで、データの関係性を客観的に判断するために「相関係数」という数値で表します。

「データ分析」ツールを用いる方法

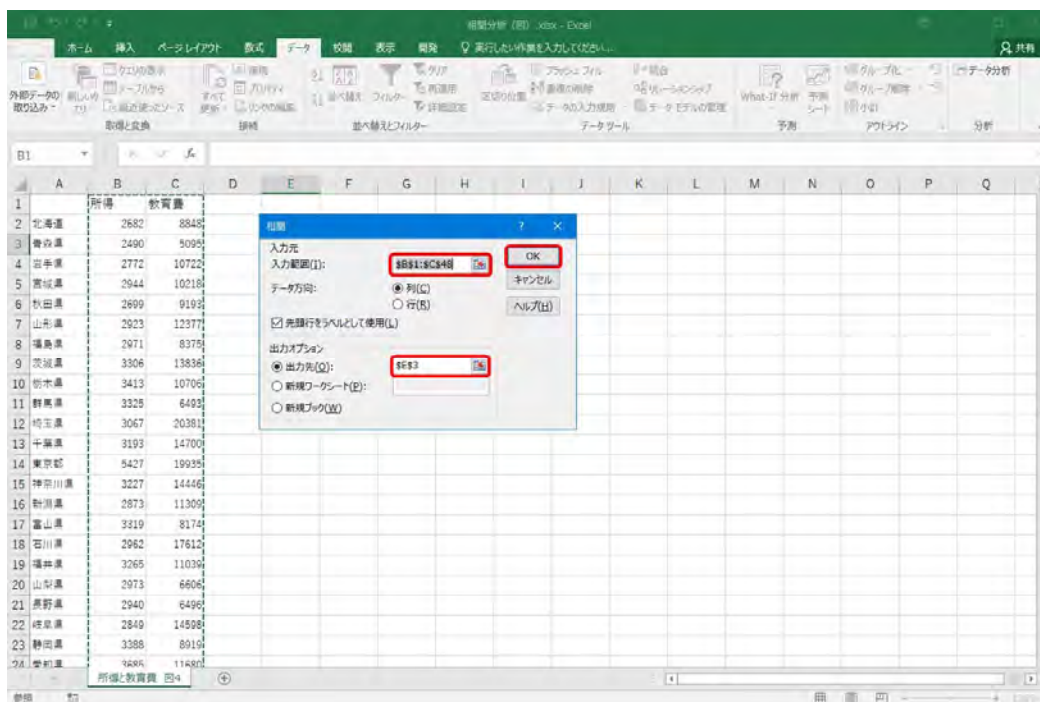
Excelで相関係数を算出する方法はいくつかありますが、ここでは「データ分析」ツールを用いる方法を紹介します。

「データ」タブ内の「データ分析」メニューを選択します。

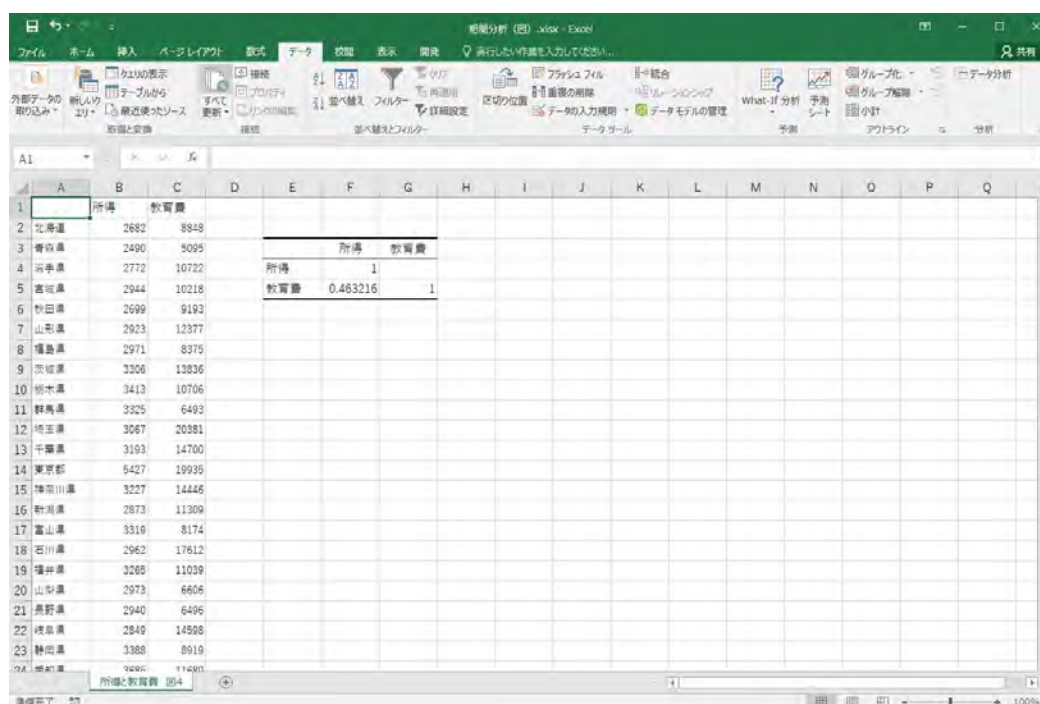
分析ツールの中から「相関」を選択します。



入力範囲のデータ列、出力先のセルを選択し OK ボタンを押します。



相関係数が算出された相関表がシート内に作成されます。



相関係数は $-1 \sim 1$ の値をとり、 -1 に近いほど負の相関があり、 1 に近いほど正の相関があります。

算出した相関係数は、 0.4632 でやや強い正の相関がみられましたが、散布図で描画した回帰直線の当てはまりの良さを示す決定係数 (R^2 : 相関係数の二乗で $0 \sim 1$ の値をとり、値が高いほど回帰直線の説明力があることを意味する。) は 0.2146 とそれほど高くないという結果となりました。

SSDSE を用いた相関分析の一例を紹介しました。

統計データを使った分析では、データ間に相関関係があるかどうか調べる機会も多く存在します。

SSDSE をダウンロード (<https://www.nstac.go.jp/use/literacy/ssdse/>) して、身近なデータの中に存在する関係性を明らかにしてみたいかどうかでしょうか。

Excel で行う相関分析 (SSDSE を用いた分析事例)

公開：令和5年3月22日

作成：独立行政法人 統計センター <https://www.nstac.go.jp/>

お問い合わせ先：統計データ分析コンペティション担当

〒162-8668 東京都新宿区若松町19-1

統計センター 統計技術・提供部 技術研究開発課

statcompe_atmark_nstac.go.jp (「_atmark_」を「@」に置き換えてください)