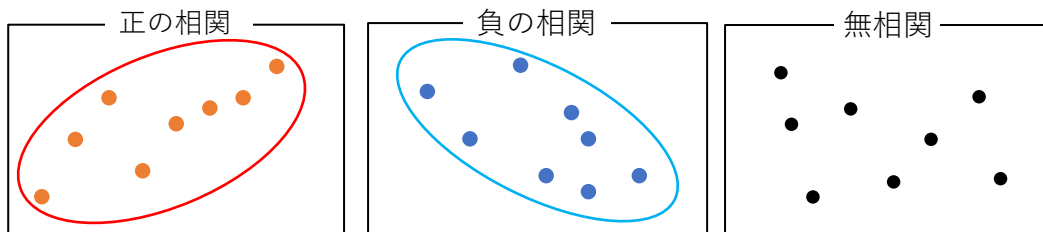


### データの関係性を確認したい

日頃の業務で「この数値を改善したいがどんな要素が影響しているのか」、「この2つのデータの関連性を数値的な根拠をもって説明したい」という場面があると思います。そんな時に使える簡単な分析手法が相関分析です。

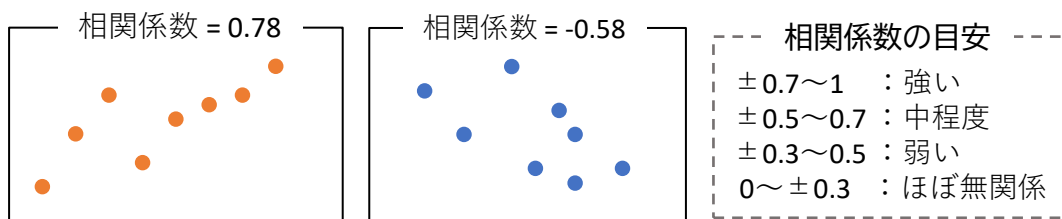
## 01 相関分析とは

片方の要素の変化にあわせてもう片方の要素も変化しているように見える関係を相関関係といいます。相関関係を見るには、まず散布図を描いてみましょう。



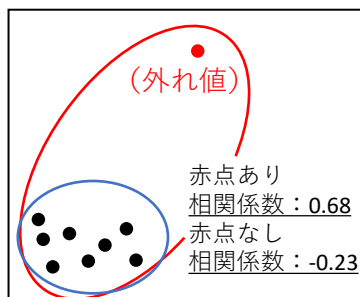
左図のようにデータの塊が右上がりになっている場合を「正の相関」、中央図のように右下がりになっている場合を「負の相関」、右図の全体的にばらばらな場合を「無相関」と言います。

また、このような見た目での判断だけでなく、どれくらい相関があるのかについて確認したい場合は、相関係数を求めてみましょう。相関係数は ExcelのCORREL関数やデータ分析ツールを用いることで簡単に求めることができます。



相関係数が正であれば正の相関、負であれば負の相関であることを示し、係数が1に近いほど強い相関、0に近いほど弱い相関であることを示します。

### 注意



直線的な関係でない場合、何か関係がありそうでも相関係数は0

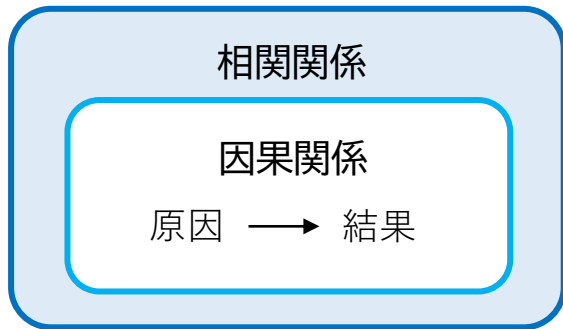


- ・外れ値に影響されやすい
- ・直線以外の関係性は正しく判定できない
- ・相関関係=因果関係ではない

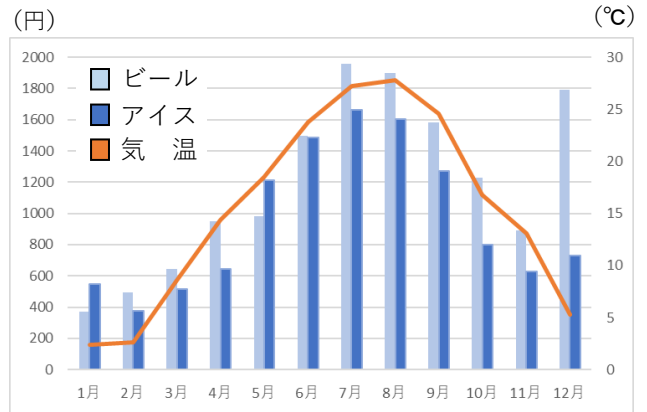


## 02 相関関係と因果関係

相関関係は2つの要素に何らかの関係があることを示していますが、因果関係は片方が原因でもう片方がその結果という関係です。相関関係と因果関係を混同すると、データについて間違っただけ解釈をしてしまうことになってしまいますので注意が必要です。



例：「ビールの支出額」と「アイスの支出額」に相関関係が見られますが、「ビールを飲むからアイスを食べる量が増える」という因果関係は通常考えられません（逆も）。ビールにもアイスにも影響する「気温」という他の要素（**交絡因子：こわらくいんし**）の存在を疑う必要があります。



福井市の2人以上世帯におけるビール・アイスの支出額と月別年平均気温(令和2年「家計調査」・気象庁データ)

## 03 因果関係を確認するには

因果関係を明らかにしていくには、事象ごとの関係性を正しく認識する必要があります。そこで、交絡因子の影響を抑え、2つの事象の関係性を正しく捉えるための代表的な方法を紹介します。

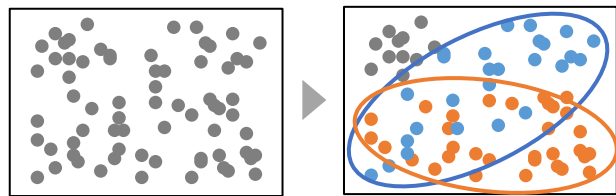
### ランダム化比較実験

調査対象を無作為に2つ以上のグループに振り分け、ある政策手段の対象とするグループ（介入群）と対象としないグループ（比較対象群）間の比較を行い、政策効果の分析を行います。対象者を無作為にグループに割り当てることで交絡因子による影響が抑えられたデータを収集することができます。

### 層別分析

収集したデータをグループ分けし、グループごとに分析を行います。交絡因子と思われる条件（年齢、性別、所得等）でグループ分けを行うと、交絡因子の影響を抑えることができます。

グループごとにデータを見ることで相関を発見！



このような統計学的なアプローチのほか、データの背景にある分野の常識・知識を理解しておくこと、インタビューなどによってインサイト（意識に表れていない本音）を探るなど、より精度の高い仮説を立ててデータの因果関係を判断することも必要です。

今回、相関関係・因果関係について簡単に触れましたが、実際に分析するには隠れた要素を考慮することが必要であり、因果関係を証明するのは難しい作業になります。

統計調査課では、分析手法や統計データの取扱いなどデータ分析を行う上で欠かせないスキルを学べる研修を実施・案内しています。また、データ分析や独自の調査の実施など、統計調査に関する相談窓口も設けています。お気軽にお問い合わせください。

