

【分析課題】課題にある売上データとクレームデータを使って、2つのデータの一部に負の相関関係があるかどうかを調べる。
EXCELを使い、以下の手順で2つのデータの相関関係を調べて下さい。

1. 分析目標の設定

クレームがある一定量を超えると売上高に影響が出てくるという仮説を検証する。

使用するデータは、以下の通り。

利用データ1 (売上データ)		
売上日	商品分類	売上額
利用データ2 (クレームデータ)		
受付日	商品分類	件数

※「データ分析の基礎用演習データ.xlsx」にデータが入っています。

【分析の進め方】

①まず、商品分類別にクレーム件数の偏差値を計算する。(2015年3月のテレビのクレーム偏差値は55。など)

②この偏差値と売上が負の相関関係が大きい商品を抽出する。

※売上とクレームの偏差値を使って、相関関係を単純に導くために、「相関売上」という概念を導入する。

相関売上とは： $(売上 \times クレーム偏差値 \times 2 \div 100)$ で算出した値。平均的なクレームの場合、売上額と等しくなる係数。

③上記計算結果を組み合わせてグラフで表現すると下記のようなになる。



2. 分析用データモデル (売上クレーム相関モデル) の設定

Excelでの分析方法をデータモデルを使って検討します。

方法は一つではありませんが、今回は、以下のデータモデルを設定し、

月単位の売上とクレーム偏差値を使って分析を実施します。

- ①

年月	商品分類	売上合計
----	------	------

 月単位の売上合計を算出する必要がある。
- ②

年月	商品分類	クレーム偏差値
----	------	---------

 月単位のクレームの偏差値を計算する必要がある。

- ③

年月	商品分類	売上額	平均売上	相関売上
----	------	-----	------	------

 2行を1行に統合する。

$$\text{相関売上} = \text{売上} \times \text{クレーム偏差値} \times 2 \div 100$$
(平均的なクレーム数の場合、≒売上となる)

◆偏差値を使う理由

月別の集計を行った場合、クレームは合計、平均、MIN、MAX、偏差値などが利用可能だが、
合計、平均、MIN、MAXは、商品の種類によって値のバラつきが出てしまうため、標準化の観点より偏差値を利用する。

3. 実際にExcelでデータを加工する手順は以下の通りです。

①日別の商品分類別売上額を月別に集計します。Excelのピボットテーブルを使用すると便利です。

年月	商品分類	売上額	平均売上
----	------	-----	------

※売上平均はグラフ化するときに必要となる。

②日別の商品分類別クレーム数を月別に集計します。

年月	商品分類	クレーム合計	クレーム平均値
----	------	--------	---------

※クレーム平均値は偏差値を計算するのに必要となる。

③上の②のデータを利用して月別にクレームの標準偏差を算出します。

年月	エアコン	標準偏差
----	------	------

※ここからは、商品分類を「エアコン」に絞って実施して下さい。

④月別クレームデータを結合 (②+③) して、クレーム偏差値を算出します。結合は、コピペで可能です。

年月	エアコン	クレーム合計	クレーム平均値	標準偏差	クレーム偏差値
----	------	--------	---------	------	---------

※ $\text{クレーム偏差値} = (\text{クレーム合計} - \text{平均値}) \div \text{標準偏差} \times 10 + 50$

⑤月別の売上情報とクレーム情報を結合 (①+④) して、相関売上を算出します。

年月	エアコン	売上額	平均売上	クレーム偏差値	相関売上
----	------	-----	------	---------	------

※ $\text{相関売上} = \text{売上} \times \text{クレーム偏差値} \times 2 \div 100$

4. Excelでグラフ化

最初に複合グラフを選択した後、売上 (棒グラフ)、平均 (折れ線グラフ)、相関売上 (棒グラフ) を設定します。