

社会人のためのデータサイエンス演習

第3週：課題の補講 Excel分析演習

講師名：橋本 武彦

第3週の内容紹介

| | |
|---|-----------------|
| 1 | ● 第3週のまとめ |
| 2 | ● ピボットテーブルによる集計 |
| 3 | ● 散布図の描画 |
| 4 | ● 相関係数の算出 |
| 5 | ● 移動平均の算出 |

※Microsoft®, Microsoft® Office Excel® は、米国 Microsoft Corporation の、米国およびその他の国における登録商標または商標です。

※本資料は、Microsoft Corporation と提携しているものではなく、また、Microsoft Corporation が許諾、後援、その他の承認をするものではありません。

※本資料の本文では、©、®、™などの表記は割愛いたします。

※本資料ではデータ分析ツールとして Microsoft® Office Excel® 2013 を利用しています。

Microsoft® Office の他のバージョンや他の分析ツールを利用している場合は、ヘルプやインターネットなどで各自で調査し、該当機能に置き換えて参照してください。

1.第3週のまとめ

比較と傾向の分析

クロス集計 / 散布図 / 相関 / 時系列

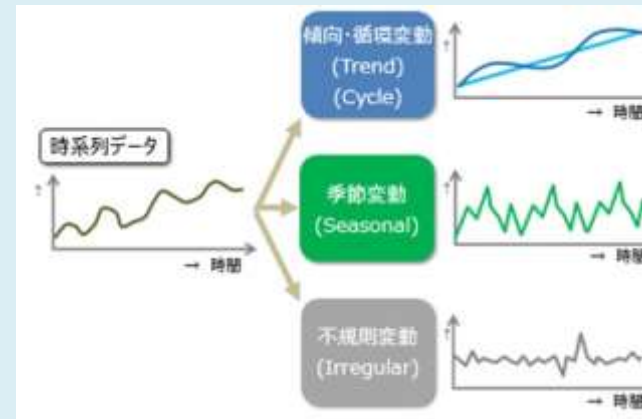
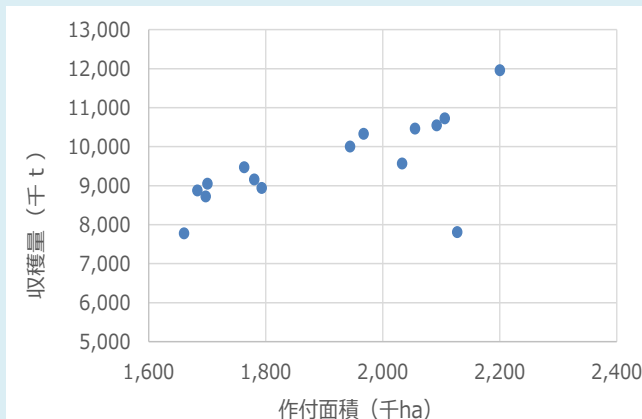
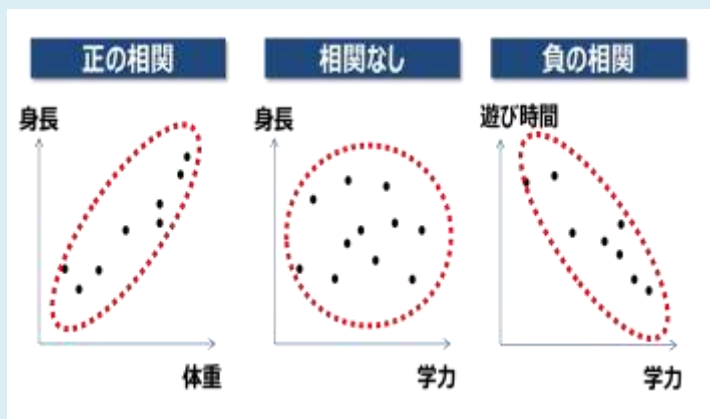
| 上段：n 下段：% | 先月 | 今月 | 計 |
|--------------|-----------------|-----------------|------------------|
| ファミリー層 | 200,000円 67% | 100,000円 33% | 300,000円 100% |
| 単身層 | 180,000円 50% | 180,000円 50% | 360,000円 100% |
| 計 | 380,000円 58% | 280,000円 42% | 660,000円 100% |

・可視化が重要

- ✓ 変数の関係性把握
- ✓ 外れ値の除外

・時系列は分解する

- ✓ 解釈が容易に
- ✓ 将来予測が可能



2.ピボットテーブルによる集計

- 下表のような売上を示すデータがあります。
- クロス集計を行い、4月と5月で、ファミリーと単身者の売上に違いがあるか観察します。

| | A | B | C |
|----|-------|-----|---------|
| 1 | 家族構成 | 売上月 | 売上 |
| 2 | ファミリー | 4月 | ¥15,000 |
| 3 | ファミリー | 4月 | ¥20,000 |
| 4 | 単身 | 4月 | ¥12,000 |
| 5 | ファミリー | 4月 | ¥18,000 |
| 6 | 単身 | 4月 | ¥9,000 |
| 7 | 単身 | 4月 | ¥5,000 |
| 8 | 単身 | 4月 | ¥10,000 |
| 9 | ファミリー | 4月 | ¥17,000 |
| 10 | 単身 | 4月 | ¥15,000 |
| 11 | ファミリー | 5月 | ¥12,000 |
| 12 | 単身 | 5月 | ¥15,000 |
| 13 | ファミリー | 5月 | ¥13,000 |
| 14 | ファミリー | 5月 | ¥12,000 |
| 15 | 単身 | 5月 | ¥16,000 |

2.ピボットテーブルによる集計

1)Excelのピボットテーブルという機能を使って、表頭に売上月、表側に家族構成を表示する集計表を作成

①挿入を選択

②集計データを選択

③ピボットテーブルをクリック

④「新規ワークシート」を選択

⑤OKをクリック

| | 家族構成 | 売上月 | 売上 |
|----|-------|-----|---------|
| 1 | ファミリー | 4月 | ¥15,000 |
| 2 | ファミリー | 4月 | ¥20,000 |
| 3 | 単身 | 4月 | ¥12,000 |
| 4 | ファミリー | 4月 | ¥18,000 |
| 5 | 単身 | 4月 | ¥9,000 |
| 6 | 単身 | 4月 | ¥5,000 |
| 7 | 単身 | 4月 | ¥10,000 |
| 8 | ファミリー | 4月 | ¥17,000 |
| 9 | 単身 | 4月 | ¥15,000 |
| 10 | ファミリー | 5月 | ¥12,000 |
| 11 | 単身 | 5月 | ¥15,000 |
| 12 | ファミリー | 5月 | ¥13,000 |
| 13 | ファミリー | 5月 | ¥12,000 |
| 14 | 単身 | 5月 | ¥16,000 |

ピボットテーブルの作成

分析するデータを選択してください。

☒ テーブルまたは範囲を選択(S)

テーブル/範囲(I): pivotテーブルによる集計!\$A\$1:\$C\$15

☐ 外部データソースを使用(U)

接続の選択(C)...

接続名:

ピボットテーブルレポートを配置する場所

☒ 新規ワークシート(N)

☐ 既存のワークシート(E)

場所(L): pivotテーブルによる集計!\$G\$31

複数のテーブルを分析するかどうかを選択

☐ このデータをデータモデルに追加する(M)

OK キャンセル

2.ピボットテーブルによる集計

3) 集計軸を設定し、各セルの合計値を算出

Excelの右側に表示されるピボットテーブルのフィールドから、上部フィールドをドラッグして下部に追加し、以下のように設定

The screenshot shows the 'PivotTable Fields' task pane with the following fields: 家族構成 (checked), 売上月 (checked), and 売上 (checked). The pane is divided into sections: 'レポートに追加するフィールドを選択してください:' (containing the checked fields), 'その他のテーブル...', and '次のボックス間でフィールドをドラッグしてください:' (containing 'フィルター', '列', and '行' sections). Red arrows and callouts explain the setup:

- ⑥** 上部の「家族構成」を「行」に追加 ⇒ クロス表の表側になる (An arrow points from '家族構成' to the '行' section.)
- ⑦** 上部の「売上月」を「列」に追加 ⇒ 表頭になる (An arrow points from '売上月' to the '列' section.)
- ⑧** 上部の「売上」を「値」に追加 ⇒ 合計値を算出 (An arrow points from '売上' to the '値' section.)

The '行' section contains '家族構成'. The '列' section contains '売上月'. The '値' section contains '合計 / 売上'.

2.ピボットテーブルによる集計

4)続いて、売上の値フィールドから計算の種類を選択して割合が表示されるように設定

⑨さらに「売上」を「値」に追加
⇒ 割合を算出

⑩下部の列の「値」を行に移動

⑪「合計/売上2」の▼をクリックして表示されるリストから「値フィールドの設定」を選択

⑫名前を「割合 / 売上」に変更する

⑬「計算の種類」を選択

⑭行方向の割合を算出する場合、「行集計に対する比率」を選択

⑮OKをクリック

値フィールドの設定

ソース名: 売上

名前の指定(C): 割合 / 売上

集計方法 計算の種類

計算の種類(A)

行集計に対する比率
計算なし
総計に対する比率
列集計に対する比率
行集計に対する比率
基準値に対する比率
親行集計に対する比率

表示 OK キャンセル

2.ピボットテーブルによる集計

5) クロス集計表が出力

| 行ラベル | 列ラベル | 4月 | 5月 | 総計 |
|-------------|------|--------|--------|---------|
| ファミリー | | | | |
| 合計 / 売上 | | 70000 | 37000 | 107000 |
| 割合 / 売上 | | 65.42% | 34.58% | 100.00% |
| 単身 | | | | |
| 合計 / 売上 | | 51000 | 31000 | 82000 |
| 割合 / 売上 | | 62.20% | 37.80% | 100.00% |
| 全体の 合計 / 売上 | | 121000 | 68000 | 189000 |
| 全体の 割合 / 売上 | | 64.02% | 35.98% | 100.00% |

- 全体の売上では5月に低下
- ファミリーと単身の総計では単身の方が売上が低い
- ファミリー、単身ともに、ほぼ同様の割合で5月の売上が低下

3. 散布図の描画

- 水稻の作付面積と収穫量の関係を、散布図を描いて観察してみましょう。

1) Excelで、右のような表を作成。
X軸に示したい項目のデータをB列、Y軸に示したい項目のデータをC列に設定

| | A | B | C | D | |
|----|-------|-----------|---------|---|--|
| 1 | 年次 | 作付面積（千ha） | 収穫量（千t） | | |
| 2 | 1990年 | 2,055 | 10,463 | | |
| 3 | 1991年 | 2,033 | 9,565 | | |
| 4 | 1992年 | 2,092 | 10,546 | | |
| 5 | 1993年 | 2,127 | 7,811 | | |
| 6 | 1994年 | 2,200 | 11,961 | | |
| 7 | 1995年 | 2,106 | 10,724 | | |
| 8 | 1996年 | 1,967 | 10,328 | | |
| 9 | 1997年 | 1,944 | 10,004 | | |
| 10 | 1998年 | 1,793 | 8,939 | | |
| 11 | 1999年 | 1,780 | 9,159 | | |
| 12 | 2000年 | 1,763 | 9,472 | | |
| 13 | 2001年 | 1,700 | 9,048 | | |
| 14 | 2002年 | 1,683 | 8,876 | | |
| 15 | 2003年 | 1,660 | 7,779 | | |
| 16 | 2004年 | 1,697 | 8,721 | | |
| 17 | | | | | |
| 18 | | | | | |
| .. | | | | | |

出典：農林水産省 作物統計より作成
http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001024932&cycleCode=0&requestSender=dsearch

3. 散布図の描画

2) ツールバーの「挿入」をクリックし、データの範囲となるセルB2からセルC16までを選択して、グラフツールの散布図を選択

②「挿入」をクリック

③ 散布図をクリック

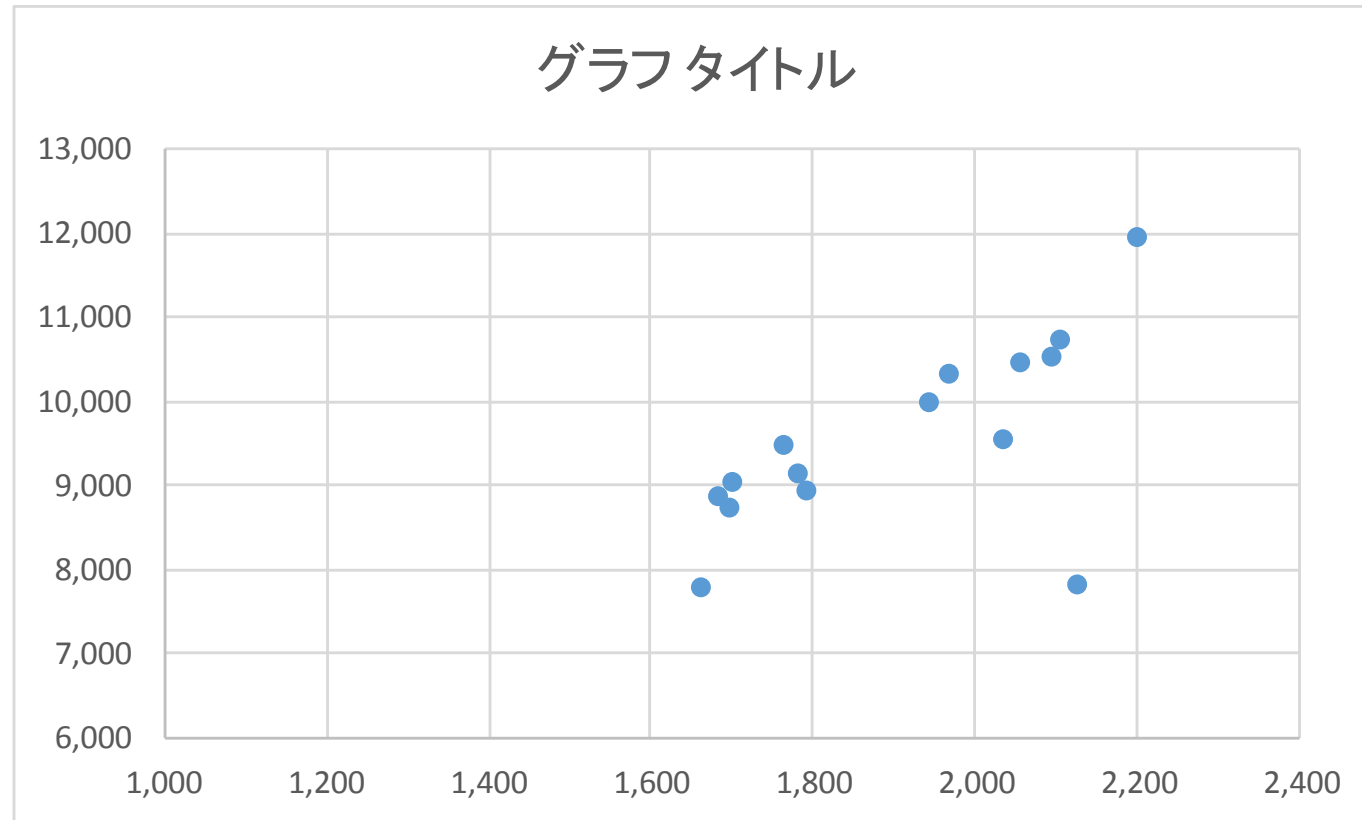
④ 左上の図形を選択

① データの範囲 (B2 から C16) を選択

| 年次 | 作付面積 (千ha) | 収穫量 (千t) |
|-------|------------|----------|
| 1990年 | 2,055 | 10,463 |
| 1991年 | 2,033 | 9,565 |
| 1992年 | 2,092 | 10,546 |
| 1993年 | 2,127 | 7,811 |
| 1994年 | 2,200 | 11,961 |
| 1995年 | 2,106 | 10,724 |
| 1996年 | 1,967 | 10,328 |
| 1997年 | 1,944 | 10,004 |
| 1998年 | 1,793 | 8,939 |
| 1999年 | 1,780 | 9,159 |
| 2000年 | 1,763 | 9,472 |
| 2001年 | 1,700 | 9,048 |
| 2002年 | 1,683 | 8,876 |
| 2003年 | 1,660 | 7,779 |
| 2004年 | 1,697 | 8,721 |

3. 散布図の描画

3) グラフ描画



※上図は結果が見やすいように軸の目盛りを変更している

- 右肩上がりの形状 ⇒ 正の相関
- 外れ値が存在

4.相関係数の算出

- 次に、相関係数を算出してみましょう。

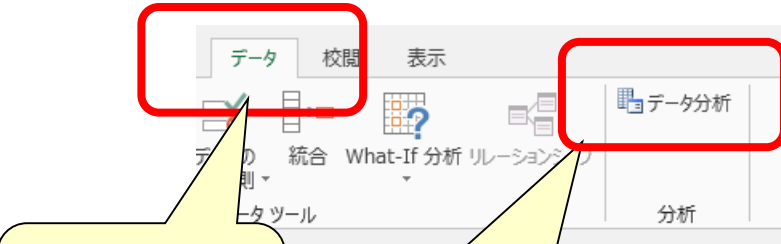
1)散布図を描いたデータと同じデータを用意

| | A | B | C | D | |
|----|-------|-----------|---------|---|--|
| 1 | 年次 | 作付面積（千ha） | 収穫量（千t） | | |
| 2 | 1990年 | 2,055 | 10,463 | | |
| 3 | 1991年 | 2,033 | 9,565 | | |
| 4 | 1992年 | 2,092 | 10,546 | | |
| 5 | 1993年 | 2,127 | 7,811 | | |
| 6 | 1994年 | 2,200 | 11,961 | | |
| 7 | 1995年 | 2,106 | 10,724 | | |
| 8 | 1996年 | 1,967 | 10,328 | | |
| 9 | 1997年 | 1,944 | 10,004 | | |
| 10 | 1998年 | 1,793 | 8,939 | | |
| 11 | 1999年 | 1,780 | 9,159 | | |
| 12 | 2000年 | 1,763 | 9,472 | | |
| 13 | 2001年 | 1,700 | 9,048 | | |
| 14 | 2002年 | 1,683 | 8,876 | | |
| 15 | 2003年 | 1,660 | 7,779 | | |
| 16 | 2004年 | 1,697 | 8,721 | | |
| 17 | | | | | |
| 18 | | | | | |
| .. | | | | | |

出典：農林水産省 作物統計より作成
http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001024932&cycleCode=0&requestSender=dsearch

4.相関係数の算出

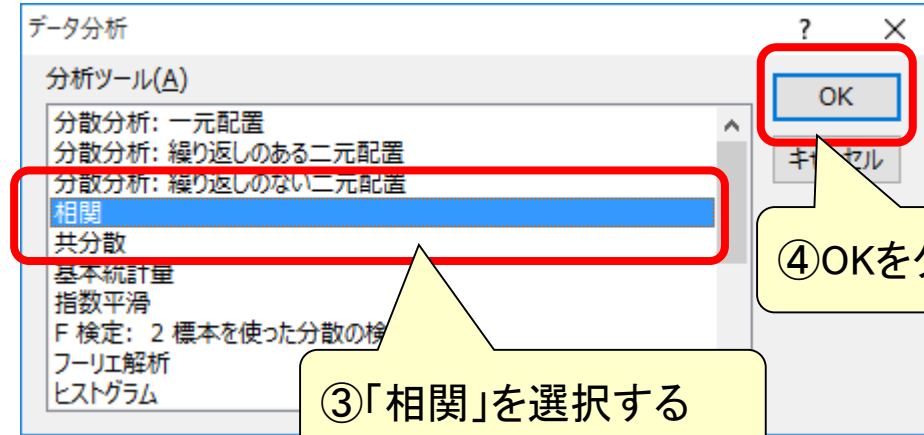
1)データ分析ツールを起動



①「データ」をクリック

②「データ分析」をクリック

2)ダイアログが表示されたら[相関]を選択

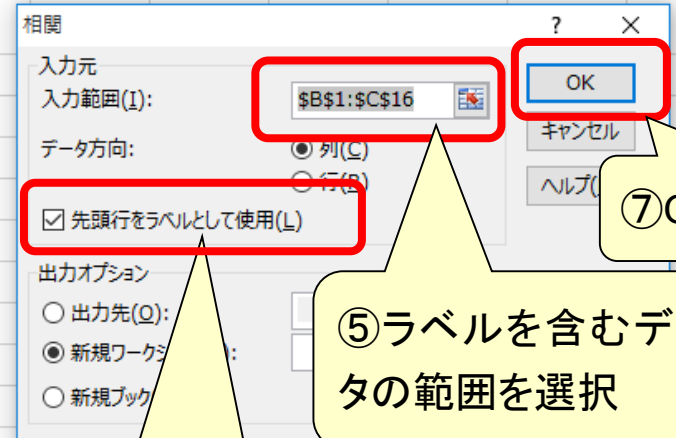


③「相関」を選択する

④OKをクリック

3)[入力範囲]にデータ範囲であるセルB1~C16を選択

| | A | B | C | D | E | F | G | H | I | J |
|----|-------|---------------|-------------|---|---|---|---|---|---|---|
| | 年次 | 作付面積 (千ha) | 収穫量 (千t) | | | | | | | |
| 1 | | | | | | | | | | |
| 2 | 1990年 | 2,055 | 10,463 | | | | | | | |
| 3 | 1991年 | 2,033 | 9,565 | | | | | | | |
| 4 | 1992年 | 2,092 | 10,546 | | | | | | | |
| 5 | 1993年 | 2,127 | 7,811 | | | | | | | |
| 6 | 1994年 | 2,200 | 11,961 | | | | | | | |
| 7 | 1995年 | 2,106 | 10,724 | | | | | | | |
| 8 | 1996年 | 1,967 | 10,328 | | | | | | | |
| 9 | 1997年 | 1,944 | 10,004 | | | | | | | |
| 10 | 1998年 | 1,793 | 8,939 | | | | | | | |
| 11 | 1999年 | 1,780 | 9,159 | | | | | | | |
| 12 | 2000年 | 1,763 | 9,472 | | | | | | | |
| 13 | 2001年 | 1,700 | 9,048 | | | | | | | |
| 14 | 2002年 | 1,683 | 8,876 | | | | | | | |
| 15 | 2003年 | 1,660 | 7,779 | | | | | | | |
| 16 | 2004年 | 1,697 | 8,721 | | | | | | | |
| 17 | | | | | | | | | | |



⑤ラベルを含むデータの範囲を選択

⑥「先頭行をラベルとして使用」にチェック

⑦OKをクリック

4.相関係数の算出

4)相関係数が出力

| | A | B | C |
|----|-----------|-------------|---------|
| 1 | | 作付面積(千ha) | 収穫量(千t) |
| 2 | 作付面積(千ha) | 1 | |
| 3 | 収穫量(千t) | 0.648739009 | 1 |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |
| 11 | | | |
| 12 | | | |

- $r=0.65$ 程度
⇒中程度の正の相関

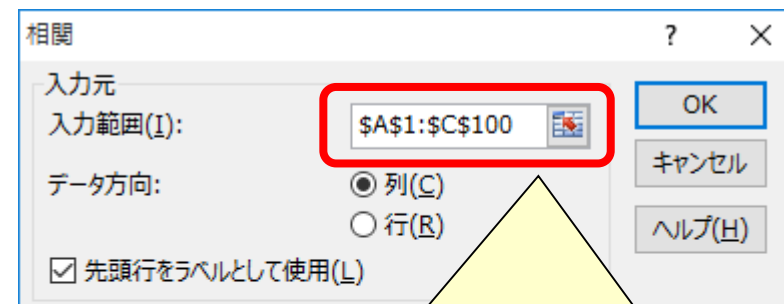
【補足1】

分析ツールアドインではなく、関数で算出することも可能

=CORREL(変量1、変量2)

【補足2】

分析ツールアドインでは、3つ以上の変数を対象に、それぞれの相関係数を一度に算出可能



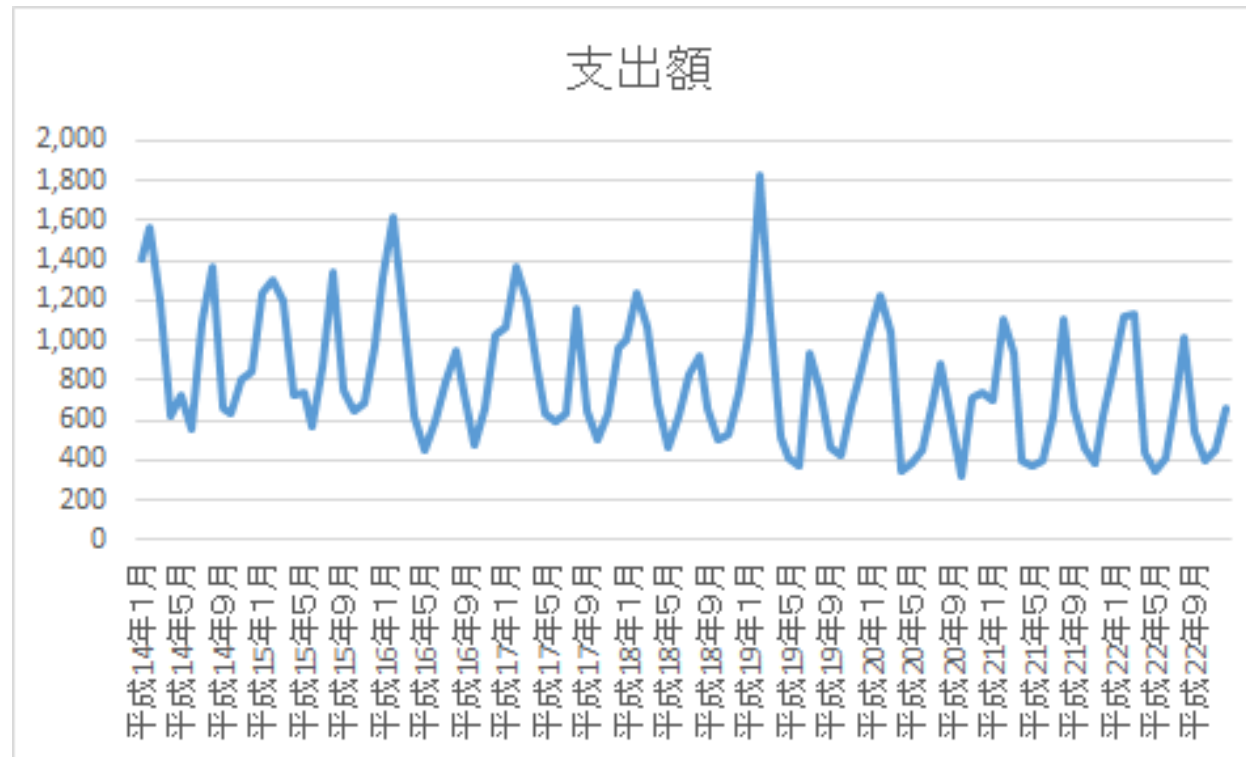
A列からC列までを指定している。
この場合、A列とB列、A列とC列、B列とC列の相関係数が一度に算出される。

5.移動平均の算出

- 下の図表は自動車教習料の支出額を時系列に示したものです。12ヶ月移動平均値を算出して季節変動を除いたグラフを作成してみましょう。

| 年月 | 支出額 |
|----------|-------|
| 平成14年1月 | 1,405 |
| 平成14年2月 | 1,563 |
| 平成14年3月 | 1,198 |
| 平成14年4月 | 617 |
| 平成14年5月 | 720 |
| 平成14年6月 | 560 |
| 平成14年7月 | 1,091 |
| 平成14年8月 | 1,369 |
| 平成14年9月 | 655 |
| 平成14年10月 | 631 |

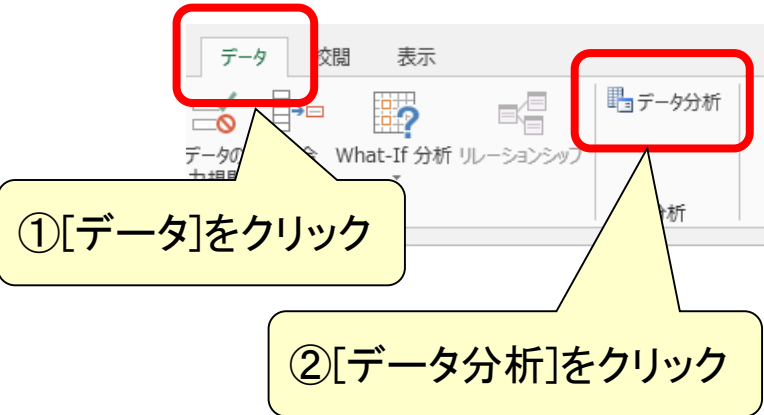
| | |
|----------|-------|
| 平成22年2月 | 1,114 |
| 平成22年3月 | 1,136 |
| 平成22年4月 | 438 |
| 平成22年5月 | 344 |
| 平成22年6月 | 410 |
| 平成22年7月 | 724 |
| 平成22年8月 | 1,016 |
| 平成22年9月 | 543 |
| 平成22年10月 | 395 |
| 平成22年11月 | 450 |
| 平成22年12月 | 659 |



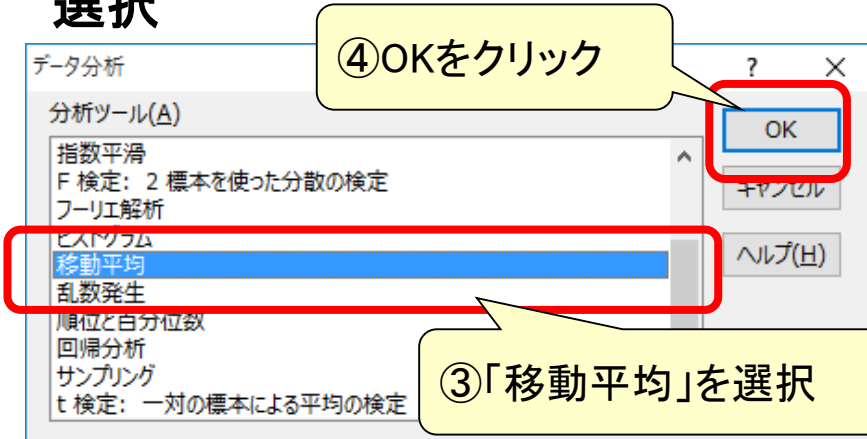
出典:総務省 家計消費状況調査(二人以上の世帯)より作成
<http://www.stat.go.jp/data/joukyou/12.htm>

5.移動平均の算出

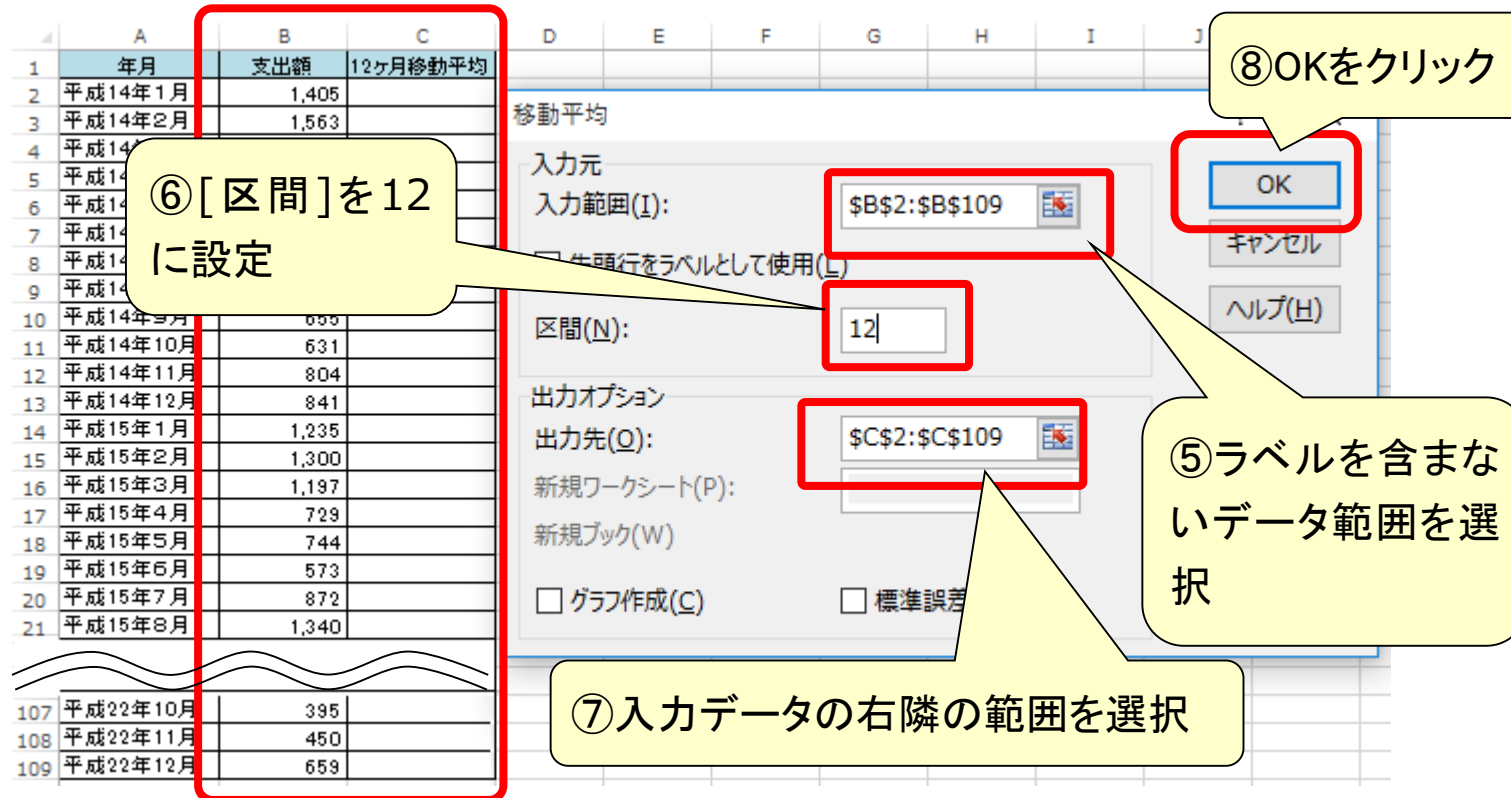
1)データ分析ツールを起動



2)ダイアログが表示されたら[移動平均] を 選択



3)[入力範囲] に 支出額の列を指定し、[出力先] に C列を指定する。 また区間を12に 設定



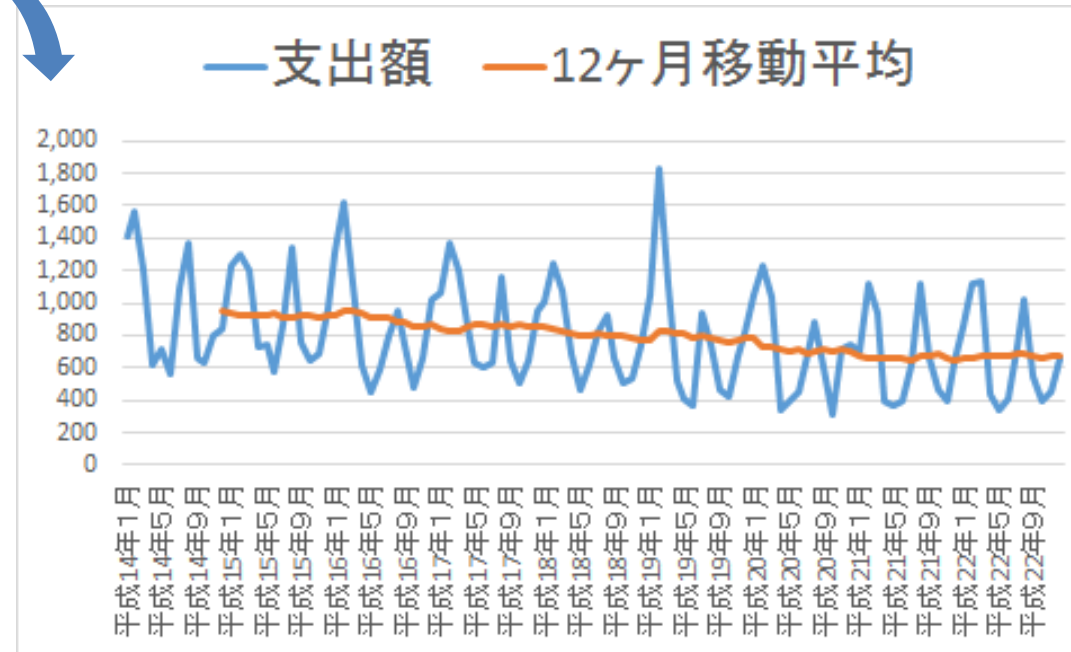
5.移動平均の算出

- 4)12か月移動平均が算出される。 ※分析ツールで求められるのは後方移動平均。
※各月とそれ以前の月の平均値を使って算出する移動平均を「後方移動平均」という。

| | A | B | C |
|-------|----------|-------|----------|
| 1 | 年月 | 支出額 | 12ヶ月移動平均 |
| 2 | 平成14年1月 | 1,405 | |
| 3 | 平成14年2月 | 1,563 | |
| 4 | 平成14年3月 | 1,198 | |
| 5 | 平成14年4月 | 617 | |
| 6 | 平成14年5月 | 720 | |
| 7 | 平成14年6月 | 560 | |
| 8 | 平成14年7月 | 1,091 | |
| 9 | 平成14年8月 | 1,369 | |
| 10 | 平成14年9月 | 655 | |
| 11 | 平成14年10月 | 631 | |
| 12 | 平成14年11月 | 824 | |
| 13 | 平成14年12月 | 841 | 955 |
| 14 | 平成15年1月 | 1,235 | 940 |
| 15 | 平成15年2月 | 1,300 | 918 |
| 16 | 平成15年3月 | 1,197 | 918 |
| 17 | 平成15年4月 | 729 | 928 |
| 18 | 平成15年5月 | 744 | 930 |
| 19 | 平成15年6月 | 573 | 931 |
| 20 | 平成15年7月 | 872 | 913 |
| 21 | 平成15年8月 | 1,340 | 910 |
| ~~~~~ | | | |
| 107 | 平成22年10月 | 395 | 666 |
| 108 | 平成22年11月 | 460 | 671 |
| 109 | 平成22年12月 | 659 | 673 |

グラフ化
すると

- 支出額は、減少傾向



5.移動平均の算出

【補足】分析ツールアドインではなく、関数で算出することも可能

| | A | B | C | D |
|----|----------|-------|----------|-------------|
| | 年月 | 支出額 | 12ヶ月移動平均 | 12ヶ月中心化移動平均 |
| 1 | | | | |
| 2 | 平成14年1月 | 1,405 | | |
| 3 | 平成14年2月 | 1,563 | | |
| 4 | 平成14年3月 | 1,198 | | |
| 5 | 平成14年4月 | 617 | | |
| 6 | 平成14年5月 | 720 | | |
| 7 | 平成14年6月 | 560 | ① | ② |
| 8 | 平成14年7月 | 1,091 | | 947 |
| 9 | 平成14年8月 | 1,369 | | 929 |
| 10 | 平成14年9月 | 655 | | 918 |
| 11 | 平成14年10月 | 631 | | 923 |
| 12 | 平成14年11月 | 804 | | 929 |
| 13 | 平成14年12月 | 841 | 955 | 930 |
| 14 | 平成15年1月 | 1,235 | 940 | 922 |
| 15 | 平成15年2月 | 1,300 | 918 | 911 |
| 16 | 平成15年3月 | 1,197 | 918 | 914 |

①12ヶ月移動平均(後方移動平均)

各月における過去12ヶ月間の平均値

C13に、=AVERAGE(B2:B13)

C14に、=AVERAGE(B3:B14)

.....

C109に、=AVERAGE(B98:B109)

②12ヶ月中心化移動平均

各月を中心にした12ヶ月間の平均値。月数は偶数のため、区間を1ヶ月ずらした平均値を2つ求め、両者の平均を取る。

C8に、=(AVERAGE(B2:B13)+
AVERAGE(B3:B14))/2

C9に、=(AVERAGE(B3:B14)+
AVERAGE(B4:B15))/2

.....

C103に、=(AVERAGE(B97:B108)+
AVERAGE(B98:B109))/2

| | | | | |
|-----|----------|-------|-----|-----|
| 100 | | | | |
| 101 | 平成22年4月 | 438 | 680 | 668 |
| 102 | 平成22年5月 | 344 | 678 | 668 |
| 103 | 平成22年6月 | 410 | 679 | 672 |
| 104 | 平成22年7月 | 724 | 688 | |
| 105 | 平成22年8月 | 1,016 | 680 | |
| 106 | 平成22年9月 | 543 | 671 | |
| 107 | 平成22年10月 | 395 | 666 | |
| 108 | 平成22年11月 | 450 | 671 | |
| 109 | 平成22年12月 | 659 | 673 | |