回帰分析（１）

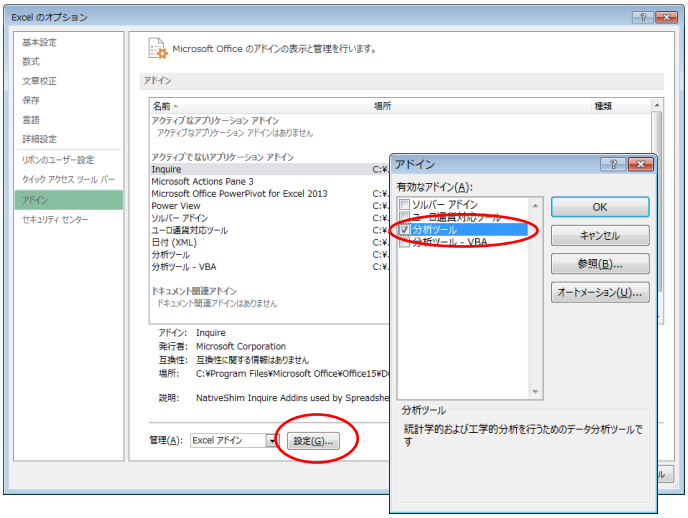
「その日の気温や湿度と、ラーメンの売上は、関係があるか」というような関係性を、「相関関係」といいます。回帰分析は、相関関係があるかないか、数学的に調べる方法の１つ。数学的な「正しさ」とか「根拠」とかはザックリ省いて、相関関係があるのかないのか、エクセルで調べてみましょう。

１．準備

通常、初期設定のままでは、回帰分析が使えるようになっていません。以下の手順で、回帰分析を使えるようにします。

①「ファイル」タブ→「オプション」→「アドイン」項目を開きます。

②「設定」ボタン→「分析ツール」にチェックを入れて、「OK」。



これで、「データ」タブに、「データ分析」というボタンが表示されます。

２．単回帰分析

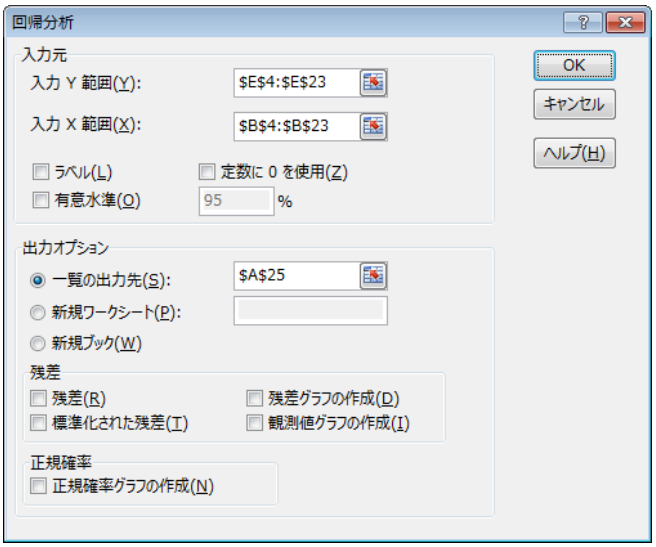
「チキン売上.xlsx」というファイルを、開いて下さい。21 号店を出店するときの参考にするため、「商圏人口・店舗面積・宣伝費と、売上高のあいだに、相関関係があるのかどうか」を、調べてみます。まずは、「商圏人口と売上高の相関関係」を、調べてみましょう。

①「データ」タブ→「データ分析」で、「回帰分析」を選択します。

②「入力 Y 範囲（結果となる方のデータ）」欄に、売上高のデータ E4～E23 セルを指定します。（入力欄をクリック→データ範囲をドラッグ　で入力できます。）

③同様に、「入力 X 範囲（原因と想像している方のデータ）」に、商圏人口のデータ B4～B23 セルを指定します。

④「一覧の出力先」に、A25 セルを指定して、「OK」。



Excel093

右のような結果が、表示されましたか？

「重相関 R」の数値を、見てください。これは「相関係数」と呼ばれるもので、0～1 の間の数値となります。一般には、0.4 以上で「相関関係がある」と言え、0.7 以上だと「強い相関関係がある」となります。

今回の結果は約 0.76 ですから、「商圏人口と売上高には、強い相関関係がある」となります。

概要

回帰統計	
重相関 R	0.758012282
重決定 R2	0.57458262
補正 R2	0.550948321
標準誤差	2403.475525
観測数	20

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	1	140439465.8	140439465.8	24.31138836	0.000107899
残差	18	103980502.8	5776694.598		
合計	19	244419968.6			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	12679.49268	3261.129024	3.888068391	0.001077694	5828.114833	19530.87	5828.115	19530.87
X 値 1	611.162132	123.9514344	4.930658005	0.000107899	350.7498316	871.5744	350.7498	871.5744

［練習問題１］

（１）店舗面積と売上高のデータで単回帰分析をおこなって、結果を A50 セル以下に表示して下さい。相関係数を、小数第２位までの数値で求めなさい。この２つのことがらに、相関関係があると言えますか。

（２）宣伝費と売上高のデータで単回帰分析をおこなって、結果を A75 セル以下に表示して下さい。相関係数を、小数第２位までの数値で求めなさい。この２つのことがらに、相関関係があると言えますか。

［解答１］

（１）相関係数＝0.19　店舗面積と売上高のあいだに、相関関係はないと言っていい。

（２）相関係数＝0.58　宣伝費と売上高のあいだに、相関関係はあると言っていい。

３．理論値

今回の回帰分析（線形回帰分析といいます）では、原因データ X と結果データ Y のあいだに、「Y=aX＋b」という関係が成り立つという前提で分析しています。

「Y＝aX＋b」覚えていますか？　中学 2 年生の数学で勉強する「１次関数」というもので、グラフにすると直線になります。a を「かたむき」、b を「切片」といいます。イヤなこと思い出させちゃいましたか(^_^;)

商圏人口と売上の回帰分析で得た、右の数値を見て下さい。切片の係数にある約 12679 が「切片 b」、X 値 1 の係数にある約 611 が「かたむき a」です。

	係数
切片	12679.49268
X 値 1	611.162132

つまり、理論値は

　　売上高＝611×商圏人口＋12679

で計算されます。

商圏人口と売上高には強い相関関係があるのですから、もし 21 号店の商圏人口が 3 万人だとしたら、

　　売上高＝611×30＋12679＝31009

およそ 3100 万円の売上高になると想像できるワケです。

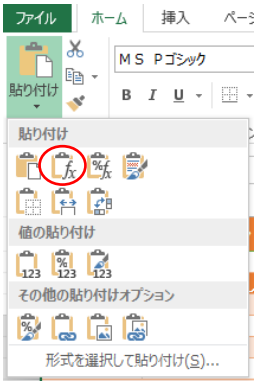
Excel093

F 列に、商圏人口と売上高の相関関係をもとにした、売上高の理論値を入力したいと思います。

①F4 セルに、理論値が計算されるような数式を、入力して下さい。

（「=$B$42*B4+$B$41」となります。「$」は絶対参照で、この数式をほかのセルにコピーしても、この部分は変わりません。）

②F4 セルをコピー→F5～F23 セルに「数式として」貼り付けします。



右図のような結果に、なりましたか？

理論値1
26,492
30,342
30,036
28,692
32,053
29,792
30,831
29,731
24,047
30,403
23,680
23,436
28,631
30,892
28,875
32,298
27,836
28,570
29,548
24,597

［練習問題２］

（１）店舗面積と売上高の相関関係をもとにした売上高の理論値が、G 列に計算されるようにしたいと思います。G 列に、適切な数式を入力して下さい。

（２）宣伝費と売上高の相関関係をもとにした売上高の理論値が、F 列に計算されるようにしたいと思います。F 列に、適切な数式を入力して下さい。

［解答２］

G4 セル→　=$B$67*C4+$B$66

F4 セル→　=$B$92*D4+$B$91

と入力して、G4～F4 セルを G5～F23 セルに「数式として」貼り付けします。

右図のような結果になれば、OK。

理論値2	理論値3
27,689	29,531
28,942	25,699
28,763	27,832
29,210	25,894
29,389	29,336
27,957	27,415
27,510	30,511
29,121	26,598
28,226	27,074
29,837	27,891
27,689	29,104
29,121	26,165
28,942	26,755
28,763	31,507
28,405	30,403
29,121	31,945
28,047	30,982
28,763	26,516
27,778	31,880
27,510	27,745

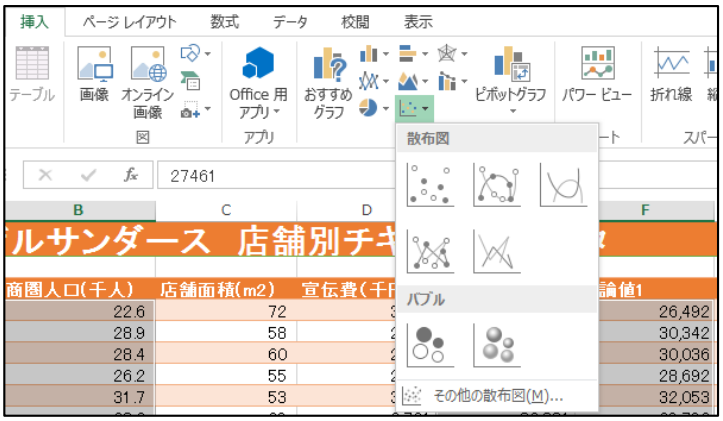
©特定非営利活動法人パソコンどーじょー

Excel093

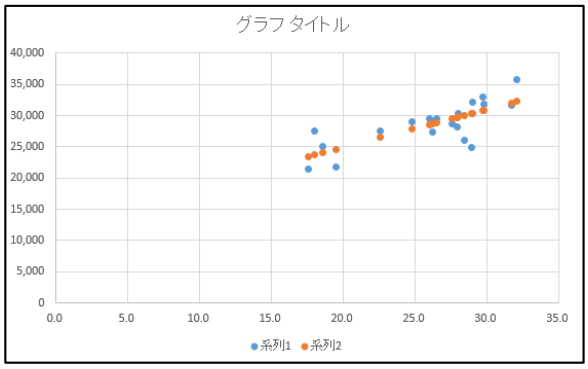
４．散布図

実際には、「商圏人口と売上高の関係」は、直線的ではないかもしれません。グラフを作って、仮説が適切か、不十分ではないか、確認してみましょう。

①商圏人口データ（B4～B23 セル）・売上高データ（E4～E23 セル）・理論値１データ（F4～F23 セル）を選択→「挿入」タブ→「散布図」で「散布図」を選択します。
（複数の領域を選択するときは、Ctrl キーを使います。）

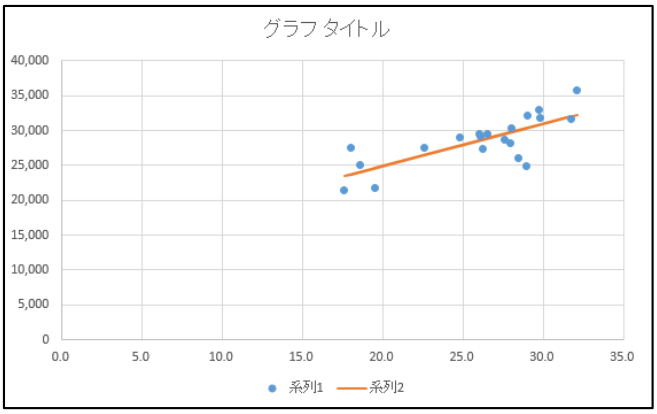


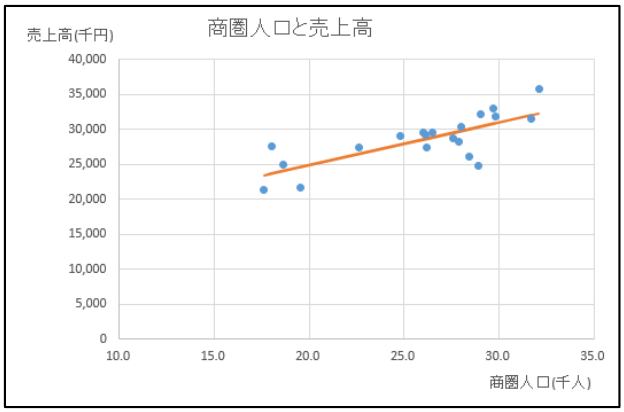
②右図のようなグラフになります。グラフを選択して、「書式」タブ→「現在の選択範囲」を「系列 2」（理論値データ）とし、「選択対象の書式設定」。



③「塗りつぶしと線」項目で、線は単色とします。また、マーカーは塗りつぶし・枠線ともなしにします。これで、理論値データのグラフは、直線になります。

④下左図のようなグラフになりましたか？　余裕があったら、各範囲の書式設定を調整して、下右図のように見やすくしてみましょう。





Excel093

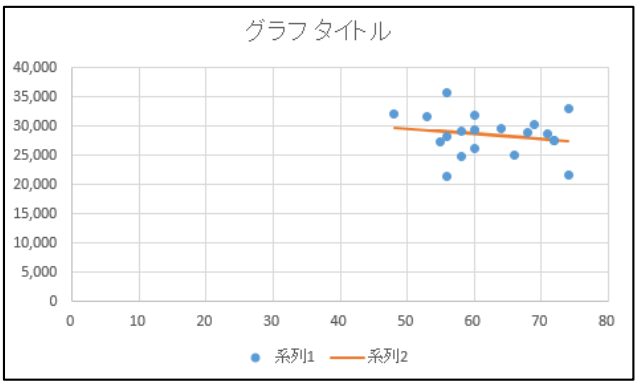
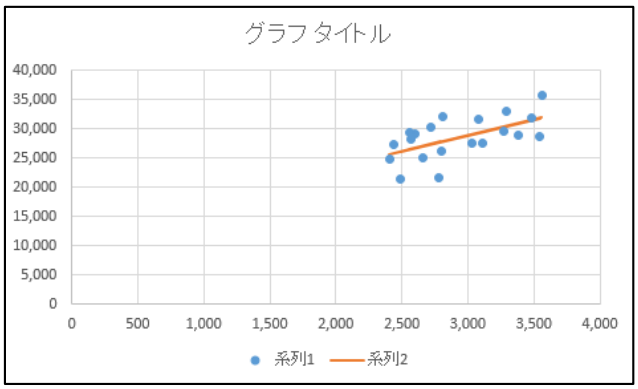
［練習問題３］

（１）店舗面積と売上高および理論値２のデータをもとに、前ページのようなグラフを作成して下さい。

（２）宣伝費と売上高および理論値３のデータをもとに、前ページのようなグラフを作成して下さい。

［解答３］

（１）



（２）