

回帰分析（2）

前回は、「商圈人口と売上高の相関関係」というように、原因（と想定していることがら）が1つでした。今回は、2つ以上の場合をアツかつてみます。
 ※わかりやすさを優先しているため、数学的な厳密さは欠いています。実際の仕事などに活用する際は、注意して下さい。

1. 用語の整理

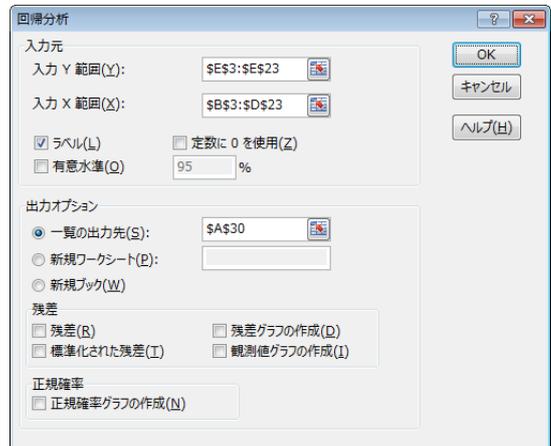
「商圈人口・店舗面積・宣伝費によって、売上高はどう変わるか」という場合……

- ・ **説明変数** = 商圈人口・店舗面積・宣伝費といった、原因と想定していることがら。
- ・ **目的変数** = 売上高のように、結果となることがら。
- ・ **単回帰分析** = 「商圈人口と売上高」のように、説明変数が1つの回帰分析。
- ・ **重回帰分析** = 「商圈人口・店舗面積・宣伝費と売上高」のように、複数の説明変数を同時にアツかう回帰分析。
- ・ **回帰式** = 目的変数の理論値を計算する式。

2. 重回帰分析（1）

「チキン売上 2.xlsx」ファイルを開いて下さい。まず、商圈人口・店舗面積・宣伝費を説明変数とします。

- 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（目的変数）」欄に、売上高のデータ E3～E23 セルを指定します。（今回は、ラベル（3行目）も含めてみます。）
- ③ 同様に、「入力 X 範囲（説明変数）」に、B3～D23 セルを指定します。
- ④ 「ラベル」にチェック。
- ⑤ 「一覧の出力先」に、A30 セルを指定して、「OK」。



30 行目以下に、右図のような結果が表示されましたか？

重相関 R が 0.84 ですから、「かなり強い相関関係がある」といえます。

概要		回帰統計						
重相関 R	0.843855034							
決定 R2	0.712081319							
修正 R2	0.658108442							
標準誤差	2087.180111							
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	3	174049337.8	58016445.95	13.19105891	0.000135931			
残差	16	70370630.71	4388164.42					
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	7904.945041	6216.682401	1.271569711	0.221699765	-5273.832924	21083.72301	-5273.832924	21083.72
商圈人口(千)	459.4059824	141.6233343	3.243857976	0.005086322	158.1778256	759.6340392	158.1778256	759.634
店舗面積(m2)	-52.21802851	81.9949396	-0.636844527	0.533232686	-226.0395355	121.6034785	-226.0395355	121.6035
宣伝費(千円)	4.089270494	1.580962347	2.586570454	0.019875308	0.737780026	7.440768951	0.737780026	7.440761

ただし、「P-値」を見て下さい。これは、「無関係である確率」です。店舗面積の P-値は 0.53（53%）とずば抜けて高いですから、目的変数としないほうがよさそうですね。

	係数	標準誤差	t	P-値
切片	7904.945041	6216.682401	1.271569711	0.221699765
商圈人口(千)	459.4059824	141.6233343	3.243857976	0.005086322
店舗面積(m2)	-52.21802851	81.9949396	-0.636844527	0.533232686
宣伝費(千円)	4.089270494	1.580962347	2.586570454	0.019875308

Excel094

3. 重回帰分析 (2)

店舗面積ははずして、商圈人口・宣伝費を説明変数として回帰分析をおこなってみます。

- ① 「データ 2」シートを選択して、「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 (目的変数)」欄に、売上高のデータ D3~D23 セルを指定します。(今回は、ラベル (3 行目) も含めてみます。)
- ③ 同様に、「入力 X 範囲 (説明変数)」に、B3~C23 セルを指定します。
- ④ 「ラベル」にチェック。
- ⑤ 「一覧の出力先」に、A30 セルを指定して、「OK」。



結果を、見てみましょう。

重相関 R=0.84、商圈人口と宣伝費の P-値がそれぞれ 0%・1%ですから、かなりよい相関関係モデルだと言えます。

概要								
回帰統計								
重相関 R	0.839519718							
重決定 R2	0.704793357							
補正 R2	0.670063164							
標準誤差	2060.188433							
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	2	172265570.1	86132785.05	20.29338997	3.13381 E-05			
残差	17	72154398.44	4244376.379					
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	4904.578472	3984.413834	1.230941031	0.235104404	-3501.799903	13310.95685	-3501.8	13310.96
商圈人口(千)	512.7710727	112.158732	4.571833718	0.000270842	276.1368328	749.4053127	276.1368	749.4053
宣伝費(千円)	3.526534705	1.287844316	2.738323771	0.014006502	0.809420705	6.243648705	0.809421	6.243649

4. 回帰式

右図のような係数から、回帰式は次のようになります。

$$\text{売上高} = 513 \times \text{商圈人口} + 3.53 \times \text{宣伝費} + 4900$$

(係数は、4桁目を四捨五入してあります。)

	係数
切片	4904.578472
商圈人口(千)	512.7710727
宣伝費(千円)	3.526534705

[練習問題 1]

- (1) E 列に、回帰式を表す数式を入力して、理論値を計算して下さい。
- (2) F 列に、実際の売上高と、回帰式から求められる理論値との差が計算されるように、数式を入力して下さい。
- (3) G 列に、F 列の値が大きい順 (実際の売上高が理論値より大きい順) に、順位が計算されるよう、関数を入力して下さい。

Excel094

[解答1]

(1) E4セルに以下を入力→コピーして、E5～E23セルに「数式として」貼り付け。

$$=B\$47*B4+B\$48*C4+B\$46$$

(2) F4セルに以下を入力→コピーして、F5～F23セルに「数式として」貼り付け。

$$=D4-E4$$

(3) G4セルに以下を入力→コピーして、G5～G23セルに「数式として」貼り付け。

$$=RANK(F4, \$F\$4: \$F\$23, 0)$$

(Rank.EQ 関数でも OK)

理論値	理論値との差	順位
27,468	-7	10
28,201	-3,307	20
29,335	-3,202	19
26,944	497	9
32,007	-409	12
28,858	1,523	6
31,747	1,307	7
28,274	-83	11
23,816	1,224	8
29,681	2,455	2
24,830	2,785	1
22,710	-1,321	16
27,453	1,789	5
32,447	-572	15
30,035	-463	13
33,912	1,870	4
29,541	-538	14
27,247	2,258	3
31,562	-2,831	17
24,714	-2,974	18

[練習問題2]

ハーゲルサンダースでは、21号店の出店を決めました。商圈人口は23.1千人です。

(1) 25行目は、「宣伝費を250万円(2,500千円)かけた場合」です。想定される売上高を、E25セルに計算して下さい。

(2) 26行目は、「目標売上高300万円(3,000千円)とする場合」です。この回帰分析モデルでは、宣伝費をいくらにすればいいと考えられますか。C26に計算して下さい。

(方程式の問題なので、苦手な方はパスして下さい。)

[解答2]

(1) 理論値の数式(E4セル)を、そのまま数式としてコピーすればOK。結果は、25,566千円となります。

(2) 以下の方程式を、C26について解きます。

$$B\$47*B26+B\$48*C26+B\$46=30000$$

C26について解くと、

$$C26=(30000-B\$47*B26-B\$46)/B\$48$$

結果は、3,757千円となります。

Excel094

[実力問題]

「アイス売上.xlsx」を開いて下さい。その日の気温・湿度と、ハーゲルダッツ・アイスクリームの売上高の、1年間の記録です。

- (1) 気温と売上高について、単回帰分析をおこなってください。相関係数（重相関 **R**）はいくつですか？
- (2) E 列に、(1) の回帰分析モデルによる理論値（回帰式）を、入力して下さい。
- (3) 気温と売上高・理論値 1 の関係を、散布図グラフで表して下さい。グラフを見て、何か感じることはありますか。
- (4) 湿度と売上高について単回帰分析をおこなってください。相関係数（重相関 **R**）はいくつですか？
- (5) F 列に、(4) の回帰分析モデルによる理論値（回帰式）を、入力して下さい。
- (6) 気温と売上高・理論値 2 の関係を、散布図グラフで表して下さい。グラフを見て、何か感じることはありますか。
- (7) 気温・湿度と売上高について、重回帰分析をおこなってください。
- (8) 今日の予想気温は 18℃、予想湿度は 45%です。(7) の回帰分析モデルをもとに、売上高を予想して下さい。

Excel094

[実力問題の解答]

(1) 相関係数は、約 0.63。相関関係があると言っていい数値です。

(2) (1) の結果、切片や気温の係数は、右図のようになります。E4セルに、以下の数式を入力し、E5セル以下にコピーします。

	係数
切片	22443.02
気温	915.6937

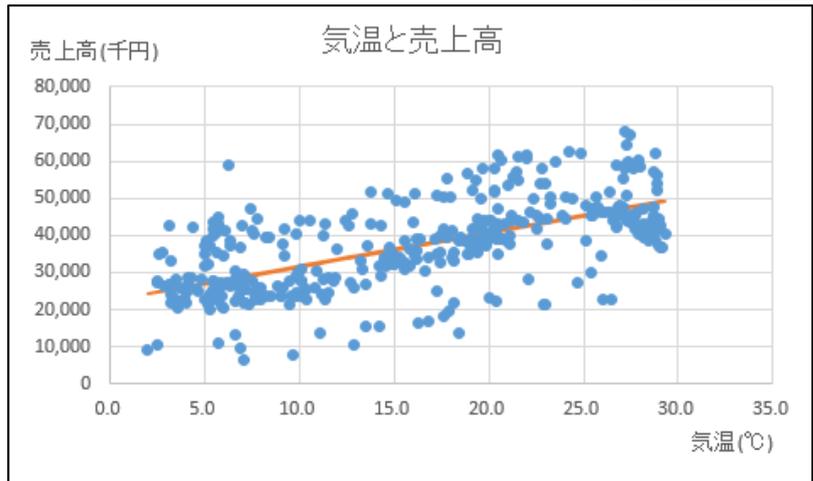
= (気温の係数)*B4+(切片)

(「気温の係数」「切片」は、絶対参照のセル番地が入力されます。)

1月1日～14日あたりまでの結果は、右図の通り。

理論値1
28,212
28,303
28,303
27,754
27,388
27,296
27,571
27,662
27,846
28,487
27,662
25,648

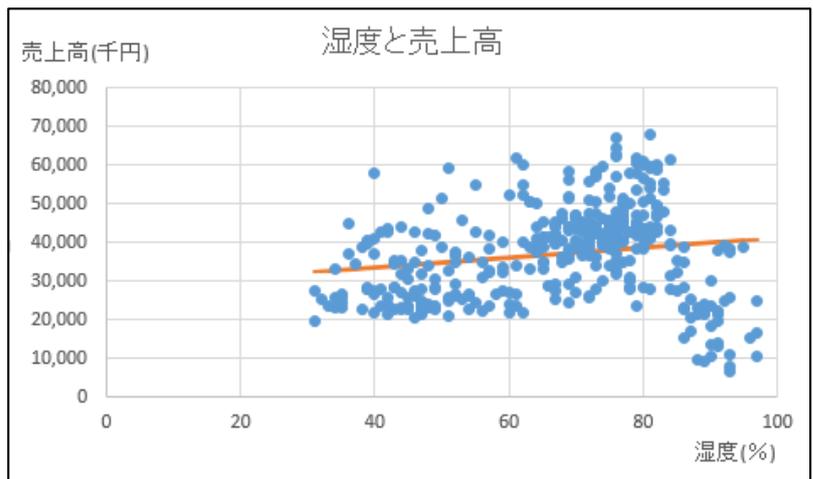
(3) 右図の通り。



(4) 相関係数は、約 0.17。相関関係があるとはいえません。

理論値2
34,940
34,811
33,001
34,294
32,872
34,423
33,518
33,001
34,940
35,586
35,198
33,389
34,681
33,906
34,552

(5) (2) と同様。1月1日～14日の結果は右図の通り。



(6) 右図の通り

(7) 右図。

(8) $1074 \times 18 - 156 \times 45 + 30350 = 42662$
 およそ、42,700円です。

概要								
回帰統計								
重相関 R	0.654566							
重決定 R2	0.428456							
補正 R2	0.425307							
標準誤差	9103.677							
観測数	366							
分散分析表								
	自由度	変動	分散	割られた分散	有意 F			
回帰	2	2.26E+10	1.13E+10	136.0611	8.03E-45			
残差	363	3.01E+10	82876937					
合計	365	5.26E+10						
係数表								
	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	30349.89	2032.753	14.93044	1.23E-39	26352.44	34347.34	26352.44	34347.34
気温	1073.898	67.51742	15.9055	1.35E-43	941.1239	1206.673	941.1239	1206.673
湿度	-156.09	34.57104	-4.51505	8.57E-06	-224.075	-88.1054	-224.075	-88.1054