

第2章 単純回帰分析(その1)

- 【課題】教科書 41 ページの家計消費支出と家計可処分所得のデータ(表 2-1)について、
 散布図を描く
 最小 2 乗パラメータ推定値を求める
 予測値を求めて散布図に描き入れる
 という 3 つのをおこなってみよう。

1. 散布図

まずは、家計可処分所得を横軸に、家計消費支出を縦軸にとった散布図を描いてみる。

☞ 手順 A 列に年、B 列に家計消費支出、C 列に家計可処分所得が入力されているものとする。
 このとき、

グラフウィザードを起動し、散布図を選択する。

データ範囲として、B2:C11 を選択する。

系列のタグをクリックし、「X の値」を C2:C11、「Y の値」を B2:B11 とする。すなわち、
 B 列と C 列の部分の入れかえである。

グラフタイトルの入力、凡例・目盛線の表示などを適切におこない、グラフを完成させる。

軸の書式設定において、

縦軸 最小値：220 最大値：290 目盛間隔：20

横軸 最小値：260 最大値：320 目盛間隔：20

という設定をおこなえば、後で示す作成見本どおりになる。

2. 最小2乗パラメータ推定値の導出

最小 2 乗方によってパラメータ推定値 \hat{a} , \hat{b} を求める。関数や分析ツールなどで求める方法もあるが、ここでは X と Y の偏差をそれぞれ計算し、その 2 乗和と交差積の和から求めていく。

☞ 手順

B13 セル、C13 セルにそれぞれ \bar{X} , \bar{Y} を求める。(関数 AVERAGE を利用する)

D 列、E 列に偏差を求める。D2 セルに「=B2 - B\$13」とし、これをコピーする。E 列についても同様におこなう。

F 列、G 列、H 列に x^2 乗、 y^2 乗、 xy をそれぞれ計算する。2 乗を表す演算子は '^' であり、F2 セルに D2 セルの 2 乗を求める場合は、「=D2^2」と入力することになる。これらの和を、F13 セル、G13 セル、H13 セルにそれぞれ計算する。合計を求める関数は SUM である。

このようにして求めたものが、 S_x^2 , S_y^2 , S_{xy} である。これらと、 \bar{X} , \bar{Y} を用いれば、パラメータ推定値 \hat{a} , \hat{b} を計算することができる。

3. 予測値と残差の計算

最小 2 乗法とは、残差の 2 乗和を最小にするようにパラメータ推定値を決定する方法である。

この残差は、ある X に対応する直線上の点(予測値または理論値) \hat{Y} を Y の実績値から引いたもの $(Y - \hat{Y})$ である。

☞ 手順

I2 セルに平成 1 年の X(C2 セル)に対応する予測値を求めるなら「=SC\$16+SC\$15*C2」(コピーの際に絶対参照を用いるので、\$がついている)とする。これを I3:I11 にコピーする。

J2 セルに「=B2-I2」とするとこの X に対する残差が求まる。これを J3:J11 にコピーする。

4. 回帰直線の散布図への書き入れ

予測値を直線で繋ぐことによって、散布図に回帰直線を描き入れることができる。

📖 手順

グラフをアクティブにした状態で、メニューバーから「グラフ」-「データの追加」を選び、I2:I11をグラフに追加する。

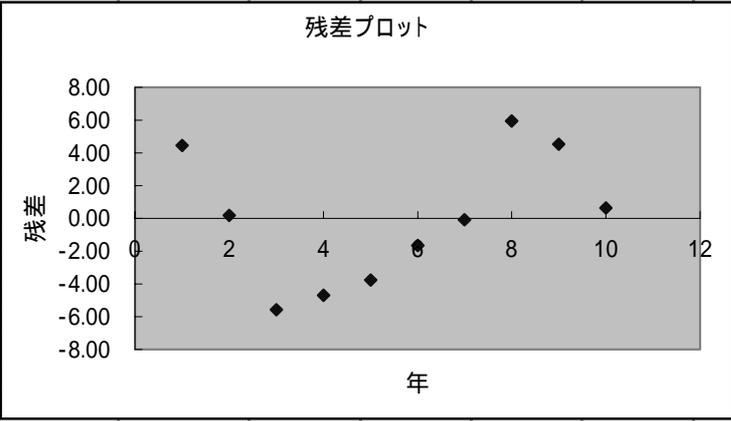
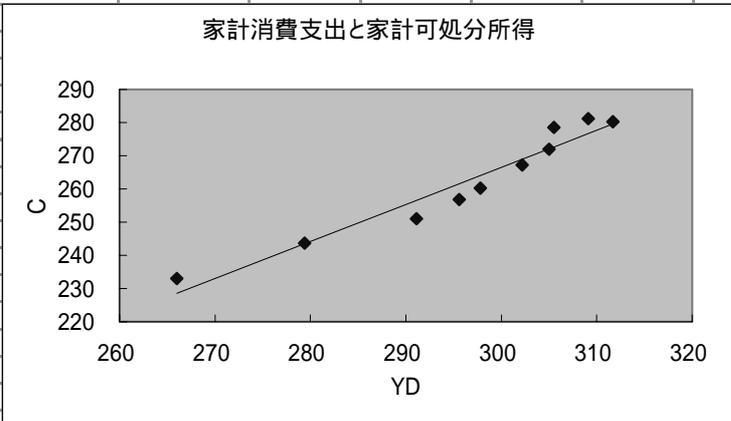
の操作で散布図上に表れたピンク色の点を直線で結ぶ。「データ系列の書式設定」の「パターン」のタグにおいて、線を「指定」にし、色を黒に変え、マーカーを「なし」にする。

5. 残差の表示

残差の状態を調べることで、回帰直線のあてはまり具合など、さまざまな情報を入手することができる。そのため、グラフに表示することが良くおこなわれる。ここでは残差 e_i を縦軸にとり、年を横軸にとった残差プロットを描いてみよう。

<作成見本>

	A	B	C	D	E	F	G	H	I	J
1	年	消費支出(Y)	所得(X)	偏差 y	偏差 x	y^2	x^2	xy	予測値	残差
2	1	233.0	266.0	-29.4	-30.3	865.54	920.52	892.60	228.54	4.46
3	2	243.7	279.4	-18.7	-16.9	350.44	286.96	317.12	243.50	0.20
4	3	251.0	291.1	-11.4	-5.2	130.42	27.46	59.84	256.57	-5.57
5	4	256.9	295.6	-5.5	-0.7	30.47	0.55	4.08	261.59	-4.69
6	5	260.3	297.8	-2.1	1.5	4.49	2.13	-3.10	264.05	-3.75
7	6	267.3	302.2	4.9	5.9	23.81	34.34	28.60	268.96	-1.66
8	7	272.0	305.0	9.6	8.7	91.78	75.00	82.96	272.09	-0.09
9	8	278.6	305.5	16.2	9.2	261.79	83.91	148.21	272.65	5.95
10	9	281.2	309.1	18.8	12.8	352.69	162.82	239.63	276.67	4.53
11	10	280.2	311.7	17.8	15.4	316.13	235.93	273.10	279.57	0.63
12										
13	平均	262.4	296.3			2427.56	1829.60	2043.05		
14										
15		b	1.12							
16		a	-68.49							
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										
35										
36										
37										
38										
39										
40										
41										
42										
43										
44										
45										
46										
47										
48										
49										
50										
51										
52										



- 【課題】教科書 41 ページの家計消費支出と家計可処分所得のデータ（表 2-1）について、決定係数を求める
 $H_0: b=0$, $H_0: a=0$ という検定をそれぞれおこなってみる。
という 2 つのことをおこなってみよう。

6. 決定係数

回帰直線のあてはまり具合を示す指標である決定係数を求める。 S_x^2 , S_y^2 , S_{xy} はパラメータ推定値の導出の時に計算されているので、それを用いることができる。

- 📖 手順 C17 セルに決定係数、C18 セルに相関係数を求めよう。
C17 セルに「=H13^2/(F13*G13)」と入力する。
C18 セルに「=H13/SQRT(F13*G13)」と入力する。

7. 回帰係数の検定

回帰係数 a, b について、1) $H_0: b=0$ 2) $H_0: a=0$ という仮説検定をおこなう。これらは t 検定となる。

📖 手順

- K 列に残差の 2 乗を求め、K13 セルにその和を求める。
K15 セルに「=K13/8」とすることによって、 s^2 を求める。
E15 セルに b の t 値を求める。「=C15/SQRT(K15/G13)」と入力することによって求められる。
E16 セルに a の t 値を求める。「=C16/SQRT(K15*(1/10+C13^2/G13))」と入力することによって求められる。
これらの値を $t_{0.95}$ と比較することによって検定がおこなわれる。自由度 8 の t 分布の $t_{0.95}$ は、「=TINV(0.05,8)」と入力することによって求められる。

8. 分析ツールの利用

Excel には統計分析を行うためのいくつかの分析ツールが付属している。これらのツールを使えば一度に詳細な分析結果を得ることができる。

- 分析ツールを最初に使用する場合にはメニューバーの「ツール」-「アドイン」を選び、分析ツールをチェックすることによって、分析ツールをアドイン(有効にすること)しなくてはならない。
アドインを行った後で、再びメニューバーから「ツール」を選ぶと、下のほうに「分析ツール」と表示される。ここで分析ツールを選び、回帰分析を選べばよい。