● **推定結果の表記方法：　回帰モデル：**

$$Y_i = \alpha + \beta X_i + u_i$$

の推定の結果， $\hat{\alpha} = 0.5$**，** $\hat{\beta} = 0.7$**，** $s_{\hat{\alpha}} = \sqrt{1.5766667} = 1.25565$**，** $s_{\hat{\beta}} = \sqrt{0.1433333} = 0.3786$**，** $\dfrac{\hat{\alpha}}{s_{\hat{\alpha}}} = 0.398$**，** $\dfrac{\hat{\beta}}{s_{\hat{\beta}}} = 1.849$**，** $s^2 = 1.433333$ (**すなわち，** $s = 1.197$)**，** $R^2 = 0.5326$**，** $\overline{R}^2 = 0.3768$ **を得た。**

これらをまとめて，

$$Y_i = \underset{\mathbf{(0.398)}}{\mathbf{0.5}} + \underset{\mathbf{(1.849)}}{\mathbf{0.7}} X_i$$

$$R^2 = 0.5326, \quad \overline{R}^2 = 0.3768, \quad s = 1.197$$

ただし，係数の推定値の下の括弧内は t **値を表すものとする。**

または，

$$Y_i = \underset{(1.256)}{\mathbf{0.5}} + \underset{(0.379)}{\mathbf{0.7}} X_i$$

$$R^2 = 0.5326, \quad \overline{R}^2 = 0.3768, \quad s = 1.197$$

ただし，係数の推定値の下の括弧内は標準誤差を表すものとする。

のように書く。

5.3.11　予測

X_0 を与えたもとで，Y_0 の予測量 $\hat{Y}_0$ は，

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0$$

となる。

ただし，Y_0 と X_0 との関係は，

$$Y_0 = \alpha + \beta X_0 + u_0$$

$$u_0 \sim N(0, \sigma^2)$$

である。u_0 は $u_1, u_2, \cdots, u_n$ から独立とする。

予測誤差 $\hat{Y}_0 - Y_0$ は，

$$\hat{Y}_0 - Y_0 = (\hat{\alpha} - \alpha) + (\hat{\beta} - \beta)X_0 - u_0$$

となる。

● 予測誤差の期待値と分散：　両辺に期待値をとると，

$$\mathbf{E}(\hat{Y}_0 - Y_0) = \mathbf{E}(\hat{\alpha} - \alpha) + \mathbf{E}(\hat{\beta} - \beta)X_0 - \mathbf{E}(u_0) = 0$$

を得る ($\hat{\alpha}$，$\hat{\beta}$ の不偏性と u_0 の仮定より)。

$\mathbf{E}(\hat{Y}_0 - Y_0) = 0$ なので，$\mathbf{E}(\hat{Y}_0) = \mathbf{E}(Y_0) = \alpha + \beta X_0$ となる。

(注意)

$\mathbf{E}(\hat{Y}_0 - Y_0) = 0$ **は正しい書き方であるが，**$\mathrm{E}(\hat{Y}_0) = Y_0$ **とは書けない。**

$Y_0 = \alpha + \beta X_0 + u_0$ **なので，**Y_0 **も確率変数なので，**$\mathbf{E}(\hat{Y}_0) = \mathbf{E}(Y_0) \neq Y_0$ **となる。**

分散について，

$$
\begin{aligned}
\mathbf{V}(\hat{Y}_0 - Y_0) &= \mathbf{E}\Big((\hat{Y}_0 - Y_0)^2\Big) \\
&= \mathbf{E}\Big((\hat{\alpha} - \alpha) + (\hat{\beta} - \beta)X_0 - u_0\Big)^2 \\
&= \mathbf{E}\Big((\hat{\alpha} - \alpha)^2\Big) + X_0^2\mathbf{E}\Big((\hat{\beta} - \beta)^2\Big) + \mathbf{E}(u_0^2) + 2X_0\mathbf{E}\Big((\hat{\alpha} - \alpha)(\hat{\beta} - \beta)\Big) \\
&\quad -2\mathbf{E}\Big((\hat{\alpha} - \alpha)u_0\Big) - 2X_0\mathbf{E}\Big((\hat{\beta} - \beta)u_0\Big)
\end{aligned}
$$

各項の期待値は，

$$\mathbf{E}\Big((\hat{\alpha}-\alpha)^2\Big) = \mathbf{V}(\hat{\alpha}) = \sigma^2\Big(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\Big)$$

$$\mathbf{E}\Big((\hat{\beta}-\beta)^2\Big) = \mathbf{V}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

$$\mathbf{E}(u_0^2) = \sigma^2$$

$$\mathbf{E}\Big((\hat{\alpha}-\alpha)(\hat{\beta}-\beta)\Big) = \mathbf{Cov}(\hat{\alpha},\ \hat{\beta}) = -\frac{\sigma^2\overline{X}}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

$$\mathbf{E}\Big((\hat{\alpha}-\alpha)u_0\Big) = 0$$

$$\mathbf{E}\Big((\hat{\beta}-\beta)u_0\Big) = 0$$

となるので，

$$\mathbf{V}(\hat{Y}_0 - Y_0) = \sigma^2\Big(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\Big) + \frac{\sigma^2 X_0^2}{\sum_{i=1}^n (X_i - \overline{X})^2} + \sigma^2 - \frac{2\sigma^2 X_0 \overline{X}}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

$$= \sigma^2\left(1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)$$

を得る。

$X_0 = \overline{X}$ **のとき，予測分散が最小になり，そのときの予測分散** $\mathrm{V}(\hat{Y}_0 - Y_0)$ **の値は** $\sigma^2\left(1 + \frac{1}{n}\right)$ **となる。**

X_0 **が** $\overline{X}$ **から離れるにつれて，予測分散は大きくなる。**

● 予測の区間推定： $\hat{Y}_0 - Y_0$ **の分布は，**

$$\hat{Y}_0 - Y_0 \sim N\left(0,\ \sigma^2\left(1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)\right)$$

となり，よって，

$$\frac{\hat{Y}_0 - Y_0}{\sigma\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}} \sim N(0, 1)$$

を得る。

一方，s^2 について，

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$$

となる。

ただし，$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}X_i)^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$ である。

$\hat{Y}_0 - Y_0$ と s^2 は独立なので (証明略)，

$$
\frac{\dfrac{\hat{Y}_0 - Y_0}{\sigma\sqrt{1 + \dfrac{1}{n} + \dfrac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}}}{\sqrt{\dfrac{(n-2)s^2}{\sigma^2}/(n-2)}} = \frac{\hat{Y}_0 - Y_0}{s\sqrt{1 + \dfrac{1}{n} + \dfrac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}} \sim t(n-2),
$$

となる。

$t_{\alpha/2}(n-2)$ を，自由度 $n-2$ の t 分布から得られた $100 \times \alpha$ % 点の値とすると，

$$
\textbf{Prob}\left(-t_{\alpha/2}(n-2) < \frac{\hat{Y}_0 - Y_0}{s\sqrt{1 + \dfrac{1}{n} + \dfrac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}} < t_{\alpha/2}(n-2) \right) = 1 - \alpha
$$

となり，

$$\textbf{Prob}\left(\hat{Y}_0 - t_{\alpha/2}(n-2)\times s\sqrt{1+\frac{1}{n}+\frac{(X_0-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}} < Y_0 < \hat{Y}_0 + t_{\alpha/2}(n-2)\times s\sqrt{1+\frac{1}{n}+\frac{(X_0-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}}\right) = 1-\alpha$$

となる。

$\hat{Y}_0$，s^2 に推定値を代入して，信頼係数 $1-\alpha$ の Y_0 の信頼区間は，

$$\left(\hat{Y}_0 - t_{\alpha/2}(n-2)\times s\sqrt{1+\frac{1}{n}+\frac{(X_0-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}},\ \ \hat{Y}_0 + t_{\alpha/2}(n-2)\times s\sqrt{1+\frac{1}{n}+\frac{(X_0-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}}\right)$$

が得られる。

数値例：　今までと同様に，以下の数値例をとりあげる。

i	X_i	Y_i	X_i^2	X_iY_i	$\hat{Y}_i$	$\hat{u}_i$
1	**5**	**4**	**25**	**20**	**4.0**	**0.0**
2	**1**	**1**	**1**	**1**	**1.2**	-0.2
3	**3**	**1**	**9**	**3**	**2.6**	-1.6
4	**2**	**3**	**4**	**6**	**1.9**	**1.1**
5	**4**	**4**	**16**	**16**	**3.3**	**0.7**
合計	$\sum X_i$ **15**	$\sum Y_i$ **13**	$\sum X_i^2$ **55**	$\sum X_iY_i$ **46**	$\sum \hat{Y}_i$ **13**	$\sum \hat{u}_i$ **0.0**
平均	$\overline{X}$ **3**	$\overline{Y}$ **2.6**				

必要な数値は，

$$\overline{X} = 3$$

$$\sum_{i=1}^{n} X_i^2 = 55$$

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = 55 - 5 \times 3^2 = 10$$

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} X_0 = 0.5 + 0.7X_0$$

$$s = \sqrt{1.433333} = 1.197$$

である。

$X_0 = 6$ **のときの，信頼係数 0.90 の** Y_0 **の信頼区間は，**$t_{0.05}(3) = 2.3534$**，**$\hat{Y}_0 = 0.5+0.7\times6 = 4.7$

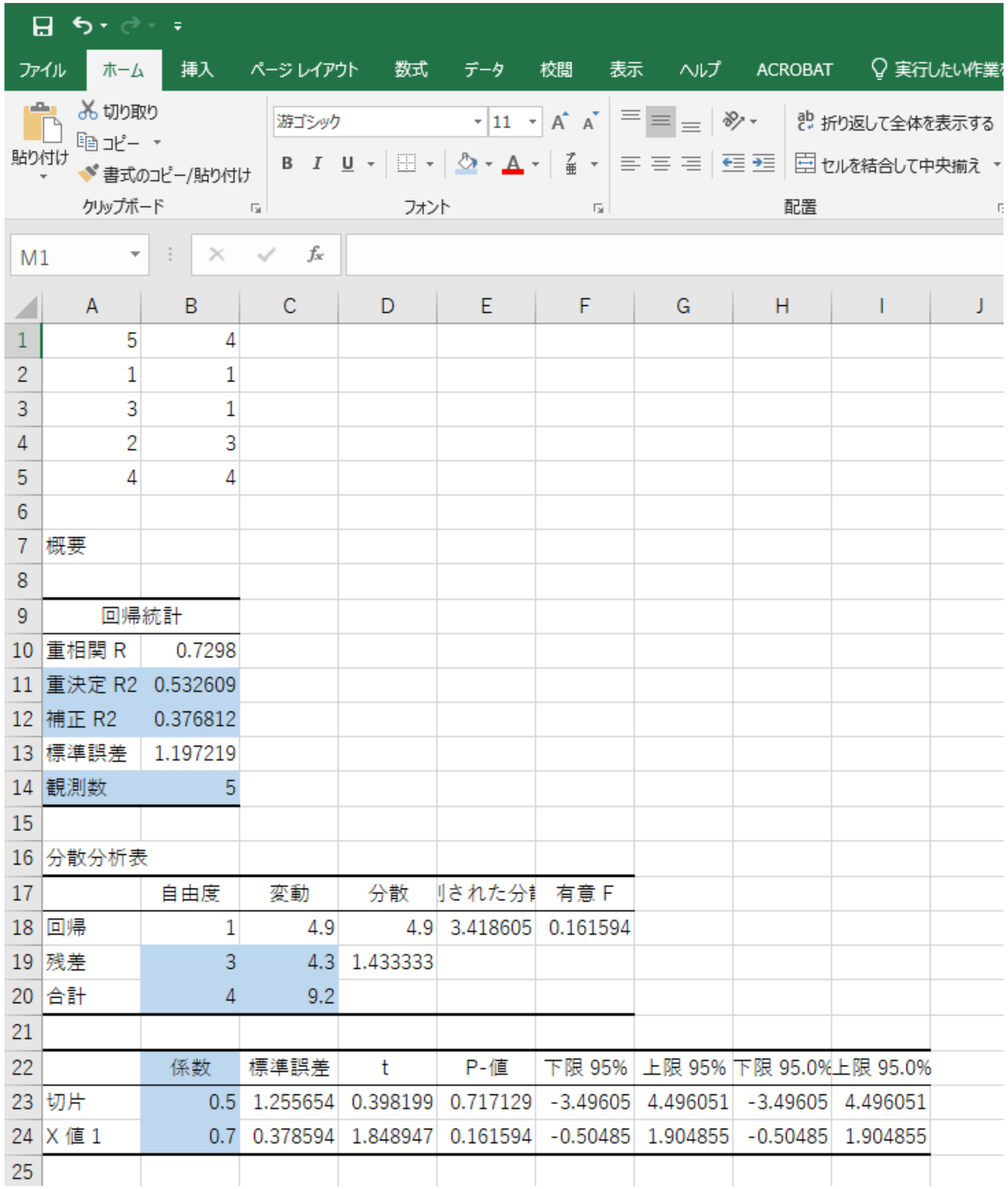
なので，

$$\left(4.7 - 2.3534 \times 1.197\sqrt{1 + \frac{1}{5} + \frac{(6-3)^2}{10}},\ \ 4.7 + 2.3534 \times 1.197\sqrt{1 + \frac{1}{5} + \frac{(6-3)^2}{10}}\right)$$
$$= (0.61775,\ \ 8.75225)$$

を得る。

5.3.12　Excel 2019 による回帰分析（その 2）

前回の推定結果の再掲（A 列が X，B 列が Y）



	A	B	C	D	E	F	G	H	I
1	5	4							
2	1	1							
3	3	1							
4	2	3							
5	4	4							
6									
7	概要								
8									
9	回帰統計								
10	重相関 R	0.7298							
11	重決定 R2	0.532609							
12	補正 R2	0.376812							
13	標準誤差	1.197219							
14	観測数	5							
15									
16	分散分析表								
17		自由度	変動	分散	された分	有意 F			
18	回帰	1	4.9	4.9	3.418605	0.161594			
19	残差	3	4.3	1.433333					
20	合計	4	9.2						
21									
22		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
23	切片	0.5	1.255654	0.398199	0.717129	-3.49605	4.496051	-3.49605	4.496051
24	X 値 1	0.7	0.378594	1.848947	0.161594	-0.50485	1.904855	-0.50485	1.904855
25									

重相関 R	0.7298

→ $\hat{Y}_i$ と Y_i との相関係数

重決定 R2	0.532609

→ 決定係数（＝$\hat{Y}_i$ と Y_i との相関係数の二乗）

補正 R2	0.376812

→ 自由度修正済み決定係数

標準誤差	1.197219

→ 回帰式の標準誤差 s

観測数	5

→ 標本数（データ数）n

標準誤差
1.255654
0.378594

→ 各係数の標準誤差（上段は $s_{\hat{\alpha}}$，下段は $s_{\hat{\beta}}$ ）

t
0.398199
1.848947

→ t値（上段は H$_0$：α＝0，下段は H$_0$：β＝0 の検定）

下限 95%	上限 95%
-3.49605	4.496051
-0.50485	1.904855

→ 95%信頼区間（上段は α，下段は β の信頼区間），F 列・G 列

下限 95.0%	上限 95.0%
-3.49605	4.496051
-0.50485	1.904855

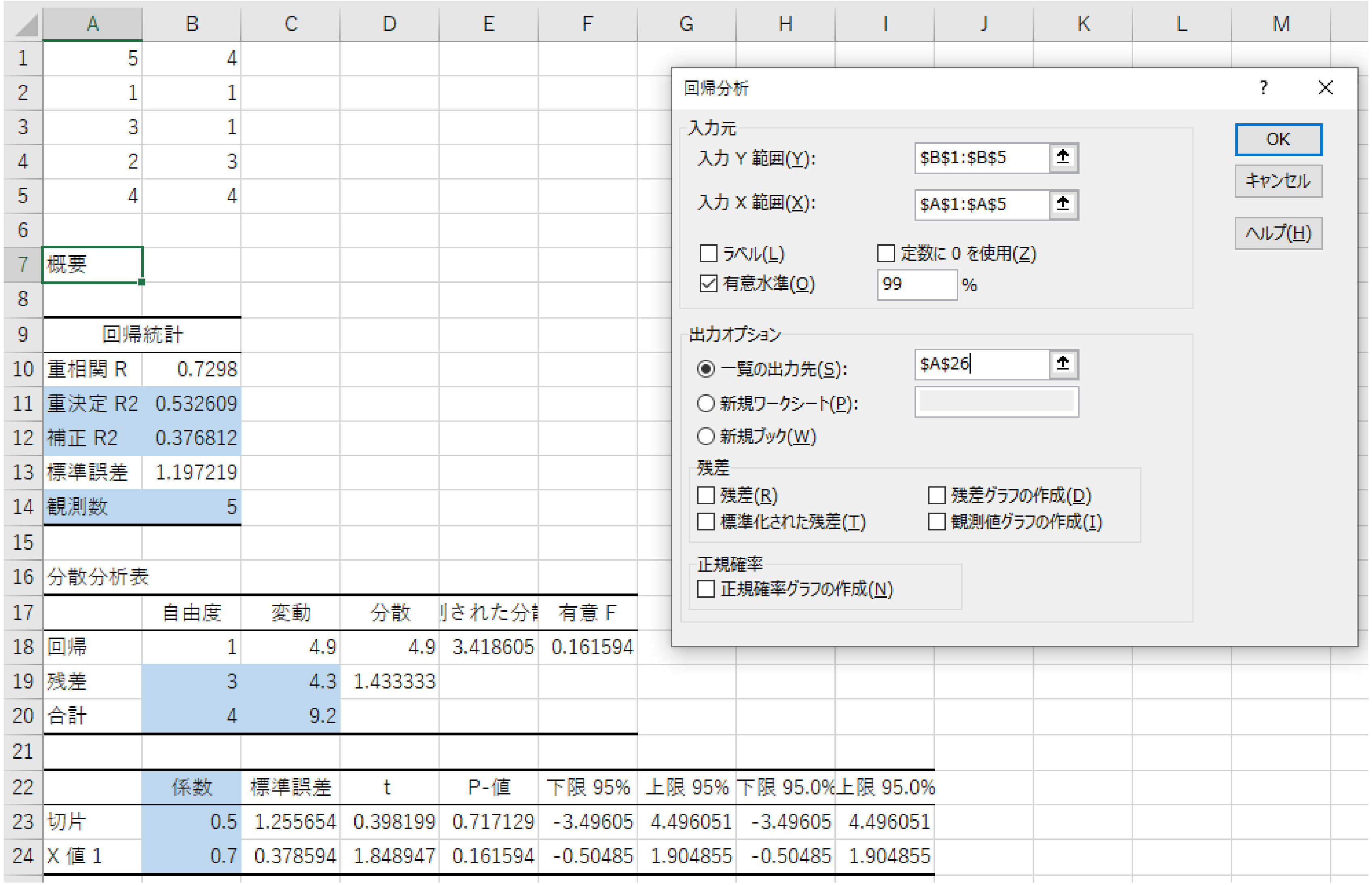
→ 95%信頼区間（上段は α，下段は β の信頼区間）

H 列・I 列（F 列・G 列と同じもの，ただし，変更可能）

● H 列・I 列の変更方法

「有意水準(O)」にチェックを入れ，その横の欄に例えば 99 とタイプする。

「一覧の出力先(S)」の横の欄には A26 から出力するように指定する（出力結果が重ならないように）。



「OK」ボタンを押すと，

26	概要								
27									
28	回帰統計								
29	重相関 R	0.7298							
30	重決定 R2	0.532609							
31	補正 R2	0.376812							
32	標準誤差	1.197219							
33	観測数	5							
34									
35	分散分析表								
36		自由度	変動	分散	された分	有意 F			
37	回帰	1	4.9	4.9	3.418605	0.161594			
38	残差	3	4.3	1.433333					
39	合計	4	9.2						
40									
41		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 99.0%	上限 99.0%
42	切片	0.5	1.255654	0.398199	0.717129	-3.49605	4.496051	-6.83416	7.83416
43	X 値 1	0.7	0.378594	1.848947	0.161594	-0.50485	1.904855	-1.51133	2.911333

と出力される。

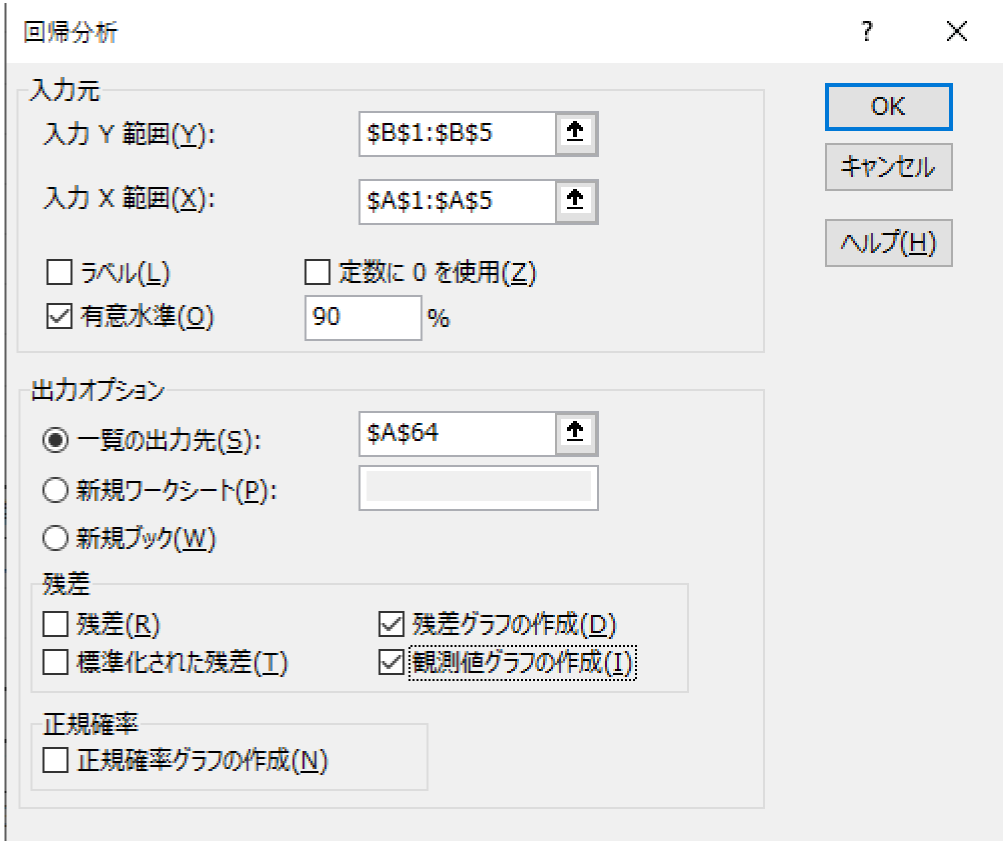
H 列・I 列が下記のように 99%信頼区間に変わる。

下限 99.0%	上限 99.0%
-6.83416	7.83416
-1.51133	2.911333

90%信頼区間にすると，次のページ。

45	概要								
46									
47	回帰統計								
48	重相関 R	0.7298							
49	重決定 R2	0.532609							
50	補正 R2	0.376812							
51	標準誤差	1.197219							
52	観測数	5							
53									
54	分散分析表								
55		自由度	変動	分散	された分散	有意 F			
56	回帰	1	4.9	4.9	3.418605	0.161594			
57	残差	3	4.3	1.433333					
58	合計	4	9.2						
59									
60		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 90.0%	上限 90.0%
61	切片	0.5	1.255654	0.398199	0.717129	-3.49605	4.496051	-2.45501	3.45501
62	X 値 1	0.7	0.378594	1.848947	0.161594	-0.50485	1.904855	-0.19097	1.590969

さらに，残差とＹの予測値を出力するためには，「残差グラフの作成(D)」，「観測値グラフの作成(I)」にチェックを入れて，「OK」ボタンを押す。



概要

回帰統計	
重相関 R	0.7298
重決定 R2	0.532609
補正 R2	0.376812
標準誤差	1.197219
観測数	5

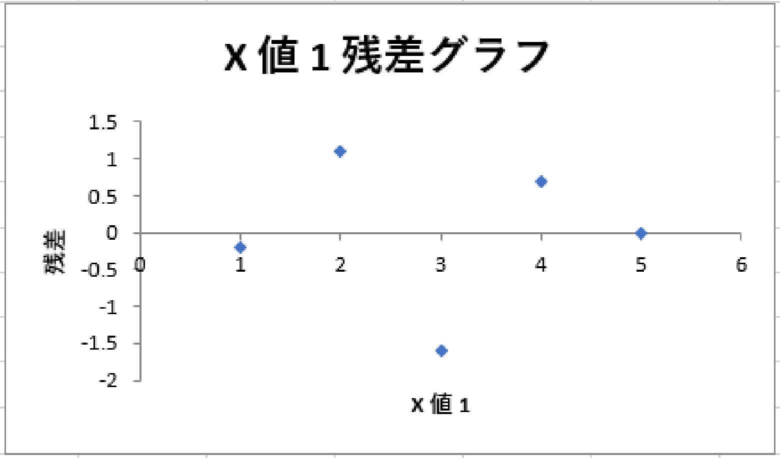
分散分析表

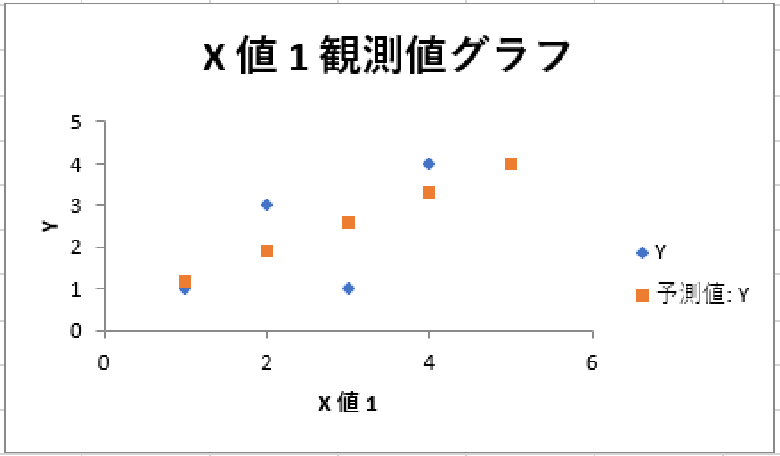
	自由度	変動	分散	された分散	有意 F
回帰	1	4.9	4.9	3.418605	0.161594
残差	3	4.3	1.433333		
合計	4	9.2			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 90.0%	上限 90.0%
切片	0.5	1.255654	0.398199	0.717129	-3.49605	4.496051	-2.45501	3.45501
X 値 1	0.7	0.378594	1.848947	0.161594	-0.50485	1.904855	-0.19097	1.590969

残差出力

観測値	予測値: Y	残差
1	4	0
2	1.2	-0.2
3	2.6	-1.6
4	1.9	1.1
5	3.3	0.7





5.4　確率的モデル：重回帰モデル

n 組のデータ $(Y_i, X_{1i}, X_{2i}, \cdots, X_{ki})$, $i = 1, 2, \cdots, n$ を用いて，k 変数の多重回帰モデルを考える。

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

ただし，X_{ji} は j 番目の説明変数の第 i 番目の観測値を表す。

u_i は誤差項 (または，攪乱項) で，同じ仮定を用いる (すなわち，$u_1, u_2, \cdots, u_n$ は互いに独立に，平均ゼロ，分散 σ^2 の正規分布に従う)。

$\beta_1, \beta_2, \cdots, \beta_k$ は推定されるべきパラメータである。

すべての i について，$X_{1i} = 1$ とすれば，β_1 は定数項として表される。

(* 再掲)　次のような関数 $S(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k)$ を定義する。

$$S(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k) = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2$$

このとき，

$$\min_{\hat{\beta}_1,\, \hat{\beta}_2,\, \cdots,\, \hat{\beta}_k} S(\hat{\beta}_1,\ \hat{\beta}_2,\ \cdots,\ \hat{\beta}_k)$$

となるような $\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ を求める。$\Longrightarrow$ 最小二乗法

最小化のためには，

$$
\begin{aligned}
\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k)}{\partial \hat{\beta}_1} &= 0 \\
\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k)}{\partial \hat{\beta}_2} &= 0 \\
&\vdots \\
\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k)}{\partial \hat{\beta}_k} &= 0
\end{aligned}
$$

を満たす $\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ **となる。**

すなわち，$\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ は，

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})X_{1i} = 0$$

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})X_{2i} = 0$$

$$\vdots$$

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})X_{ki} = 0$$

を満たす。

さらに，

$$
\begin{aligned}
\sum_{i=1}^{n} X_{1i}Y_i &= \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^{n} X_{1i}X_{2i} + \ \cdots \ + \hat{\beta}_k \sum_{i=1}^{n} X_{1i}X_{ki} \\
\sum_{i=1}^{n} X_{2i}Y_i &= \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{2i} + \hat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2 + \ \cdots \ + \hat{\beta}_k \sum_{i=1}^{n} X_{2i}X_{ki} \\
&\vdots \\
\sum_{i=1}^{n} X_{ki}Y_i &= \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{ki} + \hat{\beta}_2 \sum_{i=1}^{n} X_{2i}X_{ki} + \ \cdots \ + \hat{\beta}_k \sum_{i=1}^{n} X_{ki}^2
\end{aligned}
$$

行列表示によって，

$$\begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

が得られ，$\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ についてまとめると，

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix}$$

を解くことになる。