第7章　回帰分析

データ・マイニングI

7-1 回帰分析の考え方と最小2乗法の計算

- 相関分析と回帰分析は極めて密接な関係がある
  - 相関分析と回帰分析の違いは，相関分析では単に2変数間の線形関係の程度を計るのみであるが，回帰分析では変数間の因果関係を前提に議論をすることである

- 回帰分析のなかで最も基本的な最小2乗法を説明する
  - 例として，時間変数（タイム・トレンド）で日経平均を説明する
  - 回帰モデルのパラメータ推定値を具体的に求める方法を説明する

7-1 回帰分析の考え方と最小2乗法の計算

- 回帰分析では以下のような線形の関数関係を前提とし，線形関係（回帰直線）を規定するパラメータ（α，β）を推定する問題を考える

$$Y_t = \alpha + \beta X_t + u_t \ \ (t = 1, 2, \ \dots, T) \qquad (7\text{-}1)$$

  - 変数Y_tを従属変数あるいは被説明変数と呼び，
  - 変数X_tを独立変数あるいは説明変数と呼ぶ
    - 回帰分析では，変数X_tの値を与えれば，変数Y_tの値が一定の確率法則に従って決定されることが前提

7-1 回帰分析の考え方と最小2乗法の計算

$$Y_t = \alpha + \beta X_t + u_t \ (t = 1, 2, \ \dots, T) \qquad (7\text{-}1)$$

- **αは切片**
  - X_tが0のとき、$Y_t \approx \alpha$、つまり被説明変数はα
- **βは直線の傾き**
  - X_tが1動いたときにY_tがどのくらい動くかという感応度
- α, βは回帰直線の形状を決めるパラメータ
- **回帰式におけるu_tは回帰式の誤差項**と呼ばれる確率変数
  - 実際のデータと回帰モデルによって計算された理論値との乖離
  - 推定されたパラメータの統計学的な性質は，誤差項の性質によって大きな影響を受け，誤差項の性質によって利用すべき回帰分析の手法を選択する

15:47:22

7-1 回帰分析の考え方と最小2乗法の計算

- 自然科学系の分析では，実験によってデータを収集した場合には，α，βの値が理論的に予想されたり，真のパラメータに想定を与えることが可能な場合もある

- 経済学関係の分析では，日々の変化を観察したデータを用いる場合には値を明確に想定することは困難であり，理論的に想定されるパラメータの符号条件のみを前提に分析することが多い
  - 具体的なデータの観察を通してα，βの値を推定することが必要となる

7-1 回帰分析の考え方と最小2乗法の計算

- 回帰モデルで誤差項を置くのは，従属変数の変動を説明するための要因は，回帰モデルで明示的に取り上げることのできるものに限られないからである
  - 従属変数に影響を与える可能性のある要因をすべて取り上げることは現実的に不可能なので，理論的に想定されるモデルから得られる値と従属変数の値が一致する保証はない
  - 明示的に取り上げることができない各種の要因をまとめたものが誤差項

【Excel】パラメータ推定値の求め方

- 観測されたデータを用いてパラメータ推定値の求め方を説明する
  1. ワーク･シート上で段階的に，パラメータ推定値を求めてから，データ分析の利用方法を説明
  2. Excelには回帰分析を行うためのデータ分析も用意されている

  – 2003年の日経平均を対象に，タイム･トレンドでこれを説明する回帰モデルを考える

【Excel】パラメータ推定値の求め方

- 以下の図7-1のように，「日次原データ(2003-2011).xls」のファイルからデータをコピーする
  - ここでTIMEとして示されているのは，列Aに入力されている営業日の番号と同一のデータをコピーしたものである
  - 日経平均(2003年)を従属変数とし，TIMEを説明変数とする

	A	B	C	D	E	F	G	H	I	J	K
1			2003年	日経平均(2003年)	TIME	日経平均(2005年)	TIME	日経平均(2008年)	TIME	日経平均(2011年)	TIME
2	1		2003/1/6	8713.33	1	11517.75	1	14691.41	1	10398.10	1
3	2		2003/1/7	8656.50	2	11437.52	2	14500.55	2	10380.77	2
4	3		2003/1/8	8517.80	3	11492.26	3	14528.67	3	10529.76	3
5	4		2003/1/9	8497.93	4	11433.24	4	14599.16	4	10541.04	4
6	5		2003/1/10	8470.45	5	11539.99	5	14388.11	5	10510.68	5
7	6		2003/1/14	8553.06	6	11453.39	6	14110.79	6	10512.80	6
8	7		2003/1/15	8611.75	7	11358.22	7	13972.63	7	10589.76	7
9	8		2003/1/16	8609.17	8	11438.39	8	13504.51	8	10499.04	8
10	9		2003/1/17	8690.25	9	11487.10	9	13783.45	9	10502.86	9
11	10	01	2003/1/20	8558.82	10	11423.26	10	13861.29	10	10518.91	10
12	11		2003/1/21	8708.58	11	11405.34	11	13325.94	11	10557.10	11
13	12		2003/1/22	8611.04	12	11284.77	12	12573.05	12	10437.31	12
14	13		2003/1/23	8790.92	13	11238.37	13	12829.06	13	10274.52	13
15	14		2003/1/24	8731.65	14	11289.49	14	13092.78	14	10345.11	14
16	15		2003/1/27	8609.47	15	11276.91	15	13629.16	15	10464.42	15
17	16		2003/1/28	8525.39	16	11376.57	16	13087.91	16	10401.90	16
18	17		2003/1/29	8331.08	17	11341.31	17	13478.86	17	10478.66	17
19	18		2003/1/30	8316.81	18	11320.58	18	13345.03	18	10360.34	18
20	19		2003/1/31	8339.94	19	11387.59	19	13592.47	19	10237.92	19
21	20		2003/2/3	8500.79	20	11384.40	20	13497.16	20	10274.50	20

図7-1　回帰分析作業(1)

【Excel】パラメータ推定値の求め方

- 変数の記号を変更すると煩雑であるので，変数 Y が日経平均(2003年)を表し，変数XがTIMEを表していると考える
- 両変数間の関係を相関図によって表現すると図7-2のようになる

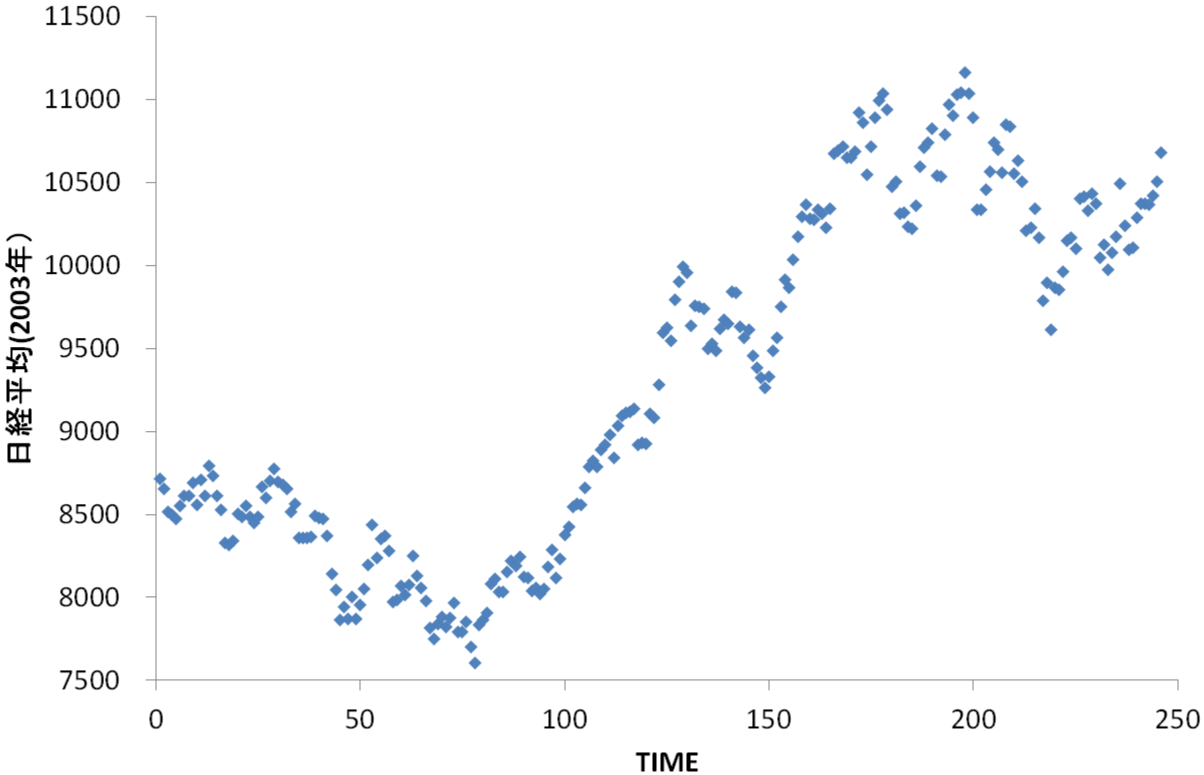


図7-2 回帰分析作業(2)

【Excel】パラメータ推定値の求め方

- 回帰直線というのは，この相関図の中に一定のルールに従って求められたパラメータ推定値，$\hat{\alpha}$と$\hat{\beta}$を用いて描かれる直線，$\hat{Y} = \hat{\alpha} + \hat{\beta}X$である
- 図7-2の画面に回帰直線を書き込んだのが図7-3である
- どのような基準で回帰直線のパラメータ$\hat{\alpha}$，$\hat{\beta}$を決めるのかが各種の回帰分析の手法である

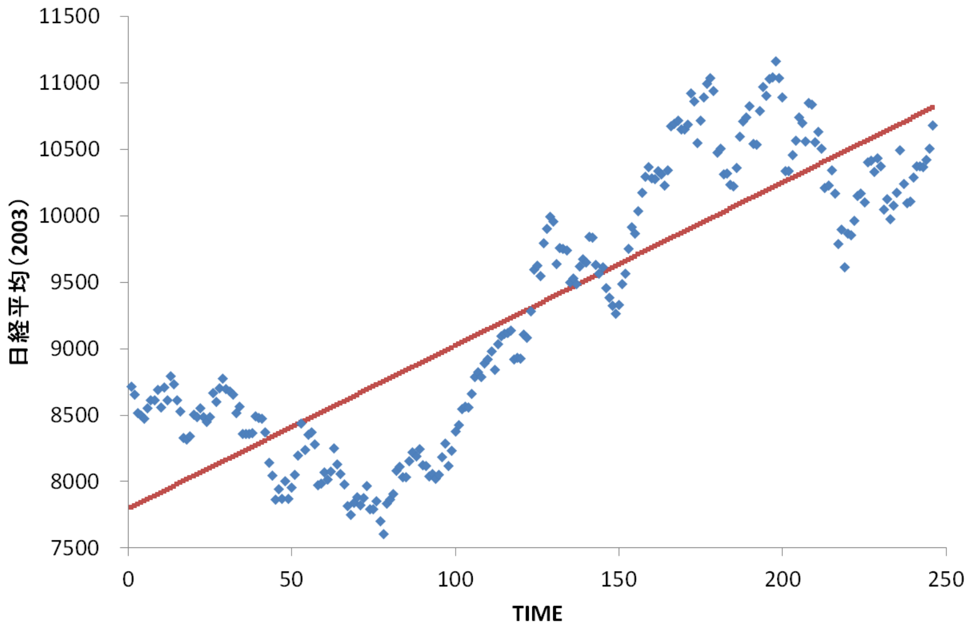


図7-3　回帰分析作業(3)

【Excel】パラメータ推定値の求め方

- 日経平均を初めとする各種金融指標の日次データを1年間を通して観察すると，年によって上昇傾向を示す年もあれば低下傾向を示す年もある
  - 回帰直線のパラメータ推定値$\hat{\beta}$が正であればその年の指標が上昇傾向であることを示し，
  - $\hat{\beta}$が負であればその年の指標が低下傾向であったことを示している

- 利用している説明変数は1営業日につき1つづつ増加する変数であるので，
  - パラメータ$\hat{\beta}$の値は平均的に1営業日あたりどれだけ従属変数の値が増加したか減少したかを表している

【Excel】 パラメータ推定値の求め方

- 回帰分析の手法の中で最も一般的に利用される最小2乗法を説明する
- 最小2乗法とは,
  - 現実のY_tと
  - 各X_tを1期からT期まで与えた場合にこれに対応する回帰直線上の値$\widehat{Y_t} = \hat{\alpha} + \hat{\beta} X_t$の差

  ⇒ 残差$e_t = Y_t - \widehat{Y_t}$を2乗した残差平方を全てのデータに関して合計した残差平方和が最小になるように $\hat{\alpha}$と$\hat{\beta}$を決めるという基準である
- これを数式的で表現すると,

$$\min_{(\hat{\alpha},\hat{\beta})} \sum_{t=1}^{T} e_t^2 = \min_{(\hat{\alpha},\hat{\beta})} \sum_{t=1}^{T} \left(Y_t - \hat{\alpha} - \hat{\beta} X_t\right)^2 \qquad (7\text{-}2)$$

となる

$\hat{\alpha}$, $\hat{\beta}$の導出

- まず，(7-2)式の右辺の括弧内は以下のように書きかえることができる.

$$Y_t - \hat{\alpha} - \hat{\beta}X_t = Y_t - \hat{\alpha} - \hat{\beta}X_t - \left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right) + \left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right)$$

$$= (Y_t - \bar{Y}) - \hat{\beta}(X_t - \bar{X}) + \left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right) \quad (7\text{-}3)$$

- この式の両辺を2乗すると,

$$\left(Y_t - \hat{\alpha} - \hat{\beta}X_t\right)^2$$

$$= (Y_t - \bar{Y})^2 + \hat{\beta}^2(X_t - \bar{X})^2 + \left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right)^2 - 2\hat{\beta}(Y_t - \bar{Y})(X_t - \bar{X})$$

$$-2(X_t - \bar{X})\left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right) + 2(Y_t - \bar{Y})\left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right) \quad (7\text{-}4)$$

という関係が得られる

$\hat{\alpha}$, $\hat{\beta}$の導出

- (7-2)式に示した残差平方の合計を計算するため(7-4)式の両辺を$t = 1, 2, \dots, T$まで合計すると

$$\sum_{t=1}^{T}\left(Y_t - \hat{\alpha} - \hat{\beta}X_t\right)^2$$

$$= \sum_{t=1}^{T}(Y_t - \bar{Y})^2 + \sum_{t=1}^{T}\hat{\beta}^2(X_t - \bar{X})^2 + \sum_{t=1}^{T}\left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right)^2$$

$$-2\hat{\beta}\sum_{t=1}^{T}(Y_t - \bar{Y})(X_t - \bar{X}) - 2\sum_{t=1}^{T}(X_t - \bar{X})\left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right)$$

$$+2\textstyle\sum_{t=1}^{T}(Y_t - \bar{Y})\left(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right) \qquad (7\text{-}5)$$

15:47:23

$\hat{\alpha}$, $\hat{\beta}$の導出

- ここで，(7-5)式の最後の2項は，

$$-2\sum_{t=1}^{T}(X_t-\bar{X})(\bar{Y}-\hat{\alpha}-\hat{\beta}\bar{X})=-2(\bar{Y}-\hat{\alpha}-\hat{\beta}\bar{X})\sum_{t=1}^{T}(X_t-\bar{X})=0$$

$$2\sum_{t=1}^{T}(Y_t-\bar{Y})(\bar{Y}-\hat{\alpha}-\hat{\beta}\bar{X})=2(\bar{Y}-\hat{\alpha}-\hat{\beta}\bar{X})\sum_{t=1}^{T}(Y_t-\bar{Y})=0$$

15:47:24

$\hat{\alpha}$，$\hat{\beta}$の導出

- よって，(7-5)式は，

$$
\begin{aligned}
&\sum_{t=1}^{T}\left(Y_t-\hat{\alpha}-\hat{\beta}X_t\right)^2\\
&=\sum_{t=1}^{T}(Y_t-\bar{Y})^2+\sum_{t=1}^{T}\hat{\beta}^2(X_t-\bar{X})^2+\sum_{t=1}^{T}\left(\bar{Y}-\hat{\alpha}-\hat{\beta}\bar{X}\right)^2\\
&\quad -2\hat{\beta}\sum_{t=1}^{T}(Y_t-\bar{Y})(X_t-\bar{X})\\
&=S_Y^2+\hat{\beta}^2S_Y^2+T\left(\bar{Y}-\hat{\alpha}-\hat{\beta}\bar{X}\right)^2-2\hat{\beta}S_{YX}\\
&=T\left(\bar{Y}-\hat{\alpha}-\hat{\beta}\bar{X}\right)^2+\left(\hat{\beta}^2S_Y^2-2\hat{\beta}S_{YX}+\frac{S_{YX}^2}{S_X^2}\right)+\left(S_Y^2-\frac{S_{YX}}{S_X^2}\right) \qquad \text{(7-6)}
\end{aligned}
$$

- ここで，S_X^2とS_Y^2は変数XとYの偏差平方であるので，

$$
S_X^2=\sum_{t=1}^{T}(X_t-\bar{X})^2\,,S_Y^2=\sum_{t=1}^{T}(Y_t-\bar{Y})^2\,,S_{YX}=\sum_{t=1}^{T}(Y_t-\bar{Y})(X_t-\bar{X})
$$

$\hat{\alpha}$，$\hat{\beta}$の導出

- (7-6)式における第3項は$\hat{\alpha}$と$\hat{\beta}$を含んでいないので，$\hat{\alpha}$と$\hat{\beta}$にどのような値を入れても変化しない
- また，第1項は括弧内の式を2乗したものなので負にはならない
- 第2項も，

$$\left(\hat{\beta} S_{YX} - \frac{S_{YX}}{S_X}\right)^2$$

と書きかえられるので負にはならない

$\hat{\alpha}$, $\hat{\beta}$の導出

- よって，初めの2項が共に0になったときに残差平方和は最小になる．このための条件は，

$$\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X} = 0$$

$$\hat{\beta} S_X - \frac{S_{YX}}{S_X} = 0$$

の両者が成り立つことである

- これからパラメータ推定値の値を求めると，

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$\hat{\beta} = \frac{S_{YX}}{S_X^2}$$

$\hat{\alpha}$，$\hat{\beta}$の導出

- 偏微分を知っている読者は，(7-1)式を$\hat{\alpha}$，$\hat{\beta}$に関して偏微分すれば，より簡便に，$\hat{\alpha}$，$\hat{\beta}$を求めることができる
  - 参考までに以下で，偏微分によるパラメータ推定値の導出法を示す
- (7-2)式を$\hat{\alpha}$，$\hat{\beta}$に関して偏微分して0とおけば（最小値の導出），

$$\frac{\partial\left(\sum_{t=1}^{T} e_t^2\right)}{\partial\hat{\alpha}} = 2\left(\sum_{t=1}^{T} \bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right)(-1) = 0 \qquad (7\text{-}7)$$

$$\frac{\partial\left(\sum_{t=1}^{T} e_t^2\right)}{\partial\hat{\beta}} = 2\left(\sum_{t=1}^{T} \bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}\right)(-X_t) = 0 \qquad (7\text{-}8)$$

の2本の式が導かれる

- これらを連立方程式として解くことによって，先と同様に$\hat{\alpha}$と$\hat{\beta}$を求めることができる

$\hat{\alpha}$, $\hat{\beta}$の導出

- 最小2乗法によってパラメータ推定値$\hat{\alpha}$, $\hat{\beta}$を求めるためには，以下の２本の式に従って計算する必要があることは理解されたことを前提とする

$$\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X} = 0 \qquad (7\text{-}9)$$

$$\hat{\beta}S_X - \frac{s_{YX}}{s_X} = 0 \qquad (7\text{-}10)$$

- (7-10)式を書き直すと，

$$\hat{\beta} = \frac{\sum_{t=1}^{T}(Y_t - \bar{Y})(X_t - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

であるので，$\hat{\alpha}$, $\hat{\beta}$を求めるためには，6-2で求めた相関係数の計算で利用したのとほぼ同じ値を計算すればよい

【Excel】 パラメータ推定値の求め方

- 図7-1に示したワーク・シートの列Eの後ろに5列挿入した上で，最小2乗法によるパラメータ推定の計算作業を展開する

	A	B	C	D	E	F	G	H	I	J	K
1			2003年	日経平均(2003年	TIME	日経平均(2005年)	TIME	日経平均(2008年)	TIME	日経平均(2011年)	TIME
2	1		2003/1/6	8713.33	1	11517.75	1	14691.41	1	10398.10	1
3	2		2003/1/7	8656.50	2	11437.52	2	14500.55	2	10380.77	2
4	3		2003/1/8	8517.80	3	11492.26	3	14528.67	3	10529.76	3
5	4		2003/1/9	8497.93	4	11433.24	4	14599.16	4	10541.04	4
6	5		2003/1/10	8470.45	5	11539.99	5	14388.11	5	10510.68	5
7	6		2003/1/14	8553.06	6	11453.39	6	14110.79	6	10512.80	6
8	7		2003/1/15	8611.75	7	11358.22	7	13972.63	7	10589.76	7
9	8		2003/1/16	8609.17	8	11438.39	8	13504.51	8	10499.04	8
10	9		2003/1/17	8690.25	9	11487.10	9	13783.45	9	10502.86	9
11	10	01	2003/1/20	8558.82	10	11423.26	10	13861.29	10	10518.91	10
12	11		2003/1/21	8708.58	11	11405.34	11	13325.94	11	10557.10	11
13	12		2003/1/22	8611.04	12	11284.77	12	12573.05	12	10437.31	12
14	13		2003/1/23	8790.92	13	11238.37	13	12829.06	13	10274.52	13
15	14		2003/1/24	8731.65	14	11289.49	14	13092.78	14	10345.11	14
16	15		2003/1/27	8609.47	15	11276.91	15	13629.16	15	10464.42	15
17	16		2003/1/28	8525.39	16	11376.57	16	13087.91	16	10401.90	16
18	17		2003/1/29	8331.08	17	11341.31	17	13478.86	17	10478.66	17
19	18		2003/1/30	8316.81	18	11320.58	18	13345.03	18	10360.34	18
20	19		2003/1/31	8339.94	19	11387.59	19	13592.47	19	10237.92	19
21	20		2003/2/3	8500.79	20	11384.40	20	13497.16	20	10274.50	20

図7-1　回帰分析作業(1)

【Excel】パラメータ推定値の求め方

- 最小2乗法によってパラメータを推定するプロセスを図7-4に計算式を示す

	A	B	C	D	E	F	G	H	I	J
1			2003年	日経平均(2003年)	TIME	日経平均偏差(2003年)	TIME偏差	偏差積	日経平均偏差平方	TIME偏差平方
2	1		37627	8713.33	1	=D2-AVERAGE(D$2:D$247)	=E2-AVERAGE(E$2:E$247)	=F2*G2	=F2*F2	=G2*G2
3	2		37628	8656.5	2	=D3-AVERAGE(D$2:D$247)	=E3-AVERAGE(E$2:E$247)	=F3*G3	=F3*F3	=G3*G3
4	3		37629	8517.8	3	=D4-AVERAGE(D$2:D$247)	=E4-AVERAGE(E$2:E$247)	=F4*G4	=F4*F4	=G4*G4
5	4		37630	8497.93	4	=D5-AVERAGE(D$2:D$247)	=E5-AVERAGE(E$2:E$247)	=F5*G5	=F5*F5	=G5*G5
6	5		37631	8470.45	5	=D6-AVERAGE(D$2:D$247)	=E6-AVERAGE(E$2:E$247)	=F6*G6	=F6*F6	=G6*G6
7	6		37635	8553.06	6	=D7-AVERAGE(D$2:D$247)	=E7-AVERAGE(E$2:E$247)	=F7*G7	=F7*F7	=G7*G7
8	7		37636	8611.75	7	=D8-AVERAGE(D$2:D$247)	=E8-AVERAGE(E$2:E$247)	=F8*G8	=F8*F8	=G8*G8
9	8		37637	8609.17	8	=D9-AVERAGE(D$2:D$247)	=E9-AVERAGE(E$2:E$247)	=F9*G9	=F9*F9	=G9*G9
10	9		37638	8690.25	9	=D10-AVERAGE(D$2:D$247)	=E10-AVERAGE(E$2:E$247)	=F10*G10	=F10*F10	=G10*G10
11	10	01	37641	8558.82	10	=D11-AVERAGE(D$2:D$247)	=E11-AVERAGE(E$2:E$247)	=F11*G11	=F11*F11	=G11*G11
12	11		37642	8708.58	11	=D12-AVERAGE(D$2:D$247)	=E12-AVERAGE(E$2:E$247)	=F12*G12	=F12*F12	=G12*G12
13	12		37643	8611.04	12	=D13-AVERAGE(D$2:D$247)	=E13-AVERAGE(E$2:E$247)	=F13*G13	=F13*F13	=G13*G13
14	13		37644	8790.92	13	=D14-AVERAGE(D$2:D$247)	=E14-AVERAGE(E$2:E$247)	=F14*G14	=F14*F14	=G14*G14
15	14		37645	8731.65	14	=D15-AVERAGE(D$2:D$247)	=E15-AVERAGE(E$2:E$247)	=F15*G15	=F15*F15	=G15*G15
16	15		37648	8609.47	15	=D16-AVERAGE(D$2:D$247)	=E16-AVERAGE(E$2:E$247)	=F16*G16	=F16*F16	=G16*G16
17	16		37649	8525.39	16	=D17-AVERAGE(D$2:D$247)	=E17-AVERAGE(E$2:E$247)	=F17*G17	=F17*F17	=G17*G17
18	17		37650	8331.08	17	=D18-AVERAGE(D$2:D$247)	=E18-AVERAGE(E$2:E$247)	=F18*G18	=F18*F18	=G18*G18
19	18		37651	8316.81	18	=D19-AVERAGE(D$2:D$247)	=E19-AVERAGE(E$2:E$247)	=F19*G19	=F19*F19	=G19*G19
20	19		37652	8339.94	19	=D20-AVERAGE(D$2:D$247)	=E20-AVERAGE(E$2:E$247)	=F20*G20	=F20*F20	=G20*G20
21	20		37655	8500.79	20	=D21-AVERAGE(D$2:D$247)	=E21-AVERAGE(E$2:E$247)	=F21*G21	=F21*F21	=G21*G21
22	21		37656	8484.9	21	=D22-AVERAGE(D$2:D$247)	=E22-AVERAGE(E$2:E$247)	=F22*G22	=F22*F22	=G22*G22

図7-4 回帰分析作業(4)

【Excel】パラメータ推定値の求め方

- 図7-5に計算結果を示す
  – これに従って各自で計算すること

	A	B	C	D	E	F	G	H	I	J
1			2003年	日経平均(200	TIME	日経平均偏差(2	TIME偏差	偏差積	日経平均偏差平方	TIME偏差平方
2	1		2003/1/6	8713.33	1	−593.25	−122.5	72672.7117	351941.5593	15006.25
3	2		2003/1/7	8656.50	2	−650.08	−121.5	78984.3101	422599.6197	14762.25
4	3		2003/1/8	8517.80	3	−788.78	−120.5	95047.5834	622168.5657	14520.25
5	4		2003/1/9	8497.93	4	−808.65	−119.5	96633.2718	653909.3658	14280.25
6	5		2003/1/10	8470.45	5	−836.13	−118.5	99081.0052	699107.7347	14042.25
7	6		2003/1/14	8553.06	6	−753.52	−117.5	88538.2036	567787.3057	13806.25
8	7		2003/1/15	8611.75	7	−694.83	−116.5	80947.3019	482784.0402	13572.25
9	8		2003/1/16	8609.17	8	−697.41	−115.5	80550.4653	486376.002	13340.25
10	9		2003/1/17	8690.25	9	−616.33	−114.5	70569.3987	379858.5099	13110.25
11	10	01	2003/1/20	8558.82	10	−747.76	−113.5	84870.3771	559139.9718	12882.25
12	11		2003/1/21	8708.58	11	−598.00	−112.5	67274.6204	357599.9647	12656.25
13	12		2003/1/22	8611.04	12	−695.54	−111.5	77552.3338	483771.1981	12432.25
14	13		2003/1/23	8790.92	13	−515.66	−110.5	56980.0572	265901.756	12210.25
15	14		2003/1/24	8731.65	14	−574.93	−109.5	62954.4655	330540.6253	11990.25
16	15		2003/1/27	8609.47	15	−697.11	−108.5	75636.0689	485957.648	11772.25
17	16		2003/1/28	8525.39	16	−781.19	−107.5	83977.5623	610252.5447	11556.25
18	17		2003/1/29	8331.08	17	−975.50	−106.5	103890.391	951593.6674	11342.25
19	18		2003/1/30	8316.81	18	−989.77	−105.5	104420.379	979637.974	11130.25
20	19		2003/1/31	8339.94	19	−966.64	−104.5	101013.527	934386.3668	10920.25
21	20		2003/2/3	8500.79	20	−805.79	−103.5	83398.9158	649292.0867	10712.25

図7-5 回帰分析作業(5)

【Excel】パラメータ推定値の求め方

- この作業で得られた結果を利用して各列の下方に展開する計算の計算式を示したのが図7-6

	A	B	C	D	E	F	G	H	I	
1			2003年	日経平均(2003年)	TIME	日経平均偏差(2003年)	TIME偏差	偏差積	日経平均偏差平方	TIME偏差平方
246	245		2003/12/29	10500.62	245	=D246-AVERAGE(D$2:D$247)	=E246-AVERAGE(E$2:E$247)	=F246*G246	=F246*F246	=G246*G246
247	246		2003/12/30	10676.64	246	=D247-AVERAGE(D$2:D$247)	=E247-AVERAGE(E$2:E$247)	=F247*G247	=F247*F247	=G247*G247
248			合計	=SUM(D2:D247)	=SUM(E2:E24	=SUM(F2:F247)	=SUM(G2:G247)	=SUM(H2:H247)	=SUM(I2:I247)	=SUM(J2:J247)
249			平均	=AVERAGE(D2:D247)	=AVERAGE(E	=AVERAGE(F2:F247)	=AVERAGE(G2:G247)	=AVERAGE(H2:H247)	=AVERAGE(I2:I247)	=AVERAGE(J2:J247)
250			分散	=VAR.P(D2:D247)	=VAR.P(E2:E2	=VAR.P(F2:F247)	=VAR.P(G2:G247)	=VAR.P(H2:H247)	=VAR.P(I2:I247)	=VAR.P(J2:J247)
251			標準偏差	=SQRT(D250)	=SQRT(E250)	=SQRT(F250)	=SQRT(G250)	=SQRT(H250)	=SQRT(I250)	=SQRT(J250)
252			傾き	=H248/J248						
253			切片	=D249-D252*E249						

図7-6 回帰分析作業(6)

【Excel】パラメータ推定値の求め方

- 計算結果を示したのが図7-7である
  - これに関しても各自で実行すること
  - 後の作業で利用するため（パラメータ推定値を求めるためには不要だが），従属変数（2003年の日経平均）の偏差平方和の計算が含まれていることにも注意すること

	A	B	C	D	E	F	G	H	I	J	K	L	M
1			2003年	日経平均(2003	TIME	日経平均偏差(2	TIME偏差	偏差積	日経平均偏差平方	TIME偏差平方	予測値	残差	残差平方
246	245		2003/12/29	10500.62	245	1194.04	121.5	145076.2699	1425739.579	14762.25	10796.6351	-296.02	87624.94
247	246		2003/12/30	10676.64	246	1370.06	122.5	167832.7633	1877073.649	15006.25	10808.899	-132.26	17492.43
248			合計	2289417.85	30381	0.00	0.00	15214018.23	259726599.64	1240557.50	2289417.85	0.00	73144074.81
249			平均	9306.58	123.50	0.00	0.00	61845.60	1055799.19	5042.92	9306.58	0.00	297333.64
250			分散	1055799.19	5042.9	1055799.186	5042.916667	1495001745	6.14779E+11	20343798.22	758465.548	297333.637	76904864247
251			標準偏差	1027.52089	71.013	1027.520893	71.01349637	38665.25242	784078.5141	4510.409984	870.899276	545.28308	277317.2628
252			傾き	12.2638557									
253			切片	7791.99044									

図7-7　回帰分析作業(7)

推定された関数

- 推定された関数は$Y = 7791.990 + 12.26X$という形で示すことができる
  - 回帰直線の傾きは正であるので，2003年の日経平均株価は基本的に上昇傾向を示し，1営業日につき平均12.26ずつ株価が上昇したことになる

【Excel】 パラメータの推定方法

- これまでは，Excelのワーク･シート上に計算過程を展開する方法で，最小2乗法のパラメータを推定した

1. 以下では，Excelのデータ分析によって最小2乗法による推定結果を求める方法を説明し，Excelの出力として示される各種の計算結果を説明する
   – 回帰分析の考え方をより詳しく説明する

2. さらに，Excelの関数を用いて，パラメータ推定する方法がある

【Excel】 パラメータの推定方法

- 「データ」,「データ分析」の順にクリックし，データ分析の選択画面で，回帰分析をクリックすると，図7-8が表示される

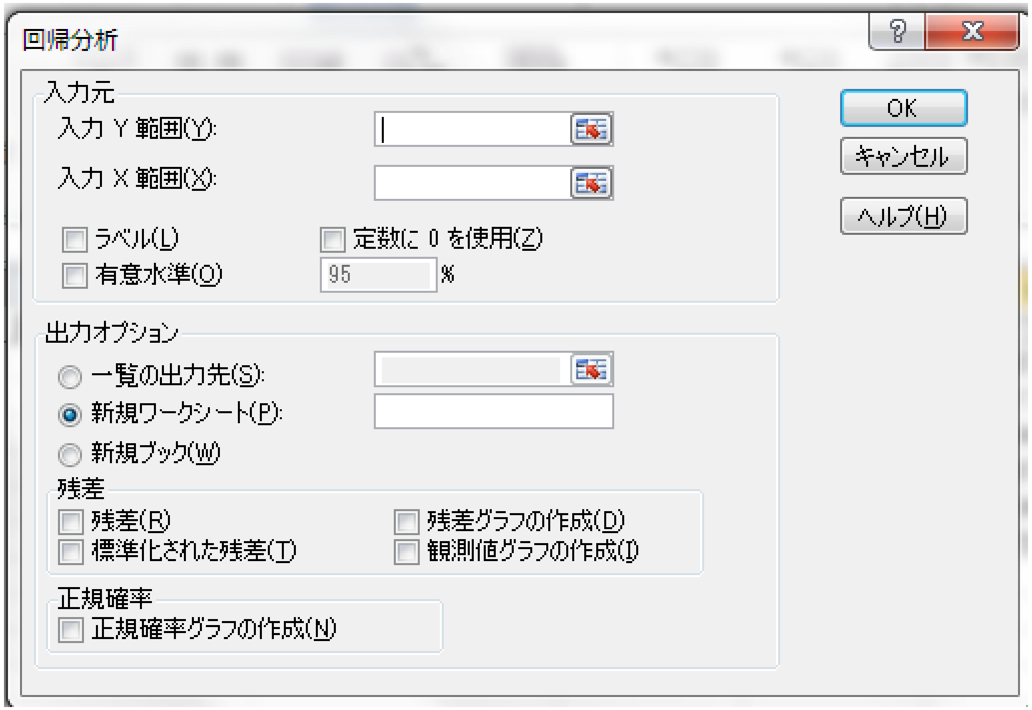


図7-8　回帰分析作業(8)

15:47:29

【Excel】パラメータの推定方法

- 「入力Y範囲(Y)」には，従属変数として利用するデータの範囲を指定するので，「D2:D247」を入力する
- 「入力X範囲(X)」には説明変数として利用するデータの範囲を指定するので，「E2:E247」を入力する
- 「出力オプション」としては，「新規のワーク・シート(P)」を選択し，右横のボックスに「日経平均(2003年)」を入力する
- 残差に関しては，「残差（R）」，「残差グラフの作成(D)」と「観測値グラフの作成(I)」のボックスをマークする

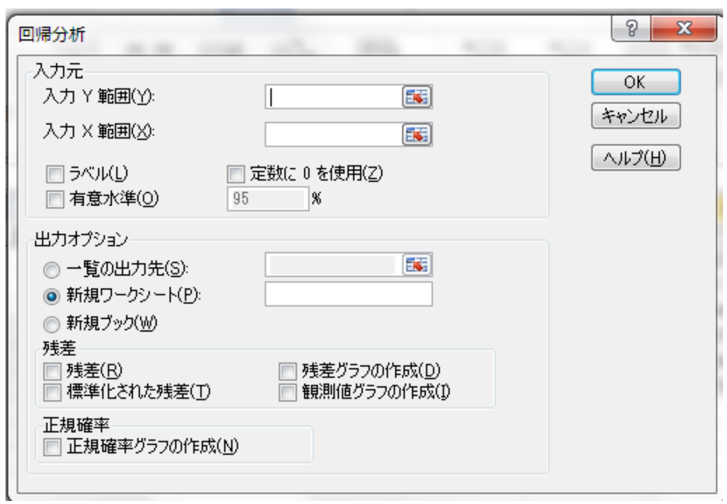


図7-8 回帰分析作業(8)

【Excel】 パラメータの推定方法

- この状況を示したのが図7-9である

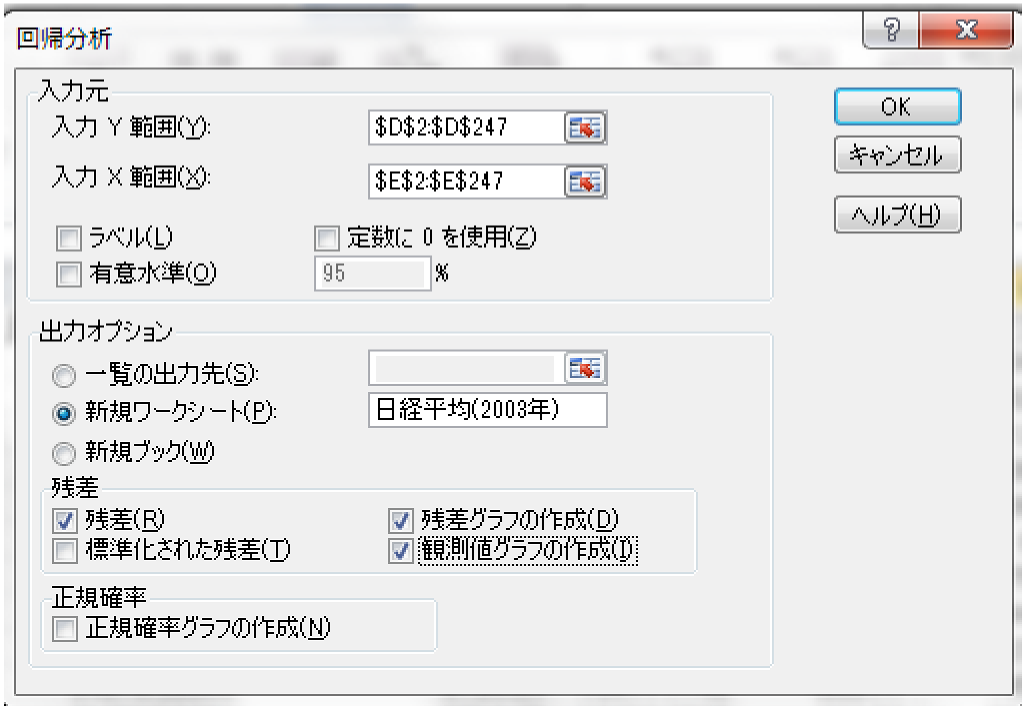


図7-9　回帰分析作業(9)

【Excel】 パラメータの推定方法

- OKをクリックすると，図7-10にその一部が示されているような結果が出力
  - 先に求めた切片と傾きに対応するパラメータ推定値は，セルB17とセルB18に示されている

概要

回帰統計	
重相関 R	0.847573
重決定 R2	0.718381
補正 R2	0.717226
標準誤差	547.5133
観測数	246

分散分析表

	自由度	変動	分散	測された分散	有意 F
回帰	1	1.87E+08	1.87E+08	622.4173	4.34E-69
残差	244	73144075	299770.8		
合計	245	2.6E+08			

	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	7791.99	70.02974	111.2669	4.2E-211	7654.05	7929.93	7654.05	7929.93
X 値 1	12.26386	0.491571	24.94829	4.34E-69	11.29559	13.23212	11.29559	13.23212

残差出力

観測値	予測値: Y	残差
1	7804.254	909.0757
2	7816.518	839.9818
3	7828.782	689.018
4	7841.046	656.8841
5	7853.31	617.1403
6	7865.574	687.4864

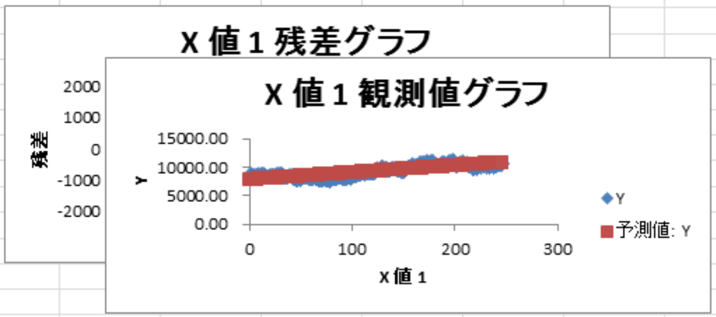


図7-10　回帰分析作業(10)

【Excel】 パラメータの推定方法

- 右側の2つのグラフが重なっていて見にくいので，x値1観測値グラフをクリックして右上に移動し2つのグラフを並べて表示すると図7-11のようにする

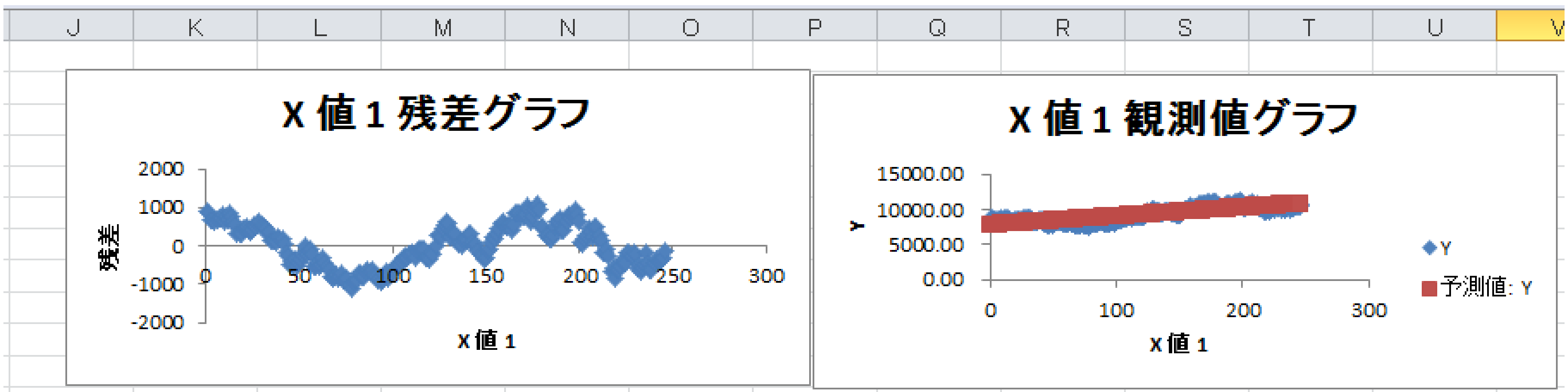


図7-11　回帰分析作業(11)

【Excel】 パラメータの推定方法

- これを見やすいように修正すると，図7-12のように示すことができる

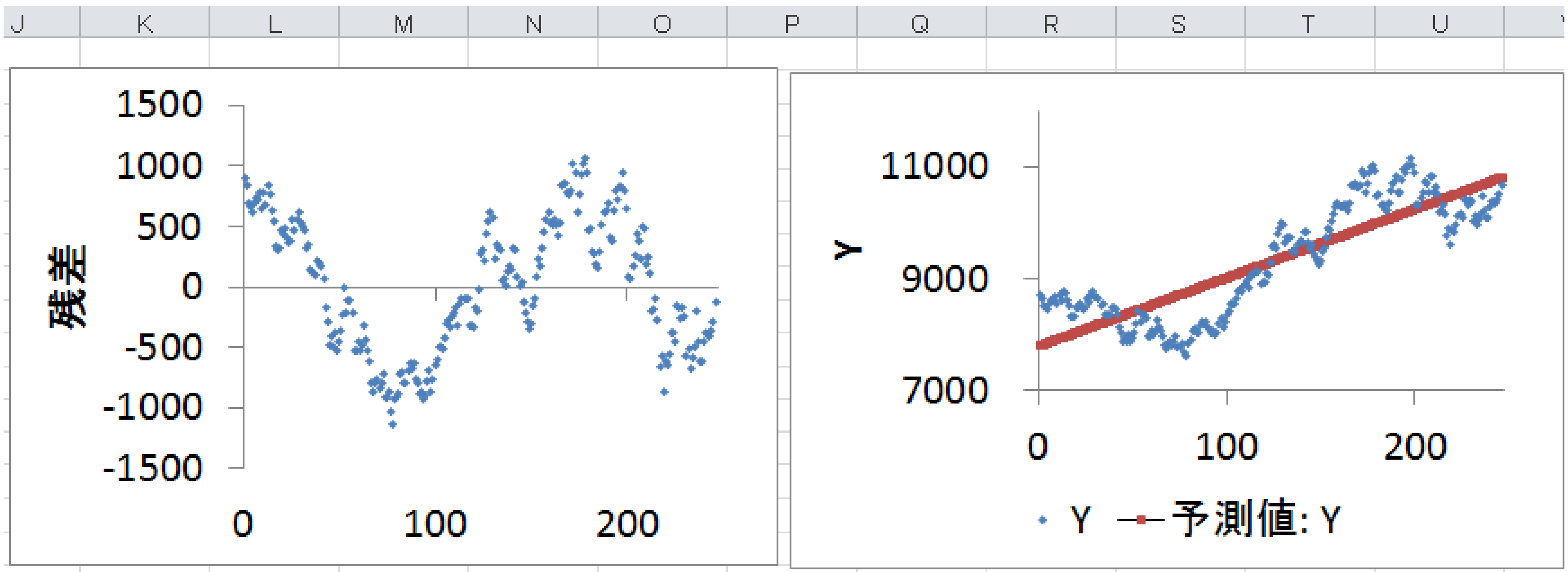


図7-12　回帰分析作業(12)

15:47:32

【Excel】 パラメータの推定方法

- 図7-11と7-12における相違は,
  1. グラフの大きさ,
  2. 画面のタイトル,
  3. 横軸の設定,
  4. 目盛軸の文字等の大きさである
- これらの調整のうち3以外は全て説明してあるので，各自で実行することとし，ここでは3の調整方法のみを説明する

【Excel】パラメータの推定方法

- 横軸にマウス･ポインターを合わせた上で，マウスを右クリックして表示される図7-13で「軸ラベル (A)」を下端/左端に設定し，フォントで，文字の大きさを変更する

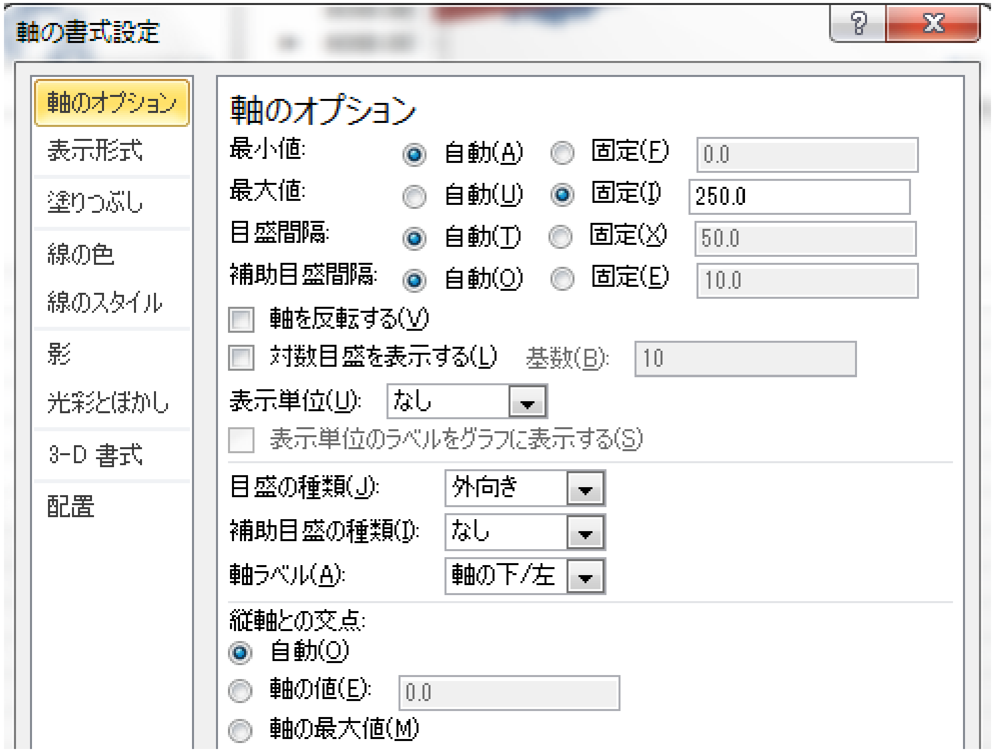


図7-13　回帰分析作業(13)

【Excel】 パラメータの推定方法

- 軸のオプションをクリックして表示される画面で図7-14に示すように入力して「閉じる」をクリックする

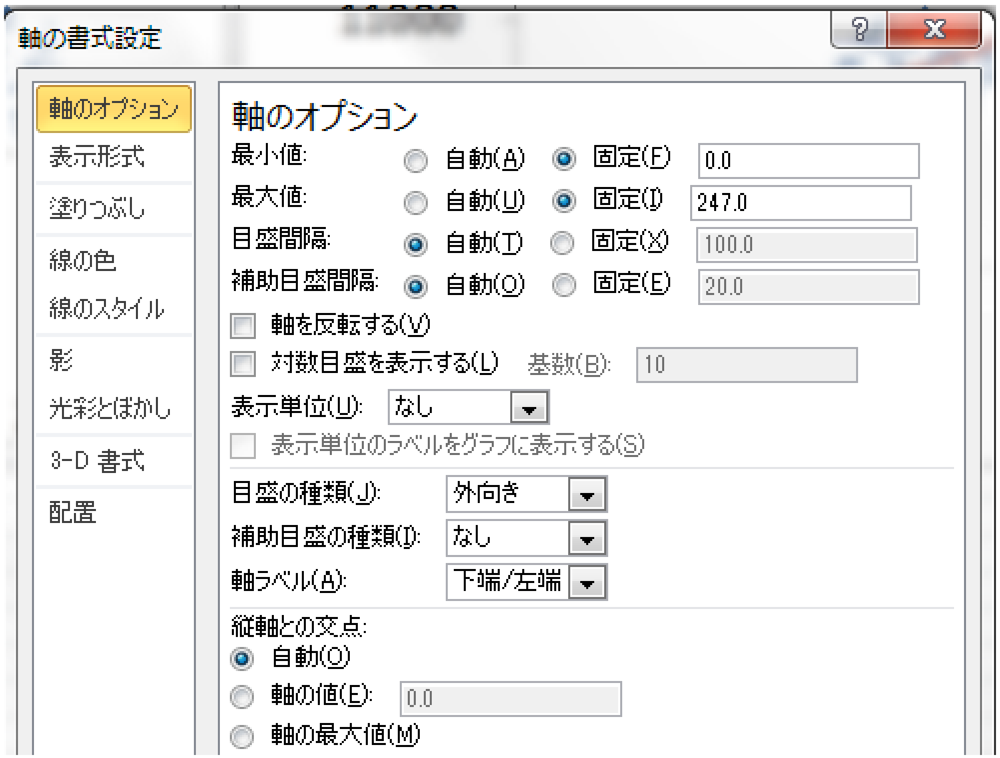


図7-14 回帰分析作業(14)

【Excel】パラメータの推定方法

- 以上の作業で，図7-12に示した図を作成することができたはずである

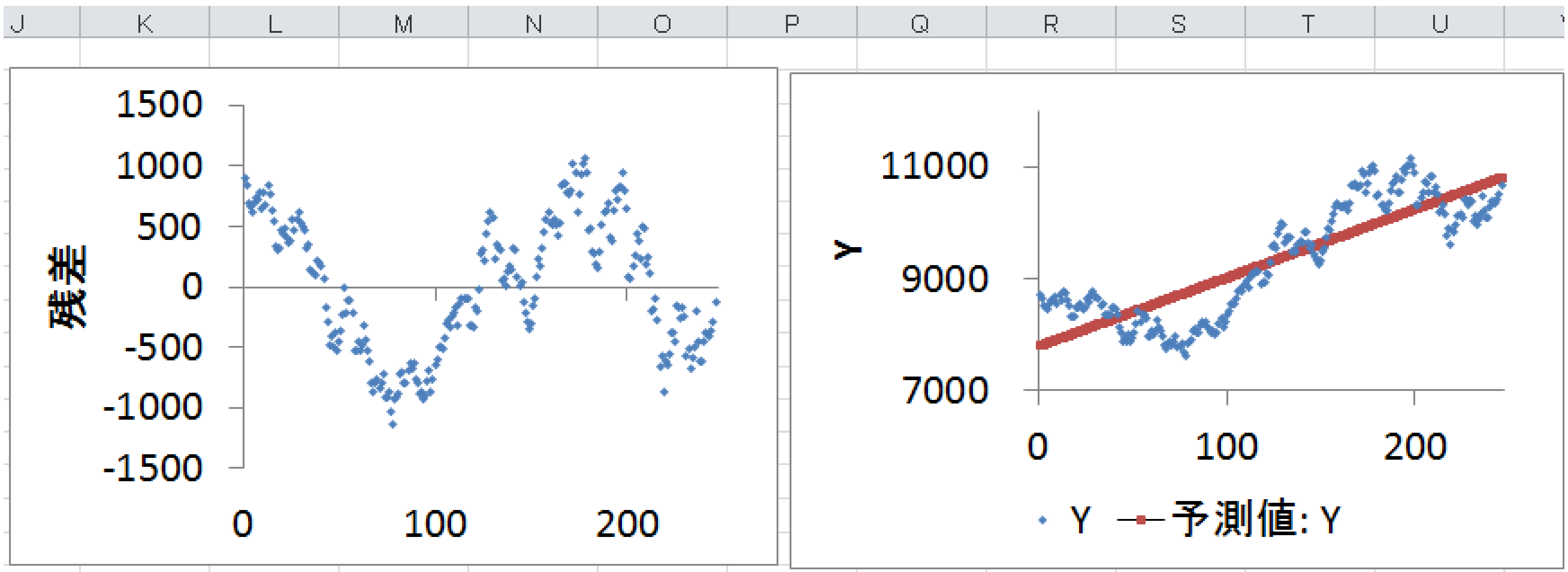


図7-12　回帰分析作業(12)

【練習問題7-1】 提出課題21：2点

- これまで，2003年の日経平均を対象に回帰分析の方法を説明してきた
- Excelの分析ツールによって，2005，2008，2011年の日経平均をTIMEで説明する回帰分析を行い，この結果のワーク･シートをMS-Word文書に貼り付けて見やすいように調整して印刷せよ
- ワーク･シート上で，回帰統計の値を計算し，Excelによる回帰分析結果と比較して確認せよ
- 残差の数値はMS-Wordで出力する必要はないが，計算はしておくこと
- 2つのグラフは共に出力し，年ごとの結果の相違に関して若干のコメントをつけること

【練習問題7-1】 提出課題21:2点

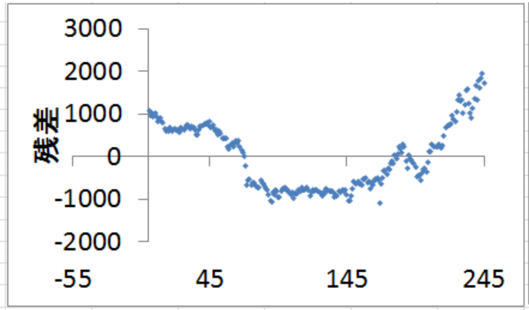
概要

回帰統計	
重相関 R	0.834168
重決定 R2	0.695836
補正 R2	0.694584
標準誤差	761.3562
観測数	245

分散分析表

	自由度	変動	分散	された分散	有意 F
回帰	1	3.22E+08	3.22E+08	555.9101	9.65E-65
残差	243	1.41E+08	579663.3		
合計	244	4.63E+08			

	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	10425.17	97.58109	106.8359	3E-206	10232.95	10617.38	10232.95	10617.38
X 値 1	16.21569	0.687754	23.57775	9.65E-65	14.86097	17.57041	14.86097	17.57041



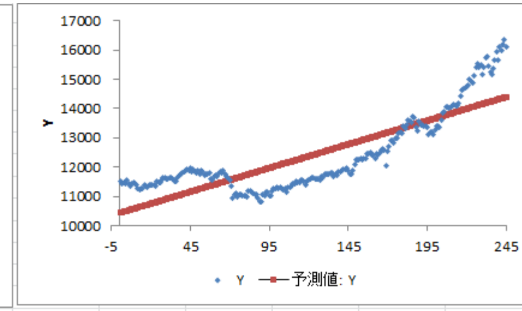


図7A-1　日経平均(2005年)の回帰分析結果と残差等グラフ

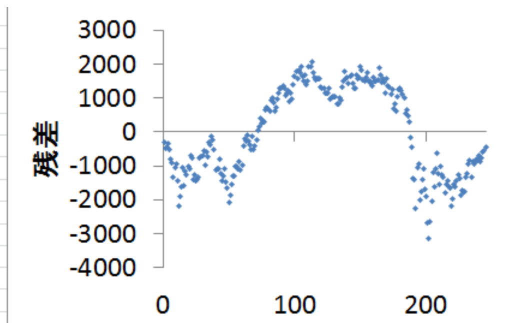
概要

回帰統計	
重相関 R	0.788466
重決定 R2	0.621679
補正 R2	0.620122
標準誤差	1298.444
観測数	245

分散分析表

	自由度	変動	分散	された分散	有意 F
回帰	1	6.73E+08	6.73E+08	399.3117	3.32E-53
残差	243	4.1E+08	1685956		
合計	244	1.08E+09			

	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	15033.69	166.4182	90.33682	5.5E-189	14705.89	15361.5	14705.89	15361.5
X 値 1	-23.4382	1.17292	-19.9828	3.32E-53	-25.7486	-21.1278	-25.7486	-21.1278



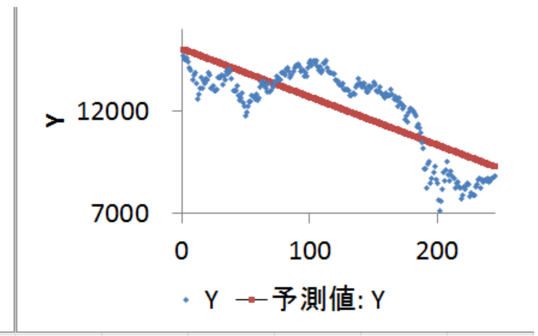


図7A-2　日経平均(2008年)の回帰分析結果と残差等グラフ

【練習問題7-1】 提出課題21：2点

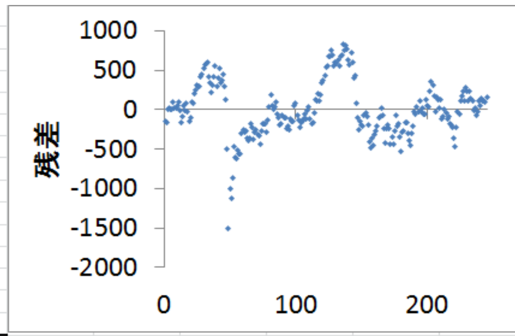
概要

回帰統計	
重相関 R	0.884132
重決定 R2	0.781689
補正 R2	0.78079
標準誤差	346.13
観測数	245

分散分析表

	自由度	変動	分散	測された分散	有意 F
回帰	1	1.04E+08	1.04E+08	870.0904	2.88E-82
残差	243	29112853	119806		
合計	244	1.33E+08			

	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	10559.43	44.3626	238.0254	6.5E-290	10472.04	10646.81	10472.04	10646.81
X 値 1	-9.22288	0.312669	-29.4973	2.88E-82	-9.83877	-8.60699	-9.83877	-8.60699



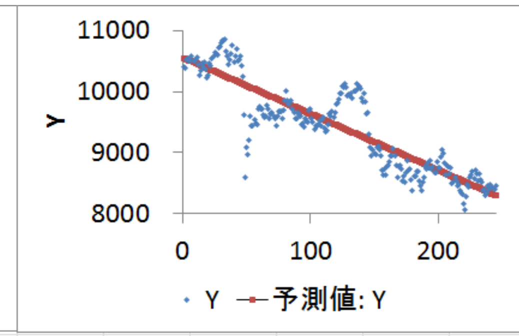


図7A-3　日経平均(2011年)の回帰分析結果と残差等グラフ

15:47:38

7-2 残差の分析

- Excelによる回帰分析の出力は，回帰式のパラメータ推定値だけでなく他にも色々ある
  - パラメータ推定値以外の統計量は，回帰統計，分散分析表という形にまとめられている

- この内，理論値と残差に関連する回帰統計の重相関R，重決定R2および補正R2について説明する
  - 予測値（理論値）と残差をワーク・シート上で計算し，これらの値をExcelのデータ分析によって得られた結果と比較して確認する

7-2　残差の分析

- 最小2乗法の予測値$\hat{Y}_t$とは，推定されたパラメータと説明変数の値を用いて従属変数の理論値を計算したものである
- 残差e_tは，現実値と予測値との差$Y_t - \hat{Y}_t$である
- これらの値を先に最小2乗法によりパラメータ推定値を求めたワーク･シート上で計算する

【Excel】 残差の分析

- セルK1に「予測値」，セルL1に「残差」と書き込む
- セルK2に予測値を計算するため，「＝$D$253+$D$252*E2」を入力して計算し，セルK247までコピーする
- セルL2に「＝D2-K2」を入力して，セルL247までコピーする
- 予測値と現実値，残差それぞれのグラフを示したのが，図7-15と7-16である
- 先のExcelの回帰分析で得られたグラフと比べれば両者は一致している

図7-15　回帰分析作業(15)

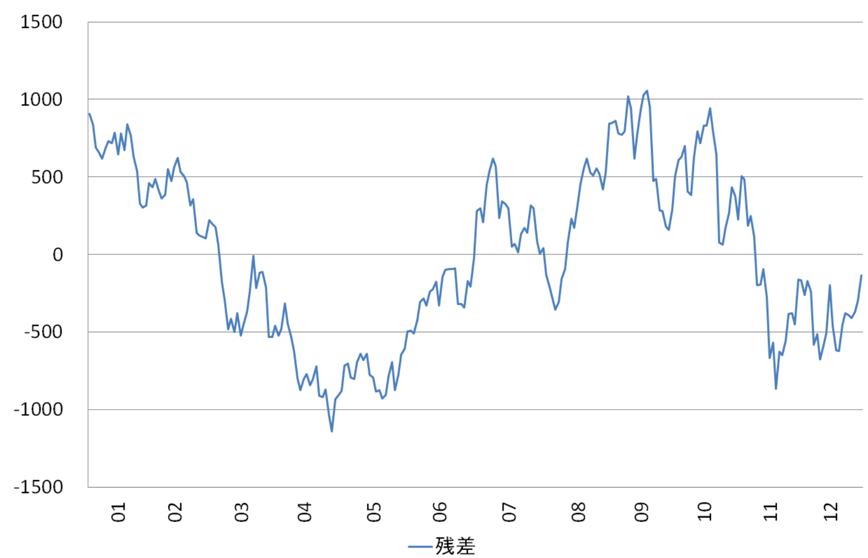


図7-16　回帰分析作業(16)

【Excel】 残差の分析

- **「観測数」は分析に利用されたデータ数**
  - 各年の金融市場の営業日数に対応しており，2003年では246になっている
- **「標準誤差」は，回帰モデルで計算された理論値と現実値との差（残差）の標準偏差**
  - 値が小さいほど回帰モデルの信頼性が高いと判断される
  - ただし，標準誤差の大小はデータの単位のとり方にも影響されるので，実際には基準化された統計量を利用して回帰モデルの信頼性を評価する
- 「重相関R」や「重決定R2」がそうした残差の標準偏差を基準化した統計量である

【Excel】 残差の分析

- ワーク･シート上で標準誤差を計算することは可能であり，すでに残差はワーク･シート上で計算しているので，これを利用する
  - 残差の和は後に証明するように0であり，残差平方を合計したものをサンプル数で割った値が，残差の分散に対応する

【Excel】 標準偏差

- 列Mに1列挿入して残差平方を計算する
- セルM1に残差平方と入力した上で，セルM2に「＝L2^2」を入力して計算し，セルM247までコピーする
- セルM248にセルM2からセルM247の合計を計算し，セルM249にこれの平均を計算すれば，残差の分散になる
- セルM250にセルM2からセルM247の分散を関数を用いて計算する
- セルM249の値の平方根をセルM251に求めれば，これが残差平方和の標準偏差である

【Excel】標準偏差

- セルM251に計算された値は545.283であるのに対して，Excelの回帰分析で求めた標準偏差の値は547.5133であり，完全には一致していない
  - 回帰分析の標準偏差は，通常は残差平方和を回帰分析における自由度で割って計算するからである
  - サンプル数ではなく自由度で割ると，推定された標準偏差が統計学的に望ましい性質である不偏性を満たすことが知られているからである
  - 自由度とは，観測値数－回帰式のパラメータ数である
- セルM250ではセルM248で求めた残差平方和を観測値数の246からパラメータ数の2を引いた244で割って求めることになる
- セルM250にセルM248の値を244で割った値を求め，これの平方根をセルM251に求めれば，547.5133となりExcelの回帰分析による標準誤差の値と一致している

重決定R2　決定係数

- 最小2乗法による推定結果から計算された従属変数の予測値と現実値とは，完全に一致することはなく，残差が存在する（グラフ等によって確認した）
- 決定係数は，この残差がどの程度の大きさであるのかを示す指標であり，
  - 決定係数が大きい場合には回帰式の当てはまりが良く，
  - 小さい場合には当てはまりが悪いことを意味している
- 回帰分析における説明変数の説明力は，従属変数の変動をどの程度説明することができるのかが基準となり，従属変数の変動を示す指標として，従属変数の偏差平方和に対応する全変動（SV_Y）を以下のように定義する

$$SV_Y = \sum_{t=1}^{T}(Y_t - \bar{Y})^2 \qquad (7\text{-}11)$$

- 決定係数によって全変動のうち，
  - どれだけが回帰式によって説明され，
  - どれだけが説明されていないかを評価する

重決定R2　決定係数

- 全変動を，回帰式における現実値と予測値との関係を用いて書きかえると，

$$\sum_{t=1}^{T}(Y_t-\bar{Y})^2=\sum_{t=1}^{T}\left(Y_t-\hat{Y}_t+\hat{Y}_t-\bar{Y}\right)^2$$

$$=\sum_{t=1}^{T}\left(e_t+\hat{Y}_t-\bar{Y}\right)^2$$

$$=\sum_{t=1}^{T}e_t^2+2e_t\left(\hat{Y}_t-\bar{Y}\right)+\left(\hat{Y}_t-\bar{Y}\right)^2$$

$$=\sum_{t=1}^{T}e_t^2+2\sum_{t=1}^{T}e_t\left(\hat{Y}_t-\bar{Y}\right)+\sum_{t=1}^{T}\left(\hat{Y}_t-\bar{Y}\right)^2$$

$$=\sum_{t=1}^{T}e_t^2+\sum_{t=1}^{T}\left(\hat{Y}_t-\bar{Y}\right)^2 \qquad \text{(7-12)}$$

重決定R2　決定係数

- ここで最後の等号が成立する理由は，

$$\sum_{t=1}^{T} e_t \hat{Y}_t = 0 \qquad (7\text{-}13)$$

$$\sum_{t=1}^{T} e_t \bar{Y} = 0 \qquad (7\text{-}14)$$

  の関係が成り立つからである

- (7-14)式は，$\sum_{t=1}^{T} e_t \bar{Y} = \bar{Y} \sum_{t=1}^{T} e_t$と書き換えることができるので，実質的には$\sum_{t=1}^{T} e_t = 0$を証明すればよい
- これは，

$$\begin{aligned}
\sum_{t=1}^{T} e_t &= \sum_{t=1}^{T} \left(Y_t - \hat{\alpha} - \hat{\beta} X_t\right) \\
&= \sum_{t=1}^{T} Y_t - T\hat{\alpha} - \hat{\beta} \sum_{t=1}^{T} X_t \\
&= \sum_{t=1}^{T} Y_t - T\left(\bar{Y} - \hat{\beta}\bar{X}\right) - \hat{\beta} \sum_{t=1}^{T} X_t \\
&= \sum_{t=1}^{T} Y_t - T\bar{Y} + \hat{\beta} T \bar{X} - \hat{\beta} \sum_{t=1}^{T} X_t \\
&= \sum_{t=1}^{T} Y_t - \sum_{t=1}^{T} Y_t + \hat{\beta} \sum_{t=1}^{T} X_t - \hat{\beta} \sum_{t=1}^{T} X_t \\
&= 0
\end{aligned}$$

重決定R2　決定係数

- (7-13)式は，

$$\sum_{t=1}^{T} e_t \hat{Y}_t = \sum_{t=1}^{T} e_t \left(\hat{\alpha} + \hat{\beta} X_t\right)$$

$$= \hat{\alpha} \sum_{t=1}^{T} e_t + \hat{\beta} \sum_{t=1}^{T} e_t X_t$$

と書き換えることできるので，$\sum_{t=1}^{T} e_t X_t = 0$を示せば(7-13)式が成立する

$$\sum_{t=1}^{T} e_t X_t = \sum_{t=1}^{T}\left(Y_t - \hat{\alpha} - \hat{\beta} X_t\right) X_t$$

と書け，これは(7-8)式と同じ式であり，最小2乗法のパラメータを導出するための条件として，0であるので，$\sum_{t=1}^{T} e_t X_t = 0$が示された

15:47:41

重決定R2　決定係数

- 決定係数を求めるためには，囲みの中の(7-12)式の両辺をSV_Yで割った

$$1 = \frac{\sum_{t=1}^{T} e_t^2}{SV_Y} + \frac{\sum_{t=1}^{T}(\hat{Y}_t - \bar{Y})^2}{SV_Y} \qquad (7\text{-}15)$$

＝モデルで説明されない比率＋モデルで説明される比率

の関係を用いる

  – (7-15)式の第1項は，残差平方の合計なので，モデルで説明されない変動に対応する
  – 第2項は，理論値と現実値の平均からの隔たりの平方和なので，モデルによって説明される変動に対応している

重決定R2　決定係数

- **決定係数は，従属変数の変動のうちモデルで説明される比率**を示すので，

$$\hat{R}^2 = \frac{\sum_{t=1}^{T}(\hat{Y}_t - \bar{Y})^2}{SV_Y} = 1 - \frac{\sum_{t=1}^{T} e_t^2}{SV_Y} \qquad (7\text{-}16)$$

  - $0 \leq \hat{R}^2 \leq 1$の関係が成り立つ
  - 決定係数は，予測値と現実値との間の相関係数を2乗した値になっている
- Excelによる回帰分析の出力にある**「重相関」は，この予測値と現実値との間の相関係数**を示している

15:47:41

補正R2　自由度修正済み決定係数

- 補正R2は，自由度修正済み決定係数といわれ，以下の式で計算される（式のkは，切片を含むパラメータの数である）

$$\hat{\hat{R}}^2 = 1 - \left(1 - \hat{R}^2\right)\frac{T-1}{T-k}$$

- **自由度修正済みの決定係数**は，決定係数をモデルの複雑さで調整した値
  - 最小2乗法では，モデルで利用する説明変数の数が多くなるにつれ，決定係数の値が大きくなることが知られている
  - モデルは可能な限り単純な方が望ましいという側面もあるので，決定係数を評価する際に，**利用した説明変数の数で決定係数を調整した指標**がこの自由度修正済みの決定係数である
  - 説明変数を追加したことに意味があり，
    - ➢ 説明力が上昇すれば自由度修正済みの決定係数の値が大きくなり，
    - ➢ 追加しても余り意味がなければ追加したことによりその値が小さくなる

【Excel】 決定係数の確認

- ワーク･シート上で実際に，重相関係数，決定係数，自由度修正済み決定係数を計算し，Excelの分析ツールによる値と同じであることを確認する
- 決定係数を計算するために必要となる情報は，すでにワーク･シート上に計算されている
  - (7-15)式の分母のSV_Yは，従属変数の偏差平方和であり，セルI248に計算されている
  - (7-15)式の分子は，残差平方和であり，列Lに入力されている残差の平方を列Mで計算し，この合計をセルM248に求めた値となる
  - セルC254に決定係数と入力した上で，セルD254に，「＝1-M248/I248」を入力して計算すれば，決定係数の値が0.718381と求まり，この値が図7-10に示したExcel回帰分析結果と一致

【Excel】 重相関係数の確認

- 重相関R：現実値と予測値との間の相関係数を計算するため，セルC255に相関係数と記入した上で，セルD255に「＝CORREL(D2:D247,K2:K247)」を入力して計算すると0.847573となる
  - Excelの回帰分析によって出力されている重相関Rの値と一致している

【Excel】 決定係数の確認

- 決定係数：セルC256に 相関係数2乗と入力した上で，セルD256に，「＝D255^2」と入力すれば，相関係数の2乗が計算されるので，この値とセルD254の決定係数の値が等しいことも確認できる

【Excel】 自由度修正済み決定係数の確認

- 補正R2：セルC257に「自由度修正済み決定係数」と記入した上でセルD257に，「＝1－(1－D254)*((246－1)/(246－2))」と入力すれば，自由度修正済み決定係数が計算される
  - この値は0.717226となり，Excelによる回帰分析結果の補正R2の値と等しい

15:47:42

残差の分布　ヒストグラム

- 残差の分布をヒストグラムの形で考えることも可能
  - 原データの分布を標準正規分布に変換した上でヒストグラムを作成しても標準正規分布の理論的な分布型と大きく異なる場合がある
  - 各年の日経平均の変動は，回帰直線で示される趨勢的な変動とこれとは異なる日々の変化に分けて考えることができる
    - 日次の変化や変化率には両者の要因が含まれている

- 回帰分析の残差を用いて標準正規分布に変換した場合，標準正規分布の理論値とどのような関係になっているのかを確認する

【Excel】 残差の分布　ヒストグラム

- 列Lの下方に，残差の合計，平均，分散，標準偏差を計算しこれらの数値を用いて，列Lに1列挿入し残差を標準正規分布に変換した系列を計算する
- 日経平均のデータを用いて計算した標準正規分布の区間区分と理論分布をコピーし，これを用いて各区間に対応する度数分布を求める

【Excel】 残差の分布　ヒストグラム

- 列Lの下に残差の合計，平均，分散，標準偏差を計算する際の計算式を示したのが図7-17である

	A	B	C	J	K	L
1			2003年	TIME偏差平方	予測値	残差
246	245		2003/12/29	=G246*G246	=$D$253+$D$252*E246	=D246−K246
247	246		2003/12/30	=G247*G247	=$D$253+$D$252*E247	=D247−K247
248			合計	=SUM(J2:J247)	=SUM(K2:K247)	=SUM(L2:L247)
249			平均	=AVERAGE(J2:J247)	=AVERAGE(K2:K247)	=AVERAGE(L2:L247)
250			分散	=VAR.P(J2:J247)	=VAR.P(K2:K247)	=VAR.P(L2:L247)
251			標準偏差	=SQRT(J250)	=SQRT(K250)	=SQRT(L250)
252			傾き			
253			切片			

図7-17　残差確率分布作業(1)

【Excel】 残差の分布 ヒストグラム

- 結果を示したのが図7-18である
- 残差の合計(その結果残差の平均も)が0になっている
  - 先に証明したように理論的な結果である

	A	B	C	J	K	L
1			2003年	TIME偏差平ﾌ	予測値	残差
246	245		2003/12/29	14762.25	10796.6351	-296.02
247	246		2003/12/30	15006.25	10808.899	-132.26
248			合計	1240557.50	2289417.85	0.00
249			平均	5042.92	9306.58	0.00
250			分散	20343798.2	758465.548	297333.637
251			標準偏差	4510.40998	870.899276	545.28308
252			傾き			

図7-18 残差確率分布作業(2)

【Excel】 残差の分布 ヒストグラム

- 上で計算された残差の平均と標準偏差を用いて，残差を標準化する計算式を示したのが図7-19である

	A	B	C	K	L	M	N
1			2003年	予測値	残差	残差平方	残差(標準化)
2	1		37627	=$D$253+$D$252*E2	=D2-K2	=L2^2	=(L2-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
3	2		37628	=$D$253+$D$252*E3	=D3-K3	=L3^2	=(L3-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
4	3		37629	=$D$253+$D$252*E4	=D4-K4	=L4^2	=(L4-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
5	4		37630	=$D$253+$D$252*E5	=D5-K5	=L5^2	=(L5-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
6	5		37631	=$D$253+$D$252*E6	=D6-K6	=L6^2	=(L6-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
7	6		37635	=$D$253+$D$252*E7	=D7-K7	=L7^2	=(L7-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
8	7		37636	=$D$253+$D$252*E8	=D8-K8	=L8^2	=(L8-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
9	8		37637	=$D$253+$D$252*E9	=D9-K9	=L9^2	=(L9-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
10	9		37638	=$D$253+$D$252*E10	=D10-K10	=L10^2	=(L10-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
11	10	01	37641	=$D$253+$D$252*E11	=D11-K11	=L11^2	=(L11-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
12	11		37642	=$D$253+$D$252*E12	=D12-K12	=L12^2	=(L12-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
13	12		37643	=$D$253+$D$252*E13	=D13-K13	=L13^2	=(L13-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
14	13		37644	=$D$253+$D$252*E14	=D14-K14	=L14^2	=(L14-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
15	14		37645	=$D$253+$D$252*E15	=D15-K15	=L15^2	=(L15-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
16	15		37648	=$D$253+$D$252*E16	=D16-K16	=L16^2	=(L16-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
17	16		37649	=$D$253+$D$252*E17	=D17-K17	=L17^2	=(L17-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
18	17		37650	=$D$253+$D$252*E18	=D18-K18	=L18^2	=(L18-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
19	18		37651	=$D$253+$D$252*E19	=D19-K19	=L19^2	=(L19-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
20	19		37652	=$D$253+$D$252*E20	=D20-K20	=L20^2	=(L20-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
21	20		37655	=$D$253+$D$252*E21	=D21-K21	=L21^2	=(L21-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
22	21		37656	=$D$253+$D$252*E22	=D22-K22	=L22^2	=(L22-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))
23	22		37657	=$D$253+$D$252*E23	=D23-K23	=L23^2	=(L23-AVERAGE(L$2:L$247))/(STDEV.P(L$2:L$247))

図7-19 残差確率分布作業(3)

【Excel】 残差の分布　ヒストグラム

- 結果を示したのが図7-20である

	A	B	C	K	L	M	N
1			2003年	予測値	残差	残差平方	残差(標準化)
2	1		1/6/2003	7804.2543	909.08	826418.63	1.67
3	2		1/7/2003	7816.5182	839.98	705569.50	1.54
4	3		1/8/2003	7828.782	689.02	474745.79	1.26
5	4		1/9/2003	7841.0459	656.88	431496.77	1.20
6	5		1/10/2003	7853.3097	617.14	380862.13	1.13
7	6		1/14/2003	7865.5736	687.49	472637.58	1.26
8	7		1/15/2003	7877.8374	733.91	538627.66	1.35
9	8		1/16/2003	7890.1013	719.07	517059.81	1.32
10	9		1/17/2003	7902.3651	787.88	620762.55	1.44
11	10	01	1/20/2003	7914.629	644.19	414982.05	1.18
12	11		1/21/2003	7926.8929	781.69	611034.79	1.43
13	12		1/22/2003	7939.1567	671.88	451427.15	1.23
14	13		1/23/2003	7951.4206	839.50	704759.30	1.54
15	14		1/24/2003	7963.6844	767.97	589771.13	1.41
16	15		1/27/2003	7975.9483	633.52	401349.77	1.16
17	16		1/28/2003	7988.2121	537.18	288560.06	0.99
18	17		1/29/2003	8000.476	330.60	109299.01	0.61
19	18		1/30/2003	8012.7398	304.07	92458.66	0.56
20	19		1/31/2003	8025.0037	314.94	99184.87	0.58
21	20		2/3/2003	8037.2676	463.52	214853.06	0.85
22	21		2/4/2003	8049.5314	435.37	189545.81	0.80

図7-20　残差確率分布作業(4)

【Excel】 残差の分布 ヒストグラム

- 日経平均のファイルから，区間の切れ目と理論分布をコピーし，列Lのデータにこの区間区分を用いて，データ，データ分析，ヒストグラムを適用して度数分布を求め，相対度数を計算する

区分	HSD	2003年残差	2003RES
−3.75	0.00009	0	0
−3.25	0.00049	0	0
−2.75	0.00240	0	0
−2.25	0.00924	0	0
−1.75	0.02783	2	0.00813
−1.25	0.06559	29	0.117886
−0.75	0.12098	37	0.150407
−0.25	0.17467	42	0.170732
0.25	0.19741	27	0.109756
0.75	0.17467	38	0.154472
1.25	0.12098	41	0.166667
1.75	0.06559	27	0.109756
2.25	0.02783	3	0.012195
2.75	0.00924	0	0
3.25	0.00240	0	0
3.75	0.00049	0	0
	0.00009	0	0
SUM	1.00000	246.00000	1

図7-21 残差確率分布作業(5)

【Excel】 残差の分布 ヒストグラム

- 相対度数のデータを用いて折れ線グラフで示す
  - ここでも山が2つに分割しているが，112ページの図5-44に比べると理論分布との差が小さくなっている

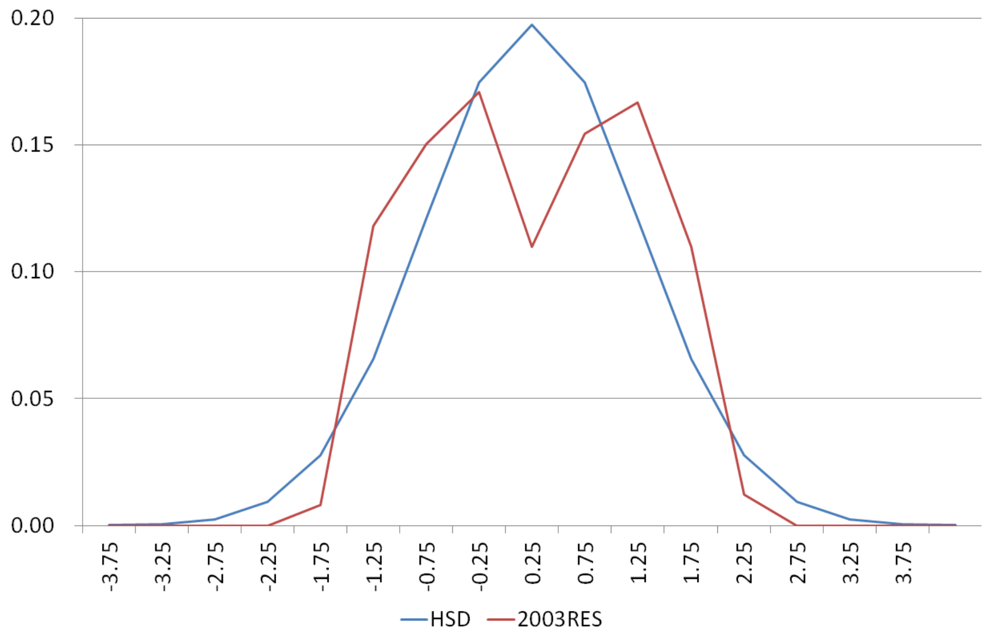


図7-22 残差確率分布作業(6)

7-3　各種金融指標のタイム・トレンドと残差 ―プロジェクトIII

1. 6-1では，日次データにより，日経平均と長期金利との間の相関係数を計算した．そこでは年毎に相関係数の値が大きく異なっていた．より長期の関係を確認するため，暦年データと月次データにより，日経平均と長期金利との間の相関係数を，1980年から2011年及び1986年から2011年に関して計算し，4つの相関係数行列を示せ．期間の相違による結果の相違に関してコメントせよ
2. これまでは，2003年の日経平均を用いて説明してきたが，2005，2008，2011年の日経平均を用いてタイム・トレンドによる回帰分析を実行し，最小2乗法の残差を計算し，これを標準正規分布に変換して，相対度数を理論分布と併せてグラフ化し，これらの結果を評価せよ．提出は2011年の結果のみでよい
3. TOPIX，長期金利，円対ドル為替レートに関しても，TIMEで説明する回帰分析を行い，残差を標準正規分布に変換して理論分布と比較して評価せよ．（結果の提出は2011年分のみで良い）

7-3　各種金融指標のタイム･トレンドと残差 ―プロジェクトIII

- 1980年から2006年の値と1986年から2006年の値が大きく異なっている
- 暦年データを用いた場合にも月次データを用いた場合にも共通に認められる
  - 日本の金融市場における株価と長期金利との間の関係には，1985年のプラザ合意による円高が進行する過程で大きく構造変化が発生した
  - 散布図を描くと明瞭に読み取れる
  - プラザ合意後の日本の金融市場にどのような変化が発生したのかを考えることは，興味深い問題である

	日経平均株価	10年国債応募者利回り
日経平均株価	1	
10年国債応募者利回り	0.085397445	1

1980年-2011年（暦年）

	日経平均株価	10年国債応募者利回り
日経平均株価	1	
10年国債応募者利回り	0.080718302	1

1980年-2011年（月次）

	日経平均株価	10年国債応募者利回り
日経平均株価	1	
10年国債応募者利回り	0.816848057	1

1986年-2011年（暦年）

	日経平均株価	10年国債応募者利回り
日経平均株価	1	
10年国債応募者利回り	0.778152202	1

1986年-2011年（月次）

7-3 各種金融指標のタイム･トレンドと残差 —プロジェクトIII

- 2005年ではグラフの形状に変化が認められる
- 2005年の回帰分析では祁提携数の値も傾きのパラメータ推定値も大きな値となっている

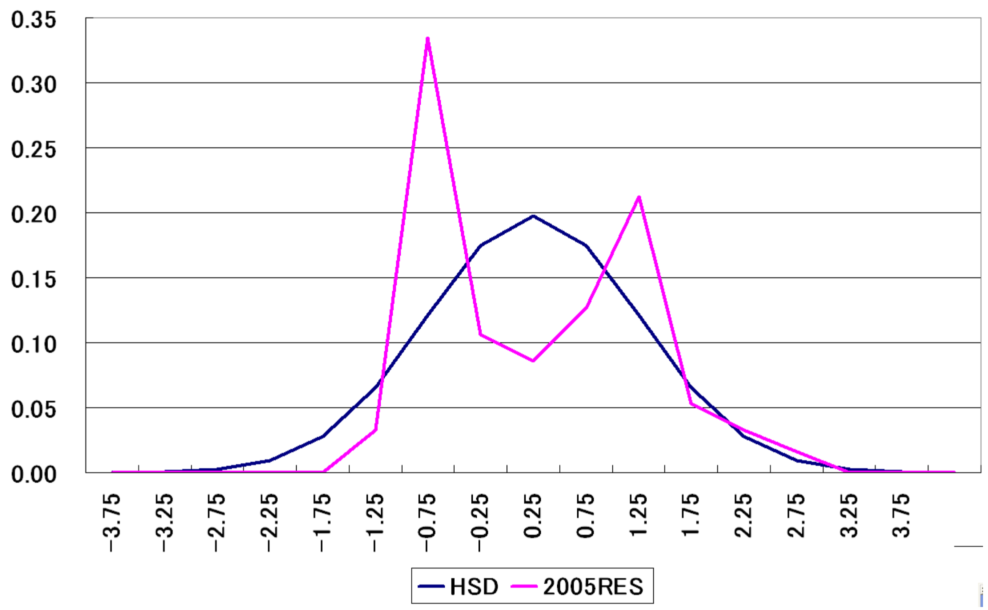


図7A-4　日経平均(2005年)残差の標準化

7-3 各種金融指標のタイム・トレンドと残差 ―プロジェクトIII

- 日経平均の場合と同様に，2003年と2005年の回帰分析における決定係数値が高い
- 残差を標準正規分布に変換した場合には，2003年と2005年特に2005年では理論分布と大きく異なっている

回帰統計	
重相関 R	0.862466
重決定 R2	0.743848
補正 R2	0.742798
標準誤差	49.01881
観測数	246

分散分析表

	自由度	変動	分散	測された分散	有意 F
回帰	1	1702560.009	1702560	708.5603	4.06E-74
残差	244	586293.9602	2402.844		
合計	245	2288853.969			

	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	773.8134	6.269756274	123.42	6.8E-222	761.4637	786.1632	761.4637	786.1632
X 値 1	1.171501	0.044010302	26.6188	4.06E-74	1.084813	1.25819	1.084813	1.25819

図7A-5 TOPIX(2003年)回帰分析結果

7-3 各種金融指標のタイム・トレンドと残差
—プロジェクトIII

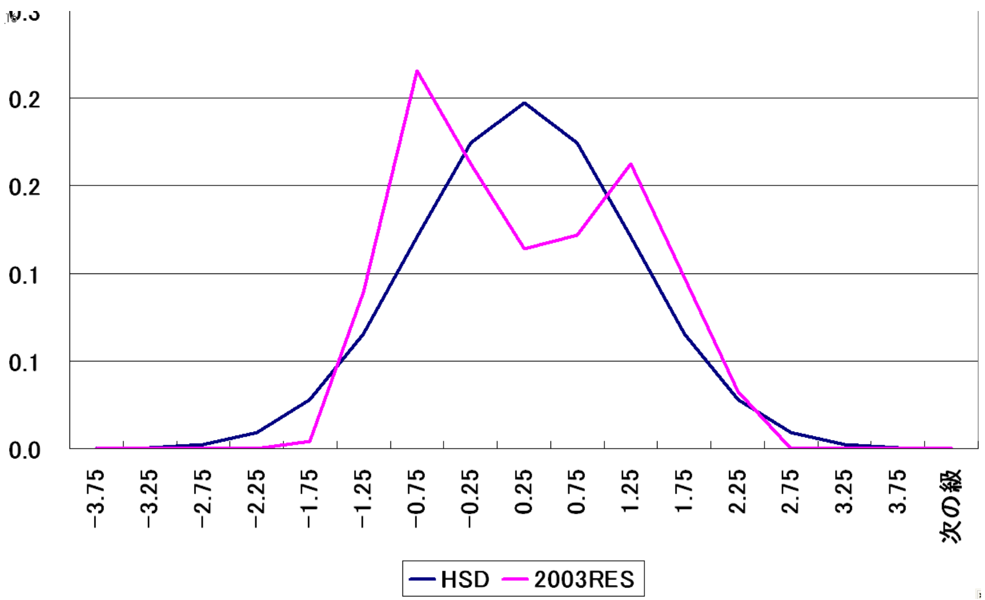


図7A-6 TOPIX(2003年)残差の標準化

7-3 各種金融指標のタイム･トレンドと残差 ―プロジェクトIII

回帰統計	
重相関 R	0.861271
重決定 R2	0.741788
補正 R2	0.740725
標準誤差	78.43573
観測数	245

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	4294736	4294736	698.0854157	2.13E-73
残差	243	1494976	6152.164		
合計	244	5789712			

	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	1039.758	10.05291	103.4285	6.6351E-203	1019.956	1059.56	1019.956	1059.56
X 値 1	1.872032	0.070853	26.42131	2.12897E-73	1.732468	2.011597	1.732468	2.011597

図7A-7　TOPIX(2005年)回帰分析結果

7-3 各種金融指標のタイム・トレンドと残差 ―プロジェクトIII

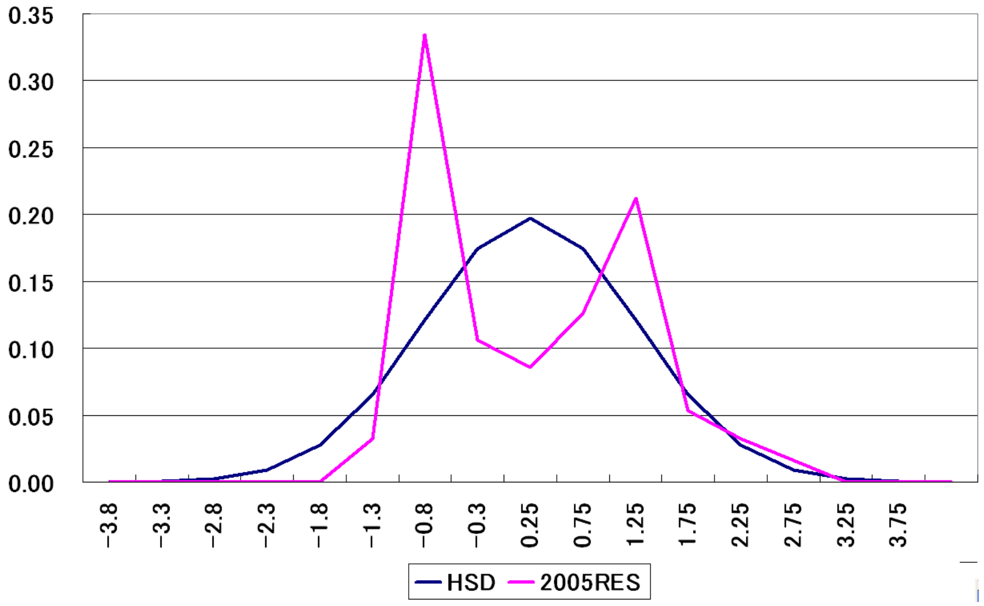


図7A-8 TOPIX(2005年)残差の標準化

7-3 各種金融指標のタイム・トレンドと残差—プロジェクトIII

- いずれの年においても決定係数値が低く，意味のある回帰分析とはいえないが，長期金利原データの分布とは形状の異なっている年が多くなっている
- 残差の分布に関しては，標準正規分布の理論分布と比較的近い形状を示している

	A	B	C	D	E	F	G	H	I
1	概要								
2									
3	回帰統計								
4	重相関 R	0.276687							
5	重決定 R2	0.076556							
6	補正 R2	0.072771							
7	標準誤差	0.165724							
8	観測数	246							
9									
10	分散分析表								
11		自由度	変動	分散	観測された分散比	有意 F			
12	回帰	1	0.555554	0.555554	20.2281988	1.06E-05			
13	残差	244	6.701294	0.027464					
14	合計	245	7.256848						
15									
16		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	切片	1.418618	0.021197	66.92575	5.3157E-159	1.376866	1.460371	1.376866	1.460371
18	X 値 1	0.000669	0.000149	4.497577	1.0635E-05	0.000376	0.000962	0.000376	0.000962

図7A-9 長期金利(2003年)回帰分析結果

7-3　各種金融指標のタイム･トレンドと残差 ―プロジェクトIII

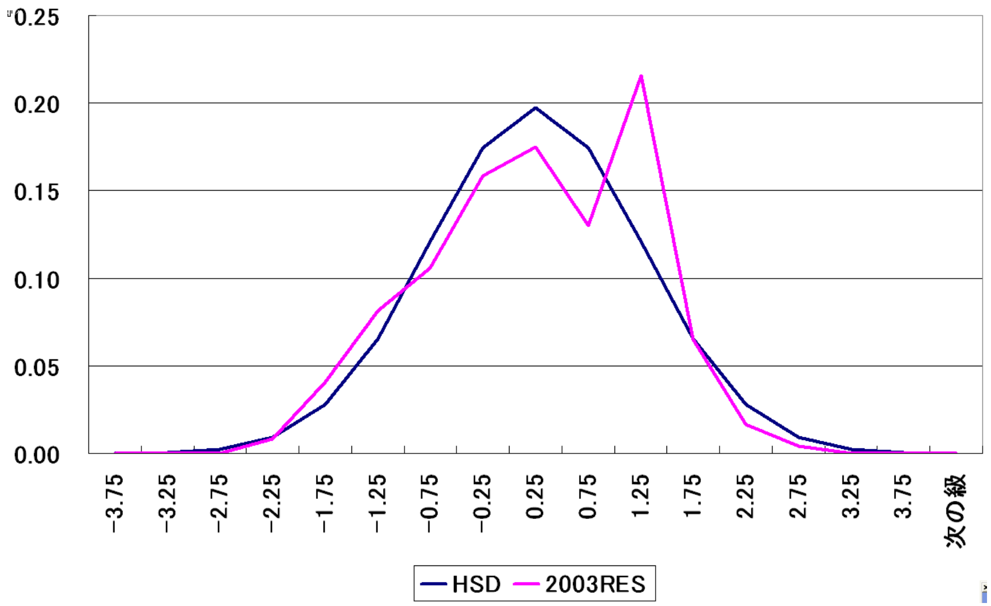


図7A-10　長期金利(2003年)残差の標準化

7-3 各種金融指標のタイム・トレンドと残差 ―プロジェクトIII

回帰統計	
重相関 R	0.467967
重決定 R2	0.218993
補正 R2	0.215779
標準誤差	0.097846
観測数	245

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	0.65233	0.65233	68.13673597	9.77E-15
残差	243	2.326442	0.009574		
合計	244	2.978772			

	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	1.293236	0.012541	103.1233	1.339E-202	1.268534	1.317938	1.268534	1.317938
X 値 1	0.00073	8.84E-05	8.254498	9.77118E-15	0.000555	0.000904	0.000555	0.000904

図7A-11 長期金利(2005年)回帰分析結果

7-3 各種金融指標のタイム・トレンドと残差 ープロジェクトIII

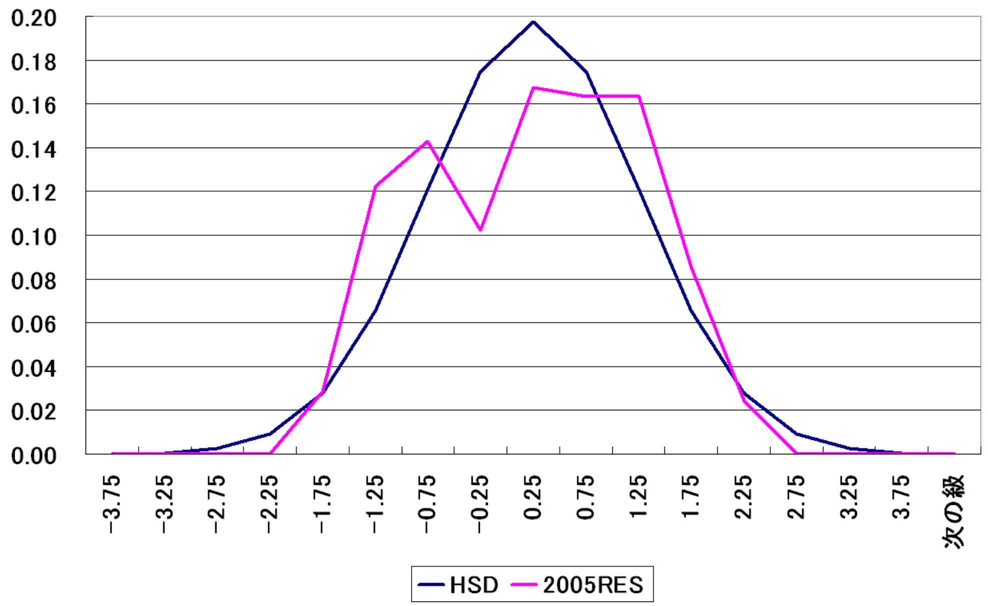


図7A-12 長期金利(2005年)残差の標準化

7-3 各種金融指標のタイム・トレンドと残差 ―プロジェクトIII

- 円対ドル為替レートの回帰分析結果では，2005年の回帰分析において決定係数の値が非常に高くなっている
- しかし，残差の分布の形状は原データの分布の形状とは大きく異なっている
- 決定係数値が大きな2005年の回帰分析の残差はほぼ標準正規分布の理論分布と合致した形状を示している

	A	B	C	D	E	F	G	H	I
1	概要								
2									
3	回帰統計								
4	重相関 R	0.163595							
5	重決定 R2	0.026763							
6	補正 R2	0.022775							
7	標準誤差	2.650638							
8	観測数	246							
9									
10	分散分析表								
11		自由度	変動	分散	測された分散	有意 F			
12	回帰	1	47.14266	47.14266	6.709856	0.010165			
13	残差	244	1714.316	7.025883					
14	合計	245	1761.458						
15									
16		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	切片	108.9013	0.33903	321.2142	0	108.2335	109.5691	108.2335	109.5691
18	X 値 1	-0.00616	0.00238	-2.59034	0.010165	-0.01085	-0.00148	-0.01085	-0.00148

図7A-13　円対ドル為替レート(2003年)回帰分析結果

7-3 各種金融指標のタイム・トレンドと残差 ―プロジェクトIII

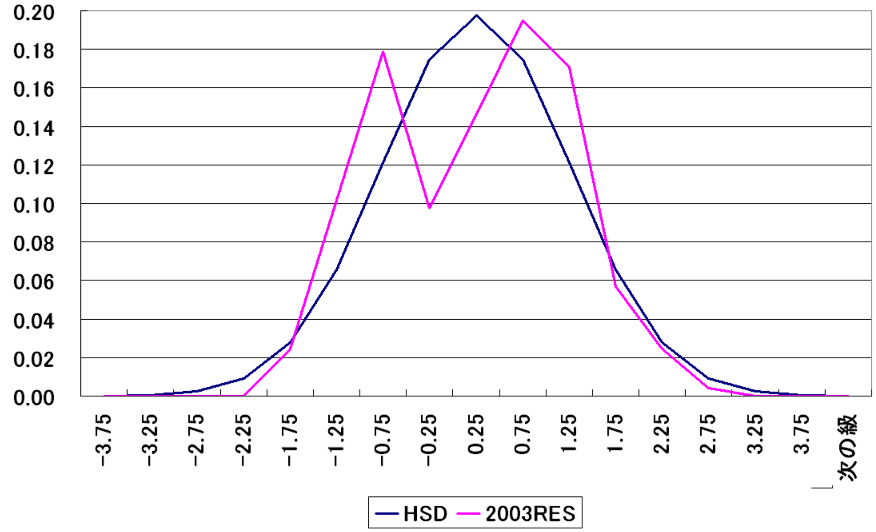


図7A-14 円対ドル為替レート(2003年)残差の標準化

7-3 各種金融指標のタイム・トレンドと残差 —プロジェクトIII

	A	B	C	D	E	F	G	H	I
1	概要								
2									
3	回帰統計								
4	重相関 R	0.944522							
5	重決定 R2	0.892122							
6	補正 R2	0.891678							
7	標準誤差	1.635411							
8	観測数	245							
9									
10	分散分析表								
11		自由度	変動	分散	測された分散	有意 F			
12	回帰	1	5374.646	5374.646	2009.538	1.7E-119			
13	残差	243	649.92	2.674568					
14	合計	244	6024.566						
15									
16		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	切片	102.0389	0.209606	486.8116	0	101.626	102.4517	101.626	102.4517
18	X 値 1	0.066225	0.001477	44.82787	1.7E-119	0.063315	0.069135	0.063315	0.069135

図7A-15 円対ドル為替レート(2005年)回帰分析結果

7-3　各種金融指標のタイム・トレンドと残差 ―プロジェクトIII

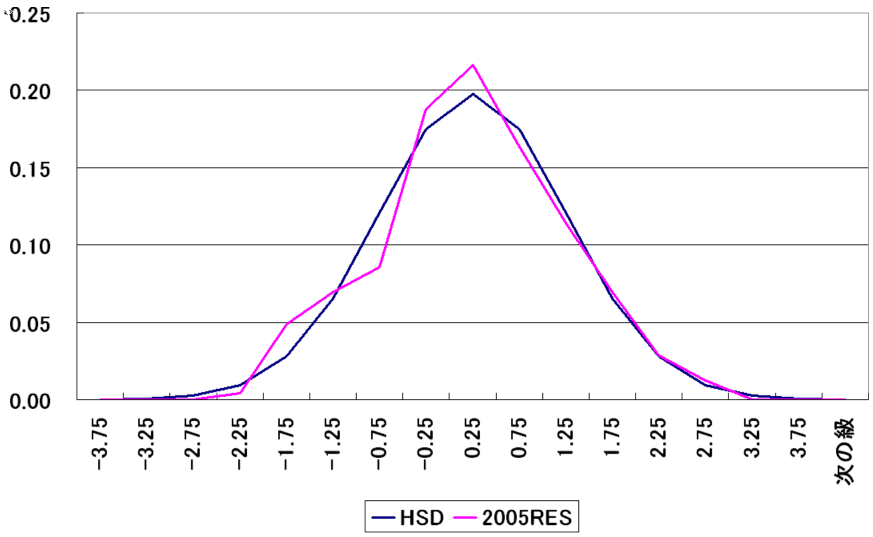


図7A-16　円対ドル為替レート(2005年)残差の標準化