4　誤差項についての検証

4．1　均一分散かどうかの検証

誤差項についての仮定のうち

$$Var(U_i) = \sigma^2 \ (i = 1, 2, \cdots, n)$$

(全ての誤差項の分散は等しい)

について検証する

この仮定が成立していないときには

＊最小二乗法による推定が必ずしも適切でない
＊有意性の検定について修正が必要となる

(第2回課題の解答)

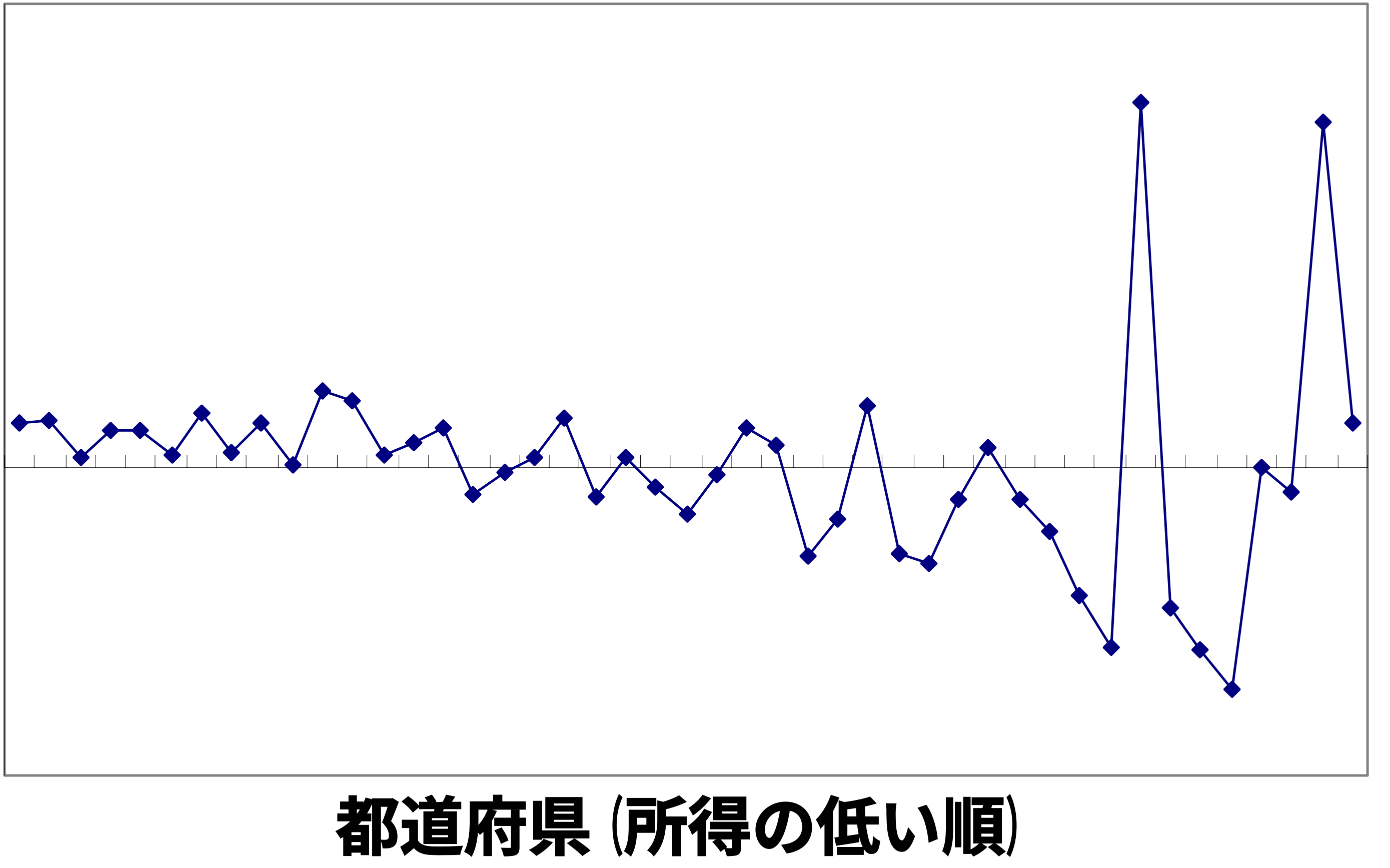
推定結果

$$Y = -497478 + 0.65X$$
$$(-3.18) \quad (45.62)$$

$$R^2 = 0.980 \quad \overline{R}^2 = 0.979$$

(　) 内は t 値

残差グラフ



◎不均一分散の検定（Whiteの検定）

帰無仮説：全ての誤差項の分散は等しい
対立仮説：少なくとも1つは異なるものがある

『手順』

(1) 最小二乗残差を求める

$$Y_i = \beta_1 \cdot 1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + U_i$$

$$(i = 1, 2, \cdots, n)$$

を普通に最小二乗法で推定し、

$\hat{U}_1, \hat{U}_2, \cdots, \hat{U}_n$ を求める

(2) 残差を被説明変数、定数項を含めた説明変数間の任意の積のうち、区別できるものを説明変数として、最小自乗法を実行決定係数を求める

単回帰の場合$(Y = \beta_1 \cdot 1 + \beta_2 X + U)$

被説明変数：$\hat{U}$

説明変数：$1, X, X^2$

	1	X
1	1	X
X	$\cancel{X}$	X^2

p ：**説明変数の個数**

$p = 3$

説明変数が２つの場合

$$(Y = \beta_1 \cdot 1 + \beta_2 X_2 + \beta_3 X_3 + U)$$

被説明変数：$\hat{U}$

説明変数：$1, X_2, X_3, X_2^2, X_3^2, X_2 X_3$

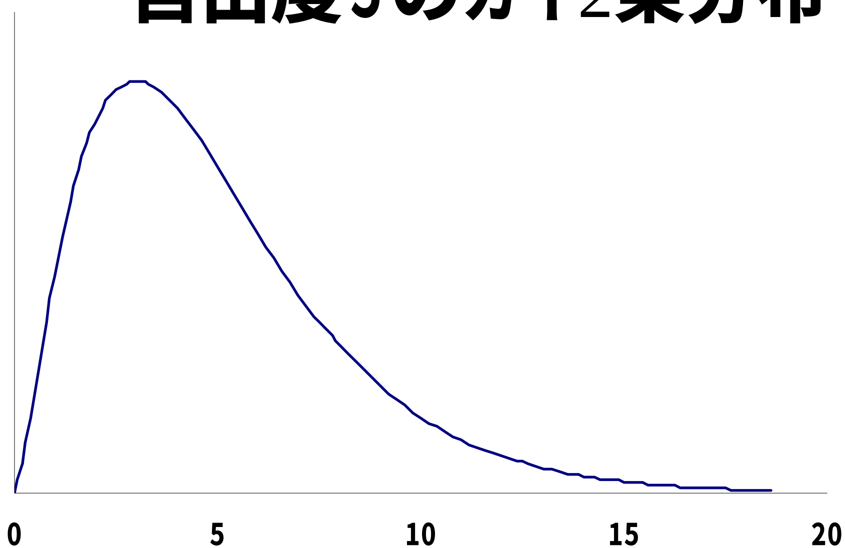
	1	X_2	X_3
1	1	X_2	X_3
X_2	$\cancel{X_2}$	X_2^2	$X_2 X_3$
X_3	$\cancel{X_3}$	$\cancel{X_3 X_2}$	X_3^2

$p = 6$

(3) nR^2 を用いて棄却か受容かを決定

nR^2 は 帰無仮説が正しいとき、n が十分大きければ自由度 p-1のカイ2乗分布で近似できる

自由度5のカイ2乗分布

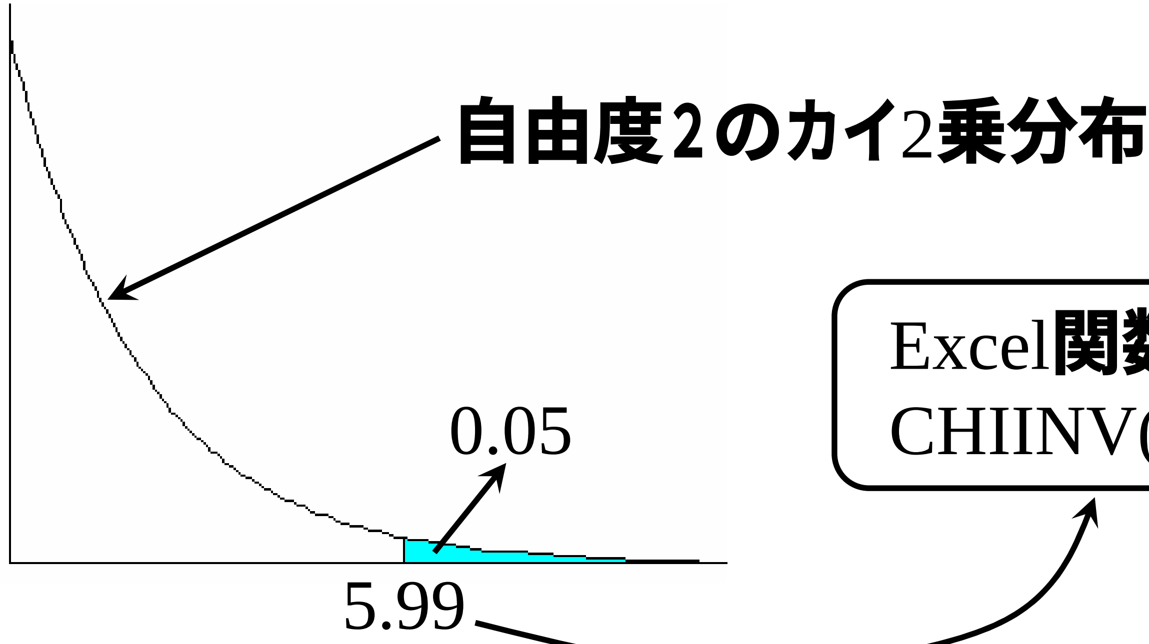


有意水準を5％とすると

自由度 p-1のカイ2乗分布の上側5％点＜nR^2

のとき帰無仮説を棄却

単回帰の場合（p = 3）

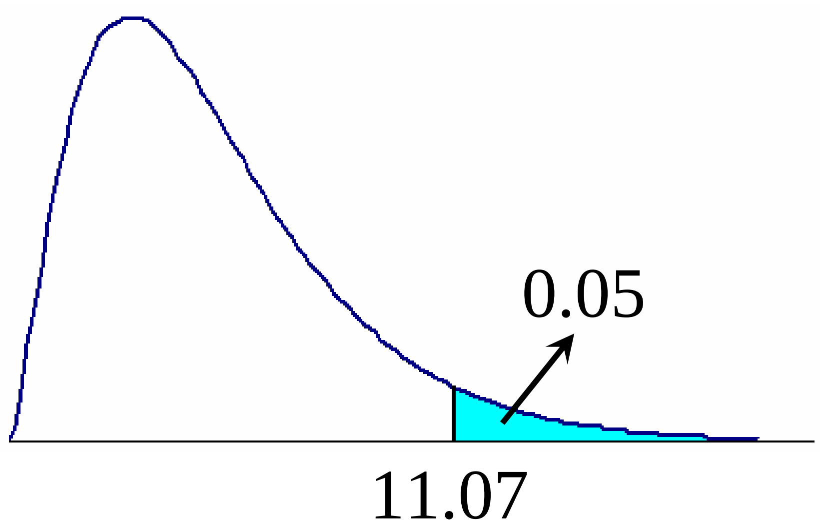


Excel関数

CHIINV(0.05,2)

説明変数が2つの場合（p = 6）

自由度5のカイ2乗分布



Excel関数
CHIINV(0.05,5)

第2回課題のデータについての不均一分散の検定

Y :民間最終消費支出　　X :県民可処分所得

$\hat{U}:\ Y=\beta_1+\beta_2X+U$ を最小二乗法で推定して得られた残差

被説明変数: $\hat{U}$　　説明変数: $1, X, X^2$

として最小二乗法で推定し得られた $nR^2=9.41$

有意水準を5%とすると

$5.99<9.41$ より帰無仮説棄却

→不均一分散が認められる

◎不均一分散が認められた場合の処理（推定）

最適な推定方法：誤差の分散 $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$ を利用した加重最小二乗法

→誤差の分散は例外的な場合を除き未知であるのでこの方法は利用できない

代替的な推定方法：誤差の分散を何らかの方法で推定し、それを利用した加重最小二乗法

→利用の仕方を間違えると通常の最小二乗法よりも悪い結果をもたらす

◎不均一分散が認められた場合の処理（検定）

通常の t 統計量を用いた有意性の検定は不均一分散が存在する場合には正しい検定方法とはいえない
→本来有意でない説明変数を有意としてしまう傾向がある

推定を通常の最小二乗法のままにしたとしても、有意性の t 検定については修正が必要

◎有意性の t 検定の修正 (Whiteの方法)

簡単化のため以下のデータを用いる：

Y	X_1	X_2	X_3
9.6	1	1	2
11.6	1	2	1
12.1	1	3	5
19.9	1	4	3
18.1	1	5	4

以後Excelを利用した場合の手順を紹介

通常の最小二乗法による推定結果

概要

回帰統計	
重相関 R	0.929
重決定 R2	0.863
補正 R2	0.726
標準誤差	2.341
観測数	5

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	69.05	34.525	6.299	0.137
残差	2	10.962	5.481		
合計	4	80.012			

	係数	標準誤差	t	P-値	下限 95%
切片	7.735	2.695	2.870	0.103	-3.860
x2	3.0625	0.925	3.309	0.080	-0.919
x3	-0.8875	0.925	-0.959	0.439	-4.869

残差出力

観測値	予測値 : y	残差
1	9.0225	0.5775
2	12.9725	-1.3725
3	12.485	-0.385
4	17.3225	2.5775
5	19.4975	-1.3975

(手順1) 最小自乗残差をデータ系列の隣に
以下のように配置する

	A	B	C	D	E
1	y	x1	x2	x3	残差
2	9.6	1	1	2	0.5775
3	11.6	1	2	1	−1.3725
4	12.1	1	3	5	−0.385
5	19.9	1	4	3	2.5775
6	18.1	1	5	4	−1.3975

(手順2) 残差の2乗を作成

	C	D	E	F
1	x2	x3	残差	残差の2乗
2	1	2	0.5775	0.334
3	2	1	−1.3725	1.884
4	3	5	−0.385	0.148
5	4	3	2.5775	6.644
6	5	4	−1.3975	1.953

=E2^2

(手順3) 残差の2乗×説明変数を作成

	F	G	H
1	残差の2乗	残差の2乗 × X2	残差の2乗 × X3
2	0.334	0.3335062	0.6670125
3	1.884	3.7675125	1.8837563
4	0.148	0.444675	0.741125
5	6.644	26.574025	19.930519
6	1.953	9.7650313	7.812025

=$F2*D2

=$F2*C2

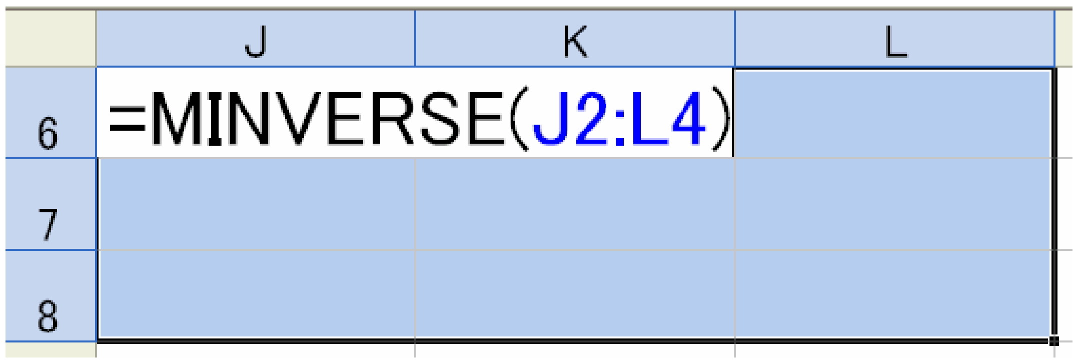
(手順4) 3行3列の空白のセルを選択し以下の入力を行い、[Ctrl]+[Shift]+[Enter]

	J	K	L	M
2	=MMULT(TRANSPOSE(B2:D6),B2:D6)			
3				
4				

＊B2:D6・・・定数項と全ての説明変数

＊3行3列・・・定数項も含めた説明変数の個数

（手順5）3行3列の空白のセルを選択し以下の入力を行い、[Ctrl]+[Shift]+[Enter]



＊J2:L4・・・（手順4）で作成した部分

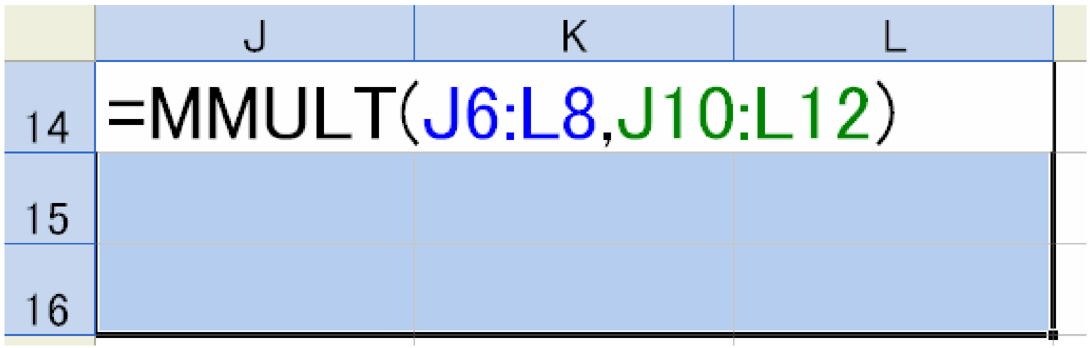
(手順6) 3行3列の空白のセルを選択し以下の入力を行い、[Ctrl]+[Shift]+[Enter]

	J	K	L	M
10	=MMULT(TRANSPOSE(B2:D6),F2:H6)			
11				
12				

＊B2:D6･･･**定数項と全ての説明変数**

＊F2:H6･･･ **(手順2)、(手順3) で作成した部分**

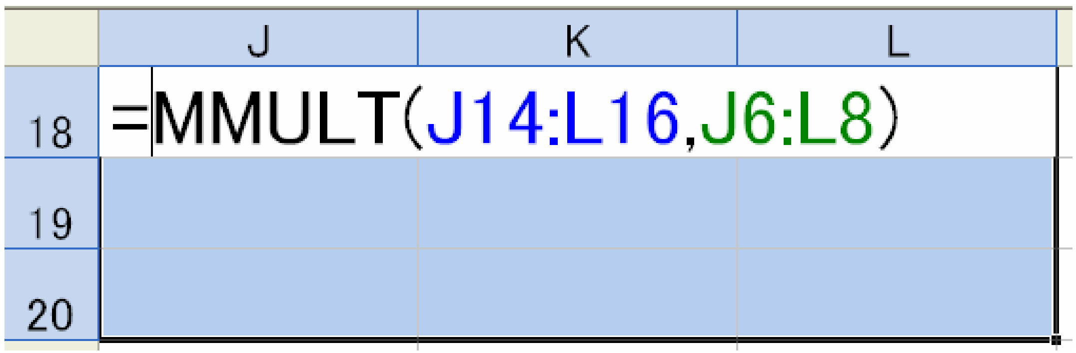
（手順7） 3行3列の空白のセルを選択し以下の入力を行い、[Ctrl]+[Shift]+[Enter]



＊J6:L8･･･（手順5）で作成した部分

＊J10:L12･･･（手順6）で作成した部分

（手順8）3行3列の空白のセルを選択し以下の入力を行い、[Ctrl]+[Shift]+[Enter]



＊J14:L16・・・（手順7）で作成した部分

＊J6:L8・・・（手順5）で作成した部分

（手順9）分析ツールの回帰分析の結果より以下の部分をコピーし空白部分に貼り付ける

	係数	標準誤差	t
切片	7.735	2.69487	2.870268
x2	3.0625	0.925422	3.3093
x3	-0.8875	0.925422	-0.95902

(手順10)(手順8)で作成した部分を利用して標準誤差を修正する

	J	K	L
18	1.55134	−0.1478	−0.3
19	−0.1478	0.27866	−0.1345
20	−0.3	−0.1345	0.16524

	J	K	L	
22		係数	標準誤差	
23	切片	7.735	1.24553	=SQRT(J18)
24	x2	3.0625	0.52788	=SQRT(K19)
25	x3	−0.8875	0.4065	=SQRT(L20)

(手順11) "係数"を"標準誤差"で割った値を"t"の部分に書き込み、更にtをzに書き換える

	K	L	M
22	係数	標準誤差	z
23	7.735	1.24553	6.21
24	3.0625	0.52788	5.801
25	−0.8875	0.4065	−2.18

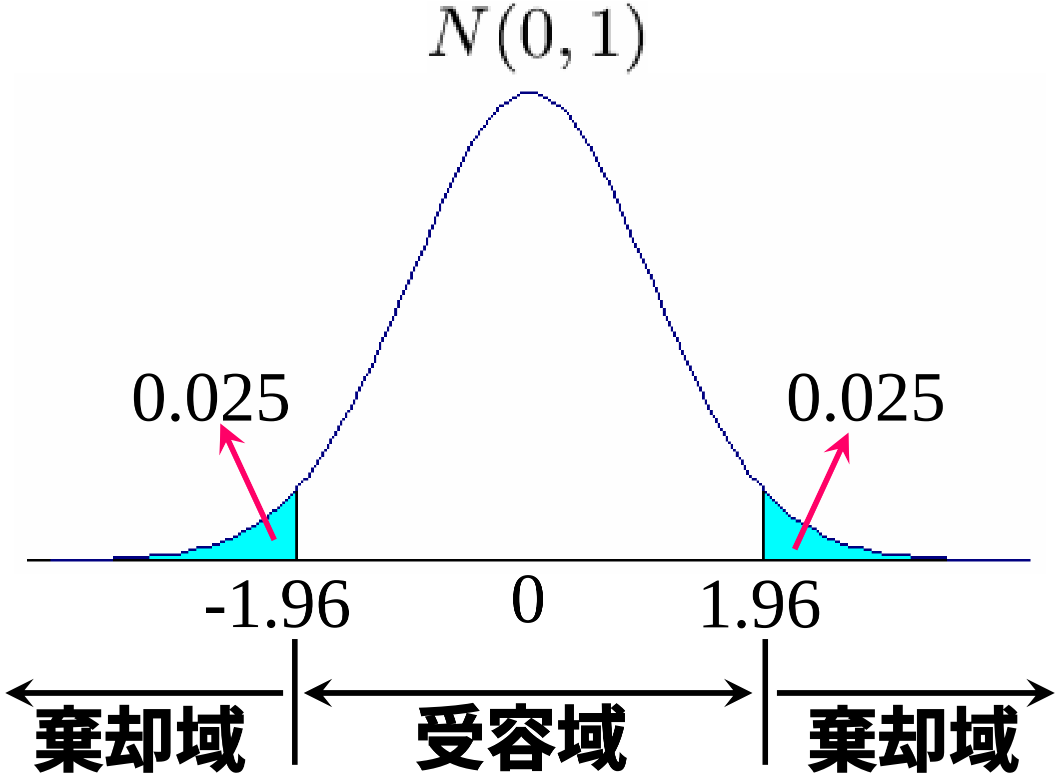
=K23/L23

=K24/L24

=K25/L25

最終的に得られた、“z 値”を用いた検定では t 分布ではなく、標準正規分布を用いる

「有意水準5％、両側検定の場合」



（注）
簡単化のためデータの個数が少ない場合で説明したが、実際には十分大きなデータの場合に、この修正は有効となる

4．2　系列相関の有無の検証

＊対象となるのは基本的に時系列データ

系列相関：異なる誤差項間の相関

＊系列相関には様々なものが存在するが
以下では1階の自己相関のみを扱う

$$U_{t+1} = \rho U_t + V_{t+1} \ (t = 1, 2, \cdots, n)$$

ただし $-1 \leq \rho \leq 1$

t+1 期の誤差は、t 期の誤差の一部と
新たに発生する“ショック”との和

V_tについて

（1） 互いに独立

（2） $E(V_t) = 0 \ (t = 1, 2, \cdots, n)$

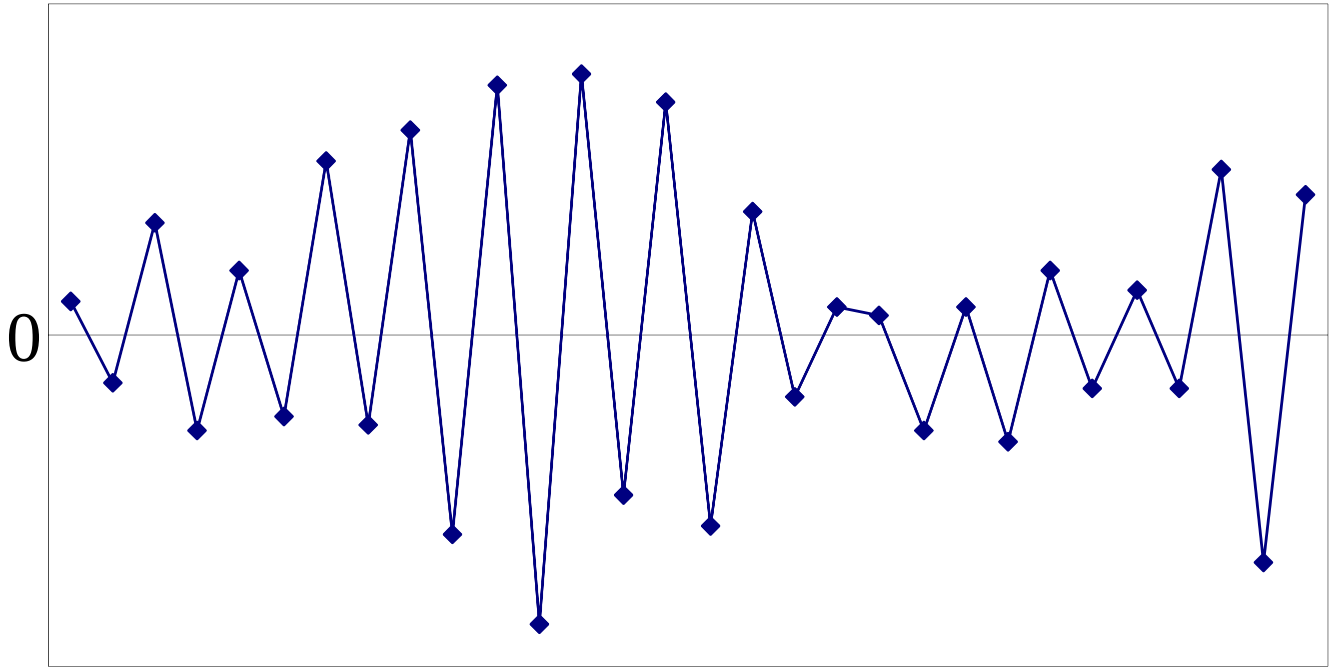
（3） $Var(V_t) = \sigma^2 \ (t = 1, 2, \cdots, n)$

（4） 過去の U_tとも独立

もしも $\rho = 0$ ならば、U_tは古典的仮定を満たす誤差項となる

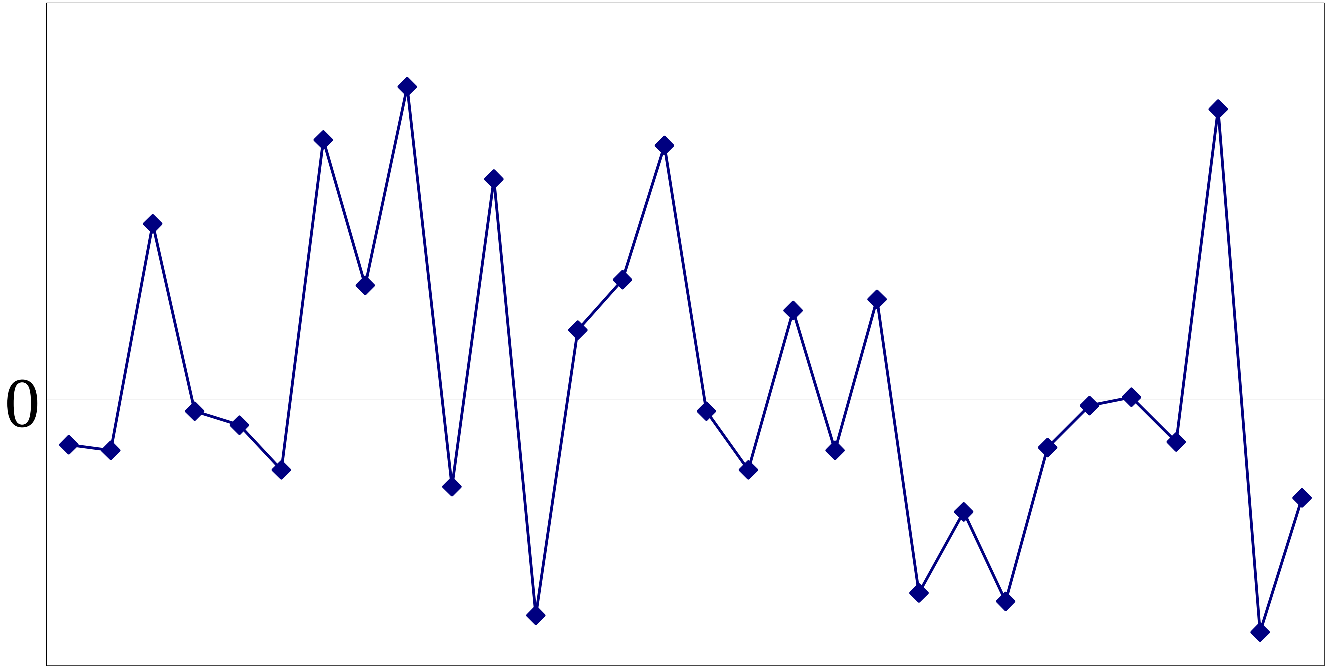
(第3回課題の解答)

$$\rho = -0.8$$



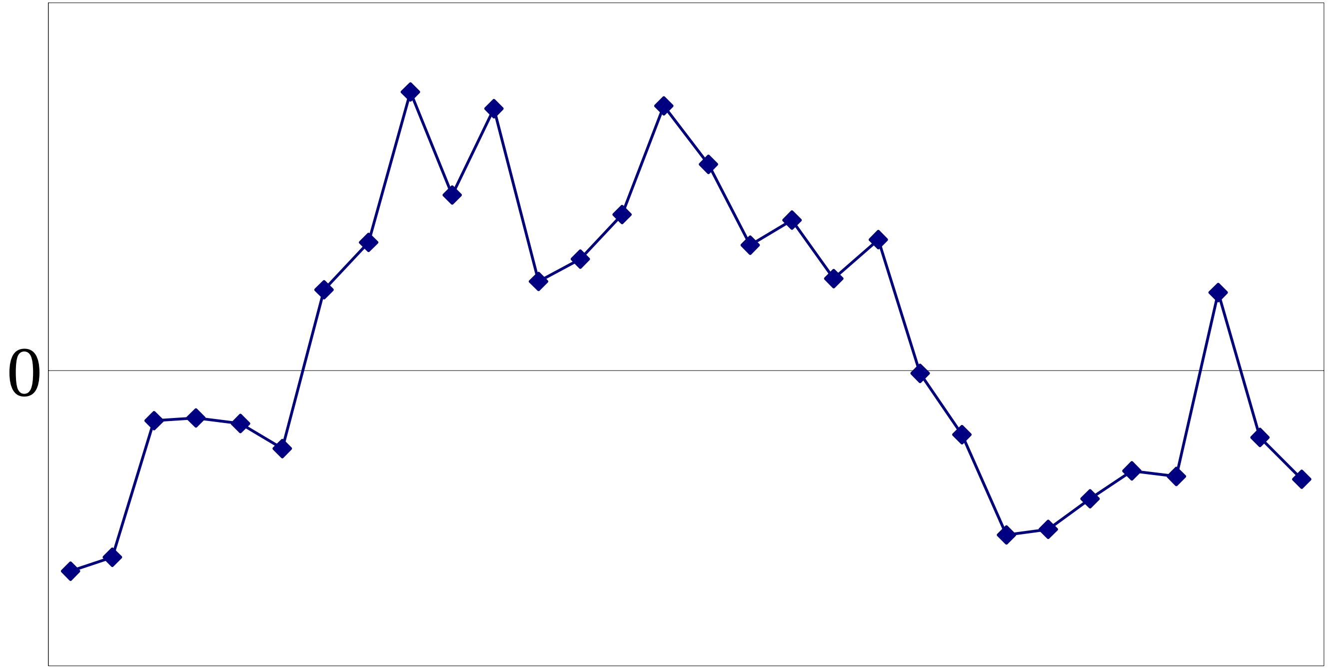
“負の系列相関の典型例”

$$\rho = 0$$



“系列相関がない場合の1つの例”

$\rho = 0.8$



“正の系列相関の典型例”

◎1階の自己相関の検定（ダービー・ワトソン検定）

ダービン・ワトソン統計量：

$$DW = \frac{\sum_{t=2}^{n}(\hat{U}_t - \hat{U}_{t-1})^2}{\sum_{t=1}^{n}\hat{U}_t^2}$$

$0 \leq DW \leq 4$ が成立し、系列相関がない

場合には2から大きく離れる確率は小さい

ダービン･ワトソン統計量の５％臨界値

n	k'=1		k'=2		k'=3		k'=4	
	D_L	D_U	D_L	D_U	D_L	D_U	D_L	D_U
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977

n：データの個数

k'：定数項以外の説明変数の個数

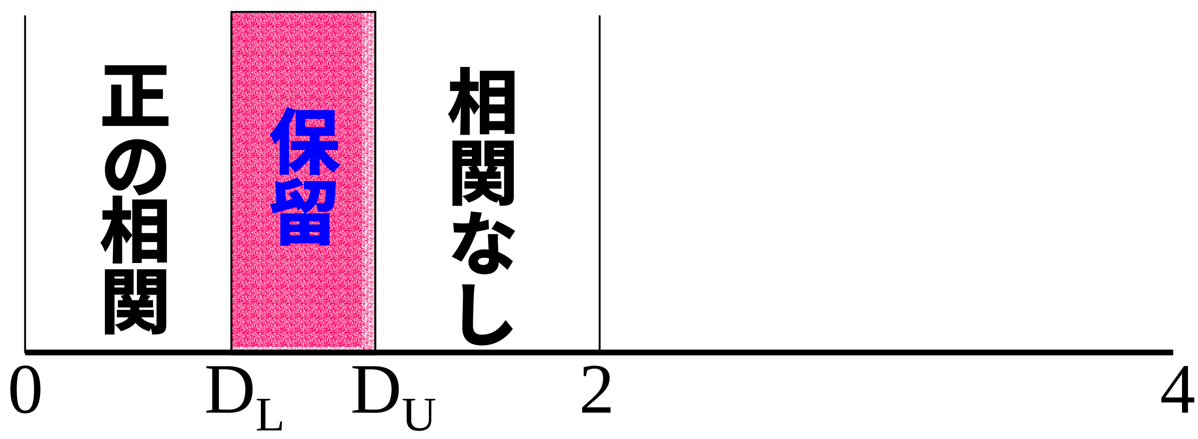
正の系列相関の検定

帰無仮説: $\rho = 0$　　**対立仮説:** $\rho > 0$

$DW < D_L$ → **帰無仮説を棄却**

$D_L < DW < D_U$ → **保留**

$D_U < DW$ → **帰無仮説を受容**



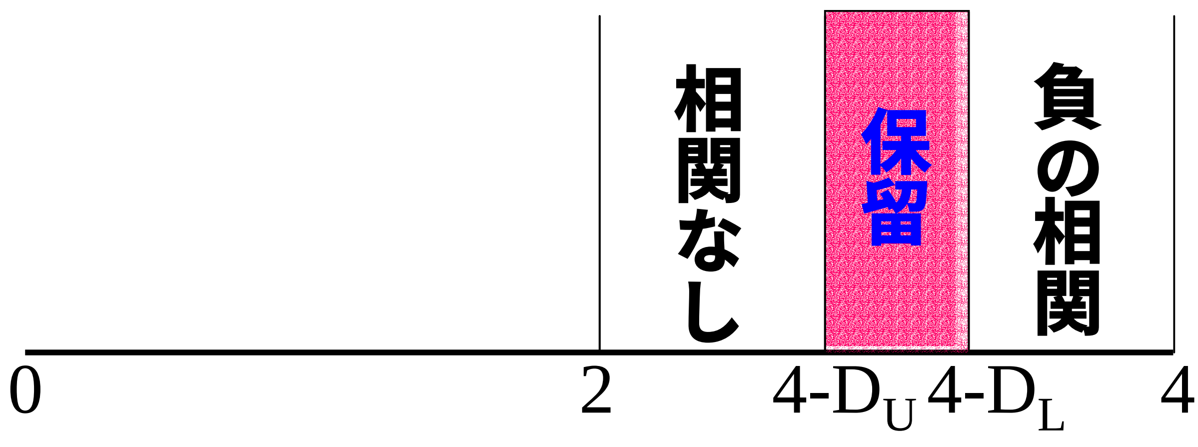
負の系列相関の検定

帰無仮説： $\rho = 0$ **対立仮説：** $\rho < 0$

$4\text{-}D_L < DW$ → **帰無仮説を棄却**

$4\text{-}D_U < DW < 4\text{-}D_L$ → **保留**

$DW < 4\text{-}D_U$ → **帰無仮説を受容**



ExcelでのDWの求め方

用いるデータ：

Y	3.7	6.1	6.1	5.6	5.3	5.7	6.6	7.2	9.8	11
X	1	2	3	4	5	6	7	8	9	10

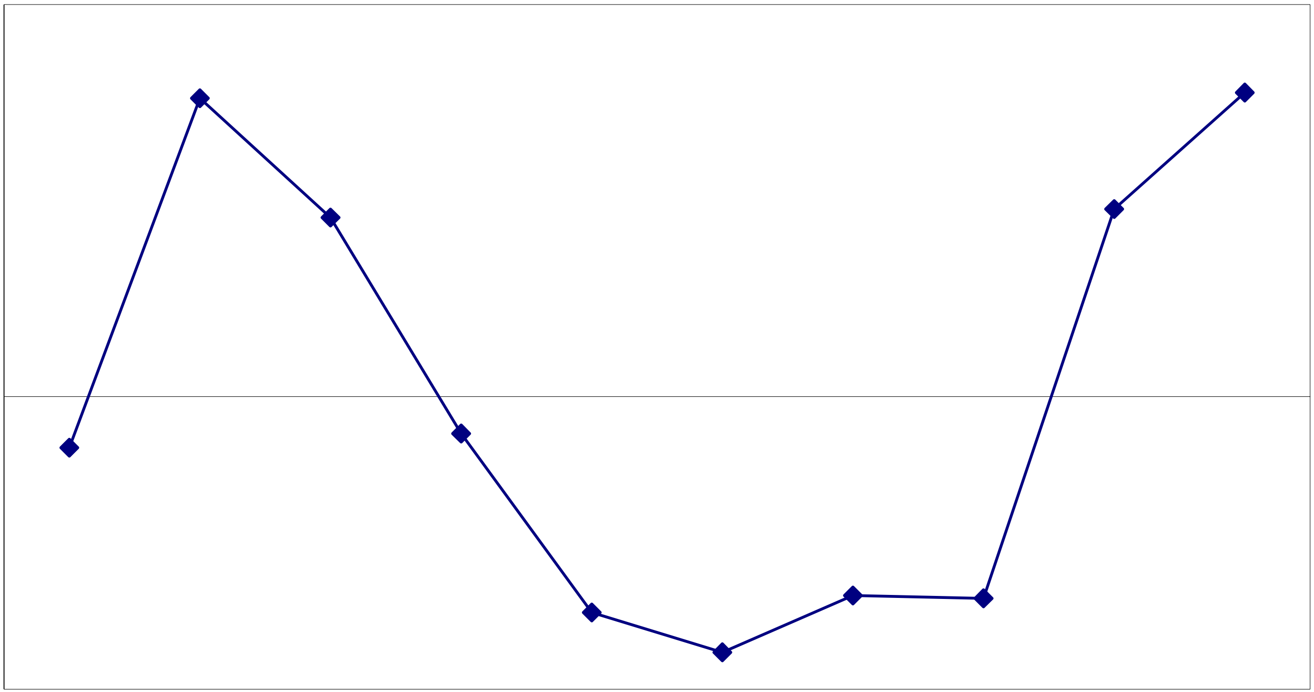
最小二乗法による推定結果：

$$\underset{(4.08)}{Y = 3.36} + \underset{(4.59)}{0.61X}$$

$$R^2 = 0.724$$

$$\overline{R}^2 = 0.690$$

最小自乗残差のグラフ



（手順1）残差と1期前の残差を以下のように配置

	A	B
1	残差	1期前残差
2	−0.269	
3	1.522	−0.269
4	0.913	1.522
5	−0.196	0.913
6	−1.105	−0.196
7	−1.315	−1.105
8	−1.024	−1.315
9	−1.033	−1.024
10	0.958	−1.033
11	1.549	0.958

(手順2)(残差－1期前の残差)の2乗を計算

	A	B	C
1	残差	1期前残差	(残差－1期前残差)の2乗
2	−0.269		
3	1.522	−0.269	3.207
4	0.913	1.522	0.371
5	−0.196	0.913	1.230
6	−1.105	−0.196	0.826
7	−1.315	−1.105	0.044
8	−1.024	−1.315	0.085
9	−1.033	−1.024	0.000
10	0.958	−1.033	3.964
11	1.549	0.958	0.349

=(A3-B3)^2

=(A4-B4)^2

以下同様

(手順3) 残差の2乗を計算

	C	D
1	(残差－1期前残差)の2乗	残差の2乗
2		0.072
3	3.207	2.316
4	0.371	0.833
5	1.230	0.039
6	0.826	1.222
7	0.044	1.728
8	0.085	1.048
9	0.000	1.067
10	3.964	0.918
11	0.349	2.400

=A2^2

=A3^2

以下同様

(手順4) DWを求める

	C	D
4	0.371	0.833
5	1.230	0.039
6	0.826	1.222
7	0.044	1.728
8	0.085	1.048
9	0.000	1.067
10	3.964	0.918
11	0.349	2.400
12	10.076	11.642
13		
14	DW	0.865

=sum(C3:C11)

=sum(D2:D11)

=C12/D12

求まったダービン･ワトソン値=0.865より、正の系列相関が疑われる

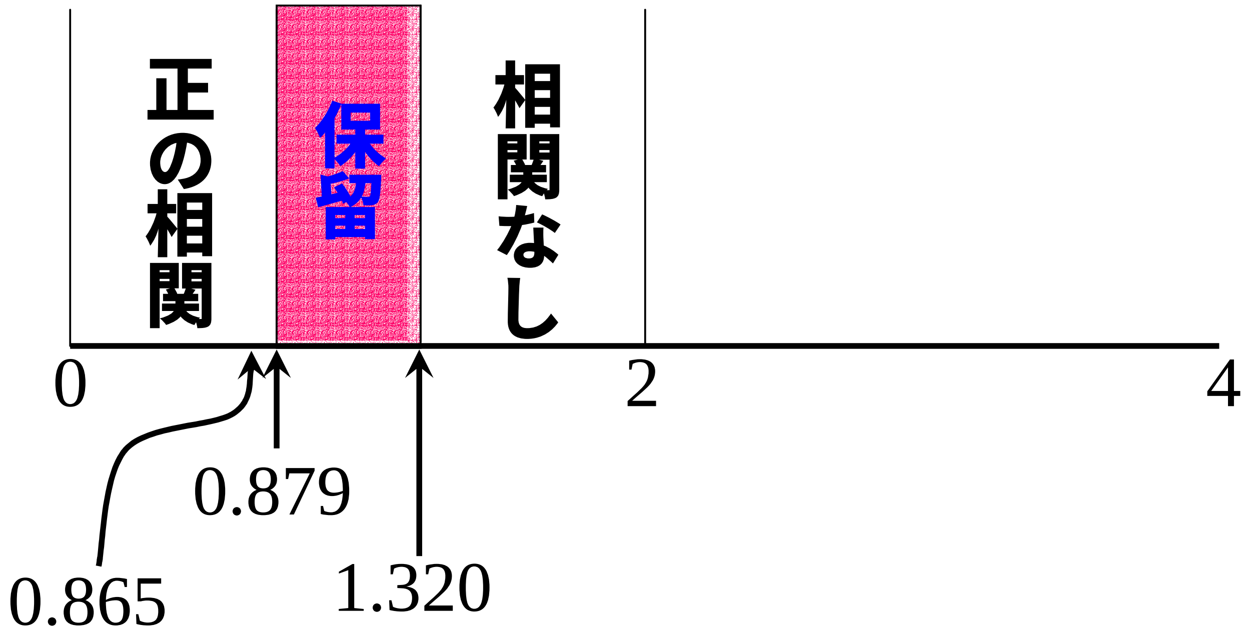
帰無仮説：$\rho = 0$　　　対立仮説：$\rho > 0$

有意水準を５％として、実際に検定を行うと

データの個数＝10
定数項を除いた説明変数の個数＝１

３２ページの数表より、D_L=0.879、D_U=1.320

DW < D_Lとなるので、帰無仮説を棄却
→正の系列相関が認められる


正の相関
保留
相関なし
0
2
4
0.879
0.865
1.320