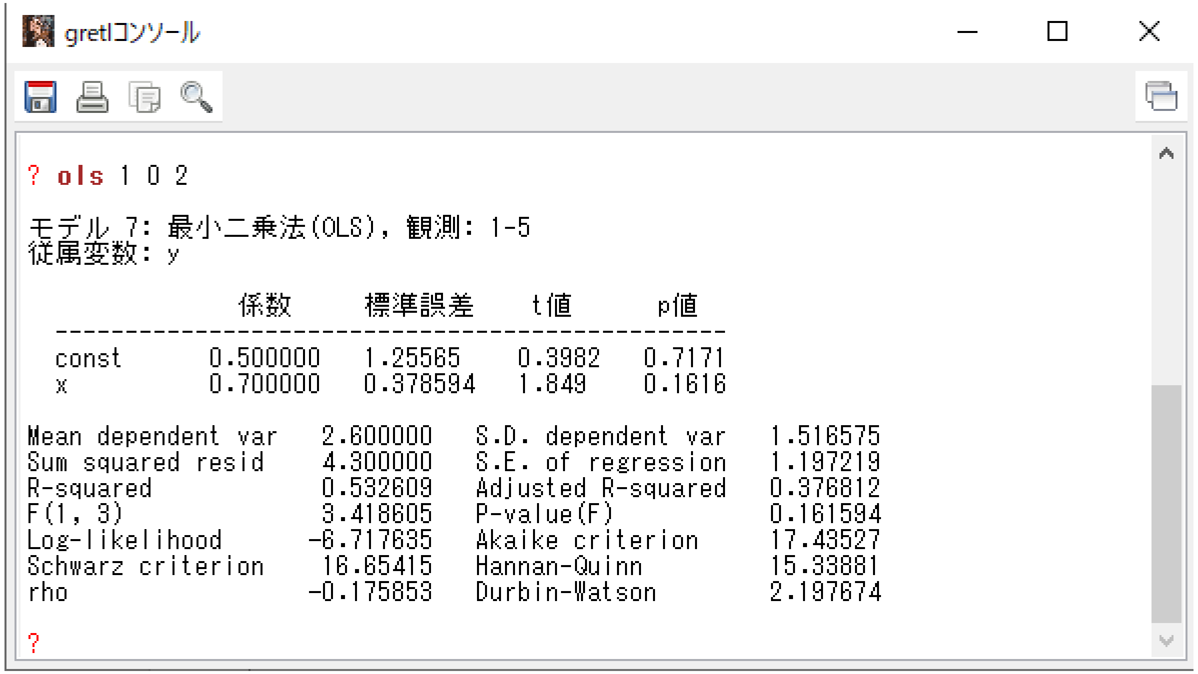
変数名でなく，最初の変数名リストの画面の「ID#」の番号でもよい。

? の後に，「ols 1 0 2」とタイプして，Enter キーを押すと，下画面のように同じ結果が得られる。

ID# の 0 が const，1 が y，2 が x である。



```
gretlコンソール

? ols 1 0 2

モデル 7: 最小二乗法(OLS), 観測: 1-5
従属変数: y

              係数       標準誤差     t値      p値
  --------------------------------------------------
  const      0.500000    1.25565     0.3982   0.7171
  x          0.700000    0.378594    1.849    0.1616

Mean dependent var   2.600000   S.D. dependent var   1.516575
Sum squared resid    4.300000   S.E. of regression   1.197219
R-squared            0.532609   Adjusted R-squared   0.376812
F(1, 3)              3.418605   P-value(F)           0.161594
Log-likelihood      -6.717635   Akaike criterion     17.43527
Schwarz criterion    16.65415   Hannan-Quinn         15.33881
rho                 -0.175853   Durbin-Watson        2.197674

?
```

● 推定結果の意味

```
モデル 1: 最小二乗法(OLS), 観測: 1-5
従属変数: y

                 係数       標準誤差      t 値       p 値
  -----------------------------------------------
  const       0.500000    1.25565     0.3982    0.7171
  x           0.700000    0.378594    1.849     0.1616

Mean dependent var   2.600000   S.D. dependent var   1.516575
Sum squared resid    4.300000   S.E. of regression   1.197219
R-squared            0.532609   Adjusted R-squared   0.376812
F(1, 3)              3.418605   P-value(F)           0.161594
Log-likelihood      -6.717635   Akaike criterion     17.43527
Schwarz criterion    16.65415   Hannan-Quinn         15.33881
rho                 -0.175853   Durbin-Watson        2.197674
```

係数の「p 値」は, t(n-k)分布について絶対値で「t 値」より大きくなる確率である。

例えば,x の係数の t 値 1.849 で,この場合は自由度 3(＝5－2)の t 分布の 1.849 より大きい確率は 0.0808,

-1.849 より小さい確率は 0.0808 なので, 両方足して p 値は 0.1616 となっている。

Mean dependent var	2.600000	→	被説明変数の平均
Sum squared resid	4.300000	→	残差平方和
R-squared	0.532609	→	決定係数
F(1, 3)	3.418605	→	定数項を除く，説明変数の係数がすべてゼロの検定統計値
Log-likelihood	−6.717635		
Schwarz criterion	16.65415		
rho	−0.175853		
S.D. dependent var	1.516575	→	被説明変数の不偏分散
S.E. of regression	1.197219	→	回帰式の標準誤差
Adjusted R-squared	0.376812	→	自由度修正済み決定係数
P-value(F)	0.161594	→	この例では，F(1, 3) 分布で 3.418605 より大きい確率
Akaike criterion	17.43527		
Hannan-Quinn	15.33881		
Durbin-Watson	2.197674	→	DW 比

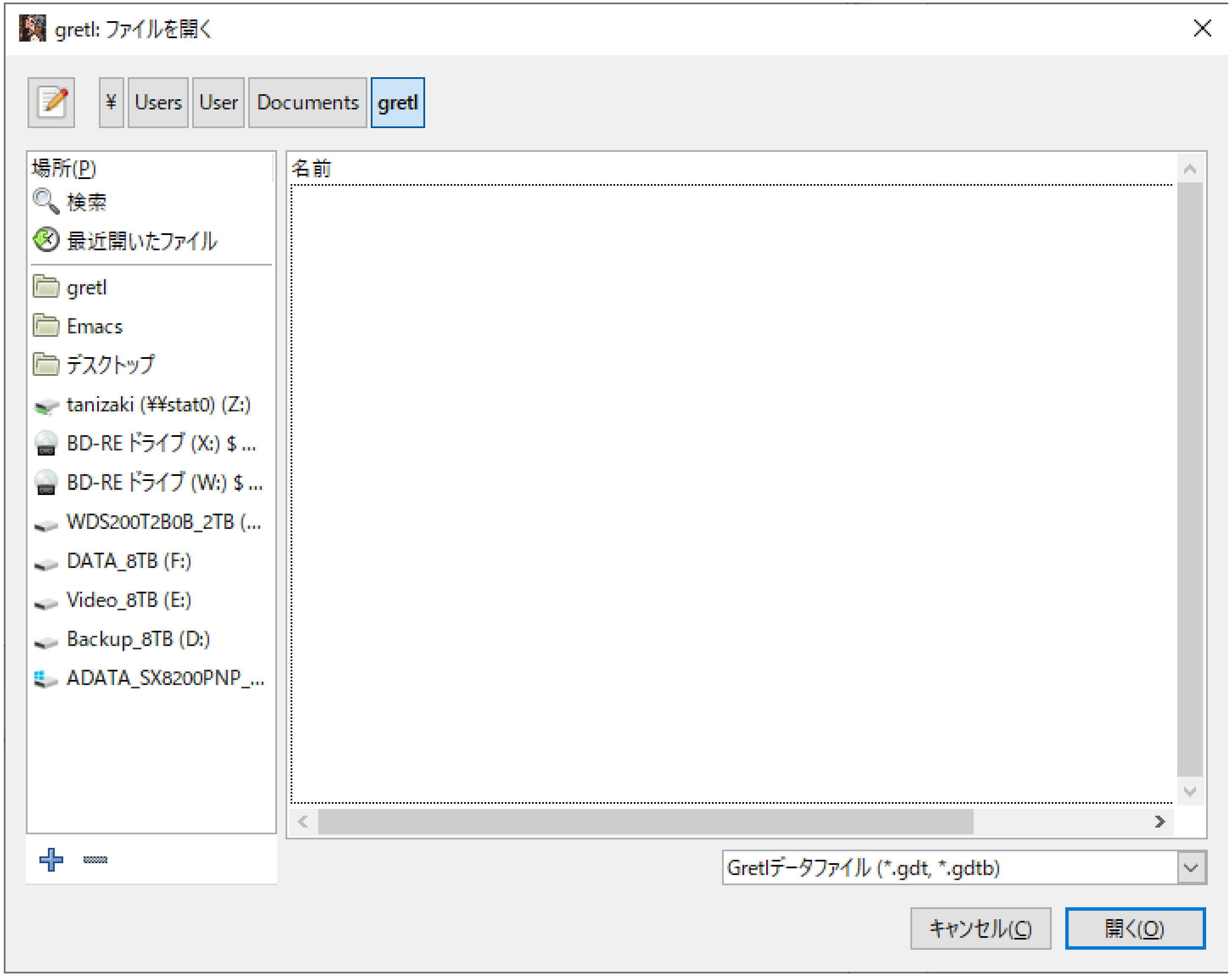
●データ入力について：　前述の「データセットの新規作成(N)」でデータ入力の方法を説明したが，Excel でデータ・ファイルを作り，gretl に読み込ませる方が便利。

次の Excel ファイルのファイル名を「data.xlsx」として保存する。

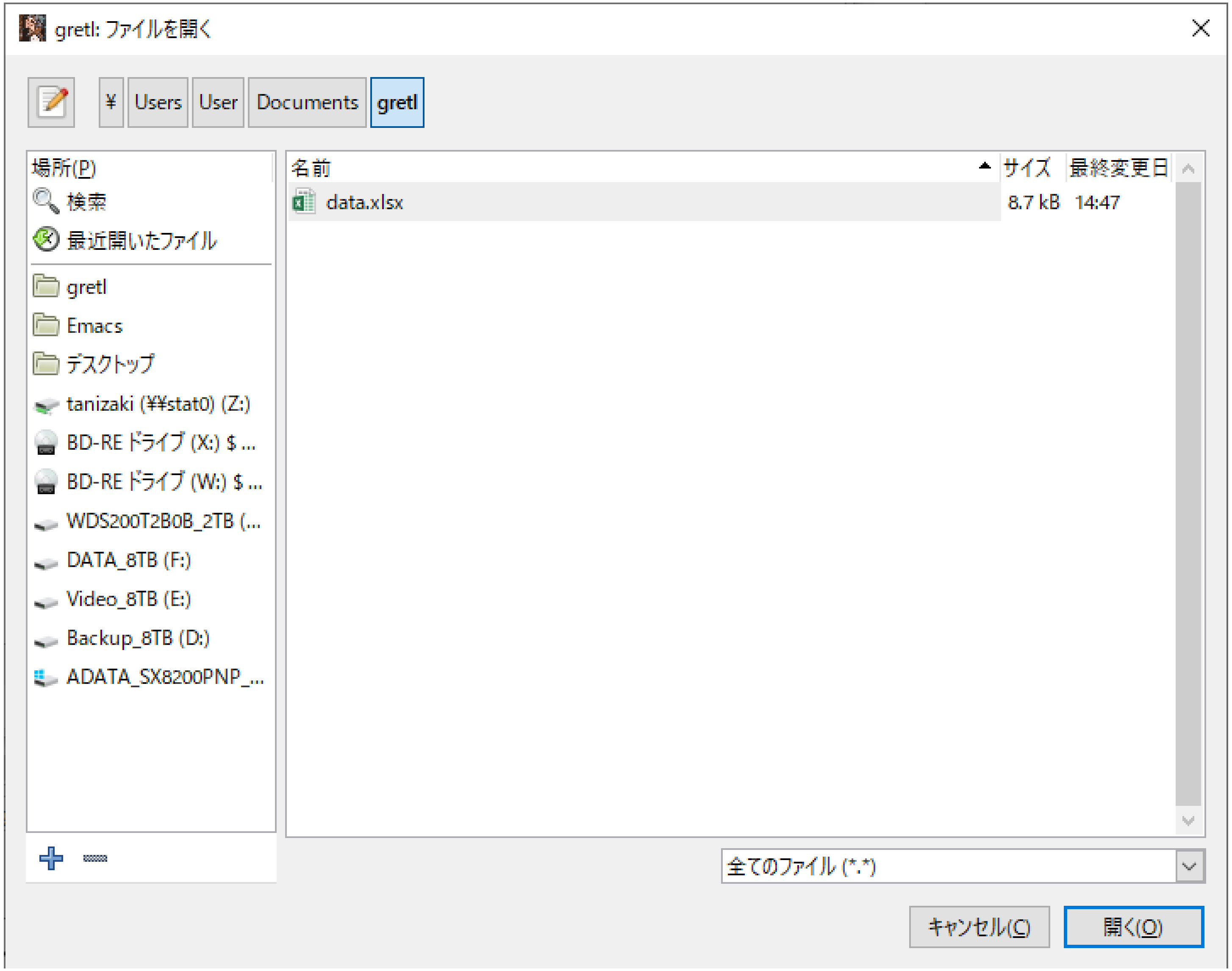
gretl のデフォルトのフォルダ（Documents¥gretl）に保存しているものとする。

	A	B
1	y	x
2	4	5
3	1	1
4	1	3
5	3	2
6	4	4

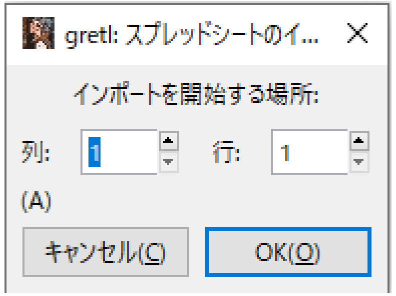
gretl で「ファイル」，「データを開く(O)」，「ユーザー・ファイル(U)」とし，次の画面になる。



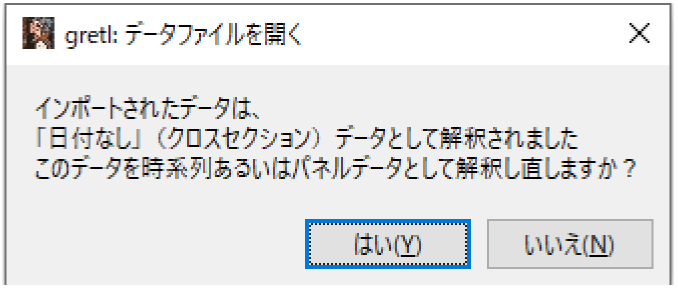
右下の「Gretl データファイル(*.gdt, *.gdtb)」のところを「全てのファイル(*.*)」にすると，data.xlsx ファイルが出てくる。



data.xlsx を選択すると次の画面が出てくる。

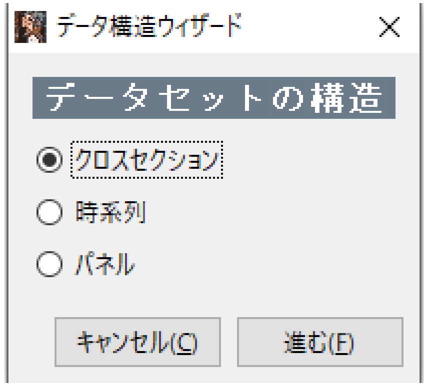


この場合は「OK(O)」で下の画面となる。

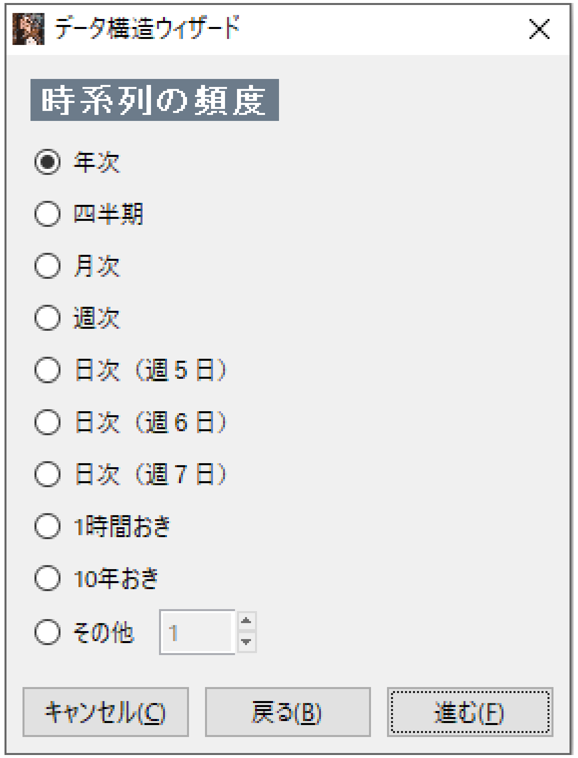


今回は，DW 比を推定結果に出力させたいので，時系列データとしてデータを読ませたい。

よって，「はい(Y)」を選択する。下記の画面へ。



「時系列」にチェックを入れて「進む(F)」を選択する。



今回は「その他」にチェックを入れて「進む(F)」を選択する。

「進む(F)」,「適用(A)」とそのまま選択していくと，変数名リストの画面（下方にアイコン付き）が出てくる。上述の「● 推定方法　その１」,「● 推定方法　その２」へ進む。

●データの変換方法について：

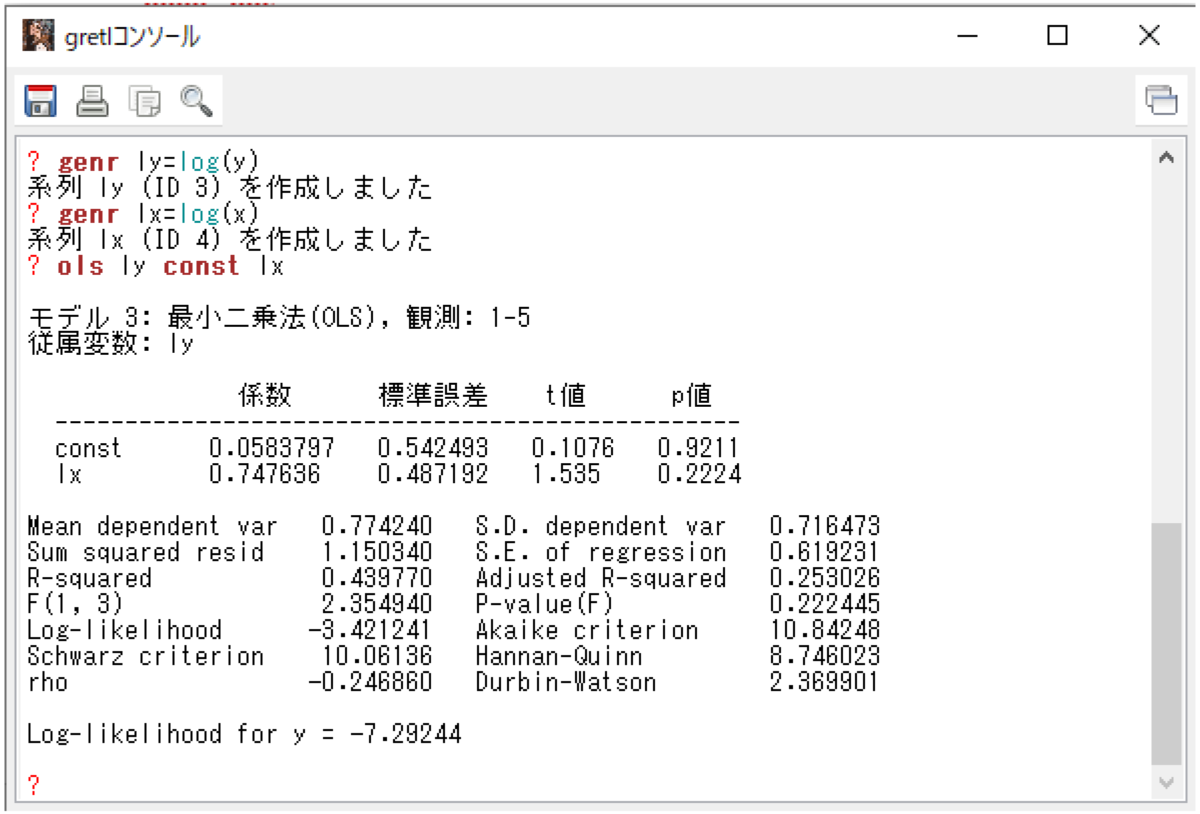
gretl コンソール画面で，

```
? genr ly=log(y)

? genr lx=log(x)

? ols ly const lx
```

と順次タイプしていくと下記の画面となる。




```
? genr ly=log(y)
系列 ly (ID 3) を作成しました
? genr lx=log(x)
系列 lx (ID 4) を作成しました
? ols ly const lx

モデル 3: 最小二乗法(OLS), 観測: 1-5
従属変数: ly

               係数         標準誤差      t値        p値
  ---------------------------------------------------------
  const       0.0583797    0.542493    0.1076    0.9211
  lx          0.747636     0.487192    1.535     0.2224

Mean dependent var    0.774240   S.D. dependent var    0.716473
Sum squared resid     1.150340   S.E. of regression    0.619231
R-squared             0.439770   Adjusted R-squared    0.253026
F(1, 3)               2.354940   P-value(F)            0.222445
Log-likelihood       -3.421241   Akaike criterion      10.84248
Schwarz criterion     10.06136   Hannan-Quinn          8.746023
rho                  -0.246860   Durbin-Watson         2.369901

Log-likelihood for y = -7.29244

?
```

$Y_i = \alpha + \beta X_i + \gamma Y_{i-1} + u_i$ を最小二乗法で推定する場合，

```
? ols y const x y(-1)
```

とタイプすると，

```
gretlコンソール

? ols y const x y(-1)

モデル 4: 最小二乗法(OLS), 観測: 2-5 (T = 4)
従属変数: y

              係数       標準誤差      t値       p値
  --------------------------------------------------------
  const      0.119048   3.50647    0.03395   0.9784
  x          0.738095   0.948504   0.7782    0.5790
  y_1        0.126984   0.816342   0.1556    0.9018

Mean dependent var   2.250000   S.D. dependent var   1.500000
Sum squared resid    4.198413   S.E. of regression   2.049003
R-squared            0.378013   Adjusted R-squared  -0.865961
F(2, 1)              0.303875   P-value(F)           0.788662
Log-likelihood      -5.772578   Akaike criterion     17.54516
Schwarz criterion    15.70404   Hannan-Quinn         13.50496
rho                 -0.162393   Durbin's h                 NA

?
```

が得られる。

7.2　誤差項：不均一分散（または，不等分散）

7.2.1　不均一分散（または，不等分散）の意味と推定方法

回帰式が

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

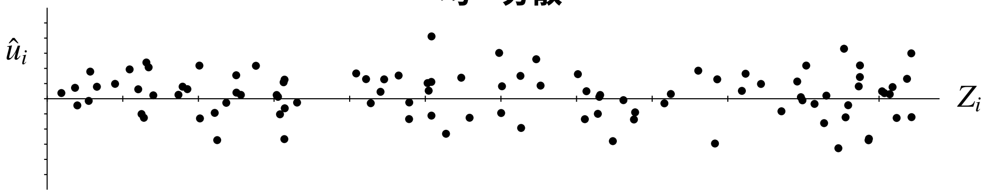
の場合を考える。X_i が外生変数，Y_i は内生変数，u_i は互いに独立な同一の分布を持つ攪乱項（最小二乗法に必要な仮定）とする。「独立な同一の分布」の意味は「攪乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に平均ゼロ，分散 σ^2 の分布する」である。

分散が i に依存する場合，代表的には，分散が他の変数（例えば，Z_i）に依存する場合，すなわち，u_i の平均はゼロ，分散は $\sigma_i^2 = \sigma_*^2 Z_i^2$ の場合は，最小二乗法の仮定に反する。ただし，Z_i は非確率変数とする。

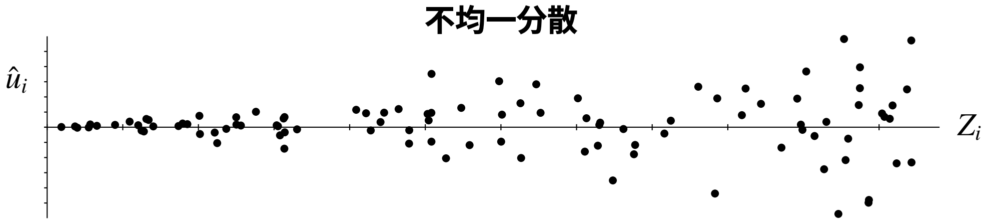
Z_i が大きくなるにつれて残差 $\hat{u}_i$ のバラツキが大きくなる場合，$\mathrm{V}(u_i) \neq \sigma^2$ となる。

下の図は $\hat{u}_i$ の範囲が Z_i に依存しない場合で，Z_i の値にかかわらず $\hat{u}_i$ の散らばりは同じ範囲となっている。

均一分散



下の図は Z_i が大きくなるにつれて $\hat{u}_i$ の範囲が広がる場合である。



図では，$n = 100$ としている。

7.2.2 不均一分散のもとで最小二乗推定量の分散について

単回帰

$$Y_i = \alpha + \beta X_i + u_i, \qquad n = 1, 2, \cdots, n$$

仮定： $\mathbf{E}(u_i) = 0$

$$\mathbf{V}(u_i) = \mathbf{E}(u_i^2) = \sigma_i^2 \qquad \longleftarrow$$ **この仮定追加**

$i \neq j$ **について，** $\mathbf{Cov}(u_i, u_j) = \mathbf{E}(u_i u_j) = 0$

不均一分散を無視して，通常の最小二乗推定量は，

$$\hat{\beta} = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} = \beta + \sum_i \omega_i u_i$$

である。ただし，$\omega_i = \dfrac{X_i - \overline{X}}{\sum_j (X_j - \overline{X})^2}$

$\mathbf{E}(\hat{\beta})$ について，$\mathbf{E}(u_i) = 0$ を利用すると，

$$\mathbf{E}(\hat{\beta}) = \mathbf{E}(\beta + \sum_i \omega_i u_i) = \beta + \sum_i \omega_i \mathbf{E}(u_i) = \beta$$

$u_1, u_2, \cdots, u_n$ の分散が不均一であっても，$\hat{\beta}$ は不偏推定量となる。

$\mathbf{V}(\hat{\beta})$ について，

$$
\begin{aligned}
\mathbf{V}(\hat{\beta}) &= \mathbf{V}(\beta + \sum_i \omega_i u_i) = \mathbf{V}(\sum_i \omega_i u_i) \\
&= \sum_i \mathbf{V}(\omega_i u_i) \qquad \longleftarrow \text{互いに独立} \\
&= \sum_i \omega_i^2 \mathbf{V}(u_i) = \sum_i \sigma_i^2 \omega_i^2 \neq \sigma^2 \sum_i \omega_i^2
\end{aligned}
$$

すなわち，$\mathbf{V}(\hat{\beta}) = \sum \sigma_i^2 \omega_i^2 = \dfrac{\sum_i \sigma_i^2 (X_i - \overline{X})^2}{\left(\sum_i (X_i - \overline{X})^2\right)^2} \neq \sigma^2 \sum_i \omega_i^2 = \dfrac{\sigma^2}{\sum_i (X_i - \overline{X})^2}$ となる。

したがって，$u_1, u_2, \cdots, u_n$ の分散が不均一であるとき，通常の最小二乗推定量 $\hat{\beta}$ は，

$$
\hat{\beta} \sim N(\beta, \sum_i \sigma_i^2 \omega_i^2)
$$

となり，標準化すると，

$$\frac{\hat{\beta} - \beta}{\sqrt{\sum_i \sigma_i^2 \omega_i^2}} \sim N(0, 1)$$

となる。

$\hat{\beta}$ の標準誤差は，$\sqrt{\sum_i s_i^2 \omega_i^2}$ とならなければならない。

s_i^2 は σ_i^2 の推定量とするが，$\sigma_i^2 = \sigma_*^2 Z_i^2$ などの仮定がなければ，s_i^2 を推定することはできない。

しかし，計量ソフトは $s^2 \sum_i \omega_i^2$ と計算する。

したがって，不均一分散にもかかわらず，最小二乗法をそのまま当てはめた場合，計量ソフトによって得られる回帰式の標準誤差，係数推定値の標準誤差，t 値，p 値，信頼区間は

すべて間違った推定結果となる。ただし，最小二乗推定量は不偏推定量となるので，係数の推定値自体は問題ない。

7.2.3　不均一分散のもとで回帰式の推定

このように，u_i の分散 $\sigma_i^2 = \sigma_*^2 Z_i^2$ の場合，分散が一定ではないため，単純には，$Y_i = \alpha + \beta X_i + u_i$ に最小二乗法を適用した場合，計量ソフトによる結果は間違ったものになる。

それでは，どのように推定すればよいのか？

u_i の分散について，$\mathrm{V}(u_i) = \sigma_i^2 = \sigma_*^2 Z_i^2$ を仮定する。回帰式は，$i = 1, 2, \cdots, n$ について，

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

とする。誤差項 u_1，u_2，$\cdots$，u_n は互いに独立であり，$u_i \sim N(0, \sigma_i^2)$ と仮定する。

$Z_i > 0$ として，回帰式の両辺を Z_i で割る。

$$\frac{Y_i}{Z_i} = \alpha\frac{1}{Z_i} + \beta\frac{X_i}{Z_i} + \frac{u_i}{Z_i}$$

すなわち，

$$Y_i^* = \alpha Z_i^* + \beta X_i^* + u_i^*$$

ただし，$Y_i^* = \dfrac{Y_i}{Z_i}$，$Z_i^* = \dfrac{1}{Z_i}$，$X_i^* = \dfrac{X_i}{Z_i}$，$u_i^* = \dfrac{u_i}{Z_i}$ とする。

このとき，新たな攪乱項 u_i^* は平均ゼロ，分散 σ_*^2 の分布となる（すなわち,「同一の」分布）。これをチェックするために，下記のように u_i^* の平均，分散を計算する。

u_i^* の平均は，

$$\mathbf{E}(u_i^*) = \mathbf{E}\left(\frac{u_i}{Z_i}\right) = \left(\frac{1}{Z_i}\right)\mathbf{E}(u_i) = 0$$

となる。u_i の仮定 $\mathrm{E}(u_i) = 0$ が使われている。

u_i^* の分散は，

$$\mathbf{V}(u_i^*) = \mathbf{V}\left(\frac{u_i}{Z_i}\right) = \left(\frac{1}{Z_i}\right)^2 \mathbf{V}(u_i) = \sigma_*^2$$

となる。u_i の仮定 $\mathrm{V}(u_i) = \sigma_*^2 Z_i^2$ が最後に使われている。このように，u_i^* の分散は，すべての i について一定 σ_*^2 となる。

誤差項 u_1，u_2，$\cdots$，u_n は互いに独立であるので，変換された誤差項 u_1^*，u_2^*，$\cdots$，u_n^* もまた互いに独立となる。

さらに，$u_i \sim N(0,\sigma_i^2)$ の仮定から，$u_i^* \sim N(0,\sigma_*^2)$ となる。

よって，$Y_i^* = \dfrac{Y_i}{Z_i}$，$Z_i^* = \dfrac{1}{Z_i}$，$X_i^* = \dfrac{X_i}{Z_i}$，$u_i^* = \dfrac{u_i}{Z_i}$ を新たな変数として，定数項なしの 2 つの説明変数で，下記の回帰式を最小二乗法に適用するればよいことがわかる。

$$Y_i^* = \alpha Z_i^* + \beta X_i^* + u_i^*$$

この回帰式は最小二乗法の仮定はすべて満たされている。

変換された変数をもとにした回帰式による推定結果では，通常の信頼区間・仮説検定などを適用することができる。

● 不均一分散の検定について：　まず，一段階目で $Y_i = \alpha + \beta X_i + u_i$ を通常の最小二乗法で推定して，残差 $\hat{u}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$ を計算する。

二段階目で，下記のように，$\hat{u}_i^2$ を Z_i^2（誤差項 u_i の分散に関係しそうな変数）に回帰させる。すなわち，

$$\hat{u}_i^2 = \gamma + \delta Z_i^2 + \epsilon_i$$

を推定し，δ の推定値 $\hat{\delta}$ の有意性の検定を行う（通常の t 検定）。