EViews

「回帰分析」

LightStone Corp

回帰分析

次の項目について学びます。

1.最小二乗法による回帰分析
2.不均一分散
3.系列相関
4.多重共線性
5.F検定と構造変化
6.最尤法によるモデル推定

最小二乗法による回帰分析

- t分布
- 回帰分析の実行
- 最小二乗法とは
- 決定係数と自由度修正済み決定係数
- 残差の分布
- 回帰の標準誤差
- 対数尤度

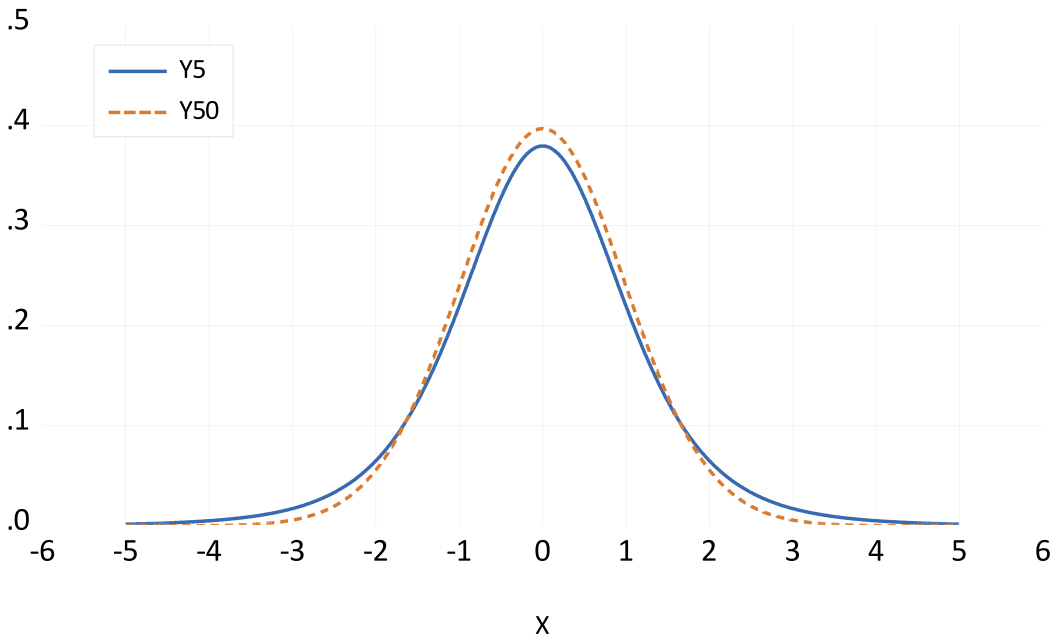
1-1.t分布

操作：eviews02に次に示すプログラムtdist.prgがあります。これを用いて、t分布について確認しましょう。

```
wfcreate(wf=mytdist) u 1000
smpl @first @first
series x=-5
smpl @first+1 @last
x=x(-1)+0.01
smpl @all
series y5=@dtdist(x,5)
series y50=@dtdist(x,50)
```

t分布の作成

操作1：x,y5,y50からなるグループオブジェクトgroup01を作成し、次のグラフ(XY-Line)を作ってください。



操作2：自由度5の時の、片側2.5%点の位置を次のコマンドで求めます。

```
show @qtdist(0.025,5)
```

練習：自由度50の時の、片側2.5%点の位置はどうなるでしょう?

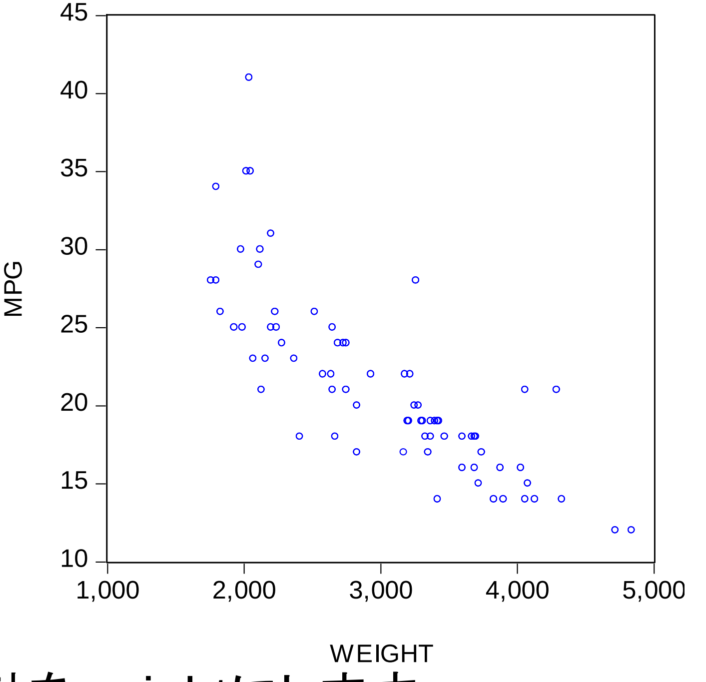
$$f(x,v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{(v\pi)^{1/2}\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}} \qquad (4)$$

1-2.回帰分析の実行

操作1:フォルダeviews02にあるauto.wf1を開きます。

これは燃費を始めとする自動車の性能に関するデータです。

操作2: weight とmpgの散布図を作成します。最初にmpgとweightのグループオブジェクトgroup01を作成します。
操作3:View/Graphとし、Specificの項目でScatterを選択します。



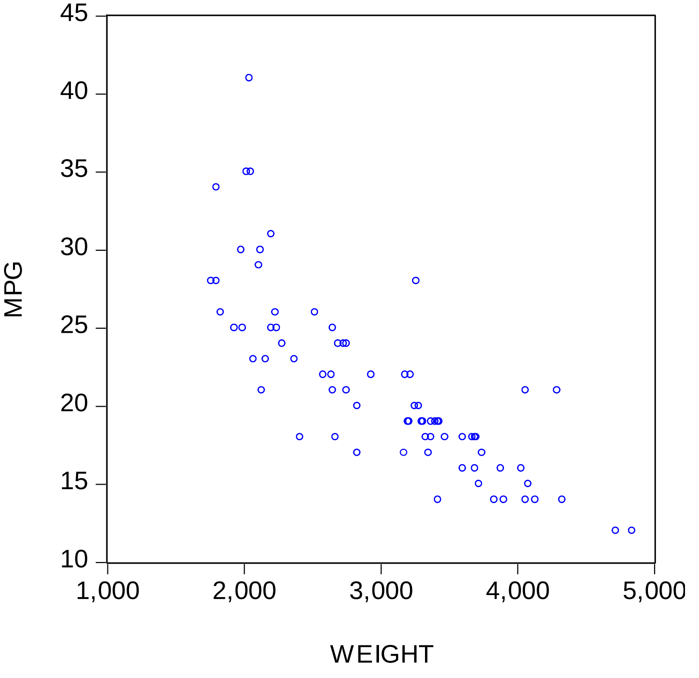
*負の相関を確認できます。横軸をweightにします。

相関係数

操作1:相関係数を求めます。group01でView/Covariance Analysisとして Correlationをチェックします。

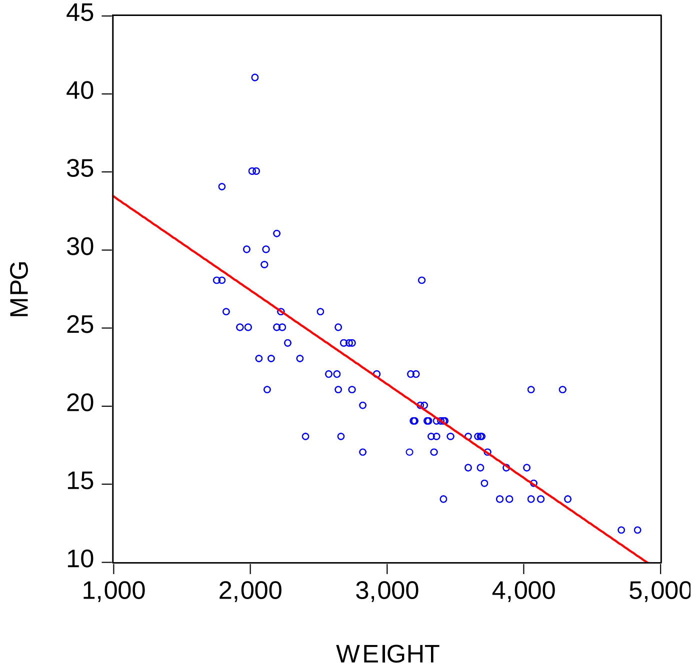
相関係数:-0.807175

車重が重いほど、燃費は悪くなるようです。



回帰直線

操作:もう一度、散布図を作成します。ただし、次のダイアログでRegression Lineを選択します。



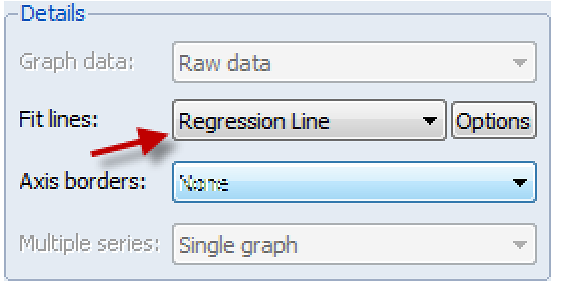


図1.回帰直線

単(純)回帰モデルの推定

操作:図1に示す回帰直線を推定します。メインメニューからQuick/Estimate Equationと操作してダイアログに次のように入力し、OKボタンをクリックします。推定式の名前はEQ01とします。

mpg c weight

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	39.44028	1.614003	24.43631	0.0000
WEIGHT	-0.006009	0.000518	-11.60251	0.0000

Variable:変数
Coefficient:推定値 (1)
Std.Error:標準誤差(推定値の標準偏差) (2)
T-Statistic:t-統計値 (3)
Prob.:p値(有意確率) (4)

係数の推定結果

情報量規準

R-squared	0.651531	Mean dependent var	21.29730
Adjusted R-squared	0.646691	S.D. dependent var	5.785503
S.E. of regression	3.438890	Akaike info criterion	5.334829
Sum squared resid	851.4693	Schwarz criterion	5.397101
Log likelihood	-195.3887	Hannan-Quinn criter.	5.359670
F-statistic	134.6182	Durbin-Watson stat	2.343928
Prob(F-statistic)	0.000000		

R-squared:決定係数 (5)
Adjusted R-squared:自由度修正済み決定係数 (6)
S.E. of regression:回帰の標準誤差 (7)
Log likelihood:対数尤度 (8)
F-statistic:F統計値 (9)
Prob(F-statistic):F統計値のp値 (10)
Mean dependent var:被説明変数の平均値
S.D. denpendent var:被説明変数の標準偏差
Durbin-Watson stat:ダービンワトソン統計量 (11)

多重回帰モデルの推定

もう少し説明力を高めましょう。

操作1:weightの二乗項weightsqと外車であることを示すforeignという変数を追加してみましょう。最初にweightsqを作成します。

```
series  weightsq=weight^2
```

操作2:counrtyは文字列変数で、米国製の場合はcountry=Domestic, 米国以外の製造メーカーの場合はcountry=Foreignとなっています。数値変数としてforeignを作成します。

```
series foreign=0
foreign=(country="Foreign")
```

操作3:counrtyとforeignのグループオブジェクトを作成して内容を確認してください。

*シリーズforeignの作成方法はこれ以外にもあります。

OLSによる推定

操作:EQ02として次の多重回帰モデルを推定します。

mpg c weight weightsq foreign

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	56.53884	6.197383	9.123019	0.0000
WEIGHT	-0.016573	0.003969	-4.175424	0.0001
WEIGHTSQ	1.59E-06	6.25E-07	2.546293	0.0131
FOREIGN	-2.203500	1.059246	-2.080253	0.0412

二乗項weightsqの意味としてどのようなことが考えられるでしょうか?

OLSによる推定

mpg c weight weightsq foreign

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	56.53884	6.197383	9.123019	0.0000
WEIGHT	-0.016573	0.003969	-4.175424	0.0001
WEIGHTSQ	1.59E-06	6.25E-07	2.546293	0.0131
FOREIGN	-2.203500	1.059246	-2.080253	0.0412

$$
\begin{aligned}
ax^2 + bx &= a\left(x^2 + \frac{b}{a}x\right) \\
&= a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a}
\end{aligned}
$$

xが2次関数に従うと仮定すると、頂点のx座標は-b/2a.

係数の操作

mpg c weight weightsq foreign

操作1:EQ02でView/Representationsと操作して係数の識別子を確認します。

MPG = C(1) + C(2)*WEIGHT + C(3)*WEIGHTSQ + C(4)*FOREIGN

b

a

操作2:mpgがweightに対して二次関数に従うと仮定して頂点のx座標を求めます。

show -c(2)/(2*c(3))

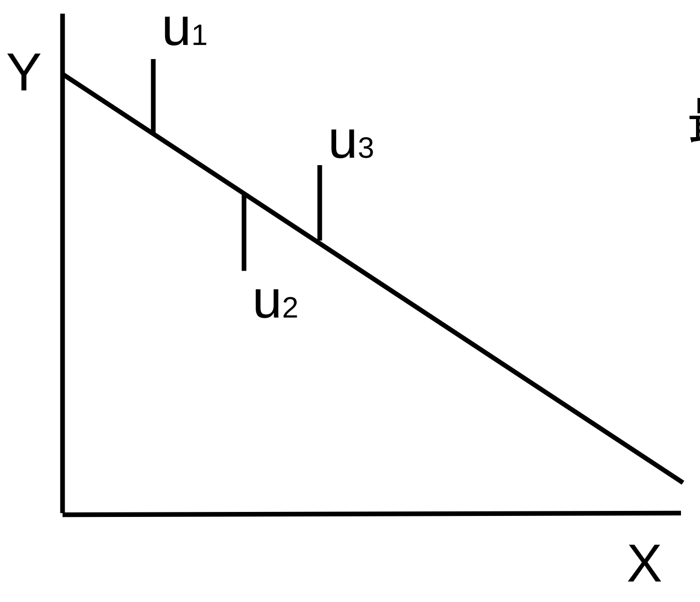
1-3.最小二乗法とは

$$Y = \alpha + \beta X$$ (直線の式)

- 回帰直線は現実のメカニズム(XとYの関係)を的確に反映している。

Y:被説明変数

X:説明変数



最小二乗法は残差uiの二乗和を最小化する

$$u_i = Y_i - \tilde{\alpha} - \tilde{\beta} X_i$$

$\tilde{\alpha}, \tilde{\beta}$:具体的な数値

最小二乗法

- 残差の絶対値の和を小さくするようにパラメータを決める

残差の二乗和をJとする。

$$J = \sum \tilde{u}_i^2 = \sum (Y_i - \tilde{\alpha} - \tilde{\beta} X_i)^2$$

$$\frac{\partial J}{\partial \tilde{\alpha}} = \sum \frac{\partial}{\partial \tilde{\alpha}} (Y_i - \tilde{\alpha} - \tilde{\beta} X_i)^2$$

$$= \sum \{-2(Y_i - \tilde{\alpha} - \tilde{\beta} X_i)\} \qquad (1)$$

$$\frac{\partial J}{\partial \tilde{\beta}} = \sum \frac{\partial}{\partial \tilde{\beta}} (Y_i - \tilde{\alpha} - \tilde{\beta} X_i)^2$$

$$= \sum \{-2X_i(Y_i - \tilde{\alpha} - \tilde{\beta} X_i)\} \qquad (2)$$

正規方程式

- 式(1),(2)をゼロにする係数を求める。

(1)より、 $$\sum\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right) = \sum Y_i - n\hat{\alpha} - \hat{\beta}\sum X_i = 0$$

(2)より、 $$\sum X_i\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right) = \sum X_i Y_i - \hat{\alpha}\sum X_i - \hat{\beta}\sum X_i^2 = 0$$

次のように整理する

$$\boxed{\begin{aligned} n\hat{\alpha} + \left(\sum X_i\right)\hat{\beta} &= \sum Y_i \\ \left(\sum X_i\right)\hat{\alpha} + \left(\sum X_i^2\right)\hat{\beta} &= \sum X_i Y_i \end{aligned}}$$

これを正規方程式と呼ぶ

推定値

- 正規方程式から、

(1)
$$\hat{\beta} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\left(\sum X_i^2\right) - n\bar{X}^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

推定値と各変数の平均値の関係として、

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

つまり、

(1)
$$\boxed{\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}}$$

このようにして観測値したデータから推定値を計算します。

確率的表現

- 撹乱項uiを用いると、

$$Y_i = \alpha + \beta X_i + u_i$$
$$\bar{Y} = \alpha + \beta \bar{X} + \bar{u}$$

(αとβは母数)

この2式を用いて、

$$\boxed{(Y_i - \bar{Y})} = \beta(X_i - \bar{X}) + (u_i - \bar{u})$$

前のスライドの式に代入すると、

$$\boxed{\hat{\beta} = \beta + \frac{\sum (X_i - \bar{X}) u_i}{\sum (X_i - \bar{X})^2}}$$

確率的表現の期待値

- 期待値を計算します。

$$E\left(\hat{\beta}\right) = E(\beta) + E\left(\frac{\sum (X_i - \bar{X})u_i}{\sum (X_i - \bar{X})^2}\right)$$

$$= \beta + \frac{\sum (X_i - \bar{X})E(u_i)}{\sum (X_i - \bar{X})^2}$$

E(ui)=0だとすると（説明は後述します）、次に示すように推定値の期待値は母数に一致します。

$$\boxed{E\left(\hat{\beta}\right) = \beta}$$

推定量の分散

- 分散は計算結果だけを表示します。

(2) $$V(\hat{\beta}) = \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}, \quad V(u_i) = \sigma^2$$

(2) $$V(\hat{\alpha}) = \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \bar{X})^2}$$

推定値の分布

- 前出の確率的表現をみてみましょう。

$$\hat{\beta} = \beta + \frac{\sum (X_i - \bar{X}) u_i}{\sum (X_i - \bar{X})^2}$$

$$\hat{\alpha} = \alpha - \sum \left\{ \frac{\bar{X}(X_i - \bar{X})}{(X_i - \bar{X})^2} - \frac{1}{n} \right\} u_i$$

撹乱項uiが正規分布の時、推定量も正規分布することを表現しています。

推定値と標準誤差

- EQ01で

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	39.44028	1.614003	24.43631	0.0000
WEIGHT	-0.006009	0.000518	-11.60251	0.0000

推定値はデータ(XiとYi)から計算しますが、標準誤差(Std.Err.)の計算には残差が大きく関係していることを覚えておきましょう。

プログラムで確認する

- EViewsプログラムolsprog.prgで計算結果を確認します。

操作:olsprogを開き、コマンドにコメントをつけましょう。そしてプログラムを実行し、コマンドの機能を確認します。

$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

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

$$V\left(\hat{\beta}\right) = \frac{\sigma^2}{\sum \left(X_i - \bar{X}\right)^2}, \quad V(u_i) = \sigma^2 \quad V(\hat{\alpha}) = \frac{\sigma^2 \sum X_i^2}{n \sum \left(X_i - \bar{X}\right)^2}$$

```
'olsprog.prg
smpl @all

table(3,3) confirm

equation eq01ndf.ls(nodf) mpg c weight

!uy=@mean(mpg)
!ux=@mean(weight)

series yd=mpg-!uy
series xd=weight-!ux

series b1=xd*yd
series b2=xd^2
```

```
'ベータの計算

!beta1=@sum(b1)/@sum(b2)

eq01.makeresids resid01

!sig2=@var(resid01)

!seb=@sqrt(!sig2/@sum(b2))
'-------------------------------
'アルファの計算

!alpha1=@mean(mpg)-!beta1*@mean(weight)

!sea=@sqrt((!sig2*@sumsq(weight))/(@obs(resid01)*@sum(b2)))
```

```
confirm(1,1)="Variable"
confirm(2,1)="C"
confirm(3,1)="WEIGHT"
confirm(1,2)="Coefficient"
confirm(1,3)="Std.Error"

confirm(2,2)=!alpha1
confirm(3,2)=!beta1

confirm(2,3)=!sea
confirm(3,3)=!seb

show  confirm
```

基準化

- 次のような計算で変数を正規分布N(0,1)に基準化できる

$$Z = \frac{X_i - \bar{X}}{SD(X)} \sim N(0,1)$$

Xiがすごく大きな値、または小さな値でも正規化すると、平均は0、分散は1になる。

$$Z = \frac{\hat{\beta} - \beta}{\sigma_{\hat{\beta}}} = \frac{\hat{\beta} - \beta}{\dfrac{\sigma}{\sqrt{\sum (X_i - \bar{X})^2}}} \sim N(0,1)$$

しかし、撹乱項uiの分散σ2(シグマ二乗)は不明。

t値

- σ2の代わりに、推定量の分散で置き換える

$$t = \frac{\hat{\beta}-\beta}{s_{\hat{\beta}}} = \frac{\hat{\beta}-\beta}{\dfrac{s}{\sqrt{\sum (X_i - \bar{X})^2}}} \quad (3)$$

tは正規分布ではなく、自由度m=n-2のt分布に従う。

操作：次のコマンドでEQ01のWEIGHTのt値を求めてみましょう

```
scalar myt=eq01.@coefs(2)/eq01.@stderr(2)
```

p値

操作:EQ01のp値を求めます。回帰分析の自由度は、推定に用いるデータの個数74から、推定するパラメータの個数2を引いた72です。

(4) scalar myp=@ctdist(myt,72)*2

帰無仮説と対立仮説

- 変数foreignを例に考えましょう。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	56.53884	6.197383	9.123019	0.0000
WEIGHT	-0.016573	0.003969	-4.175424	0.0001
WEIGHTSQ	1.59E-06	6.25E-07	2.546293	0.0131
FOREIGN	-2.203500	1.059246	-2.080253	0.0412

$$t = \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}}$$

p値が(0.0412/2)の時のt値はいくつになりますか?自由度は(データの個数-定数項を含む変数の数)です。

帰無仮説H0: $\beta = 0$

対立仮説H1: $\beta \neq 0$

$$t = \frac{\hat{\beta}}{s_{\hat{\beta}}}$$

4.決定係数

（標本）相関係数/(標本)単相関係数

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}} \qquad -1 \leq r \leq 1$$

回帰直線の決定係数

$$R^2 = \frac{\hat{Y}\text{で説明された部分}}{Y\text{の全変動}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad 0 \leq R^2 \leq 1$$

マクロ経済分析の場合：
　　0.8（まあまあ良い）、0.9(良い)、0.5以下(あまり良くない)
クロスセクションデータ分析の場合:
　　0.5(極めて良い)
決定係数と単相関係数（単純回帰モデル）の関係

$$R^2 = r^2$$

自由度修正済み決定係数

- 説明変数を増やすと残差が小さくなり、R2が1に近くなる

(5)
$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

- 自由度修正済み決定係数

(6)
$$\boxed{\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2/(n-K)}{\sum\left(Y_i - \bar{Y}\right)^2/(n-1)}}$$

説明変数を増やしても、K（説明変数の個数）により自由度修正済み決定係数は、小さくなるとは限らない。

自由度修正済み決定係数

- 応用

R2が小さい場合、修正R2は負になることもある。

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-K}$$

n=10,K=3,R2=0.1の場合、修正R2は-0.157

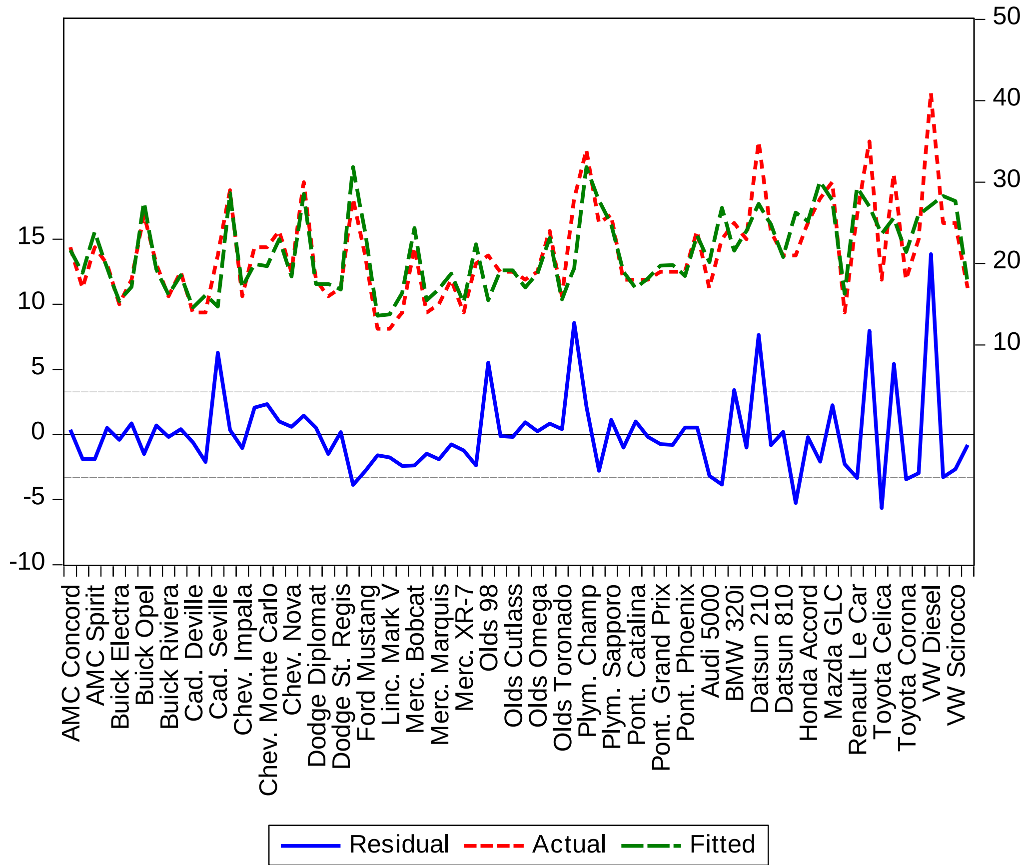
クロスセクションデータの分析や、被説明変数が階差
データの場合にこのような現象が生じることがある。

残差の分布

- 視覚的に残差の分布を確認します。

操作1：EQ02でResidsボタンをクリックして残差プロットを作成します。Freezeボタンをクリックし、rvfplot1として保存します。

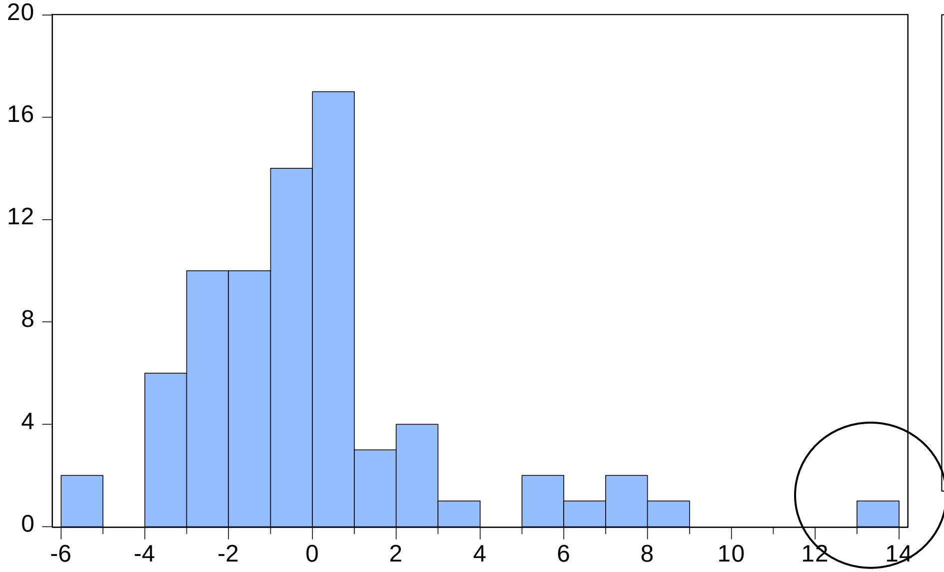
操作2:残差を取り出してみましょう。EQ02でProc/Make Residual Series…と操作します。名前は「u」とします。



残差プロット

操作：uの記述統計量を求めます。シリーズuのウィンドウでView/Descriptive Statistics & Tests/Histogram & Statsと操作し、平均値が0、最大値が14に近いことを確認します。

ヒストグラムを見ると右裾の厚い分布のようです。



Series: U
Sample 1 74
Observations 74

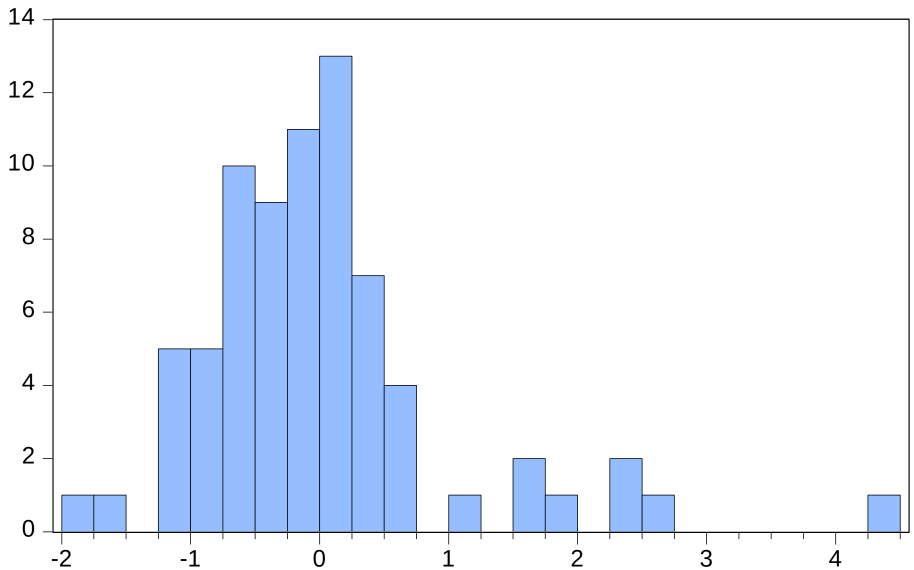
Mean	-7.94e-16
Median	-0.301969
Maximum	13.85172
Minimum	-5.636126
Std. Dev.	3.214491
Skewness	1.709467
Kurtosis	7.337345
Jarque-Bera	94.04682
Probability	0.000000

外れ値

標準化した残差を求めて、正規分布曲線の両側1%(左右0.5%) に含まれるデータを探します。

操作：次のコマンドで標準化した残差u1を求め、分布を確認します。

series u1= (u-@mean(u))/@stdev(u)



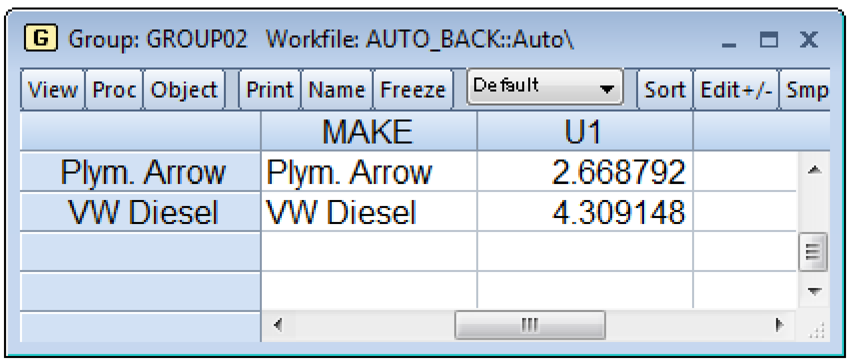
Series: U1
Sample 1 74
Observations 74

Mean	1.80e-17
Median	-0.093940
Maximum	4.309148
Minimum	-1.753349
Std. Dev.	1.000000
Skewness	1.709467
Kurtosis	7.337345
Jarque-Bera	94.04682
Probability	0.000000

外れ値

操作1：製造メーカーmakeと標準化残差u1のグループオブジェクトgroup02を作成します。
操作2:コマンドウィンドウでu1が標準正規分布の両側1%(左右0.5%)に含まれるという条件を設定します。

smpl @all if u1>2.58



2つの車は理論値に比べ、実質的に燃費がかなり優れていることが分かります。

回帰の標準誤差

- モデルの当てはまりが良いほど、残差は正規分布に近づき、誤差の分散は小さくなる。

(7)　S.E of regression　$$s = \sqrt{\frac{\hat{\epsilon}'\hat{\epsilon}}{(T-k)}}$$

操作:次のコマンドをコマンドウィンドウに入力し、eq02の回帰の標準誤差を求めてみましょう。

```
series u2=u^2
show @sqrt(@sum(u2)/(74-4))
```

対数尤度

- モデルの当てはまりが良いほど、残差を標準化したものは標準正規分布に近づく。

⇩

対数尤度は大きくなる

対数尤度は最小二乗法以外の推定量の場合にも利用できる、便利なモノサシ

(8) $$ll = -\frac{T}{2}(1 + \log(2\pi) + \log(\hat{\epsilon}'\hat{\epsilon})/T)$$

まとめ

- 最小二乗法推定量
- 推定値の標準誤差
- 推定量の確率的表現
- t値とp値
- 決定係数と自由度修正済み決定係数
- 残差の分布

MEMO

2.不均一分散

- 標準的仮定と不均一分散
- 不均一分散の仮説検定
- 標準誤差のオプション
- 加重最小二乗法
- クラスターへの対応

標準的仮定

1.説明変数は確率変数ではなく、固定した値を持つ。
2.データが無限に増えると、説明変数の偏差の二乗和も無限大になる。
3.誤差項の期待値はゼロである。
4.誤差項の分散は一定である。 ⇦ 不均一分散の問題
5.誤差項に系列相関はない。
6.誤差項は正規分布にしたがう。

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

仮定4	$V(u_i) = E(u_i^2) = \sigma^2$	(均一分散)

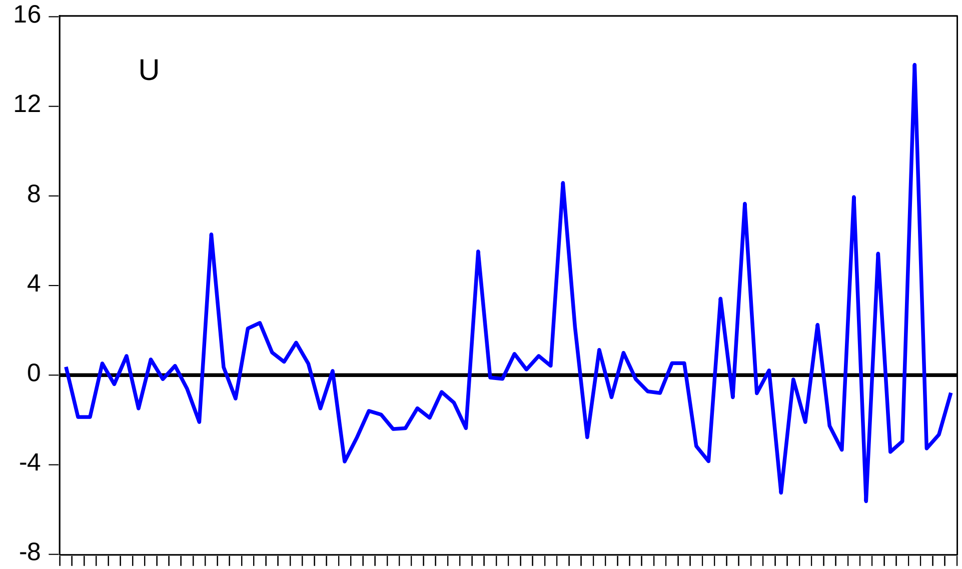
$$V(u_i) = E(u_i^2) = \sigma_i^2$$ (不均一分散)

不均一分散

- 回帰モデルEQ02の推定後に診断を行います。

残差は均一だと考えていいでしょうか?

推定後のモデルについて色々と調査することを診断(diagnostics)といいます。



不均一分散

操作：EQ02でView/Residual Diagnostics/Heteroskedasticityと操作し、ダイアログでBreush-Pagan-Godfreyを選択し、OKボタンをクリックします。

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	3.314783	Prob. F(3,70)	0.0248
Obs*R-squared	9.204926	Prob. Chi-Square(3)	0.0267
Scaled explained SS	26.09939	Prob. Chi-Square(3)	0.0000

帰無仮説は「分散は均一である」(H0:Constant variance)です。p値はいずれも0.05より小さくなり、帰無仮説を棄却しています。

⇩

不均一分散であることが分かりました

Breusch-Pagan(1979) and Godfrey(1978)

- BPGによるラグランジュ乗数検定

残差の分散(不均一分散)の定式化

$$\sigma_t^2 = \sigma^2 h(z_t' \alpha) \tag{3}$$

帰無仮説:分散均一

一般的にz_tは元の回帰式の説明変数。

具体的な手順: 元の回帰式の残差の二乗系列を、定数項を含む変数z_t(複数の変数)に回帰させる(補助回帰)。この時のESSを次のように計算した値がカイ二乗の検定統計量となる。自由度は変数z_tの数。

$$\frac{ESS}{2\hat{\sigma}^4} \sim \chi^2_{(n)} \quad \left(R^2 = \frac{ESS}{TSS} \right) \tag{5}$$

Breusch-Pagan(1979) and Godfrey(1978)

Koenker(1981)

$$Obs \times R^2$$

R2は補助回帰の決定係数。

$$\begin{aligned} \log(m1) &= c(1) + c(2) \times \log(ip) + c(3) \times tb3 \\ resid\text{^}2 &= c(1) + c(2) \times \log(ip) + c(3) \times tb3 \end{aligned}$$

F検定：補助回帰式のF値検定。

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	3.314783	Prob. F(3,70)	0.0248
Obs*R-squared	9.204926	Prob. Chi-Square(3)	0.0267
Scaled explained SS	26.09939	Prob. Chi-Square(3)	0.0000

Koenker

BPG

不均一分散

帰無仮説は棄却される。

$$V(\hat{\beta}) = \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \Rightarrow \boxed{\frac{\sigma_i^2}{\sum (X_i - \bar{X})^2}}$$

推定値の分散の計算式が異なる！

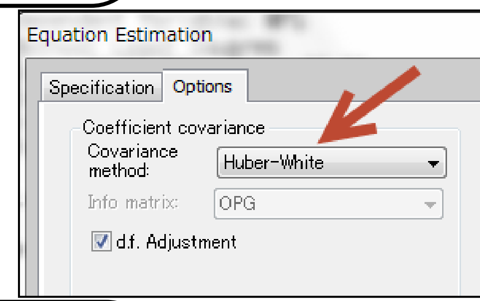
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	56.53884	6.197383	9.123019	0.0000
WEIGHT	-0.016573	0.003969	-4.175424	0.0001
WEIGHTSQ	1.59E-06	6.25E-07	2.546293	0.0131
FOREIGN	-2.203500	1.059246	-2.080253	0.0412

標準誤差を利用するt値とp値も額面通りに理解できなくなります。

不均一分散だと判定したら

- 推定のオプション

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	56.53884	6.197383	9.123019	0.0000
WEIGHT	-0.016573	0.003969	-4.175424	0.0001
WEIGHTSQ	1.59E-06	6.25E-07	2.546293	0.0131
FOREIGN	-2.203500	1.059246	-2.080253	0.0412



オプション:Huber-White
EQ02のEstimationダイアログのOptionsタブを
利用します

White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	56.53884	5.522689	10.23756	0.0000
WEIGHT	-0.016573	0.003436	-4.822897	0.0000
WEIGHTSQ	1.59E-06	5.27E-07	3.019811	0.0035
FOREIGN	-2.203500	1.029480	-2.140402	0.0358

不均一分散

- オプション:white

White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	56.53884	5.522689	10.23756	0.0000
WEIGHT	-0.016573	0.003436	-4.822897	0.0000
WEIGHTSQ	1.59E-06	5.27E-07	3.019811	0.0035
FOREIGN	-2.203500	1.029480	-2.140402	0.0358

オプションを利用すると、標準誤差が変わります。しかし、推定値は変わりません。

$$\hat{v} = \hat{V}\left(\sum_{j=1}^{N} u_i' u_i\right)\hat{V}$$

$$\hat{V} = \left(\frac{-\partial^2 \ln L}{\partial \beta^2}\right)^{-1}$$

操作:ワークファイルautoを上書き保存して閉じます。加重最小二乗法は他のワークファイルで練習します。

不均一分散への対応

- 加重最小二乗法という推定手法がありますが、加重データを準備する必要がありますので、実用は難しい。

通常の最小二乗法 $$J = \sum (Y_i - \alpha - \beta X_i)^2$$

加重最小二乗法 $$J = \sum \frac{1}{\sigma_i^2}(Y_i - \alpha - \beta X_i)^2$$

加重最小二法は次の仮説が成立するような変数を利用する。

$\sigma_i = cZ_i$ または $\sigma_i^2 = c^2 Z_i^2$

加重最小二乗法

- 仮に分散が既知とした場合

$$J = \sum \frac{1}{\sigma_i^2}(Y_i - \alpha - \beta X_i)^2$$

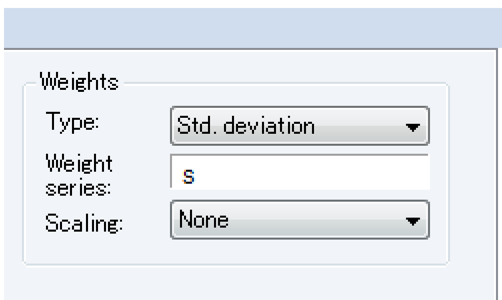
次のコマンドを順番に実行して、最小二乗法と加重最小二乗法の推定結果を確認しましょう。

操作1:ワークファイルvwlsxmpl.wf1を開きます。Equationオブジェクトolsとして次のモデルを推定します。

y c x

加重最小二乗法

操作：次にEquationオブジェクトwlsとして同じモデルを推定します。ただし、加重シリーズとしてs(標準偏差)を利用しますので、Estimationダイアログ のOptionsタブで次の図のように設定します。



OLS

Variable	Coefficient	Std. Error
C	0.160714	0.134871
X	0.980952	0.026708

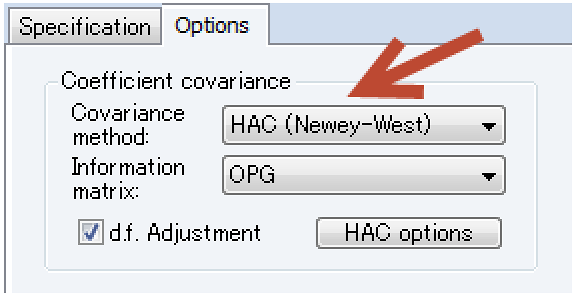
WLS

Weighting series: S
Weight type: Standard deviation (no scaling)

Variable	Coefficient	Std. Error
C	0.113855	0.112008
X	0.982468	0.037074

不均一分散

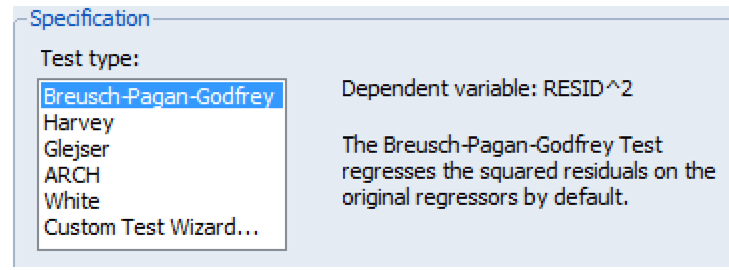
- Whiteのオプションは不均一分散のみを仮定、不均一分散と系列相関の両方を想定する場合はHAC(Newey-West)オプションを利用する。



その他の検定手法

- Whiteのオプションは不均一分散のみを仮定、不均一分散と系列相関の両方を想定する場合はHAC(Newey-West)オプションを利用する。

- 不均一分散のその他の検定



Harvey(1976)

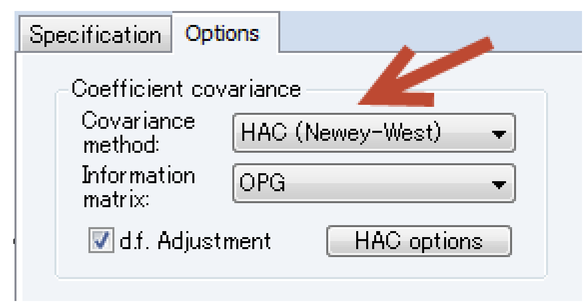
帰無仮説:分散均一

対立仮説: $\sigma_t^2 = \exp(z_t'\alpha)$

Glejser(1969)

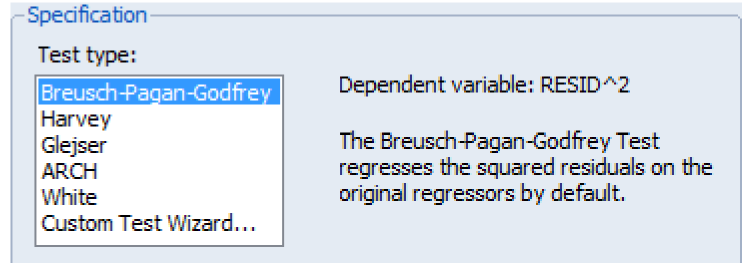
帰無仮説:分散均一

対立仮説: $\sigma_t^2 = (\sigma^2 + z_t' a)^m, m = 1,$



その他の検定手法

- 不均一分散のその他の検定



ARCH (Engle 1982)
帰無仮説:ARCH効果なし

対立仮説: $e_t^2 = \beta_0 + (\sum_{s=1}^{q} \beta_s e_{t-s}^2) + v_t$

White(1980)
帰無仮説:分散均一

対立仮説:定式化なし

$$y_t = b_1 + b_2 x_t + b_3 z_t + e_t$$

$$e_t^2 = \alpha_0 + \alpha_1 x_t + \alpha_2 z_t + \alpha_3 x_t^2 + \alpha_4 z_t^2 + \alpha_5 x_t z_t + v_t$$

クラスターへの対応

- サンプルがグループ(クラスター)化されている場合

$$E(\epsilon_i \epsilon_j) \neq 0$$ iとjは同じグループ

$$E(\epsilon_i \epsilon_h) = 0$$ iとhは異なるグループ

このような場合、不均一分散と同様、標準誤差の計算式が標準のものとは異なる

クラスターへの対応

操作：ptersen_cluster.wf1を利用して次に示すモデルを推定します。

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

EQ01:デフォルトの設定でOLS推定を行う

EQ02:クラスターIDとしてfirmidを利用する

EQ03:さらにyearをスペース区切りでクラスタに追加する

まとめ

- 不均一分散の検定:帰無仮説は「均一分散」
- Whiteのオプション
- 加重最小二乗法
- クラスターへの対応

3.系列相関

$$Y_i = \alpha + \beta X_i + u_i \quad (i = 1, 2, \cdots, n) \qquad (7)$$

iは時間的な前後関係を持つものと考えます。

仮定5　$\mathrm{COV}(u_i, u_j) = 0$　(系列相関なし)

系列相関の要因

$$Y_i = \alpha + \beta X_i + \gamma W_i + v_i \quad (i = 1, 2, \cdots, n) \quad (8)$$

Wiが系列相関を持っているとき、(8)式でなく、(7)式を推定すると、

$\gamma W_i + v_i$ は強い系列相関をしめす。

系列相関の影響

$$Y_i = \alpha + \beta X_i + u_i \quad (i = 1, 2, \cdots, n) \qquad (7)$$

$$\sigma_{\hat{\beta}}^{+2} = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} + \frac{\sum \sum (X_i - \bar{X})(X_j - \bar{X})\sigma_{ij}}{\left\{ \sum (X_i - \bar{X})^2 \right\}^2}$$

第2項の分子の符号によって、分散が増減する。

もし、符号が正であるとする。ソフトウェアで計算する標準誤差は第1項の部分だけになるので、真の分散を過小評価する結果になる。

t値が大きくなるので、係数は有意であると判定する可能性が高くなる

ダービン=ワトソン統計量

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \quad (i = 1, 2, \cdots, n) \quad (9)$$

$$u_i = \rho u_{i-1} + \varepsilon_i \quad (10)$$

$$\boxed{H_0 : \rho = 0, \quad H_1 : \rho \neq 0}$$

$$\hat{\rho} = \frac{\sum_2^n \hat{u}_i \hat{u}_{i-1}}{\sum_2^n \hat{u}_i^2}$$

しかし、ρ の分布は標本数が大きい場合しか分かっていない。

標本数が15
個以上なら ⇨ $\boxed{DW = \frac{\sum_2^n (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_1^n \hat{u}_i^2}}$ $\quad DW \simeq 2(1-\rho)$

DWの分布

DWの分布は標本数nと定数項以外の説明変数の数mに依存します。

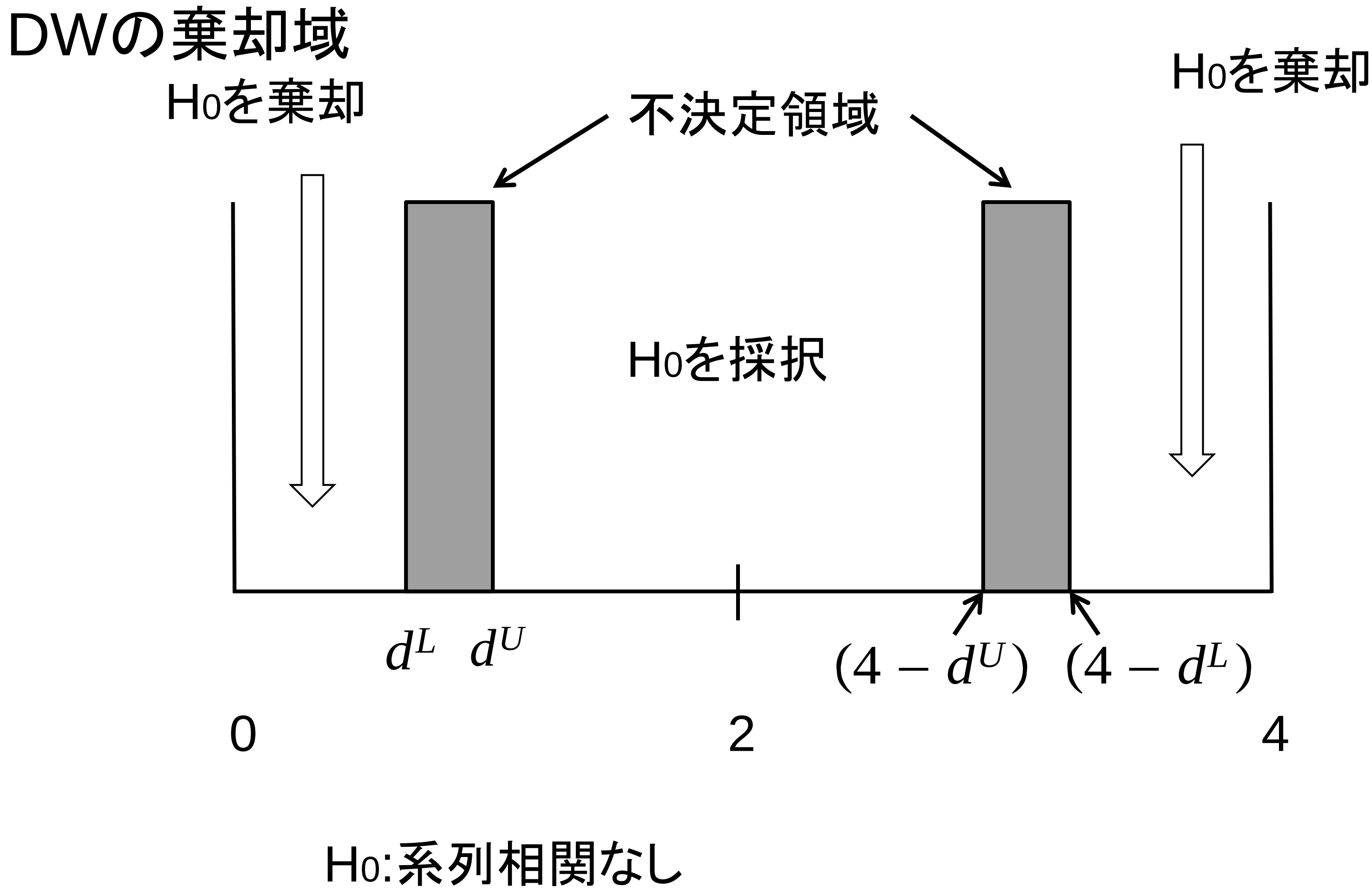
(11)
$$d_{n,m} = DW = \frac{\sum_2^n (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_1^n \hat{u}_i^2}$$

$d_{n,m}$の分布は2を中心とした対称分布です。

有意水準をαとすると、この分布の下方のα×100%の臨界値$d_{n,m,\alpha}$は表として計量経済学の教科書の付録などとして用意されています。

$$P(d_{n,m} < d_{n,m,\alpha}) = \alpha$$

DWの分布



ダービン=ワトソン統計量

操作:ワークファイルklein.wf1を開いて次に示す消費関数EQ01を推定します。

consump c wagegovt

$$H_0 : \rho = 0, \quad H_1 : \rho \neq 0$$

Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	40.84699	3.192183	12.79594	0.0000
WAGEGOVT	2.507440	0.595717	4.209111	0.0004

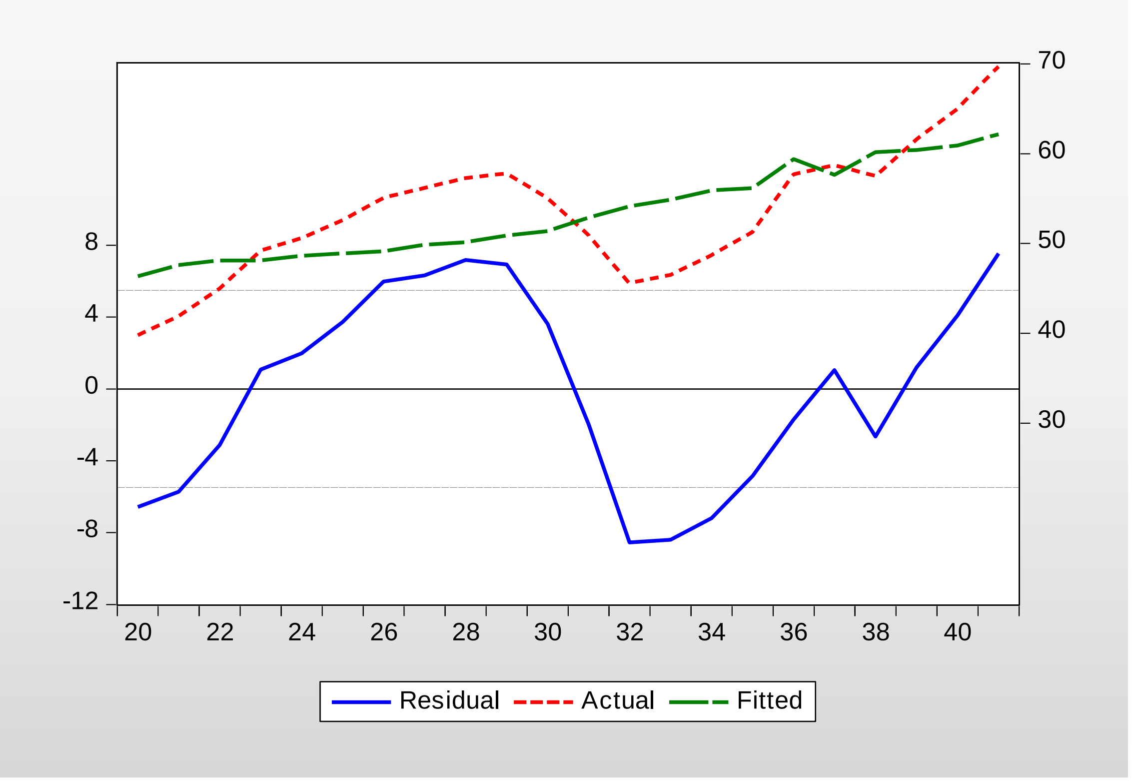
R-squared	0.469730	Mean dependent var	53.35000
Adjusted R-squared	0.443216	S.D. dependent var	7.347740
S.E. of regression	5.482733	Akaike info criterion	6.327592
Sum squared resid	601.2072	Schwarz criterion	6.426778
Log likelihood	-67.60352	Hannan-Quinn criter.	6.350958
F-statistic	17.71661	Durbin-Watson stat	0.321800
Prob(F-statistic)	0.000431		

問:ダービン=ワトソン統計値は0.3218、データの個数は22、変数の数は1個です。有意水準5%の両側検定で考えた時、d_L=1.12,d_U=1.31が分かっています。系列相関は存在するでしょうか?

ダービン=ワトソン統計量

■DW検定でチェックできるのは1次の系列相関だけです。
■説明変数に被説明変数のラグ項を含む場合は利用できません。

操作:系列相関ありという示唆を得ました。残差をプロットしてその様子を確認しましょう。EQ01でResidsボタンをクリックします。



一般的な系列相関の検定

■Breusch-Godfrey LM Test

$$Y_i = \alpha + \beta X_i + u_i \quad (i = 1, 2, \cdots, n)$$

$$u_i = \gamma X_i + \left(\sum_{s=1}^{p} \alpha_s u_{i-s}\right) + v_i$$

操作:EQ01でView/Residual Diagnostics/Serial Correlation LM Test…と操作します。ラグは自己相関の存在を想定した整数を入力します。ここでは2とします。

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	24.88483	Prob. F(2,18)	0.0000
Obs*R-squared	16.15668	Prob. Chi-Square(2)	0.0003

*帰無仮説は系列相関なし。

自由度はラグ次数

残差の自己相関

$$Y_i = \alpha + \beta X_i + u_i \quad (i = 1, 2, \cdots, n)$$

操作:EQ01でView/Residual Diagnostics/Correlogram–Q-statistics と操作し、ラグを12とします。

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.756	0.756	14.370	0.000
		2	0.408	−0.382	18.765	0.000
		3	0.056	−0.223	18.853	0.000
		4	−0.188	−0.035	19.885	0.001
		5	−0.435	−0.414	25.756	0.000
		6	−0.568	−0.080	36.392	0.000
		7	−0.612	−0.205	49.569	0.000
		8	−0.561	−0.271	61.421	0.000
		9	−0.385	0.086	67.433	0.000
		10	−0.069	0.149	67.644	0.000
		11	0.196	−0.188	69.493	0.000
		12	0.324	−0.140	75.047	0.000

*帰無仮説は自己相関なし。

*点線より上は「ゼロと有意に異なる」

arma誤差項

操作:次のように手順に入力して系列相関を考慮したモデルEQ02を推定します。

consump c wagegovt ar(1)

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (i = 1, 2, \cdots, n) \qquad (9)$$

$$u_i = \rho u_{i-1} + \varepsilon_i \qquad (10)$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	69.88511	39.94381	1.749585	0.0972
WAGEGOVT	0.474856	1.621232	0.292898	0.7729
AR(1)	0.934632	0.112000	8.344890	0.0000

R-squared	0.829063	Mean dependent var	53.99524
Adjusted R-squared	0.810070	S.D. dependent var	6.860865
S.E. of regression	2.990031	Akaike info criterion	5.160008
Sum squared resid	160.9251	Schwarz criterion	5.309226
Log likelihood	-51.18009	Hannan-Quinn criter.	5.192392
F-statistic	43.65097	Durbin-Watson stat	0.848400
Prob(F-statistic)	0.000000		

Inverted AR Roots	.93

ar項に関する詳細は「時系列分析」の講習会でご説明します。

4.多重共線性

多重回帰モデルにおいて相関の強い変数をともに説明変数として利用すると、多重共線性が発生します。ほとんどの場合、計算はそのまま終了してしまいますので、常に注意が必要です。

多重共線性による問題点:

1.サンプルを増減すると推定値が大きく変化する。
2.説明変数を入れ替えると推定値が大きく異なる。
3.推定値の標準誤差が大きくなり、有意性の検定に問題が生じる。
4.R2やF値は大きいが、個別のt値は小さい。

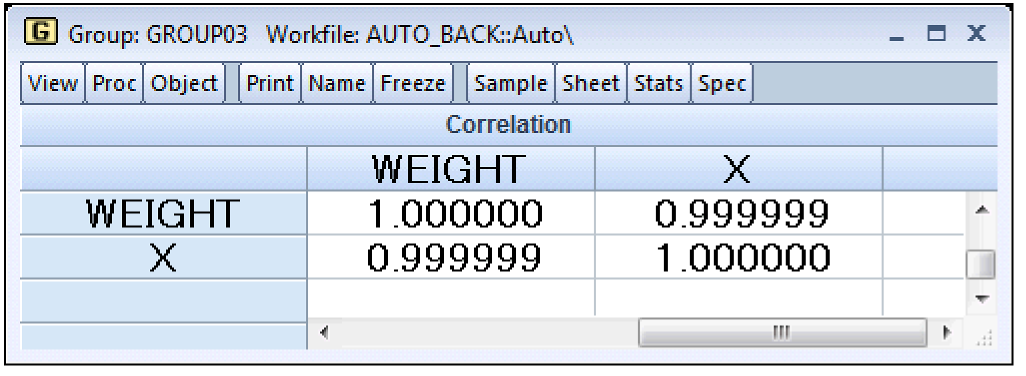
多重共線性

例えば、次のような変数を作成して、多重共線性について確認しましょう。変数Xとweightは線形関係にあります。

操作1:ワークファイルautoを再び開きます。次のようにしてweightと相関の強い変数xを作成します。rndは一様分布の関数です。

```
series x=rnd*3+10+weight
```

操作2:xとweightのグループを作成し、View/Covariance Analysisと操作し、ダイアログでCorrelationだけ選択します。



	WEIGHT	X
WEIGHT	1.000000	0.999999
X	0.999999	1.000000

多重共線性

極めて相関の強い変数Xを追加して推定を行います。

操作:次のモデルEQ03を推定します。

mpg c weight x

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	43.79199	5.588896	7.835534	0.0000
WEIGHT	0.369652	0.461806	0.800448	0.4261
X	-0.375678	0.461827	-0.813460	0.4187

R-squared	0.654749	Mean dependent var	21.29730
Adjusted R-squared	0.645024	S.D. dependent var	5.785503
S.E. of regression	3.446997	Akaike info criterion	5.352580
Sum squared resid	843.6069	Schwarz criterion	5.445988
Log likelihood	-195.0454	Hannan-Quinn criter.	5.389841
F-statistic	67.32374	Durbin-Watson stat	2.332086
Prob(F-statistic)	0.000000		

F値は大きくなっていますが、weightとxのパラメータは有意ではありません

多重共線性の原因

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (i = 1, 2, \dots n)$$

多重回帰モデルの正規方程式は、

$$n\hat{\beta}_1 + \left(\sum X_{2i}\right)\hat{\beta}_2 + \left(\sum X_{3i}\right)\hat{\beta}_3 = \sum Y_i$$

$$\left(\sum X_{2i}\right)\hat{\beta}_1 + \left(\sum X_{2i}\right)^2\hat{\beta}_2 + \left(\sum X_{2i}X_{3i}\right)\hat{\beta}_3 = \sum X_{2i}Y_i$$

$$\left(\sum X_{3i}\right)\hat{\beta}_1 + \left(\sum X_{2i}X_{3i}\right)\hat{\beta}_2 + \left(\sum X_{3i}^2\right)\hat{\beta}_3 = \sum X_{3i}Y_i$$

ここで、次のようなケースを考える。

$$X_{2i} = c + kX_{3i}$$

仮にc=0として多重回帰モデルの正規方程式を求めると、

多重共線性の原因

変数間に線形関係があると、

$$n\hat{\beta}_1 + \left(\sum X_{3i}\right)\hat{\beta}_2 + \left(\sum X_{3i}\right)\hat{\beta}_3 = \sum Y_i$$

$$k\left(\sum X_{3i}\right)\hat{\beta}_1 + k^2\left(\sum X_{3i}^2\right)\hat{\beta}_2 + k\left(\sum X_{3i}^2\right)\hat{\beta}_3 = k\sum X_{3i}Y_i$$

$$\left(\sum X_{3i}\right)\hat{\beta}_1 + k\left(\sum X_{3i}^2\right)\hat{\beta}_2 + \left(\sum X_{3i}^2\right)\hat{\beta}_3 = \sum X_{3i}Y_i$$

2行目は3行目のk倍ですから、推定値は一意に決まらない。

相関係数が完全に1という事はない。しかし、推定値の分散は、

例えば、

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma^2}{\sum \hat{v}_i^2} \Leftarrow$$

X2とX3の相関が強いと分母は小さくなる。

$$(X_{2i} = \gamma_1 + \gamma_2 X_{3i} + v_i)$$

結論:推定値の分散が大きくなってしまう。

分散拡大要因

分散拡大要因を確認するVIFコマンドで多重共線性の状態を確認できます。

操作：EQ03でView/Coefficient Diagnostics/Variance Inflation Factorsと操作します。

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	31.23576	194.5369	NA
WEIGHT	0.213265	12900981	791441.4
X	0.213284	12994060	791441.4

判定：
最大のCentered VIFが10よりも大きい
Centered VIFの平均が1に比べてかなり大きい

⇒ 多重共線性あり

対応方法

相関の強い説明変数の中から一つ外して、回帰を行います。

変数選択

- 帰無仮説H_0:β=0が棄却されることを望む

帰無仮説が棄却されるとき(p値が0.05よりも小さいとき)、推定値 $\hat{\beta}$ は有意である(ゼロでない)という。

t検定は説明変数としてのXの妥当性を検証していることになり、変数選択の役目を果たすことになる。

有意水準は研究者が自分で決めるものであるが、5%が最も標準的である。その次によく用いられるのが1%である。例外的に10%,20%とする場合もある。

推定結果の評価と対策

- 符号条件

最初に係数推定値の符号条件を満たし、係数が有意であれば、その推定式は及第であり、予測などの分析に利用できる。

$$mpg = 56.54 - 0.0166 \cdot weight + 1.59 \times 10^{-6} \cdot weightsq - 2.20 \cdot foreign$$

問:符号条件は妥当だと考えられますか?

推定結果の表現

- 推定式を報告書に記述する場合は次のように書きます

$$Y_i = 56.54 - 1.66 \times 10^{-2} X_i + 1.59 \times 10^{-6} W_i - 2.20 Z_i$$

(9.12)　(-4.18)　(2.55)　(-2.08)

$$R^2 = 0.691, \bar{R}^2 = 0.678, s = 3.2827$$

カッコの中にはt値を書きます。自由度が30以上の場合、t値が2以上であれば、有意になります。研究者によっては、カッコ内に標準誤差を書くこともあります。「カッコ内はt値を示す」という一文を付け加えましょう。

変数の過不足とその影響

- 不必要な変数を含むと、推定量の分散を増大させる。逆に必要な変数を除くことは推定量の不偏性や一致性を失う。
- 過少定式化による誤りの方が過剰定式化による誤りよりはるかに深刻である。

MEMO

F検定と構造変化

K個の説明変数を持つ多重回帰モデル

$$
\begin{aligned}
Y_i \;=\; & \beta_1 + \beta_2 X_{2i} + \cdots + \beta_{K-G} X_{K-G,i} + \\
& \beta_{K-G+1} X_{K-G+1,i} + \cdots + \beta_K X_{Ki} + u_i \quad (i = 1, 2, \ldots n)
\end{aligned} \tag{11}
$$

$$
\boxed{
\begin{aligned}
H_0 &: \beta_{K-G+1} = \cdots = \beta_K = 0 \ (G\text{個の制約}) \\
H_1 &: H_0\text{でない}
\end{aligned}
}
$$

(11)式はH_1モデル。H_0モデルは次のようになる。

H_0モデル

$$
Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_{K-G} X_{K-G,i} + v_i \quad (i = 1, 2, \ldots n) \tag{12}
$$

F検定と構造変化

H_1モデル(11式)とH_0モデル(12式)をOLSで推定し、次式について考える。

$$\boxed{\sum \hat{v}_i^2 \geq \sum \hat{u}_i^2}$$

$\sum \hat{v}_i^2 - \sum \hat{u}_i^2$ がH_0の正しさを判断する基準になる。

(9) $$F = \frac{\left(\sum \hat{v}_i^2 - \sum \hat{u}_i^2\right)/G}{\sum \hat{u}_i^2/(n-K)}$$

F統計値は自由度G,n-KのF分布に従う。 $F \sim F_{G,n-K}$

（2つの自由度に依存）

$$\boxed{\begin{array}{l} F \geq F_{G,n-K,\alpha}\text{の時、有意水準}100\alpha\%\text{で}H_0\text{を棄却} \\ F < F_{G,n-K,\alpha}\text{の時、有意水準}100\alpha\%\text{で}H_0\text{を採択} \end{array}}$$

F検定と構造変化

決定係数による表現

一般的に、 $R^2 = \dfrac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = 1 - \dfrac{\hat{u}_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}$

H0モデルの決定係数: R_0^2 $\qquad \sum \hat{v}_i^2 = (1 - R_0^2) \sum y_i^2$

H1モデルの決定係数: R_1^2 $\qquad \sum \hat{u}_i^2 = (1 - R_1^2) \sum y_i^2$

$$F = \frac{\left(R_1^2 - R_0^2\right)/G}{\left(1 - R_1^2\right)/(n - K)}$$

よく用いられる例

$H_0 : \beta_2 = \cdots = \beta_K = 0$ ($G - 1$個の制約)

$H_1 : H_0$でない

操作:EQ03のF値とp値を確認してください。

線形制約の検定

複数の係数について特定の値を考える

H₁モデル:K=5の場合

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i \qquad (13)$$

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

$H_0 : \beta_1 = \beta_1^*, \beta_2 = \beta_2^*$ ($G = 2$個の制約)

$H_1 : H_0$でない

*具体的な数値を仮定します。

$$Y_i = \beta_1^* + \beta_2^* X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + v_i$$

$$Y_i' = Y_i - \beta_1^* - \beta_2^* X_{2i}$$

H₀モデル

$$Y_i' = \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + v_i \qquad (14)$$

ワルド検定

操作1:coef_test.wf1を開きます。次に示すコブ･ダグラス型の生産関数をEQ01として推定します。

$$\log Q = A + \alpha \log L + \beta \log K + \epsilon$$

log(q) c log(l) log(k)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.327939	0.410601	-5.669595	0.0000
LOG(L)	1.591175	0.167740	9.485970	0.0000
LOG(K)	0.239604	0.105390	2.273498	0.0331

操作2:一次同次であるとい仮説をワルド検定で検定します。EQ01でView/Coefficient Diagnostics/Wald-Coefficient Restrictions… として次のように入力します。

c(2)+c(3)=1

ワルド検定

Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
t-statistic	10.95526	22	0.0000
F-statistic	120.0177	(1, 22)	0.0000
Chi-square	120.0177	1	0.0000

Null Hypothesis: C(2)+C(3)=1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(2) + C(3)	0.830779	0.075834

Restrictions are linear in coefficients.

一次同次という帰無仮説は棄却されています。

F統計値は次のようなケースでは仮説検定には利用できません。
■非線形モデル
■ARMAモデル
■Newey-West, Whiteの利用時。

ワルド統計量

次の線形回帰モデルを仮定します。

$$y = X\beta + \epsilon$$

線形制約を次のように記述します。

$$H_0 \ : R\beta - r = 0$$

Rはq×k行列(qはH0下の制約数)、rはqベクトル。
ワルド統計量は、H0下で漸近的に自由度qのカイ二乗分布に従います。

$$W = (Rb - r)'\left(Rs^2(X'X)^{-1}R'\right)^{-1}(Rb - r)$$

線形回帰モデルの撹乱項がiidの場合、

$$F = \frac{W}{q}$$

構造変化の検定

構造変化:ある時点から定数項あるいは係数が変化しているか?

第n_1期によって標本期間が2つに分けられるモデル

$$前半: Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_{1i} \quad (i = 1, 2, \ldots n_1)$$

$$後半: Y_i = \beta_1' + \beta_2' X_{2i} + \beta_3' X_{3i} + u_{2i} \quad (i = n_1 + 1, n_1 + 2, \ldots, n)$$

H_0:構造変化なし

$$\beta_1 = \beta_1',\ \beta_2 = \beta_2',\ \beta_3 = \beta_3'$$

つまり、H0モデルは全期間で、

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + v_i \quad (i = 1, 2, \ldots n) \quad (15)$$

構造変化の検定

H1モデル $Y_i = \beta_1 D1_i + \beta_2 D1_i X_{2i} + \beta_3 D1_i X_{3i}$

$$+ \beta_1' D2_i + \beta_2' D2_i X_{2i} + \beta_3' D2_i X_{3i} + u_i \qquad (16)$$

$$D1_i = \begin{cases} 1 \\ 0 \end{cases} \qquad D2_i = \begin{cases} 0 & 1 \le i \le n_1 \\ 1 & n_1 + 1 \le i \le n \end{cases}$$

$$F = \frac{(u'u - (u_1' u_1 + u_2' u_2))/k}{(u_1' u_1 + u_2' u_2)/(T - 2k)}$$

チャウ(Chow)テスト

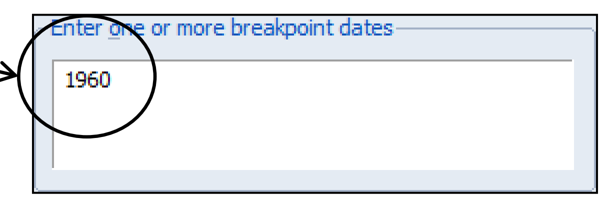
構造変化の時点に注意

EQ01の推定期間は1947年から1971年です。
そこで、次のように考えた場合の検定方法を実行します。

前半:1947-1959
後半:1960-1971

操作:EQ01でView/Stability Diagnostics/Chow Breakpoint Test…と操作します。

後半の始期
を入力します。



チャウ(Chow)テスト

3つの検定統計量について

Chow Breakpoint Test: 1960
Null Hypothesis: No breaks at specified breakpoints
Varying regressors: All equation variables
Equation Sample: 1947 1971

F-statistic	4.980645	Prob. F(3,19)	0.0102
Log likelihood ratio	14.50531	Prob. Chi-Square(3)	0.0023
Wald Statistic	14.94194	Prob. Chi-Square(3)	0.0019

帰無仮説:
設定した時点で構造
変化はない

F-statistic:残差がiidであることを前提とする
LR:尤度比検定統計量。
(自由度は(m-1)k。kはパラメータの個数、mはサブサンプル数。)
Wald Statistic:ワルド統計量

チャウテスト

- プログラムを利用する chow.prg

$$F = \frac{(u'u - (u_1'u_1 + u_2'u_2))/k}{(u_1'u_1 + u_2'u_2)/(T-2k)}$$

```
'チャウ検定と同じ結果を得る
smpl @all
!obs=@obssmpl

'全期間モデルの推定
equation eq01.ls log(q) c log(l) log(k)
!k=eq01.@ncoefs
eq01.makeresids u
```

チャウテスト

```
'1960年をブレークポイントとするサンプルオブジェクトの作成
sample s1 @first 1959
sample s2 1960 @last

'期間ごとのモデル推定
smpl s1
equation eq021.ls log(q) c log(l) log(k)
eq021.makeresids u1

smpl s2
equation eq022.ls log(q) c log(l) log(k)
eq022.makeresids u2
```

チャウテスト

```
'検定統計量の作成
smpl @all
series usq=u^2
series u12=u1^2
series u22=u2^2

!u2=@sum(usq)
!u12=@sum(u12)
!u22=@sum(u22)

!num=(!u2-(!u12+!u22))/!k

!dnm=(!u12+!u22)/(!obs-2*!k)
```

チャウテスト

```
'手作業による検定統計量
!fv=!num/!dnm

show !fv

'コマンドによる検定結果の表示
eq01.chow 1960
```

チャウ(Chow)テスト

例えば、ダイアログに 1960 1970 と入力すると、サブサンプルの期間を1950-1959, 1960-1969, 1970-1994に分割して構造変化の診断を行います。

EViewsはこの他に、Quant-Andrews Breakpoint Test, Multiple Break point Testなどをサポートしています。

6.最尤法によるモデル推定

- プロビットモデルの限界効果を求める

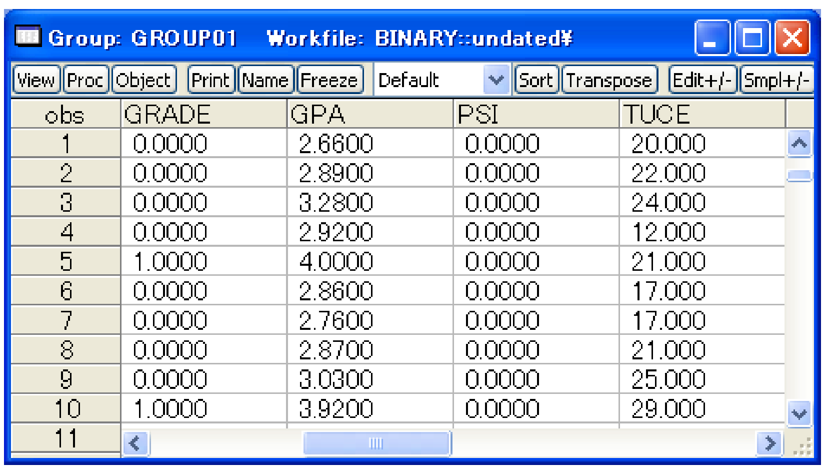
操作：ワークファイルbinary.wf1を開きます。被説明変数grade、説明変数gpa、tuce、psiの意味を確認します。32人の生徒に関するデータです。時系列データではありません。

grade：合格者を示すダミー
gpa:成績評価点の平均、tuce:過去の試験の得点、psi：講義方法のダミー

分析の狙い

- 合格確率に最も大きな影響を及ぼす変数は何か?その変数を見つけること。

操作1:grade,gpa, psi, tuceの変数からなるグループオブジェクトgroup01を作成し、内容を確認します。



obs	GRADE	GPA	PSI	TUCE
1	0.0000	2.6600	0.0000	20.000
2	0.0000	2.8900	0.0000	22.000
3	0.0000	3.2800	0.0000	24.000
4	0.0000	2.9200	0.0000	12.000
5	1.0000	4.0000	0.0000	21.000
6	0.0000	2.8600	0.0000	17.000
7	0.0000	2.7600	0.0000	17.000
8	0.0000	2.8700	0.0000	21.000
9	0.0000	3.0300	0.0000	25.000
10	1.0000	3.9200	0.0000	29.000
11				

被説明変数gradeは合格1または不合格を示す0のダミー変数。説明変数psiもダミー変数です。

操作2:Proc/Make Equationとしてprobitモデル「eq_probit」を作成します。

推定結果

- バイナリモデルの推定結果の特徴的なところだけを説明します。

Dependent Variable: GRADE
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 02/18/13 Time: 15:56
Sample: 1 32
Included observations: 32
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

収束までに要した繰り返し計算回数。

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-7.452320	2.542472	-2.931131	0.0034
GPA	1.625810	0.693882	2.343063	0.0191
TUCE	0.051729	0.083890	0.616626	0.5375
PSI	1.426332	0.595038	2.397045	0.0165

係数は前述した理由により限界効果とは理解できません。ただし、密度関数は正なので限界効果の正負は、推定値の符号に左右されることは分かる。

$$\frac{\partial \mathrm{E}(y_i|x_i,\beta)}{\partial x_{ij}} = f(-x_i'\beta)\beta_j$$

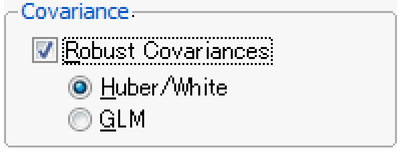
$$f(x) = dF(x)/dx$$

例）係数が正なら、この変数が増加すれば、確率は高くなる。

推定結果

- Estimationのオプションタブにある係数共分散行列について

Dependent Variable: GRADE
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 02/18/13 Time: 15:56
Sample: 1 32
Included observations: 32
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives



デフォルト

Newton-Raphson,Quadratic Hill Climing法… $\hat{H}^{-1}$

BHHH法… $(\hat{g}\hat{g}')^{-1}$

Huber/White

疑似最尤法で(yの誤った分布に対して)堅牢な疑似最尤標準誤差を計算します。

GLM

yが指数分布族に属し、yiの条件付き平均が線形部x'iβを非線形変換したものになる。

推定結果

LR Statistics（尤度比統計量）

$$LR = -2(l_r - l_u)$$

尤度比検定の自由度は制約の数です。

McFadden R-squared	0.377478	Mean dependent var	0.343750
S.D. dependent var	0.482559	S.E. of regression	0.386128
Akaike info criterion	1.051175	Sum squared resid	4.174660
Schwarz criterion	1.234392	Log likelihood	-12.81880
Hannan-Quinn criter.	1.111907	Deviance	25.63761
Restr. deviance	41.18346	Restr. log likelihood	-20.59173
LR statistic	15.54585	Avg. log likelihood	-0.400588
Prob(LR statistic)	0.001405		
Obs with Dep=0	21	Total obs	32
Obs with Dep=1	11		

帰無仮説は「係数は同時にゼロ」

Obs with…0と1を取る観測値の個数。

データの個数

どの程度、あたるか?

- 推定したモデルの適合度をExcelを使って調べます。

操作1:Forecast機能で予測確率を計算します。シリーズ名をgrade_fとします。
操作2:Excelファイル「適合度」を開きます。
操作3:gradeの値をA列にコピー&ペーストします。実際に合格した人の人数が「実際の合格者数」の上に入ります。
操作4:C列には実際に合格した人のうち、計算確率が50%以上の人の数のセルに1を入力します。

	A列		C列
19	0	0.030851	
20	0	0.593402	
21	1	0.657186	1
22	0	0.061929	
23	1	0.904539	1
24	0	0.273191	
25	0	0.84745	
26	1	0.834195	1
27	1	0.488726	
28	1	0.642407	1
29	0	0.328673	
30	1	0.840017	1
31	1	0.952245	1
32	0	0.539959	
33	1	0.123544	
34	11		8
35	実際の合格者数		実際の合格者で、合格確率50%以上の生徒
36			
37			適合度
38			0.727272727

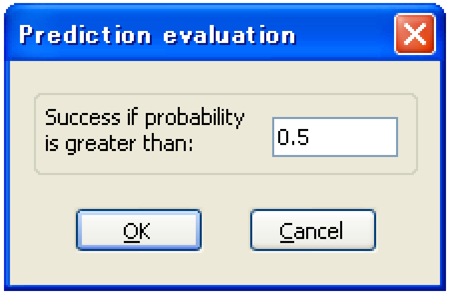
適合度=C列の人数/A列の人数　　　適合度は8/11=0.7273

EViewsで適合度を見る

- 推定したモデルの適合度を調べます。

操作1：eq_probitでView/Expectation-Prediction Evaluationと操作します。

操作2：1(合格)として判定する境界値を入力します。普通はデフォルトの0.5とします。



Expectation-Prediction Evaluation for Binary Specification
Equation: EQ_PROBIT
Date: 02/14/08 Time: 10:23
Success cutoff: C = 0.5

	Estimated Equation			Constant Probabi...		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	18	3	21	21	11	32
P(Dep=1)>C	3	8	11	0	0	0
Total	21	11	32	21	11	32
Correct	18	8	26	21	0	21
% Correct	85.71	72.73	81.25	100.00	0.00	65.63
% Incorrect	14.29	27.27	18.75	0.00	100.00	34.38
Total Gain*	-14.29	72.73	15.63			
Percent Ga...	NA	72.73	45.45			

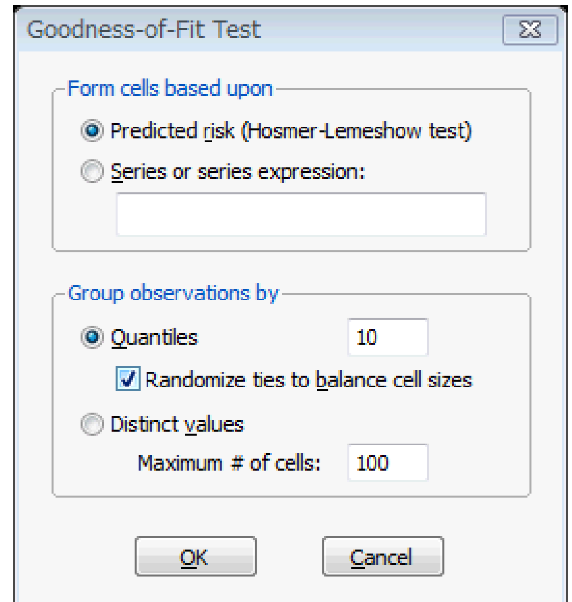
ゼロを正しく認識したもの(特異度)…18/21=85.71%
1を正しく認識したもの(感度)…8/11=72.73%
全体で正しく認識したのも…26/32=81.25%

Excelの値と同じですか？

適合度検定

操作1：eq_probitでView/Goodness-of –Fit Test…と操作します。

操作2：デフォルトの状態でOKボタンをクリックします。



2つのカイ二乗値を算出していますが、結論が異なっています。帰無仮説は「期待度数(モデル)と観測度数は統計的に等しい」

データ数が少ない場合はこのようなことが生じやすい。

H-L Statistic	6.9086	Prob. Chi-Sq(8)	0.5465
Andrews Statistic	20.6045	Prob. Chi-Sq(10)	0.0240

限界効果の計算

- 最初に変数GPAの限界効果を求めます。32人のデータを集めて推定しましたが、彼らのデータの平均を使って、Xiβ'(インデックス)を求めます。それを使って限界効果を計算します(平均値の回りでの限界効果)。

サンプルの限界効果:@dnorm(X'β)*(GPAの係数)。

成績が平均の人を想定し、その人たちの限界効果を求めます。

限界効果を求めるためにはコマンドウィンドウまたはプログラムファイル(ウィンドウ)を利用します。

限界効果の計算

操作1：File/New/Programとしてプログラム画面を表示します。
操作2：次のコマンドを入力します。

```
series me_gpa=@dnorm(c(1) + c(2)*@mean(GPA) +
                      c(3)*@mean(PSI) + c(4)*@mean(TUCE))*c(2)
```

操作3：入力が完了したら、Saveボタンをクリックしてプログラムファイルを Data9に保存します。名前は「prog_me」とします。
操作4：「Run」ボタンをクリックしてプログラムを実行します。
操作5：シリーズオブジェクトme_gpaの値を確認します。

me_gpa:0.5333

得点が平均的な人の場合、GPAの得点が1ポイント上がるとほぼ間違いなく合格することが分かります。

パラメータと限界効果

操作：TUCEと定数項について、パラメータ、z統計量、限界効果を次の表に記入してください。定数項には限界効果はありません。ダミー変数PSIの限界効果は計算方法が異なりますので後で計算します。

説明変数	係数	Z統計量	P値	限界効果
GPA	**1.6258**	**2.343063**	**	**0.5333**
ダミー変数PSI				
TUCE				
定数項				不要です

サンプルサイズ：
Log likelihood:
擬似R^2:

*(有意水準10%), **(有意水準5%), ***(有意水準1%)

パラメータと限界効果

TUCEの限界効果を追加

説明変数	係数	Z統計量	P値	限界効果
GPA	**1.6258**	**2.34306**	**	**0.5333**
ダミー変数 PSI				
TUCE	**0.0517**	**0.6166**	**ns**	**0.0170**
定数項	**-7.4523**	**-2.9311**	***	不要です

サンプルサイズ:32
Log likelihood:-12.828
擬似R^2:0.3775

*(有意水準10%), **(有意水準5%), ***(有意水準1%)

ダミー変数の限界効果

- 簡単な例を用いて説明します。

$$y_i^* = a + b_1 x_i + b_2 Dummy_i + u_i$$

ダミー変数の効果は、

ダミーが1の場合: $P(y_i = 1) = \Phi(a + b_1 x_i + b_2)$

ダミーが0の場合: $P(y_i = 0) = \Phi(a + b_1 x_i)$

これらの差として求めます。

$$\Phi(a + b_1 x_i + b_2) - \Phi(a + b_1 x_i)$$

限界効果の計算

ダミー変数が1とゼロの場合に分けて、それぞれ計算します。
PSI=1の場合を「xbd1」、PSI=0の場合を「xbd0」、そしてPSIの限界効果をme_psiに格納するコードを追加します。分布関数の微分は行いませんので注意してください。

```
‘ psiが1
series xbd1=@cnorm(c(1) + c(2)*@mean(GPA) + c(3) +
                                   c(4)*@mean(TUCE))

‘psiが0
series xbd0=@cnorm(c(1) + c (2)*@mean(GPA)  +
                                   c (4)*@mean(TUCE))

‘最後に両者の差を計算します
series me_psi=xbd1- xbd0
```

パラメータと限界効果

TUCEの限界効果を追加

説明変数	係数	Z統計量	P値	限界効果
GPA	**1.6258**	**2.3431**	**	**0.5333**
ダミー変数 PSI	**1.4263**	**2.3970**	**	**0.4644**
TUCE	**0.0517**	**0.6166**	**ns**	**0.0170**
定数項	**-7.4523**	**-2.9311**	***	不要です

サンプルサイズ:32
Log likelihood:-12.828
擬似R^2:0.3775

*(有意水準10%), **(有意水準5%), ***(有意水準1%)

全体のまとめ

- 最小二乗法
- 標準正規分布とt分布
- 標準的仮定(不均一分散、系列相関)
- 経済データにおける構造変化
- 最尤法による質的選択モデルの推定

山本 拓著,計量経済学の構成

I 基礎編:回帰分析

- 最小二乗法:直線のあてはめ
- 単純回帰分析
- 多重回帰モデル

II 応用編:計量経済学

- モデルの関数型と特殊な変数
- F検定と構造変化の検定
- 分布ラグ・モデル
- 標準的仮定の意味と不均一分散
- 撹乱項の系列相関
- 説明変数と撹乱項の相関
- 同時方程式モデル

EViews関連情報

EViewsの基本的な用法や役立つ情報を次のウェブサイトで公開しています。

http://www.lightstone.co.jp/eviews

補足:推定2

限界効果:線形回帰モデルの場合

$$Y = \alpha + \beta X + u$$

$$\beta = \frac{dY}{dX}$$

⇒ Xが1単位増えるとYはβだけ増える

$$\log Y = \alpha + \beta \log X + u$$

$$\frac{dY}{Y} = \beta \frac{dX}{X}$$

$$\beta = \frac{dY}{Y} / \frac{dX}{X}$$

⇒ Xが1%増えるとYはβ%増える

$$Y = \alpha + \beta \log X + u$$

$$dY = \beta \frac{dX}{X}$$

$$\beta = dY / \frac{dX}{X}$$

⇒ Xが1%増えるとYはβだけ増える