統計解析論特殊講義

Gretl を使った計量経済学３

中川　満

単回帰に相当する部分を実習する

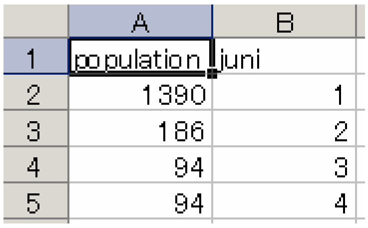
１．変数の変換

教科書例 2.10（p.57）の Ziph の法則の例を使って変数の変換を説明する．

Ziph の法則とは，変数値×変数順位　一定の法則のことである．これを回帰分析で検証していく．変数の値を x，順位を r とすると，$xr \approx const.$ なので，$\log x + \log r \approx const.$ となり，$\log x \approx const. - \log r$ となる．定数を α，誤差項（等号が完全に成立しないずれ）を u とすると，$\log x = \alpha + \beta \log r + u$ として，$\beta = -1$ となることを意味する．この回帰分析をやってみよう．なお，この場合の log は自然対数でも常用対数でも良いが，gretl の log は自然対数になっている．

① p.57 の表２－４のデータを以下のように入力．（全部で１２市）．単位は千人．

ここで第二変数になぜ rank という英語ではなく，juni というローマ字を入れたかというと，rank という変数名を使って読み込ませようとしたら，以下のメッセージを gretl が出したから．



	A	B
1	population	juni
2	1390	1
3	186	2
4	94	3
5	94	4

図１

scanning for variable names...

line: population,rank

'rank' refers to a function and may not be used as a variable name

第３行目に'rank'という名前は，rank という関数名は gretl が使うから，変数名には使ってくれるな（こういうのを予約語という）という趣旨のメッセージがある．gretl の関数名と同じ名前は事後的に避けなければならない．

② このデータを gretl に読み込ませる．（２の１．１参照）

③ logx のデータを作るために，図２のようにメニューバーの「Add」→「Logs of selected variables」を選択．

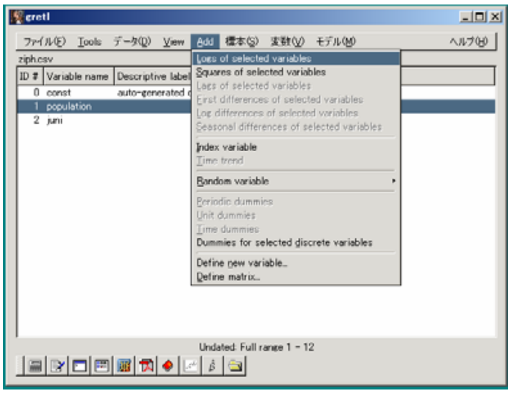


図２

④ すると，下に，population（人口）の log 値として l_population という変数が付け加えられる．

⑤ ③と同様にして，juni を選択してメニューバーの「Add」→「Logs of selected variables」を選択．

⑥ juni（順位）の log 値 l_juni ができる．

⑦ 散布図を書く．図３の様にメニューバーから「View」→「Graph specified vars」→「X-Y scatter」を選ぶ．

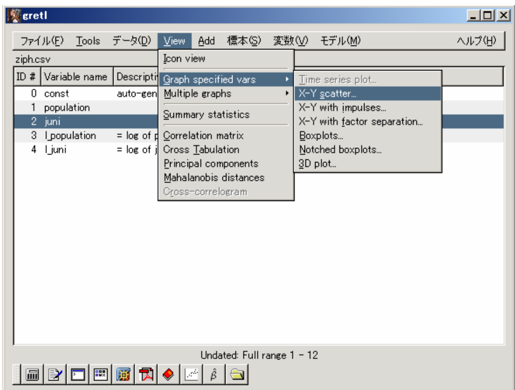


図３

⑧ 次のウィンドウに対しては，図４のように「X-axis variable（X 軸の変数）」には「l_juni」を左クリッ

クしてから「Choose→」ボタンを押して選択し，「Y-axis variable（Y 軸の変数）」には回帰式の左辺の「l_population」を左クリックして「Add→」ボタンをおす．その後「OK」

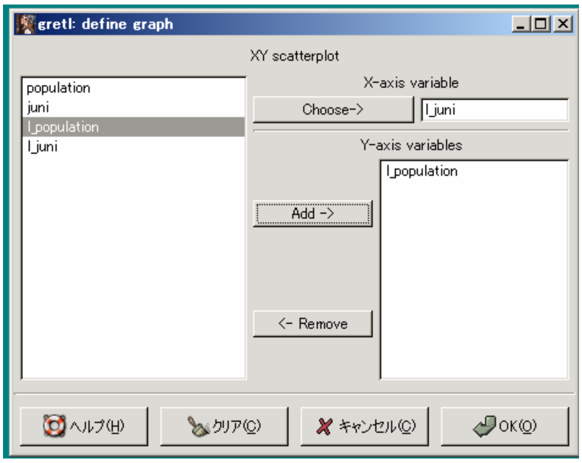


図４

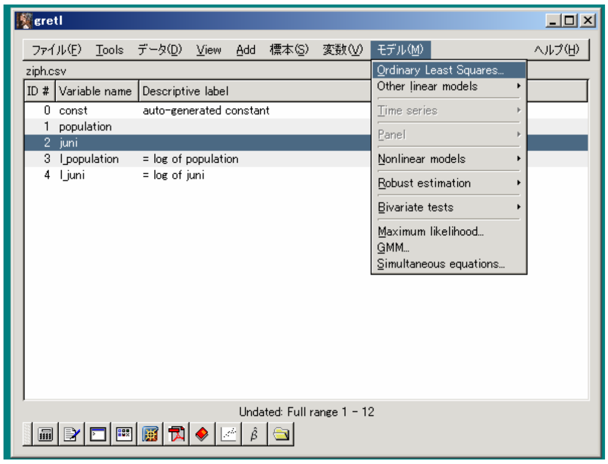


図５

ボタンを押す．散布図が得られるが，直線当てはめがある程度成功していることがみてとれる．

⑨　今度は回帰分析を行う．図５のようにメニューバーの「モデル」→「Ordinary least squares」を選択する．次に出てきたウィンドウでは図６のように「Dependent variable（従属変数）」に回帰式の左辺の「l_population（人口の log 値）」を左クリックで選択した後，「Choose→」ボタンで入力し，「Independent variables（独立変数）」には「l_juni（順位の log 値）」を選択し「Add→」ボタンを押すことで，入力する．その後「OK」ボタンを押す．

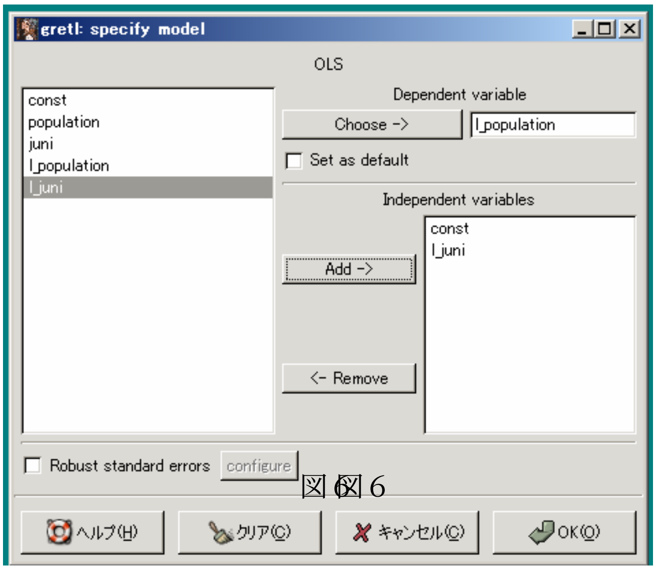


図６

⑩　プリント１の回帰分析に関するところを参考にしてこの回帰分析の結果をよんで見よう．

Model 2: OLS estimates using the 12 observations 1-12

Dependent variable: l_population

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	6.53039	0.275560	23.699	<0.00001 ***
l_juni	-1.23506	0.151739	-8.139	0.00001 ***

Mean of dependent variable = 4.47328

Standard deviation of dep. var. = 1.00148

Sum of squared residuals = 1.44692

Standard error of residuals = 0.380383

Unadjusted R-squared = 0.86885

Adjusted R-squared = 0.855735

Degrees of freedom = 10

Log-likelihood = -4.33443

(Log-likelihood for population = -58.0138)

Akaike information criterion (AIC) = 12.6689

Schwarz Bayesian criterion (BIC) = 13.6387

Hannan-Quinn criterion (HQC) = 12.3098

⑪　どのくらいこの回帰が成功しているかを，実際の値とモデルの予測値とをグラフにしてみよう．（プリント１，図７参照）

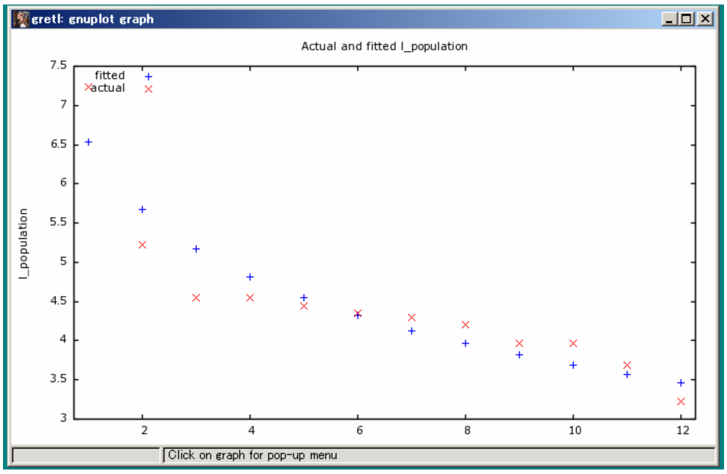


図７

２．区間推定と検定

さて，ziph の法則によれば，$\beta=-1$である．データはこれを支持するのだろうか？誤差項の存在によって，βの推定値は真の値から揺らいでいて隔たっている．だから，βが－１であっても，推定値は－１であるとは限らない．で，問題は，本当の値が－１であって，推定値が－1.23 というのはあり得ることなのだろうか？その判定は，一つは直接検定を行うのであるが，真の値が０以外の場合，gretl ではうまくいかない．間接的に区間推定を使用して判定する．

区間推定とは，ある確率でパラメータの真の値がその中に入っている（的中している）ような区間（信頼区間）を求めるものである．よく使われるのは９５%信頼区間である．

①　gretl の OLS(最小二乗法)の結果の書いてある「モデルウィンドウ」を左クリックして，メニューバーから「Analysis」→「Confidence intervals for coefficients」を選択する（図８参照）．その結果，以下のようなメッセージが書いてあるウィンドウが表示される．．

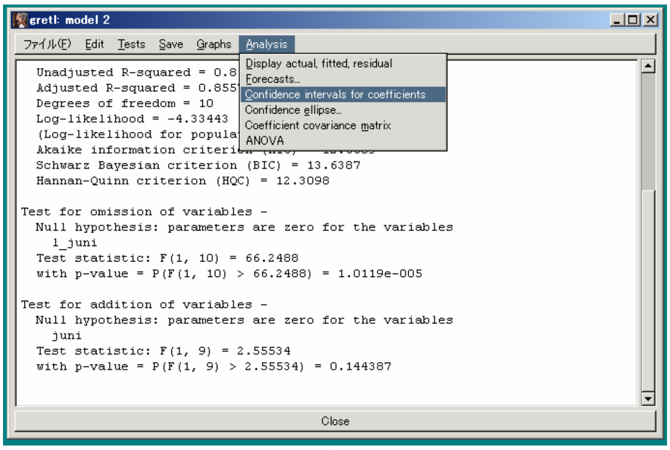


図８

t(10, .025) = 2.228

VARIABLE	COEFFICIENT	95% CONFIDENCE INTERVAL
const	6.53039	(5.91640, 7.14437)
l_juni	-1.23506	(-1.57315, -0.896959)

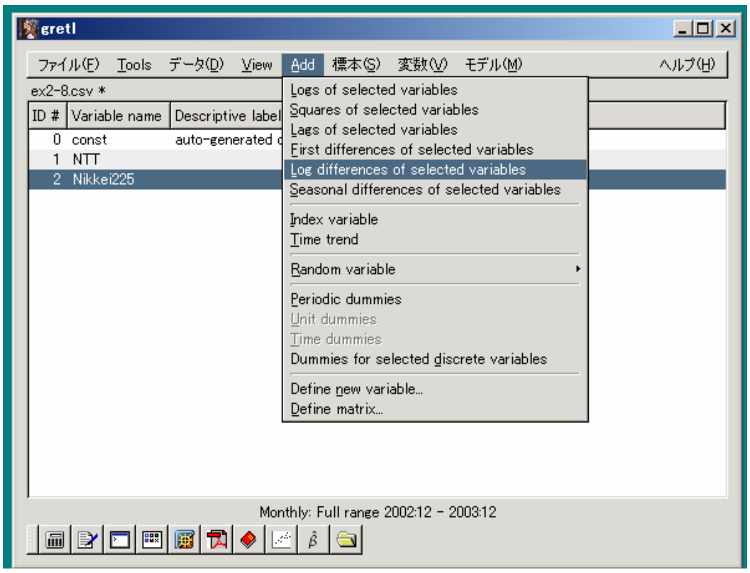
この意味は定数項の９５%信頼区間は(5.92, 7.14)，$\log r$ の係数の信頼区間は，(−1.57,−0.90)であるということである．

② これを元に検定する．$\log r$ の係数として想定される値−１はこの信頼区間に含まれているので，有意水準（１００−９５）%＝５%で係数＝−１という仮説は受容される．よって，このデータは，Ziph の法則を覆す(refute)ことはできない．

問題

１．教科書第２章練習問題８

ヒント：ex2-8.csv を私のＨＰからダウンロードし，gretl に読み込ませる．その後，Time Series データであることを gretl にわからせる．すると，メニューバーの「Add」→「Log differences of selected variables」が現れる．「データ」→「Select All」を利用して両変数を選択してから，このメニューバーの操作をすると，対数収益率，または，複利収益率が得られる．



２．上記の問題のそれぞれの回帰モデルにおける各係数の９５%信頼区間を求めよ．