第 9 回：単回帰モデルの係数の検定

北村　友宏

2020 年 7 月 3 日

本日の内容

1. 標準誤差
2. 仮説検定

消費関数の推定

いま整理・加工・分析している都道府県別・男女別データセットを用いて，ケインズ型消費関数

$$c_i = \beta_0 + \beta_1 y_i + u_i$$

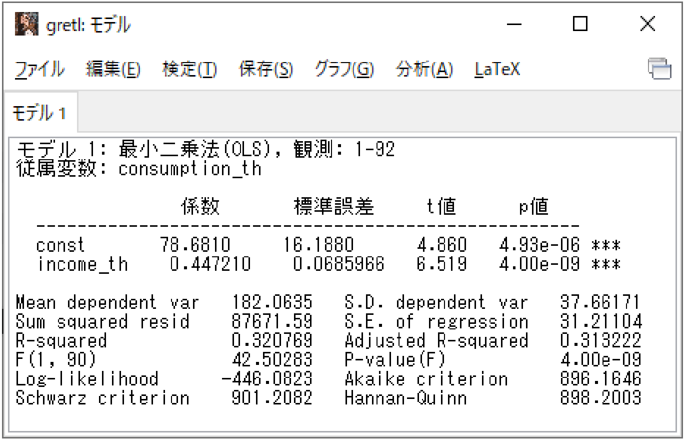
- c_i：消費支出
- y_i：可処分所得

を推定する．

実習 1

1. gretl を起動.
2. 「ファイル」→「データを開く」→「ユーザー・ファイル」と操作.
3. 消費 2009.gdt を選択し，「開く」をクリック.
4. gretl のメニューバーから「モデル」→「通常の最小二乗法」と操作.

5. 出てきたウィンドウ左側の変数リストにある consumption_th をクリックし，3 つの矢印のうち上の青い右向き矢印をクリック．
    - 推定式の左辺の変数（被説明変数，従属変数）が consumption_th（千円単位の消費支出）となる．
6. 「デフォルトとして設定」にチェック．
    - gretl を終了するまでの間，次回以降「通常の最小二乗法」での推定を行う際に，いま選択した変数が自動的に被説明変数（従属変数）に入力される．
7. ウィンドウ左側の変数リストにある income_th をクリックし，3 つの矢印のうち真ん中の緑の右向き矢印をクリック．
    - 推定式の右辺の変数（説明変数，独立変数）が income_th（千円単位の可処分所得）となる．
    - 最初から説明変数リストに入っている const は推定式の切片（定数項）のこと．
8. 「OK」をクリックすると，結果が新しいウィンドウに表示される．



このような画面が表示されれば成功.「gretl: モデル」のウィンドウはまだ閉じない！

出力結果の見方

- 係数: 回帰係数推定値
- 標準誤差: 回帰係数の標準誤差
- t 値: 「回帰係数が 0」という帰無仮説の両側 t 検定における検定統計量の実現値（t 値）
- p 値: 両側 p 値
- R-squared: 決定係数

標準誤差

- 推定量の標準偏差の推定値を標準誤差（standard error）という．
- 回帰係数の OLS 推定量 $\hat{\beta}_0$ と $\hat{\beta}_1$ の（デフォルトの）標準誤差は，それぞれ

$$\text{s.e.}\left(\hat{\beta}_0\right) = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2} \cdot \left\{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right\}},$$

$$\text{s.e.}\left(\hat{\beta}_1\right) = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2} \cdot \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

$\Rightarrow$ これらの標準誤差は，任意の i について $V(u_i \mid x_i)$ が一定（均一分散）の場合のみ正しい．

頑健標準誤差

- $V(u_i \mid x_i)$ が一定でないことを（条件付き）不均一分散（heteroskedasticity）という．
- 不均一分散があっても厳密な標準誤差を求めるために，頑健標準誤差（robust standard error）が開発されている．
- gretl では，例えば White の頑健標準誤差などを出力できる．
  - 「gretl: モデル推定」ダイアログボックスの，「頑健標準誤差」をチェックすればよい．

- 経済学分野の実証分析では，誤差項 u_i に不均一分散があることを前提として頑健標準誤差を計算する場合が多い.
- 頑健標準誤差のほうがデフォルトの標準誤差より大きくなることもあれば，小さくなることもある.

仮説検定

- $y_i = \beta_0 + \beta_1 x_i + u_i$ の y_i や x_i は様々な値をとり，観測される前はどのような値になるかが不確定（確率変数，random variable）.
- y_i や x_i を用いて計算する $\bar{y}$ や $\bar{x}$ の値も不確定.
- $y_i, \bar{y}, x_i, \bar{x}$ を用いて計算する $\hat{\beta}_0$ や $\hat{\beta}_1$ の値も不確定.

⇒ 例えば回帰係数 β_1 の推定値として $\hat{\beta}_1 = 0.44721$ という値が得られても，「推定値 $\hat{\beta}_1$ は真の β_1 の値と必ずしも同じではなく，真の β_1 は 0 で，その推定値 $\hat{\beta}_1$ は様々な値をとりうる中でたまたま 0.44721 になった」可能性もある.

⇒ 仮説検定（hypothesis testing）を行い，「真の β_0 や β_1 が 0 かどうか」を検証する.

gretl などの統計解析ソフトで線形回帰モデルを推定すると，各回帰係数 β_j（単回帰の場合 $j = 0, 1$）について，

$$H_0 : \beta_j = 0 \quad \text{vs} \quad H_1 : \beta_j \neq 0$$

を検定するのに必要な情報が出力される．

- 回帰分析では，通常は両側検定を行う．

- $\boxed{H_0\text{（係数は 0）棄却}}$
  ➡「その回帰係数は統計的に有意に 0 と異なる」と判断.
  - 「その説明変数は被説明変数と統計的に有意に相関している」と解釈.
  - 定数項の検定の場合は「定数項は統計的に有意に 0 と異なる」と解釈.
- $\boxed{H_0\text{（係数は 0）採択}}$
  ➡「その回帰係数は 0 と異なるとは言えない」と判断.
  - 「その説明変数は被説明変数と相関しているとは言えない」と解釈.
  - 定数項の検定の場合は「定数項は統計的に有意に 0 と異なるとは言えない」と解釈.

p 値による判断

- p 値が 0.1 以下（未満）: 有意水準 10%で H_0 を棄却.
- p 値が 0.05 以下（未満）: 有意水準 5%で H_0 を棄却.
- p 値が 0.01 以下（未満）: 有意水準 1%で H_0 を棄却.

gretl では，モデル推定結果の各説明変数の行の右端にアスタリスク（∗）が表示され，∗の個数を見れば，「有意水準何%で『回帰係数は 0』の H_0 を棄却できるか」が分かる．

- （アスタリスクなし）: 有意水準 10%でも「係数は 0」の H_0 採択．
- *: 有意水準 10%で，「係数は 0」の H_0 棄却．
- **: 有意水準 5%で，「係数は 0」の H_0 棄却．
- ***: 有意水準 1%で，「係数は 0」の H_0 棄却．

t 値による判断

$\beta_j = 0$ という H_0 を検定するための t 検定統計量は，

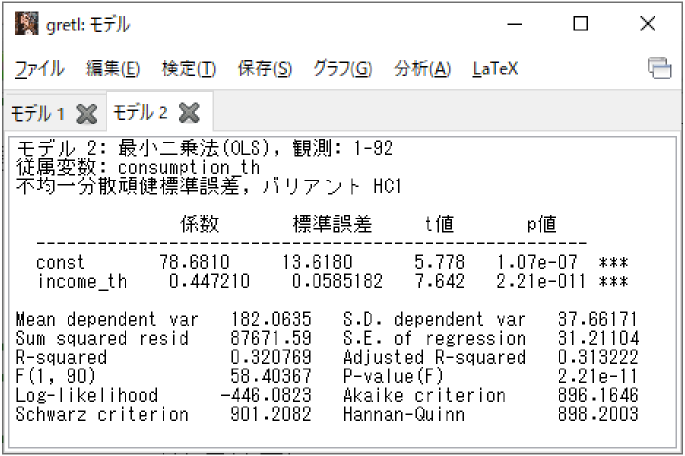
$$t = \frac{\hat{\beta}_j}{\text{s.e.}\left(\hat{\beta}_j\right)} \sim t(n-k-1).$$

- 観測値数が十分に大きいとき，t 値の絶対値がほぼ 2 を超えていれば，H_0 を棄却と判断（大雑把な判断）．
  ➥ 「有意水準何%で H_0 を棄却できるか」を厳密に判断するには，t 値ではなく p 値を見る．

実習 2

1. gretl のメニューバーから「モデル」→「通常の最小二乗法」と操作．説明変数（独立変数）は必ず前回の選択内容が記録されており，被説明変数（従属変数）は前回「デフォルトとして設定」にチェックしていれば前回の選択内容が記録されている．
2. 従属変数の入力ボックスに consumption_th が入力されていなければ，出てきたウィンドウ左側の変数リストにある consumption_th をクリックし，3 つの矢印のうち上の青い右向き矢印をクリック．
   - 推定式の左辺の変数（被説明変数，従属変数）が consumption_th（千円単位の消費支出）となる．

3. 「頑健標準誤差を使用する」にチェック．
    - 不均一分散に対して頑健な，White の標準誤差が計算され，推定式の誤差項 u_i の分散に関する仮定が誤っていても，より厳密な分析ができるようになる．
4. 「OK」をクリックすると，結果が表示される．



このような画面が表示されれば成功.「gretl: モデル」のウィンドウはまだ閉じない！

ケインズ型消費関数推定結果

- 所得の係数
  - 0.44721（符号は正）
  - 限界消費性向の推定値
  - t 値は 7.642, p 値は 2.21×10^{-11}
    ➡ 仮に「income_th の係数が 0」だとすると，
    7.642 という t 値は 2.21×10^{-11}, つまりほぼ 0%の確率（1%を下回る確率）でしか出てこない．
    ➡ 有意水準 1%で，係数ゼロの H_0 棄却．
    ➡ 所得は消費と統計的に有意に相関している．

- 定数項
  - 78.681（符号は正）
  - 基礎消費の推定値
  - t 値は 5.778, p 値は 1.07×10^{-7}
    ➡ 仮に「定数項が 0」だとすると，5.778 という t 値は 1.07×10^{-7}, つまりほぼ 0%の確率（1%を下回る確率）でしか出てこない.
    ➡ 有意水準 1%で，係数ゼロの H_0 棄却.
    ➡ 定数項は統計的に有意に 0 と異なる.

実習 3

1. 「モデル 2」が表示されている状態で，「gretl: モデル」のウィンドウのメニューバーから「ファイル」→「名前を付けて保存」と操作．
2. 「標準テキスト」を選び，「OK」をクリック．
3. 消費関数推定結果 2.txt という名前で「2020 ミクロデータ分析 1」フォルダに保存．すると，表示された推定結果をそのままテキストファイルで保存できる．本日の作業はここまで．