

第 12 回 : 回帰係数の均一性の 検定

【教科書第 8 章】

北村 友宏

2020 年 12 月 25 日

本日の内容

1. 回帰係数の均一性の検定方法
2. gretl を用いた検定の実習

グループごとに回帰係数は異なるか？

- ▶ d_i をダミー変数とする.
- ▶ n 個の個体のうち, $d_i = 0$ となる個体が n_0 個, $d_i = 1$ となる個体が n_1 個あるとする.

このとき,

- ▶ $d_i = 0$ のグループのみについて y_i を x_i に回帰するモデルは,

$$y_i = \beta_0 + \beta_X x_i + u_{0i}, \quad i = 1, 2, \dots, n_0. \quad (1)$$

- ▶ $d_i = 1$ のグループのみについて y_i を x_i に回帰するモデルは,

$$y_i = \tilde{\beta}_0 + \tilde{\beta}_X x_i + u_{1i}, \quad i = 1, 2, \dots, n_1. \quad (2)$$

▶ 一般に,

$$\beta_0 \neq \tilde{\beta}_0, \quad \beta_X \neq \tilde{\beta}_X.$$

すなわち, 回帰係数はグループごとに異なる.

(1) と (2) を 1 本の式で表すと,

$$y_i = \beta_0 + \beta_X x_i + \beta_D d_i + \beta_{DX} d_i x_i + u_i, \quad (3)$$
$$i = 1, 2, \dots, n.$$

$\tilde{\beta}_0 = \beta_0 + \beta_D$, $\tilde{\beta}_X = \beta_X + \beta_{DX}$ とすれば,

▶ $d_i = 0$ のとき

u_i を u_{0i} に書き換えると,

$$\begin{aligned} y_i &= \beta_0 + \beta_X x_i + \beta_D \cdot 0 + \beta_{DX} \cdot 0 \cdot x_i + u_{0i} \\ &= \beta_0 + \beta_X x_i + u_{0i}. \end{aligned} \tag{1}$$

▶ $d_i = 1$ のとき

u_i を u_{1i} に書き換えると,

$$\begin{aligned} y_i &= \beta_0 + \beta_X x_i + \beta_D \cdot 1 + \beta_{DX} \cdot 1 \cdot x_i + u_{1i} \\ &= \beta_0 + \beta_X x_i + \beta_D + \beta_{DX} x_i + u_{1i} \\ &= \underbrace{\beta_0 + \beta_D}_{=\tilde{\beta}_0} + \underbrace{(\beta_X + \beta_{DX})}_{=\tilde{\beta}_X} x_i + u_{1i} \\ &= \tilde{\beta}_0 + \tilde{\beta}_X x_i + u_{1i}. \end{aligned} \tag{2}$$

- ▶ $\tilde{\beta}_0 = \beta_0 + \beta_D, \tilde{\beta}_X = \beta_X + \beta_{DX}$ なので,

$$\beta_D = 0 \text{ かつ } \beta_{DX} = 0$$



$$\beta_0 = \tilde{\beta}_0 \text{ かつ } \beta_X = \tilde{\beta}_X.$$

すなわち、「 β_D と β_{DX} がともにゼロであること」と「 $d_i = 0$ のグループと $d_i = 1$ のグループで回帰係数が同じであること」は同値.

- ▶ β_D は (3) の d_i の係数, β_{DX} は (3) の $d_i x_i$ の係数.



(3) を推定し,

$$H_0 : \beta_D = 0 \text{ and } \beta_{DX} = 0 \quad \text{vs} \quad H_1 : \beta_D \neq 0 \text{ or } \beta_{DX} \neq 0$$

を検定すれば, $d_i = 0$ のグループと $d_i = 1$ のグループで回帰係数が異なるかどうかを検定できる.

- ▶ H_0 棄却 \Rightarrow 両グループで回帰係数が統計的に有意に異なる.
- ▶ H_0 採択 \Rightarrow 両グループの回帰係数には統計的に有意な差がない.

複数の係数に関する仮説検定なので, F 検定 (またはカイ二乗検定) を行う.

gretl での複数係数ゼロの検定

1. OLS 推定（メニューバーから「モデル」→「通常の最小二乗法」）を実行した後，出てきた「gretl: モデル」のウィンドウのメニューバーから「検定」→「変数を取り除く」と操作.
2. 出てきた「gretl: モデルの検定」ウィンドウ左側の変数リストにある変数のうち，係数をゼロと想定したい変数の名前をクリックし，2つの矢印のうち上の緑の右向き矢印をクリック.
3. ラジオボタンの「変数を減らしたモデルを推定する」をクリックして選ぶ.
4. 「OK」をクリック.

F 検定とカイ二乗検定

- ▶ 複数の係数に関する仮説検定は、カイ二乗検定でも可能.
- ▶ カイ二乗検定は、観測値数が十分に大きいときに行う（カイ二乗分布を近似的に当てはめる）.



- ▶ F 検定は観測値数が小さいときでも大きいときでも、（分布の仮定が正しければ）より厳密な判断が可能.
- ▶ gretl で線形回帰モデル推定後に複数係数ゼロの検定を実行すると、 F 検定が行われる.

⇒ この授業の実習では、 F 検定を行う。

中古マンションデータを用いた回帰係数の均一性の検定

まず,

$$\begin{aligned} price_i = & \beta_0 + \beta_1 minutes_i + \beta_2 age_i + \beta_3 area_i + \beta_4 d_i \\ & + \beta_5 d_i \cdot minutes_i + \beta_6 d_i \cdot age_i + \beta_7 d_i \cdot area_i + u_i \end{aligned} \quad (4)$$

- ▶ $price_i$: 中古マンション価格 (万円)
- ▶ $minutes_i$: 最寄り駅までの所要時間 (分)
- ▶ age_i : 築年数 (年)
- ▶ $area_i$: 面積 (m^2)
- ▶ d_i : ワンルームダミー
- ▶ i : 中古マンション番号

を推定する.

続いて,

$$H_0 : (\beta_4, \beta_5, \beta_6, \beta_7)' = (0, 0, 0, 0)'$$

vs $H_1 : (\beta_4, \beta_5, \beta_6, \beta_7)' \neq (0, 0, 0, 0)'$

を検定する.

- ▶ 帰無仮説 : $\beta_4, \beta_5, \beta_6, \beta_7$ が全てゼロ.
- ▶ 対立仮説 : $\beta_4, \beta_5, \beta_6, \beta_7$ のうち少なくとも1つはゼロでない.



ワンルームのグループとワンルーム以外のグループで偏回帰係数が異なるかどうかを検定できる.

gretl での変数の積

以下の方法で変数の積を作成できる。

1. gretl のメニューバーから「追加」→「新規変数の定義」と操作。
2. 出てきた「gretl: 変数の追加」ダイアログボックスの入力ボックスに

(付けたい変数名)=(変数の定義式)

を入力し、「OK」をクリック。

- ▶ 積の演算子は、* (半角アスタリスク)。

使える演算子などについては、「gretl: 変数の追加」ダイアログボックスの「ヘルプ」をクリックすれば参照できる (英語)。

実習 1

まず，ワンルームダミーと他の説明変数との積の変数を作成する．

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. setagayaapartment.gdt を選択し，「開く」をクリック．
4. gretl のメニューバーから「追加」 → 「新規変数の定義」と操作．
5. 出てきたダイアログボックスの入力ボックスに
$$\text{onekrminutes} = \text{onekr} * \text{minutes}$$
と入力し，「OK」をクリック．
 - ▶ 「onekrminutes」という変数が作成され，「onekr と minutes を掛けたもの」と定義される．

6. 出てきたダイアログボックスの入力ボックスに

$\text{onekrage} = \text{onekr} * \text{age}$

と入力し、「OK」をクリック。

- ▶ 「onekrage」という変数が作成され、「onekr と age を掛けたもの」と定義される。

7. 出てきたダイアログボックスの入力ボックスに

$\text{onekrage} = \text{onekr} * \text{area}$

と入力し、「OK」をクリック。

- ▶ 「onekrarea」という変数が作成され、「onekr と area を掛けたもの」と定義される。

8. gretl のメニューバーから「ファイル」→「データを保存」と操作し，setagayaapartment.gdt を上書き保存.
9. Ctrl キーを押しながら「id」「onekr」「onekrminutes」「onekrage」「onekrarea」の5つを左クリックして選択し，その上で右クリック→「データ（値）を表示」と操作すると，選択した5変数の観測値リストが新規ウィンドウにて表示される.

The screenshot shows a window titled "gretl: データ表示" (gretl: Data Display). The window contains a table with the following data:

	id	onekr	onekrminutes	onekrage	onekrarea
1	1	1	5	28	15
2	2	0	0	0	0
3	4	0	0	0	0
4	5	0	0	0	0
5	6	1	2	18	20
6	7	0	0	0	0
7	8	1	6	1	25
8	9	0	0	0	0
9	10	0	0	0	0
10	11	0	0	0	0
11	12	0	0	0	0
12	13	0	0	0	0
13	14	0	0	0	0
14	15	0	0	0	0
15	16	1	5	10	20
16	17	1	3	28	10
17	18	0	0	0	0
18	19	1	5	11	15
19	21	0	0	0	0
20	22	0	0	0	0
21	23	0	0	0	0
22	24	0	0	0	0
23	25	0	0	0	0
24	26	1	6	13	15
25	27	1	8	25	20
26	28	0	0	0	0

このような画面が表示されれば成功。確認したら閉じる。

実習 2

続いて、(4) を推定する。

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。
2. 出てきたウィンドウ左側の変数リストにある price_10th 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 price_10th（万円単位の中古マンション価格）となる。
3. 「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。

4. ウィンドウ左側の変数リストにある minutes をクリックした後、Ctrl キーを押しながら、area, onekr, age, onekrminutes, onekrage, onekrarea 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 minutes, area, onekr, age, 「onekr と minutes の積」, 「onekr と age の積」, 「onekr と area の積」の7つとなる。
 - ▶ 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。
5. 「頑健標準誤差を使用する」にチェック。これで、推定式の誤差項 u_i のバラつき（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。
6. 「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。

gretl: モデル

ファイル 編集(E) 検定(I) 保存(S) グラフ(G) 分析(A) LaTeX

モデル 1

モデル 1: 最小二乗法(OLS), 観測: 1-194
 従属変数: price 10th
 不均一分散頑健標準誤差, バリエーション HCl

	係数	標準誤差	t値	p値	
const	1750.40	186.105	9.405	1.97e-017	***
minutes	-40.1940	9.85799	-4.162	4.82e-05	***
area	63.7856	3.20758	19.88	6.79e-048	***
onekr	-952.286	466.350	-2.042	0.0426	**
age	-63.0556	4.39335	-14.35	6.22e-032	***
onekrminutes	45.5570	24.2528	1.878	0.0619	*
onekrage	19.1397	9.23662	2.072	0.0396	**
onekrarea	-0.378739	14.2903	-0.02650	0.9789	
Mean dependent var	3762.577	S.D. dependent var	2150.961		
Sum squared resid	94313959	S.E. of regression	712.0845		
R-squared	0.894378	Adjusted R-squared	0.890403		
F(7, 186)	315.2357	P-value(F)	1.45e-99		
Log-likelihood	-1545.419	Akaike criterion	3106.839		
Schwarz criterion	3132.982	Hannan-Quinn	3117.425		

このような画面が表示されれば成功。まだ作業があるので、「gretl: モデル」のウィンドウは**まだ閉じない!**

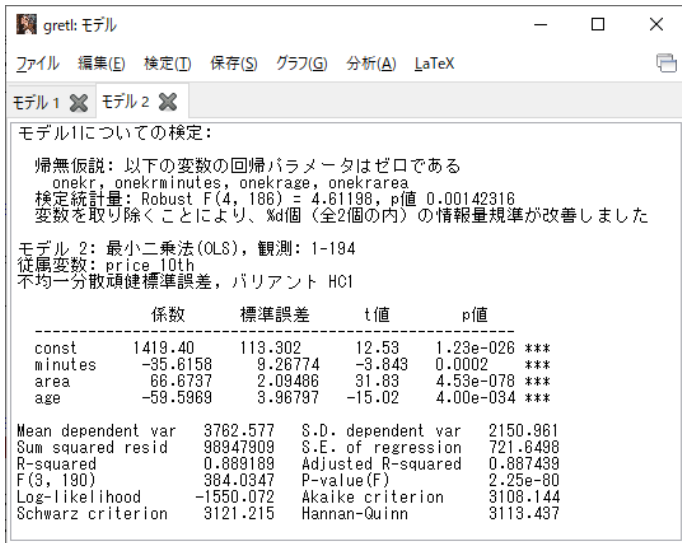
7. 出力された「gretl: モデル」のウィンドウのメニューバーから「ファイル」→「名前を付けて保存」と操作.
8. 「標準テキスト」を選び、「OK」をクリック.
9. results20201225_1.txt という名前で 2020microdatag フォルダに保存. すると, 表示された推定結果をそのままテキストファイルで保存できる.

実習 3

次に、「(4) の $\beta_4, \beta_5, \beta_6, \beta_7$ が全てゼロ」を検定する。

1. 「gretl: モデル」のウィンドウのメニューバーから「検定」→「変数を取り除く」と操作。
2. 出てきたウィンドウ左側の変数リストにある `onekr` をクリックした後、Ctrl キーを押しながら、`onekrminutes`, `onekrage`, `onekrarea` をクリックし、2つの矢印のうち上の緑の右向き矢印をクリック。
 - ▶ 「`onekr` の係数、『`onekr` と `minutes` の積』の係数、『`onekr` と `age` の積』の係数、『`onekr` と `area` の積』の係数が全てゼロ」が検定される。

3. ラジオボタンの「変数を減らしたモデルを推定する」が選ばれていなければ、「変数を減らしたモデルを推定する」をクリック。
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。まだ作業があるので、「gretl: モデル」のウィンドウは**まだ閉じない!**

5. 「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.
6. 「標準テキスト」を選び、「OK」をクリック.
7. results20201225_2.txt という名前で 2020microdatag 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.

回帰係数の均一性の検定結果

▶ 頑健 F 統計量

- ▶ 4.61198.
- ▶ Robust $F(4, 186)$ のカッコ内は F 分布の自由度.
- ▶ p 値は 0.00142316.
- ▶ 有意水準 1% で, 4 個の係数がゼロという H_0 棄却.
↳ ワンルームのグループとそれ以外のグループで, モデルの偏回帰係数が統計的に有意に異なる.

本日の作業はここまで.