第 9 回：回帰分析の応用
【教科書第 7 章第 1 節～第 3 節】

北村　友宏

2025 年 11 月 25 日

本日の内容

1. 二乗項を含む回帰
2. ダミー変数を含む回帰
3. 回帰係数の均一性の検定

二乗項を含む回帰

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i, \\
E(u_i \mid x_i) &= 0, \\
E(u_i u_j \mid x_i) &= 0 \quad (i \neq j), \\
V(u_i \mid x_i) &= \sigma^2, \\
i &= 1, 2, \cdots, n
\end{aligned}
$$

を推定することを考える．

- 二乗項が入っているので厳密には非線形回帰モデルであるが，x_i^2 は「二乗」まで含めて 1 つの変数と考えれば，線形回帰モデルと同様の手法で推定できる．

微分による回帰係数の解釈

- x_i が 1 単位変化すると y_i が何単位変化するかを表すものを限界効果（marginal effect）という.

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$$

を x_i で微分すると,

$$\frac{dy_i}{dx_i} = \beta_1 + 2\beta_2 x_i.$$

これが限界効果.

$\Rightarrow$ この場合，限界効果は x_i の値に依存して変化する.

- 二乗項のないモデル $y_i = \beta_0 + \beta_1 x_i + u_i$ なら，限界効果は $\dfrac{dy_i}{dx_i} = \beta_1$ で，x_i にかかわらず一定.

β_1 も β_2 も統計的に有意に 0 と異なれば，以下の解釈ができる（$x_i \geq 0$ の場合）.

- $\beta_1 > 0, \beta_2 > 0$
  - 限界効果は正で，x_i が大きくなるにつれて限界効果が大きくなる.
- $\beta_1 > 0, \beta_2 < 0$
  - x_i が小さいときは限界効果は正であるが，x_i が大きくなるにつれて限界効果が小さくなり，やがて限界効果は負に変わる.
- $\beta_1 < 0, \beta_2 > 0$
  - x_i が小さいときは限界効果は負であるが，x_i が大きくなるにつれて限界効果が大きくなり，やがて限界効果は正に変わる.
- $\beta_1 < 0, \beta_2 < 0$
  - 限界効果は負で，x_i の値が大きくなるにつれて限界効果が（負の方向に）小さくなる.

「$\beta_1 > 0, \beta_2 < 0$」の場合と「$\beta_1 < 0, \beta_2 > 0$」の場合において，限界効果の符号が変わる点での説明変数 x_i の値は，

$$\frac{dy_i}{dx_i} = 0 \Leftrightarrow \beta_1 + 2\beta_2 x_i = 0 \Leftrightarrow x_i = -\frac{\beta_1}{2\beta_2}.$$

$\Rightarrow$ 一次項の偏回帰係数 β_1 と二乗項の偏回帰係数 β_2 の OLS 推定値 $\hat{\beta}_1, \hat{\beta}_2$ を求め，

$$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

を計算すれば，限界効果の符号が変わる点での説明変数の値が求まる．

gretl での変数の二乗の作成方法

gretl の画面で，二乗を作成したい変数をクリックして選択し，メニューバーから「追加」→「選択された変数の二乗」と操作すると，選択した変数の二乗が作成される．

- 二乗の変数名は，「sq_（元の変数名）」となる．

二乗項を含むミンサー方程式の推定

年齢によって 1 年働くことによる年収の増え方が異なることをコントロールしたうえで，「修学年数が増えると，年収がどれだけ増えるのか」を分析するためのモデル（ミンサー方程式）

$$\ln income_i = \beta_0 + \beta_Y yeduc_i + \beta_E exper_i + \beta_{EE} exper_i^2 + u_i$$

- $income_i$: 年収（万円）
- $yeduc_i$: 修学年数（年）
- $exper_i$: 就業可能年数（年）
- i : 個人番号

を推定する．

➥ 「年収の対数値」を「修学年数」と「就業可能年数」と「就業可能年数の 2 乗」に回帰する．

- 就業可能年数
  - 最後の学校を卒業してからの年数

$$\text{就業可能年数} = \text{年齢} - \text{修学年数} - 6.$$

※ 小学校に入学する年齢が 6 歳のため，6 を引いている．

  - 熟練度を表す．
    ⇒ 賃金に影響を与える．

$$\Downarrow$$

修学年数が年収に与える純粋な効果を計測するには，熟練度（を表す就業可能年数) をコントロールする必要がある．

モデル推定結果


```
モデル 1: 最小二乗法(OLS), 観測: 1-4299
従属変数: lincome

              係数          標準誤差       t値         p値
  ---------------------------------------------------------
  const       2.48550       0.110782      22.44    1.64e-105 ***
  yeduc       0.117547      0.00706026    16.65    2.31e-060 ***
  exper       0.196174      0.00749354    26.18    2.75e-140 ***
  exper2     -0.00638115    0.000316188  -20.18    1.32e-086 ***

Mean dependent var   5.290452   S.D. dependent var   0.895883
Sum squared resid    2736.906   回帰の標準誤差             0.798267
R-squared            0.206603   Adjusted R-squared   0.206049
F(3, 4295)           372.8097   P-value(F)           3.4e-215
Log-likelihood      -5129.400   Akaike criterion     10266.80
Schwarz criterion    10292.27   Hannan-Quinn         10275.79
```

二乗項を含むミンサー方程式推定結果

- 就業可能年数の一次項の係数
  - 0.196174（符号は正）
  - 有意水準 1%で，「係数は 0」の H_0 が棄却される（5%や 10%でも棄却される）．
    ➥ 就業可能年数の一乗は年収と統計的に有意に相関している．
- 就業可能年数の二乗項の係数
  - −0.00638115（符号は負）
  - 有意水準 1%で，「係数は 0」の H_0 が棄却される（5%や 10%でも棄却される）．
    ➥ 就業可能年数の二乗も年収と統計的に有意に相関している．

就業可能年数の一次項の係数と二乗項の係数がどちらも統計的に有意に 0 と異なり,

$$-\frac{\hat{\beta}_E}{2\hat{\beta}_{EE}} \approx 15.4$$

なので,もう 1 つの説明変数の修学年数を一定としたうえで,以下のような解釈ができる.

- 就業可能年数が 15.4 年に達するまでは,就業可能年数が長くなると年収も高くなる.また,就業可能年数が長くなるにつれ,年収の増え方は次第に小さくなる.
- 就業可能年数が 15.4 年に達した後は,就業可能年数が長くなると年収が低下する.また,就業可能年数が長くなるにつれ,年収の減り方は次第に大きくなる.

ダミー変数

- ある事柄が当てはまるなら 1, 当てはまらないなら 0 とする変数をダミー変数（dummy variable）という.
  - e.g., 男性である・ない，女性である・ない，働いている・いない，自家用車を持っている・いない，朝食を食べている・いない，など.

ダミー変数を含む回帰

線形回帰モデル

$$
\begin{aligned}
y_i &= \beta_0 + \beta_X x_i + \beta_D d_i + u_i, \\
E(u_i \mid x_i, d_i) &= 0, \\
E(u_i u_j \mid x_i, d_i) &= 0 \quad (i \neq j), \\
V(u_i \mid x_i, d_i) &= \sigma^2, \\
i &= 1, 2, \cdots, n
\end{aligned}
$$

を推定することを考える．

- d_i はダミー変数（0 と 1 の値のみをとる）．
  - e.g., 男性ダミー（男性＝ 1，女性＝ 0）
  - e.g., 女性ダミー（女性＝ 1，男性＝ 0）

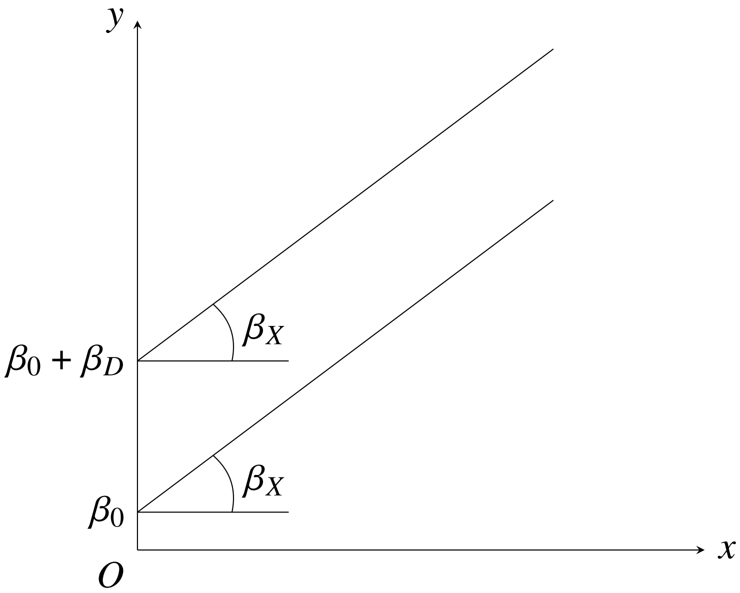
- $d_i = 0$ のとき

$$y_i = \beta_0 + \beta_X x_i + u_i = \underbrace{\beta_0}_{\text{切片}} + \beta_X x_i + u_i.$$

- $d_i = 1$ のとき

$$\begin{aligned} y_i &= \beta_0 + \beta_X x_i + \beta_D + u_i \\ &= \underbrace{(\beta_0 + \beta_D)}_{\text{切片}} + \beta_X x_i + u_i. \end{aligned}$$

⇒ ダミー変数の値が 0 か 1 かによって，縦軸切片が変化する．
⇒ ダミー変数の偏回帰係数 β_D の OLS 推定値 $\hat{\beta}_D$ を求めれば，$d_i = 1$ の場合は $d_i = 0$ の場合と比べて y_i がどの程度異なるかが分かる．


y
β_X
$\beta_0 + \beta_D$
β_X
β_0
x
O

説明変数にダミー変数を含む場合の注意

- すべての個体について，ダミー変数の値の合計が 1 になるような複数のダミー変数を作成した場合は，そのうち 1 つを除外して説明変数に用いる．
  - e.g.,「男性ダミー＋女性ダミー＝ 1」➡ 男性ダミーか女性ダミーどちらか 1 つを説明変数に用いる．

$$\Downarrow$$

- 除外したダミー変数が表すものを基準として，ダミー変数の（偏）回帰係数は基準と比較してどの程度，被説明変数に対する影響度合いが異なるか，という解釈．
  - e.g., 男性ダミーを除外して女性ダミーを説明変数に用いた場合，女性は男性と比べて被説明変数がどの程度異なるかが分かる．

ダミー変数の合計

個人番号	男性	女性	男性 + 女性
1	1	0	1
2	1	0	1
⋮	⋮	⋮	⋮
45	1	0	1
46	0	1	1
47	0	1	1
⋮	⋮	⋮	⋮
63	0	1	1

$\Downarrow$

「男性」というダミー変数と「女性」というダミー変数が完全に相関する.

- 除外するダミー変数（基準）を変更しても，残る全てのダミー変数を説明変数として用いる限り，ダミー以外の説明変数の偏回帰係数の推定値，標準誤差，t 値，p 値は変わらない．
- gretl では，すべての個体についてダミー変数の値の合計が 1 になるようなダミー変数を全て説明変数に選んだ場合，ダミー変数のうち 1 つが自動的に除外されて結果が表示される．

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，

$$y_i = \beta_0 + \beta_X x_i + u_{0i}, \quad i = 1, 2, \cdots, n_0. \tag{1}$$

- $d_i = 1$ 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，

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

- 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，

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

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

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

$$\Updownarrow$$

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

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.

$$\Downarrow$$

(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 $\Longrightarrow$ 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 $\Longrightarrow$ 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．

※ これと同値の検定を，教科書に掲載されているチョウ検定（Chow 検定）でも実行できる．

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）.

$$\Updownarrow$$

- 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線形制約の検定を実行すると，F 検定が行われる.

$\Rightarrow$ この授業では，F 検定の例で説明する.

ミンサー方程式の回帰係数の均一性の検定

「修学年数が増えると，年収がどれだけ増えるのか」が男女間で異なるかどうかを試すモデル

$$\ln income_i = \beta_0 + \beta_Y yeduc_i + \beta_F female_i + \beta_{FY} female_i \cdot yeduc_i + u_i$$

- $income_i$：年収（万円）
- $yeduc_i$：修学年数（年）
- $female_i$：女性ダミー（女性＝ 1, 男性＝ 0）
- i：個人番号

を推定する．

続いて,

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

を検定する.

- 帰無仮説：β_F, β_{FY} がどちらもゼロ.
- 対立仮説：β_F, β_{FY} のうち少なくとも 1 つはゼロでない.

$$\Downarrow$$

女性と男性で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）.

ミンサー方程式の推定結果

gretl: モデル1

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

```
モデル 1: 最小二乗法(OLS), 観測: 1-4286
従属変数: lincome

                  係数         標準誤差       t値        p値
  -----------------------------------------------------------------
  const          5.34690      0.120920      44.22     0.0000     ***
  yeduc          0.0240947    0.00853344     2.824    0.0048     ***
  female        -2.07920      0.192386     -10.81     7.03e-027  ***
  female_yeduc   0.0902285    0.0137996      6.538    6.94e-011  ***

Mean dependent var   5.260657   S.D. dependent var   0.936133
Sum squared resid    2899.053   回帰の標準誤差              0.822820
R-squared            0.227977   Adjusted R-squared   0.227436
F(3, 4282)           421.4897   P-value(F)           6.5e-240
Log-likelihood      -5243.723   Akaike criterion     10495.45
Schwarz criterion    10520.90   Hannan-Quinn         10504.44
```

回帰係数の均一性の検定結果

gretl: モデル2

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

```
モデル1についての検定:

  帰無仮説: 以下の変数の回帰パラメータはゼロである
    female, female_yeduc
  検定統計量: F(2, 4282) = 565.448, p値 1.20894e-218
  変数を取り除くことにより、%d個 (全0個の内) の情報量規準が改善しました

モデル 2: 最小二乗法(OLS), 観測: 1-4286
従属変数: lincome

                 係数          標準誤差       t値          p値
  ---------------------------------------------------------------
  const        4.19033       0.105054      39.89    3.76e-296 ***
  yeduc        0.0772308     0.00751143    10.28    1.64e-024 ***

Mean dependent var    5.260657   S.D. dependent var    0.936133
Sum squared resid     3664.707   回帰の標準誤差               0.924900
R-squared             0.024082   Adjusted R-squared    0.023855
F(1, 4284)            105.7149   P-value(F)            1.64e-24
Log-likelihood       -5745.965   Akaike criterion      11495.93
Schwarz criterion     11508.66   Hannan-Quinn          11500.43
```

- 検定統計量の実現値は 565.488, p 値は 1.20894×10^{-218}.
  ➡ 仮に「female と female×yeduc の係数がどちらも 0」だとすると，565.488 という検定統計値は 1.20894×10^{-218}, 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，「female と female×yeduc の係数がどちらも 0」の H_0 が棄却される（5%や 10%でも棄却される）.
  ➡ female と female×yeduc の係数のうち少なくとも一方は 0 でないと判断される.
  ➡ 女性と男性で，モデルの回帰係数が統計的に有意に異なる.
  ➡ 「修学年数が増えると，年収がどれだけ増えるのか」が男女間で統計的に有意に異なる.

今日のキーワード

限界効果，ダミー変数

次回までの準備

- 今回の講義スライドを読み直す.
- 「提出課題 6」に取り組む.
- 教科書第 7 章第 4 節〜第 5 節を読む.