第 6 章: 重回帰分析：発展

Jeffrey Wooldridge (2016).
Introductory Econometrics: A Modern Approach
Seventh Edition. Cengage Learning.

2026-03-06

準備

必要なパッケージの読み込み

- wooldridge パッケージの読み込み

```
library(wooldridge)
```

6-1 データのスケーリングと OLS 統計量

単位の変更の影響

- 従属変数 y や独立変数 x の単位を変更（スケーリング）した場合、OLS 推定量や統計量はどのように変化するか？
- **例：** y を円から千円にする（1000 で割る）、x をマイルからキロメートルにするなど。

従属変数 y のスケーリング

- y を定数 c 倍（例えば $y^* = cy$）した場合：
  - 係数 $\hat{\beta}_j$ はすべて c 倍になる。
  - 標準誤差 $\mathrm{se}(\hat{\beta}_j)$ も c 倍になる。
  - **t 統計量は変化しない。**
  - **決定係数 R^2 は変化しない。**

独立変数 x_j のスケーリング

- 独立変数 x_k を定数 c 倍（例えば $x_k^* = cx_k$）した場合：
  - 該当する係数 $\hat{\beta}_k$ は $1/c$ 倍になる。
  - 他の係数 $\hat{\beta}_j$ $(j \neq k)$ は変化しない。
  - 標準誤差も同様に変化するため、**t 統計量は変化しない。**
  - **決定係数 R^2 は変化しない。**

標準化係数 (Beta Coefficients)

- 変数の単位に依存しない係数比較を行いたい場合、変数を**標準化**（平均を引いて標準偏差で割る）する。
- すべての変数を標準化して回帰を行うと、得られる係数は**標準化係数**（Beta Coefficients）と呼ばれる。
- 「x が 1 標準偏差変化したとき、y が何標準偏差変化するか」を表す。

6-2 関数形式のさらなる検討

対数形式 (Log Forms)

- 対数形式をとることで、非線形な関係を線形モデルで扱える。
- 対数の差は近似的に変化率（パーセント）を表す。

モデル	従属変数	独立変数	解釈 (β_1)
Level-Level	y	x	$\Delta y = \beta_1 \Delta x$
Level-Log	y	$\log(x)$	$\Delta y \approx (\beta_1/100)\%\Delta x$
Log-Level	$\log(y)$	x	$\%\Delta y \approx (100\beta_1)\Delta x$
Log-Log	$\log(y)$	$\log(x)$	$\%\Delta y \approx \beta_1 \%\Delta x$ (弾力性)

対数形式の利点

1. 係数が変化率（弾力性）として解釈しやすくなる。
2. 変数の分布が右に裾を引いている場合（所得など)、対数をとることで正規分布に近づき、外れ値の影響を緩和できる。
3. 不均一分散を緩和することが多い。

2 次形式 (Quadratic Forms)

- 限界効果が一定でない場合に使用。
- モデル: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$
- 限界効果: $\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x$
- $\beta_2 > 0$: 下に凸（U 字型）
- $\beta_2 < 0$: 上に凸（逆 U 字型）
- 頂点（極値）：$x^* = -\beta_1/(2\beta_2)$

2 次形式の例 (hprice1)

```
data(hprice1)
# 住宅面積 (sqrft) の 2 次形式
res <- lm(price ~ lotsize + sqrft + I(sqrft^2) + bdrms, data=hprice1)
summary(res)$coefficients[3:4, ]
```

```
##                  Estimate   Std. Error    t value   Pr(>|t|)
## sqrft        -8.622611e-03 6.953558e-02 -0.1240029 0.90161283
## I(sqrft^2)    2.741764e-05 1.425207e-05  1.9237662 0.05781174
```

交差項 (Interaction Terms)

- ある独立変数の効果が、別の独立変数の値に依存する場合。
- モデル: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2) + u$
- x_1 の限界効果: $\frac{\partial y}{\partial x_1} = \beta_1 + \beta_3 x_2$
- x_2 の値によって x_1 の効果が変化することを表現できる。

6-3 決定係数 R^2 と自由度修正決定係数 $\bar{R}^2$

通常の R^2 の問題点

- モデルに独立変数を追加すると、たとえその変数に説明力がなくても、R^2 は決して減少せず、通常は増加する。
- モデルの複雑さを考慮した指標が必要。

自由度修正決定係数 (Adjusted R-squared)

- 推定に要した自由度を考慮した決定係数：

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\hat{\sigma}^2}{SST/(n-1)}$$

- 変数を追加しても、$\hat{\sigma}^2$（誤差分散の推定値）が減少しなければ $\bar{R}^2$ は増加しない。
- $\bar{R}^2$ は負になることもある。

モデル選択における注意

- R^2 や $\bar{R}^2$ を最大化することだけが、モデル選択の目的ではない。
- 重要なのは、理論的背景や、内生性（$E[u|x]=0$）の排除である。
- 非線形な変換（例：y vs $\log(y)$）をしたモデル同士の R^2 を直接比較してはいけない。

6-4 予測と残差分析

予測 (Prediction)

- 推定されたモデルを用いて、与えられた x の値に対する y の期待値を推定する。
- `predict()` 関数を使用して計算可能。
- 信頼区間 (Confidence Interval) と予測区間 (Prediction Interval) の違いに注意。

残差分析 (Residual Analysis)

- 推定された残差 $\hat{u} = y - \hat{y}$ を観察することで、モデルの妥当性をチェックする。
- 特定のサンプル（アウトライヤー）が結果に大きな影響を与えていないか確認する。

まとめ

- スケーリングは係数の値を変化させるが、t 統計量や R^2 には影響しない。
- 対数、2 次形式、交差項により、現実的な非線形関係をモデル化できる。
- $\bar{R}^2$ はモデルの複雑さを考慮するが、過信は禁物。
- 予測と残差分析は、モデルの実用性と妥当性の検証に不可欠。