

第 7 章: 重回歸分析：質的情報

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)
```

7-1 ダミー変数の基礎

質的情報 (Qualitative Information)

- ▶ これまでは、所得、教育年数、価格などの**量的変数**を扱ってきた。
- ▶ 現実には、性別、人種、業界、地域、政策の有無などの**質的情報**も重要である。
- ▶ これを 0 と 1 で表したものを**ダミー変数** (または二値変数, Binary variables) と呼ぶ。

ダミー変数の定義

- ▶ 例: 女性であれば1、男性であれば0をとる変数 *female*
- ▶ モデル: $wage = \beta_0 + \delta_0 female + \beta_1 educ + u$
- ▶ 男性の期待値 ($female = 0$):
 $E(wage | female = 0, educ) = \beta_0 + \beta_1 educ$
- ▶ 女性の期待値 ($female = 1$):
 $E(wage | female = 1, educ) = (\beta_0 + \delta_0) + \beta_1 educ$
- ▶ δ_0 は、教育年数が同じ男女間の賃金格差を表す。

ダミー変数の罠 (Dummy Variable Trap)

- ▶ カテゴリが G 個ある場合、含めるダミー変数は $G - 1$ 個にする必要がある。
- ▶ すべてのカテゴリのダミー変数と定数項を同時に入れると、完全な多重共線性が発生する。
- ▶ 入れなかったカテゴリが**基準グループ** (Benchmark/Reference group) となる。

7-2 複数のカテゴリを持つダミー変数

例: 学歴レベル

- ▶ 「中卒」「高卒」「大卒」の3カテゴリがある場合。
- ▶ ダミー変数: `highschool`, `university` の2つを作成。
- ▶ 基準グループ: 「中卒」
- ▶ モデル: $y = \beta_0 + \delta_1 \text{highschool} + \delta_2 \text{university} + \dots + u$
- ▶ δ_2 は「中卒」と比較した「大卒」の効果。
- ▶ 「高卒」と「大卒」の差を検定するには、 $\delta_2 - \delta_1 = 0$ かどうかを調べる。

7-3 交差項とダミー変数

ダミー変数同士の交差項

- ▶ 例: 性別 (female) と 既婚 (married)
- ▶ 4つのグループ (独身男性、独身女性、既婚男性、既婚女性) の差を表現できる。
- ▶ `female * married` という交差項を入れることで、結婚の効果が男女で異なるかどうかを検証できる。

量的変数とダミー変数の交差項

▶ モデル:

$$\log(\text{wage}) = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \delta_1 (\text{female} \cdot \text{educ}) + u$$

▶ 男性の教育効果: β_1

▶ 女性の教育効果: $\beta_1 + \delta_1$

▶ $\delta_1 \neq 0$ であれば、教育の収益率に男女差があることになる。

7-4 線形確率モデル (LPM)

従属変数がダミー変数の場合

- ▶ 例: 就業の有無、住宅ローンの承認、政策の効果。
- ▶ 従属変数 y が 0 または 1 をとるモデル。
- ▶ $P(y = 1|x) = E(y|x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$
- ▶ 係数 β_j は、 x_j が 1 単位変化したときの $y = 1$ となる確率の変化を表す。

LPM の利点と欠点

- ▶ **利点:** 解釈が非常に容易である。
- ▶ **欠点:**
 1. 予測値 \hat{y} が 0 未満や 1 を超えることがある。
 2. 誤差項に不均一分散が必ず発生する。
 3. 限界効果が常に一定という仮定が不自然な場合がある。
- ▶ (より高度な手法として Logit や Probit モデルがある)

LPM の推定例 (mroz)

```
data(mroz)
# 既婚女性の労働供給 (inlf=1: 労働市場に参加)
res_lpm <- lm(inlf ~ nwifeinc + educ + exper + age + kidslt6, data=mroz)
summary(res_lpm)$coefficients[1:4, ]
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	0.769832596	0.135006985	5.702169	1.703869e-08
## nwifeinc	-0.003258636	0.001455510	-2.238827	2.546088e-02
## educ	0.039128569	0.007363805	5.313635	1.419779e-07
## exper	0.022210504	0.002144369	10.357595	1.394203e-23

7-5 政策評価とプログラム評価

処置効果 (Treatment Effect)

- ▶ 政策やプログラム（処置）に参加したことの効果を測定する。
- ▶ $y = \beta_0 + \delta_0 \text{prog} + \dots + u$
- ▶ ここで prog は参加した人に対して 1 をとるダミー変数。
- ▶ 理想的には、参加がランダムに割り当てられていれば、 δ_0 は因果関係を表す。

まとめ

- ▶ ダミー変数は質的情報を回帰モデルに取り込むための強力なツール。
- ▶ 解釈の際は常に「基準グループ」を意識する。
- ▶ 交差項を用いることで、グループ間の傾きの違いを表現できる。
- ▶ 従属変数がダミー変数の場合は線形確率モデル (LPM) として扱うことができる。