

重回帰分析

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

5. 定性的変数

Ch.7 重回帰分析: 定性的変数

1. 定性的情報の表現
2. 単一ダミーの説明変数
3. 複数カテゴリーのダミー変数
4. ダミー変数の交差項
5. 2値従属変数
6. 政策分析とプログラム評価
7. 離散従属変数による回帰結果の解釈*

7-1&2 定性的情報とダミー変数

◆ 定性的情報

- 例: 性別、人種、産業、地域、格付け...etc.
- ◆ 定性的情報を組込む→ダミー変数を使用
- ◆ これらを説明変数または被説明変数として表示

◆ 単一ダミー説明変数

$$wage = \beta_0 + \delta_0 female + \beta_1 educ + u$$

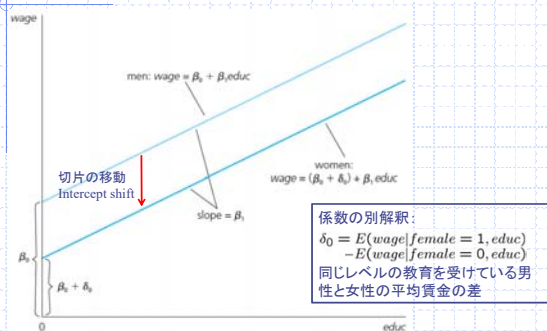
- ◆ female: ダミー変数
 - 女性のとき female=1
 - 男性のとき female=0
- ◆ δ_0 : (他の要因を固定して) 女性賃金の男性との差

7-1 定性的情報の表現

TABLE 7.1
A Partial Listing of the Data in WAGE1.RAW

person	wage	educ	exper	female	married
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
...	...	...	...	...	...
525	11.56	16	5	0	1
526	3.50	14	5	1	0

7-2 単一ダミーの説明変数



7-2 単一ダミーの説明変数

◆ ダミー変数の罠 Dummy variable trap

$$wage = \beta_0 + \gamma_0 male + \delta_0 female + \beta_1 educ + u$$

- ◆ 完全共線性のためモデル推定不可能
- ダミー変数使用の際、1グループは必ず除外
 - ◆ $wage = \beta_0 + \delta_0 female + \beta_1 educ + u$ ← 基本グループは男性
 - ◆ $wage = \beta_0 + \gamma_0 male + \beta_1 educ + u$ ← 基本グループは女性
- 他の方: 切片の除外
 - ◆ $wage = \gamma_0 male + \delta_0 female + \beta_1 educ + u$
 - 係数間の差の検定が困難
 - 切片なしでは R^2 は非有効

7-2 単一ダミーの説明変数

◆例7.1: 賃金方程式

$$\widehat{wage} = -1.57 \begin{pmatrix} .72 \end{pmatrix} - 1.81 \begin{pmatrix} .26 \end{pmatrix} female + .572 \begin{pmatrix} .049 \end{pmatrix} educ \\ + .025 \begin{pmatrix} .012 \end{pmatrix} exper + .141 \begin{pmatrix} .021 \end{pmatrix} tenure \quad n = 526, R^2 = .364$$

- 教育、職歴、勤続が一定のもとで、女性の賃金は男性より\$1.81少ない
- この結果は女性への差別を意味するのか？
 - 性別は制御されていない他の生産性特性と相関する可能性があり、必ずしも女性への差別を意味しない

7-2 単一ダミーの説明変数

◆ダミー変数による各集団平均の比較

$$\widehat{wage} = 7.10 \begin{pmatrix} .21 \end{pmatrix} - 2.51 \begin{pmatrix} .26 \end{pmatrix} female \quad n = 526, R^2 = .116$$

- 他の要因一定でなければ、女性は男性より\$2.51低い
 - 男性の平均賃金と女性の平均賃金の差は\$2.51

■考察

- 平均値の差の有意性を容易に検定
- 他の要因が一定でなければ、男女間の賃金格差はさらに大きい
 - 賃金格差の一部は、男性と女性の間に存在する教育、職歴、勤続の違い、等々による

7-2 単一ダミーの説明変数

◆例7.3: 訓練補助金が訓練時間に及ぼす影響

$$\widehat{hrsemp} = 46.67 \begin{pmatrix} 43.41 \end{pmatrix} + 26.25 \begin{pmatrix} 5.59 \end{pmatrix} \text{grant} - 0.98 \begin{pmatrix} 3.54 \end{pmatrix} \log(sales) \\ - 6.07 \begin{pmatrix} 3.88 \end{pmatrix} \log(employ), \quad n = 105, R^2 = .237$$

助成金受給企業のダミー変数

■プログラム評価の一例

- 治療群Treatment group (=助成金受給企業)
- 対照群Control group (=助成金なしの企業)
- 関心の結果に対するプログラムの効果は因果的か？

7-2 単一ダミーの説明変数

◆7-2a log(y)とダミー説明変数

$$\widehat{\log(price)} = -1.35 \begin{pmatrix} .65 \end{pmatrix} + .168 \begin{pmatrix} .038 \end{pmatrix} \log(lotsize) + .707 \begin{pmatrix} .093 \end{pmatrix} \log(sqft) \\ + .027 \begin{pmatrix} .029 \end{pmatrix} bdrms + .054 \begin{pmatrix} .045 \end{pmatrix} \text{colonial} \quad n = 88, R^2 = .649$$

- コロニアルスタイルの家についてのダミー変数

■部分効果と解釈

$$\Rightarrow \frac{\Delta \log(price)}{\Delta \text{colonial}} = \frac{\% \Delta price}{\Delta \text{colonial}} = 5.4\%$$

- コロニアルスタイルのダミー変数は0→1と変化
- 住宅価格は5.4%ポイント上昇

7-3 複数カテゴリーのダミー変数

◆複数カテゴリーのダミー変数

- ダミー変数による各カテゴリーを定義
- 1カテゴリーを基本グループとして除外

$$\log(wage) = .321 \begin{pmatrix} .100 \end{pmatrix} + .213 \begin{pmatrix} .055 \end{pmatrix} \text{marrmale} - .198 \begin{pmatrix} .058 \end{pmatrix} \text{marrfem} \\ - .110 \begin{pmatrix} .056 \end{pmatrix} \text{singfem} + .079 \begin{pmatrix} .007 \end{pmatrix} educ + .027 \begin{pmatrix} .005 \end{pmatrix} exper - .00054 \begin{pmatrix} .00011 \end{pmatrix} exper^2 \\ + .029 \begin{pmatrix} .007 \end{pmatrix} tenure - .00053 \begin{pmatrix} .00023 \end{pmatrix} tenure^2 \quad n = 2,725, R^2 = .0422$$

- 他の要因と一定として、既婚女性は独身男性(基本グループ)よりも19.8%低い賃金

7-3 複数カテゴリーのダミー変数

◆7-3a ダミー変数による順序情報の組込

■例: 市の信用格付けと地方債金利

$$\widehat{MBR} = \beta_0 + \beta_1 \text{CR}_i + \text{other factors}$$

信用格付け
(0 = worst, 4 = best)

- 信用格付けは順序情報のみのため、この特定化は不十分
 - この情報を組込むためのより良い方法は、ダミーの定義
- $$MBR = \beta_0 + \delta_1 \text{CR}_1 + \delta_2 \text{CR}_2 + \delta_3 \text{CR}_3 + \delta_4 \text{CR}_4 + \text{other factors}$$
- CR_i = 特定の格付けが適用されるかどうかを示すダミー
 - 例: $\text{CR} = 1$ の場合は $\text{CR}_1 = 1$ 、そうでない場合は $\text{CR}_1 = 0$
 - 各効果は最低格付け(=基本グループ)と比較して測定

7-4 ダミー変数の交差項

◆7-4a ダミー変数間の交差項

- ダミー変数交差項はグループ細分化とほぼ同じ

• 例: (7.11) & (7.14)の比較

- 例: 女性 (*fem*) & 既婚 (*mar*)

$$y = \beta_0 + \delta_1 fem + \delta_2 mar + \delta_3 fem \cdot mar + \beta_1 x + u$$

- If $fem = 0$ and $mar = 0$, $y = \beta_0 + \beta_1 x + u$
- If $fem = 0$ and $mar = 1$, $y = \beta_0 + \delta_2 + \beta_1 x + u$
- If $fem = 1$ and $mar = 0$, $y = \beta_0 + \delta_1 + \beta_1 x + u$
- If $fem = 1$ and $mar = 1$, $y = \beta_0 + \delta_1 + \delta_2 + \delta_3 + \beta_1 x + u$
- Base groupは未婚男性

7-4 ダミー変数の交差項

◆7-4b 異なる傾き

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

- 男性: β_0 = 男性の切片 β_1 = 男性の傾き
- 女性: $\beta_0 + \delta_0$ = 女性の切片 $\beta_1 + \delta_1$ = 女性の傾き

- 帰無仮説

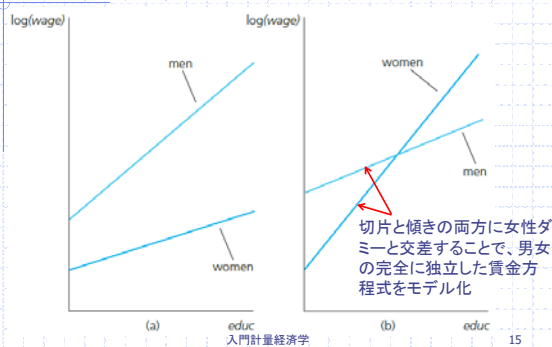
$$H_0: \delta_1 = 0$$

- 教育収益率は男女間で等しい

$$H_0: \delta_0 = 0, \delta_1 = 0$$

- 賃金方程式全体が男女間で等しい

7-4 ダミー変数の交差項



7-4 ダミー変数の交差項

◆交差項を含む賃金方程式

$$\log(wage) = .389 + .227 female - .082 educ - .0056 female \cdot educ + .029 exper - .00058 exper^2 + .032 tenure - .00059 tenure^2, n = 526, R^2 = .441$$

- 質問: この結果は、同じ教育、経験、勤続では女性給与の方が低いという証拠はないことを意味するか?
- 答え: No. これは教育 = 0 のときの効果に過ぎず、交差項を教育 = 12.5 (= 平均教育) 等、中心化する必要
- 仮説「教育収益率は男女間で同じ」に反する証拠なし

7-4 ダミー変数の交差項

◆7-4c グループ間回帰関数の差異の検定

- 非制約モデル(全要因の交差項を含む)

$$cumgpa = \beta_0 + \delta_0 female + \beta_1 sat + \delta_1 female \cdot sat + \beta_2 hspers + \delta_2 female \cdot hspers + \beta_3 tothrs + \delta_3 female \cdot tothrs + u$$

- 制約モデル(全グループで同じ回帰式)

$$cumgpa = \beta_0 + \beta_1 sat + \beta_2 hspers + \beta_3 tothrs + u$$

- 帰無仮説

$$H_0: \delta_0 = 0, \delta_1 = 0, \delta_2 = 0, \delta_3 = 0$$

- すべての交差項の効果はゼロ、すなわち、男女間で同じ回帰係数が適用

7-4 ダミー変数の交差項

- 非制約モデルの推定

$$cumgpa = 1.48 - .353 female + .0011 sat + .00075 female \cdot sat - .0085 hspers - .00055 female \cdot hspers + .0023 tothrs - .00012 female \cdot tothrs$$

- 個別検定では「交差項効果ゼロ」という仮説を棄却できない

- F値による同時検定 Joint test

$$F = \frac{(SSR_p - SSR_{ur})/q}{SSR_{ur}/(n - k - 1)} = \frac{(85.515 - 78.355)/4}{78.355/(366 - 7 - 1)} \approx 8.18$$

- 同時検定では仮説を棄却

7-4 ダミー変数の交差項

◆Chow検定(F値を計算する別方法)

■手法

1. グループ別に回帰式を推定
2. 各推定結果のSSRを合計、非制約SSRを得る
3. 全グループを対象に回帰式を推定、制約SSRを得る
4. 非制約&制約SSRを用いてF値を計算

■メリット

- 説明変数の数を制限

■前提: グループ間で均一誤差分散を仮定

7-5 2値従属変数

◆従属変数が2値のときの線形回帰

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$$

$$\Rightarrow E(y|x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

■従属変数値が0 or 1のとき

$$E(y|x) = 1 \cdot P(y = 1|x) + 0 \cdot P(y = 0|x)$$

■線形確率モデルLinear probability model (LPM)

$$\Rightarrow P(y = 1|x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

$$\Rightarrow \beta_j = \Delta P(y = 1|x) / \Delta x_j$$

- 線形確率モデルでは、係数はy = 1である確率に対する説明変数の効果を表す

7-5 2値従属変数

◆例: 既婚女性の労働参加

$$\begin{aligned} \widehat{inlf} = & .586 - .0034 \text{nwfeinc} + .038 \text{educ} + .039 \text{exper} \\ & - .00060 \text{exper}^2 - .016 \text{age} + .262 \text{kidslt6} \\ & + .0130 \text{kidsge6}, n = 753, R^2 = .264 \end{aligned}$$

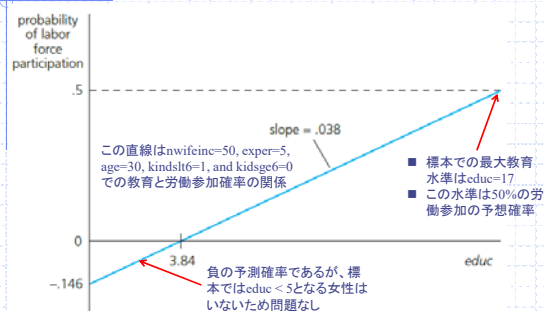
就業 = 1
非就業 = 0

配偶者年収 (\$1000)

有意ではない

6歳未満の子供の人数が1人増えると、その女性が働く確率は26.2%低下

7-5 2値従属変数



7-5 2値従属変数

■線形確率モデルのデメリット

- 予測確率は1以上、0未満になり得る
- 限界的な確率効果が論理的に不可能なときも
- 線形確率モデルは必然的に不均一分散

■Bernoulli変数の分散

$$Var(y|x) = P(y = 1|x) [1 - P(y = 1|x)]$$

■不均一分散の一致標準誤差HCSEを計算する必要

■線形確率モデルのメリット

- 簡単な推定と解釈
- 実際、合理的に良い推定効果と予測

7-6 政策分析とプログラム評価

◆例: 労働者生産性への職業訓練助成効果

$$\begin{aligned} \widehat{\log(scrap)} = & 4.99 - .052 \text{grant} - .455 \log(sales) \\ & + .639 \log(employ), n = 50, R^2 = .072 \end{aligned}$$

廃棄率

■grant: 訓練助成受給企業=1、非受給企業=0

■Treatment group: 受給企業、Control group: 非受給企業

■解釈

- 生産性への明確な訓練助成効果はない。ただし...
 - 助成金給付は先着順のため、これはランダムではない
 - 労働者生産性が低い企業ほど、生産性向上の機会と考えると早く申請した可能性

7-6 政策分析とプログラム評価

◆自己選択問題Self-selection problems

- プログラム参加と内生性
 - ◆ 政策・プログラム評価例において、政策プログラムへの参加状況は結果に影響を及ぼす特性と関連
 - 個人が個々の特性や見通しに応じて自らの行動(プログラム参加等)を決定すること起因
- ◆ 実験評価
 - 実験ではプログラム参加or不参加はランダムに決定
 - $y = \beta_0 + \beta_1 \text{partic} + u$
 - プログラム参加のダミー変数は、結果yに影響を及ぼす他の要因uとは無関係
 - 説明変数が多ければ内生性問題を緩和する可能性大

入門計量経済学

25

7-6 政策分析とプログラム評価

◆内生的ダミー説明変数の例

- 非白人顧客は差別されているか？

$$\text{approved} = \beta_0 + \beta_1 \text{nonwhite} + \beta_2 \text{income} + \beta_3 \text{wealth} + \beta_4 \text{credrate} + u$$

信用承認 人種ダミー 信用格付

- 融資承認に重要な特性(職業、失業等)を制御することが重要
- 非白人ダミーと関連する重要な特性を除外すると、見せかけの「差別の証拠」が生じる

入門計量経済学

26