高知工科大学 経済・マネジメント学群

計量経済学

9. 交差項の利用

やない ゆうき
矢内 勇生

https://yukiyanai.github.io

yanai.yuki@kochi-tech.ac.jp

このトピックの目標

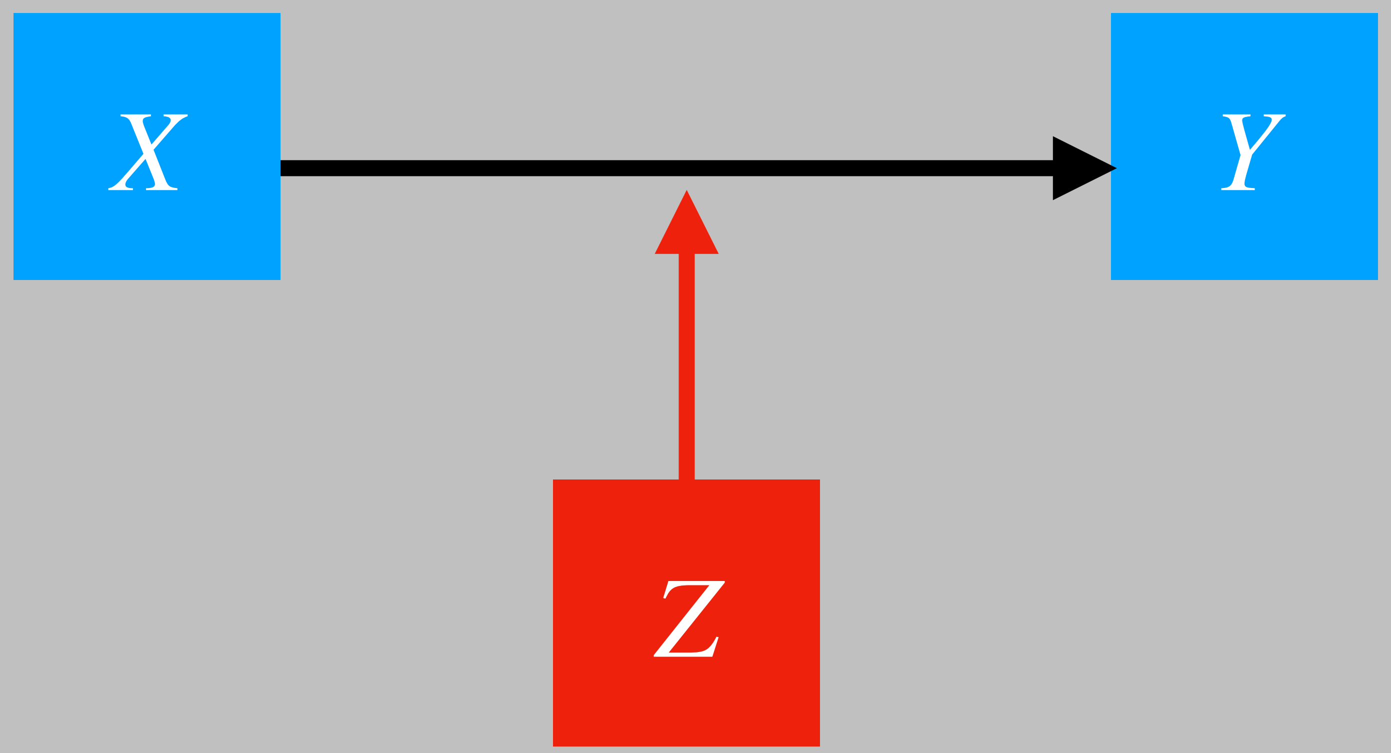
- 回帰分析で交差項を使う方法を理解する
  - 交差項とは？
  - 交差項を使って回帰分析の結果を解釈する方法

交差項を含む回帰分析

説明変数が応答変数に与える影響は一定か？

- 応答変数 Y
- 説明変数 X
- 回帰モデル：$Y_i \sim \text{Normal}(\beta_0 + \beta_1 X_i, \sigma)$
  - モデルの仮定（の1つ）：β_1 はある1つの値（定数）
    - つまり、X が1単位変化するのに応じた Y の変化量は一定

一定とは限らないのでは？



- Z：調整変数（説明変数の1種）
  - Z の値によって、「X が Y に与える影響」が変わる

調整変数を含む回帰モデル

- 回帰モデル：$Y_i \sim \text{Normal}(\mu_i = \beta_0 + \beta_1 X_i, \sigma)$
  - β_0 は Z の関数：$\beta_0 = \gamma_0 + \gamma_2 Z_i$ とする
  - β_1は Z の関数：$\beta_1 = \gamma_1 + \gamma_3 Z_i$ とする

$$
\begin{aligned}
\mu_i &= \beta_0 + \beta_1 X_i \\
&= (\gamma_0 + \gamma_2 Z_i) + (\gamma_1 + \gamma_3 Z_i) X_i \\
&= \gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i
\end{aligned}
$$

  - よって、回帰モデルは、

$$
Y_i \sim \text{Normal}(\gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i, \sigma)
$$

調整変数を含む回帰モデル（続）

- $Y_i \sim \text{Normal}(\gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i, \sigma)$
  - Y を X と Z と XZ に回帰する重回帰
    - γ_k $(k = 0,1,2,3)$ を推定し、**そこから β_1 を推定する**
  - $\gamma_3 X_i Z_i$：交差項, 交互作用項 (interaction term)
    - Rでは、`lm(Y ~ X * Z, data = d)`

Z がダミー変数の場合

$$Y_i \sim \text{Normal}(\mu_i = \gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i, \sigma)$$

- $Z_i = 0$ のとき：

$$\hat{Y}_i = \gamma_0 + \gamma_1 X_i + \gamma_2 \cdot 0 + \gamma_3 \cdot X_i \cdot 0 = \gamma_0 + \gamma_1 X_i$$

- $Z_i = 1$ のとき：

$$\begin{aligned}\hat{Y}_i &= \gamma_0 + \gamma_1 X_i + \gamma_2 \cdot 1 + \gamma_3 \cdot X_i \cdot 1 = \gamma_0 + \gamma_1 X_i + \gamma_2 + \gamma_3 X_i \\ &= (\gamma_0 + \gamma_2) + (\gamma_1 + \gamma_3) X_i\end{aligned}$$

★ Z の値によって、

- 回帰直線の「切片」が変わる（$\gamma_2 \neq 0$ のとき）: γ_0 or $\gamma_0 + \gamma_2$
- 回帰直線の「傾き」が変わる（$\gamma_3 \neq 0$ のとき）: γ_1 or $\gamma_1 + \gamma_3$

Z がダミー変数の場合（続）

$Y_i \sim \text{Normal}(\mu_i = \gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i, \sigma)$

- $Z_i = 0$ のとき：切片は γ_0 、傾き（X が Y に与える影響）は γ_1
- $Z_i = 1$ のとき：切片は $\gamma_0 + \gamma_2$ 、傾き（X が Y に与える影響）は $\gamma_1 + \gamma_3$

★ 推定された偏回帰係数の意味

- γ_0 ：$Z = 0$ のときの回帰直線の切片
- γ_1 ：$Z = 0$ のときの回帰直線の傾き
- γ_2 ：$Z = 1$ のときと $Z = 0$ のときとの回帰直線の切片の差
- γ_3 ：$Z = 1$ のときと $Z = 0$ のときとの回帰直線の傾きの差

Z が量的変数の場合

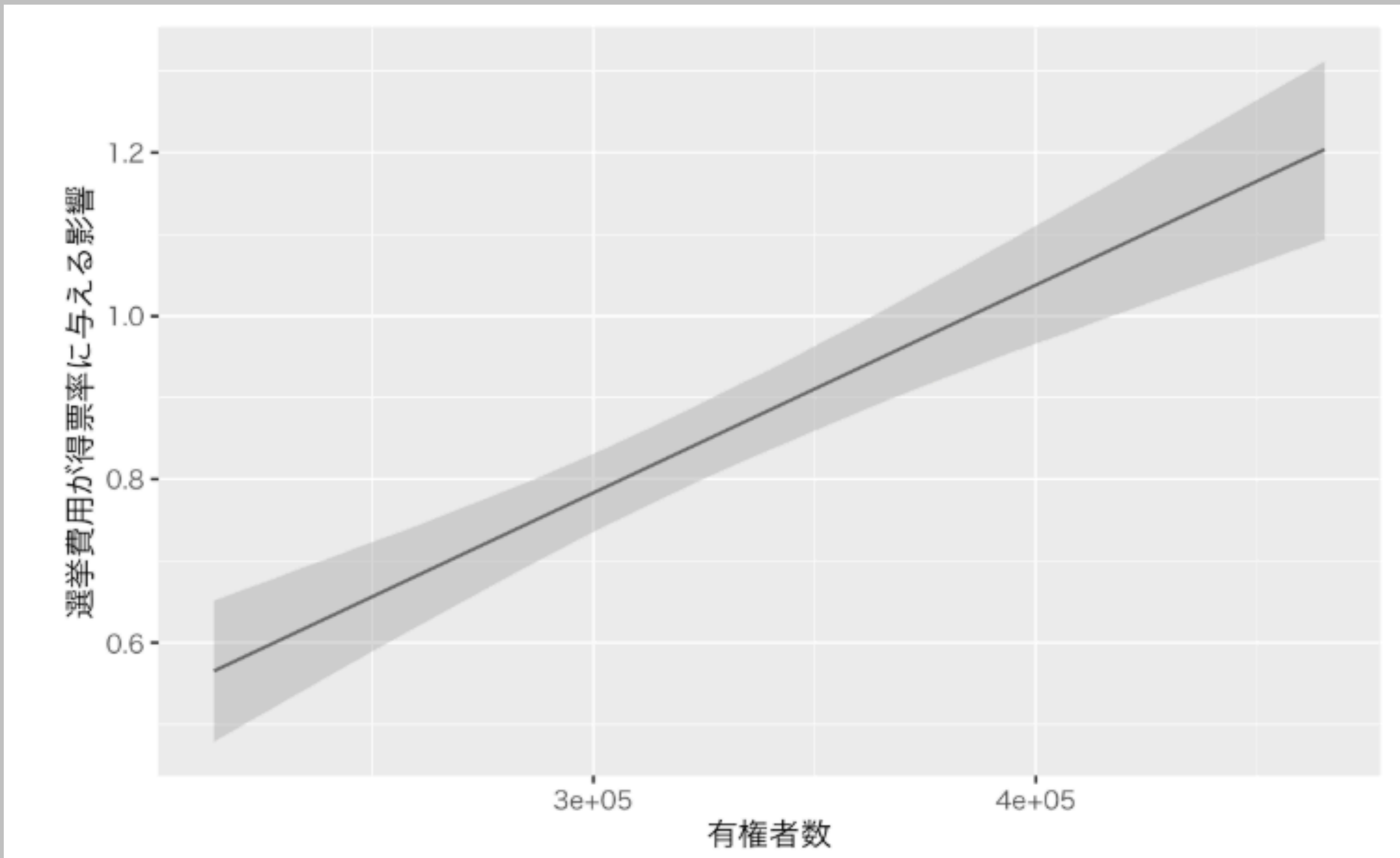
$Y_i \sim \text{Normal}(\mu_i = \gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i, \sigma)$

- X-Y 平面における回帰直線の切片：$\gamma_0 + \gamma_2 Z_i$
- X-Y 平面における回帰直線の傾き：$\gamma_1 + \gamma_3 Z_i$

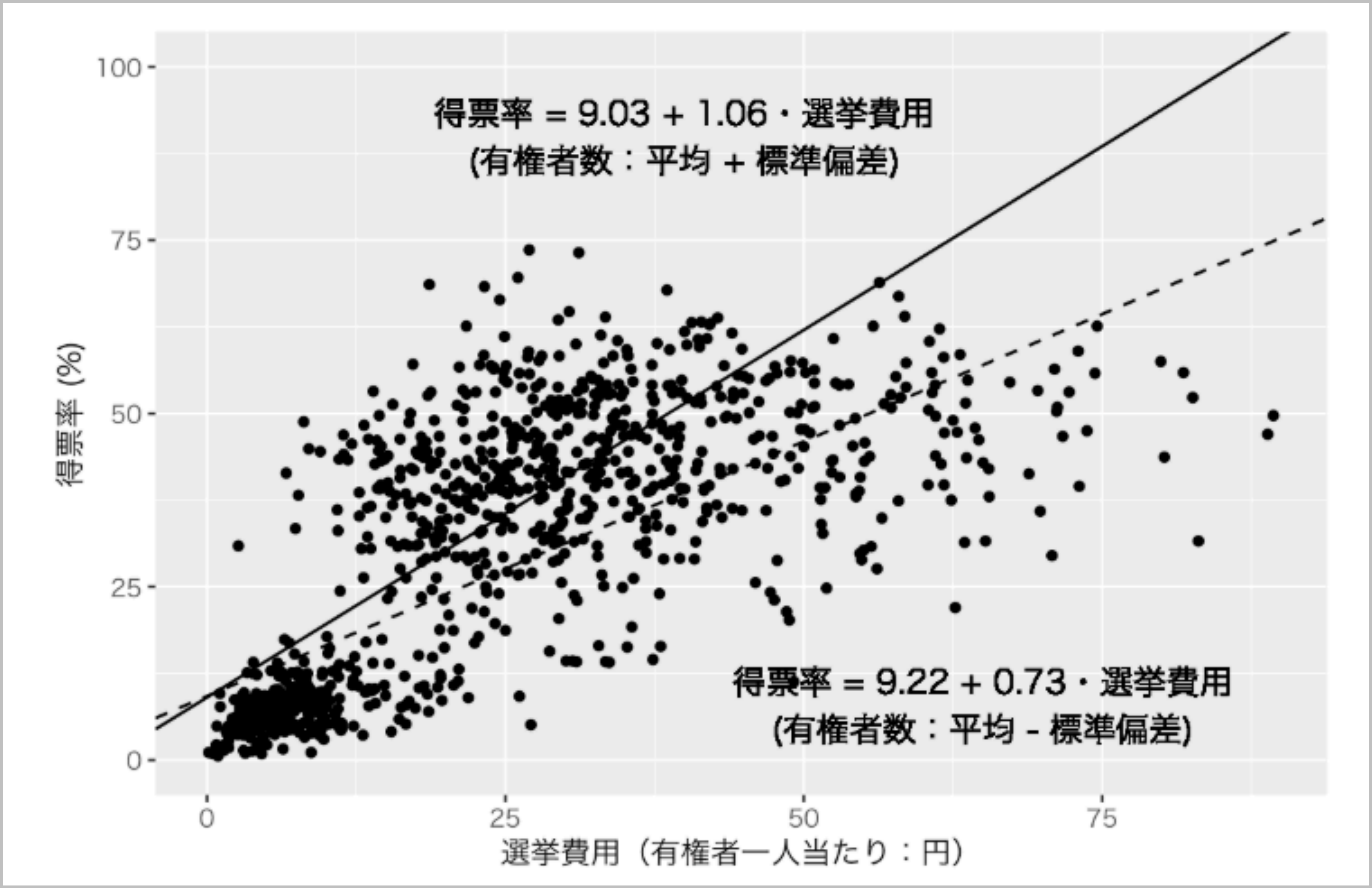
★ 切片も傾きも Z_i の値によって変わる！

- $Z_i = 0$ のとき：切片は γ_0 、傾き（X が Y に与える影響）は γ_1
- $Z_i \neq 0$ のとき：切片も傾き（X が Y に与える影響）も、Z_i の値による
  - **回帰係数だけを見ても、X が Y に与える影響はわからない！！！**
- Z_i **を横軸、**$\gamma_1 + \gamma_3 Z_i$**（**X **が** Y **に与える影響）を縦軸にした図を作る！(A)**
- Z_i **をいくつかの値に固定して、複数の回帰直線を図示する！ (B)**

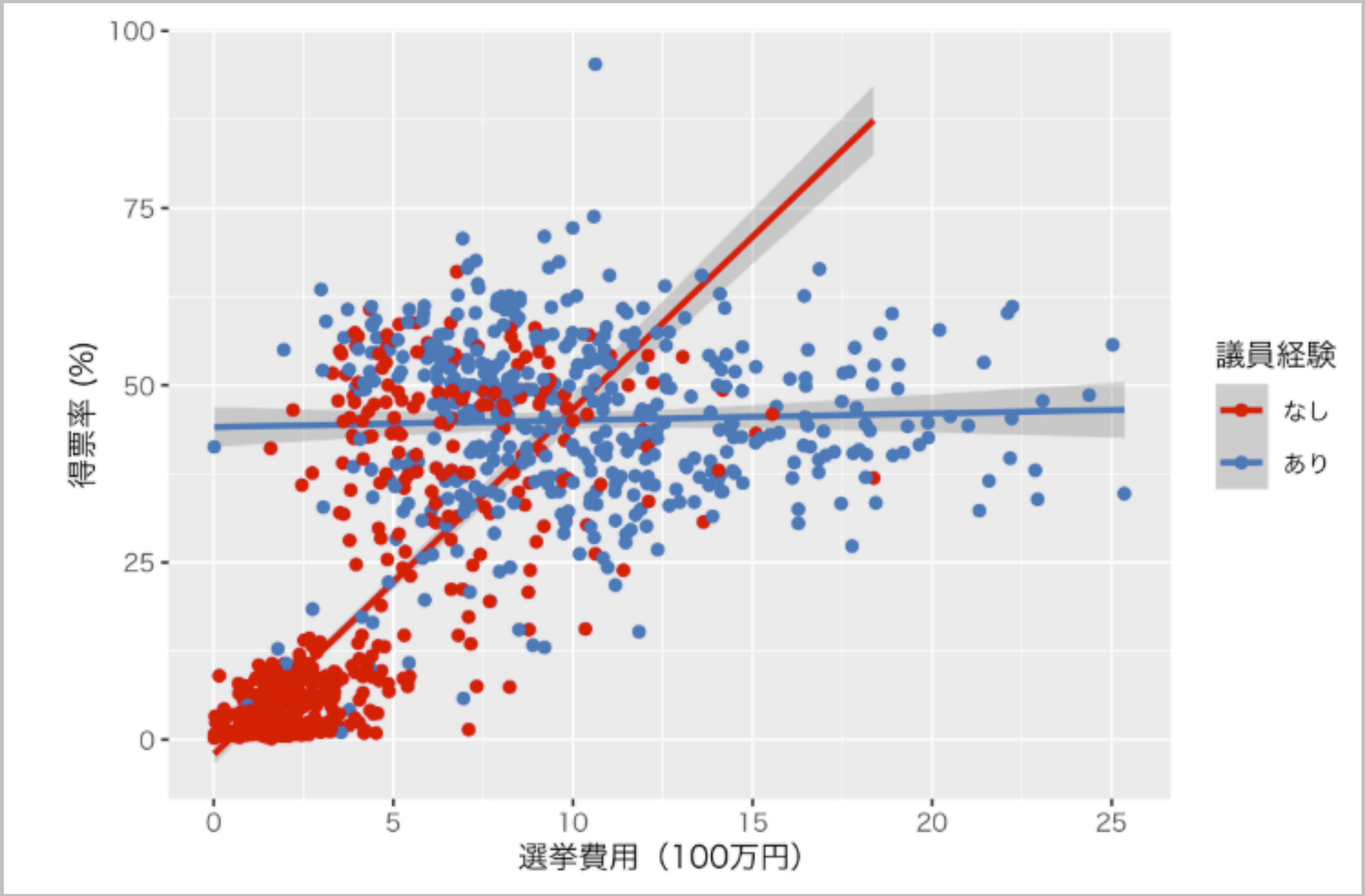
(A) の例：説明変数が選挙費用, 調整変数が有権者数の場合



(B) の例：説明変数が選挙費用, 調整変数が有権者数の場合



(B) の例：説明変数が選挙費用, 調整変数が議員経験の場合



Z が量的変数の場合（続）

$$Y_i \sim \text{Normal}(\mu_i = \gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i, \sigma)$$

★ 推定された偏回帰係数の意味

- γ_0：$Z = 0$ のときの回帰直線の切片
- γ_1：$Z = 0$ のときの回帰直線の傾き
- γ_2：？（説明しようと思えば説明できるが、わかりにくい）
- γ_3：？（説明しようと思えば説明できるが、わかりにくい）

- Z_i が0をとらない変数だったら？？？
  - それぞれの偏回帰係数に意味がない：回帰係数を表で提示しても、読者に意味が伝わりにくい

説明変数を中心化する

$Y_i \sim \text{Normal}(\tilde{\mu}_i = \tilde{\gamma}_0 + \tilde{\gamma}_1 \tilde{X}_i + \tilde{\gamma}_2 \tilde{Z}_i + \tilde{\gamma}_3 \tilde{X}_i \tilde{Z}_i, \sigma)$

- $\tilde{X}_i$：X_i を中心化したもの
- $\tilde{Z}_i$：Z_i を中心化したもの

★推定された偏回帰係数の意味

- $\tilde{\gamma}_0$：$\tilde{Z} = 0$ のとき、すなわち Z が平均値のときの回帰直線の切片
- $\tilde{\gamma}_1$：$\tilde{Z} = 0$ のとき、すなわち Z が平均値のときの回帰直線の傾き
- $\tilde{\gamma}_2$：？（説明しようと思えば説明できるが、わかりにくい）
- $\tilde{\gamma}_3$：？（説明しようと思えば説明できるが、わかりにくい）

- 中心化することによって、$\tilde{\gamma}_0$ と $\tilde{\gamma}_1$ の意味だけは解釈可能になることが保証される

交差項と交差項の元となる変数

- XZ には、X も Z も含まれている
  - $Y_i \sim \text{Normal}(\gamma_0 + \gamma_3 X_i Z_i, \sigma)$ でいいのでは？
    - ✦ 一般的には、ダメ！（Brambor et al. 2006 を参照）
- $\gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i$ (1) と $\gamma_0 + \gamma_3 X_i Z_i$ (2) は何が違うのか？
  - (2) は (1) の式に $\gamma_1 = 0$ かつ $\gamma_2 = 0$ という制約を加えている：強い仮定
    - 「$\gamma_1 = 0$ かつ $\gamma_2 = 0$ である」という理論的根拠があれば (2) を使っても良い
    - そうでなければ、(1) を推定する
  - 同様の理由で、一部だけ除くのもダメ：詳しくは実習で

交差項は除いていいの？

- $\gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \gamma_3 X_i Z_i$ (3) と $\gamma_0 + \gamma_1 X_i + \gamma_2 Z_i$ (4) は何が違うのか？
  - (4) は (3) の式に $\gamma_3 = 0$ という制約を加えている：強い仮定（？）
  - 「$\gamma_3 = 0$である」という理論的根拠がない限り、交差項を常に使うべきでは？
    - 「当てはまりの良いモデル」と「シンプルだが有用なモデル（予測性能が良いことが多い）」のバランスを考える

交差項のまとめ

- （交差項がなくてもそうだが、交差項がある場合は特に）説明変数を中心化すべき
- 交差項を使うときは、交差項を構成するそれぞれの変数も説明変数に加える（理論的に正当化できる場合は除く）
- 交差項があるときは、**偏回帰係数だけでは意味がわからない**（結果を表で示すだけで不十分！）
  - 効果を**可視化**（作図）する！

次のトピック

多重比較