第 5 章：重回帰

Jonathan Roth

数理計量経済学 I
ブラウン大学

アウトライン

1. 重回帰と OLS の導出
2. 回帰と因果関係
3. 回帰に関するその他のトピック

1 つの「説明変数」を超えて

- これまでは、1 つのスカラー X_i に対して、条件付き期待値関数 (CEF) $E[Y_i|X_i = x] \approx \alpha + x\beta$ を近似する方法として回帰を扱ってきました。
    - そして、推定対象 (α, β) を OLS で推定する方法を示しました。

1つの「説明変数」を超えて

- これまでは、1つのスカラー X_i に対して、条件付き期待値関数（CEF）$E[Y_i|X_i = x] \approx \alpha + x\beta$ を近似する方法として回帰を扱ってきました。
  - そして、推定対象 (α, β) を OLS で推定する方法を示しました。
- 次に、これを一般化して、ベクトル $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{iK})'$ に対して $E[Y_i|\mathbf{X}_i = \mathbf{x}] \approx \mathbf{x}'\boldsymbol{\beta}$ を近似・推定する方法を見ていきます。
  - 注意：いつものように、ベクトルや行列は太字で表記します。

1つの「説明変数」を超えて

- これまでは、1つのスカラー X_i に対して、条件付き期待値関数（CEF）$E[Y_i|X_i = x] \approx \alpha + x\beta$ を近似する方法として回帰を扱ってきました。
  - そして、推定対象 (α, β) を OLS で推定する方法を示しました。
- 次に、これを一般化して、ベクトル $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{iK})'$ に対して $E[Y_i|\mathbf{X}_i = \mathbf{x}] \approx \mathbf{x}'\boldsymbol{\beta}$ を近似・推定する方法を見ていきます。
  - 注意：いつものように、ベクトルや行列は太字で表記します。
- これには主に2つの動機があります：

1 つの「説明変数」を超えて

- これまでは、1 つのスカラー X_i に対して、条件付き期待値関数（CEF）$E[Y_i|X_i = x] \approx \alpha + x\beta$ を近似する方法として回帰を扱ってきました。
  - そして、推定対象 (α, β) を OLS で推定する方法を示しました。
- 次に、これを一般化して、ベクトル $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{iK})'$ に対して $E[Y_i|\mathbf{X}_i = \mathbf{x}] \approx \mathbf{x}'\boldsymbol{\beta}$ を近似・推定する方法を見ていきます。
  - 注意：いつものように、ベクトルや行列は太字で表記します。
- これには主に 2 つの動機があります：

1. 回帰を用いて因果的効果を識別したいが、条件付き非交絡性は複数のコントロール変数を入れなければ妥当ではない場合。
   - ブラウン大対 URI の例では、高校の GPA、世帯所得、SAT スコア、人種などをコントロールしたいかもしれません。

1つの「説明変数」を超えて

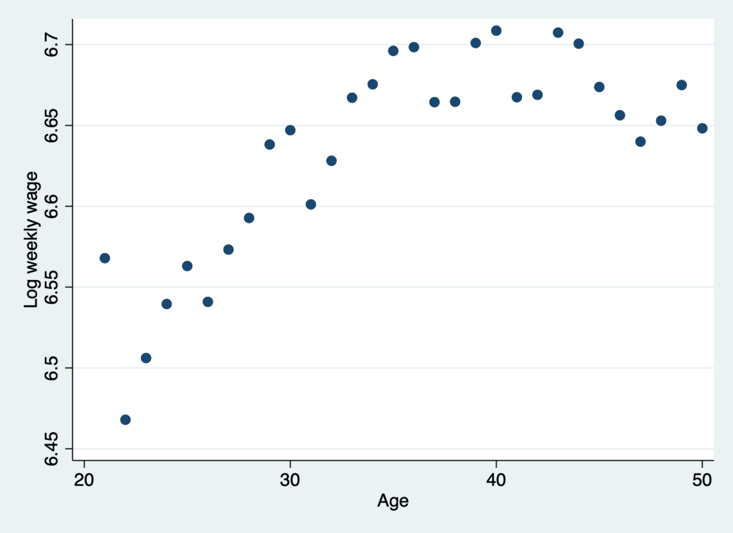
- これまでは、1つのスカラー X_i に対して、条件付き期待値関数（CEF）$E[Y_i|X_i = x] \approx \alpha + x\beta$ を近似する方法として回帰を扱ってきました。
  - そして、推定対象 (α, β) を OLS で推定する方法を示しました。

- 次に、これを一般化して、ベクトル $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{iK})'$ に対して $E[Y_i|\mathbf{X}_i = \mathbf{x}] \approx \mathbf{x}'\boldsymbol{\beta}$ を近似・推定する方法を見ていきます。
  - 注意：いつものように、ベクトルや行列は太字で表記します。

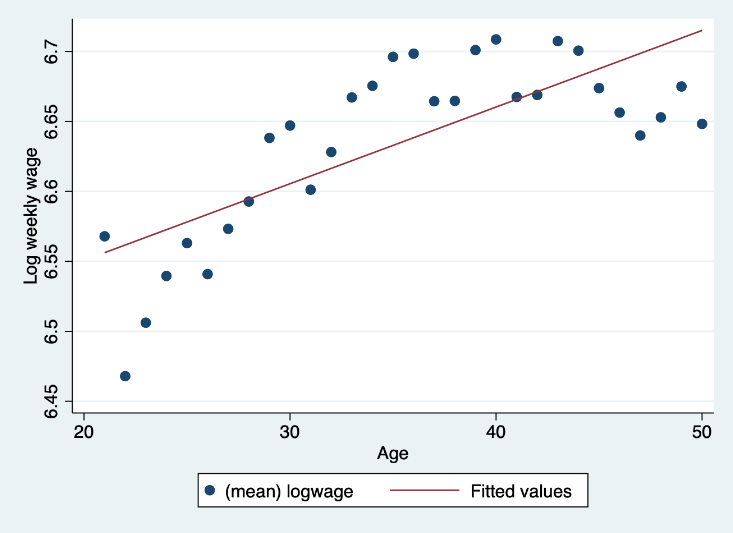
- これには主に2つの動機があります：

1. 回帰を用いて因果的効果を識別したいが、条件付き非交絡性は複数のコントロール変数を入れなければ妥当ではない場合。
   - ブラウン大対 URI の例では、高校の GPA、世帯所得、SAT スコア、人種などをコントロールしたいかもしれません。
2. 非線形な CEF の近似が欲しい場合：例 $E[Y_i \mid X_i] \approx \alpha + X_i\beta + X_i^2\gamma$
   - $\mathbf{X}_i = (1, X_i, X_i^2)'$ と設定することで、回帰を「騙して」これを行わせることができます。

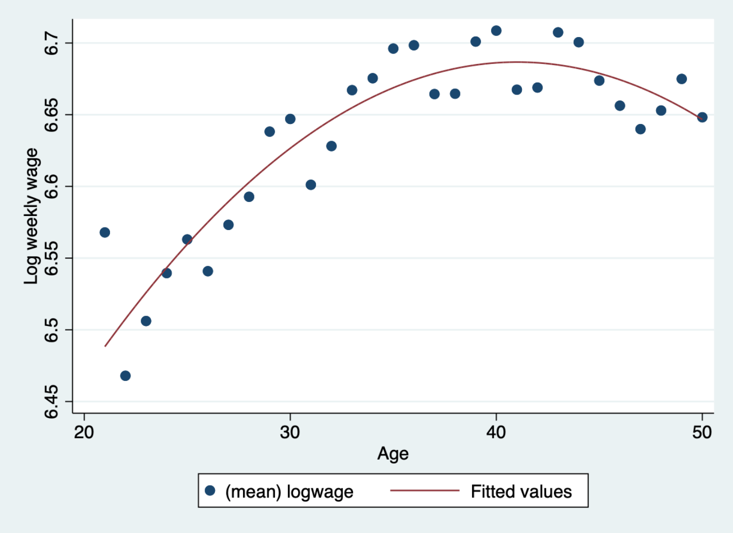
年齢別・対数賃金



OLS 回帰（線形フィット）



OLS 回帰（2 次式フィット）



最小二乗問題としての重回帰

- 単変量 OLS では、以下を解きました：

$$(\alpha, \beta) = \arg\min_{a,b} E[(Y_i - (a + bX_i))^2]$$

CEF が線形であれば $E[Y|X] = \alpha + \beta X$ となり、そうでなければ $\alpha + \beta X$ が最良の線形近似を与えることを見ました。

最小二乗問題としての重回帰

- 単変量 OLS では、以下を解きました：

$$(\alpha, \beta) = \arg\min_{a,b} E[(Y_i - (a + bX_i))^2]$$

CEF が線形であれば $E[Y|X] = \alpha + \beta X$ となり、そうでなければ $\alpha + \beta X$ が最良の線形近似を与えることを見ました。

- ここで多変量の場合を考えます：

$$\boldsymbol{\beta} = \arg\min_{\mathbf{b}} E\left[(Y_i - \mathbf{X}_i'\boldsymbol{b})^2\right]$$

- 単変量の場合と同様の手順で、CEF が $\mathbf{X}$ に関して線形であれば $E[Y|\mathbf{X}] = \mathbf{X}'\boldsymbol{\beta}$ となり、そうでなければ $\mathbf{X}'\boldsymbol{\beta}$ が CEF を平均二乗誤差（MSE）の意味で最小化する近似となることを示せます。

回帰係数の算出

- 母集団回帰係数 $\boldsymbol{\beta}$ は以下の最小二乗問題を解くことで得られます：

$$\boldsymbol{\beta} = \arg\min_{\mathbf{b}} E[(Y_i - \mathbf{X}_i'\mathbf{b})^2]$$

回帰係数の算出

- 母集団回帰係数 $\boldsymbol{\beta}$ は以下の最小二乗問題を解くことで得られます：

$$\boldsymbol{\beta} = \arg\min_{\mathbf{b}} E[(Y_i - \mathbf{X}_i'\mathbf{b})^2]$$

- $\boldsymbol{\beta}$ について微分（勾配）をとり、0 と置きます：

$$\frac{d}{d\boldsymbol{\beta}} E[(Y_i - \mathbf{X}_i'\boldsymbol{\beta})^2] = \mathbf{0}$$

回帰係数の算出

- 母集団回帰係数 $\boldsymbol{\beta}$ は以下の最小二乗問題を解くことで得られます：

$$\boldsymbol{\beta} = \arg\min_{\mathbf{b}} E[(Y_i - \mathbf{X}_i'\mathbf{b})^2]$$

- $\boldsymbol{\beta}$ について微分（勾配）をとり、0 と置きます：

$$\frac{d}{d\boldsymbol{\beta}} E[(Y_i - \mathbf{X}_i'\boldsymbol{\beta})^2] = \mathbf{0}$$

$$\Rightarrow E[\frac{d}{d\boldsymbol{\beta}}(Y_i - \mathbf{X}_i'\boldsymbol{\beta})^2] = \mathbf{0}$$

回帰係数の算出

- 母集団回帰係数 $\boldsymbol{\beta}$ は以下の最小二乗問題を解くことで得られます：

$$\boldsymbol{\beta} = \arg\min_{\mathbf{b}} E[(Y_i - \mathbf{X}_i'\mathbf{b})^2]$$

- $\boldsymbol{\beta}$ について微分（勾配）をとり、0 と置きます：

$$\begin{aligned}
&\frac{d}{d\boldsymbol{\beta}} E[(Y_i - \mathbf{X}_i'\boldsymbol{\beta})^2] = \mathbf{0} \\
\Rightarrow &E[\frac{d}{d\boldsymbol{\beta}}(Y_i - \mathbf{X}_i'\boldsymbol{\beta})^2] = \mathbf{0} \\
\Rightarrow &E[-2\mathbf{X}_i(Y_i - \mathbf{X}_i'\boldsymbol{\beta})] = \mathbf{0}
\end{aligned}$$

回帰係数の算出

- 母集団回帰係数 $\boldsymbol{\beta}$ は以下の最小二乗問題を解くことで得られます：

$$\boldsymbol{\beta} = \arg\min_{\mathbf{b}} E[(Y_i - \mathbf{X}_i'\mathbf{b})^2]$$

- $\boldsymbol{\beta}$ について微分（勾配）をとり、0 と置きます：

$$\begin{aligned}
&\frac{d}{d\boldsymbol{\beta}} E[(Y_i - \mathbf{X}_i'\boldsymbol{\beta})^2] = \mathbf{0} \\
\Rightarrow & E[\frac{d}{d\boldsymbol{\beta}}(Y_i - \mathbf{X}_i'\boldsymbol{\beta})^2] = \mathbf{0} \\
\Rightarrow & E[-2\mathbf{X}_i(Y_i - \mathbf{X}_i'\boldsymbol{\beta})] = \mathbf{0} \\
\Rightarrow & E[\mathbf{X}_i Y_i] = E[\mathbf{X}_i\mathbf{X}_i']\boldsymbol{\beta}
\end{aligned}$$

母集団と標本における重回帰

- $\boldsymbol{\beta}$ について解くと、母平均を用いた式が得られます：

$$\boldsymbol{\beta} = E[\mathbf{X}_i \mathbf{X}_i']^{-1} E[\mathbf{X}_i Y_i]$$

  - 少し計算すれば、$\mathbf{X}_i = (1, X_i)'$ のときに単変量の公式 $\beta = \frac{Cov(X_i, Y_i)}{Var(X_i)}$ および $\alpha = E[Y_i] - E[X_i]\beta$ に一致することを示せます。

母集団と標本における重回帰

- $\boldsymbol{\beta}$ について解くと、母平均を用いた式が得られます：

$$\boldsymbol{\beta} = E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} E[\boldsymbol{X}_i Y_i]$$

  - 少し計算すれば、$\mathbf{X}_i = (1, X_i)'$ のときに単変量の公式 $\beta = \frac{Cov(X_i, Y_i)}{Var(X_i)}$ および $\alpha = E[Y_i] - E[X_i]\beta$ に一致することを示せます。

- $\boldsymbol{\beta}$ を推定するために、母平均を標本平均に置き換えます：

母集団と標本における重回帰

- $\boldsymbol{\beta}$ について解くと、母平均を用いた式が得られます：

$$\boldsymbol{\beta} = E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} E[\boldsymbol{X}_i Y_i]$$

  - 少し計算すれば、$\mathbf{X}_i = (1, X_i)'$ のときに単変量の公式 $\beta = \frac{Cov(X_i, Y_i)}{Var(X_i)}$ および $\alpha = E[Y_i] - E[X_i]\beta$ に一致することを示せます。

- $\boldsymbol{\beta}$ を推定するために、母平均を標本平均に置き換えます：

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i Y_i \right)$$

母集団と標本における重回帰

- $\boldsymbol{\beta}$ について解くと、母平均を用いた式が得られます：

$$\boldsymbol{\beta} = E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} E[\boldsymbol{X}_i Y_i]$$

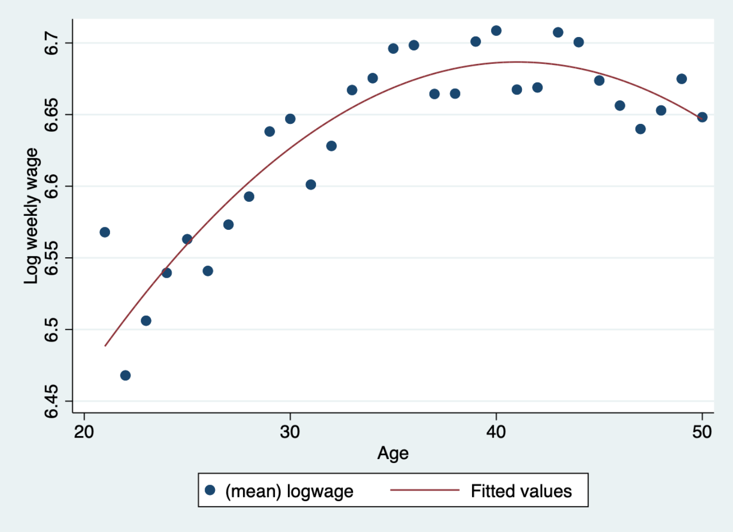
  - 少し計算すれば、$\mathbf{X}_i = (1, X_i)'$ のときに単変量の公式 $\beta = \frac{Cov(X_i, Y_i)}{Var(X_i)}$ および $\alpha = E[Y_i] - E[X_i]\beta$ に一致することを示せます。

- $\boldsymbol{\beta}$ を推定するために、母平均を標本平均に置き換えます：

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i Y_i \right)$$

- これで、任意のベクトル $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{iK})'$ に対して $E[Y_i \mid \mathbf{X}_i] \approx \mathbf{X}_i' \boldsymbol{\beta}$ を推定する一般的な方法が手に入りました。

対数賃金の年齢に対する２次回帰



定数項	5.8591
年齢 (Age)	0.0403
年齢の２乗 (Age^2)	-0.0005

2 次回帰係数の解釈

- 2 次式のフィットでは、以下が得られます：

$$E[Y_i|X_i = x] \approx \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 x + \hat{\boldsymbol{\beta}}_2 x^2$$

2 次回帰係数の解釈

- 2 次式のフィットでは、以下が得られます：

$$E[Y_i|X_i = x] \approx \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 x + \hat{\boldsymbol{\beta}}_2 x^2$$

- $E[Y_i|X_i = x]$ の傾きはいくらでしょうか？ 微分すると：

2 次回帰係数の解釈

- 2 次式のフィットでは、以下が得られます：

$$E[Y_i|X_i = x] \approx \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 x + \hat{\boldsymbol{\beta}}_2 x^2$$

- $E[Y_i|X_i = x]$ の傾きはいくらでしょうか？ 微分すると：

$$\frac{d}{dx} E[Y_i|X_i = x] \approx \hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 x$$

2 次回帰係数の解釈

- 2 次式のフィットでは、以下が得られます：

$$E[Y_i|X_i = x] \approx \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 x + \hat{\boldsymbol{\beta}}_2 x^2$$

- $E[Y_i|X_i = x]$ の傾きはいくらでしょうか？ 微分すると：

$$\frac{d}{dx} E[Y_i|X_i = x] \approx \hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 x$$

- 重回帰から推定される微分、この場合は $\hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 x$ は、しばしば「限界効果（marginal effect）」と呼ばれます。
  - この用語は少し注意が必要です。これは単に CEF の推定された微分であり、必ずしも因果的な効果であるとは限りません。

例題の係数の解釈

定数項 ($\hat{\beta}_0$)	5.8591
年齢 ($\hat{\beta}_1$)	0.0403
年齢の 2 乗 ($\hat{\beta}_2$)	-0.0005

- 対数所得の平均の、年齢に関する推定された傾きは何でしょうか？

例題の係数の解釈

定数項 ($\hat{\beta}_0$)	5.8591
年齢 ($\hat{\beta}_1$)	0.0403
年齢の 2 乗 ($\hat{\beta}_2$)	-0.0005

- 対数所得の平均の、年齢に関する推定された傾きは何でしょうか？
- $\hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 \cdot Age$

例題の係数の解釈

定数項 ($\hat{\beta}_0$)	5.8591
年齢 ($\hat{\beta}_1$)	0.0403
年齢の 2 乗 ($\hat{\beta}_2$)	-0.0005

- 対数所得の平均の、年齢に関する推定された傾きは何でしょうか？
- $\hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 \cdot Age = 0.0403 - 2 \times 0.0005 \cdot Age =$

例題の係数の解釈

定数項 ($\hat{\beta}_0$)	5.8591
年齢 ($\hat{\beta}_1$)	0.0403
年齢の 2 乗 ($\hat{\beta}_2$)	-0.0005

- 対数所得の平均の、年齢に関する推定された傾きは何でしょうか？
- $\hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 \cdot Age = 0.0403 - 2 \times 0.0005 \cdot Age = 0.0403 - 0.001 \cdot Age.$

例題の係数の解釈

定数項 ($\hat{\beta}_0$)	5.8591
年齢 ($\hat{\beta}_1$)	0.0403
年齢の 2 乗 ($\hat{\beta}_2$)	-0.0005

- 対数所得の平均の、年齢に関する推定された傾きは何でしょうか？
- $\hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 \cdot Age = 0.0403 - 2 \times 0.0005 \cdot Age = 0.0403 - 0.001 \cdot Age.$
- 推定された所得が最大になる年齢はいくつでしょうか？

例題の係数の解釈

定数項 ($\hat{\beta}_0$)	5.8591
年齢 ($\hat{\beta}_1$)	0.0403
年齢の 2 乗 ($\hat{\beta}_2$)	-0.0005

- 対数所得の平均の、年齢に関する推定された傾きは何でしょうか？
- $\hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 \cdot Age = 0.0403 - 2 \times 0.0005 \cdot Age = 0.0403 - 0.001 \cdot Age.$
- 推定された所得が最大になる年齢はいくつでしょうか？

$$0.0403 - 0.001 \cdot Age = 0$$

例題の係数の解釈

定数項 ($\hat{\beta}_0$)	5.8591
年齢 ($\hat{\beta}_1$)	0.0403
年齢の 2 乗 ($\hat{\beta}_2$)	-0.0005

- 対数所得の平均の、年齢に関する推定された傾きは何でしょうか？
- $\hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 \cdot Age = 0.0403 - 2 \times 0.0005 \cdot Age = 0.0403 - 0.001 \cdot Age.$
- 推定された所得が最大になる年齢はいくつでしょうか？

$$0.0403 - 0.001 \cdot Age = 0 \Rightarrow Age = 0.0403/0.001 = 40.3$$

行列代数を用いた多変量 OLS の書き換え

- 以下を示しました：

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i Y_i \right)$$

- この公式は、行列表記を用いることでより簡潔に表現されます。

行列代数を用いた多変量 OLS の書き換え

- 以下を示しました：

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i Y_i \right)$$

- この公式は、行列表記を用いることでより簡潔に表現されます。
- $\boldsymbol{X}$ を $N \times K$ 行列とし、その i 行 k 列の要素を X_{ik} とします。

行列代数を用いた多変量 OLS の書き換え

- 以下を示しました：

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i Y_i \right)$$

- この公式は、行列表記を用いることでより簡潔に表現されます。
- $\boldsymbol{X}$ を $N \times K$ 行列とし、その i 行 k 列の要素を X_{ik} とします。同様に $\boldsymbol{Y} = (Y_1, \ldots, Y_N)'$ とします。

行列代数を用いた多変量 OLS の書き換え

- 以下を示しました：

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i Y_i \right)$$

- この公式は、行列表記を用いることでより簡潔に表現されます。
- $\boldsymbol{X}$ を $N \times K$ 行列とし、その i 行 k 列の要素を X_{ik} とします。同様に $\boldsymbol{Y} = (Y_1, \dots, Y_N)'$ とします。
- 例えば、$\boldsymbol{X}_i = (1, X_i)'$ かつ $N = 3$ であれば：

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \end{pmatrix}$$

行列代数を用いた多変量 OLS の書き換え

- 以下を示しました：

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i Y_i \right)$$

- この公式は、行列表記を用いることでより簡潔に表現されます。

- $\boldsymbol{X}$ を $N \times K$ 行列とし、その i 行 k 列の要素を X_{ik} とします。同様に $\boldsymbol{Y} = (Y_1, \ldots, Y_N)'$ とします。

- 例えば、$\boldsymbol{X}_i = (1, X_i)'$ かつ $N = 3$ であれば：

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \end{pmatrix} \qquad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

行列代数を用いた多変量 OLS の書き換え

- 以下を示しました：

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i Y_i \right)$$

- この公式は、行列表記を用いることでより簡潔に表現されます。

- $\boldsymbol{X}$ を $N \times K$ 行列とし、その i 行 k 列の要素を X_{ik} とします。同様に $\boldsymbol{Y} = (Y_1, \dots, Y_N)'$ とします。

- 例えば、$\boldsymbol{X}_i = (1, X_i)'$ かつ $N = 3$ であれば：

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \end{pmatrix} \qquad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

- この表記を用いると、$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ と書けることが示せます。

多変量 OLS の漸近的性質

- 母集団の CEF に関する仮説検定を行うには、$\hat{\beta}$ の漸近分布を導出する必要があります。

多変量 OLS の漸近的性質

- 母集団の CEF に関する仮説検定を行うには、$\hat{\beta}$ の漸近分布を導出する必要があります。
- 単変量 OLS と同様に、$\hat{\beta}$ が一致性を持ち、漸近的に正規分布に従うことを示せます。

多変量 OLS の漸近的性質

- 母集団の CEF に関する仮説検定を行うには、$\hat{\beta}$ の漸近分布を導出する必要があります。
- 単変量 OLS と同様に、$\hat{\beta}$ が一致性を持ち、漸近的に正規分布に従うことを示せます。
- 証明は単変量の場合と非常に似ているため、ここでは結果のみを示します。

多変量 OLS の漸近的性質

- 母集団の CEF に関する仮説検定を行うには、$\hat{\boldsymbol{\beta}}$ の漸近分布を導出する必要があります。
- 単変量 OLS と同様に、$\hat{\boldsymbol{\beta}}$ が一致性を持ち、漸近的に正規分布に従うことを示せます。
- 証明は単変量の場合と非常に似ているため、ここでは結果のみを示します。
- 一致性：$\hat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}$。

多変量 OLS の漸近的性質

- 漸近正規性：

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}(0, \boldsymbol{\Sigma}),$$

ここで $\boldsymbol{\Sigma} = E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} Var(\boldsymbol{X}_i \varepsilon_i) E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1}$ であり、$\varepsilon_i = Y_i - \boldsymbol{X}_i' \boldsymbol{\beta}$ です。

多変量 OLS の漸近的性質

- 漸近正規性：
$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}(0, \boldsymbol{\Sigma}),$$
ここで $\boldsymbol{\Sigma} = E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} Var(\boldsymbol{X}_i \varepsilon_i) E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1}$ であり、$\varepsilon_i = Y_i - \boldsymbol{X}_i' \boldsymbol{\beta}$ です。

- $\boldsymbol{\Sigma}$ は母平均を標本平均に置き換えることで推定できます：
$$\hat{\boldsymbol{\Sigma}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \hat{\varepsilon}_i^2 \right) \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1}$$
ただし $\hat{\varepsilon}_i = Y_i - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}$ です。

多変量 OLS の漸近的性質

- 漸近正規性：
$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}(0, \boldsymbol{\Sigma}),$$
ここで $\boldsymbol{\Sigma} = E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1} Var(\boldsymbol{X}_i \varepsilon_i) E[\boldsymbol{X}_i \boldsymbol{X}_i']^{-1}$ であり、$\varepsilon_i = Y_i - \boldsymbol{X}_i' \boldsymbol{\beta}$ です。

- $\boldsymbol{\Sigma}$ は母平均を標本平均に置き換えることで推定できます：
$$\hat{\boldsymbol{\Sigma}} = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \hat{\varepsilon}_i^2 \right) \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \boldsymbol{X}_i' \right)^{-1}$$
ただし $\hat{\varepsilon}_i = Y_i - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}$ です。

- 注意：$\hat{\boldsymbol{\Sigma}}$ は行列です。
  - $\hat{\boldsymbol{\beta}}_j$ の標準誤差は $\sqrt{\hat{\boldsymbol{\Sigma}}_{jj}} / \sqrt{N}$ です。
  - 非対角要素は、$\hat{\boldsymbol{\beta}}_j$ と $\hat{\boldsymbol{\beta}}_k$ の間の共分散に対応します。

例：年齢別・対数賃金

変数	係数	標準誤差 (SE)
定数項 (β_0)	5.8591	0.1409
年齢 (β_1)	0.0403	0.0077
年齢の 2 乗 (β_2)	-0.0005	0.0001

- β_2 の信頼区間は何でしょうか？

例：年齢別・対数賃金

変数	係数	標準誤差 (SE)
定数項 (β_0)	5.8591	0.1409
年齢 (β_1)	0.0403	0.0077
年齢の 2 乗 (β_2)	-0.0005	0.0001

- β_2 の信頼区間は何でしょうか？

$$\hat{\beta}_2 \pm 1.96 \times SE_{\beta_2} =$$

例：年齢別・対数賃金

変数	係数	標準誤差 (SE)
定数項 (β_0)	5.8591	0.1409
年齢 (β_1)	0.0403	0.0077
年齢の 2 乗 (β_2)	-0.0005	0.0001

- β_2 の信頼区間は何でしょうか？

$$\hat{\beta}_2 \pm 1.96 \times SE_{\beta_2} = -0.0005 \pm 1.96 \times 0.0001$$

例：年齢別・対数賃金

変数	係数	標準誤差 (SE)
定数項 (β_0)	5.8591	0.1409
年齢 (β_1)	0.0403	0.0077
年齢の 2 乗 (β_2)	-0.0005	0.0001

- β_2 の信頼区間は何でしょうか？

$$\hat{\beta}_2 \pm 1.96 \times SE_{\beta_2} = -0.0005 \pm 1.96 \times 0.0001 = [-0.0007, -0.0003]$$

複数の変数をコントロールする

- 多変量 OLS を用いると、より柔軟な関数形式（例：2 次式）を許容するだけでなく、同時に複数の変数を条件とした CEF を近似することが可能になります。

複数の変数をコントロールする

- 多変量 OLS を用いると、より柔軟な関数形式（例：2 次式）を許容するだけでなく、同時に複数の変数を条件とした CEF を近似することが可能になります。

- 例：テキサス州の各郡における過去 3 回の回の大統領選挙（2012 年、2016 年、2020 年）の得票率データがあります。

複数の変数をコントロールする

- 多変量 OLS を用いると、より柔軟な関数形式（例：2 次式）を許容するだけでなく、同時に複数の変数を条件とした CEF を近似することが可能になります。

- 例：テキサス州の各郡における過去 3 回の回の大統領選挙（2012 年、2016 年、2020 年）の得票率データがあります。

- Y_i を 2020 年のバイデン得票率、X_{i1} を 2016 年のクリントン得票率、X_{i2} を 2012 年のオバマ得票率とします。

複数の変数をコントロールする

- 多変量 OLS を用いると、より柔軟な関数形式（例：2 次式）を許容するだけでなく、同時に複数の変数を条件とした CEF を近似することが可能になります。
- 例：テキサス州の各郡における過去 3 回の回の大統領選挙（2012 年、2016 年、2020 年）の得票率データがあります。
- Y_i を 2020 年のバイデン得票率、X_{i1} を 2016 年のクリントン得票率、X_{i2} を 2012 年のオバマ得票率とします。
- 以下の回帰を推定します：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

複数の変数をコントロールする

- 多変量 OLS を用いると、より柔軟な関数形式（例：2 次式）を許容するだけでなく、同時に複数の変数を条件とした CEF を近似することが可能になります。
- 例：テキサス州の各郡における過去 3 回の回の大統領選挙（2012 年、2016 年、2020 年）の得票率データがあります。
- Y_i を 2020 年のバイデン得票率、X_{i1} を 2016 年のクリントン得票率、X_{i2} を 2012 年のオバマ得票率とします。
- 以下の回帰を推定します：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

- これは以下を近似しています：

$$E[\text{バイデン得票率} | \text{クリントン得票率}, \text{オバマ得票率}] \approx \beta_0 + \beta_1 \times \text{クリントン得票率} + \beta_2 \times \text{オ}$$

OLS 推定値

変数	係数	標準誤差 (SE)
定数項	0.05	0.01
クリントン	1.39	0.13
オバマ	-0.56	0.13

OLS 推定値

変数	係数	標準誤差 (SE)
定数項	0.05	0.01
クリントン	1.39	0.13
オバマ	-0.56	0.13

- オバマが 50%、クリントンが 60%の票を得た郡におけるバイデンの予測得票率は？

OLS 推定値

変数	係数	標準誤差 (SE)
定数項	0.05	0.01
クリントン	1.39	0.13
オバマ	-0.56	0.13

- オバマが 50%、クリントンが 60%の票を得た郡におけるバイデンの予測得票率は？

$$\hat{\beta}_0 + \hat{\beta}_1 0.6 + \hat{\beta}_2 0.5 =$$

OLS 推定値

変数	係数	標準誤差 (SE)
定数項	0.05	0.01
クリントン	1.39	0.13
オバマ	-0.56	0.13

- オバマが 50%、クリントンが 60%の票を得た郡におけるバイデンの予測得票率は？

$$\hat{\beta}_0 + \hat{\beta}_1 0.6 + \hat{\beta}_2 0.5 = 0.05 + 1.39 \times 0.6 - 0.56 \times 0.5 =$$

OLS 推定値

変数	係数	標準誤差 (SE)
定数項	0.05	0.01
クリントン	1.39	0.13
オバマ	-0.56	0.13

- オバマが 50%、クリントンが 60%の票を得た郡におけるバイデンの予測得票率は？

$$\hat{\beta}_0 + \hat{\beta}_1 0.6 + \hat{\beta}_2 0.5 = 0.05 + 1.39 \times 0.6 - 0.56 \times 0.5 = 0.604$$

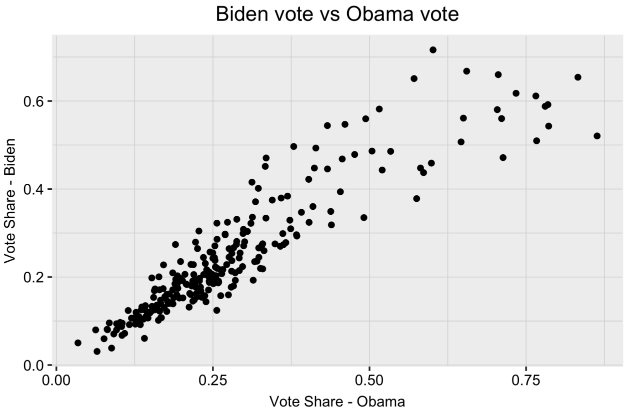
OLS 推定値

変数	係数	標準誤差 (SE)
定数項	0.05	0.01
クリントン	1.39	0.13
オバマ	-0.56	0.13

- オバマが 50%、クリントンが 60%の票を得た郡におけるバイデンの予測得票率は？

$$\hat{\beta}_0 + \hat{\beta}_1 0.6 + \hat{\beta}_2 0.5 = 0.05 + 1.39 \times 0.6 - 0.56 \times 0.5 = 0.604$$

- オバマ得票率の係数が負であることに注目してください。
- これは、オバマが健闘した場所ほどバイデンが苦戦したということでしょうか？！



- データを見ると、バイデンの得票率とオバマの得票率には非常に強い正の相関があります。
- では、何が起きているのでしょうか？！

回帰係数の解釈

- 多変量 OLS が CEF を次のように近似していることを思い出しましょう：

$$E[Y_i|\boldsymbol{X}_i = \boldsymbol{x}] \approx \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2$$

回帰係数の解釈

- 多変量 OLS が CEF を次のように近似していることを思い出しましょう：

$$E[Y_i|\boldsymbol{X}_i = \boldsymbol{x}] \approx \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2$$

- したがって、β_2 は偏微分の推定値です：

$$\frac{\partial}{\partial x_{i2}} E[Y_i|\boldsymbol{X}_i = \boldsymbol{x}] \approx \beta_2,$$

  すなわち、X_{i1} を一定に保ったまま X_{i2} を変化させたときの CEF の変化です。

回帰係数の解釈

- 多変量 OLS が CEF を次のように近似していることを思い出しましょう：

$$E[Y_i|\mathbf{X}_i = \mathbf{x}] \approx \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2$$

- したがって、β_2 は偏微分の推定値です：

$$\frac{\partial}{\partial x_{i2}} E[Y_i|\mathbf{X}_i = \mathbf{x}] \approx \beta_2,$$

  すなわち、X_{i1} を一定に保ったまま X_{i2} を変化させたときの CEF の変化です。

- もし $\beta_2 < 0$ ならば、クリントンの得票率が同じ郡の間では、オバマの得票率が低かった場所ほどバイデンの得票率が高かったことを意味します。

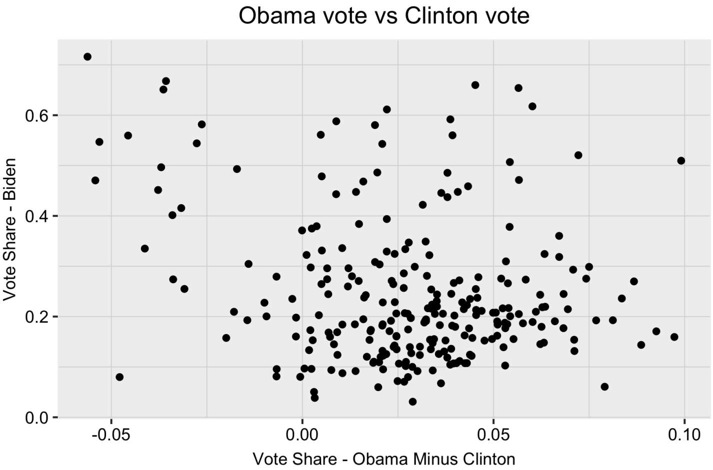
回帰係数の解釈

- 多変量 OLS が CEF を次のように近似していることを思い出しましょう：
$$E[Y_i|\boldsymbol{X}_i = \boldsymbol{x}] \approx \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2$$

- したがって、β_2 は偏微分の推定値です：
$$\frac{\partial}{\partial x_{i2}} E[Y_i|\boldsymbol{X}_i = \boldsymbol{x}] \approx \beta_2,$$
すなわち、X_{i1} を一定に保ったまま X_{i2} を変化させたときの CEF の変化です。

- もし $\beta_2 < 0$ ならば、クリントンの得票率が同じ郡の間では、オバマの得票率が低かった場所ほどバイデンの得票率が高かったことを意味します。

- 言い換えれば、2012 年から 2016 年の間に民主党の得票率が上昇した場所ほど、バイデンが健闘したということです。


Obama vote vs Clinton vote
0.6
0.4
0.2
0.0
Vote Share - Biden
-0.05
0.00
0.05
0.10
Vote Share - Obama Minus Clinton

このアイデアの一般化

- フリッシュ＝ウォー＝ロヴェル（**Frisch-Waugh-Lovell, FWL**）定理は、重回帰の係数を解釈する一般的な方法を与えてくれます。

このアイデアの一般化

- フリッシュ＝ウォー＝ロヴェル（Frisch-Waugh-Lovell, FWL）定理は、重回帰の係数を解釈する一般的な方法を与えてくれます。以下の回帰を考えます：

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \varepsilon_i$$

このアイデアの一般化

- フリッシュ＝ウォー＝ロヴェル（**Frisch-Waugh-Lovell, FWL**）定理は、重回帰の係数を解釈する一般的な方法を与えてくれます。以下の回帰を考えます：

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \varepsilon_i$$

- FWL 定理によれば、OLS 係数 $\hat{\beta}_2$ は以下の手順で得られます：

このアイデアの一般化

- フリッシュ＝ウォー＝ロヴェル（**Frisch-Waugh-Lovell, FWL**）定理は、重回帰の係数を解釈する一般的な方法を与えてくれます。以下の回帰を考えます：

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \varepsilon_i$$

- FWL 定理によれば、OLS 係数 $\hat{\beta}_2$ は以下の手順で得られます：

- 1) X_{i2} を X_{i1} と定数項に回帰する：

$$X_{i2} = \gamma_0 + X_{i1}\gamma_1 + u_i$$

このアイデアの一般化

- フリッシュ＝ウォー＝ロヴェル（**Frisch-Waugh-Lovell, FWL**）定理は、重回帰の係数を解釈する一般的な方法を与えてくれます。以下の回帰を考えます：

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \varepsilon_i$$

- FWL 定理によれば、OLS 係数 $\hat{\beta}_2$ は以下の手順で得られます：
- 1) X_{i2} を X_{i1} と定数項に回帰する：

$$X_{i2} = \gamma_0 + X_{i1}\gamma_1 + u_i$$

- 2) 各ユニットについて、ステップ 1) で得られた係数を用いて X_{i2} を予測する：

$$\hat{X}_{i2} = \hat{\gamma}_0 + X_{i1}\hat{\gamma}_1$$

このアイデアの一般化

- **フリッシュ＝ウォー＝ロヴェル（Frisch-Waugh-Lovell, FWL）定理は、重回帰の係数を解釈する一般的な方法を与えてくれます。以下の回帰を考えます：**

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \varepsilon_i$$

- FWL 定理によれば、OLS 係数 $\hat{\beta}_2$ は以下の手順で得られます：
- 1) X_{i2} を X_{i1} と定数項に回帰する：

$$X_{i2} = \gamma_0 + X_{i1}\gamma_1 + u_i$$

- 2) 各ユニットについて、ステップ 1) で得られた係数を用いて X_{i2} を予測する：

$$\hat{X}_{i2} = \hat{\gamma}_0 + X_{i1}\hat{\gamma}_1$$

- 3) Y_i を OLS 残差 $X_{i2} - \hat{X}_{i2}$ に回帰して、$\hat{\beta}_2$ を得る：

$$Y_i = \alpha + (X_{i2} - \hat{X}_{i2})\beta_2 + v_i$$

選挙データを用いたイラスト

- オバマ得票率をクリントン得票率に回帰します：

切片 ($\hat{\gamma}_0$)	0.03
クリントン ($\hat{\gamma}_1$)	0.98

選挙データを用いたイラスト

- オバマ得票率をクリントン得票率に回帰します：

切片 ($\hat{\gamma}_0$)	0.03
クリントン ($\hat{\gamma}_1$)	0.98

- クリントン得票率を用いてオバマ得票率を予測します：

$$\hat{X}_{i2} =$$

選挙データを用いたイラスト

- オバマ得票率をクリントン得票率に回帰します：

切片 ($\hat{\gamma}_0$)	0.03
クリントン ($\hat{\gamma}_1$)	0.98

- クリントン得票率を用いてオバマ得票率を予測します：

$$\hat{X}_{i2} = 0.03 + 0.98 X_{i1}$$

選挙データを用いたイラスト

- オバマ得票率をクリントン得票率に回帰します：

切片 ($\hat{\gamma}_0$)	0.03
クリントン ($\hat{\gamma}_1$)	0.98

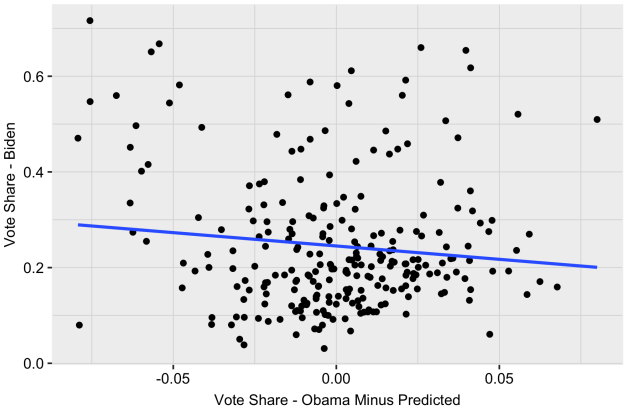
- クリントン得票率を用いてオバマ得票率を予測します：

$$\hat{X}_{i2} = 0.03 + 0.98 X_{i1}$$

- バイデン得票率を $X_{i2} - \hat{X}_{i2}$ に回帰します：

切片 ($\hat{\alpha}$)	0.25
オバマ（予測値除去後）($\hat{\beta}_2$)	-0.56

- 推定値 $\hat{\beta}_2 = -0.56$ は、以前得たものと全く同じです！



- 回帰直線の傾きはまさに $\hat{\beta}_2 = -0.56$ となります。
- FWL 定理は、重回帰係数を可視化・解釈するための容易な方法を提供します。

モデルの適合度の尺度

- 2 次項を加えることで、賃金と年齢の CEF の近似は改善したでしょうか？ これをどのように測定すればよいでしょうか？

モデルの適合度の尺度

- 2 次項を加えることで、賃金と年齢の CEF の近似は改善したでしょうか？　これをどのように測定すればよいでしょうか？

- 適合度を測る 1 つの方法は母集団 R^2 です。回帰 $Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + \varepsilon_i$ について：

$$R^2 = \frac{Var(\boldsymbol{X}_i'\boldsymbol{\beta})}{Var(Y_i)}$$

モデルの適合度の尺度

- 2 次項を加えることで、賃金と年齢の CEF の近似は改善したでしょうか？ これをどのように測定すればよいでしょうか？

- 適合度を測る 1 つの方法は母集団 R^2 です。回帰 $Y_i = \boldsymbol{X}_i' \boldsymbol{\beta} + \varepsilon_i$ について：

$$R^2 = \frac{Var(\boldsymbol{X}_i' \boldsymbol{\beta})}{Var(Y_i)}$$

- 直感的には、母集団 R^2 は Y_i の分散のうち、$\boldsymbol{X}_i' \beta$ によって説明される割合を測定します。
  - $Cov(\mathbf{X}_i' \boldsymbol{\beta}, e_i) = 0$ なので、$R^2 = 1 - \frac{Var(e_i)}{Var(Y_i)}$ とも書けます。

モデルの適合度の尺度

- 2 次項を加えることで、賃金と年齢の CEF の近似は改善したでしょうか？ これをどのように測定すればよいでしょうか？

- 適合度を測る 1 つの方法は母集団 R^2 です。回帰 $Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + \varepsilon_i$ について：

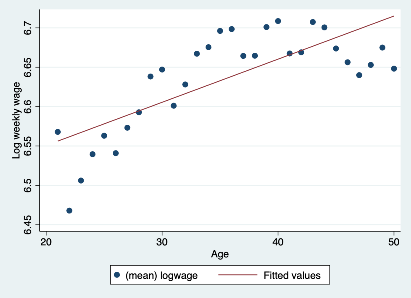
$$R^2 = \frac{Var(\boldsymbol{X}_i'\boldsymbol{\beta})}{Var(Y_i)}$$

- 直感的には、母集団 R^2 は Y_i の分散のうち、$\boldsymbol{X}_i'\beta$ によって説明される割合を測定します。
  - $Cov(\mathbf{X}_i'\boldsymbol{\beta}, e_i) = 0$ なので、$R^2 = 1 - \frac{Var(e_i)}{Var(Y_i)}$ とも書けます。

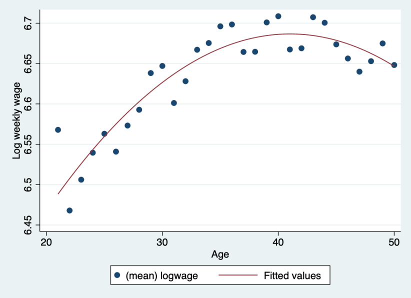
- R^2 を推定するには、母集団の値を標本対応物に置き換えます：

$$\hat{R}^2 = \frac{\frac{1}{N}\sum_i(\mathbf{X}_i'\hat{\boldsymbol{\beta}} - \bar{\mathbf{X}}_i'\hat{\boldsymbol{\beta}})^2}{\frac{1}{N}\sum_i(Y_i - \bar{Y})^2} = 1 - \frac{\frac{1}{N}\sum_i\hat{\varepsilon}_i^2}{\frac{1}{N}\sum_i(Y_i - \bar{Y})^2}$$

賃金・年齢の例における R^2



(a) $\hat{R}^2 = 0.44$



(b) $\hat{R}^2 = 0.73$

- 線形フィットは年齢による平均所得の変動の 44% を説明しますが、2 次式フィットは 73% を説明します。

$\hat{R}^2$ に関する注意

- 注意：より複雑なモデルにすると、標本の $\hat{R}^2$ は必ず上昇します。なぜでしょうか？

$\hat{R}^2$ に関する注意

- 注意：より複雑なモデルにすると、標本の $\hat{R}^2$ は必ず上昇します。なぜでしょうか？

- 線形フィットの係数は、以下を最小化します：

$$\frac{1}{N}\sum_i (\underbrace{Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)}_{\hat{\varepsilon}_{\text{線形}}})^2$$

一方、2 次式フィットの係数は以下を最小化します：

$$\frac{1}{N}\sum_i (\underbrace{Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2)}_{\hat{\varepsilon}_{2\text{次}}})^2$$

2 次式の最小化において $\hat{\beta}_2 = 0$ とすることも可能なので、最小化された値は 2 次式の方が必ず（弱く）小さくなります。

$\hat{R}^2$ に関する注意

- 注意：より複雑なモデルにすると、標本の $\hat{R}^2$ は必ず上昇します。なぜでしょうか？

- 線形フィットの係数は、以下を最小化します：

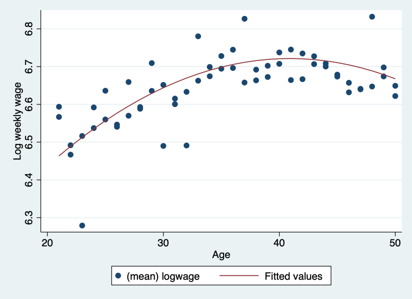
$$\frac{1}{N}\sum_i \underbrace{(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))}_{\hat{\varepsilon}_{\text{線形}}}{}^2$$

一方、2 次式フィットの係数は以下を最小化します：

$$\frac{1}{N}\sum_i \underbrace{(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2))}_{\hat{\varepsilon}_{2\text{次}}}{}^2$$

2 次式の最小化において $\hat{\beta}_2 = 0$ とすることも可能なので、最小化された値は 2 次式の方が必ず（弱く）小さくなります。

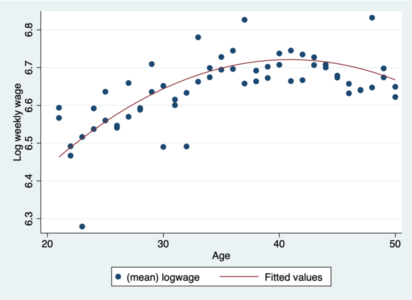
- しかし、より複雑なモデルが常に優れていると言えるでしょうか？



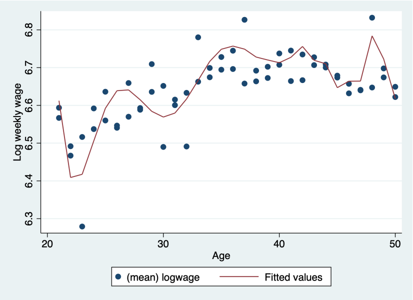
(c) 2 次式, $\hat{R}^2 = 0.44$

(d) 20 次多項式, $\hat{R}^2 = 0.70$

- サイズ 10,000 のサンプルをとり、2 次式と 20 次多項式をフィットさせたとします。

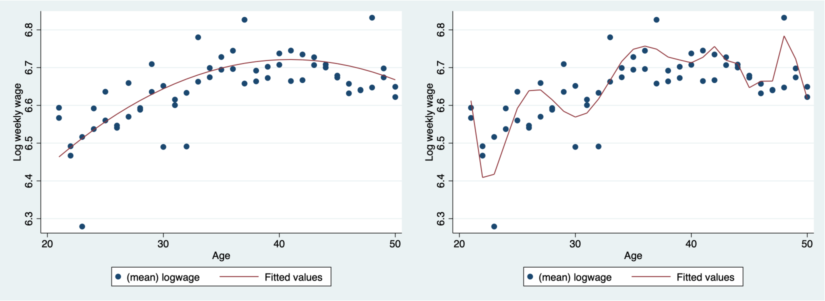


(e) 2 次式, $\hat{R}^2 = 0.44$



(f) 20 次多項式, $\hat{R}^2 = 0.70$

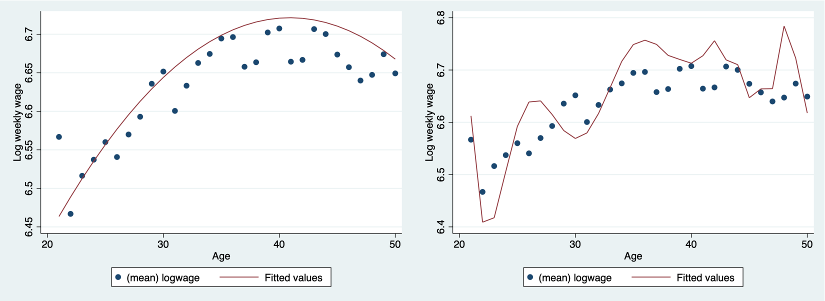
- サイズ 10,000 のサンプルをとり、2 次式と 20 次多項式をフィットさせたとします。
- 20 次多項式の方が R^2 は高いですが、これは妥当に見えますか？



(g) 2 次式, $\hat{R}^2 = 0.44$ (h) 20 次多項式, $\hat{R}^2 = 0.70$

- サイズ 10,000 のサンプルをとり、2 次式と 20 次多項式をフィットさせたとします。
- 20 次多項式の方が R^2 は高いですが、これは妥当に見えますか？
- いいえ、あまりに「ギザギザ」すぎます。サンプルの特定の点に過剰に適合してしまっています。

- 新しいサンプルを抽出し、最初のデータセットで学習したモデルの予測精度をテストしてみます。



(i) 2 次式　　(j) 20 次多項式

- 新しいサンプルを抽出し、最初のデータセットで学習したモデルの予測精度をテストしてみます。
- 2 次式のフィットは、新しいデータに対してもよく汎化（適用）できています。
- しかし、20 次多項式の成績は非常に悪いです。特定のサンプルの特徴に「過学習（overfit）」してしまい、新しいサンプルには汎化できないのです。

どのように過学習を避けるか？

- 課題は、CEF の主要な特徴を捉えるのに十分なほど豊かでありながら、過学習しない程度に単純なモデルを選ぶことです。

どのように過学習を避けるか？

- 課題は、CEF の主要な特徴を捉えるのに十分なほど豊かでありながら、過学習しない程度に単純なモデルを選ぶことです。
- これを助けるためのいくつかのツールがありますが、どれも完璧ではありません。

どのように過学習を避けるか？

- 課題は、CEF の主要な特徴を捉えるのに十分なほど豊かでありながら、過学習しない程度に単純なモデルを選ぶことです。
- これを助けるためのいくつかのツールがありますが、どれも完璧ではありません。
- 自由度修正済み $\hat{R}^2$（Adjusted R^2）は、変数の数に応じてペナルティを加える修正です。
  - 通常の $\hat{R}^2$ よりは一般に優れていますが、これが最大のモデルがベストであるとは限りません。

どのように過学習を避けるか？

- 課題は、CEF の主要な特徴を捉えるのに十分なほど豊かでありながら、過学習しない程度に単純なモデルを選ぶことです。

- これを助けるためのいくつかのツールがありますが、どれも完璧ではありません。

- 自由度修正済み $\hat{R}^2$（Adjusted R^2）は、変数の数に応じてペナルティを加える修正です。
  - 通常の $\hat{R}^2$ よりは一般に優れていますが、これが最大のモデルがベストであるとは限りません。

- 交差検証（Cross validation）：「標本外」での予測精度に基づいてモデルの複雑さを選びます。

どのように過学習を避けるか？

- 課題は、CEF の主要な特徴を捉えるのに十分なほど豊かでありながら、過学習しない程度に単純なモデルを選ぶことです。
- これを助けるためのいくつかのツールがありますが、どれも完璧ではありません。
- 自由度修正済み $\hat{R}^2$（Adjusted R^2）は、変数の数に応じてペナルティを加える修正です。
  - 通常の $\hat{R}^2$ よりは一般に優れていますが、これが最大のモデルがベストであるとは限りません。
- 交差検証（Cross validation）：「標本外」での予測精度に基づいてモデルの複雑さを選びます。
  - データを 2 つに分け、半分でモデルを学習させ、もう半分で予測の良さを確認します。
  - 標本外の予測が最も良くなるようなモデルの複雑さを選びます。

どのように過学習を避けるか？

- 課題は、CEF の主要な特徴を捉えるのに十分なほど豊かでありながら、過学習しない程度に単純なモデルを選ぶことです。
- これを助けるためのいくつかのツールがありますが、どれも完璧ではありません。
- 自由度修正済み $\hat{R}^2$（Adjusted R^2）は、変数の数に応じてペナルティを加える修正です。
  - 通常の $\hat{R}^2$ よりは一般に優れていますが、これが最大のモデルがベストであるとは限りません。
- 交差検証（Cross validation）：「標本外」での予測精度に基づいてモデルの複雑さを選びます。
  - データを 2 つに分け、半分でモデルを学習させ、もう半分で予測の良さを確認します。
  - 標本外の予測が最も良くなるようなモデルの複雑さを選びます。
  - 交差検証は現代の機械学習（ML）手法の基礎となっています。ML は非常に強力であり、最近の研究では因果推論の問題にも ML が拡張されています（が、この講義の範囲外です）。

実務における過学習の回避

- 上記のツールはモデル選択に有用ですが、実務ではモデル選択はより経験的（heuristic）に行われることが多いです。

実務における過学習の回避

- 上記のツールはモデル選択に有用ですが、実務ではモデル選択はより経験的（heuristic）に行われることが多いです。
- 研究者は通常、最も重要だと考える変数を含む単純なモデル（例：線形や 2 次式）から始めます。

実務における過学習の回避

- 上記のツールはモデル選択に有用ですが、実務ではモデル選択はより経験的（heuristic）に行われることが多いです。
- 研究者は通常、最も重要だと考える変数を含む単純なモデル（例：線形や 2 次式）から始めます。
- そして、変数を加えたり減らしたり、高次の項を加えたりして、結論が「頑健（robust)」であるかを評価します。

実務における過学習の回避

- 上記のツールはモデル選択に有用ですが、実務ではモデル選択はより経験的（heuristic）に行われることが多いです。
- 研究者は通常、最も重要だと考える変数を含む単純なモデル（例：線形や 2 次式）から始めます。
- そして、変数を加えたり減らしたり、高次の項を加えたりして、結論が「頑健（robust）」であるかを評価します。
- 一般に、モデルの仕様を少し変えても結論が変わらなければ、その結論に対する信頼が高まります。

アウトライン

1. 重回帰と OLS の導出 ✓
2. 回帰と因果関係
3. 回帰に関するその他のトピック

回帰と因果関係の出会い

- 重回帰は、条件付き非交絡性の下で因果的効果を推定するためによく用いられます。

回帰と因果関係の出会い

- 重回帰は、条件付き非交絡性の下で因果的効果を推定するためによく用いられます。

- 条件付き非交絡性の下では、以下が成り立つことを思い出してください：

$$CATE(\boldsymbol{x}) = E[Y_i|D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$$

回帰と因果関係の出会い

- 重回帰は、条件付き非交絡性の下で因果的効果を推定するためによく用いられます。
- 条件付き非交絡性の下では、以下が成り立つことを思い出してください：

$$CATE(\boldsymbol{x}) = E[Y_i|D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$$

- CEF を以下のように線形近似するのが一般的です：

$$E[Y_i|D_i, \boldsymbol{X}_i] \approx D_i\beta + \boldsymbol{X}_i'\boldsymbol{\gamma}$$

回帰と因果関係の出会い

- 重回帰は、条件付き非交絡性の下で因果的効果を推定するためによく用いられます。
- 条件付き非交絡性の下では、以下が成り立つことを思い出してください：

$$CATE(\boldsymbol{x}) = E[Y_i|D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$$

- CEF を以下のように線形近似するのが一般的です：

$$E[Y_i|D_i, \boldsymbol{X}_i] \approx D_i\beta + \boldsymbol{X}_i'\boldsymbol{\gamma}$$

- すると条件付き非交絡性は $CATE(\boldsymbol{x}) \approx \beta$ を意味します。
  - これは x に依存しないため、$\beta \approx ATE$ でも期待できます。

回帰と因果関係の出会い

- 重回帰は、条件付き非交絡性の下で因果的効果を推定するためによく用いられます。
- 条件付き非交絡性の下では、以下が成り立つことを思い出してください：

$$CATE(\boldsymbol{x}) = E[Y_i|D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$$

- CEF を以下のように線形近似するのが一般的です：

$$E[Y_i|D_i, \boldsymbol{X}_i] \approx D_i\beta + \boldsymbol{X}_i'\boldsymbol{\gamma}$$

- すると条件付き非交絡性は $CATE(\boldsymbol{x}) \approx \beta$ を意味します。
  - これは x に依存しないため、$\beta \approx ATE$ でも期待できます。
- したがって、重回帰

$$Y_i = D_i\beta + \boldsymbol{X}_i'\boldsymbol{\gamma} + \varepsilon_i$$

を推定すれば、$\hat{\boldsymbol{\beta}}$ を ATE の推定値として解釈できます。

Dale and Krueger

- Dale & Krueger は（覚えている通り）、より選択度の高い大学に通うことが所得に与える影響に興味を持っています。
    - 彼らは「College and Beyond (C&B)」調査から、所得と大学への出願情報を得ました。

Dale and Krueger

- Dale & Krueger は（覚えている通り）、より選択度の高い大学に通うことが所得に与える影響に興味を持っています。
    - 彼らは「College and Beyond (C&B)」調査から、所得と大学への出願情報を得ました。
- C&B 調査は、1978 年に高校を卒業した 30 の大学の学生をカバーしており、重要な変数を含んでいます：

Dale and Krueger

- Dale & Krueger は（覚えている通り）、より選択度の高い大学に通うことが所得に与える影響に興味を持っています。
  - 彼らは「College and Beyond (C&B)」調査から、所得と大学への出願情報を得ました。

- C&B 調査は、1978 年に高校を卒業した 30 の大学の学生をカバーしており、重要な変数を含んでいます：

- 1996 年時点の所得

- SAT スコア、クラス順位、世帯所得、人種などの大学出願時および人口統計学的変数

- 大学への出願結果 — すなわち、どの大学に出願し、どこに合格したかのセット

Institution	School-average SAT score in 1978
Barnard College	1210
Bryn Mawr College	1370
Columbia University	1330
Denison University	1020
Duke University	1226
Emory University	1150
Georgetown University	1225
Hamilton College	1246
Kenyon College	1155
Miami University (Ohio)	1073
Northwestern University	1240
Oberlin College	1227
Pennsylvania State University	1038
Princeton University	1308
Rice University	1316
Smith College	1210
Stanford University	1270
Swarthmore College	1340
Tufts University	1200
Tulane University	1080
University of Michigan (Ann Arbor)	1110
University of North Carolina (Chapel Hill)	1080
University of Notre Dame	1200
University of Pennsylvania	1280
Vanderbilt University	1162
Washington University	1180
Wellesley College	1220
Wesleyan University	1260
Williams College	1255
Yale University	1360

選択バイアスへの対処

- Dale & Krueger は条件付き非交絡性、すなわち $D_i \perp\!\!\!\perp (Y_i(\cdot))|X_i$ を仮定しています。ここで D_i は進学した大学の平均 SAT スコア、X_i はコントロール変数のセットです。

選択バイアスへの対処

- Dale & Krueger は条件付き非交絡性、すなわち $D_i \perp\!\!\!\perp (Y_i(\cdot))|X_i$ を仮定しています。ここで D_i は進学した大学の平均 SAT スコア、X_i はコントロール変数のセットです。

- 彼らは以下の形式の回帰を推定します：

$$ln(Y_i) = D_i\beta + \boldsymbol{X}_i'\boldsymbol{\gamma} + \varepsilon_i$$

ここで Y_i は 1996 年の所得です。

選択バイアスへの対処

- Dale & Krueger は条件付き非交絡性、すなわち $D_i \perp\!\!\!\perp (Y_i(\cdot))|X_i$ を仮定しています。ここで D_i は進学した大学の平均 SAT スコア、X_i はコントロール変数のセットです。

- 彼らは以下の形式の回帰を推定します：

$$ln(Y_i) = D_i\beta + \boldsymbol{X}_i'\boldsymbol{\gamma} + \varepsilon_i$$

  ここで Y_i は 1996 年の所得です。

- 条件付き非交絡性が成り立ち、回帰による CEF の近似が適切であれば、β は平均 SAT スコアの高い大学に通うことの（近似的な）因果的効果となります。

最初のステップ：SAT スコアと属性を X_i に含める

Variable	Full sample 1
School-average SAT score/100	0.076 (0.016)
Predicted log(parental income)	0.187 (0.024)
Own SAT score/100	0.018 (0.006)
Female	−0.403 (0.015)
Black	−0.023 (0.035)
Hispanic	0.015 (0.052)
Asian	0.173 (0.036)
Other/missing race	−0.188 (0.119)
High school top 10 percent	0.061 (0.018)
High school rank missing	0.001 (0.024)
Athlete	0.102 (0.025)

- $\hat{\beta} = 0.076$ は、平均 SAT が 100 ポイント高い大学に通うことで、対数賃金が約 7.6% 上昇することを示しています。

この推定値を信じますか？

- これらのコントロール変数を用いても、なぜ非交絡性が満たされない可能性があるのでしょうか？

この推定値を信じますか？

- これらのコントロール変数を用いても、なぜ非交絡性が満たされない可能性があるのでしょうか？
  - どの大学に合格するかは、観察されない重要な要因に依存するかもしれません。例えば、より良い出願エッセイを書いた学生は、より多くの大学に合格し、どこに進学したかに関わらずより多く稼ぐかもしれません。

この推定値を信じますか？

- これらのコントロール変数を用いても、なぜ非交絡性が満たされない可能性があるのでしょうか？
  - どの大学に合格するかは、観察されない重要な要因に依存するかもしれません。例えば、より良い出願エッセイを書いた学生は、より多くの大学に合格し、どこに進学したかに関わらずより多く稼ぐかもしれません。
- この懸念に対処するため、Dale and Krueger は第 2 の分析を行いました：学生が出願し合格した大学のセットをコントロールします。

この推定値を信じますか？

- これらのコントロール変数を用いても、なぜ非交絡性が満たされない可能性があるのでしょうか？
  - どの大学に合格するかは、観察されない重要な要因に依存するかもしれません。例えば、より良い出願エッセイを書いた学生は、より多くの大学に合格し、どこに進学したかに関わらずより多く稼ぐかもしれません。
- この懸念に対処するため、Dale and Krueger は第 2 の分析を行いました：学生が出願し合格した大学のセットをコントロールします。
  - C&B データがあるからこそ、これが可能になりました。

「固定効果」の導入

- 「出願し合格した大学のセットをコントロールする」とは、具体的にどういうことでしょうか？

「固定効果」の導入

- 「出願し合格した大学のセットをコントロールする」とは、具体的にどういうことでしょうか？

- ブラウン、エール、URI の 3 校があり、全員がこの 3 校に出願したという簡略化されたケースを考えましょう。

「固定効果」の導入

- 「出願し合格した大学のセットをコントロールする」とは、具体的にどういうことでしょうか？
- ブラウン、エール、URI の 3 校があり、全員がこの 3 校に出願したという簡略化されたケースを考えましょう。
- サンプル内の学生は全員 URI に合格しましたが、ブラウンとエールの合否は異なるとします。4 つの可能性があります：

「固定効果」の導入

- 「出願し合格した大学のセットをコントロールする」とは、具体的にどういうことでしょうか？
- ブラウン、エール、URI の 3 校があり、全員がこの 3 校に出願したという簡略化されたケースを考えましょう。
- サンプル内の学生は全員 URI に合格しましたが、ブラウンとエールの合否は異なるとします。4 つの可能性があります：
  ケース 1：全校に合格
  ケース 2：URI とブラウンのみ合格
  ケース 3：URI とエールのみ合格
  ケース 4：URI のみ合格

「固定効果」の導入

- 「出願し合格した大学のセットをコントロールする」とは、具体的にどういうことでしょうか？
- ブラウン、エール、URI の 3 校があり、全員がこの 3 校に出願したという簡略化されたケースを考えましょう。
- サンプル内の学生は全員 URI に合格しましたが、ブラウンとエールの合否は異なるとします。4 つの可能性があります：
  ケース 1：全校に合格
  ケース 2：URI とブラウンのみ合格
  ケース 3：URI とエールのみ合格
  ケース 4：URI のみ合格
- ケース 1 であれば $X_{i1} = 1$、そうでなければ 0 となるような変数を作成できます。X_{i2} から X_{i4} も同様です。

「固定効果」の導入

- 「出願し合格した大学のセットをコントロールする」とは、具体的にどういうことでしょうか？

- ブラウン、エール、URI の 3 校があり、全員がこの 3 校に出願したという簡略化されたケースを考えましょう。

- サンプル内の学生は全員 URI に合格しましたが、ブラウンとエールの合否は異なるとします。4 つの可能性があります：
  ケース 1：全校に合格
  ケース 2：URI とブラウンのみ合格
  ケース 3：URI とエールのみ合格
  ケース 4：URI のみ合格

- ケース 1 であれば $X_{i1} = 1$、そうでなければ 0 となるような変数を作成できます。X_{i2} から X_{i4} も同様です。

- これらの変数 $X_{i1}, \ldots, X_{i4}$ は、しばしば合格した学校のセットに対する「固定効果（fixed effects）」と呼ばれます。

「固定効果」の導入

- すると、CEF を以下のように近似できます：

$$E[Y_i|\boldsymbol{X}_i, D_i] = D_i\beta_D + X_{i1}\beta_1 + X_{i2}\beta_2 + X_{i3}\beta_3 + X_{i4}\beta_4$$

「固定効果」の導入

- すると、CEF を以下のように近似できます：

$$E[Y_i|\boldsymbol{X}_i, D_i] = D_i\beta_D + X_{i1}\beta_1 + X_{i2}\beta_2 + X_{i3}\beta_3 + X_{i4}\beta_4$$

- これにより、どの大学に合格したかによって平均的な結果が異なることを許容しつつ、その中での大学の選択度 (D_i) の影響を捉えられます。

「固定効果」の導入

- すると、CEF を以下のように近似できます：

$$E[Y_i|\boldsymbol{X}_i, D_i] = D_i\beta_D + X_{i1}\beta_1 + X_{i2}\beta_2 + X_{i3}\beta_3 + X_{i4}\beta_4$$

- これにより、どの大学に合格したかによって平均的な結果が異なることを許容しつつ、その中での大学の選択度 (D_i) の影響を捉えられます。

- 直感的には、β_D は、同じ大学のセットに合格した学生たちの間での、エリート校進学による平均的な差を表しています。

		Private			Public			
Applicant Group	Student	School I	School II	School III	School IV	School V	School VI	1996 Earnings
	1		Reject	**Admit**		Admit		110,000
A	2		Reject	**Admit**		Admit		100,000
	3		Reject	Admit		**Admit**		110,000
B	4	**Admit**			Admit		Admit	60,000
	5	Admit			Admit		**Admit**	30,000
C	6		**Admit**					115,000
	7		**Admit**					75,000
D	8	Reject			**Admit**	Admit		90,000
	9	Reject			Admit	**Admit**		60,000

- 「出願者グループ（Applicant groups）」は全員、同じセットの大学に出願し、合格しました。
- 実際に進学した大学がハイライトされています。

実務における出願結果のコントロール

- 実際には、Dale and Krueger のデータには、全く同じ大学のセットに出願し合格した学生はそれほど多くありません。

実務における出願結果のコントロール

- 実際には、Dale and Krueger のデータには、全く同じ大学のセットに出願し合格した学生はそれほど多くありません。
- 近似として、彼らは大学を平均 SAT スコアに基づいて 25 ポイント刻みのビンにグループ分けしました。

実務における出願結果のコントロール

- 実際には、Dale and Krueger のデータには、全く同じ大学のセットに出願し合格した学生はそれほど多くありません。
- 近似として、彼らは大学を平均 SAT スコアに基づいて 25 ポイント刻みのビンにグループ分けしました。
- そして、このビンに基づいた出願・合格のセットをコントロールしました（例：X_{i1} は、SAT 1350 の大学に不合格になり、SAT 1250 の大学 2 校に合格したことに対応するかもしれません）。

Variable	Basic model: no selection controls		Matched-applicant model
	Full sample	Restricted sample	Similar school-SAT matches*
	1	2	3
School-average SAT score/100	0.076	0.082	$\boxed{-0.016}$
	(0.016)	(0.014)	$\boxed{(0.022)}$
Predicted log(parental income)	0.187	0.190	0.163
	(0.024)	(0.033)	(0.033)
Own SAT score/100	0.018	0.006	−0.011
	(0.006)	(0.007)	(0.007)
Female	−0.403	−0.410	−0.395
	(0.015)	(0.018)	(0.024)
Black	−0.023	−0.026	−0.057
	(0.035)	(0.053)	(0.053)
Hispanic	0.015	0.070	0.020
	(0.052)	(0.076)	(0.099)
Asian	0.173	0.245	0.241
	(0.036)	(0.054)	(0.064)
Other/missing race	−0.188	−0.048	0.060
	(0.119)	(0.143)	(0.180)
High school top 10 percent	0.061	0.091	0.079
	(0.018)	(0.022)	(0.026)
High school rank missing	0.001	0.040	0.016
	(0.024)	(0.026)	(0.038)
Athlete	0.102	0.088	0.104
	(0.025)	(0.030)	(0.039)

- 出願コントロールを加えると、$\hat{\beta} = -0.016$ となります。

Variable	Basic model: no selection controls		Matched-applicant model
	Full sample	Restricted sample	Similar school-SAT matches*
	1	2	3
School-average SAT score/100	0.076	0.082	−0.016
	(0.016)	(0.014)	(0.022)
Predicted log(parental income)	0.187	0.190	0.163
	(0.024)	(0.033)	(0.033)
Own SAT score/100	0.018	0.006	−0.011
	(0.006)	(0.007)	(0.007)
Female	−0.403	−0.410	−0.395
	(0.015)	(0.018)	(0.024)
Black	−0.023	−0.026	−0.057
	(0.035)	(0.053)	(0.053)
Hispanic	0.015	0.070	0.020
	(0.052)	(0.076)	(0.099)
Asian	0.173	0.245	0.241
	(0.036)	(0.054)	(0.064)
Other/missing race	−0.188	−0.048	0.060
	(0.119)	(0.143)	(0.180)
High school top 10 percent	0.061	0.091	0.079
	(0.018)	(0.022)	(0.026)
High school rank missing	0.001	0.040	0.016
	(0.024)	(0.026)	(0.038)
Athlete	0.102	0.088	0.104
	(0.025)	(0.030)	(0.039)

- 出願コントロールを加えると、$\hat{\beta} = -0.016$ となります。
- これは、平均 SAT が 100 ポイント高い大学に通うことによる対数賃金の収益が約 -1.6% であることを示していますが、有意ではありま

条件付き非交絡性の再評価

- ここでも、なぜ条件付き非交絡性が満たされない可能性があるのでしょうか？

条件付き非交絡性の再評価

- ここでも、なぜ条件付き非交絡性が満たされない可能性があるのでしょうか？
- 懸念されるのは、出願結果を条件とした後でも、大学の選択が所得と相関する未観察要因に依存していることです。
    - 例えば、より選択度の高い大学に進む学生は、異なるキャリア志向（例：実業界かアカデミアか）を持っているかもしれません。

条件付き非交絡性の再評価

- ここでも、なぜ条件付き非交絡性が満たされない可能性があるのでしょうか？
- 懸念されるのは、出願結果を条件とした後でも、大学の選択が所得と相関する未観察要因に依存していることです。
    - 例えば、より選択度の高い大学に進む学生は、異なるキャリア志向（例：実業界かアカデミアか）を持っているかもしれません。
    - ランクの低い大学に進むのは、そこが特定の分野で非常に優れたプログラムを持っている場合だけかもしれません。

条件付き非交絡性の再評価

- ここでも、なぜ条件付き非交絡性が満たされない可能性があるのでしょうか？
- 懸念されるのは、出願結果を条件とした後でも、大学の選択が所得と相関する未観察要因に依存していることです。
  - 例えば、より選択度の高い大学に進む学生は、異なるキャリア志向（例：実業界かアカデミアか）を持っているかもしれません。
  - ランクの低い大学に進むのは、そこが特定の分野で非常に優れたプログラムを持っている場合だけかもしれません。
- C&B 調査の対象校はもともと選択度の高い大学ばかりです。この結果は「選択度の高い大学」と「非常に選択度の高い大学」の間の効果として解釈すべきものです。

欠落変数バイアス

- C&B の例では、条件付き非交絡性が成り立つために必要なすべての変数をコントロールできていないのではないか、という不安が残ります。

欠落変数バイアス

- C&B の例では、条件付き非交絡性が成り立つために必要なすべての変数をコントロールできていないのではないか、という不安が残ります。
- いくつかの変数を含めるのを忘れた場合、推定される係数はどのようにバイアスを受けるのでしょうか？

欠落変数バイアス

- C&B の例では、条件付き非交絡性が成り立つために必要なすべての変数をコントロールできていないのではないか、という不安が残ります。
- いくつかの変数を含めるのを忘れた場合、推定される係数はどのようにバイアスを受けるのでしょうか？
- この疑問に答えるために、「欠落変数バイアス（OVB）」の公式を導出します。

欠落変数バイアス – 最も単純なケース

- X_{i1} を条件として非交絡性が成り立つとし、以下を推定したいとします：

$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + e_i$$

  β_D は（CEF の近似が良ければ）ATE を近似します。
- もし X_{i1} を含めず、$Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ を推定した場合：

欠落変数バイアス – 最も単純なケース

- X_{i1} を条件として非交絡性が成り立つとし、以下を推定したいとします：

$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + e_i$$

  β_D は（CEF の近似が良ければ）ATE を近似します。

- もし X_{i1} を含めず、$Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ を推定した場合：

$$\tilde{\beta}_D = \frac{Cov(Y_i, D_i)}{Var(D_i)}$$

欠落変数バイアス – 最も単純なケース

- X_{i1} を条件として非交絡性が成り立つとし、以下を推定したいとします：

$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + e_i$$

β_D は（CEF の近似が良ければ）ATE を近似します。

- もし X_{i1} を含めず、$Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ を推定した場合：

$$\tilde{\beta}_D = \frac{Cov(Y_i, D_i)}{Var(D_i)} = \frac{Cov(\beta_0 + \beta_D D_i + \beta_1 X_{i1} + e_i, D_i)}{Var(D_i)}$$

欠落変数バイアス – 最も単純なケース

- X_{i1} を条件として非交絡性が成り立つとし、以下を推定したいとします：

$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + e_i$$

  β_D は（CEF の近似が良ければ）ATE を近似します。
- もし X_{i1} を含めず、$Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ を推定した場合：

$$\tilde{\beta}_D = \frac{Cov(Y_i, D_i)}{Var(D_i)} = \frac{Cov(\beta_0 + \beta_D D_i + \beta_1 X_{i1} + e_i, D_i)}{Var(D_i)}$$
$$= \frac{\beta_D Var(D_i) + \beta_1 Cov(X_{i1}, D_i) + 0}{Var(D_i)}$$

欠落変数バイアス – 最も単純なケース

- X_{i1} を条件として非交絡性が成り立つとし、以下を推定したいとします：

$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + e_i$$

β_D は（CEF の近似が良ければ）ATE を近似します。

- もし X_{i1} を含めず、$Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ を推定した場合：

$$\begin{aligned}
\tilde{\beta}_D &= \frac{Cov(Y_i, D_i)}{Var(D_i)} = \frac{Cov(\beta_0 + \beta_D D_i + \beta_1 X_{i1} + e_i, D_i)}{Var(D_i)} \\
&= \frac{\beta_D Var(D_i) + \beta_1 Cov(X_{i1}, D_i) + 0}{Var(D_i)} \\
&= \beta_D + \beta_1 \frac{Cov(X_{i1}, D_i)}{Var(D_i)}
\end{aligned}$$

ここで回帰の 1 階条件 $E[e_i] = E[D_i e_i] = 0$ を用いています。

OVB の評価

- （母集団）回帰 $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ は以下を与えます：

$$\tilde{\beta}_D = \beta_D + \beta_1 \gamma_D \text{（ただし } \gamma_D = \frac{Cov(X_{i1}, D_i)}{Var(D_i)}\text{）}$$

ここで β_D は私たちが本当に知りたかった係数です。

OVB の評価

- （母集団）回帰 $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ は以下を与えます：

$$\tilde{\beta}_D = \beta_D + \beta_1 \gamma_D \text{（ただし } \gamma_D = \frac{Cov(X_{i1}, D_i)}{Var(D_i)}\text{）}$$

  ここで β_D は私たちが本当に知りたかった係数です。
- バイアスは β_1 と γ_D の積に依存します。

OVB の評価

- （母集団）回帰 $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ は以下を与えます：

$$\tilde{\beta}_D = \beta_D + \beta_1 \gamma_D \text{（ただし } \gamma_D = \frac{Cov(X_{i1}, D_i)}{Var(D_i)}\text{）}$$

  ここで β_D は私たちが本当に知りたかった係数です。
- バイアスは β_1 と γ_D の積に依存します。
- β_1：D_i が与えられた下で X_{i1} が Y_i をどの程度予測するか。

OVB の評価

- （母集団）回帰 $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ は以下を与えます：

$$\tilde{\beta}_D = \beta_D + \beta_1 \gamma_D \text{（ただし } \gamma_D = \frac{Cov(X_{i1}, D_i)}{Var(D_i)}\text{）}$$

  ここで β_D は私たちが本当に知りたかった係数です。
- バイアスは β_1 と γ_D の積に依存します。
- β_1：D_i が与えられた下で X_{i1} が Y_i をどの程度予測するか。
- γ_D：X_{i1} を D_i に回帰したときの傾き（D_i と X_{i1} の相関）。

OVB の評価

- （母集団）回帰 $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ は以下を与えます：

$$\tilde{\beta}_D = \beta_D + \beta_1 \gamma_D \text{（ただし } \gamma_D = \frac{Cov(X_{i1}, D_i)}{Var(D_i)}\text{）}$$

  ここで β_D は私たちが本当に知りたかった係数です。
- バイアスは β_1 と γ_D の積に依存します。
- β_1：D_i が与えられた下で X_{i1} が Y_i をどの程度予測するか。
- γ_D：X_{i1} を D_i に回帰したときの傾き（D_i と X_{i1} の相関）。
- 結論：欠落した変数 X_{i1} が Y_i とも D_i とも強く相関している場合、$\tilde{\beta}_D$ は大きなバイアスを持ちます。

OVB の評価

- （母集団）回帰 $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i$ は以下を与えます：

$$\tilde{\beta}_D = \beta_D + \beta_1 \gamma_D \text{（ただし } \gamma_D = \frac{Cov(X_{i1}, D_i)}{Var(D_i)}\text{）}$$

  ここで β_D は私たちが本当に知りたかった係数です。
- バイアスは β_1 と γ_D の積に依存します。
- β_1：D_i が与えられた下で X_{i1} が Y_i をどの程度予測するか。
- γ_D：X_{i1} を D_i に回帰したときの傾き（D_i と X_{i1} の相関）。
- 結論：欠落した変数 X_{i1} が Y_i とも D_i とも強く相関している場合、$\tilde{\beta}_D$ は大きなバイアスを持ちます。
  - $\beta_1 = 0$ または $\gamma_D = 0$ であれば、OVB は発生しません！

有限標本における OVB 公式

- 母集団回帰における係数の関係を示しました：

$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_i + e_i \tag{1}$$

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i \tag{2}$$

これらは OVB 公式で結ばれています：

$$\tilde{\beta}_D = \beta_D + \beta_1 \frac{Cov(X_{i1}, D_i)}{Var(D_i)}$$

有限標本における OVB 公式

- 母集団回帰における係数の関係を示しました：

$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_i + e_i \tag{1}$$

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \varepsilon_i \tag{2}$$

これらは OVB 公式で結ばれています：

$$\tilde{\beta}_D = \beta_D + \beta_1 \frac{Cov(X_{i1}, D_i)}{Var(D_i)}$$

- 実は、これら 2 つの回帰の OLS 推定値の間にも全く同じ関係が成り立ちます（すべての式の母集団平均を標本平均に置き換えるだけです）：

$$\hat{\tilde{\beta}}_D = \hat{\beta}_D + \hat{\beta}_1 \frac{\widehat{Cov}(X_{i1}, D_i)}{\widehat{Var}(D_i)}$$

OVB のイラスト

- Angrist and Pischke の教科書では、Dale and Krueger の変形版として、D を「私立大学に通うかどうか」、X_{i1} を「学生の SAT スコア」としたケースを考えています。

OVB のイラスト

- Angrist and Pischke の教科書では、Dale and Krueger の変形版として、D を「私立大学に通うかどうか」、X_{i1} を「学生の SAT スコア」としたケースを考えています。
- 彼らは以下の回帰を推定しています：

$$Y_i = D_i\beta_D + X_{i1}\beta_1 + \varepsilon_i$$

OVB のイラスト

- Angrist and Pischke の教科書では、Dale and Krueger の変形版として、D を「私立大学に通うかどうか」、X_{i1} を「学生の SAT スコア」としたケースを考えています。
- 彼らは以下の回帰を推定しています：

$$Y_i = D_i \beta_D + X_{i1} \beta_1 + \varepsilon_i$$

- もし X_{i1} を条件として非交絡性が成り立ち（かつ CEF が近似的に線形なら）、係数 β_D は私立大学に通うことの因果的効果に対応します。

OVB のイラスト

- Angrist and Pischke の教科書では、Dale and Krueger の変形版として、D を「私立大学に通うかどうか」、X_{i1} を「学生の SAT スコア」としたケースを考えています。

- 彼らは以下の回帰を推定しています：

$$Y_i = D_i \beta_D + X_{i1}\beta_1 + \varepsilon_i$$

- もし X_{i1} を条件として非交絡性が成り立ち（かつ CEF が近似的に線形なら）、係数 β_D は私立大学に通うことの因果的効果に対応します。

- ここで、SAT スコアをコントロールするのを忘れたらどうなるかを考えてみましょう。

- A&P が SAT をコントロールしたときに得た結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.095	0.052
SAT スコア /100 ($\hat{\beta}_1$)	0.048	0.009
定数項	[...]	[...]

- もし回帰から SAT スコアのコントロールを除外したら何が起きるでしょうか？ $\hat{\beta}_D$ はどう変わるでしょうか？

- A&P が SAT をコントロールしたときに得た結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.095	0.052
SAT スコア /100 ($\hat{\beta}_1$)	0.048	0.009
定数項	[...]	[...]

- もし回帰から SAT スコアのコントロールを除外したら何が起きるでしょうか？ $\hat{\beta}_D$ はどう変わるでしょうか？

- これを計算するには、X_{i1} (SAT スコア /100) を D_i (私立大学) に回帰した係数を知る必要があります：

- A&P が SAT をコントロールしたときに得た結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.095	0.052
SAT スコア /100 ($\hat{\beta}_1$)	0.048	0.009
定数項	[...]	[...]

- もし回帰から SAT スコアのコントロールを除外したら何が起きるでしょうか？ $\hat{\beta}_D$ はどう変わるでしょうか？

- これを計算するには、X_{i1} (SAT スコア /100) を D_i (私立大学) に回帰した係数を知る必要があります：

変数	係数
私立大学 ($\hat{\gamma}_D$)	.83
定数項	[...]

- A&P が SAT をコントロールしたときに得た結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.095	0.052
SAT スコア /100 ($\hat{\beta}_1$)	0.048	0.009
定数項	[...]	[...]

- もし回帰から SAT スコアのコントロールを除外したら何が起きるでしょうか？ $\hat{\beta}_D$ はどう変わるでしょうか？

- これを計算するには、X_{i1} (SAT スコア /100) を D_i (私立大学) に回帰した係数を知る必要があります：

変数	係数
私立大学 ($\hat{\gamma}_D$)	.83
定数項	[...]

- したがって、もし X_{i1} を除外すれば、私立大学の推定係数は $\hat{\gamma}_D \times \hat{\beta}_1 = 0.83 \times 0.048 \approx 0.04$ だけ大きくなります。

- 実際に SAT スコアを除外して回帰を実行すると、私立大学の係数は 0.04 大きくなります。

- SAT スコアを含めた結果：

変数	係数	標準誤差 (SE)
私立大学 (β_D)	0.095	0.052
SAT スコア /100 (β_1)	0.048	0.009
定数項	[...]	[...]

- SAT スコアを除外した結果：

変数	係数	標準誤差 (SE)
私立大学 ($\tilde{\beta}_D$)	0.135	0.055
定数項	[...]	[...]

- 実際に SAT スコアを除外して回帰を実行すると、私立大学の係数は 0.04 大きくなります。

- SAT スコアを含めた結果：

変数	係数	標準誤差 (SE)
私立大学 (β_D)	0.095	0.052
SAT スコア /100 (β_1)	0.048	0.009
定数項	[...]	[...]

- SAT スコアを除外した結果：

変数	係数	標準誤差 (SE)
私立大学 ($\tilde{\beta}_D$)	0.135	0.055
定数項	[...]	[...]

- どのような場合に、SAT スコアの除外がより大きな問題になるでしょうか？

- 実際に SAT スコアを除外して回帰を実行すると、私立大学の係数は 0.04 大きくなります。

- SAT スコアを含めた結果：

変数	係数	標準誤差 (SE)
私立大学 (β_D)	0.095	0.052
SAT スコア /100 (β_1)	0.048	0.009
定数項	[...]	[...]

- SAT スコアを除外した結果：

変数	係数	標準誤差 (SE)
私立大学 ($\tilde{\beta}_D$)	0.135	0.055
定数項	[...]	[...]

- どのような場合に、SAT スコアの除外がより大きな問題になるでしょうか？
  - SAT スコアが所得とより強く関連している場合（$\hat{\beta}_1$ が大きい）

- 実際に SAT スコアを除外して回帰を実行すると、私立大学の係数は 0.04 大きくなります。

- SAT スコアを含めた結果：

変数	係数	標準誤差 (SE)
私立大学 (β_D)	0.095	0.052
SAT スコア /100 (β_1)	0.048	0.009
定数項	[...]	[...]

- SAT スコアを除外した結果：

変数	係数	標準誤差 (SE)
私立大学 ($\tilde{\beta}_D$)	0.135	0.055
定数項	[...]	[...]

- どのような場合に、SAT スコアの除外がより大きな問題になるでしょうか？
  - SAT スコアが所得とより強く関連している場合（$\hat{\beta}_1$ が大きい）
  - 私立大学への進学が SAT スコアとより強く相関している場合（$\hat{\gamma}_D$ が大きい）

欠落変数バイアス公式 – 複数変数の場合

- さて、Y_i, D_i, および X_{i1} は観察されていますが、非交絡性は $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$ を条件としない限り成り立たないとします。

欠落変数バイアス公式 – 複数変数の場合

- さて、Y_i, D_i, および X_{i1} は観察されていますが、非交絡性は $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$ を条件としない限り成り立たないとします。例：Y_i は所得、D_i は大学の選択度、X_{i1} は高校の GPA、そして X_{i2} は SAT スコア。

欠落変数バイアス公式 – 複数変数の場合

- さて、Y_i, D_i, および X_{i1} は観察されていますが、非交絡性は $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$ を条件としない限り成り立たないとします。例：Y_i は所得、D_i は大学の選択度、X_{i1} は高校の GPA、そして X_{i2} は SAT スコア。

- 本来であれば $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ を推定したいところです。

欠落変数バイアス公式 – 複数変数の場合

- さて、Y_i, D_i, および X_{i1} は観察されていますが、非交絡性は $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$ を条件としない限り成り立たないとします。例：Y_i は所得、D_i は大学の選択度、X_{i1} は高校の GPA、そして X_{i2} は SAT スコア。

- 本来であれば $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ を推定したいところです。

- しかし X_{i2} が観察できないため、代わりに $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \tilde{\beta}_1 X_{i1} + \varepsilon_i$ を推定します。

欠落変数バイアス公式 – 複数変数の場合

- さて、Y_i, D_i, および X_{i1} は観察されていますが、非交絡性は $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$ を条件としない限り成り立たないとします。例：Y_i は所得、D_i は大学の選択度、X_{i1} は高校の GPA、そして X_{i2} は SAT スコア。

- 本来であれば $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ を推定したいところです。

- しかし X_{i2} が観察できないため、代わりに $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \tilde{\beta}_1 X_{i1} + \varepsilon_i$ を推定します。

- $\tilde{\beta}_D$ と β_D はどのように関連しているでしょうか？

欠落変数バイアス公式 – 複数変数の場合

- さて、Y_i, D_i, および X_{i1} は観察されていますが、非交絡性は $\boldsymbol{X}_i = (X_{i1}, X_{i2})'$ を条件としない限り成り立たないとします。例：Y_i は所得、D_i は大学の選択度、X_{i1} は高校の GPA、そして X_{i2} は SAT スコア。

- 本来であれば $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ を推定したいところです。

- しかし X_{i2} が観察できないため、代わりに $Y_i = \tilde{\beta}_0 + \tilde{\beta}_D D_i + \tilde{\beta}_1 X_{i1} + \varepsilon_i$ を推定します。

- $\tilde{\beta}_D$ と β_D はどのように関連しているでしょうか？ 答え：

$$\tilde{\beta}_D = \beta_D + \beta_2 \gamma_D$$

  ここで γ_D は、以下の回帰から得られる係数です：

$$X_{i2} = \gamma_0 + \gamma_D D_i + \gamma_1 X_{i1} + u_i$$

- これは以前の OVB 公式と似ていますが、すべてにおいて X_{i1} がコントロールされている点が異なります。

バイアスの評価（再び）

- 多変量の OVB 公式は $\tilde{\beta}_D = \beta_D + \beta_2 \gamma_D$ です。ここで β_D は（X_{i2} をコントロールできた場合に）得たかった係数です。

バイアスの評価（再び）

- 多変量の OVB 公式は $\tilde{\beta}_D = \beta_D + \beta_2 \gamma_D$ です。ここで β_D は（X_{i2} をコントロールできた場合に）得たかった係数です。以前と同様に、バイアスは β_2 と γ_D の積となります。

バイアスの評価（再び）

- 多変量の OVB 公式は $\tilde{\beta}_D = \beta_D + \beta_2\gamma_D$ です。ここで β_D は（X_{i2} をコントロールできた場合に）得たかった係数です。以前と同様に、バイアスは β_2 と γ_D の積となります。
- $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ であったことを思い出してください。
  $\Longrightarrow$ したがって、β_2 は、X_{i1} と D_i をコントロールした後の X_{i2} と Y_i の相関を表します。

バイアスの評価（再び）

- 多変量の OVB 公式は $\tilde{\beta}_D = \beta_D + \beta_2\gamma_D$ です。ここで β_D は（X_{i2} をコントロールできた場合に）得たかった係数です。以前と同様に、バイアスは β_2 と γ_D の積となります。

- $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ であったことを思い出してください。
$\Longrightarrow$ したがって、β_2 は、X_{i1} と D_i をコントロールした後の X_{i2} と Y_i の相関を表します。

- また $X_{i2} = \gamma_0 + \gamma_D D_i + \gamma_1 X_{i1} + u_i$ です。
$\Longrightarrow$ したがって、γ_D は、X_{i1} をコントロールした後の D_i と X_{i2} の相関を表します。

バイアスの評価（再び）

- 多変量の OVB 公式は $\tilde{\beta}_D = \beta_D + \beta_2 \gamma_D$ です。ここで β_D は（X_{i2} をコントロールできた場合に）得たかった係数です。以前と同様に、バイアスは β_2 と γ_D の積となります。

- $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ であったことを思い出してください。
  $\Longrightarrow$ したがって、β_2 は、X_{i1} と D_i をコントロールした後の X_{i2} と Y_i の相関を表します。

- また $X_{i2} = \gamma_0 + \gamma_D D_i + \gamma_1 X_{i1} + u_i$ です。
  $\Longrightarrow$ したがって、γ_D は、X_{i1} をコントロールした後の D_i と X_{i2} の相関を表します。

- まとめ：欠落した変数 X_{i2} が、X_{i1} をコントロールした後でもアウトカム Y_i および処置 D_i の両方と相関している場合、$\tilde{\beta}_D$ には大きなバイアスが生じます。

バイアスの評価（再び）

- 多変量の OVB 公式は $\tilde{\beta}_D = \beta_D + \beta_2 \gamma_D$ です。ここで β_D は（X_{i2} をコントロールできた場合に）得たかった係数です。以前と同様に、バイアスは β_2 と γ_D の積となります。

- $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ であったことを思い出してください。
  $\Longrightarrow$ したがって、β_2 は、X_{i1} と D_i をコントロールした後の X_{i2} と Y_i の相関を表します。

- また $X_{i2} = \gamma_0 + \gamma_D D_i + \gamma_1 X_{i1} + u_i$ です。
  $\Longrightarrow$ したがって、γ_D は、X_{i1} をコントロールした後の D_i と X_{i2} の相関を表します。

- まとめ：欠落した変数 X_{i2} が、X_{i1} をコントロールした後でもアウトカム Y_i および処置 D_i の両方と相関している場合、$\tilde{\beta}_D$ には大きなバイアスが生じます。
  - いずれかの相関が 0 であれば、OVB は 0 です。

バイアスの評価（再び）

- 多変量の OVB 公式は $\tilde{\beta}_D = \beta_D + \beta_2 \gamma_D$ です。ここで β_D は（X_{i2} をコントロールできた場合に）得たかった係数です。以前と同様に、バイアスは β_2 と γ_D の積となります。

- $Y_i = \beta_0 + \beta_D D_i + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ であったことを思い出してください。
$\Longrightarrow$ したがって、β_2 は、X_{i1} と D_i をコントロールした後の X_{i2} と Y_i の相関を表します。

- また $X_{i2} = \gamma_0 + \gamma_D D_i + \gamma_1 X_{i1} + u_i$ です。
$\Longrightarrow$ したがって、γ_D は、X_{i1} をコントロールした後の D_i と X_{i2} の相関を表します。

- まとめ：欠落した変数 X_{i2} が、X_{i1} をコントロールした後でもアウトカム Y_i および処置 D_i の両方と相関している場合、$\tilde{\beta}_D$ には大きなバイアスが生じます。
    - いずれかの相関が 0 であれば、OVB は 0 です。
    - 相関が同じ符号であれば OVB は正になり、そうでなければ負になります。

欠落変数バイアスのイラスト

- A&P の教科書では、再び Dale and Krueger の例を考えます。D_i は私立大学に通うかどうかです。

欠落変数バイアスのイラスト

- A&P の教科書では、再び Dale and Krueger の例を考えます。D_i は私立大学に通うかどうかです。

- X_{i2} を学生の SAT スコアとし、$\boldsymbol{X}_{i1}$ を合格した大学のセットを示すベクトルの固定効果とします。

欠落変数バイアスのイラスト

- A&P の教科書では、再び Dale and Krueger の例を考えます。D_i は私立大学に通うかどうかです。
- X_{i2} を学生の SAT スコアとし、$\boldsymbol{X}_{i1}$ を合格した大学のセットを示すベクトルの固定効果とします。
- 彼らは以下の回帰を推定しています：

$$Y_i = D_i\beta_D + \boldsymbol{X}'_{i1}\boldsymbol{\beta}_1 + X_{i2}\beta_2 + \varepsilon_i$$

欠落変数バイアスのイラスト

- A&P の教科書では、再び Dale and Krueger の例を考えます。D_i は私立大学に通うかどうかです。
- X_{i2} を学生の SAT スコアとし、$\boldsymbol{X}_{i1}$ を合格した大学のセットを示すベクトルの固定効果とします。
- 彼らは以下の回帰を推定しています：

$$Y_i = D_i \beta_D + \boldsymbol{X}_{i1}' \boldsymbol{\beta}_1 + X_{i2} \beta_2 + \varepsilon_i$$

- もし $\boldsymbol{X}_{i1}$ と X_{i2} を条件として非交絡性が成り立てば、係数 β_D は私立大学への進学による因果的効果に対応します。

欠落変数バイアスのイラスト

- A&P の教科書では、再び Dale and Krueger の例を考えます。D_i は私立大学に通うかどうかです。
- X_{i2} を学生の SAT スコアとし、$\boldsymbol{X}_{i1}$ を合格した大学のセットを示すベクトルの固定効果とします。
- 彼らは以下の回帰を推定しています：

$$Y_i = D_i\beta_D + \boldsymbol{X}_{i1}'\boldsymbol{\beta}_1 + X_{i2}\beta_2 + \varepsilon_i$$

- もし $\boldsymbol{X}_{i1}$ と X_{i2} を条件として非交絡性が成り立てば、係数 β_D は私立大学への進学による因果的効果に対応します。
- ここで、SAT スコアのコントロールを忘れたらどうなるか考えてみましょう。

- A&P が SAT スコアと合格大学セットの両方をコントロールした結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.003	0.039
SAT スコア /100 ($\hat{\beta}_2$)	0.033	0.007
合格大学セット ($\hat{\boldsymbol{\beta}}_1$)	[...]	[...]

- 回帰から SAT スコアのコントロールを除外した場合、$\hat{\beta}_D$ はどう変わるでしょうか？

- A&P が SAT スコアと合格大学セットの両方をコントロールした結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.003	0.039
SAT スコア /100 ($\hat{\beta}_2$)	0.033	0.007
合格大学セット ($\hat{\boldsymbol{\beta}}_1$)	[...]	[...]

- 回帰から SAT スコアのコントロールを除外した場合、$\hat{\beta}_D$ はどう変わるでしょうか？

- これを計算するには、X_{i2} (SAT スコア /100) を D_i (私立大学) と $\boldsymbol{X}_{i1}$ (合格大学セット) に回帰した係数を知る必要があります：

- A&P が SAT スコアと合格大学セットの両方をコントロールした結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.003	0.039
SAT スコア /100 ($\hat{\beta}_2$)	0.033	0.007
合格大学セット ($\hat{\boldsymbol{\beta}}_1$)	[...]	[...]

- 回帰から SAT スコアのコントロールを除外した場合、$\hat{\beta}_D$ はどう変わるでしょうか？

- これを計算するには、X_{i2} (SAT スコア /100) を D_i (私立大学) と $\boldsymbol{X}_{i1}$ (合格大学セット) に回帰した係数を知る必要があります：

変数	係数
私立大学 ($\hat{\gamma}_D$)	.12
合格大学セット ($\hat{\boldsymbol{\gamma}}_1$)	[...]

- A&P が SAT スコアと合格大学セットの両方をコントロールした結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.003	0.039
SAT スコア /100 ($\hat{\beta}_2$)	0.033	0.007
合格大学セット ($\hat{\boldsymbol{\beta}}_1$)	[...]	[...]

- 回帰から SAT スコアのコントロールを除外した場合、$\hat{\beta}_D$ はどう変わるでしょうか？

- これを計算するには、X_{i2} (SAT スコア /100) を D_i (私立大学) と $\boldsymbol{X}_{i1}$ (合格大学セット) に回帰した係数を知る必要があります：

変数	係数
私立大学 ($\hat{\gamma}_D$)	.12
合格大学セット ($\hat{\boldsymbol{\gamma}}_1$)	[...]

- X_{i2} を除外することによる変化は $\hat{\gamma}_D \times \hat{\beta}_2$ なので、除外後の推定係数は

- A&P が SAT スコアと合格大学セットの両方をコントロールした結果は以下の通りです：

変数	係数	標準誤差 (SE)
私立大学 ($\hat{\beta}_D$)	0.003	0.039
SAT スコア /100 ($\hat{\beta}_2$)	0.033	0.007
合格大学セット ($\hat{\boldsymbol{\beta}}_1$)	[...]	[...]

- 回帰から SAT スコアのコントロールを除外した場合、$\hat{\beta}_D$ はどう変わるでしょうか？

- これを計算するには、X_{i2} (SAT スコア /100) を D_i (私立大学) と $\boldsymbol{X}_{i1}$ (合格大学セット) に回帰した係数を知る必要があります：

変数	係数
私立大学 ($\hat{\gamma}_D$)	.12
合格大学セット ($\hat{\boldsymbol{\gamma}}_1$)	[...]

- X_{i2} を除外することによる変化は $\hat{\gamma}_D \times \hat{\beta}_2$ なので、除外後の推定係数は $0.033 \times .12 \approx 0.004$ だけ大きくなります。

- 実際に SAT スコアを除外して回帰を実行すると、私立大学の係数は 0.004 だけ大きくなっています。

- SAT スコアを含めた結果：

変数	係数	標準誤差 (SE)
私立大学 (β_D)	0.003	0.039
SAT スコア /100 (β_2)	0.033	0.007
合格大学セット ($\boldsymbol{\beta}_1$)	[...]	[...]

- SAT スコアを除外した結果：

変数	係数	標準誤差 (SE)
私立大学 ($\tilde{\beta}_D$)	0.007	0.038
合格大学セット ($\tilde{\boldsymbol{\beta}}_X$)	[...]	[...]

- どのような場合に、SAT スコアの除外がより重要になるでしょうか？

- 実際に SAT スコアを除外して回帰を実行すると、私立大学の係数は 0.004 だけ大きくなっています。

- SAT スコアを含めた結果：

変数	係数	標準誤差 (SE)
私立大学 (β_D)	0.003	0.039
SAT スコア /100 (β_2)	0.033	0.007
合格大学セット ($\boldsymbol{\beta}_1$)	[...]	[...]

- SAT スコアを除外した結果：

変数	係数	標準誤差 (SE)
私立大学 ($\tilde{\beta}_D$)	0.007	0.038
合格大学セット ($\tilde{\boldsymbol{\beta}}_X$)	[...]	[...]

- どのような場合に、SAT スコアの除外がより重要になるでしょうか？
  - （合格大学セットをコントロールした後でも）SAT スコアが所得とより強く関連している場合、すなわち $\hat{\beta}_2$ が大きい。

- 実際に SAT スコアを除外して回帰を実行すると、私立大学の係数は 0.004 だけ大きくなっています。

- SAT スコアを含めた結果：

変数	係数	標準誤差 (SE)
私立大学 (β_D)	0.003	0.039
SAT スコア /100 (β_2)	0.033	0.007
合格大学セット ($\boldsymbol{\beta}_1$)	[...]	[...]

- SAT スコアを除外した結果：

変数	係数	標準誤差 (SE)
私立大学 ($\tilde{\beta}_D$)	0.007	0.038
合格大学セット ($\tilde{\boldsymbol{\beta}}_X$)	[...]	[...]

- どのような場合に、SAT スコアの除外がより重要になるでしょうか？
  - （合格大学セットをコントロールした後でも）SAT スコアが所得とより強く関連している場合、すなわち $\hat{\beta}_2$ が大きい。
  - （合格大学セットをコントロールした後でも）私立大学への進学が SAT スコアとより強く相関している場合、すなわち $\hat{\gamma}_D$ が大きい。

異質的処置効果のモデル化

- しばしば、処置効果が異質的（＝グループで異なる）かに興味があります。例えば、エリート校の効果は家庭の所得で異なるでしょうか？

異質的処置効果のモデル化

- しばしば、処置効果が異質的（＝グループで異なる）かに興味があります。例えば、エリート校の効果は家庭の所得で異なるでしょうか？
- 条件付き非交絡性の下では：

$$CATE(\boldsymbol{x}) = E[Y_i|D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$$

異質的処置効果のモデル化

- しばしば、処置効果が異質的（＝グループで異なる）かに興味があります。例えば、エリート校の効果は家庭の所得で異なるでしょうか？
- 条件付き非交絡性の下では：

$$CATE(\boldsymbol{x}) = E[Y_i|D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$$

- ここで以下のように近似します：

$$E[Y_i|D_i = d, \boldsymbol{X}_i = \boldsymbol{x}] \approx \beta_D d + \beta_{DX}(x_1 \times d) + \boldsymbol{x}'\boldsymbol{\gamma}$$

異質的処置効果のモデル化

- しばしば、処置効果が異質的（＝グループで異なる）かに興味があります。例えば、エリート校の効果は家庭の所得で異なるでしょうか？
- 条件付き非交絡性の下では：

$$CATE(\boldsymbol{x}) = E[Y_i|D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$$

- ここで以下のように近似します：

$$E[Y_i|D_i = d, \boldsymbol{X}_i = \boldsymbol{x}] \approx \beta_D d + \beta_{DX}(x_1 \times d) + \boldsymbol{x}'\boldsymbol{\gamma}$$

- すると $CATE(\boldsymbol{x}) \approx \beta_D + \beta_{DX} x_1$。
  つまり $X_{i1} = x_1$ の人の効果は、近似的に $\beta_D + \beta_{DX} x_1$ です。

異質的処置効果のモデル化

- しばしば、処置効果が異質的（＝グループで異なる）かに興味があります。例えば、エリート校の効果は家庭の所得で異なるでしょうか？
- 条件付き非交絡性の下では：

$$CATE(\boldsymbol{x}) = E[Y_i|D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$$

- ここで以下のように近似します：

$$E[Y_i|D_i = d, \boldsymbol{X}_i = \boldsymbol{x}] \approx \beta_D d + \beta_{DX}(x_1 \times d) + \boldsymbol{x}'\boldsymbol{\gamma}$$

- すると $CATE(\boldsymbol{x}) \approx \beta_D + \beta_{DX} x_1$。
  つまり $X_{i1} = x_1$ の人の効果は、近似的に $\beta_D + \beta_{DX} x_1$ です。
- X_{i1} による異質性に興味がある場合は、以下を推定します：

$$Y_i = \beta_D D_i + \beta_{DX}(D_i \times X_{i1}) + \boldsymbol{\gamma}'\boldsymbol{X}_i + \varepsilon_i$$

Dale and Krueger – 異質的効果

- DK は以下の形式の回帰を推定しています：

$$Y_i = \beta_D D_i + \beta_{DX}(D_i \times X_{i1}) + \boldsymbol{\gamma}' \boldsymbol{X}_i + \varepsilon_i$$

  ここで D_i は大学の平均 SAT スコア $/100$、X_{i1} は世帯所得の対数、$\boldsymbol{X}_i$ は進学した大学のダミー変数やその他のコントロール変数を含みます。

Dale and Krueger – 異質的効果

- DK は以下の形式の回帰を推定しています：

$$Y_i = \beta_D D_i + \beta_{DX}(D_i \times X_{i1}) + \boldsymbol{\gamma}' \boldsymbol{X}_i + \varepsilon_i$$

ここで D_i は大学の平均 SAT スコア /100、X_{i1} は世帯所得の対数、$\boldsymbol{X}_i$ は進学した大学のダミー変数やその他のコントロール変数を含みます。

  - $D_i \times X_{i1}$ という変数は、しばしば交差項（interaction term）と呼ばれます。

Dale and Krueger – 異質的効果

- DK は以下の形式の回帰を推定しています：

$$Y_i = \beta_D D_i + \beta_{DX}(D_i \times X_{i1}) + \boldsymbol{\gamma}' \boldsymbol{X}_i + \varepsilon_i$$

  ここで D_i は大学の平均 SAT スコア /100、X_{i1} は世帯所得の対数、$\boldsymbol{X}_i$ は進学した大学のダミー変数やその他のコントロール変数を含みます。
  - $D_i \times X_{i1}$ という変数は、しばしば交差項（interaction term）と呼ばれます。

- 世帯所得が 10 万ドルの学生にとって、平均 SAT が 100 ポイント高い大学に通うことの推定効果は：

Dale and Krueger – 異質的効果

- DK は以下の形式の回帰を推定しています：

$$Y_i = \beta_D D_i + \beta_{DX}(D_i \times X_{i1}) + \boldsymbol{\gamma}' \boldsymbol{X}_i + \varepsilon_i$$

  ここで D_i は大学の平均 SAT スコア $/100$、X_{i1} は世帯所得の対数、$\boldsymbol{X}_i$ は進学した大学のダミー変数やその他のコントロール変数を含みます。
  - $D_i \times X_{i1}$ という変数は、しばしば交差項（interaction term）と呼ばれます。

- 世帯所得が 10 万ドルの学生にとって、平均 SAT が 100 ポイント高い大学に通うことの推定効果は：

$$\hat{\beta}_D + \hat{\beta}_{DX} \times ln(100,000)$$

OF MALE AND FEMALE FULL-TIME WORKERS

Variable	Parameter estimates: Basic model: no selection controls (1)	Parameter estimates: Matched-applicant model* (2)
School-average SAT score/100	0.701	0.537
	(0.185)	(0.224)
Predicted log(parental income)	0.915	0.819
	(0.212)	(0.247)
Predicted log of parental income * school SAT score/100	−0.063	−0.056
	(0.019)	(0.023)
Own SAT score/100	0.018	−0.011
	(0.006)	(0.007)
High school top 10 percent	0.062	0.080
	(0.019)	(0.026)
High school rank missing	0.005	0.018
	(0.024)	(0.038)
Athlete	0.104	0.105
	(0.025)	(0.040)

Variable	Parameter estimates: Basic model: no selection controls (1)	Parameter estimates: Matched-applicant model* (2)
School-average SAT score/100	0.701	0.537
	(0.185)	(0.224)
Predicted log(parental income)	0.915	0.819
	(0.212)	(0.247)
Predicted log of parental income * school SAT score/100	−0.063	−0.056
	(0.019)	(0.023)
Own SAT score/100	0.018	−0.011
	(0.006)	(0.007)
High school top 10 percent	0.062	0.080
	(0.019)	(0.026)
High school rank missing	0.005	0.018
	(0.024)	(0.038)
Athlete	0.104	0.105
	(0.025)	(0.040)

- DK のデータでは、世帯所得の下位 10
- 下位 10

OF MALE AND FEMALE FULL-TIME WORKERS

	Parameter estimates	
	Basic model: no selection controls	Matched-applicant model*
Variable	1	2
School-average SAT score/100	0.701	0.537
	(0.185)	(0.224)
Predicted log(parental income)	0.915	0.819
	(0.212)	(0.247)
Predicted log of parental income $*$ school SAT score/100	-0.063	-0.056
	(0.019)	(0.023)
Own SAT score/100	0.018	-0.011
	(0.006)	(0.007)
High school top 10 percent	0.062	0.080
	(0.019)	(0.026)
High school rank missing	0.005	0.018
	(0.024)	(0.038)
Athlete	0.104	0.105
	(0.025)	(0.040)

- DK のデータでは、世帯所得の下位 10
- 下位 10

$$\hat{\beta}_D + \hat{\beta}_{DX} x =$$

OF MALE AND FEMALE FULL-TIME WORKERS

Variable	Parameter estimates: Basic model: no selection controls (1)	Parameter estimates: Matched-applicant model* (2)
School-average SAT score/100	0.701	0.537
	(0.185)	(0.224)
Predicted log(parental income)	0.915	0.819
	(0.212)	(0.247)
Predicted log of parental income * school SAT score/100	−0.063	−0.056
	(0.019)	(0.023)
Own SAT score/100	0.018	−0.011
	(0.006)	(0.007)
High school top 10 percent	0.062	0.080
	(0.019)	(0.026)
High school rank missing	0.005	0.018
	(0.024)	(0.038)
Athlete	0.104	0.105
	(0.025)	(0.040)

- DK のデータでは、世帯所得の下位 10
- 下位 10

$$\hat{\beta}_D + \hat{\beta}_{DX} x = 0.537 - 0.056 \times 8.86 = 0.041$$

OF MALE AND FEMALE FULL-TIME WORKERS

	Parameter estimates	
	Basic model: no selection controls	Matched-applicant model*
Variable	1	2
School-average SAT score/100	0.701	0.537
	(0.185)	(0.224)
Predicted log(parental income)	0.915	0.819
	(0.212)	(0.247)
Predicted log of parental income * school SAT score/100	−0.063	−0.056
	(0.019)	(0.023)
Own SAT score/100	0.018	−0.011
	(0.006)	(0.007)
High school top 10 percent	0.062	0.080
	(0.019)	(0.026)
High school rank missing	0.005	0.018
	(0.024)	(0.038)
Athlete	0.104	0.105
	(0.025)	(0.040)

- DK では、世帯所得の中央値（50
- 中央値の学生にとって、平均 SAT が 100 ポイント高い大学に通う推定効果は？

$$\hat{\beta}_D + \hat{\beta}_{DX} x =$$

	Parameter estimates	
	Basic model: no selection controls	Matched-applicant model*
Variable	1	2
School-average SAT score/100	0.701	0.537
	(0.185)	(0.224)
Predicted log(parental income)	0.915	0.819
	(0.212)	(0.247)
Predicted log of parental income * school SAT score/100	−0.063	−0.056
	(0.019)	(0.023)
Own SAT score/100	0.018	−0.011
	(0.006)	(0.007)
High school top 10 percent	0.062	0.080
	(0.019)	(0.026)
High school rank missing	0.005	0.018
	(0.024)	(0.038)
Athlete	0.104	0.105
	(0.025)	(0.040)

- DK では、世帯所得の中央値（50
- 中央値の学生にとって、平均 SAT が 100 ポイント高い大学に通う推定効果は？

$$\hat{\beta}_D + \hat{\beta}_{DX} x = 0.537 - 0.056 \times 10.39 = -0.045$$

異質性の説明

- 上記の結果は、より選択度の高い大学に通う収益が、貧しい家庭の学生にとってはプラスですが、裕福な家庭の学生にとってはマイナスであることを示しています。

異質性の説明

- 上記の結果は、より選択度の高い大学に通う収益が、貧しい家庭の学生にとってはプラスですが、裕福な家庭の学生にとってはマイナスであることを示しています。
- なぜ貧しいバックグラウンドの学生にとって、エリート校進学がより重要に見えるのでしょうか？

異質性の説明

- 上記の結果は、より選択度の高い大学に通う収益が、貧しい家庭の学生にとってはプラスですが、裕福な家庭の学生にとってはマイナスであることを示しています。
- なぜ貧しいバックグラウンドの学生にとって、エリート校進学がより重要に見えるのでしょうか？
- 明確な理由は分かっていません... 家庭のコネクションが少ない学生にとって、ネットワーキングがより重要だからでしょうか？

係数の線形結合

- 上記の DK の例では、世帯所得が x の学生に対する推定された $CATE(x)$ は $\hat{\beta}_D + \hat{\beta}_{DX}x$ でした。

係数の線形結合

- 上記の DK の例では、世帯所得が x の学生に対する推定された $CATE(x)$ は $\hat{\beta}_D + \hat{\beta}_{DX}x$ でした。
- この $CATE(x)$ について信頼区間を構築したり、仮説検定を行いたい場合はどうすればよいでしょうか？

係数の線形結合

- 数講義前に、以下を示しました：

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow_d \mathrm{N}(0, \boldsymbol{\Sigma})$$

係数の線形結合

- 数講義前に、以下を示しました：

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}(0, \boldsymbol{\Sigma})$$

- ここで $\beta_1 + \beta_2 x$ に興味があるとしましょう。

係数の線形結合

- 数講義前に、以下を示しました：

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}(0, \boldsymbol{\Sigma})$$

- ここで $\beta_1 + \beta_2 x$ に興味があるとしましょう。$g(\boldsymbol{\beta}) = \beta_1 + \beta_2 x$ は連続なので、連続写像定理より：

係数の線形結合

- 数講義前に、以下を示しました：

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}(0, \boldsymbol{\Sigma})$$

- ここで $\beta_1 + \beta_2 x$ に興味があるとしましょう。$g(\boldsymbol{\beta}) = \beta_1 + \beta_2 x$ は連続なので、連続写像定理より：

$$\sqrt{N}(\hat{\beta}_1 + \hat{\beta}_2 x - (\beta_1 + \beta_2 x)) \to_d \mathrm{N}(0, \sigma_x^2),$$

ここで $\sigma_x^2 = \Sigma_{11} + x^2 \Sigma_{22} + 2x\Sigma_{12}$ です。

係数の線形結合

- 数講義前に、以下を示しました：

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}(0, \boldsymbol{\Sigma})$$

- ここで $\beta_1 + \beta_2 x$ に興味があるとしましょう。$g(\boldsymbol{\beta}) = \beta_1 + \beta_2 x$ は連続なので、連続写像定理より：

$$\sqrt{N}(\hat{\beta}_1 + \hat{\beta}_2 x - (\beta_1 + \beta_2 x)) \to_d \mathrm{N}(0, \sigma_x^2),$$

ここで $\sigma_x^2 = \Sigma_{11} + x^2\Sigma_{22} + 2x\Sigma_{12}$ です。

- 以前の講義で、$\boldsymbol{\Sigma}$ の一致推定量 $\hat{\boldsymbol{\Sigma}}$ を（標本対応物を代入することで）導出しました。

係数の線形結合

- 数講義前に、以下を示しました：

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}(0, \boldsymbol{\Sigma})$$

- ここで $\beta_1 + \beta_2 x$ に興味があるとしましょう。$g(\boldsymbol{\beta}) = \beta_1 + \beta_2 x$ は連続なので、連続写像定理より：

$$\sqrt{N}(\hat{\beta}_1 + \hat{\beta}_2 x - (\beta_1 + \beta_2 x)) \to_d \mathrm{N}(0, \sigma_x^2),$$

  ここで $\sigma_x^2 = \Sigma_{11} + x^2\Sigma_{22} + 2x\Sigma_{12}$ です。

- 以前の講義で、$\boldsymbol{\Sigma}$ の一致推定量 $\hat{\boldsymbol{\Sigma}}$ を（標本対応物を代入することで）導出しました。

- したがって、$\beta_1 + \beta_2 x$ に対する信頼区間を $\hat{\beta}_1 + \hat{\beta}_2 x \pm 1.96\hat{\sigma}_x/\sqrt{N}$ として形成できます。ここで $\hat{\sigma}_x^2 = \hat{\Sigma}_{11} + x^2\hat{\Sigma}_{22} + 2x\hat{\Sigma}_{12}$ です。

例題

- $\hat{\beta}_D = 0.537$, $\hat{\beta}_{DX} = -0.056$ でした。$\hat{\Sigma}/N$ の一部が以下のとき：

	$\hat{\beta}_D$	$\hat{\beta}_{DX}$
$\hat{\beta}_D$	0.0050	-0.0025
$\hat{\beta}_{DX}$	-0.0025	0.0005

例題

- $\hat{\beta}_D = 0.537$, $\hat{\beta}_{DX} = -0.056$ でした。$\hat{\Sigma}/N$ の一部が以下のとき：

	$\hat{\beta}_D$	$\hat{\beta}_{DX}$
$\hat{\beta}_D$	0.0050	-0.0025
$\hat{\beta}_{DX}$	-0.0025	0.0005

- $\hat{\Sigma}_{11}/N = 0.0050$, $\hat{\Sigma}_{22}/N = 0.0005$, $\hat{\Sigma}_{12}/N = -0.0025$。

例題

- $\hat{\beta}_D = 0.537$, $\hat{\beta}_{DX} = -0.056$ でした。$\hat{\Sigma}/N$ の一部が以下のとき：

	$\hat{\beta}_D$	$\hat{\beta}_{DX}$
$\hat{\beta}_D$	0.0050	-0.0025
$\hat{\beta}_{DX}$	-0.0025	0.0005

- $\hat{\Sigma}_{11}/N = 0.0050$, $\hat{\Sigma}_{22}/N = 0.0005$, $\hat{\Sigma}_{12}/N = -0.0025$。
- $CATE(10.39)$ に対する 95% 信頼区間は：

例題

- $\hat{\beta}_D = 0.537$, $\hat{\beta}_{DX} = -0.056$ でした。$\hat{\Sigma}/N$ の一部が以下のとき：

	$\hat{\beta}_D$	$\hat{\beta}_{DX}$
$\hat{\beta}_D$	0.0050	-0.0025
$\hat{\beta}_{DX}$	-0.0025	0.0005

- $\hat{\Sigma}_{11}/N = 0.0050$, $\hat{\Sigma}_{22}/N = 0.0005$, $\hat{\Sigma}_{12}/N = -0.0025$。
- $CATE(10.39)$ に対する 95% 信頼区間は：

$$\hat{\beta}_D + \hat{\beta}_{DX} x \pm 1.96 \sqrt{\hat{\Sigma}_{11}/N + x^2 \hat{\Sigma}_{22}/N + 2x\hat{\Sigma}_{12}/N} =$$

例題

- $\hat{\beta}_D = 0.537$, $\hat{\beta}_{DX} = -0.056$ でした。$\hat{\Sigma}/N$ の一部が以下のとき：

	$\hat{\beta}_D$	$\hat{\beta}_{DX}$
$\hat{\beta}_D$	0.0050	-0.0025
$\hat{\beta}_{DX}$	-0.0025	0.0005

- $\hat{\Sigma}_{11}/N = 0.0050$, $\hat{\Sigma}_{22}/N = 0.0005$, $\hat{\Sigma}_{12}/N = -0.0025$。
- $CATE(10.39)$ に対する 95% 信頼区間は：

$$\hat{\beta}_D + \hat{\beta}_{DX}x \pm 1.96\sqrt{\hat{\Sigma}_{11}/N + x^2\hat{\Sigma}_{22}/N + 2x\hat{\Sigma}_{12}/N} =$$
$$(0.537 - 0.056 \times 10.39) \pm$$
$$1.96 \times \sqrt{0.0050 + 10.39^2 \times 0.0005 + 2 \times 10.39 \times (-0.0025)} =$$

例題

- $\hat{\beta}_D = 0.537$, $\hat{\beta}_{DX} = -0.056$ でした。$\hat{\Sigma}/N$ の一部が以下のとき：

	$\hat{\beta}_D$	$\hat{\beta}_{DX}$
$\hat{\beta}_D$	0.0050	-0.0025
$\hat{\beta}_{DX}$	-0.0025	0.0005

- $\hat{\Sigma}_{11}/N = 0.0050$, $\hat{\Sigma}_{22}/N = 0.0005$, $\hat{\Sigma}_{12}/N = -0.0025$。
- $CATE(10.39)$ に対する 95% 信頼区間は：

$$\hat{\beta}_D + \hat{\beta}_{DX}x \pm 1.96\sqrt{\hat{\Sigma}_{11}/N + x^2\hat{\Sigma}_{22}/N + 2x\hat{\Sigma}_{12}/N} =$$

$$(0.537 - 0.056 \times 10.39) \pm$$

$$1.96 \times \sqrt{0.0050 + 10.39^2 \times 0.0005 + 2 \times 10.39 \times (-0.0025)} =$$

$$[-0.21, 0.12]$$

lincom コマンド

- Stata の `lincom` コマンドは、上記の議論を一般化し、係数の任意の線形結合（すなわち $a_1\beta_1 + \cdots + a_k\beta_k$ という形式のパラメータ）をテストできるようにしています。

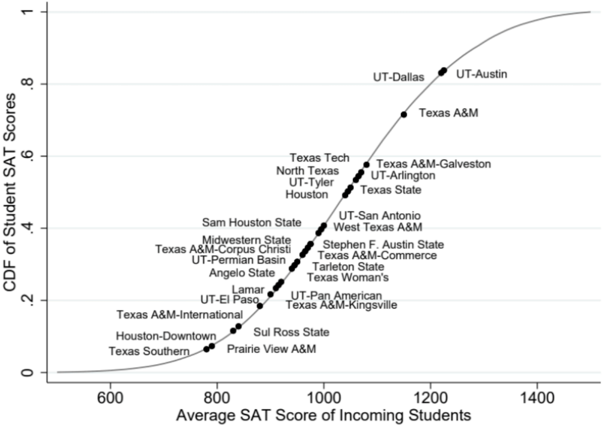
lincom コマンド

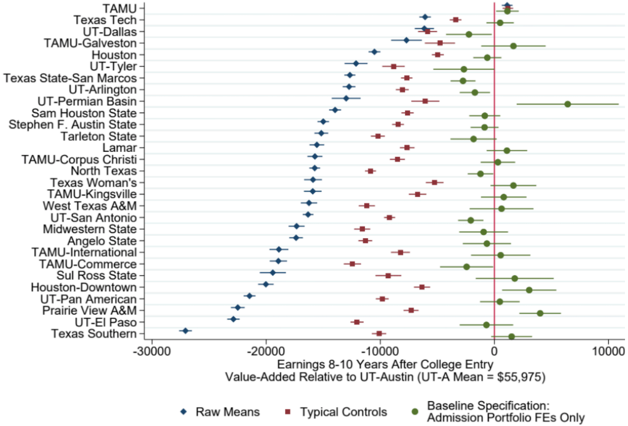
- Stata の `lincom` コマンドは、上記の議論を一般化し、係数の任意の線形結合（すなわち $a_1\beta_1 + \cdots + a_k\beta_k$ という形式のパラメータ）をテストできるようにしています。
- 同様の（しかし少し複雑な）漸近的議論を用いることで、係数の非線形結合（例：$\beta_1\beta_2 + \beta_3^2$）に関する仮説もテストできます。これは Stata の `nlcom` コマンドで実行可能です。

さらなる証拠 – Mountjoy and Hickman (2021)

- Mountjoy and Hickman は、テキサス州のすべての公立 4 年制大学のデータを用いて、Dale and Krueger の現代版を行いました。
- サンプルは、1999 年から 2008 年にテキサス州の高校を卒業した全学生を含みます。
- 非常に大きなサンプルのため、出願・合格した大学の正確なセットをコントロールすることができ、各大学ごとの効果を推定することができました（42.2 万人が 4 年制大学に進学し、そのうち 12.6 万人が複数の大学に合格していました）。
- 合格した中で最も選択度の高い大学に入学した学生は 55

Figure 1: Selectivity Across Colleges and SAT Score Variation Within Colleges



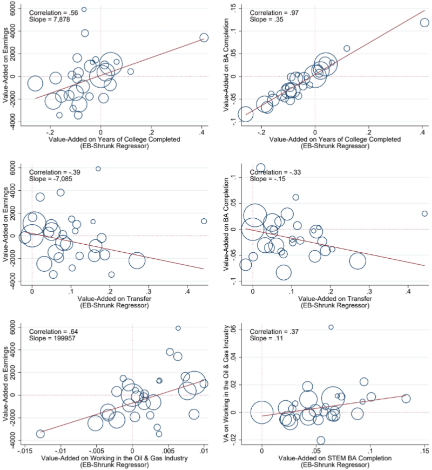


- 大学間で所得に有意な差はありますが、それほど大きくはありません。
- プレステージが高いことが必ずしも高い所得を意味するわけではありません（例：テキサス大オースティン校 vs パーミアンベースン校）。

付加価値の高い大学は何をしているのか

大学卒業率が高い、STEM 分野の卒業率が高い、石油・ガス産業への就職率が高いなど

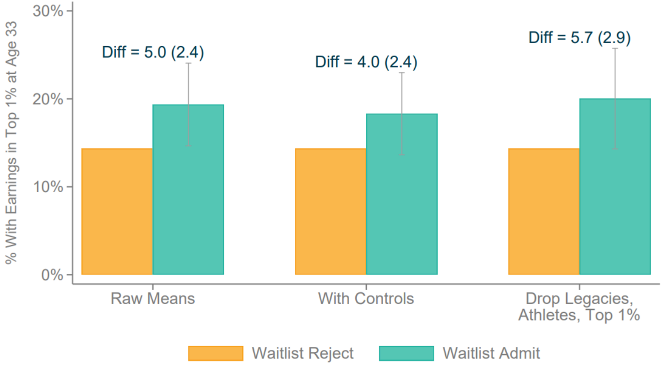
Figure 15: Other Potential Mechanisms: Persistence, Transfer, and Industry of Employment



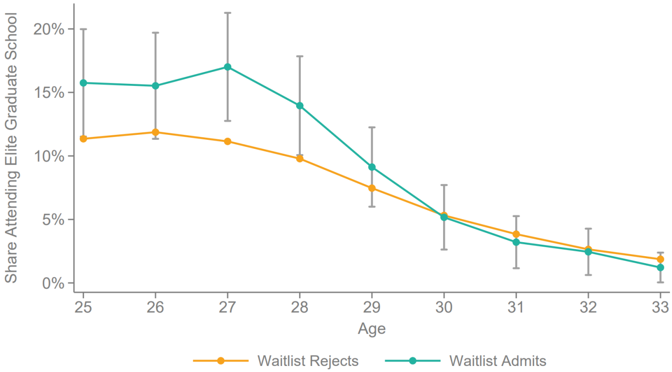
Chetty, Deming, Friedman (2025)

- いくつかの Ivy+ 校（アイビーリーグ、スタンフォード、MIT、シカゴ、デューク）の合格データと納税所得データを組み合わせて分析しました。
- 補欠合格（ウェイトリスト）から合格した学生と、合格しなかった学生を比較しました。
  - ある Ivy+ 校での補欠合格は、他の Ivy+ 校での結果を予測しません。
- Dale and Krueger と同様に、彼らは平均所得に対する大きな影響は見出しませんでした。
- しかし、「エリート」なアウトカムに対しては影響がありました。所得上位 1

(a) Earnings in Top 1% at Age 33



(b) Elite Graduate School Attendance



アウトライン

1. 重回帰と OLS の導出 ✓
2. 回帰と因果関係 ✓
3. 回帰に関するその他のトピック

その他のトピック 1：多重検定（多重比較）

- グループ間での因果的効果の違いをテストする方法を見てきました。

その他のトピック 1：多重検定（多重比較）

- グループ間での因果的効果の違いをテストする方法を見てきました。
- ここでは「貧しい学生 vs 裕福な学生」という 1 つの次元の異質性を考えました。しかし、他のグループについての異質性にも興味があるかもしれません。
  - 男性の方が女性よりも効果が大きいか？

その他のトピック 1：多重検定（多重比較）

- グループ間での因果的効果の違いをテストする方法を見てきました。
- ここでは「貧しい学生 vs 裕福な学生」という 1 つの次元の異質性を考えました。しかし、他のグループについての異質性にも興味があるかもしれません。
  - 男性の方が女性よりも効果が大きいか？
  - アスリートの方が非アスリートよりも効果が大きいか？

その他のトピック 1：多重検定（多重比較）

- グループ間での因果的効果の違いをテストする方法を見てきました。
- ここでは「貧しい学生 vs 裕福な学生」という 1 つの次元の異質性を考えました。しかし、他のグループについての異質性にも興味があるかもしれません。
  - 男性の方が女性よりも効果が大きいか？
  - アスリートの方が非アスリートよりも効果が大きいか？
  - 裕福な家庭の男性アスリートの方が、貧しい家庭の女性非アスリートよりも効果が大きいか？

その他のトピック 1：多重検定（多重比較）

- グループ間での因果的効果の違いをテストする方法を見てきました。
- ここでは「貧しい学生 vs 裕福な学生」という 1 つの次元の異質性を考えました。しかし、他のグループについての異質性にも興味があるかもしれません。
  - 男性の方が女性よりも効果が大きいか？
  - アスリートの方が非アスリートよりも効果が大きいか？
  - 裕福な家庭の男性アスリートの方が、貧しい家庭の女性非アスリートよりも効果が大きいか？
- テストしたい仮説の数は非常に多くなる可能性があります！

その他のトピック 1：多重検定（多重比較）

- グループ間での因果的効果の違いをテストする方法を見てきました。
- ここでは「貧しい学生 vs 裕福な学生」という 1 つの次元の異質性を考えました。しかし、他のグループについての異質性にも興味があるかもしれません。
  - 男性の方が女性よりも効果が大きいか？
  - アスリートの方が非アスリートよりも効果が大きいか？
  - 裕福な家庭の男性アスリートの方が、貧しい家庭の女性非アスリートよりも効果が大きいか？
- テストしたい仮説の数は非常に多くなる可能性があります！
- これは「多重検定（multiple hypothesis testing）」問題と呼ばれる状況を引き起こします。

多重検定（続き）

- 思い出してください：p 値は、（漸近的に）処置効果がないという帰無仮説の下で、$p < 0.05$ となる確率が 5

多重検定（続き）

- 思い出してください：p 値は、（漸近的に）処置効果がないという帰無仮説の下で、$p < 0.05$ となる確率が 5
- したがって、全人口に対して有意な効果があるかテストする場合、真に効果がなければ、棄却される確率は 5

多重検定（続き）

- 思い出してください：p 値は、（漸近的に）処置効果がないという帰無仮説の下で、$p < 0.05$ となる確率が 5
- したがって、全人口に対して有意な効果があるかテストする場合、真に効果がなければ、棄却される確率は 5
- しかし、まず男性について有意な効果をテストし、次に女性についてもテストしたとしましょう。

多重検定（続き）

- 思い出してください：p 値は、（漸近的に）処置効果がないという帰無仮説の下で、$p < 0.05$ となる確率が 5
- したがって、全人口に対して有意な効果があるかテストする場合、真に効果がなければ、棄却される確率は 5
- しかし、まず男性について有意な効果をテストし、次に女性についてもテストしたとしましょう。
- どちらのグループにも有意な効果が全くない場合、少なくとも 1 つの有意な結果が見つかる確率はどの程度でしょうか？

- 男性と女性のサンプルが独立に抽出されているとします。
- 有意な結果の数を N_{sig} とします。

$$P(N_{sig} \geq 1) =$$

- 男性と女性のサンプルが独立に抽出されているとします。
- 有意な結果の数を N_{sig} とします。

$$P(N_{sig} \geq 1) = 1 - P(N_{sig} = 0)$$

- 男性と女性のサンプルが独立に抽出されているとします。
- 有意な結果の数を N_{sig} とします。

$$
\begin{aligned}
P(N_{sig} \geq 1) &= 1 - P(N_{sig} = 0) \\
&= 1 - P(\text{女性が有意でない かつ 男性が有意でない})
\end{aligned}
$$

- 男性と女性のサンプルが独立に抽出されているとします。
- 有意な結果の数を N_{sig} とします。

$$
\begin{aligned}
&P(N_{sig} \geq 1) = 1 - P(N_{sig} = 0) \\
&= 1 - P(\text{女性が有意でない かつ 男性が有意でない}) \\
&= 1 - P(\text{女性が有意でない})P(\text{男性が有意でない})
\end{aligned}
$$

- 男性と女性のサンプルが独立に抽出されているとします。
- 有意な結果の数を N_{sig} とします。

$$
\begin{aligned}
&P(N_{sig} \geq 1) = 1 - P(N_{sig} = 0) \\
&= 1 - P(\text{女性が有意でない かつ 男性が有意でない}) \\
&= 1 - P(\text{女性が有意でない})P(\text{男性が有意でない}) \\
&= 1 - .95^2
\end{aligned}
$$

- 男性と女性のサンプルが独立に抽出されているとします。
- 有意な結果の数を N_{sig} とします。

$$
\begin{aligned}
&P(N_{sig} \geq 1) = 1 - P(N_{sig} = 0) \\
&= 1 - P(\text{女性が有意でない かつ 男性が有意でない}) \\
&= 1 - P(\text{女性が有意でない})P(\text{男性が有意でない}) \\
&= 1 - .95^2 = 0.0975
\end{aligned}
$$

- したがって、どちらのグループにも真の効果がない場合でも、約 10

- 男性と女性のサンプルが独立に抽出されているとします。
- 有意な結果の数を N_{sig} とします。

$$
\begin{aligned}
&P(N_{sig} \geq 1) = 1 - P(N_{sig} = 0) \\
&= 1 - P(\text{女性が有意でない かつ 男性が有意でない}) \\
&= 1 - P(\text{女性が有意でない})P(\text{男性が有意でない}) \\
&= 1 - .95^2 = 0.0975
\end{aligned}
$$

- したがって、どちらのグループにも真の効果がない場合でも、約 10
- 同じ論理で、K 個の独立なグループについて帰無仮説をテストし、真の効果がすべて 0 である場合、帰無仮説を棄却してしまう確率は

- 男性と女性のサンプルが独立に抽出されているとします。
- 有意な結果の数を N_{sig} とします。

$$
\begin{aligned}
&P(N_{sig} \geq 1) = 1 - P(N_{sig} = 0) \\
&= 1 - P(\text{女性が有意でない かつ 男性が有意でない}) \\
&= 1 - P(\text{女性が有意でない})P(\text{男性が有意でない}) \\
&= 1 - .95^2 = 0.0975
\end{aligned}
$$

- したがって、どちらのグループにも真の効果がない場合でも、約 10
- 同じ論理で、K 個の独立なグループについて帰無仮説をテストし、真の効果がすべて 0 である場合、帰無仮説を棄却してしまう確率は $1 - 0.95^K$ となります。

K 個の独立なグループ（処置効果は 0）において、少なくとも 1 つのサブグループで有意な効果が見つかる確率：

K	$1 - 0.95^K$
1	0.05
2	0.0975

K 個の独立なグループ（処置効果は 0）において、少なくとも 1 つのサブグループで有意な効果が見つかる確率：

K	$1-0.95^K$
1	0.05
2	0.0975
3	

K 個の独立なグループ（処置効果は 0）において、少なくとも 1 つのサブグループで有意な効果が見つかる確率：

K	$1-0.95^K$
1	0.05
2	0.0975
3	0.1426
5	

K 個の独立なグループ（処置効果は 0）において、少なくとも 1 つのサブグループで有意な効果が見つかる確率：

K	$1 - 0.95^K$
1	0.05
2	0.0975
3	0.1426
5	0.2262
10	

K 個の独立なグループ（処置効果は 0）において、少なくとも 1 つのサブグループで有意な効果が見つかる確率：

K	$1-0.95^K$
1	0.05
2	0.0975
3	0.1426
5	0.2262
10	0.4013
100	

K 個の独立なグループ（処置効果は 0）において、少なくとも 1 つのサブグループで有意な効果が見つかる確率：

K	$1-0.95^K$
1	0.05
2	0.0975
3	0.1426
5	0.2262
10	0.4013
100	0.9941

多重検定のシミュレーション

- 「Current Population Survey」の平均時給の調査データがあります。
- 確率 $1/2$ で 1 となる偽の処置 D_i を生成します。

多重検定のシミュレーション

- 「Current Population Survey」の平均時給の調査データがあります。
- 確率 $1/2$ で 1 となる偽の処置 D_i を生成します。
- この処置の真の因果的効果は何でしょうか？

多重検定のシミュレーション

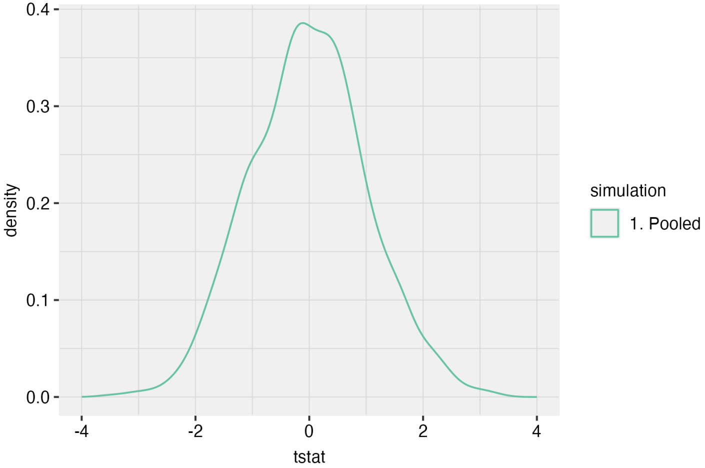
- 「Current Population Survey」の平均時給の調査データがあります。
- 確率 $1/2$ で 1 となる偽の処置 D_i を生成します。
- この処置の真の因果的効果は何でしょうか？ もちろん 0 です。

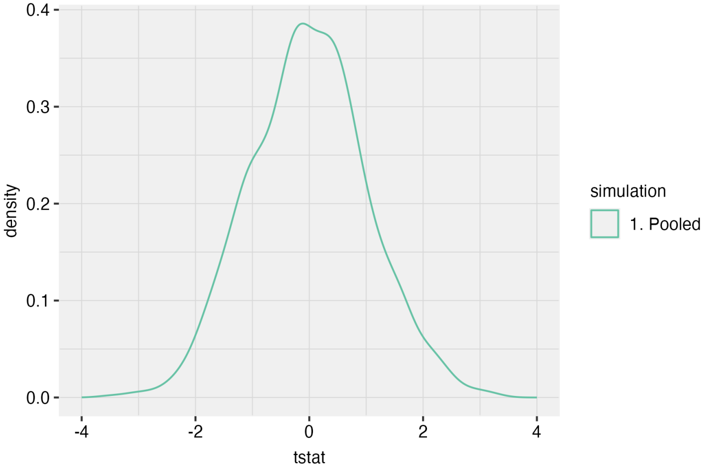
多重検定のシミュレーション

- 「Current Population Survey」の平均時給の調査データがあります。
- 確率 1/2 で 1 となる偽の処置 D_i を生成します。
- この処置の真の因果的効果は何でしょうか？ もちろん 0 です。
- この偽の処置を 1000 回シミュレーションし、以下を推定します：
  1. 全州をプールした処置効果
  2. 最初の 10 州それぞれの処置効果
  3. 全 50 州それぞれの処置効果

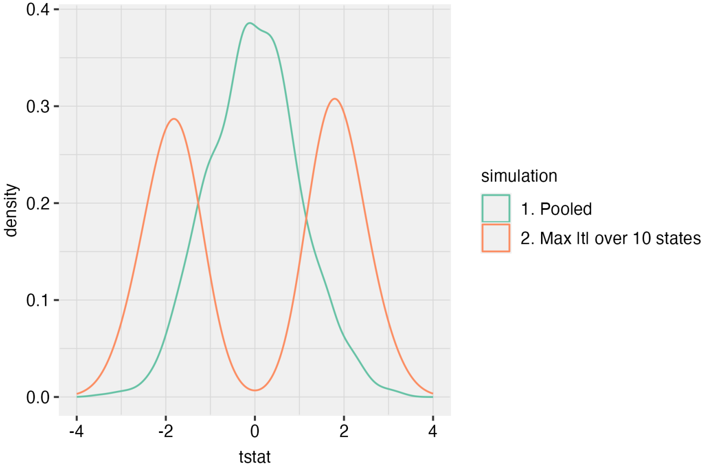
多重検定のシミュレーション

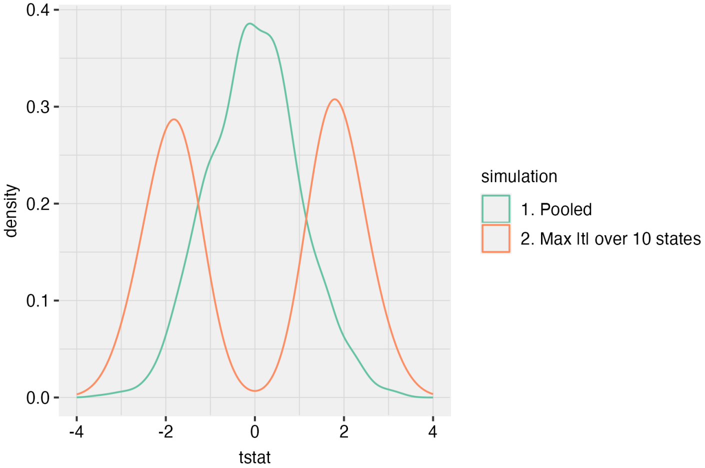
- 「Current Population Survey」の平均時給の調査データがあります。
- 確率 1/2 で 1 となる偽の処置 D_i を生成します。
- この処置の真の因果的効果は何でしょうか？ もちろん 0 です。
- この偽の処置を 1000 回シミュレーションし、以下を推定します：
  1. 全州をプールした処置効果
  2. 最初の 10 州それぞれの処置効果
  3. 全 50 州それぞれの処置効果
- そして、少なくとも 1 つの有意な推定値が得られるシミュレーションの割合を計算します。


0.4
0.3
0.2
0.1
0.0
density
-4
-2
0
2
4
tstat
simulation
1. Pooled

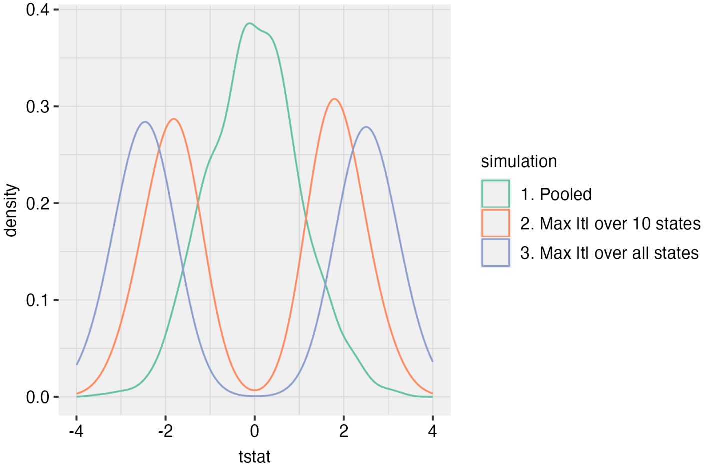


- 全州をプールしてテストした場合、6

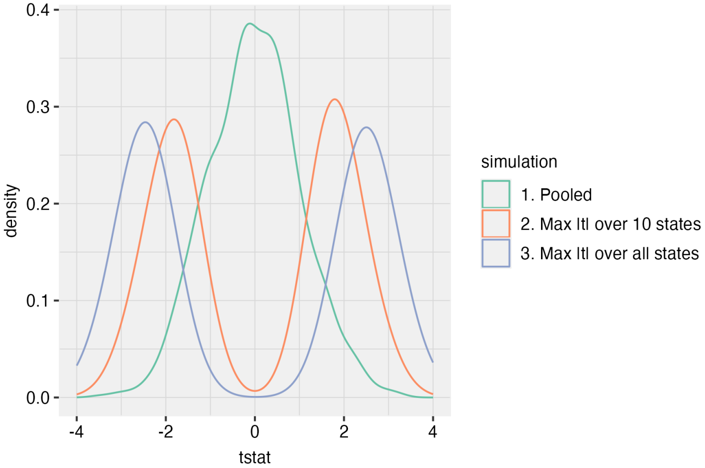

0.4
0.3
0.2
0.1
0.0
density
-4
-2
0
2
4
tstat
simulation
1. Pooled
2. Max |t| over 10 states



- 10 州を個別にテストした場合、41



0.4
0.3
0.2
0.1
0.0
density
-4
-2
0
2
4
tstat
simulation
1. Pooled
2. Max |t| over 10 states
3. Max |t| over all states



- 50 州を個別にテストした場合、93

多重検定の補正

- 多重検定の最も単純な補正方法は、ボンフェローニ（Bonferroni）補正です。

多重検定の補正

- 多重検定の最も単純な補正方法は、ボンフェローニ（Bonferroni）補正です。

- K 個の個別の仮説を 5

多重検定の補正

- 多重検定の最も単純な補正方法は、ボンフェローニ（Bonferroni）補正です。
- K 個の個別の仮説を 5
- すると：

$$P(N_{sig} > 0) \leq \sum_k P(\text{k が有意}) = K(0.05/K) = 0.05$$

多重検定の補正

- 多重検定の最も単純な補正方法は、ボンフェローニ（Bonferroni）補正です。
- K 個の個別の仮説を 5
- すると：

$$P(N_{sig} > 0) \leq \sum_k P(\text{k が有意}) = K(0.05/K) = 0.05$$

- これは仮説が独立でない場合（例：グループが重複している場合）でも機能します。

多重検定の補正

- 多重検定の最も単純な補正方法は、ボンフェローニ（Bonferroni）補正です。
- K 個の個別の仮説を 5
- すると：

$$P(N_{sig} > 0) \leq \sum_k P(\text{k が有意}) = K(0.05/K) = 0.05$$

- これは仮説が独立でない場合（例：グループが重複している場合）でも機能します。
- 欠点：仮説の数が多いと、個々の仮説に対する検出力が非常に低くなります（例：99.9 また、このテストは一般に保守的であり、有意な効果が見つかる確率は 5

ボンフェローニ補正なし、および補正ありで、少なくとも 1 つの仮説を棄却する確率：

シミュレーション	補正なし	補正あり
プール	0.06	0.05
10 州	0.41	0.05
50 州	0.93	0.04

その他のトピック 2：結合仮説

- 時として、いずれかのサブグループ（例：いずれかの州）で効果があるかどうかだけを知りたい場合があります。

その他のトピック 2：結合仮説

- 時として、いずれかのサブグループ（例：いずれかの州）で効果があるかどうかだけを知りたい場合があります。
- すべてのサブグループで処置効果が 0 である、という結合帰無仮説をテストすることができます。

その他のトピック 2：結合仮説

- 時として、いずれかのサブグループ（例：いずれかの州）で効果があるかどうかだけを知りたい場合があります。
- すべてのサブグループで処置効果が 0 である、という結合帰無仮説をテストすることができます。
- もしこれを棄却できれば、少なくとも 1 つのサブグループで処置効果が 0 ではない強い証拠があると結論づけられます。

その他のトピック 2：結合仮説

- 時として、いずれかのサブグループ（例：いずれかの州）で効果があるかどうかだけを知りたい場合があります。
- すべてのサブグループで処置効果が 0 である、という結合帰無仮説をテストすることができます。
- もしこれを棄却できれば、少なくとも 1 つのサブグループで処置効果が 0 ではない強い証拠があると結論づけられます。
- どのようにこれを行うのでしょうか？

- $\hat{t}_k$ を仮説 k の t 統計量とします。帰無仮説 k の下で、$\hat{t}_k \to_d$

- $\hat{t}_k$ を仮説 k の t 統計量とします。帰無仮説 k の下で、$\hat{t}_k \to_d N(0,1)$ です。

- 各仮説のサンプルが独立である場合、K 個すべての帰無仮説が満たされていれば、$(\hat{t}_1, \ldots, \hat{t}_K)' \to_d N(0, \mathbf{I})$ となります。

- F 統計量 $F = \sum_k \hat{t}_k^2$ を考えます。

- $\hat{t}_k$ を仮説 k の t 統計量とします。帰無仮説 k の下で、$\hat{t}_k \to_d N(0,1)$ です。

- 各仮説のサンプルが独立である場合、K 個すべての帰無仮説が満たされていれば、$(\hat{t}_1, \ldots, \hat{t}_K)' \to_d N(0, \mathbf{I})$ となります。

- F 統計量 $F = \sum_k \hat{t}_k^2$ を考えます。連続写像定理より、K 個すべての帰無仮説が満たされていれば：

$$F \to_d \sum_k Z_k^2 \text{（ただし } Z \sim N(0, \mathbf{I})\text{）}$$

  これは、自由度 K のカイ二乗（χ^2）分布と呼ばれます。

- したがって、F をカイ二乗分布の 95

- このアプローチは F 検定と呼ばれます。

すべての州で処置効果が 0 であるという結合帰無仮説を棄却する確率：

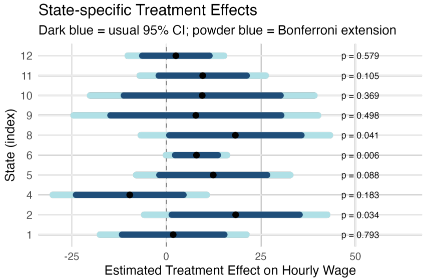
シミュレーション	個別のテスト	結合テスト
プール	0.06	0.06
10 州	0.41	0.04
50 州	0.93	0.05

F 検定はボンフェローニよりも検出力が高い場合がある

- 偽の処置を受けた場合に賃金に 6 ドル加算されるようにシミュレーション設定を変更します。
- ボンフェローニを用いると、少なくとも 1 つの州で有意な効果が見つかるのは 39

F 検定はボンフェローニよりも検出力が高い場合がある

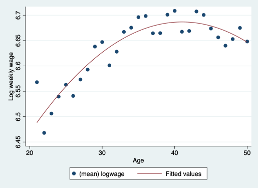
- 偽の処置を受けた場合に賃金に 6 ドル加算されるようにシミュレーション設定を変更します。
- ボンフェローニを用いると、少なくとも 1 つの州で有意な効果が見つかるのは 39
- あるシミュレーションの例です：



ボンフェローニ補正後、州レベルの効果はどれも有意ではありません。しかし、結合テストの p 値は 0.004 です。

複数の回帰係数のテスト

- 複数の回帰係数が 0 であるという帰無仮説をテストしたい場合があります（例：所得と年齢の例で、線形項と 2 次項の両方が 0 であるか)。



- この問題に対しても F 検定を使用できます。係数間の相関を考慮に入れる必要があります。

- 以前の講義で $\sqrt{N}(\hat{\beta}-\beta_0) \rightarrow_d N(0,\mathbf{\Sigma})$ を示しました。帰無仮説 $H_0 : \beta = \beta_0$ をテストするには、$F = \sqrt{N}\hat{\mathbf{\Sigma}}^{-\frac{1}{2}}(\hat{\beta}-\beta_0)$ を使用します。

- これは Stata の `test` 関数で実装されています。例えば `test age agesq` は、年齢と年齢の 2 乗の係数がともに 0 である（＝所得が年齢に依存しない）という帰無仮説をテストします。

```
. reg logwage age agesq, r

Linear regression                               Number of obs     =         30
                                                F(2, 27)          =      30.81
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8412
                                                Root MSE          =     .02622

------------------------------------------------------------------------------
             |               Robust
     logwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0403444   .0076871     5.25   0.000     .0245718     .056117
       agesq |   -.000492   .0001011    -4.87   0.000    -.0006994   -.0002846
       _cons |   5.859108   .1409008    41.58   0.000     5.570003    6.148212
------------------------------------------------------------------------------

.
. test age agesq

 ( 1)  age = 0
 ( 2)  agesq = 0

       F(  2,    27) =   30.81
            Prob > F =    0.0000
```