講義の超詳細まとめ：欠落変数の問題と因果関係の推定

今回の講義は、主に「欠落変数バイアス」の問題と、その発展としての「因果関係の推定」という2つの大きなテーマで構成されていました。

1. 欠落変数バイアス（Omitted Variable Bias: OVB）

a. 問題の所在

- 実際の分析では、どの変数を説明変数（X）としてモデルに含めるかが重要になる。
- 特に関心のある1つのX（例：クラスの大きさ）がY（例：テストの点数）に与える影響を知りたい場合が多い。
- しかし、関心のあるXとYだけで回帰分析を行うと、間違った結果を導く可能性が非常に高い。

b. 欠落変数の定義 講義では、問題を引き起こす「欠落変数」を以下の2つの条件を満たすものとして定義しました。

1. 被説明変数 Y に対して影響を与えているが、現在のモデルの説明変数として含まれていない（= 省略されている）。
2. モデルに既に含まれている説明変数 X と相関がある。

- 具体例（クラスの大きさと学力）：
  - モデル：テストスコア = $\beta_0 + \beta_1 \cdot STR$ (生徒教師比率) + U
  - 関心：STRがテストスコアに与える影響（β_1）。
  - 欠落変数（Uに含まれる）：教師の質、生徒の家庭環境、教育設備など。
  - 問題: これらの欠落変数は、テストスコア（Y）に影響を与える（条件1）。さらに、STR（X）とも相関する可能性が高い（例：裕福な地域はSTRが低く、家庭環境も良い）（条件2）。

c. OVBがもたらす数学的な問題

- 真のモデルが $Y = X\beta + Z\gamma + U$ であるとします（Zが欠落変数）。
- しかし、分析者が $Y = X\beta + V$ というモデルを（Zを省略して）推定した場合、推定量 $\hat{\beta}$ は以下のようになります。$\hat{\beta} = (X'X)^{-1}X'Y$
- この Y に真のモデルを代入すると、$\hat{\beta} = \beta + (X'X)^{-1}X'Z\gamma + (X'X)^{-1}X'U$
- この推定量の期待値（あるいは確率極限）を考えると、$\hat{\beta}$ は $\beta + (X'X)^{-1}X'Z\gamma$ に収束します。
- $(X'X)^{-1}X'Z\gamma$ の部分がバイアスです。
- このバイアスは、以下の理由で発生します。
  1. Z と X に相関がある（$X'Z \neq 0$）。
  2. Z が Y に影響を与える（$\gamma \neq 0$）。
- もし X と Z の相関がゼロ（$X'Z = 0$）であれば、たとえ Z を省略してもバイアスは発生しません（これが条件2が重要な理由です）。
- このバイアスは、サンプルサイズ（n）を大きくしても消えないため、$\hat{\beta}$ は一致性（**consistency**）を失います。

d. OVBへの対処法**1**：条件付き平均の独立性（**CMI**）

- 理想は Z のデータを取得してモデルに含めることですが、データがない場合もあります。
- もし関心があるのが X の係数（β_1）だけで、他の変数（W）の係数はどうでもよい場合、より緩い仮定で β_1 を正しく推定できる可能性があります。
- **CMI**の仮定: $E[V | X, W] = g(W)$
  - V は省略された変数 Z を含む誤差項です。
  - この仮定は、「W を条件付け（コントロール）すれば、誤差項 V の期待値は X には依存しない」ということを意味します。

- 結果:
    - 真のモデルが $Y = \beta_0 + \beta_1 X + \alpha W + \delta Z + U$ で、Z を省略した $V = \delta Z + U$ を考えます。
    - もし $E[V | X, W] = \gamma_0 + \gamma_1 W$ のように W だけの関数で書ける（CMIが成立する）場合、推定するモデルは $Y = (\beta_0 + \gamma_0) + \beta_1 X + (\alpha + \gamma_1) W + \epsilon$ と書き換えられます。
    - この新しい誤差項 ϵ は、$E[\epsilon | X, W] = 0$ を満たします（OLSの仮定1が成立）。
    - その結果、X の係数 β_1 は正しく（一致性を持って）推定できます。
    - ただし、W の係数は $(\alpha + \gamma_1)$ というバイアスのかかった値として推定されます。

2. 回帰モデルの解釈（各種）

a. 2項変数（ダミー変数）モデル
- D を 0 または 1 の値をとるダミー変数（例：男性=0, 女性=1）とします。
- モデル： $Y = \beta_0 + \beta_1 D + U$
- $E[Y | D=0] = \beta_0$ （D=0 のグループの Y の平均値）
- $E[Y | D=1] = \beta_0 + \beta_1$ （D=1 のグループの Y の平均値）
- したがって、$\beta_1 = E[Y | D=1] - E[Y | D=0]$ となり、β_1 は**2**つのグループ間の平均値の差を意味します。

b. 多項式モデルと対数変換
- X^2, X^3 などをモデルに含めることで、非線形の関係を捉えることができます。
- 対数変換（$\log(Y)$ や $\log(X)$）もよく使われます。
    - 経済変数の多く（GDP、消費など）は正の値しか取らないため、対数をとることで分布を正規分布に近づけることができます。
    - $\log(Y)$ と $\log(X)$ の両方を使う「log-logモデル」では、係数は「弾力性（Xが1%変化した時にYが何%変化するか）」として解釈できます。

c. 相互作用項（交差項）モデル
- $X_1 \times X_2$ のような変数の掛け算をモデルに含めます。
- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + U$
- このモデルでは、X_1 が Y に与える影響（$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$）が、X_2 の値によって変化するようになります。
- これは「差分の差分法（DID）」などでも利用されます。

3. 因果関係（**Causal Inference**）の推定

a. OLSは「相関」であり「因果」ではない
- OLS（最小二乗法）が推定する β は、本質的に X と Y の共分散（相関）を見ています。
- 相関関係は因果関係を意味しません（例：身長と体重。身長が高いから体重が重い、とは必ずしも言えない）。
- OLSが因果効果を推定するための最も重要な仮定は「仮定1：$E[U|X] = 0$」です。これは、説明変数 X と誤差項 U が無相関である（Xが外生的である）ことを意味します。

b. RCT（ランダム化比較試験）**vs** 観察データ
- 例：肥料（**X**）と収穫量（**Y**）
    - U（誤差項）には「土地の豊かさ」「日当たり」などが含まれます。
- **RCT (**実験データ**)**:
    - 研究者が、各農地（区画）に与える肥料の量（X）をランダムに割り当てます。

    - ランダム化により、X（肥料）は U（土地の豊かさ）と**無関係（独立）**になります。
    - $E[U|X] = 0$ が成立するため、OLSで推定した β は X が Y に与える因果効果となります。
- 観察データ:
    - 農家が、自分の判断で肥料の量（X）を決定します。
    - 農家は「日当たりが悪い・土地が痩せている（Uが悪い）場所」に、より多くの肥料（**X**）を与える可能性があります。
    - この場合、X と U が相関してしまい（$E[U|X] \neq 0$）、OLSの結果はバイアス（内生性の問題）を持ち、因果関係とは言えません。
- **RCT**の限界: 社会科学ではRCTの実施が困難な場合があります（例：日銀が為替介入をランダムに行うのはコストが大きすぎる、倫理的な問題（ECMOの実験例など））。

c. ルービンの因果モデル（**RCM**）**/** 潜在的結果

- 因果関係を定義するための枠組みです。
- D_i: 個人 i が処置（例：職業訓練）を受けたかどうか（受けた=1, 受けてない=0）。
- $Y_i(1)$: 個人 i がもし処置を受けた場合の結果（賃金）。
- $Y_i(0)$: 個人 i がもし処置を受けなかった場合の結果（賃金）。
- 因果効果（個人）: $\tau_i = Y_i(1) - Y_i(0)$
- 根本的な問題: 各個人 i について、$Y_i(1)$ と $Y_i(0)$ の片方しか観測できません。もし処置を受けたら（$D_i=1$） $Y_i(1)$ が観測され、$Y_i(0)$ は観測されません（反実仮想）。

d. 推定したい主要なパラメータ

1. **ATE (Average Treatment Effect;** 平均処置効果**)**: $E[\tau_i] = E[Y(1) - Y(0)]$
    - 集団全体（訓練を受けた人＋受けなかった人）に対する処置の平均的な効果。
2. **ATT (Average Treatment effect on the Treated;** 処置群の平均処置効果**)**: $E[\tau_i | D=1] = E[Y(1) - Y(0) | D=1]$
    - 実際に処置を受けた人たちにとっての平均的な効果。
3. CATE (Conditional ATE): 特定の共変量 X を持つ集団での平均効果。
4. LATE (Local ATE): 特定の操作変数 Z に反応した人たち（Compliers）への平均効果。

e. 因果効果の推定方法

- ケース**1**：**RCT**が実施された場合
    - ランダム化により、「処置の割り当て D」と「潜在的結果 (Y(0), Y(1))」は独立になります。
    - この独立性（$D \perp (Y(0), Y(1))$）のおかげで、$ATE = E[Y(1)] - E[Y(0)] = E[Y(1) | D=1] - E[Y(0) | D=0]$
    - 観測される Y は $E[Y | D=1] = E[Y(1) | D=1]$、$E[Y | D=0] = E[Y(0) | D=0]$ なので、$ATE = E[Y | D=1] - E[Y | D=0]$
    - 結論: RCTの下では、ATEは「処置グループの Y の平均値」から「非処置グループの Y の平均値」を単純に引き算するだけで求められます。
    - これは $Y_i = \beta_0 + \tau D_i + U_i$ という単純なOLS回帰でも推定でき、τ がATEになります。
- ケース**2**：観察データの場合（交絡変数の調整）
    - RCTがないため、D と (Y(0), Y(1)) は独立ではありません。
    - 交絡変数（**Confounder**）: 処置 D と 結果 (Y(0), Y(1)) の両方に影響を与える変数 X（例：訓練前の収入、学歴、やる気など）。
    - 対処法のための仮定:
        1. 条件付き独立の仮定 **(CIA) / Unconfoundedness**: $(Y(0), Y(1)) \perp D \mid X$（交絡変数 X を条件付ければ（= X の値が同じ人たちの中では）、処置 D はランダムに割り当てられたかのように独立である）
        2. オーバーラップ **(Overlap / Common Support)**: $0 < P(D=1 | X=x) < 1$ （X のど

      の値においても、処置を受けた人と受けていない人の両方が存在する）
  - 推定（回帰による調整）:
    - CIAの仮定のもとでは、X を条件としたATEは \tau(X) = E[Y(1) | X] - E[Y(0) | X] = E[Y | D=1, X] - E[Y | D=0, X]
    - ATEは、この \tau(X) を X について平均（期待値）をとったもの（E_X[\tau(X)]）になります。
    - 具体的な推定手順:
      1. m_1(X) = E[Y | D=1, X] を推定する。（例：D=1 のサンプルだけを使い、Y を X に回帰する）
      2. m_0(X) = E[Y | D=0, X] を推定する。（例：D=0 のサンプルだけを使い、Y を X に回帰する）
      3. \tau(X) = m_1(X) - m_0(X) を計算する。
      4. これを全サンプルで平均したものがATEの推定量となる。
    - （講義はここで終了し、次回は傾向スコア・マッチングなどの話に移るようでした。）