🚀 講義のまとめ：因果関係の推定と操作変数法

第1部：因果関係の推定

まず前半は、ある介入（例：職業訓練）が結果（例：賃金）にどないな影響を与えるか、その「因果関係」をどないして調べるか、という話でした。

1. 基本的な枠組み：潜在的結果モデル

- 潜在的結果（**Potential Outcome**）：
    - 介入を受けた場合の結果を y_1、受けなかった場合の結果を y_0 と定義します。これはあくまで「潜在的」な変数です。
    - 例えば、i さんが訓練を受けたら（D_i=1）y_i = y_{1i} が観測され、y_{0i} は観測されへん。逆に訓練を受けへんかったら（D_i=0）y_i = y_{0i} が観測され、y_{1i} は観測されへん。
    - この関係は、一つの式 y_i = (1-D_i)y_{0i} + D_i y_{1i} (式73) とも書けます。
- 推定したい指標:
    - **ATE (Average Treatment Effect; 平均処置効果)**:
        - 集団全体の平均的な効果 ATE = E[y_1 - y_0] (式68)。
        - これが0より大きければ、介入には平均的に効果があるっちゅうことです。
    - **ATT (Average Treatment Effect on the Treated; 処置群の平均処置効果)**:
        - 介入を「受けた人たち」における平均的な効果 ATT = E[y_1 - y_0 | [span_6](start_span)D=1]。
    - **ATU (Average Treatment Effect on the Untreated)**:
        - 介入を「受けなかった人たち」が、もし受けていたらどうやったか、という効果 E[y_1 - y_0 | [span_7](start_span)D=0] も定義上は考えられます。

2. 推定方法 (1)：RCT（ランダム化比較試験）

一番ありがたい、理想的な状況です。

- 仮定: 介入 D（職業訓練を受けるかどうか）と、潜在的結果 y_1, y_0 が独立であること。
    - これは「この人は効果が出そう」とか「この人はあかんそう」とか関係なく、サイコロを振るみたいにランダムに介入を割り当てた状況（RCT）で成立します。
- 推定:
    - この仮定があると、E[y_1] = E[y_1 | D=1] = E[y | [span_11](start_span)D=1] となります。
    - 同様に、E[y_0] = E[y_0 | D=0] = E[y | [span_12](start_span)D=0] となります。
    - その結果、ATEは「介入を受けた人の平均 y」と「受けなかった人の平均 y」の単純な差で推定できます。ATE = E[y | D=1] - E[y | D=0] (式81)
- 回帰分析での推定:
    - y_i = \beta_0 + \tau D_i + u_i (式84の変形) という回帰式を考えます。
    - RCTの仮定（y_1, y_0 が D と独立）があれば、誤差項 u と D が直交し（E[u|D]=0）、OLS（最小二乗法）で推定した係数 \tau が ATE と一致します。

3. 推定方法 (2)：RCT以外（交絡変数がある場合）

現実の社会科学ではRCTは難しく、介入 D と結果 y の両方に影響を与える交絡変数 **X**（例：訓練前

の収入や能力）が存在します。

- 必要な**2**つの仮定:
    1. 条件付き独立性の仮定 **(CIA)**: X の値を固定すれば（X を条件づければ）、D と y_1, y_0 は独立になる。
    2. オーバーラップの仮定: どんな X の値であっても、介入を受ける確率 $P(D=1|X)$ が 0 や 1 にはならへん。（※後述の傾向スコア法で分母が0になるのを防ぐため）

この仮定のもとで、主に3つの推定方法が紹介されました。

① 回帰による調整 **(Regression Adjustment)**

- 考え方:
    - $M_1(X) = E[y_1 | X=x]$（X の値が x の人の y_1 の期待値）
    - $M_0(X) = E[y_0 | X=x]$（X の値が x の人の y_0 の期待値）
    - これらを推定できれば、$ATE = E_X [M_1(X) - M_0(X)]$ として、全 X について平均すればATEが出ます。
- 推定ステップ:
    1. D=1（介入あり）のグループだけを使って、y を X に回帰し、$M_1(X)$ を推定します。
    2. D=0（介入なし）のグループだけを使って、y を X に回帰し、$M_0(X)$ を推定します。
    3. 推定した $\hat{M}_1(X)$ と $\hat{M}_0(X)$ の差を、全サンプルで平均します。
- 特徴:
    - y と X の関係（例：線形 $y = \alpha + \beta X$）を正しく仮定（特定化）する必要があります。これが間違うと推定結果も間違うてしまいます。
    - ただし、オーバーラップの仮定は不要です。

② 傾向スコア・マッチング **(Propensity Score Matching / IPW)**

- 傾向スコア **(Propensity Score)**: $P(X) = P(D=1 | X=x)$。
    - その人の X の値に基づき、介入を受ける確率を定義したものです。
- 考え方:
    - $E[y_1]$ と $E[y_0]$ を、傾向スコアの逆数で重み付けする（IPW: Inverse Probability Weighting）ことで推定します。
    - 導出の結果、以下の関係が成り立ちます。
        - ■
        - ■
- 推定ステップ:
    1. まず $P(X)$ を推定する必要があります。データには $P(X)$ はないので、プロビットモデルやロジットモデルを使って D を X で説明し、予測確率 $\hat{P}(X)$ を計算します。
    2. $ATE = E[\frac{D \cdot y}{\hat{P}(X)} - \frac{(1-D) \cdot y}{1-\hat{P}(X)}]$ (式101, 102) をサンプルの平均で計算します。
- 特徴:
    - 回帰調整と違い、y と X の関数形（M_1, M_0）を仮定せんでええのが強みです。
    - ただし、傾向スコア $P(X)$ のモデル（プロビットなど）を正しく特定化する必要があります。

③ 二重にロバストな推定法 **(Doubly Robust Estimation)**

- 考え方: 上記2つの「ええとこ取り」をする方法です。
    - 回帰調整で使うモデル (M_1, M_0)
    - 傾向スコアで使うモデル ($P(X)$)
    - この**「どっちか一方」が正しく特定化されていれば**、ATEを正しく推定できる、という非

    常に頑健（ロバスト）な方法です。

- 推定:
  - （講義では複雑な式 (式105, 106) が提示されましたが）
  - もし **M** のモデルが間違っていても **P** が正しければ、追加で足した補正項の期待値が0になり、結果的に傾向スコア推定量と一致して、正しい値に収束します。
  - もし **P** のモデルが間違っていても **M** が正しければ、こちらも上手いこと調整され、正しい値に収束します。

4. 因果推論の限界

- 外的妥当性 **(External Validity)**: 京都市での職業訓練の効果が、北海道やアフリカで同じとは限らへん、という問題です。
- 一般均衡効果 **(General Equilibrium Effect)**:
  - 実験的に「一部の人」に訓練した場合の効果と、制度として「日本全国民」に訓練した場合の効果は違う可能性があります。
  - みんなが訓練を受けたら、技能の希少価値がなくなって賃金が上がらんかもしれへん、ということです。

第**2**部：操作変数法 **(IV**法**)**

後半は、OLS（最小二乗法）が使えへん状況、すなわち説明変数 X と誤差項 U が相関してしまう「内生性」がある場合に、どないして正しい係数を推定するか、という話でした。

1. なぜ内生性が起きるのか？

- 欠落変数 **(Omitted Variable)**: 本当はモデルに必要な変数 Z がデータになく、モデルから欠落している場合。Z が X とも y とも相関していると、Z の影響が誤差項 U に含まれてしまい、X と U が相関します。
- 同時方程式 **(Simultaneity)**: 需要と供給のように、y と X がお互いに影響しあって決まる場合。
- 測定誤差 **(Measurement Error)**: 説明変数 X に測定誤差がある場合。

2. 測定誤差（詳細）

- 真のモデル: $y = \beta_0 + \beta_1 X^* + U$ （X^* は真の値）
- 観測されるデータ: $X = X^* + V$ （V は測定誤差）
- 推定するモデル: 観測できる X を使うと、式は $y = \beta_0 + \beta_1 X + (\boldsymbol{U - \beta_1 V})$ となります。
- 問題点: 新しい誤差項 $\epsilon = (U - \beta_1 V)$ と、説明変数 $X = (X^* + V)$ は、共通の V を含んでいるため、必ず相関してしまいます ($Cov(X, \epsilon) \neq 0$)。
- 結果: OLSで推定すると、β_1 の推定値は間違った値（過小推定）になります。

3. 操作変数 **(Instrumental Variable, IV)** とは？

この内生性の問題を解決するのが「操作変数 Z」です。

- 操作変数 **Z** が満たすべき**2**つの条件:
  1. 関連性 **(Relevance)**: Z は、内生変数 X と相関している ($Cov(Z, X) \neq 0$)。
  2. 外生性 **(Exogeneity)**: Z は、誤差項 U とは相関していない ($Cov(Z, U) = 0$)。

- 用語の整理:
    - 内生変数: 誤差項 U と相関する変数（この例では X と Y）。
    - 外生変数: 誤差項 U と相関しない変数（この例では Z）。

4. IV推定量の導出 (X, Z が各1つの場合)

- 元のモデル $y = \beta_0 + \beta_1 X + U$ の両辺と Z の共分散 $Cov(Z, \cdot)$ をとります。
  $Cov(Z, y) = Cov(Z, \beta_0 + \beta_1 X + U)$
- 共分散の性質から、$Cov(Z, y) = \beta_1 Cov(Z, X) + Cov(Z, U)$ となります。
- ここでIVの条件を使います。
    - 外生性 ($Cov(Z, U) = 0$) より、第2項は0になります。
    - 関連性 ($Cov(Z, X) \neq 0$) より、$Cov(Z, X)$ は0ではないので、両辺を割ることができます。
- **IV**推定量: $\beta_1 = \frac{Cov(Z, y)}{Cov(Z, X)}$
- データ（標本）からは、それぞれの標本共分散 S_{ZY} と S_{ZX} の比として推定できます。
  $\hat{\beta}_{1, IV} = \frac{S_{ZY}}{S_{ZX}}$

5. 2段階最小二乗法 (TSLS or 2SLS)

IV推定量を計算する、もう一つの（そしてより一般的な）方法です。
- 考え方: 内生変数 X のうち、Z と相関している「良い部分」（外生的な部分）だけを抽出して回帰に使う、という考え方です。
- 推定ステップ:
    - 第**1**段階 **(1st Stage)**:
        - 内生変数 X を、操作変数 Z にOLS回帰します。
        -
        - この回帰から、X の予測値 $\hat{X} = \hat{\pi}_0 + \hat{\pi}_1 Z$ を作ります。
        - この $\hat{X}$ は、X の変動のうち Z で説明できる「良い」部分だけを取り出したものと解釈できます。
    - 第**2**段階 **(2nd Stage)**:
        - 元の y を、元の X の代わりに予測値 $\hat{X}$ を使ってOLS回帰します。
        -
        - この時に得られた係数 $\hat{\beta}_1$ が、TSLS推定量です。
- 特徴:
    - X と Z が1変数ずつの場合、このTSLSで得られる推定量は、先に導出したIV推定量 ($\frac{S_{ZY}}{S_{ZX}}$) と完全に一致します。

連絡事項
- 来週月曜日に中間試験。
- 範囲は今日の講義（操作変数法）まで。
- 持ち込みは一切なし。

こんな感じで、講義の流れと重要なポイントを詳細に追ってみました。どないでしょうか？