回帰分析１：フレームワーク

講師：遠山祐太

最終更新：2024-11-16

はじめに

モチベーション：観察研究 (observational study)

- RCTは強力なツール。しかし、社会科学においてRCTが常に実施できるわけではない。
- **観察データ**: 経済主体の実際の行動から得られるデータ。
- ポイント：観察データにおいては処置が無作為に割り当てられていない
- 今日(＆今後)のテーマ：観察データを用いて、どのように因果効果の推定を行うか？

リサーチデザイン(Research Design)

- **リサーチデザイン**：観察データから因果効果をバイアスなく推定するための分析デザイン。
  - 識別戦略 (identification strategy) と呼ぶことも。
- 直観的に言うとコントロール・トリートメントグループを **「上手」に比較**する方法。
- リサーチデザインは問い・背景・設定・データによって異なってくる。
  - 状況により、適用するツールに必要な仮定が満たされるか否かが異なる。

今回のテーマ：マッチングから線形回帰分析

- **マッチング**による処置効果の識別
  - 選択バイアスをもたらす**観察可能な特徴を統制する**
  - 平たく言うと、「同じ属性を持っているが、処置が異なる人同士を比較する」
- このアイデアは、**最小二乗法（ordinary least squares; OLS）**で**線形回帰モデル（linear regression model）**を推定することに帰着する。

授業プラン

- Part 1: フレームワーク
  - マッチング（matching）と selection on observables
  - 線形回帰モデル：モデル、推定、統計的推論

- Part 2: 応用編
  - 仮定の吟味：欠落変数バイアス、多重共線性
  - 実証研究：テレビに出ると政治家になりやすいのか？レーガンの例。

- 補足資料：
  - 回帰分析の実践的トピック、マッチング推定量の詳細
  - Rを用いた回帰分析のコード

Selection on Observables・マッチング

マッチングのアイデア

- 問題点：処置群と統制群を単純比較すると、その差は処置以外の差にも起因するかも。。。
- 解決策：処置群と統制群で、**同じ特徴 ¥(X¥) をもつ個体**を比較する -> **マッチング**
- もし観察される特徴が処置選択とアウトカムに影響するならば、それらの要素を統制すれば選択バイアスを取り除けるはず。
  - 例：両親の所得が、自身の進学選択と賃金アウトカムに影響する。
- マッチングの仮定
  1. Selection on observables
  2. Overlapping

仮定１(/２)：Selection on Observables

- 条件付き無視可能性（conditional ignorability）・条件付き交換可能性（conditional exchangability）とも呼ばれる。
- X_{i} を観察される特徴（**共変量（covariate）**）とする。
    - 年齢・所得・教育年数・人種など
- **仮定１**： $$D_{i}\perp(Y_{0i},Y_{1i})\left|X_{i}\right.$$
- 意味： X_{i} で**条件付けると**、処置割り当てと潜在アウトカムは独立

仮定２(/２)：群間の重複（Overlapping）

- positivity（正値性）・common supportとも呼ばれる。
- **仮定２**： $$P(D_{i}=1|X_{i}=x)\in(0,1)\ \forall x$$
- 意味： (x) が与えられたら、それに対して処置群と統制群の両方の人を観察できなければならない。
- $(P(D_{i}=1|X_{i}=x))$ は**傾向スコア（propensity score）**とよばれる。

処置効果パラメータの識別

- 仮定１により $$¥begin{aligned} E[Y_{1i}|D_{i} & =1,X_{i}]=E[Y_{1i}|D_{i}=0,X_{i}]=E[Y_{1i}|X_{i}], ¥¥ E[Y_{0i}|D_{i} & =1,X_{i}]=E[Y_{0i}|D_{i}=0,X_{i}]=E[Y_{0i}|X_{i}].¥end{aligned}$$
- ¥(X_i¥) で条件づけてしまえば、RCTにおける識別の議論基本同じである。

続・識別

- ¥(X_{i}=x¥) で条件付けたATTは $$¥begin{aligned} E[Y_{1i}-Y_{0i}|D_{i}=1,X_{i}] & =E[Y_{1i}|D_{i}=1,X_{i}]-E[Y_{0i}|D_{i}=1,X_{i}]¥¥ & =E[Y_{1i}|D_{i}=1,X_{i}]-E[Y_{0i}|D_{i}=0,X_{i}]¥¥ & =¥underbrace{E[Y_{i}|D_{i}=1,X_{i}]}_{X_iをもつ処置群の期待値}-¥underbrace{E[Y_{i}|D_{i}=0,X_{i}]}_{X_iをもつ統制群の期待値}¥end{aligned}$$
  - 直感： ¥(X_i¥) で条件付けたあと、処置群と統制群を比較する。
- 仮定２がここで必要となる。
  - もしある ¥(X_i = x¥) となる人が処置・統制群両方にいないと、比較できない！

ATT・ATEの識別

- 今、 ¥(X_i = x¥) ごとの処置効果 ¥(E[Y_{1i}-Y_{0i}|D_{i}=1,X_{i}]¥) がわかっている。
- ¥(X_i = x¥) に関してその平均をとる(積分を取る)と、ATTがわかる。(次スライド参照)
- また、条件付き独立性(仮定１)から、 $$E[Y_{1i}-Y_{0i}|D_{i}=1,X_{i}=x] = E[Y_{1i}-Y_{0i}|D_{i}=0,X_{i}=x] = E[Y_{1i}-Y_{0i}|X_{i}=x]$$
- 意味： Conditional on ¥(x¥), ¥(ATE(x) = ATT(x) = ATC(x)¥).

(参考) ATT \(E[Y_{1i}-Y_{0i}|D_{i}=1]\)

- ATTは次のように計算される。 $$¥begin{aligned} ATT & =E[Y_{1i}-Y_{0i}|D_{i}=1]¥¥ & =¥int E[Y_{1i}-Y_{0i}|D_{i}=1,X_{i}=x]f_{X_{i}}(x|D_{i}=1)dx¥¥ & =E[Y_{i}|D_{i}=1]-¥int¥left(E[Y_{i}|D_{i}=0,X_{i}=x]¥right)f_{X_{i}}(x|D_{i}=1)¥end{aligned}$$

(参考) ATT \(E[Y_{1i}-Y_{0i}]\)

- ATEは次のように計算される。 $$¥begin{aligned} ATE= & E[Y_{1i}-Y_{0i}]¥¥ = & ¥int E[Y_{1i}-Y_{0i}|X_{i}=x]f_{X_{i}}(x)dx¥¥ = & ¥int E[Y_{1i}|D_{i}=1,X_{i}=x]f_{X_{i}}(x)dx +¥int E[Y_{0i}|D_{i}=0,X_{i}=x]f_{X_{i}}(x)dx ¥¥ = & ¥int E[Y_{i}|D_{i}=1,X_{i}=x]f_{X_{i}}(x)dx +¥int E[Y_{i}|D_{i}=0,X_{i}=x]f_{X_{i}}(x)dx ¥¥ ¥end{aligned}$$

識別から推定へ

- ふたつの条件付き期待値 ¥(E[Y_{i}|D_{i}=1,X_{i}=x]¥) と ¥(E[Y_{i}|D_{i}=0,X_{i}=x]¥) を推定しなければならない。
- 幾つかの方法

  1. 回帰：パラメトリックおよびノンパラメトリック
  2. 最近傍マッチング（nearest neighborhood matching）
  3. 傾向スコアマッチング（propensity score matching）

- 本授業ではパラメトリック回帰(線形回帰)のみを紹介する。
- マッチング推定量の詳細については補論または教科書を参照すること。

Loading [MathJax]/extensions/MathZoom.js

マッチングから線形回帰モデルへ

- 以下のように仮定する $$¥begin{aligned} E[Y_{i}|D_{i}=0,X_{i}=x] & =¥beta'x_{i},¥¥ E[Y_{i}|D_{i}=1,X_{i}=x] & =¥beta'x_{i}+¥tau. ¥end{aligned}$$
  - ¥(¥beta'x_{i} = ¥beta_0 + ¥beta_1 x_{1i} + ¥cdots + ¥beta_{K} x_{Ki}¥)
- ここで、処置効果は ¥(¥tau¥) で与えられる。
- 書き換えると、線形回帰モデル $$y_{i}=¥beta'x_{i}+¥tau D_{i}+¥epsilon_{i},¥ E[¥epsilon_i | D_i, x_i] = 0$$ となり、パラメータ ¥(¥tau¥) を最小二乗法により推定する

線形回帰：フレームワーク

回帰の枠組み

- **線形回帰モデル（linear regression model）**
  $$Y_{i}=\beta_{0}+\beta_{1}X_{1i}+\cdots+\beta_{K}X_{Ki}+\epsilon_{i}.$$
    - \(i\) ：観測の添字（ \(i = 1,\cdots, N\) ）
    - \(Y_i\) ：**従属変数（dependent variable）**または**被説明変数**
    - \(X_{ki}\) ： **独立変数（independent variable）**または**説明変数（explanatory variable）**
    - \(\epsilon_i\) ： **誤差項（error term）**
    - \(\beta\) ：**係数（coefficient）**

- データ（標本） ： \(\{ Y_i , X_{i1}, \ldots, X_{iK} \}_{i=1}^N\)

- ゴール：データを用いて係数 \(\beta\) を推定

回帰係数の解釈

- 線形回帰モデル $$Y_{i}=\beta_{0}+\beta_{1}X_{1i}+\cdots+\beta_{K}X_{Ki}+\epsilon_{i}$$
- 係数 β_k ：**他が一定に保たれたときの（ceteris paribus）** Y に対する X_k の効果
- すなわち、もし X_k が連続な確率変数ならば $$\frac{\partial Y}{\partial X_k} = \beta_k$$
- もし β_k をバイアスなく推定できれば、 Y に対する X_k の**因果効果**がわかる。

最小二乗法（ordinary least squares; OLS）

- 最小二乗推定量（OLS estimator）は残差平方和（sum of squared residuals）を最小化する。 $$¥min_{¥beta_0, ¥cdots, ¥beta_K} ¥frac{1}{N} ¥sum_{i=1}^N ¥left(Y_i - (¥beta_0 + ¥beta_1 X_{i1} + ¥cdots + ¥beta_K X_{iK}) ¥right)^2$$
- (一定の条件下で)一階条件を解くことで、最小二乗推定量 ¥(¥hat{¥beta}¥) が得られる。
  - 行列表記で ¥(¥hat{¥beta}=(X'X)^{-1} X'Y¥) となる。(教科書参照)
- OLSのイメージ図：

OLSの別の見方：残差回帰(Residual Regression)

- モデル： $$Y_i = ¥beta_0 + ¥alpha D_i + ¥beta_1 X_{1i} + ¥cdots + ¥beta_K X_{Ki} + ¥epsilon_i$$
- ここで処置変数 ¥(D_i¥) の係数 ¥(¥alpha¥) に興味があるとしよう。
- 残差回帰という方法によって、OLS推定量 ¥(¥hat{¥alpha}¥) を以下のように与えることができる。

Frisch-Waugh-Lovell Theorem (FWL定理)

1. ¥(D_i¥) を他のすべての説明変数 ¥(1, X_{1i}, ¥ldots, X_{Ki}¥) に回帰する。その残差を ¥(¥hat{u}_i^D¥) とする。

2. ¥(Y_i¥) を他のすべての説明変数 ¥(1, X_{1i}, ¥ldots, X_{Ki}¥)　に回帰する。その残差を ¥(¥hat{u}_i^Y¥) とする。

3. ¥(¥hat{u}_i^Y¥) を ¥(¥hat{u}_i^D¥) に 定数項抜きで回帰することで、OLS 推定量 ¥(¥hat{¥alpha}¥) は以下として与えられる。 $$ ¥hat{¥alpha} = ¥frac{¥sum_i ¥hat{u}_i^D ¥hat{u}_i^Y }{¥sum_i (¥hat{u}_i^D)^2 } $$

- 補足：定数項なしの線形回帰モデル ¥(y_i = ¥beta x_i + ¥epsilon_i¥) のOLS推定量は ¥(¥hat{¥beta}= ¥frac{¥sum_i y_i x_i}{¥sum_i x_i^2}¥)

FWL定理が役立つところ

- 特定の係数のみに興味があるときに、計算コストを減らすことができる。
  - 例：パネルデータの固定効果モデル ¥(y_{it} = ¥beta_0 + ¥beta_x x_{it} + ¥theta_i + ¥epsilon_{it}¥).
- 特定の係数が「どのようなデータの変動から推定されているか？」を見るのに有用。多重共線性のところで再び。
- Double machine learning (Chernozhukov et al 2018): コントロール変数が大量にあるときに、機械学習をうまく使って、処置変数の係数を推定する。

最小二乗法の仮定

1. **ラングムサンプル**： ¥(¥{ Y_i , X_{i1}, ¥ldots, X_{iK} ¥}¥) はi.i.d.（独立同一分布）なサンプルである。
2. **平均独立（mean independence）**： ¥(¥epsilon_i¥) の条件付き期待値が0である。
   $$E[¥epsilon_i | X_{i1}, ¥ldots, X_{iK}] = 0$$
   - この条件から ¥(Cov(X_{ik},¥epsilon_i)=0 ¥ ¥forall k¥) (説明変数と誤差項の間に相関がない)
3. ¥(Y_i¥) および ¥(X_{ik}¥) の4次モーメントが有限である。：大きい特異値が発生しにくい
4. **完全な多重共線性がない（No perfect multicollinearity）**：説明変数の間に線形の関係がない（互いに独立である）。

線形回帰モデルの実践的トピック

Loading [MathJax]/extensions/MathZoom.js

線形回帰モデルにおける定式化

- 実証分析においてよく用いられる定式化
  1. 非線形項
  2. 対数変換
  3. ダミー（カテゴリカル）変数
  4. 交差項

非線形項（non-linear term）

- ¥(Y¥) と ¥(X¥) の非線形的な関係を線形和の形で書くと、例えば $$Y_i = ¥beta_0 + ¥beta_1 X_i + ¥beta_2 X_i^2 + ¥beta_3 X_i^3 + ¥epsilon_i$$
- 誤差項 ¥(¥epsilon_i¥) が線形的な加法の形で現れるかぎり、最小二乗法で係数を推定できる。
  - 多項式の項が多いと、多重共線性が問題になりうる。
  - ¥(¥log(x)¥) や ¥(¥sqrt{x}¥) といった別の非線形項を使ってもよい。

対数変換

- `log` を使うと係数 ¥(¥beta¥) のスケール解釈が変わる。

被説明変数	説明変数	解釈
¥(Y¥)	¥(X¥)	¥(X¥) 一単位の増加は ¥(Y¥) を ¥(¥beta¥) 単位変化させる。
¥(¥log Y¥)	¥(X¥)	¥(X¥) 一単位の増加は ¥(Y¥) を ¥(100¥beta¥%¥) 変化させる。
¥(Y¥)	¥(¥log X¥)	¥(X¥) の ¥(1¥%¥) の増加は ¥(Y¥) を ¥(¥beta/100¥) 単位変化させる。
¥(¥log Y¥)	¥(¥log X¥)	¥(X¥) の ¥(1¥%¥) の増加は ¥(Y¥) を ¥(¥beta¥%¥) 変化させる。

ダミー変数

- **ダミー変数（dummy variable）**は0と1の二値をとる変数。質的情報を表現する。
- 例：人種ダミー $$¥begin{aligned} white_i = ¥left¥{ ¥begin{array}{ll} 1 & ¥text{if white} ¥¥ 0 & ¥text{otherwise} ¥end{array} ¥right. ¥end{aligned}$$
- ダミー変数の係数は、カテゴリー間でのアウトカム ¥(Y¥) の差異を捉えている。 $$Y_i = ¥beta_0 + ¥beta_1 white_i + ¥epsilon_i$$ 係数 ¥(¥beta_1¥) は、白人・非白人間での ¥(Y¥) の差異を捉えている。

Loading [MathJax]/extensions/MathZoom.js

交差項（interaction term）

- 回帰モデルに説明変数間の相互作用を加えることができる。
- 例えば、 $$wage_i = ¥beta_0 + ¥beta_1 educ_i + ¥beta_2 white_i + ¥beta_3 educ_i ¥times white_i + ¥epsilon_i,$$ ただし ¥(wage_i¥) は個人 ¥(i¥) の賃金を、 ¥(educ_i¥) は個人 ¥(i¥) の教育年数を表す。
- ¥(educ_i¥) の効果は $$¥frac{¥partial wage_i}{¥partial educ_i} = ¥beta_1 + ¥beta_3 white_i$$
- 人種間での教育年数の異質効果を許容する。

適合度の評価

- 線形回帰モデル適合度の評価指標として、 ¥(R^2¥) **（決定係数）** がよく使われる。
- ¥(R^2¥)の定義 $$R^2 = ¥frac{SSE}{SST} = ¥frac{¥sum_i (¥hat{y}_i - ¥bar{y})^2}{¥sum_i (y_i - ¥bar{y})^2} $$
  - 予測値: ¥(¥hat{y}_i = ¥hat{¥beta}_0 + ¥hat{¥beta}_1 X_{i1} + ¥cdots + ¥hat{¥beta}_K X_{iK}¥)
  - 標本平均 ¥(¥bar{y} = ¥frac{1}{N} ¥sum_i y_i¥)
  - SST (total sum of squares): 被説明変数 ¥(y¥) のサンプルにおける変動
  - SSE (explained sum of squares): 予測値 ¥(¥hat{y}¥) のサンプルにおける変動
- ¥(R^2 ¥in [0,1]¥). ¥(R^2¥) が大きいほど適合度が高い。

\(R^2\) の解釈

- ¥(R^2¥) は回帰モデルによって表される ¥(Y¥) の変動の割合を捉える。
- ¥(R^2¥) をどの程度気にするかは目的次第。
- もし回帰係数の推定(因果効果)が目的ならば、 ¥(R^2¥) はさほど重要ではない。
  - (上級向け： Oster (2019) の coefficient stability)
- もし被説明変数の予測が目的ならば、R^2 はある程度重要。
  - ただし、 ¥(R^2¥) はサンプル内での適合度。説明変数を増やせば増やすほど大きくなる。
  - モデル予測を考える際には、in-sample と out-of-sample 両方を考えるべき (過学習問題)

OLS推定量の統計的性質

統計的推論

- (頻度論において)推定量は**確率変数**である。
  - 抽出された標本に依存して、推定**値**が決まる。
  - 標本はランダムにサンプルされる。

- (図示) データの生成過程 (data generating process, DGP) $$y_i = ¥beta_0 + ¥beta_1 x_i + ¥epsilon_i, ¥epsilon_i ¥sim F_x(¥cdot), x_i ¥sim F_x (¥cdot)$$

最小二乗推定量の統計的推論

- 抽出された標本に依存するという点で、最小二乗推定量は**確率変数**である。
- 統計的推論を行い、最小二乗推定値の統計的不確実性を測らなければならない。
- 以下の流れ
  - 最小二乗推定量の理論的性質
  - 統計的推論：検定と信頼区間
  - 分散均一性（homoskedasticity） vs. 分散不均一性（heteroskedasticity）

最小二乗推定量の理論的性質

1. **不偏性（unbiasedness）**：説明変数 ¥(X¥) で条件付けたとき、最小二乗推定量 ¥(¥hat{¥beta}¥) の期待値が真のパラメータ ¥(¥beta¥) に一致する。 $$E[¥hat{¥beta} | X] = ¥beta$$

1. **一致性（consistency）**：最小二乗推定量 ¥(¥hat{¥beta}¥) が真のパラメータ ¥(¥beta¥) に確率収束する。 $$¥hat{¥beta}¥overset{p}{¥longrightarrow}¥beta$$
   - サンプルサイズが大きくなるにつれ、 ¥(¥hat{¥beta}¥) と真の値 ¥(¥beta¥) が一致する確率が 1 に近づく。

1. **漸近正規性（asymptotic normality）**：次ページ

Loading [MathJax]/extensions/MathZoom.js

最小二乗推定量の漸近分布

- 先述の仮定の元で、最小二乗推定量は**漸近正規性（asymptotic normality）**をもつ。
  $$¥sqrt{N}(¥hat{¥beta}-¥beta)¥overset{d}{¥rightarrow}N¥left(0,V ¥right)$$
  - ¥(¥hat{¥beta},¥beta¥) は ¥((K+1)¥times1¥) 係数ベクトル。 ¥(¥beta = (¥beta_0, ¥beta_1,¥ldots, ¥beta_K)¥)
  - ¥(¥overset{d}{¥rightarrow}¥) は分布収束を意味する。
  - 右辺は多変量正規分布
  - ¥(V¥) は **漸近分散（asymptotic variance）** とよばれる
    $$¥underbrace{V}_{(K+1)¥times(K+1)} =
    E[¥mathbf{x}_{i}¥mathbf{x}_{i}']^{-1}E[¥mathbf{x}_{i}¥mathbf{x}_{i}'¥epsilon_{i}^
    ここで、 ¥(¥mathbf{x}_{i}=¥left(1, X_{i1},¥cdots,X_{iK} ¥right)'¥) は ¥
    ((K+1)¥times 1¥) ベクトルである。

Loading [MathJax]/extensions/MathZoom.js

漸近分布の解釈

- 推定量 ¥(¥hat{¥beta}¥) の分布は多変量正規分布で**近似**できる。 $$¥hat{¥beta} ¥sim N(¥beta, V / N)$$
- 各々の係数 ¥(¥beta_k¥) もまた、正規分布に近似的に従う。(¥(V_{kk}¥) は 行列 ¥(V¥) の ¥((k,k)¥) 要素) $$¥hat¥beta_k ¥sim N(¥beta_k, V_{kk} / N)$$
- この近似の精度はサンプルサイズが大きいほど高くなる。
- 漸近分布を用いることで、係数の推定量に関する統計的推測を行うことができる。

漸近分散の推定

- ¥(V¥) は推定されるべき未知の量である。
- ¥(V¥) 推定方法として、その標本対応 ¥(¥hat{V}¥) を考えると $$¥hat{V}=¥left(¥frac{1}{N}¥sum_{i=1}^{N}¥mathbf{x}_{i}¥mathbf{x}_{i}'¥right)^{-1}¥left(¥frac{1}{N}¥sum_{i=1}^{N}¥mathbf{x}_{i}¥mathbf{x}_{i}'¥hat{¥epsilon}_{i}^{2}¥right)¥left(¥f
{N}¥sum_{i=1}^{N}¥mathbf{x}_{i}¥mathbf{x}_{i}'¥right)^{-1},$$ ただし ¥(¥hat{¥epsilon}_i = y_i - (¥hat{¥beta}_0 + ¥cdots + ¥hat{¥beta}_K X_{iK})¥) は残差である。
    - 大数の法則より ¥(¥hat{V}¥) は ¥(V¥) に確率収束する。

標準誤差 (Standard Error)

- （漸近）**標準誤差** の定義： ¥(SE(¥hat{¥beta_k}) = ¥sqrt{¥hat{V}_{kk} / N }¥)
    - ¥(¥hat{V}_{kk}¥) は 行列 ¥(¥hat{V}¥) の ¥((k,k)¥) 要素
- 標準誤差は、最小二乗推定量 ¥(¥hat{¥beta_k}¥) の標準偏差に対する推定量

分散均一性と分散不均一性

- 回帰分析においては、しばしば以下のような**分散均一性（homoskedasticity）**を仮定する。$$¥sigma^2 = Var(¥epsilon_i)$$ このとき、漸近分散行列は以下 $$V = E[¥mathbf{x}_{i}'¥mathbf{x}_{i}]^{-1}¥sigma^{2}$$
- 一方で、上記の仮定がないケースを**分散不均一性（heteroskedasticity）**と呼ぶ。
  - 例：誤差項 ¥(¥epsilon_i¥) の分散が ¥(X_{i}¥) に依存する場合。
  - この場合の漸近分散行列はすでに導出した通り。

標準誤差に関する実践アドバイス

- 分散不均一性を許容した上での標準誤差を、**分散不均一性に頑健な標準誤差（heteroskedasticity-robust standard error）**とよぶ。
- 多くの統計パッケージ（RやStataも含む）で、初期設定では分散均一性の仮定に基づいて最小二乗推定量の標準誤差が計算される。
- しかし、分散不均一性があった場合、分散均一性の仮定の下で標準誤差は**過小推定**される。
- 最小二乗法では、**必ず分散不均一性に頑健な標準誤差を用いること。**

線形回帰モデルの統計的推論

仮説検定

- 係数に関する何らかの仮説を考える。
  - ¥(X¥) は本当に ¥(Y¥) に影響を及ぼしているのか？
  - 生産技術は規模に対して収穫一定なのか？

- データを用いて最小二乗法を行い、係数の推定値・標準誤差が得られる。

- 上述の仮説が母集団において本当に成立していそうかを **仮説検定** を用いて吟味する。

仮説検定の 3 Step

- Step 1：帰無仮説 ¥(H_{0}¥) と対立仮説 ¥(H_{1}¥) を考える。(¥(k¥) は分析者が決定)
  $$H_{0}:¥beta_{1}=k,¥ H_{1}:¥beta_{1}¥neq k$$
- Step 2：推定結果に基づいて、**t統計量**の値 (いわゆるt値)計算する。
  $$t_{n}=¥frac{¥hat{¥beta_1}-k}{SE(¥hat{¥beta_1})}$$
- Step 3：もし $$|t_{n}|>C$$ ならば ¥(¥alpha¥) %の有意水準で帰無仮説 ¥(H_0¥) を棄却する。 ¥(C¥) は標準正規分布の ¥(100-¥alpha/2¥) 分位点である。もし上の条件が成り立たないならば、 ¥(H_0¥) を**棄却することはできない。**

仮説検定の考え方

- 仮に帰無仮説 ¥(H_{0}:¥beta_{1}=k¥) が正しいとしよう。
- すると、t統計量 ¥(t_{n}=¥frac{¥hat{¥beta_1}-k}{SE(¥hat{¥beta_1})}¥) という確率変数は(漸近的に) 標準正規分布に従う。
- もし、手元のデータに基づいて得られたOLSの推定値及び標準誤差の値を用いて計算したt値が**極端な値**を取るならば、そもそもそも仮説設定 ¥(H_0¥) が間違っていたことを示唆する。
- この**極端**の度合いを決めるのが、標準正規分布の分位点。(以下図示)

仮説検定のポイント

- ¥(k = 0¥) のとき ¥(|t_{n}|>1.96¥) であれば、 ¥(¥hat{¥beta}¥) は5%の水準で**統計的に有意**といわれる。
  - 英訳：statistically significant
  - 厳密には ¥(|t_{n}|>C=1.95996...¥approx 1.96¥)。 標準正規分布の97.5%(=100-2.5)分位点。

- 係数の**経済的な有意性（economic significance）**についても検討するべき。

- 事例１：小さいが統計的に有意な係数
  - 標本の大きさ ¥(N¥) が大きくなればなるほど、標準誤差 ¥(SE¥) は減少する。

- 事例２：大きいが統計的に有意でない係数
  - その説明変数は、おそらく重要な（経済的に意味のある）効果をもたらしている。
  - しかし、手持ちの標本では正確に効果を推定することができない。

F検定

- 複数のパラメータを含むような複合仮説を検定することがある。例えば、 $$H_{0}:¥beta_{1} + ¥beta_2 = 0,¥ H_{1}:¥beta_{1} + ¥beta_2 ¥neq 0$$
- この場合、**F検定**を実施する。

- 詳細は割愛

信頼区間

- 95%信頼区間（confidence interval）推定量 $$¥begin{aligned} CI_{n} &= ¥left¥{ k:¥left|¥frac{¥hat{¥beta}_{1}-k}{SE(¥hat{¥beta}_{1})}¥right| ¥leq 1.96¥right¥} ¥¥ &= ¥left[¥hat{¥beta}_{1}-1.96¥times SE(¥hat{¥beta}_{1}),¥ ¥hat{¥beta_{1}}+1.96¥times SE(¥hat{¥beta}_{1})¥right] ¥end{aligned}$$

信頼区間の解釈

- **解釈**：この推定**量**は,95%の確率で真のパラメタを含む。
    - たくさんのサンプルを持ってきて95%信頼区間を構築したら、それら全ての信頼区間のうち95%が真のパラメータの値を含む。
- 注意：実際のサンプルに基づいて計算した95%信頼区間の推定**値**(実現値)は、真のパラメタを含むか否かのどちらか。
- 「推定した信頼区間(推定値)は95%で真の値を含む」とは言えない！！
    - (頻度論においては)真のパラメタは固定されたもの。

Loading [MathJax]/extensions/MathZoom.js

信頼区間の解釈に関する図示 (95%信頼区間)

- (1) 推定値がゼロに近く、5%有意水準で統計的に有意
- (2) 推定値がゼロに近くないが、5%有意水準で統計的に有意ではない
- (3) 5%有意水準で統計的に有意でないが点推定値大,
- (4) ギリギリ5%有意水準で統計的に有意

最後に：推定結果の解釈としての3S

- 3S: 符号(Sign)、推定値(Size)、統計的有意性 (Statistical Significance)
- 特に統計的有意性に引きづられがち。
- 必ず推定値について議論すること！ その際には、被説明変数の記述統計(平均など）が重要

Loading [MathJax]/extensions/MathZoom.js