

松井啓之

9/10AM

計量経済学

アメリカの大学（経済学）では必修
日本は必修ではない

2日間 8コマ 1単位分

基本的な統計学の知識はあるものとして講義する

実証分析 evidenceベース 医療の分野でも扱われている。
田中先生著の計量経済学の本で扱われているデータセットを含む

相関と因果

偏差値が高い大学の親の年収は高額
お金があるからいい大学にいける?
いい大学に行っているからお金がある?

因果関係、因果硬貨を示さなければならない
⇒ 医療の分野で最初に取り上げられた

反事実

医薬品を投与した場合の因果関係

ピアソン⇒相関関係しかわからないと主張

実証、実験は現実的には難しい
自然実験

計量経済学のソフトウェアは“Stata”が現在のスタンダード

経済学 時系列解析が多かった

gretl 経済学分野で主に使用されている統計解析ソフトウェア

田中先生、鹿野先生

インストール済みのデータは教育用として公開されているもの。
一部、個人入手する場合は正規の手続きが必要なもの（SRDQ）もあり。

9/10PM

中心極限定理
大数の原理

推定、推測統計

普遍性

一致性

効率性（有効性） 最小分散

母集団を知りたい

標本平均が従う分布

平均の検定

母平均がわかる \Rightarrow 正規分布 (z分布)

母平均がわからない \Rightarrow t分布

分散検定 カイ二乗分布

パラメトリック検定 標本分布がわかっている場合
ノンパラメトリック検定 標本分布がわかっていない場合

ダミー変数

ジップの法則

相関係数の基準 0.7

決定率 $R^2 \leftarrow$ 強い相関 $0.7^2 = 0.49 \approx 0.5 \dots 50\%$

直線の上に乗ることが大前提

直線の上にのらないと、相関がないとはいきれない

最小二乗法 (OLS: Ordinary Least Square)

普遍性、一致性、効率性を満たしている。ただし、ガウスマルコフ性に従うこと？

$\sum e_i \cdot x_i = 0 \rightarrow$ 誤差と変数xの内積ゼロ \rightarrow 誤差と変数xは直交している \rightarrow 独立である

説明変数 x が非確率変数であること \rightarrow 古典

x が確立変数になる \rightarrow 現代的な考え方 \rightarrow 期待値 $E(y|x)$ 条件付期待値、条件付確率 \rightarrow 因果関係が発生する

$$E(y|x) = E(\alpha) + \beta E(x) + E(u)$$

古典的ならば

$$= \alpha + \beta x$$

記述統計 (データありき)

推測統計 (標本から母集団の構造を推定する)

$\hat{\beta}$ と $\hat{\beta}$ の誤差はt分布に従う

統計的に意味があるかどうか検定する

\Rightarrow 仮説として立てた式の意味があることを示すことができる

仮説検定には、従う分布がわからなければいけない

有意差検定

t検定で2以上であれば帰無仮説は棄却される (0.05未満) \Rightarrow 係数に意味がある
P一値を見た場合、帰無仮説 (係数がゼロである) の起こり得る確率を見て確認

重回帰分析

3つ以上の母集団の平均を比較する場合は分散分析

F検定

(ANOVA : analysis of variance)

モデル式全体のチェック F検定

F0.05未満で棄却=有意

モデルの適合度

重決定R2（重回帰ではadjusted R-squared 調整済み適合度）

赤池情報量は小さいほうが良い

適合度合いとモデルの複雑さから判断している

足立区の中古マンション

価格(万円)、最寄駅からの時間、築年数、広さ、1Kかどうか

価格が何によって決まるのかを調べたい

頑健標準誤差

セッションにアイコンとして保存

分散均一性 → 現実的には、値が大きくなるにしたがって分散も大きくなると考える方が自然では？

重み付き回帰分析、加重最小二乗法 (WLS:Weighted Least Square)

頑健標準誤差を使うことで不均一分散へ対応する

「分析」で残差を確認

誤差均一の場合は、社会的データではあまりない

残差の正規性の分析は、データ数が増えるほどシビアになることが知られている。

時系列データの場合はダービー・ワトソン検定

Chow検定、時系列の時によくやる。統計量の構造のチェック。時系列である時点から構造（傾き）が変わったかどうか。

非線形かどうかのチェック方法がある

変数を減らす方が早く見つけ出せる（減少法）

重回帰分析

説明変数を減らしていく、より適合するモデルがあるかを探す

説明変数と非説明変数の意味合いも考えながらモデルを組み立て、求めたパラメータの値の意味合いを考える。

- ・検定を用いて値に意味を持つのかを確認
- ・モデルの意味合いの検定の確認
- ・モデルの適合度合いを確認
- ・誤差均一性の確認
- ・統計量に構造があるかどうか
- ・変数間の多重共線性の確認

多重共線性

説明変数間で線形関係があるとき、変数間に高い相関があるとき、解を一意に定めることができない

チェックする方法 VIF (Variance Inflation Factor)

現在は、多重共線性があってもよいと考える人もいる。昔はどうるさく言われなくなっている。

20180910_memo.txt Tanaka 7-1_incomeを用いて復習

income = $\alpha + \beta_1 \text{edu} + \beta_2 D(\text{ダミー}) + \beta_3 \text{edu}*D$
ダミー：1女性、0男性

Akaike criterion (AIC)
schwarz criterion (BIC)

対数変換すると、掛け算が足し算になるほかに、経済分野での弾力性を求めることができる

非線形の回帰

- ・プロビット：正規分布の累積分布関数
- ・ロジット：ロジスティック関数

非説明変数が0～1になる

均一分散ではないため、OLS推定できない⇒最尤推定

ロジスティック関数 スコアリング マーケティング分野で活用

birthwtのデータで分析

csvは行列指定できない⇒あらかじめデータを成型する必要有

限界効果、限界効用

ロジットモデルの係数 直線の傾き

モデルの違いで係数の符号に気を付ける（傾向が同じかどうか）

ロジットモデルのような非線形では、いろいろな検定をできない。カイニ乗検定ができるくらい。

<http://tylervigen.com/spurious-correlations>

因果関係、互恵効果、両者に影響する第3の要因

互恵効果 — ミクロマクロリンク

両者に影響する第3の要因 — 疑似相関

因果推論 反実仮想

因果推論の根本問題（観察対象について潜在的結果は1つしか観察できない）

前後比較、個体間比較

無作為化比較実験（RCT）・・・因果関係を証明する方法

他の手法もあるが、完全に証明できない

無作為に よく似た 規模が大きい 外部からの意図がない

実験が最良の方法

ただ、いつも実験ができるわけではない

社会科学は実験できない問題が多い（倫理的問題も含めて）

準実験、自然実験（たまたまRCTに似た条件のデータが得られた場合）

代表的な手法でわかりやすいものを紹介する

今井耕介 プリンストン大 社会経済学の責任者
森田 穀 若手の法学者 アメリカで学位取得
伊藤公一朗 京大で教鞭をとられていた

操作変数法はどの本も取り上げていない、難しい

誤差項と説明変数 x は独立（相関なし）

値 : x, y 確率変数 : X, Y

条件付き確率分布で扱う

条件付き確率の下での期待値 条件付き期待値

条件付き分散

条件付き期待値関数

誤植あり : $E(a+bY|X)=a+bE(Y|X)$

繰り返し期待値の法則

$E(u_i|X_i)=0$

外生性

X という条件の下で Y になる確率・・・因果関係 ($X \Rightarrow Y$) の回帰式で表される
 X と誤差との間に相関がない、直交している、因果なし

独立ではない⇒内生性

効率性（分散が一番小さい）

$Y=a+bX+u$

$E(X, u) \neq 0$ であれば、 X と u が独立ではない

X の中に u へ影響を与えるものが含まれている

それを取り除く z を探し出す必要がある

・・・操作変数法

$E(u|z)=0$

$E(X|z) \neq 0 \quad E(X, z) < 0$

これを満たす z を bX の項に加える

例（経験則）：兄弟の数

操作変数法は非常に難しい（ z を見つけ出すことが難しい）

心理学と経済学では手法が違う

心理学：因子分析 パス解析 共分散構造分析

回帰不連続デザイン

構造変化のところだけのデータで分析する

条件を揃えたい。

例えば、教育課程が改訂された影響を調べたい。

3月生まれと4月生まれの人だけをサンプルとする。

これは、経済状況などは変わらずに

教育課程が変わったことだけの影響を抽出するため。

傾向スコア・マッチング

医学系統でよく使われる。（ソフトウェア：ジャンプ）

バランスの取れた部分標本

共変量の値が等しい ← 集団全体で見たら原因の変数以外が同じように見える
様々な方法あり

20180910_memo.txt

施策を受ける確率を想定する

確率が近い人のグループで分析する

確率の近さは、確率の値で等分するのではなく、データ数でなるべく等分する。

Tanaka_10-1_income

0 -0.241

0.241-0.290

0.290-0.4

0.4 -1.0

グルーピングごとに年収の平均の傾向を示す

大学に行けるレベルでグルーピング

(大学に行った/行かなかった) で (年収の平均) を比較

大学に行きそうな確率ごとにグルーピングした群ごとに比較しても傾向が同じになることを確認する

大学に行った人は年収が高く、行かない人は年収が低い

要因と結果

- 2-3変速したからインプット回転数が上がった
- × インプット回転数が上がったから2-3変速した

プールの人数を同じにすることが重要

ロジットモデルやプロビットモデルを使う場合は離散値ではなく連続値が必要

パネル・データ分析
差分の差分法

Tanaka_9_2_life_xt

生活満足度が変わったかどうか

差分に対して回帰分析をかける 固定効果モデル

時間により変化しない個別効果が存在するとして個別効果を消去させて回帰パラメータを推定する。

ハウスマン検定 固定効果モデルと変量効果モデルの評価

時系列で差分を取る。その平均で違いがあるかどうか。→ 平均差分法

パネルデータを集めることが難しい（実用上）

操作変数法

ミンサー方程式

相関行列で相関をチェック

$$I(income) = a + b \text{edu} + u$$

$$\text{edu} = \text{edu(paedu)}$$

$$I(income) = a + b_1 \text{edu} + b_2 \text{paedu} + u$$

20180910_memo.txt

+ -
paeduは内生性 (eduにも影響する)

中の変数に対して回帰分析をし、その回帰変数を用いて再度回帰分析をする。

操作変数法、二段階変数法

b2を推定後、b1を推定する。

$$I(income) = a + b1 \text{ edu_est} + u$$
$$\text{edu_est} = a2 + b2 \text{ paedu}$$

操作変数をいろいろ試して効果があるものを探す。

探すことが非常に難しい。

関係ない操作変数を入れると回帰パラメータのt検定が悪化する。

内生変数に対する操作変数をきちんと選ぶことができれば、傾向を表すことができる。

田中先生の教科書のデータセットはインストール済み

2015年に出た本はどの本も良い

アメリカの本に比べて薄い、内容が薄い。