第 15 回：まとめ

北村　友宏

2021 年 1 月 22 日

本日の内容

1. ミクロデータとは
2. 係数の仮説検定
3. 同時方程式モデル
4. パネルデータ分析
5. 2 値応答モデル
6. レポート・論文用の表の作成
7. 今年度の授業で出た質問と回答

ミクロ・集計データ

- 個人，家計，事業所，企業などの観測単位からなるデータをミクロデータ（micro data）という．
  - 個人の所得，消費・個別市場での財の価格・数量，個人の成績
- ミクロデータを市町村，都道府県，国などの単位で合計または平均したデータを集計データ（aggregate data）という．
  - 個人の所得，消費の各都道府県における平均

この授業では「ミクロデータを都道府県別や市町村別に集計したもの」を「広義のミクロデータ」とする．
$\Rightarrow$ 期末レポートでは「広義のミクロデータ」を用いてよい．

係数の仮説検定

係数の仮説検定における帰無仮説と対立仮説は，

- 帰無仮説：「その係数は 0」
- 対立仮説：「その係数は 0 と異なる」

- p 値が 0.1 を超える・アスタリスクなし:
  - 有意水準 10%でも「その係数は 0」の帰無仮説を採択.
- p 値が 0.1 以下（未満）・*：
  - 有意水準 10%で「その係数は 0」の帰無仮説を棄却.
- p 値が 0.05 以下（未満）・**：
  - 有意水準 5%で「その係数は 0」の帰無仮説を棄却（有意水準 10%でも棄却できる）.
- p 値が 0.01 以下（未満）・***：
  - 有意水準 1%で「その係数は 0」の帰無仮説を棄却（有意水準 10%や 5%でも棄却できる）.

※検定統計量が連続型の確率分布（正規分布，t 分布，カイ二乗分布，F 分布など）に従う場合，「以上」と「超える」，「以下」と「未満」は区別しなくて良い.

- 帰無仮説が棄却された場合

  ➥ 対立仮説を採択し，「その係数は統計的に有意に 0 と異なる」と判断.

  - 「その説明変数は被説明変数と統計的に有意に相関している」と解釈.
  - 定数項の検定の場合は「定数項は統計的に有意に 0 と異なる」と解釈.

- 帰無仮説が採択された場合

  ➥ 「その係数は 0 と異なるとは言えない」と判断.

  - 「その説明変数は被説明変数と相関しているとは言えない」と解釈.
  - 定数項の検定の場合は「定数項は統計的に有意に 0 と異なるとは言えない」と解釈.

何を帰無仮説としているのか，そして棄却・採択の意味を考えること！

同時方程式モデル

- 需要・供給関数のモデルのように，「$x_i \to y_i \to x_i \to y_i \to \cdots$」のように決定され，$x_i$ と y_i が相互依存関係にあるモデル．

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$
$$x_i = \alpha_0 + \alpha_1 y_i + w_i.$$

- 説明変数と誤差項が相関するので，それに対処するために2段階最小二乗法（など）を用いて推定する．

パネルデータ

- 複数の個体を複数の時点にわたり，一定の時間間隔で観測したデータをパネルデータ（panel data）という．
  - e.g., 47 都道府県，1999 年〜2014 年，5 年間隔
  - e.g., 12 市場，2010 年 1 月〜12 月，1 ヶ月間隔

必ず複数の個体と複数の時点の数値がデータに含まれていること！

パネルデータのモデル

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 x_{it} + \mu_i + \varepsilon_{it}, \\ i &= 1, 2, \cdots, n, \\ t &= 1, 2, \cdots, T. \end{aligned}$$

- μ_i ：個別効果（individual effect）
  - 個体に特有で時間を通じて一定の効果
- ε_{it} ：その他要因
- 説明変数と独立でない個別効果を固定効果（fixed effect）という.
- 説明変数と独立な個別効果を変量効果（random effect）という.

固定効果モデル

- 個別効果が説明変数と独立でないことを仮定したモデルを固定効果モデル（fixed effect model）という．
- 固定効果モデルの推定方法
  - 1. Least Squares Dummy Variable (LSDV)
（個別効果 μ_i の部分に「係数 × 個体ダミー変数」を充てる）
  - 2. Within 推定
（変数変換して個別効果 μ_i を消去する）

変量効果モデル

- 個別効果が説明変数と独立であることを仮定したモデルを変量効果モデル（random effect model）という．
- 変量効果モデルの推定方法
  - 実行可能な一般化最小二乗法

2 値応答モデル

- 被説明変数がダミー変数の場合のモデル．

$$
\begin{aligned}
y_i &= \begin{cases} 1 & \text{if } y_i^* > 0, \\ 0 & \text{otherwise}, \end{cases} \\
y_i^* &= \beta_0 + \beta_1 x_i + u_i, \\
u_i \mid x_i &\sim F(.).
\end{aligned}
$$

y_i^* は潜在変数（latent variable）．観測不可能な変数で，y_i の値を決定づける．

- 最尤法で推定する．

- 誤差項 u_i には，説明変数 x_i を所与とした条件付き分布を仮定する．
    - e.g., 標準正規分布，ロジスティック分布
    - 誤差項の条件付き分布を標準正規分布と仮定した2値応答モデルを2値プロビット・モデル（binary probit model）という．
    - 誤差項の条件付き分布をロジスティック分布と仮定した2値応答モデルを2値ロジット・モデル（binary logit model）という．

注：“$y_i = \cdots$” の式ではなく “$y_i^* = \cdots$” の式の誤差項の分布を仮定している．

2 値プロビット・モデル

2 値プロビット・モデルは,

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$y_i^* = \beta_0 + \beta_1 x_i + u_i,$$

$$u_i \mid x_i \sim N(0, 1).$$

あるいは,

$$P(y_i = 1 \mid x_i) = \Phi(\beta_0 + \beta_1 x_i).$$

という書き方もある.

2 値ロジット・モデル

2 値ロジット・モデルは,

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$y_i^* = \beta_0 + \beta_1 x_i + u_i,$$

$$u_i \mid x_i \sim \Lambda(.).$$

あるいは,

$$P(y_i = 1 \mid x_i) = \Lambda(\beta_0 + \beta_1 x_i).$$

という書き方もある.

レポート・論文用の記述統計表の作成

見やすく，理解しやすい表を載せるには，

- 表番号と表のタイトルをつける．
- 変数名は統計解析ソフトでの変数名そのままではなく，分かりやすいように書き直す．
- 単位がある変数は単位を明記する．
- 小数の数値はあまり細かく表示せず，小数第 1〜4 位程度まで示せば十分．縦方向に見たとき，小数点の位置が揃うようにする．
- 観測値数が全変数について同じ場合，表の下に「注（Note）」として「観測値数は＊＊＊である」などと注記してもよい．

例えば以下のような表を載せればよい.

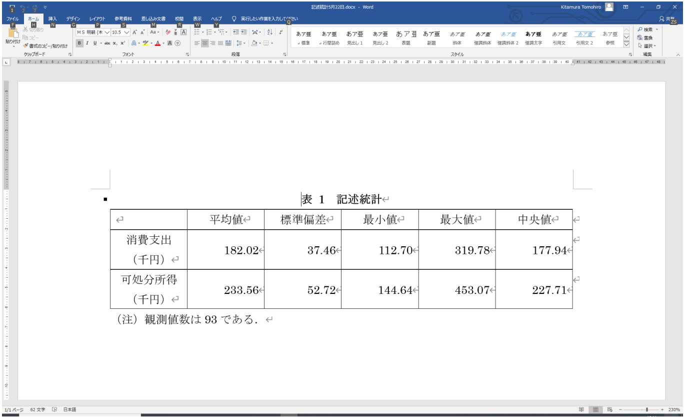

表 1　記述統計

	平均値	標準偏差	最小値	最大値	中央値
消費支出（千円）	182.02	37.46	112.70	319.78	177.94
可処分所得（千円）	233.56	52.72	144.64	453.07	227.71

（注）観測値数は 93 である.

レポート・論文用のモデル推定結果表の作成（線形モデルの場合）

見やすく，理解しやすい表を載せるには，

- 表番号と表のタイトルをつける．
- 最低限，以下の情報を載せる．
  - 係数推定値
  - t 値または z 値または標準誤差または p 値のどれか
  - R^2 または $\bar{R}^2$ のどちらか（2SLS では第 1 段階のみ）
  - 観測値数
- 有意性を示すアスタリスクを付けた場合は，表の下に「注（Note）」として「表中の***，**，*はそれぞれ有意水準 1%，5%，10%で統計的に有意であることを表す」などと注記する．

- 仮説検定に用いた標準誤差の種類や，頑健なら何に対して頑健なのかを，表の下に「注（Note）」として「誤差項同士のクラスター構造に対して頑健な標準誤差を用いている」などと注記する．
- 観測値数は，表の下に「注（Note）」として「観測値数は〇〇〇である」などと注記してもよい
  - e.g., 1 つの表に複数のモデルを載せ，観測値数が全モデルについて同じ場合など．
- 変数名は統計解析ソフトでの変数名そのままではなく，分かりやすいように書き直す．
- 小数の数値はあまり細かく表示せず，小数第 2〜4 位程度まで示せば十分．特に係数推定値，標準誤差，t 値，z 値，p 値は縦方向に見たとき，可能な限り小数点の位置が揃うようにする．
  - t 値，z 値は小数第 2 位まで，p 値，R^2，$\bar{R}^2$ は小数第 3 位まで示せば十分．

- 2 段階最小二乗法（2SLS）での推定結果を載せる場合は，第 1 段階推定と第 2 段階推定両方の結果を載せる.
  - 第 1 段階推定の結果は，用いた操作変数が操作変数として機能しているかを確認するために重要.
- 2 段階最小二乗法（2SLS）における第 2 段階の推定結果では，R^2 や $\bar{R}^2$ は載せなくてよい（解釈ができないため）.

2 段階最小二乗法（2SLS）の第 1 段階推定結果については，例えば以下の表を載せればよい．

表 1　第 1 段階推定結果

	係数	t 値	
大田ダミー	4.08	0.04	
北足立ダミー	−24.75	−0.24	
葛西ダミー	−14.58	−0.14	
豊島ダミー	−3.67	−0.04	
淀橋ダミー	−2.00	−0.02	
世田谷ダミー	9.00	0.09	
板橋ダミー	−63.50	−0.61	
多摩ニュータウンダミー	−20.42	−0.20	
月	339.88	10.91	***
月の二乗	−26.21	−11.23	***
定数項	−275.09	−2.45	**
自由度修正済み決定係数	0.522		

（注1）　表中の***，**はそれぞれ有意水準 1%，5%で統計的に有意であることを表す.

（注2）　デフォルトの標準誤差を用いている．

（注3）　観測値数は 108 である．

2 段階最小二乗法（2SLS）の第 2 段階推定結果については，例えば以下の表を載せればよい．

表 2　第 2 段階（需要関数）推定結果

	係数	t 値	
価格	−3.44	−5.41	***
大田ダミー	3542.55	4.75	***
北足立ダミー	−277.08	−0.37	
葛西ダミー	−683.35	−0.92	
豊島ダミー	−707.03	−0.95	
淀橋ダミー	−162.88	−0.22	
世田谷ダミー	−878.20	−1.18	
板橋ダミー	−829.01	−1.11	
多摩ニュータウンダミー	−1034.76	−1.39	
定数項	2782.38	4.48	***

（注1）　表中の***は有意水準 1%で統計的に有意であることを表す．

（注2）　デフォルトの標準誤差を用いている．

（注3）　観測値数は 108 である．

レポート・論文用のモデル推定結果表の作成（2 値応答モデルの場合）

- レポートや論文に載せるための 2 値応答モデルの推定結果表を作成する際には，最低限，以下の情報を載せればよい.
  - 係数推定値
  - 限界効果推定値
  - 「係数ゼロ仮説の検定のための z 値または p 値」または「係数の標準誤差」のどれか
  - 対数尤度
  - 観測値数
- 定数項の限界効果は存在しない.

レポートや論文に，2 値プロビット・モデルと 2 値ロジット・モデルの推定結果を並列して載せて比較したいときは，例えば以下のような表を載せればよい.

表 1 : 2 値応答モデル推定結果

	2 値プロビット・モデル				2 値ロジット・モデル			
	偏回帰係数	限界効果	z 値		偏回帰係数	限界効果	z 値	
出場時間率	−1.54	−0.60	−2.17	**	−2.44	−0.59	−2.09	**
得点率	−0.12	−0.05	−0.03		−0.46	−0.11	−0.08	
定数項	0.29		0.94		0.45		0.93	
対数尤度	−20.16				−20.19			

（注 1）表中の**は有意水準 5%で統計的に有意であることを表す.
（注 2）モデルの定式化に対して頑健な標準誤差を用いている.
（注 3）観測値数は 33 である.

今年度の授業で出た質問（その 1）

質問：整理した Excel ファイルでは変数が A 列から E 列までの 5 個しかないはずなのに，それを gretl に読み込もうとすると，31 個の変数が見つかった，などと表示され，読み込めない．

$$\Downarrow$$

回答：元ファイルからのコピー・貼り付けの際に余分な情報が貼り付けられ，

$$31 - 5 = 26 \text{ 個}$$

の目に見えない余分な変数が作成されてしまっている．この場合は，整理した Excel ファイルの F 列から AE 列までの 26 列を選択し，Delete キーを押して情報を消去して上書き保存すれば，gretl で読み込めるようになる．

今年度の授業で出た質問（その 2）

質問：2 段階最小二乗法の第 1 段階推定において，内生説明変数を操作変数だけでなく，システムに登場する全ての外生変数に回帰する理由が分からない．

$\Downarrow$

回答：結論としては，システム全体で見れば，内生説明変数は，操作変数だけでなく，推定したい式に入っていない外生変数も含め，システムに登場する全ての外生変数に依存して変化するため．

- ポイント 同時方程式モデルでは，複数の方程式に沿って各変数の値が決まる．

- 例えば授業の実習での需要・供給モデル

$$
\begin{aligned}
\text{需要}:& q_{it} = \beta_{D0} + \beta_{DP} p_{it} + \sum_{m=2}^{9} \beta_m d_{mi} + u_{Dit}, \\
\text{供給}:& p_{it} = \beta_{S0} + \beta_{SQ} q_{it} + \beta_T t + \beta_{TT} t^2 + u_{Sit}.
\end{aligned}
$$

では，システム内の需要関数と供給関数が連立方程式を成している．

- 連立方程式なので，例えば供給関数を内生変数である価格 p_{it} について解いた式で表すことができる．

- 供給関数の説明変数の取引数量 q_{it} に需要関数を代入すると，

$$
\begin{aligned}
p_{it} &= \beta_{S0} + \beta_{SQ}\left(\beta_{D0} + \beta_{DP}p_{it} + \sum_{m=2}^{9}\beta_m d_{mi} + u_{Dit}\right) \\
&\quad + \beta_T t + \beta_{TT}t^2 + u_{Sit} \\
&= \beta_{S0} + \beta_{SQ}\beta_{D0} + \beta_{SQ}\beta_{DP}p_{it} + \sum_{m=2}^{9}\beta_{SQ}\beta_m d_{mi} \\
&\quad + \beta_{SQ}u_{Dit} + \beta_T t + \beta_{TT}t^2 + u_{Sit}.
\end{aligned}
$$

- よって，

$$
\begin{aligned}
p_{it} - \beta_{SQ}\beta_{DP}p_{it} =& \beta_{S0} + \beta_{SQ}\beta_{D0} + \sum_{m=2}^{9}\beta_{SQ}\beta_m d_{mi} \\
&+ \beta_{SQ}u_{Dit} + \beta_T t + \beta_{TT}t^2 + u_{Sit} \\
\Leftrightarrow (1 - \beta_{SQ}\beta_{DP})p_{it} =& \beta_{S0} + \beta_{SQ}\beta_{D0} + \sum_{m=2}^{9}\beta_{SQ}\beta_m d_{mi} \\
&+ \beta_{SQ}u_{Dit} + \beta_T t + \beta_{TT}t^2 + u_{Sit}.
\end{aligned}
$$

- したがって，

$$p_{it} = \frac{\beta_{S0} + \beta_{SQ}\beta_{D0}}{1 - \beta_{SQ}\beta_{DP}} + \sum_{m=2}^{9} \frac{\beta_{SQ}\beta_m}{1 - \beta_{SQ}\beta_{DP}} d_{mi} + \frac{\beta_T}{1 - \beta_{SQ}\beta_{DP}} t + \frac{\beta_{TT}}{1 - \beta_{SQ}\beta_{DP}} t^2 + \frac{\beta_{SQ} u_{Dit} + u_{Sit}}{1 - \beta_{SQ}\beta_{DP}},$$

となり，供給関数から取引数量 q_{it} が消去され，

  - 被説明変数：価格 p_{it}
  - 説明変数：システムに登場する全ての外生変数 d_{mi}, t, t^2

となる．この形を誘導形（reduced form）という．

- この時点で，システム全体で見れば，その設定上，内生変数である価格 p_{it} は，操作変数になりうる月 t と月の二乗 t^2 だけでなく，各市場ダミー d_{mi} も含め全ての外生変数に依存して変化することが分かる．
- そこで，この例での 2 段階最小二乗法の第 1 段階推定では，供給関数の誘導形を

$$p_{it} = \underbrace{\pi_0 + \sum_{m=2}^{9} \pi_m d_{mi} + \pi_T t + \pi_{TT} t^2}_{u_{Dit}\text{と無相関}} + \underbrace{v_i}_{u_{Dit}\text{と相関}},$$

として，価格 p_{it} をシステムに登場する全ての外生変数 d_{mi}, t, t^2 に回帰する．

- ここで，第 1 段階推定の説明変数は外生変数なので，推定したい需要関数の誤差項 u_{Dit} とは定義上相関しないので，この推定を行えば，第 1 段階推定の被説明変数（価格 p_{it}）の変動のうち需要関数の誤差項 u_{Dit} と相関する部分と相関しない部分を切り離すことができる．