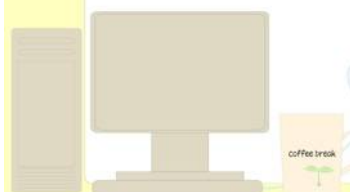


産業研究所・パネル共同研究拠点共催
統計分析コンピュータ講座
パネル分析講座・実習編

2010年6月実施



1

アウトライン(理論編)

1. パネルデータとは何か？
2. パネルデータの利点
3. 事例:大学の教育効果の検証
4. パネルデータの難点
5. 日本におけるパネル調査
6. 実習編:各事例の背景

2

事例1: 米国労働者パネルデータ

- Natioanl Longitudinal Survey:
 - 米国・労働統計局が作成しているパネルデータ
 - 1960、1970年代から、一定年齢層を追跡調査
 - 雇用・所得・消費動向を調査
 - ここでは「組合加入者の賃金プレミアム」を計測
- Do-file: NLSY-panel.do
 - ※データの読み込み: stata社提供のデータ
 - use <http://stata-press.com/data/r11/nlswork>
 - 個体識別番号: idcode, 時間識別番号: year

3

Stataによるパネルデータ分析

- パネルデータの認証
 - tsset [個体識別番号] [時間識別番号]
 - **tsset idcode year**
- データ構造の記述
 - **xtides**

4

Stataによるパネルデータ分析

```
. xtset
idcode: 1, 2, ..., 5159
year: 68, 69, ..., 88
Delta(year) = 1 unit
Span(year) = 21 periods
(idcode*year uniquely identifies each observation)
```

全サンプル数 → n = 4711
データ期間 → T = 15

Distribution of T_i:

	min	5%	25%	50%	75%	95%	max
	1	1	3	5	9	13	15

Freq.	Percent	Cum.	Pattern
136	2.89	2.89	1.....
114	2.42	5.31	1
89	1.89	7.20	1.11
87	1.85	9.04	11
86	1.83	10.87	111111.1.11.1.11.1.11
61	1.29	12.16	11.1.11
56	1.19	13.35	11.....
54	1.15	14.50	1.1.11
54	1.15	15.64	1.11.1.11.1.11
3974	84.36	100.00	(other patterns)
4711	100.00		xxxxxx.x.xx.x.xx.x.xx

1年分しかデータがないサンプル

5

Stataによるパネルデータ分析

- 労働組合状況遷移行列の作成
 - xttrans

		t 時点		
		No	Yes	
t-1 時点	No	XX	ZZZ	
	Yes	YY	WW	

```
. xttrans union if year==87|year==88
```

1 if union	1 if union		Total
	0	1	
0	93.06	6.94	100.00
1	19.79	80.21	100.00
Total	74.92	25.08	100.00

6

Stataによるパネルデータ分析

- 変数の概要: des
- 回帰モデル

$$\ln_wage_{it} = a + b_1 \cdot grade_i + b_2 \cdot union_{it} + b_3 \cdot age_{it} + b_4 \cdot ttl_exp_{it} + b_5 \cdot tenure_{it} + b_6 \cdot black_i + v_i + u_{it}$$
 (対数賃金) (就学年数) (組合ダミー) (年齢)
 (就業年数) (勤続年数) (黒人ダミー) (固定効果)
 - 固定効果: v_i
 - 時間を通じて変化しない観察不能な特性、能力など
 - プーリング回帰のときは無視される

7

パネル回帰コマンド

- xtreg y x1 x2 x3, fe
 - 固定効果モデル
 - cf. areg y x1 x2 x3, absorb(idcode) は同義
- xtreg y x1 x2 x3, re
 - 変量効果モデル

8

プーリング
回帰モデル

```
. reg ln_wage grade union age ttl_exp tenure black
```

Source	SS	df	MS	
Model	1434.18864	6	239.03144	
Residual	2712.45487	19001	.142753269	
Total	4146.64351	19007	.218164019	

	Number of obs =	19008
	F(6, 19001) =	1674.44
	Prob > F =	0.0000
	R-squared =	0.3459
	Adj R-squared =	0.3457
	Root MSE =	0.37783

ln_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
grade	-.0709392	.0011936	59.43	0.000	-.0685997 - .0732787
union	-.1709792	.0065974	25.92	0.000	-.1580477 - .1839106
age	-.0098343	.0006099	-16.12	0.000	-.0110298 - .0086388
ttl_exp	.0309557	.0010061	30.77	0.000	.0289837 .0329278
tenure	.015023	.0008784	17.10	0.000	.0133012 .0167448
black	-.1092001	.0062051	-17.60	0.000	-.1213627 - .0970375
_cons	.8508772	.0216822	39.24	0.000	.8083781 .8933763

固定効果
モデル

```
. xtreg ln_wage grade union age tenure black ,fe
```

note: grade omitted because of collinearity
note: black omitted because of collinearity

Fixed-effects (within) regression

Group variable: idcode	Number of obs	=	19008
	Number of groups	=	4132
R-sq: within	=	0.1259	Obs per group: min
between	=	0.1526	avg
overall	=	0.1300	max
			1
			4.6
			12

corr(u_i, Xb) = **0.1257**

F(3,14873) = **713.78**
Prob > F = **0.0000**

ln_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
grade	(omitted)				
union	-.0995394	.007003	14.21	0.000	-.0858127 - .1132662
age	-.0096535	.0004893	19.73	0.000	-.0086943 - .0106126
tenure	.0175694	.0008112	21.66	0.000	.0159794 .0191593
black	(omitted)				
_cons	1.360583	.0140832	96.61	0.000	1.332978 1.388188

sigma_u = .40555343
sigma_e = .25651274
rho = .7142567 (fraction of variance due to u_i)

F test that all u_i=0: F(4131, 14873) = **6.88** Prob > F = **0.0000**

Omitされる
のは何故？

Grade, blackを
browseで確認

9

練習問題(1)

Do-fileを作成しよう

– retail-density.doファイルを完成させよ

- retail-density.csvを読み込む
- パネルデータとして認証する(tsset)
- プーリング回帰と固定効果回帰で次の式を推定

$$- \text{dor}_{it} = a + b_1 \text{fspace}_{it} + b_2 \text{pcp}_{it} + b_3 \text{cvp}_{it} + b_4 \text{pdp}_{it} + v_i + u_{it}$$

- » dor: 小売店舗数／人口
- » fspace: 1住宅あたり床面積(－)
- » pcp: 自家用車保有台数／人口(－)
- » cvp: 商用車数／人口(+)
- » pdp: 人口／面積(人口が増えると密度は下がる, －)

10

パネルデータの作成と分析

最終的なデータ・フォーマット

個体識別 番号	年月	変数(1)	変数(2)	属性(1)	属性(2)	属性(3)
1	1990	Y ¹ _{1,1990}	Y ² _{1,1990}	X ¹ ₁	X ² _{1,1990}	Z ¹ _{1,1990}
1	1991	Y ¹ _{1,1991}	Y ² _{1,1991}	X ¹ ₁	X ² _{1,1991}	Z ¹ _{1,1991}
1	⋮	⋮	⋮	⋮	⋮	⋮
1	1995	Y ¹ _{1,1995}	Y ² _{1,1995}	X ¹ ₁	X ² _{1,1995}	Z ¹ _{1,1995}
2	1990	Y ² _{1,1990}	Y ² _{2,1990}	X ¹ ₂	X ² _{2,1990}	Z ¹ _{2,1990}
2	1991	Y ² _{1,1991}	Y ² _{2,1991}	X ¹ ₂	X ² _{2,1991}	Z ¹ _{2,1991}
2	⋮	⋮	⋮	⋮	⋮	⋮
2	1995	Y ² _{1,1995}	Y ² _{2,1995}	X ¹ ₂	X ² _{2,1995}	Z ¹ _{2,1995}
3	1990	Y ³ _{1,1990}	Y ³ _{2,1990}	X ¹ ₃	X ² _{3,1990}	Z ¹ _{3,1990}
3	1991	Y ³ _{1,1991}	Y ³ _{2,1991}	X ¹ ₃	X ² _{3,1991}	Z ¹ _{3,1991}
3	⋮	⋮	⋮	⋮	⋮	⋮
3	1995	Y ³ _{1,1995}	Y ³ _{2,1995}	X ¹ ₃	X ² _{3,1995}	Z ¹ _{3,1995}
4	1990	Y ⁴ _{1,1990}	Y ⁴ _{2,1990}	X ¹ ₄	X ² _{4,1990}	Z ¹ _{4,1990}
4	1991	Y ⁴ _{1,1991}	Y ⁴ _{2,1991}	X ¹ ₄	X ² _{4,1991}	Z ¹ _{4,1991}
4	⋮	⋮	⋮	⋮	⋮	⋮
4	1995	Y ⁴ _{1,1995}	Y ⁴ _{2,1995}	X ¹ ₄	X ² _{4,1995}	Z ¹ _{4,1995}

Stata入門 P.63

11

パネルデータの作成と分析

- 実際のデータは以下のようなフォーマットが多い

→ Stataで接続してデータ構築

個体識別 番号	年月	変数(1)
1	1990	Y ¹ _{1,1990}
2	1990	Y ² _{1,1990}
3	1990	Y ³ _{1,1990}
4	1990	Y ⁴ _{1,1990}

個体識別 番号	年月	変数(1)
1	1991	Y ¹ _{1,1991}
2	1991	Y ² _{1,1991}
3	1991	Y ³ _{1,1991}
4	1991	Y ⁴ _{1,1991}

個体識別 番号	年月	変数(1)
1	1995	Y ¹ _{1,1995}
2	1995	Y ² _{1,1995}
3	1995	Y ³ _{1,1995}
4	1995	Y ⁴ _{1,1995}

個体識別 番号	年月	変数(1)
1	1990	Y ¹ _{1,1990}
1	1991	Y ¹ _{1,1991}
1	⋮	⋮
1	1995	Y ¹ _{1,1995}

個体識別 番号	年月	変数(1)
2	1990	Y ² _{1,1990}
2	1991	Y ² _{1,1991}
2	⋮	⋮
2	1995	Y ² _{1,1995}

個体識別 番号	年月	変数(1)
3	1990	Y ³ _{1,1990}
3	1991	Y ³ _{1,1991}
3	⋮	⋮
3	1995	Y ³ _{1,1995}

個体識別 番号	年月	変数(1)
4	1990	Y ⁴ _{1,1990}
4	1991	Y ⁴ _{1,1991}
4	⋮	⋮
4	1995	Y ⁴ _{1,1995}

Stata入門
P.69-70

12

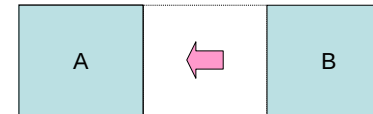
実習: データの構築

- Do-file: data-create.do
- CSV-file
 - fspace.csv
 - car.csv
 - land.csv
 - pop.csv
 - retailer.csv

13

Step 1: 横方向のデータ接続

- fspace.csvとpassenger_car.csvの接続
 - 横方向に接続: merge



- fspace.csvを読み込む。
- “prefecture year”でsortして、保存(retail-density.dta)
- car.csvを読み込んで、
- merge [キー変数] using retail-density.dta

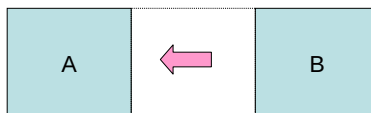
Stata入門 P.73

14

Step 1: 横方向のデータ接続

[merge] (キー変数) using B.dta

- ファイルAにファイルBを接続する場合
 - AとBの個体を識別する変数(例では都道府県番号 prefectureと年次year)であらかじめ、ソートして、ファイルBを保存しておく。
 - ファイルAも、prefectureでsortして、上記を実行。



15

Step 1: 横方向のデータ接続

“_merge”について

- mergeコマンドを使うと自動的に作成される。
- 接続状況を示す変数
 - 1のとき: もとのファイル(A.dta)にのみ存在するデータ
 - 2のとき: 接続するファイル(B.dta)にのみ存在するデータ
 - 3のとき: 両方のファイル(A.dtaとB.dta)に存在する

Stata入門 P.77

16

Step 1: 横方向のデータ接続

- 例) 以下の2つのファイルをmergeで接続
– “_merge”は接続状況を示す。

even.dta

number	Even
5	10
6	12
7	14
8	16

odd.dta

number	Odd
1	1
2	3
3	5
4	7
5	9

use even.dta
merge number using odd.dta

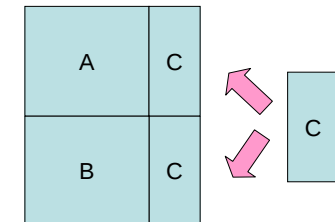
number	Even	Odd	_merge
5	10	9	3
6	12		1
7	14		1
8	16		1
1		1	2
2		3	2
3		5	2
4		7	2

- Exercise
– retail-density.dtaに commercial_vehicleを接続しよう

17

Step 2: 時間変動のないデータ接続

- land.csvの接続
– 横方向に接続: merge



- キー変数はprefectureのみにする
- merge [キー変数] using retail-density.dta

Stata入門 P.73

18

Step 3: データ形式の変換

- データ・フォーマットの変換
– retail-density.dtaはlong形式、
pop.csvはwide形式

図表 6-11 LONG形式のデータ (例)

fid	year	labor	slsprofit	head-g
6501	1994	80493	0.0174	13
6501	1995	78388	0.0215	13
6501	1996	75590	0.0312	13
6501	1997	72193	0.0196	13
6501	1998	70375	0.0042	13
6501	1999	66046	-0.0304	13
6501	2000	58739	0.0084	13
6501	2001	54017	0.0140	13
6501	2002	48590	-0.0232	13
6502	1994	74558	0.0165	14
6502	1995	73463	0.0215	14

図表 6-12 WIDE形式のデータ (例)

fid	labor1994	labor1995	labor1996	labor1997	labor...
6501	80493	78388	75590	72193	...
6502	74558	73463	7170	8441	...
6503	49842	8421	7752	7372	...

Stata入門 P.80

19

Step 3: データ形式の変換

- データ・フォーマットの変換
– Long形式⇒Wide形式
• reshape wide population, i(prefecture) j(year)
– Wide形式⇒Long形式
• reshape long population, i(prefecture) j(year)
- Exercise
– retailer.csvをLONG形式に変更し、retail-density.dtaに接続しよう

20

練習問題(2)

- 都道府県別の生産額、資本、労働データを接続して、都道府県パネルデータ(prod-prefec.dta)作成
- pref-data-const.doを完成させよ
 - "k-prefec1993.csv"
 - "k-prefec1994.csv"
 - "y-prefec.csv" ←long形式に変換
 - "l-prefec.csv" ←long形式に変換
 - Step 1 1と2を接続(append, k-prefec.dta)
 - Step 2 k-prefecとy-prefec.csvを接続(prod-prefec.dta)
 - Step 3 prod-prefec.dtaとl-prefec.dtaを接続

Stata入門 P.87

21

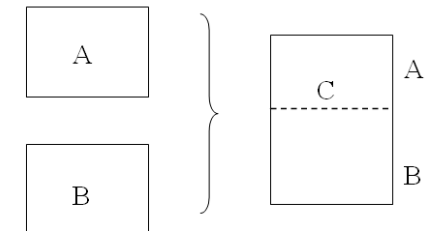
データの縦方向の接続

- データを縦方向に接続
 - appendコマンド

例)

use A.dta

append using B.dta



Stata入門 P.72

22

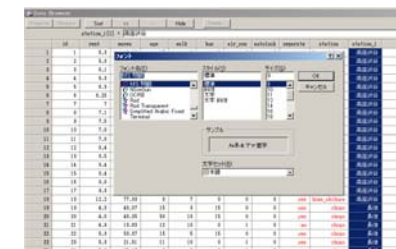
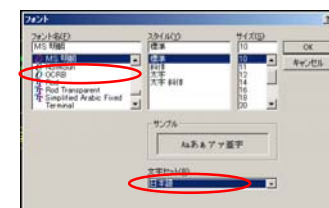
練習問題 (3)

- paneldata-exercise.pdf
 - 地域別企業別海外従業者数、企業別従業者のデータを接続してパネル回帰を実施する
 - データ: firm-data.zip

23

その他のtips (1)

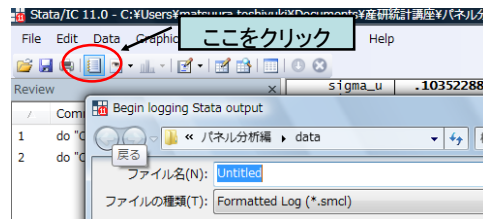
- Stata上で日本語を表示させるには？
 - “Result”ウインドウ上で、右クリック⇒”Fons”を選択
 - MS明朝、文字セット「日本語」を選択
 - ブラウザも同様に変更可能



24

その他のtips (2)

- 結果をファイルに保存(ログをとる)



- Log fileの閲覧
 - 「File」→「view」→「choose file to view」
→「browse」