

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

2010年6月実施

慶應義塾大学 産業研究所
専任講師 松浦寿幸

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

パネルデータとは？ データ形式の様々

時系列データ:特定の個体のデータを時間で追跡したデータ

	1970	1971	...	1980	...	1990
A						

横断面データ:ある一時点の様々な個体に関するデータ

	A	B	C	...	D
1990					

3

プーリングデータ

- 複数年次のクロスセクション・データを束ねたデータ (個体は追跡できない)

	1990		1991		1992		1993
90-a		91-e		92-i		93-m	
90-b		91-f		92-j		93-n	
90-c		91-g		92-k		93-o	
90-d		91-h		92-l		93-p	
⋮		⋮		⋮		⋮	

4

パネルデータ

- 複数の個体について、数時点わたって追跡調査したデータ

	1970	1971	...	1980	...	1990
A						
B						
C						
D						
⋮						
Z						

5

パネルデータ利用の利点

- サンプル数が増える。
- 横断面・時系列データでは困難な仮説の検証が可能となる場合がある。

6

大学はレジャーランド？

- 「就職市場における大学の銘柄効果」
安部由起子

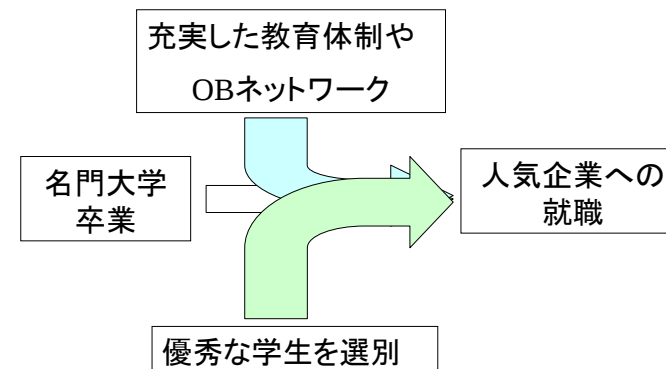
『雇用慣行の変化と女性労働』中馬宏之・駿河輝和編 所収、
東京大学出版会、pp.151-170 (1997)

- 「人気企業内定者の多くが名門大学出身」...学歴重視の傾向は続く。
- 「入学は難しいが卒業は簡単」、「大学はレジャーランド」との批判も

7

名門大学と人気企業への就職

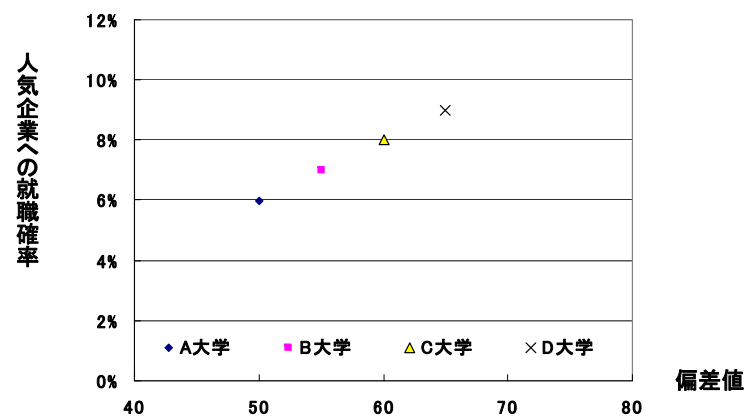
- 経済学における大学の役割とは？



8

横断面データによる分析

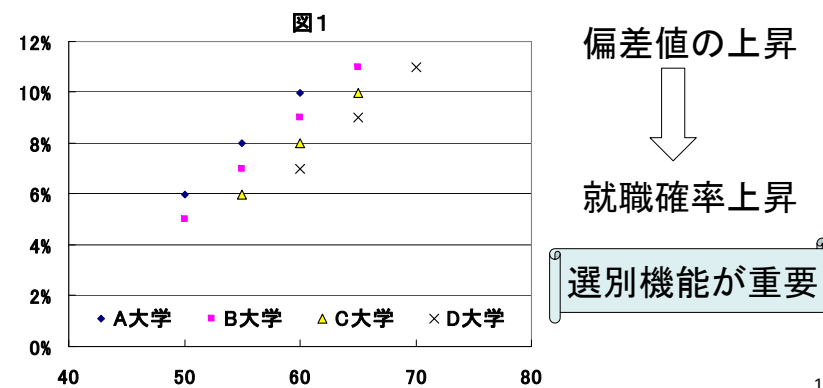
- 選別機能か教育機能を識別できない。



9

パネルデータの利用

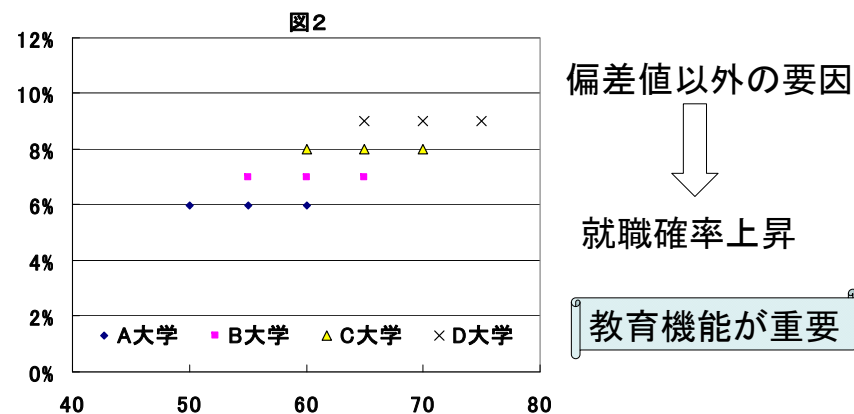
- 入学時の偏差値と就職市場での評価



10

パネルデータの利用

- 偏差値の変化が就職確率に影響しないとき



11

階差モデルによる推定

- 偏差値(X)の変化と人気企業への就職確率(Y)の変化

$$\Delta Y_{it} = a + b\Delta X_{it} + u_{it}$$

$$\Delta Y_{it} = Y_{it} - Y_{it-1} : \text{就職確率の変化}$$

$$\Delta X_{it} = X_{it} - X_{it-1} : \text{偏差値の変化}$$

12

固定効果モデル

- 大学・学部の変数を用いたモデル

$$Y_{it} = a + bX_{it} + \gamma_1 D_1 + \cdots + \gamma_n D_n + u_{it}$$

- ダミー変数とは？

$$D_i = \begin{cases} 1 & \text{if 第 } i \text{ 大学・学部 のとき} \\ 0 & \text{otherwise} \end{cases}$$

13

ダミー変数

id		年	偏差値	就職確率	D ₁	D ₂	D ₃	...
1	関学・経	1990	65	15%	1	0	0	...
1	関学・経	1991	66	16%	1	0	0	...
2	立命・営	1990	63	14%	0	1	0	...
2	立命・営	1991	62	13%	0	1	0	...
3	関大・総	1990	60	12%	0	0	1	...
3	関大・総	1991	61	12%	0	0	1	...

14

固定効果モデルと階差モデルの関連

- T=2のとき、両者は等しくなる。

$$Y_{it} = a + bX_{it} + \gamma_1 D_1 + \cdots + \gamma_n D_n + u_{it}$$

$$\rightarrow Y_{it-1} = a + bX_{it-1} + \gamma_1 D_1 + \cdots + \gamma_n D_n + u_{it-1}$$

$$\Delta Y_{it} = a + b\Delta X_{it} + \Delta u_{it}$$

$$\Delta Y_{it} = Y_{it} - Y_{it-1},$$

$$\Delta X_{it} = X_{it} - X_{it-1},$$

$$\Delta u_{it} = u_{it} - u_{it-1}$$

15

固定効果モデルの利点

$$Y_{it} = a + bX_{it} + \gamma_1 D_1 + \cdots + \gamma_{n-1} D_{n-1} + u_{it}$$

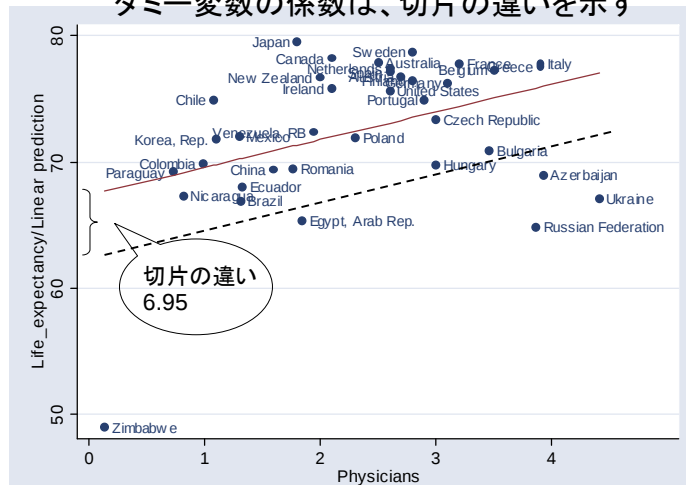
- (分析者にとって)観察不能な個体固有効果

→ 大学・学部固有のブランド力・教育効果の指標と考えることができる

16

ダミー変数の回帰分析

ダミー変数の係数は、切片の違いを示す



参考: 変量効果モデル

- 企業特殊要因を確率変数とみなしたモデル
– 誤差項の一部と考えたケース

$$Y_{it} = a + bX_{it} + v_i + u_{it}$$

確率変数: 誤差項として扱う

- 固定効果モデルでは...

$$Y_{it} = a + bX_{it} + v_i + u_{it}$$

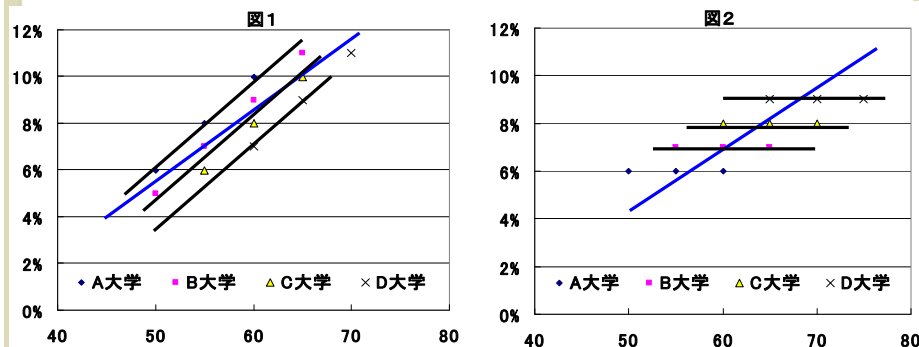
確定的な値と仮定

確率変数

18

推定結果の見方

- 固定効果を考慮しない回帰
(プーリング回帰): 図1,2ともにb: 大きい
- 固定効果モデル: 図1→b大きい、図2→b小さい



推定結果

- Y: ダイヤモンド社「人気ベスト100」への大学・学部別就職確率
- X: 入学時点の偏差値(河合塾)
– 国立と私立のサンプルを分割して分析

□ 推定結果(係数bの値)

- プーリング回帰: 0.170
- 固定効果モデル: 0.036
- 偏差値の変化の影響は小さい

20

考察

- 偏差値の変化  就職確率の変化

□結論

- 選別機能より「教育」機能が強い

➤留意点

- ここでの「教育」にはOBネットワークなどの「ブランド」も含む。

21

パネルデータ利用の難点

- データ数が多くなるため、コンピュータに高い計算能力が求められる。
- データセット作成(統計調査)に多大な費用がかかる。
 - データソースが限定される

22

日本における主なパネル調査

- 経済産業省「企業活動基本調査」
- 厚生労働省「21世紀出生児縦断調査」「21世紀成年縦断調査」
- 有価証券報告書・企業財務データ(日経NEEDS)
- 家計経済研究所「消費生活に関するパネル調査」
- 慶応大学「就業と生活に関するパネル調査」
 - ピンク: 大学院生は利用可能
 - 赤色: 学部生も利用可能

※東京大学社研アーカイブにも関連データ多数

<http://ssjda.iss.u-tokyo.ac.jp/>

23

実習編: 各事例の背景

- 米国の労働者パネル・データ
National Longitudinal Surveyの個票データ
- 都道府県パネルデータによる
小売店舗密度の分析: データ構築の実習
 - 小売店舗密度: 小売店舗数 / 人口

24

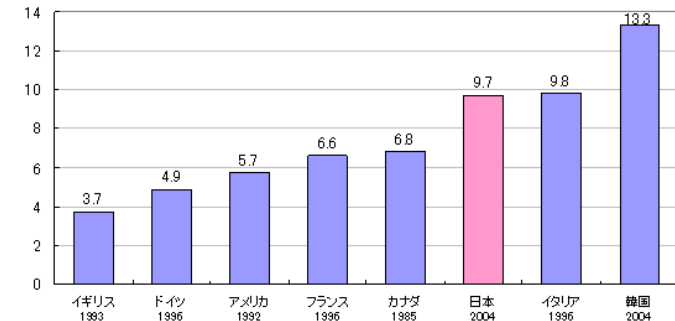
実習編: 各事例の背景

- National Longitudinal Surveyとは？
 - 米国労働統計局が1970年ごろより、収集している統計調査
 - 調査開始時点の、ある一定の年齢層を抽出し、その後の所得・消費・雇用など、経済社会生活状況を追跡調査したもの
 - 米国を代表する個人を対象とするパネル調査の一つ

25

小売店舗密度の分析

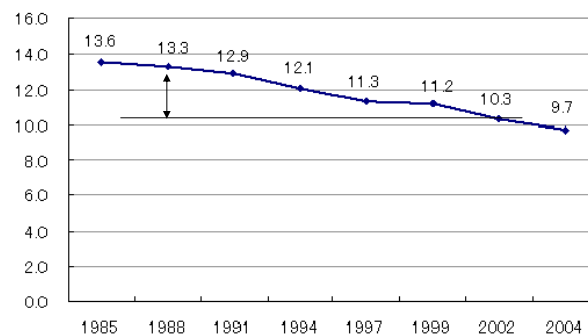
- OECD主要国の小売店舗密度



26

小売店舗密度の分析

- わが国の小売店舗密度の推移



1988年⇒2002年の間で、3ポイント低下

27

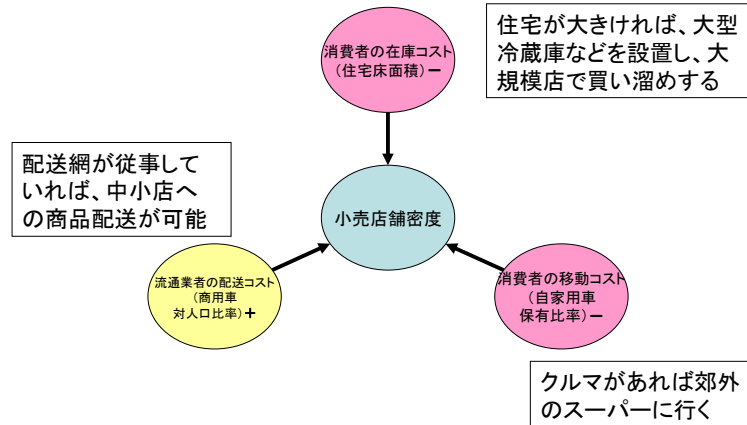
小売店舗密度の分析

- なぜ小売店舗密度が下がっているのか？
 - 大規模店が増加して、小規模店が駆逐された？
 - 規制改革で大規模小売店が増えた？
- 参考文献
 - 松井健二・成生達彦, 2003「わが国の小売店舗密度に関するパネル分析」『マーケティング・サイエンス』Vol.12, No.1,2, pp.44-61
 - Flath, D., 2003, "Regulation, Distribution Efficiency, and Retail Density," NBER Working Paper, No.9450

28

小売店舗密度の分析

どの要因が最も大きなインパクトを持つか？



29

小売店舗密度の分析

・ 推計式

$$-DOR_{it} = a + b_1 FSPACE_{it} + b_2 PCP_{it} + b_3 CVP_{it} + b_4 PDP_{it} + v_i + u_{it}$$

- ・ DOR: 小売店舗数／人口
- ・ FSPACE: 1住宅あたり床面積(－)
- ・ PCP: 人口当たり自家用車保有台数(－)
- ・ CVP: 人口当たりトラック台数(＋)
- ・ PDP: 人口密度(人口が増えると密度は下がる, －)
- ✓5時点の都道府県パネルデータ

30

プーリング回帰モデル

. reg dor fspace pcv cvp pdp

Source	SS	df	MS		Number of obs =
Model	686.609252	4	171.652313		230
Residual	266.63875	225	1.18506111		F(4, 225) = 144.85
Total	953.248001	229	4.16265503		Prob > F = 0.0000
					R-squared = 0.7203
					Adj R-squared = 0.7153
					Root MSE = 1.0886

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
dor					
fspace	-.0221256	.0048994	4.52	0.000	-.0124711
pcv	-.0115206	.0008587	-13.42	0.000	-.0132127
cvp	.0321643	.002135	15.07	0.000	.0279571
pdp	-.1956353	.1043322	-1.88	0.062	-.4012284
_cons	8.970683	.4547158	19.73	0.000	8.074637

期待される符号と異なる！

固定効果モデル

xtreg dor fspace pcv cvp pdp, fe

Fixed-effects (within) regression
Group variable: prefecture

R-sq: within = 0.9288
between = 0.0496
overall = 0.0123

corr(u_i, xb) = -0.6217

Number of obs = 230
Number of groups = 46
obs per group: min = 5
avg = 5.0
max = 5

F(4,180) = 587.43
Prob > F = 0.0000

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
fspace	-.0444384	.0167126	-2.66	0.009	-.0774162
pcv	-.0108663	.0006663	-16.31	0.000	-.0121812
cvp	.0142042	.0031657	4.49	0.000	.0079575
pdp	-1.232561	1.268765	-0.97	0.333	-3.736126
_cons	20.44907	2.074932	9.86	0.000	16.35475

sigma_u = 2.5870204
sigma_e = .31301031
rho = .98557201 (fraction of variance due to u_i)

F test that all u_i=0: F(45, 180) = 56.48 Prob > F = 0.0000

期待どおりの符号になった！

31

どの要因が影響したか？

. tabstat fspace pcv cvp pdp if prefecture!=47, by(year)

Summary statistics: mean
by categories of: year

year	fspace	pcv	cvp	pdp
1988	102.7526	253.41	193.7835	1.277184
1994	106.262	352.3422	199.2968	1.268632
1997	107.9808	405.7068	189.8851	1.263655
1999	107.0716	431.3407	180.1249	1.263426
2002	106.5989	465.1061	169.1008	1.267735
Total	106.1332	381.9812	186.4382	1.268126

1) PCPの係数: -0.01

2) PCPの変化: 211

1) × 2) = -2.11

1988年⇒2002年の
低下幅: 3ポイント

うち、2.11は自家用車
保有比率の上昇による

PCPは、211
上昇している

32

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^2_1	$X^2_{2,1990}$	$Z_{1(1)1990}$
2	1991	$Y^2_{1,1991}$	$Y^2_{2,1991}$	X^2_1	$X^2_{2,1991}$	$Z_{1(1)1991}$
2	⋮	⋮	⋮	⋮	⋮	⋮
2	1995	$Y^2_{1,1995}$	$Y^2_{2,1995}$	X^2_1	$X^2_{2,1995}$	$Z_{1(1)1995}$
3	1990	$Y^3_{1,1990}$	$Y^3_{2,1990}$	X^3_1	$X^3_{2,1990}$	$Z_{2(1)1990}$
3	1991	$Y^3_{1,1991}$	$Y^3_{2,1991}$	X^3_1	$X^3_{2,1991}$	$Z_{2(1)1991}$
3	⋮	⋮	⋮	⋮	⋮	⋮
3	1995	$Y^3_{1,1995}$	$Y^3_{2,1995}$	X^3_1	$X^3_{2,1995}$	$Z_{2(1)1995}$
4	1990	$Y^4_{1,1990}$	$Y^4_{2,1990}$	X^4_1	$X^4_{2,1990}$	$Z_{2(1)1990}$
4	1991	$Y^4_{1,1991}$	$Y^4_{2,1991}$	X^4_1	$X^4_{2,1991}$	$Z_{2(1)1991}$
4	⋮	⋮	⋮	⋮	⋮	⋮
4	1995	$Y^4_{1,1995}$	$Y^4_{2,1995}$	X^4_1	$X^4_{2,1995}$	$Z_{2(1)1995}$

Stata入門 P.63

33

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

34

実習編までのwarming up(宿題)

- 以下2つのdo-fileを完成させよ
 - NLSY-reg.do
 - Stata社提供の米国・個人パネル調査のデータ・ファイル <http://stata-press.com/data/r11/nlswork.dta>を使った回帰分析
 - 黒人ダミー-blackを作成せよ。人種はrace, race==1のとき白人、2のとき黒人、3のとき、その他
 - 賃金(ln wage)を被説明変数とする回帰分析を実施し、労働組合加入の賃金プレミアムを計測せよ。説明変数は、grade:就学年数、union:組合加入ダミー、age:年齢、tenure:勤続年数、black:黒人ダミーとする
 - retail-density.do
 - retail-density.csvを読み込む
 - 以下の回帰式をregコマンドで推定せよ
 - $dor = a + b_1 \text{ fspace} + b_2 \text{ pc} + b_3 \text{ cvp} + b_4 \text{ pdp} + u$

35