

パネル・データの分析

内容

- パネル・データとは
- pooled cross section dataの分析
- DID (Difference in Differences)モデル
- パネル・データの分析
 - 階差モデル (first difference model)
 - 固定効果モデル (fixed effects model)
 - 変量効果モデル (random effects model)
- パネル分析の実際
 - データ・セットの作成
 - 推定
 - Hausman 検定

パネル・データとは

- 計量分析におけるデータの構造
 - クロスセクション・データ
 - 時系列データ
 - パネル・データ
- パネル・データ
 - クロスセクション・データ × 時系列
 - 同一の家計（企業…）を時系列的に追ったデータ
 - 都道府県別データ（県民所得，物価，地価等のデータ）を数年分あつめたデータ→疑似パネルデータとよばれる
 - 本当の意味でのパネルデータではなく，集計データの時系列だから
- pooled cross section data
 - 複数時点のクロスセクション・データのあつまり
 - 調査対象は異なる時点では同一ではない
 - パネル・データとは異なる

パネル・データとは(2)

wagepan.xls : 賃金や労働時間についてのパネル・データ

nr	year	black	exper	hisp	hours	lwage	married
13	1980	0	1	0	2672	1.19754	0
13	1981	0	2	0	2320	1.85306	0
13	1982	0	3	0	2940	1.344462	0
13	1983	0	4	0	2960	1.433213	0
13	1984	0	5	0	3071	1.568125	0
13	1985	0	6	0	2864	1.699891	0
13	1986	0	7	0	2994	-0.72026	0
13	1987	0	8	0	2640	1.669188	0
17	1980	0	4	0	2484	1.675962	0
17	1981	0	5	0	2804	1.518398	0
17	1982	0	6	0	2530	1.559191	0
17	1983	0	7	0	2340	1.72541	0
17	1984	0	8	0	2486	1.622022	0
17	1985	0	9	0	2164	1.608588	0
17	1986	0	10	0	2749	1.572385	0
17	1987	0	11	0	2476	1.820334	0
18	1980	0	4	0	2332	1.515963	1
18	1981	0	5	0	2116	1.735379	1
18	1982	0	6	0	2500	1.631744	1
18	1983	0	7	0	2474	1.998229	1

個人を識別する変数：同一個人は同じ値

特定の個人についての複数年のデータ：
人種ダミー，経験年数，労働時間，賃金，結婚ダミー

時間を通じて変化しない変数もある（人種ダミー等）。労働時間，賃金は年によって異なる値。

パネル・データ利用の利点

- サンプル数の増加，変数の変動性が大きくなる
 - 推計値の信頼度が高まる
- クロスセクション・データやマクロ時系列データでは個々の主体の異質性をうまく捉えられない。
 - パネル・データを利用することで克服できる場合もある
- 動学的効果を捉えることができるかもしれない
 - マクロ時系列データ
 - 集計された経済主体の行動（個々の主体の意思決定がうまく捉えられない）
 - 社会資本整備の効果→地価に反映される→社会資本整備計画のアナウンスの前後のデータがあれば

Pooled cross section data の分析

- cps78-85.xls
- 教育の収益率の推計
 - 1978年と1985年の2時点のデータ
 - 2時点で調査対象は異なる (pooled cross section data)
 - 被説明変数：賃金
 - 説明変数：教育年数, 経験年数, 女性ダミー等
- 複数時点のデータを単純にプールして分析することの問題点
 - 78年と85年を単純に比較できない
 - 全般的な賃金変化, マクロ経済ショックの存在
 - 時点ダミー (定数項ダミー) を用いて対処
 - 78年と85年で教育の収益率の違いがあるかもしれない
 - 時点ダミーと教育年数(EDUC)の交差項を用いる

poold cross section dataの分析

78年と85年の2時点のデータ

data: CPS78_85.xls
被説明変数: lwage

```
lm(formula = lwage ~ y85 + educ + y85educ + exper +  
expersq + union + female + y85fem)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4.589e-01	9.345e-02	4.911	1.05e-06	***
y85	1.178e-01	1.238e-01	0.952	0.3415	
educ	7.472e-02	6.676e-03	11.192	< 2e-16	***
y85educ	1.846e-02	9.354e-03	1.974	0.0487	*
exper	2.958e-02	3.567e-03	8.293	3.27e-16	***
expersq	-3.994e-04	7.754e-05	-5.151	3.08e-07	***
union	2.021e-01	3.029e-02	6.672	4.03e-11	***
female	-3.167e-01	3.662e-02	-8.648	< 2e-16	***
y85fem	8.505e-02	5.131e-02	1.658	0.0977	.

y85: 時点ダ
ミー (85年
なら1)

y85educ
=y85*educ

y85fem
=y85*fem

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4127 on 1075 degrees of freedom

Multiple R-squared: 0.4262, Adjusted R-squared: 0.4219

F-statistic: 99.8 on 8 and 1075 DF, p-value: < 2.2e-16

pooled cross section dataの分析(2)

前ページの回帰分析の結果について

- y85の係数 0.1178
 - 78年に比べ85年の賃金は11.78%高い
 - y85educの係数 0.01846
 - 教育の収益率（教育年数の1年の増加が何%の賃金の増加をもたらすか）は78年に比べ、85年は1.846%ポイント高くなっている
 - educの係数は0.07472→78年の教育の収益率は7.472%→85年の教育の収益率は $7.422 + 1.846 = 9.593\%$
 - y85femの係数 0.0851
 - 女性ダミーの係数が78年に比べ85年には0.0851大きくなる
 - femaleの係数 -0.3167 78年において、女性は他の変数が同じ値をとる男性に比べ賃金が31.67%低い
 - しかし、85年になると男性との賃金格差は8.51%ポイント縮小して、23.16% ($= -0.3167 + 0.0851 = -0.2316$)になった。
- (→定数項ダミー，傾きのダミーの使い方)

DID estimator (difference in differences estimator)

- 日本語訳は「差分の差分法」
- 少なくとも2時点のクロスセクションデータがあり（パネルデータである必要はない），そのうち一方は2時点間の間にある変数の影響を受けているが（このグループをtreatment group: 処置群とよぶ），他方は影響を受けていないとする（このグループをcontrol group: 対照群とよぶ）。
- DIDは，treatment group とcontrol groupの反応の差をみることで，ある変数の影響をとらえようとする手法
- 疑似的な実験
- 例）社会資本整備の影響をはかるために，社会資本整備の恩恵を受ける地域の世帯とそうでない地域の世帯の比較を行う

DID estimator (2)

- kielmc.xls
 - ゴミ焼却場の建設が住宅価格に与える影響
 - 1978年, 81年の2時点のデータ (同一の住宅を追跡したものではない)
- 1978年時点では焼却場の建設の噂なし, 81年時点ではあり (実際には85年に建設)
 - 2時点間の間にあるeventが生じた
 - 焼却場から離れた地域はeventの影響を受けない (control group)
 - 焼却場の近くの住宅はeventの影響を受ける (treatment group)
- 一種の自然実験
 - eventの実施の前後でcontrol groupに比較したtreatment group の反応の違いをみる
 - 完全な実験であれば, control group とtreatment groupは同質な集団でなければならない。このケースでは, 一般にゴミ焼却場は中心部から離れた地域に建設されるので, ゴミ焼却場が建設されようがされまいが, 住宅価格は安くなることに注意。ここでは, その効果をダミー変数 (ゴミの焼却場の近く=辺鄙な場所) で処理すれば, 両グループはほぼ同質とみなせるとして分析する。

問題 1

- データ：kielmc.xls

- 住宅価格方程式の推計せよ

被説明変数： 住宅価格 (rprice: 1978年実質)

説明変数：

nearinc (住宅がゴミ焼却場の近くなら1) ,

y81(81年なら1というダミー変数)

y81nrinc (=y81 * nearinc)

住宅の質を表す変数 (広さ, 築年数等) , 周辺環境 等のその他の変数 (これらの変数は, nearinc=1と0の地域の住宅をほぼ同質の住宅とみなすための制御変数)

- 比較のため, 次の方程式も推計する

- 時点ダミーを用いずに単純にプールしたデータで回帰分析を行う
- 78年だけ, 81年だけのデータで回帰で回帰を行う

注) 全サンプルではなくサブセットで回帰を行うには `lm( )` のoptionで `subset=...` を使います。例えば81年のデータだけを用いるには

`lm(y ~ x1 + x2 + x3, subset=( y81 == 1))`

とします。(y81 == 1) は論理式で, y81が1に等しいときTRUEを返します。78年だけのデータを用いたい場合は (y81==0)に変更する。

パネル・データの分析

- crime2.xls

- 米国46市の失業率(unem)と犯罪発生率(crmrte), 1982年と1987年の2時点
- 同一の市を追跡したパネル・データ
- 犯罪発生率と失業率の関係を分析

他の条件が一定なら, 失業率の上昇は犯罪発生率を高めるはず→失業率の係数は正であるはず

- crmrte を被説明変数, unemを説明変数にしてOLSを, 各年ごと, およびプールしたデータで行うと, 失業率と犯罪発生率の間に予想した関係は見いだせるだろうか。

Rでの分析

- crime2.xls

- 欠損値がある（. が欠損値） → importの際に注意

- サブセットでの回帰

- d87: 87年なら1, 82年なら0というダミー変数を用いる
 - crime2がattachされているとして
 - 87年だけの回帰
 - `lm(y ~ x1 + x2 + x3, subset = (d87 == 1))`
 - 82年だけの回帰
 - `lm(y ~ x1 + x2 + x3, subset = (d87 == 0))`

犯罪発生率の回帰分析

- 82年

```
lm(formula = crmrte ~ unem, subset = (d87 == 0))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	84.569	10.904	7.756	9.07e-10	***
unem	1.307	1.027	1.272	0.21	

82年：失業率の係数は正だが、有意ではない

- 87年

```
lm(formula = crmrte ~ unem, subset = (d87 == 1))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	128.378	20.757	6.185	1.8e-07	***
unem	-4.161	3.416	-1.218	0.23	

87年：失業率が高いと犯罪発生率が低下？

- pool data

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	103.2434	8.0587	12.81	<2e-16	***
unem	-0.3077	0.9317	-0.33	0.742	

プールしたデータではマイナスの係数（ただし有意ではない）

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

パネル・データの分析 階差モデル

- 各都市の犯罪発生率に結び付く，各都市固有の観察不可能な変数があるかもしれない。
 - この変数は観察不可能だが，2時点間で変化しないとする。
 - この変数の影響を考慮するにはどうすれば良いか？
 - まず，次のようなモデルを考える

$$y_{i,t} = \alpha_i + \beta x_{i,t} + \gamma D_t + u_{i,t}$$

- α_i : 都市 i の固有の性質（観察不可能）
 - $x_{i,t}, y_{i,t}$: 都市 i , 時点 t の説明変数と被説明変数
 - D_t : 時点ダミー（1987年なら1 : マクロ経済ショックがあるかもしれない）
- 階差をとると問題は解決

$$\Delta y_{i,t} = \gamma + \beta \Delta x_{i,t} + \Delta u_{i,t}$$

観察不可能な変数 α_i が消去される

問題 2

- データセット : crime2.xls
- 1. 次の回帰を82年だけ, 87年だけ, 82年と87年のプールしたデータで行え
 - 被説明変数 : crmrte (人口1000人当たりの犯罪数)
 - 説明変数 : unem (失業率)
- 2. 次の回帰を行い, 失業率と犯罪発生率の関係について検討せよ
 - 被説明変数 ccrmrte (2時点間のcrmrte の変化)
 - 説明変数 cunem (2時点間のunemの変化)

階差モデルによる回帰の結果

```
lm(formula = ccrmrte ~ cunem)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-36.912	-13.369	-5.507	12.446	52.915

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	15.4022	4.7021	3.276	0.00206	**
cunem	2.2180	0.8779	2.527	0.01519	*

係数は正で有意

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 20.05 on 44 degrees of freedom  
(46 observations deleted due to missingness)
```

```
Multiple R-squared:  0.1267,    Adjusted R-squared:  0.1069
```

```
F-statistic: 6.384 on 1 and 44 DF,  p-value: 0.01519
```

パネル・データの分析方法

次のモデルを考える

$$y_{it} = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \cdots + \beta_k x_{k,it} + u_{it}$$

i : 個人 i

t : 時点 t

y_{it} : 被説明変数 個人 i , 時点 t

$x_{j,it}$: 説明変数 $x_j(j=1,2,\dots,k)$ 個人 i , 時点 t

u_{it} : 誤差項 個人 i , 時点 t

u_{it} の想定 $\rightarrow$ pooling dataでのOLS, fixed effects model, random effects model で異なる

固定効果モデル (fixed effects model) の想定

$$y_{it} = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \cdots + \beta_k x_{k,it} + u_{it}$$

$$u_{it} = \alpha_i + v_{it}$$

$$E[v_{it}] = 0, \text{var}[v_{it}] = \sigma^2, \text{cov}[v_{it}, x_{j,it}] = 0$$

$$\text{but } \text{cov}[\alpha_i, x_{j,it}] \neq 0$$

$$\Rightarrow \text{cov}[u_{it}, x_{j,it}] \neq 0$$

α_i : individual effect

個人 i のindividual effectと説明変数の間に相関があると、誤差項 u_{it} と説明変数が独立だという最小二乗法的前提が満たされない

例) 賃金方程式の推計で、個人の生来の資質 (individual effect) は観察できない。しかし、個人の生来の資質は学歴と相関があるかもしれない。

変量効果モデル (random effects model) の 想定

$$y_{it} = \beta_0 + \beta_1 x_{1,it} + \cdots + \beta_k x_{k,it} + u_{it}$$

$$u_{it} = \alpha_i + v_{it}$$

$$\text{cov}(\alpha_i, x_{j,it}) = 0$$

α_i : individual effect

- fixed effects model random effects modelの違いは、individual effectと説明変数に相関があるかないかという想定の違い
- もともとは、文字通りの意味で、fixed effects は非確率変数（固定値）、random effectsは確率変数という想定だった。今となってはミスリーディングな用語である。

パネル分析 その他

- 一般的には, individual effects は個々の主体 (クロスセクション) に関連
- 場合によっては, 時点特有のショックをコントロールするために, 時点に関連させる場合もある
- unobserved effects model
 - 個々の主体に特徴的な, 観察されない変数の影響をうまく処理するための手法
 - 狭い意味でのパネル・データの分析に限定されない

FE(fixed effects) modelの推定方法

$$y_{it} = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \cdots + \beta_k x_{k,it} + \alpha_i + v_{it}$$

各変数のtに関しての平均を求め、平均からの乖離をとる
(within estimatorともよばれる)

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{1,it} + \beta_2 \ddot{x}_{2,it} + \cdots + \beta_k \ddot{x}_{k,it} + \ddot{v}_{it}$$

ただし、 $\ddot{y}_{it} = y_{i,t} - \bar{y}_i$, $\ddot{x}_{it} = x_{i,t} - \bar{x}_i$

あるいは、1階の階差モデルを考える

FD(first differenced) estimator

$$\Delta y_{it} = \beta_1 \Delta x_{1,it} + \beta_2 \Delta x_{2,it} + \cdots + \beta_k \Delta x_{k,it} + \Delta v_{it}$$

ただし $\Delta y_{it} = y_{it} - y_{it-1}$, $\Delta x_{j,it} = x_{j,it} - x_{j,it-1}$

こうした変形で、individual effectを表す項は消去され、他のパラメータを推定できる。

FE modelの推定方法(2)

- 自由度
 - N 個のクロスセクション, T 時点のデータ, k 個の説明変数
 - $N*(T-1)-k$ の自由度になることに注意
 - 一階の階差をとると初期時点の階差データはない。平均からの乖離をとると, T 時点の観測のうち独立なのは $T-1$ 個になる
- 説明変数
 - 時間を通じて一定の値をとる変数の効果は分析できない
 - 例) 教育の収益率
 - 学歴や性別, 人種は一定の値 $\rightarrow$ 教育の収益率, 性・人種の違いが賃金に与える影響は, FEモデルでは推定できない
- FD estimator (階差モデルの推定量) か within estimator か
 - 誤差項の性質が異なる
 - 時点の数が少ないときはあまり大きな問題ではない
 - v_{it} の系列相関が高い時 $\rightarrow$ FD estimator
 - そうでなければ within estimator

RE (random effects) modelの推定

$$u_{it} = \alpha_i + v_{it}$$

誤差項の分散共分散行列
OLSの前提が満たされない

$$\text{cov}(x_{j,it}, \alpha_i) = 0$$

$$E[\alpha_i] = 0, E[v_{it}] = 0, \text{var}[\alpha_i] = \sigma_\alpha^2, \text{var}[v_{it}] = \sigma_v^2$$

$$\text{cov}[\alpha_i, v_{jt}] = 0 \text{ (for all } i, j, t), \text{cov}[v_{it}, v_{js}] = 0 \text{ (if } i \neq j \text{ or } t \neq s)$$

$$\text{cov}[\alpha_i, \alpha_j] = 0 \text{ (if } i \neq j)$$

$$\Rightarrow E(u_{it}^2) = \sigma_\alpha^2 + \sigma_v^2, \quad E(u_{it}u_{is}) = \sigma_\alpha^2 \text{ (for } t \neq s) \text{ for each } i$$

Pooled dataでOLSを行うと、誤差項の均一分散の前提が崩れる→
RE model : 誤差項の分散共分散行列の想定からGLS(Generalized Least Square)で推定 (詳しい説明は上級の計量経済学の教科書を参照すること)

パネル分析の実際

- wagepan.xls
 - 545 obs. × 8年 = 4360 obs.
 - まず、クロスセクションデータとしてimportし、その後、パネル・データに変換する
- 回帰式の推定
 - pooling model
 - fixed effects model
 - within estimator
 - first difference estimator
 - random effects model
 - FEモデルかREモデルか

パネル分析：R

- パッケージplmが必要
 - パッケージplmをインストールし、さらにlibrary(plm)でロードする
- データをクロスセクションデータとして読み込み、`pdata.frame()`でパネルデータに変換
 - `pdata.frame(データ, index=c(ID,period))`
 - `index`でクロスセクションIDと時間変数を指定（次ページ参照）
 - `plm.data()`は古くて推奨されなくなった
- 回帰分析
 - `plm(モデル式, データ名, モデルの指定)`

通常の線形回帰`lm()`の場合、データセットをattachしておけば、`lm()`中にデータセット名の指定は不要だったが、`plm()`ではattachできないので、必ずデータセット名を指定する必要がある。

R: クロスセクションデータ形式のファイルをパネルデータに変換し、パネル分析を行う

パッケージplmをインストール（インストールされていなければ）library(plm)でパッケージplmをロード。wagepanという dataframe形式のデータがあるとして

```
wagepan1 <- pdata.frame(wagepan, index = c("nr", "year"))
```

```
# pdata.frame( ) でパネルデータへ変換
```

```
# index はクロスセクションIDと時点を表す変数の指定
```

```
wage1_fe <- plm(lwage ~ exper, data=wagepan1, model="within" )
```

```
# plm(モデル式, データの指定, 分析方法の指定)
```

```
# model: 固定効果"within", 階差モデル"fd", ランダム効果"random",  
プーリングモデル"pooling"
```

```
# plm の結果をwage1_fe に保存
```

R: クロスセクションデータ形式のファイルをパネルデータに変換し, パネル分析を行う (2)

```
summary(wage1_fe) # 結果の要約
```

```
fixef(wage1_fe)
```

```
# 固定効果の出力 (クロスセクションごとの定数項)
```

```
# ランダム効果, プーリングデータでの推定
```

```
wage1_re <- plm(lwage ~ educ + exper, data=wagepan1, model =  
"random")
```

```
wage1_pool <- plm(lwage ~ educ + exper, data = wagepan1, model =  
"pooling")
```

- 結果の要約はsummary() で取り出す

R : panel分析で用いられる関数

変数 x : $x_{i,t}$ 個人 i , 時点 t

$\text{lag}(x)$: $x_{i,t-1}$ を返す

$\text{diff}(x)$: $x_{i,t} - x_{i,t-1}$ を返す

$\text{between}(x)$: $b_i = \frac{1}{T} \sum_{t=1}^T x_{i,t}$

$\text{Between}(x)$: $B_{i,t} = b_i$

$\text{Within}(x)$: $w_{i,t} = x_{i,t} - B_{i,t}$

問題 3

- データ : wagepan.xls

- 1980年から87年の8時点のデータ
- 賃金, 教育年数, 経験年数, 人種ダミー, 結婚ダミー, ...

1. 次のモデルをpooled OLSで推定せよ

- 被説明変数 $\ln wage$ (賃金の対数値)
- 説明変数 $educ$ (教育年数), $exper$ (経験年数), $expersq (= exper^2)$, $black$ (黒人ダミー), $hisp$ (ヒスパニック・ダミー), 結婚ダミー, 組合ダミー

2. random effects model で推計し, pooled OLSとの結果と比較せよ。

3. fixed effects modelで回帰を行え (注意: 学歴, 人種ダミー等を説明変数に加えることはできない)。

```
plm(formula = lwage ~ exper + expersq + married + union + educ +
      black + hisp, data = wagepan1, model = "pooling")
```

Balanced Panel: n = 545, T = 8, N = 4360

pooling model

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)	
(Intercept)	-0.03470561	0.06456900	-0.5375	0.5910	
exper	0.08917906	0.01011105	8.8200	< 2.2e-16	***
expersq	-0.00284866	0.00070736	-4.0272	5.742e-05	***
married	0.10766559	0.01569647	6.8592	7.897e-12	***
union	0.18007255	0.01712053	10.5179	< 2.2e-16	***
educ	0.09938779	0.00467760	21.2476	< 2.2e-16	***
black	-0.14384172	0.02355950	-6.1055	1.114e-09	***
hisp	0.01569798	0.02081119	0.7543	0.4507	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 1236.5

Residual Sum of Squares: 1005.8

R-Squared: 0.18659

Adj. R-Squared: 0.18528

F-statistic: 142.613 on 7 and 4352 DF, p-value: < 2.22e-16

```
plm(formula = lwage ~ exper + expersq + married + union,
data = wagepan1, model = "within")
```

fixed effects model (within)

Balanced Panel: n = 545, T = 8, N = 4360

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)	
exper	0.11684669	0.00841968	13.8778	< 2.2e-16	***
expersq	-0.00430089	0.00060527	-7.1057	1.422e-12	***
married	0.04530333	0.01830968	2.4743	0.01339	*
union	0.08208713	0.01929073	4.2553	2.138e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

Total Sum of Squares: 572.05

Residual Sum of Squares: 470.2

R-Squared: 0.17804

Adj. R-Squared: 0.059852

F-statistic: 206.375 on 4 and 3811 DF, p-value: < 2.22e-16

```
plm(formula = lwage ~ exper + expersq + married + union, data =  
wagepan1, model = "fd")
```

first difference model

Balanced Panel: n = 545, T = 8, N = 4360

Observations used in estimation: 3815

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)	
(Intercept)	0.1157500	0.0195867	5.9096	3.729e-09	***
expersq	-0.0038824	0.0013863	-2.8005	0.005128	**
married	0.0381377	0.0229283	1.6633	0.096325	.
union	0.0427878	0.0196575	2.1767	0.029566	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 751.19

Residual Sum of Squares: 748.04

R-Squared: 0.0042035

Adj. R-Squared: 0.0034196

F-statistic: 5.36242 on 3 and 3811 DF, p-value: 0.0011047

```
plm(formula = lwage ~ exper + expersq + married + union + educ +
black + hisp,
data =wagepan1, model = "random")
```

random effects model

Balanced Panel: n = 545, T = 8, N = 4360

Coefficients:

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	-0.10746420	0.11070573	-0.9707	0.3316880
exper	0.11211949	0.00826087	13.5724	< 2.2e-16 ***
expersq	-0.00406885	0.00059183	-6.8751	6.195e-12 ***
married	0.06279512	0.01677285	3.7439	0.0001812 ***
union	0.10737885	0.01783001	6.0224	1.719e-09 ***
educ	0.10122461	0.00891329	11.3566	< 2.2e-16 ***
black	-0.14413069	0.04761483	-3.0270	0.0024698 **
hisp	0.02015107	0.04260112	0.4730	0.6362008

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 656.91

Residual Sum of Squares: 539.82

R-Squared: 0.17824

Adj. R-Squared: 0.17692

Chisq: 943.951 on 7 DF, p-value: < 2.22e-16

RE model か FE modelか

- 時間を通じて一定の変数の効果に興味がある場合はFE model を使った分析はできない
賃金方程式で教育の収益率をはかる場合→ RE model で推定
- RE model (変量効果モデル) かFE model (固定効果モデル) かの統計的検定

Hausman test

RE modelの残差と説明変数に相関があるかどうかの検定

Eviews ではRE modelのestimation outputのmenuから

View/ Fixed/Random Effects Testing/ Correlated Random Effects
-Hausman Test をたどれば自動的に検定してくれる

(H0: RE modelが正しい)

Rの場合: `phtest(FEモデル名, REモデル名)`

パッケージplmが必要

Stata fixed effects model の結果をメモリーにsaveする。menuから
Statistics/ Postestimation/ Specification/...