数理社会学会
第6回ワンステップアップセミナー

# パネルデータ分析の基礎と応用

三輪　哲
（東北大学）

# アウトライン

1. 横断的な問いと縦断的な問い
2. パネルデータとは
3. パネルデータ分析の意義
4. パネルデータの記述的分析
5. パネルデータの回帰分析１
6. パネルデータの回帰分析２
7. 誤差の分散共分散
8. さいごに

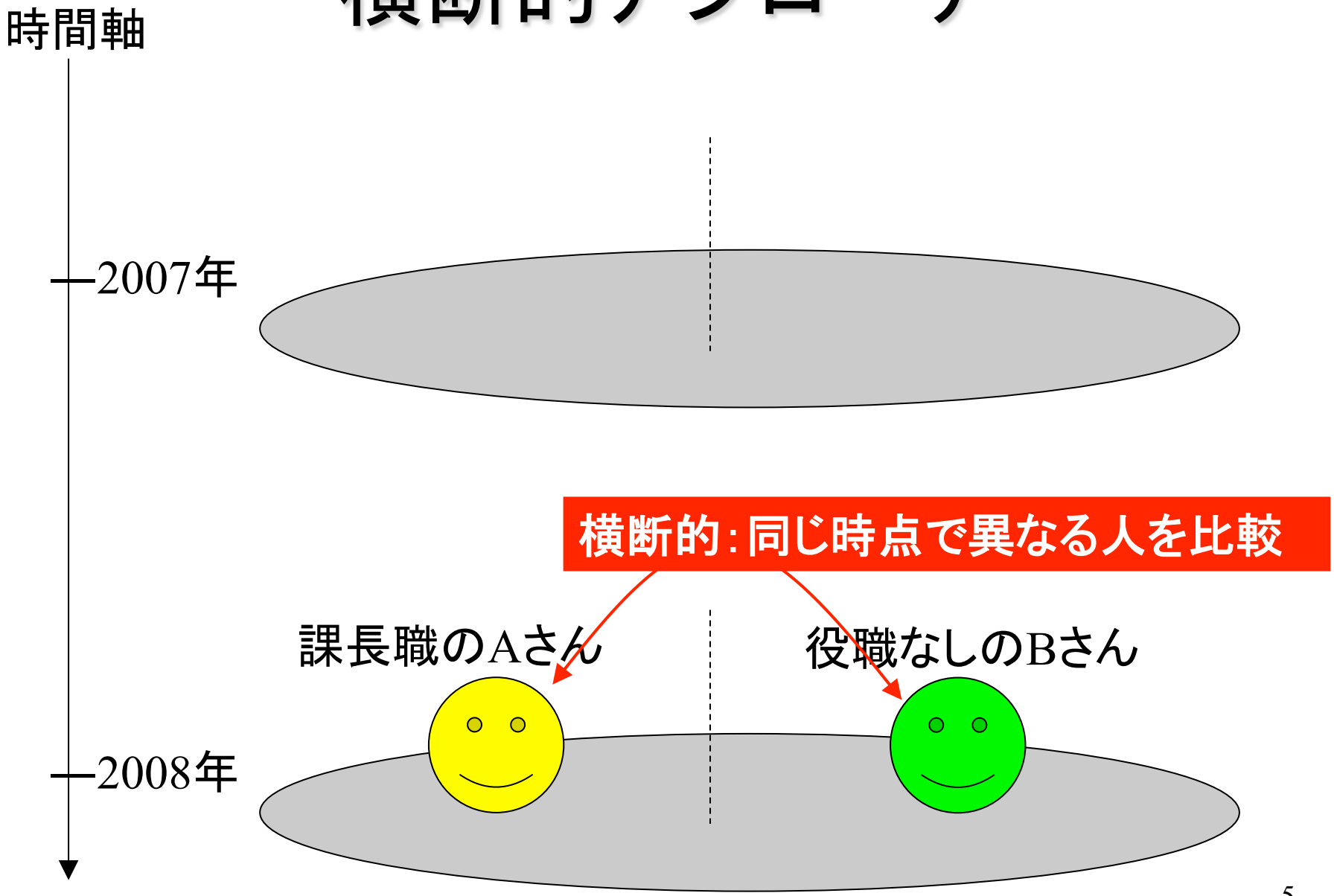
Section 1.

# 横断的な問いと縦断的な問い

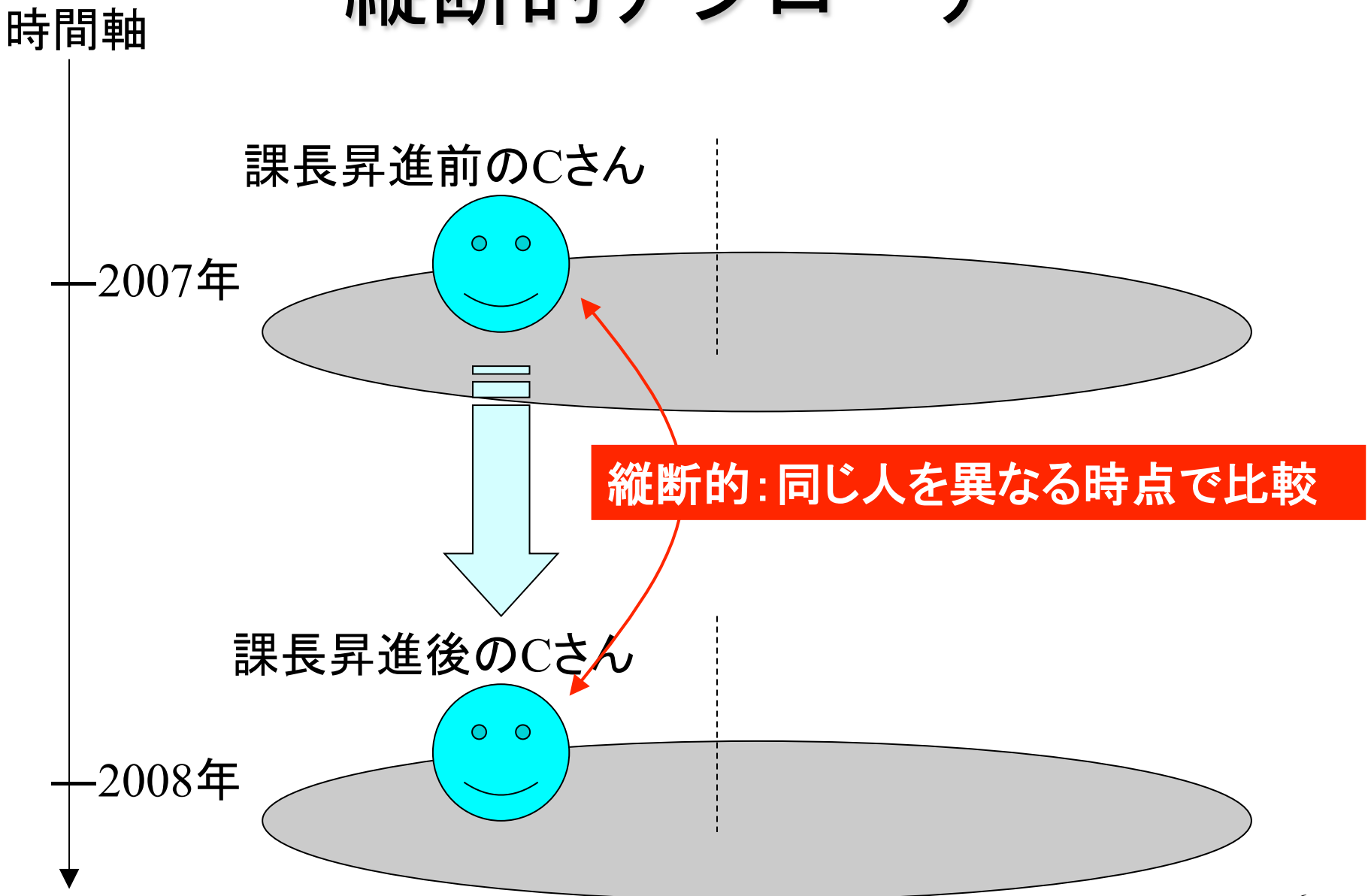
# 関連を見る2つのアプローチ

- 例：会社での役職と、主観的地位との間に、関係はあるのだろうか？
- 2つのアプローチ
  - 「役職の高い人は、役職の低い人よりも、自身の地位をより高く評価する」（横断的アプローチ）
  - 「高い役職へ昇進すると、自身の地位を高く評価するようになる」（縦断的アプローチ）

# 横断的アプローチ



# 縦断的アプローチ



# 縦断的仮説と横断的仮説

- 説明変数＝役職、被説明変数＝主観的地位
- 横断的な（betweenの）仮説
  - 「～なら（～という個体において）、～である」
  - 「役職の高い人は、自身の地位を高く評価する」
- 縦断的な（withinの）仮説
  - 「～すると、～になる」
  - 「昇進すると、自身の地位を高く評価するようになる」

# 横断的仮説

- ある1時点での、異なる個人間での、役職と主観的地位の関係を取り上げている

- 時間的変化についての仮説ではない
  - 仮説が肯定されても、無前提には変化について語れない
  - 例:「役職が変われば、主観的地位が変わる」とは無前提には言えない

# 縦断的仮説

- ある時点t1で、役職が低い人が
  その後の時点t2までに、出世して役職が高くなると…
- 時点t1よりも時点t2の方において、主観的地位が高くなる？

- 検証のためには、同一個体を追跡した複数時点の情報が必要　⇒パネルデータ

Section 2.

# パネルデータとは

# パネルデータとは

- 同一の個体が、複数時点にわたって観察されているデータのことを、パネルデータと呼ぶ

| id | 2007年の主観的地位 | 2007年の職業威信 |
|---|---|---|
| 1 | 6 | 52.2 |
| 2 | 5 | 45.6 |
| 3 | 2 | 47.2 |
| 4 | 8 | 64.6 |
| 5 | 7 | 84.3 |

# パネルデータとは

- 同一の個体が、複数時点にわたって観察されているデータのことを、パネルデータと呼ぶ

ワイド形式

| id | 2007年の主観的地位 | 2007年の職業威信 | 2008年の主観的地位 | 2008年の職業威信 |
|---|---|---|---|---|
| 1 | 6 | 52.2 | 5 | 47.2 |
| 2 | 5 | 45.6 | 6 | 53.1 |
| 3 | 2 | 47.2 | 3 | 47.2 |
| 4 | 8 | 64.6 | 8 | 42.4 |
| 5 | 7 | 84.3 | 6 | 84.3 |

# パネルデータとは

- パネルデータは、しばしば、このような形式に変換されて扱われることが多い

ロング形式

| id | 調査年 | 主観的地位 | 職業威信 |
|---|---|---|---|
| 1 | 2007 | 6 | 52.2 |
| 1 | 2008 | 5 | 47.2 |
| 2 | 2007 | 5 | 45.6 |
| 2 | 2008 | 6 | 53.1 |
| 3 | 2007 | 2 | 47.2 |
| 3 | 2008 | 3 | 47.2 |
| 4 | 2007 | 8 | 64.6 |
| 4 | 2008 | 8 | 42.4 |
| 5 | 2007 | 7 | 84.3 |
| 5 | 2008 | 6 | 84.3 |

# パネルデータとは

- パネルデータの表記

  i で個体をあらわし、

  t で時間をあらわす

  特にパネル調査のときに調査の次数のことをwaveと呼ぶ

  個体ごとに、かつ時間とともに変わる変数を $X_{it}$ としてあらわす

  主観的地位を変数Y、職業威信を変数Xとすると…

  $Y_{2,2007}$ は 5

  $X_{4,2008}$ は 42.4

| id | 調査年 | 主観的地位 | 職業威信 |
|---|---|---|---|
| 1 | 2007 | 6 | 52.2 |
| 1 | 2008 | 5 | 47.2 |
| 2 | 2007 | 5 | 45.6 |
| 2 | 2008 | 6 | 53.1 |
| 3 | 2007 | 2 | 47.2 |
| 3 | 2008 | 3 | 47.2 |
| 4 | 2007 | 8 | 64.6 |
| 4 | 2008 | 8 | 42.4 |
| 5 | 2007 | 7 | 84.3 |
| 5 | 2008 | 6 | 84.3 |

# パネルデータをつくる

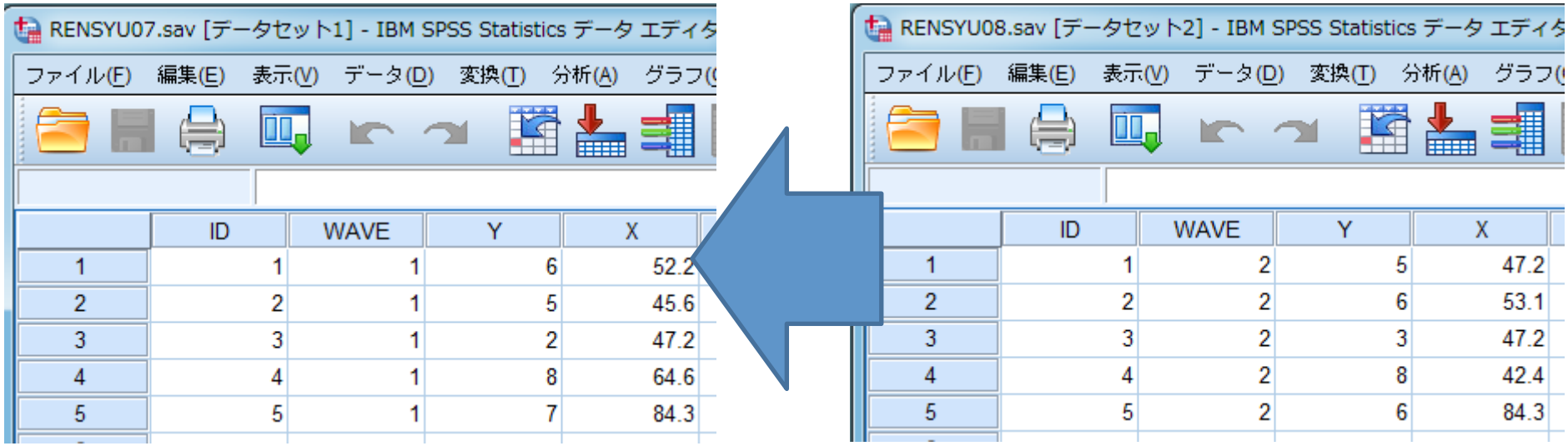
- 前のスライドのような、パネルデータ分析をしやすいロング形式のデータが、常に用意されているとは限らない
- そこで、自身でロング形式のデータへと変換できるようになることが、まず重要！

【手順】

１）データを合併して

２）ヨコ（wide）からタテ（long）へと変換

# 1）データの合併




第1波の調査データがあり、それと同じ対象者に対して追跡調査した第2波の調査データを合併する場合、「変数の追加」をおこなう

# 1）データの合併

## SPSSでのやりかた

*まず変数名を重複しないように変えておく（idの変数名は変えない、idの順にソートしておく）.

```
dataset activate データセット1.
ren var (Y=Y1)(X=X1)(WAVE=WAVE1).
exe.
dataset activate データセット2.
ren var (Y=Y2)(X=X2)(WAVE=WAVE2).
exe.

*そのあとで「右から」データ合併.
dataset activate データセット1.
match files /file=*
  /file='データセット2'.
exe.
```

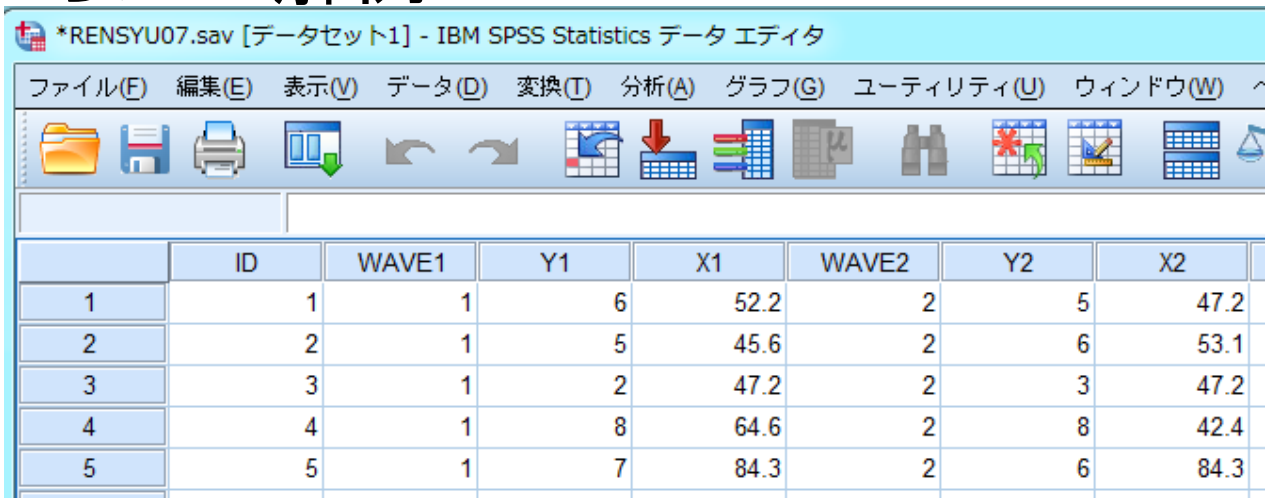
## STATAでのやりかた

```
* 07data no hensu mei wo kaeteoku
rename Y Y1
rename X X1
rename WAVE WAVE1
* 08data no hensu mei wo kaeteoku
rename Y Y2
rename X X2
rename WAVE WAVE2
save D:\RENSYU08M.dta
* migi kara data gappei
merge 1:1 ID using D:\RENSYU08M.dta
```

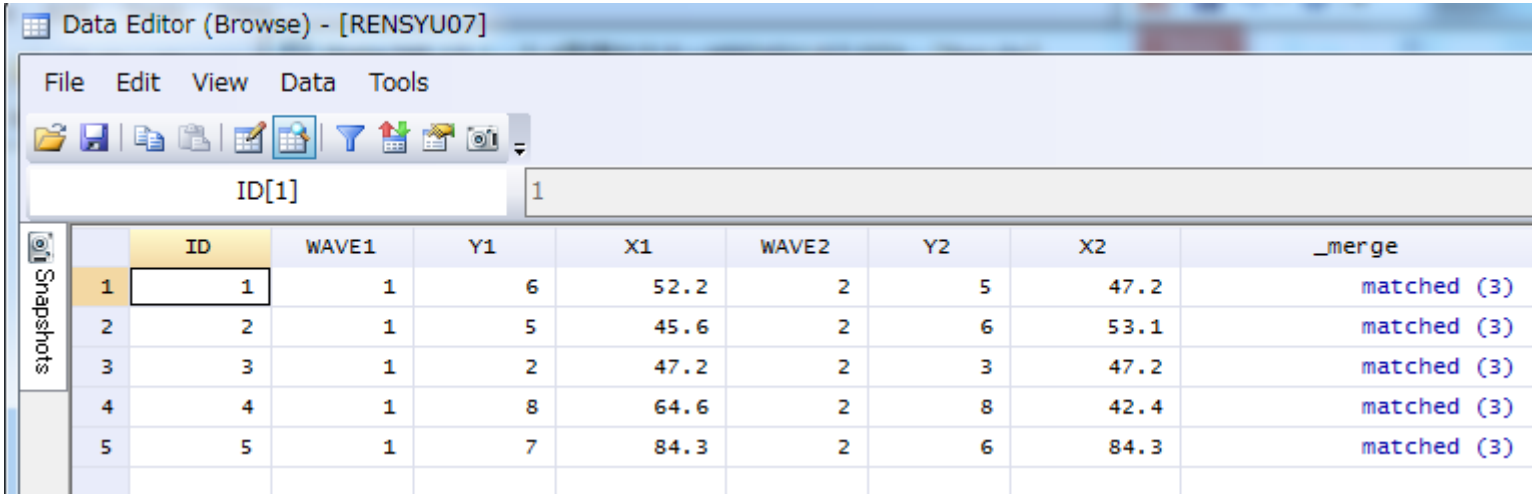
STATAでは、MERGEのあとの1:1を、1:mやm:1に変えることで、id番号の1対多対応付けや多対1対応付けが可能

# 1）データの合併

- これらが正解例



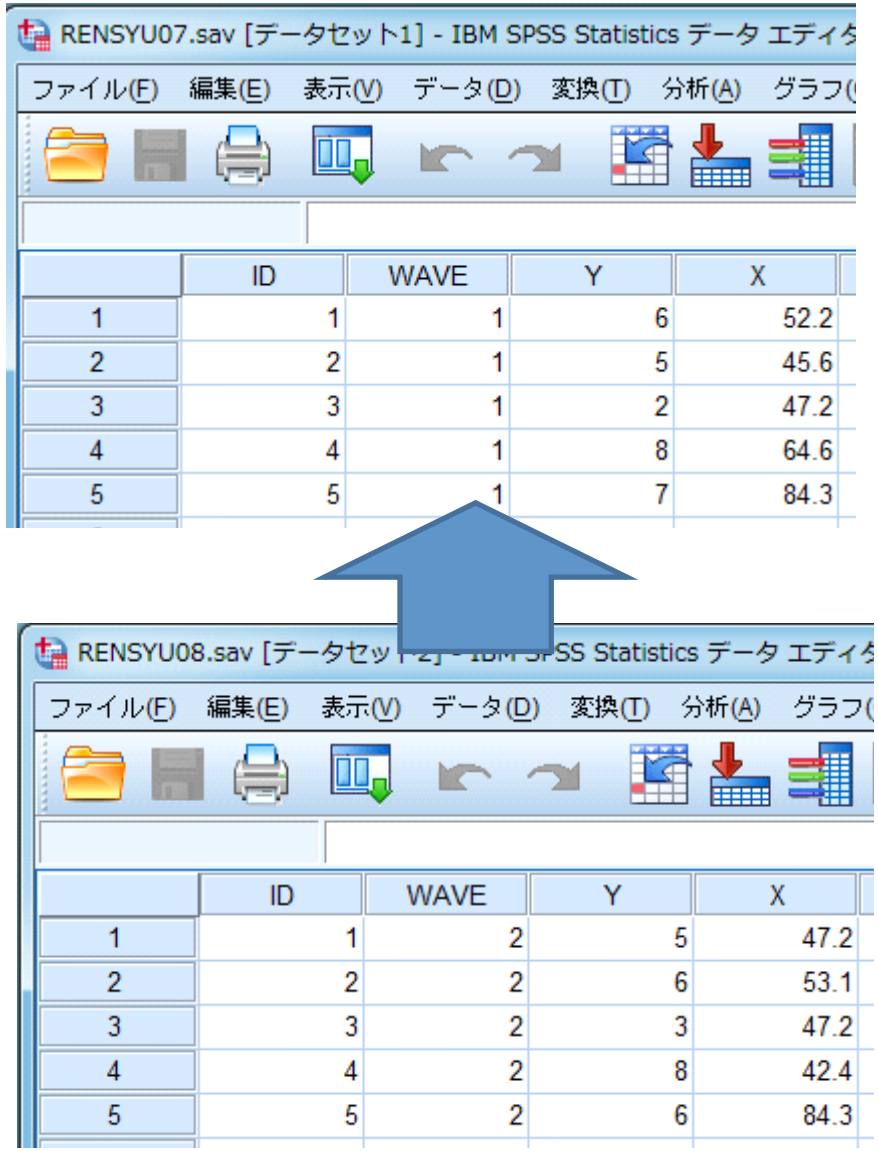

| | ID | WAVE1 | Y1 | X1 | WAVE2 | Y2 | X2 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 6 | 52.2 | 2 | 5 | 47.2 |
| 2 | 2 | 1 | 5 | 45.6 | 2 | 6 | 53.1 |
| 3 | 3 | 1 | 2 | 47.2 | 2 | 3 | 47.2 |
| 4 | 4 | 1 | 8 | 64.6 | 2 | 8 | 42.4 |
| 5 | 5 | 1 | 7 | 84.3 | 2 | 6 | 84.3 |




| | ID | WAVE1 | Y1 | X1 | WAVE2 | Y2 | X2 | _merge |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 6 | 52.2 | 2 | 5 | 47.2 | matched (3) |
| 2 | 2 | 1 | 5 | 45.6 | 2 | 6 | 53.1 | matched (3) |
| 3 | 3 | 1 | 2 | 47.2 | 2 | 3 | 47.2 | matched (3) |
| 4 | 4 | 1 | 8 | 64.6 | 2 | 8 | 42.4 | matched (3) |
| 5 | 5 | 1 | 7 | 84.3 | 2 | 6 | 84.3 | matched (3) |

# 1’）データの合併その2

- 合併でも、同一の対象者群を、別のケースであるとみなして、「ケースの追加」をおこなうこともできる



# 1’）データの合併その2

## SPSSでのやりかた

*もし変数名がそろっていなければ、先にそろえておく.

*下からデータ合併.
dataset activate データセット1.
add files /file=*
  /file='データセット2'.
exe.

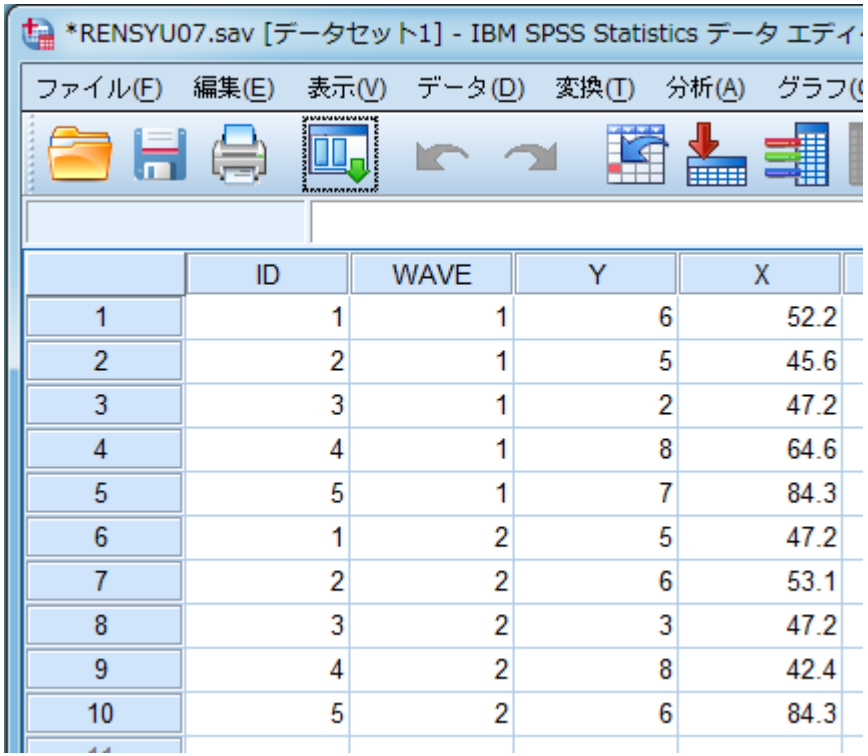
## STATAでのやりかた

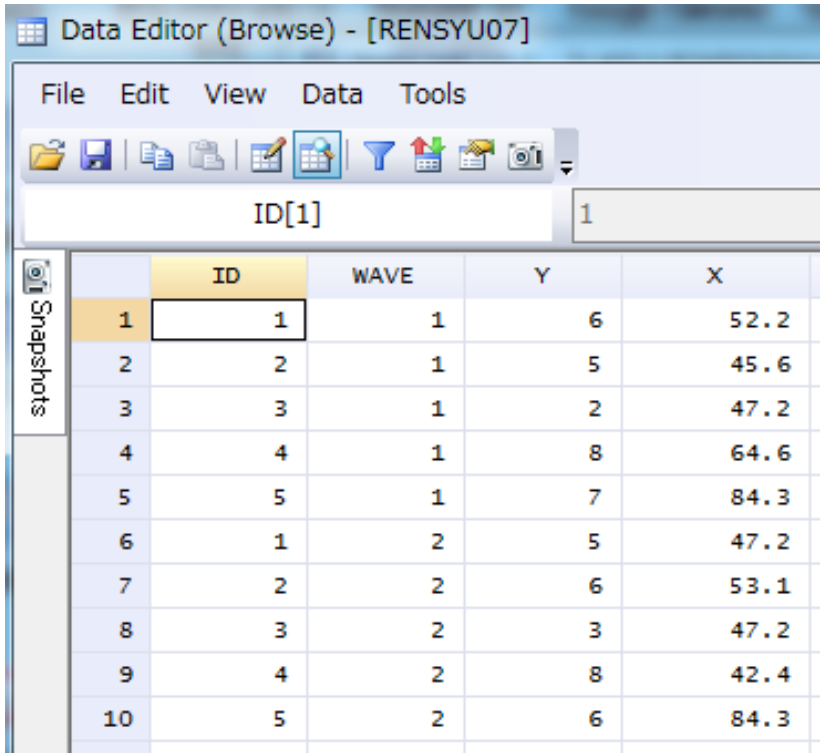
*hensu mei wo soroeteoku

*sita kara data gappei
append using D:\SUURI\RENSYU08.dta

# 1’）データの合併その2

- これらが正解例

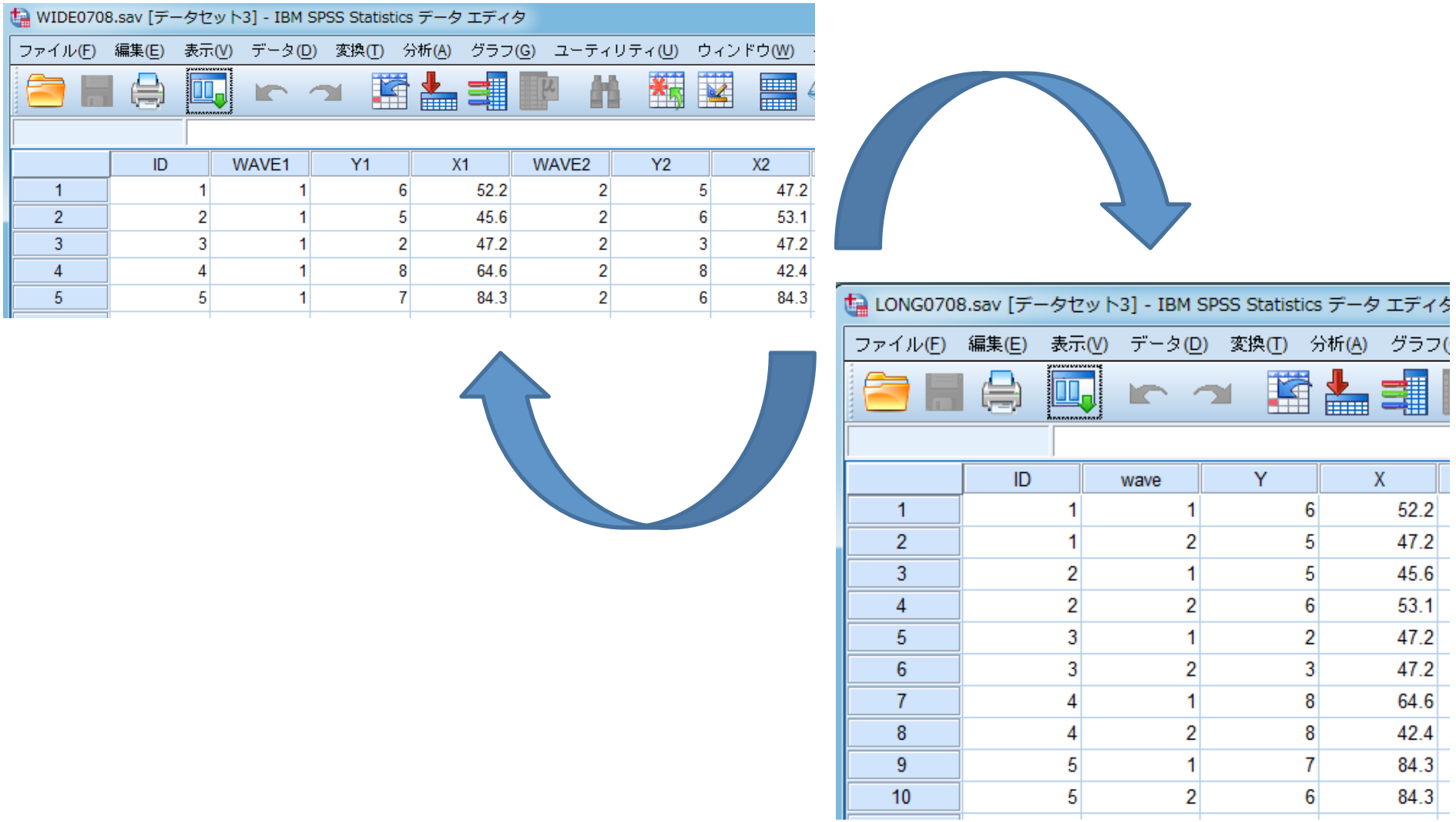


# 1）データの合併

- match file/merge　と　add file/append　では、どちらをおこなうべきか？
- いきなりロング形式のデータがつくれるという点で、　add file/append　のほうが都合がよい
  - 合併するファイル間で、変数名がそろっていなければならない
- ただし変数の加工しやすさや、分析法によっては、ワイド形式のほうが都合がよいことも
  - 合併するファイル間で、変数名がそろってはいけない

# 2）ヨコ（wide）からタテ（long）へ



# 2）ヨコ（wide）からタテ（long）へ

**SPSSでのやりかた**

```
*変数からケースへと変換.
varstocases
  /make Y from Y1 Y2
  /make X from X1 X2
  /make WAVE from WAVE1 WAVE2
  /keep=ID
  /null=keep.
```

**STATAでのやりかた**

```
/* yoko kara tate */
reshape long Y X ,i(ID) j(WAVE)
```

# 2’）タテ（long）からヨコ（wide）へ

**SPSSでのやりかた**

```
*ケースから変数へ.
casestovars
  /id=ID
  /index=WAVE.
```

**STATAでのやりかた**

```
/* tate kara yoko */
reshape wide Y X ,i(ID) j(WAVE)
```

# パネルデータをつくる

パネル調査によって得られるデータは、調査回数が重なるにつれて変数が膨大になり、PCの動きは重くなってしまう。
実際にパネルデータ分析するときは、変換を施したり、必要な変数だけに絞ることも多い。

**SPSSでのやりかた**

*変数の削除.

delete variables *VARNAME*.

*ケースの削除（下のかっこ内の条件にあてはまるケース以外を削除）.

select if (*VARNAME=1*).

**STATAでのやりかた**

*変数の削除.

drop *VARNAME*.

*ケースの削除（下のかっこ内の条件にあてはまるケースを削除、かっこは省略可）.

drop if (*VARNAME==1)*.

# 用いるパネルデータの調査概要

「働き方とライフスタイルの変化に関する全国調査」

(Japanese Life course Panel Survey; JLPS)

（東京大学社会科学研究所パネル調査）

- 大規模な全国調査
- 総合性のある社会調査
- 公開を前提とした調査
- 国際比較のできるような設計
- パネル調査（＝追跡調査）として設計

# 調査の概要　JLPS-M 2007

調査時期　2007年1月～
調査対象　2006年末時点で満35歳～40歳の
　　日本人男女
調査地域　日本全国
抽出台帳　住民基本台帳と選挙人名簿
抽出方法　層化2段無作為抽出法
　　全国をブロック（北海道、東北、関東…）と
　　市町村規模により層化し、271地点を無
作　　為に抽出
　　各地点で10名程度を無作為に抽出
※回収できない場合、同様の属性の予備対象に依頼

# JLPS-Mの回収状況

第1波（2007）では、
有効回収票数　1,433（アタック数の40.4%）
　※毎年調査に回答してもらうことを依頼した

第3波（2009）では、
有効回収票数　1,164（アタック数の86.0%）

より詳しくは、下記URL参照
http://ssjda.iss.u-tokyo.ac.jp/gaiyo/PM030g.html

# 今回使うのはsubsetデータ

- JLPS-Mの本物のデータも配布したが、今日用いるのはそれから作ったsubsetデータ

  変数を少なく絞り、それらのいずれもが3時点で有効回答された対象者に限定し、そこからさらに無作為抽出した400名のデータ（balanced）

【変数】 ID: 個人ID番号、 FEMALE: 女性ダミー、 EDUY: 教育年数、 PHED: 親高等教育ダミー、 WAVE: 調査の波、 SS: 主観的地位、 OPS95: 職業威信、 WSELF: 自営・経営、 WPART: 非正規、 HINC: 世帯年収、 HLTH: 主観的健康、 LSAT: 生活満足度、 SPOUSE: 配偶者有ダミー ※M_がついたものは、個人内平均値、D_がついたものは個人内平均からの偏差

Section 3.

# パネルデータ分析の意義

# パネルデータ分析の強み

（１）観察されない異質性の統制が可能

（２）個人レベルでの変化の分析が可能

（３）より精確な因果推論が可能

（４）情報量が豊富であることによる技術的強み

　自由度が多い、多重共線性の緩和、推計上の利点

# パネルデータ分析の弱み

クロスセクションデータ分析と比べると…

パネルデータ分析の弱みは、一切ない

- ただし、データの扱いや分析モデルの選択で難しさはあるので、とっつきにくいのは確か

パネル「調査」の弱み（困難）

（１）調査対象者の脱落（attrition）

（２）標本の代表性

（３）調査コストが高い

# 強み（１）：観察されない異質性の統制

- とりわけ、固定効果モデルならではの強み
  - 固定効果モデルとは？
    - 変数の個人内のばらつきだけを利用した回帰モデル

$$Y_{it} = \beta X_{it} + \gamma Z_i + u_i + \varepsilon_{it}$$

Y, Xは時間依存変数、Zは時間不変（個人レベル）変数とする、uは観察されない個人効果、εは誤差

– 上記の式の各項から、個人内平均を引くと・・・

$$(Y_{it} - \bar{Y}_i) = \beta(X_{it} - \bar{X}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

観察されない個人効果と、観察された個人効果が消え、個人内で変わりうるXの効果と誤差からYが決まっていることに

# 強み（１）：観察されない異質性の統制

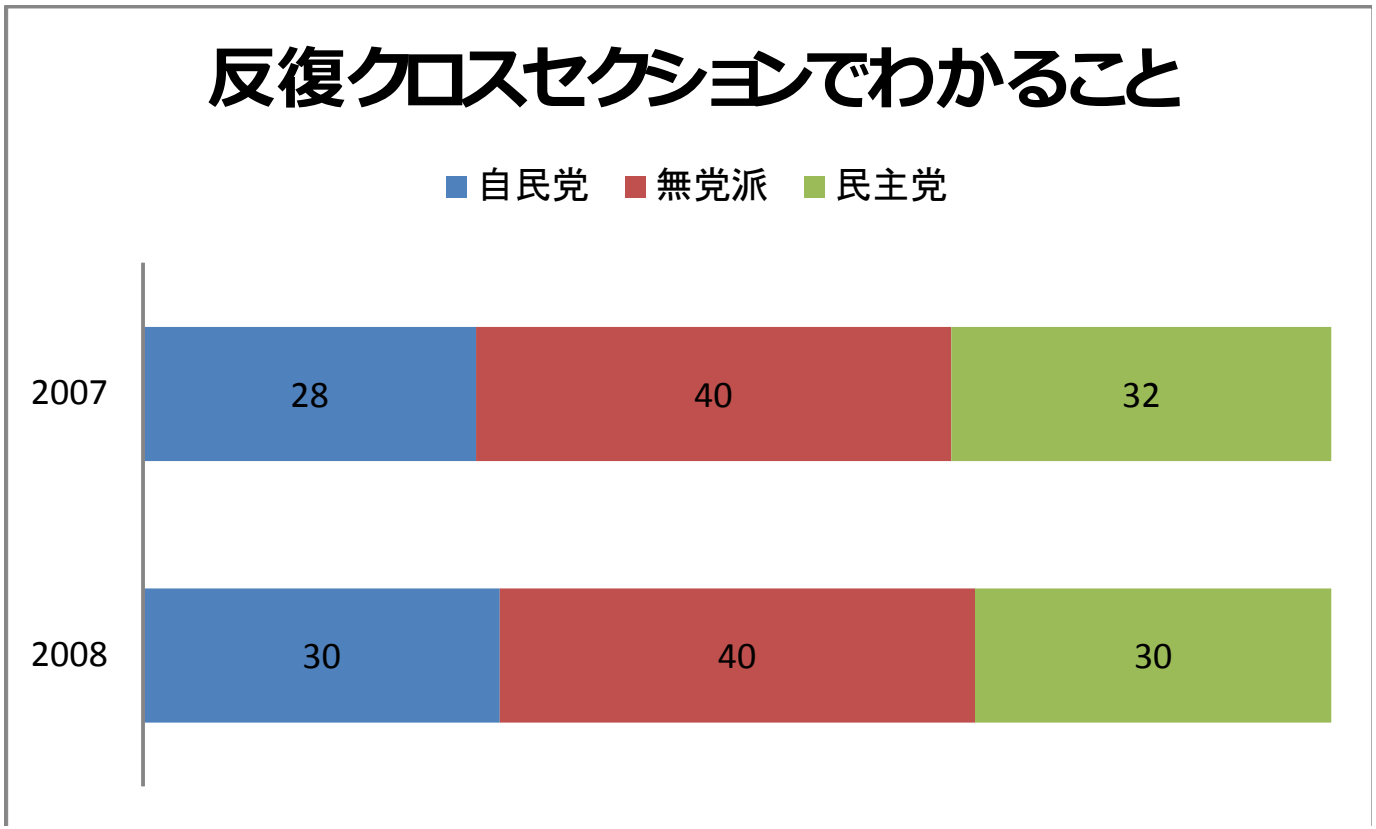
  - すると、もはや個人レベルの違いを示す変数（個人内では変わらない変数）は、Yの変化には何ら影響を与えない
  - 個人内で変わりうる変数も、個人内平均からの偏差をとったことで、個人間変動が取り除かれている
- このような、Yの変化を、Xの変化のみで説明しようとするモデルが、固定効果モデル

– 固定効果モデルのありがたみ

- 我々が分析するにあたり、統制すべき個人差のある変数をモデルに入れ損ねたり、それが調査票に入っていなかったり、測ることがそもそもできないなどで、悩まされる
- 固定効果モデルによって、そうした悩みなしに、Xの変化がいかほどYの変化をうながすか偏りのない推定ができる

# 強み（2）：個人レベルでの変化

- 変化をとらえるのには、反復クロスセクション調査でも十分と思われるかもしれない・・・



# 強み（２）：個人レベルでの変化

- 反復クロスセクションでは、社会の変化は見える
  - ただそれは、パネルデータでも見える（下表の周辺度数）
- しかし、誰が変化したか、なぜ変化したかには迫りえない
  - パネルなら、個人レベルで変化をとらえ、分析可能に

パネル調査でわかること

| | | 2008 | | |
|---|---|---|---|---|
| | | 自民党 | 無党派 | 民主党 |
| 2007 | 自民党 | 25 | 5 | 0 |
| | 無党派 | 4 | 21 | 15 |
| | 民主党 | 1 | 14 | 15 |

# 強み（３）：より精確な因果推論

- 縦断的なデータは、因果推論になじみやすい
- 他の個人差要因に撹乱されることなく、関連をとらえられる
- さらに、変数間に因果順序を仮定するのではなく、実際に測定された前後関係を活かして分析することもできる（例：ラグ変数の使用）
- そのうえ、個人のサンプルサイズが大きくなれば、偶発的な時間的共変動をとらえる危険も減る
- これらにより、パネルデータは、より精確な因果推論へと近づけさせてくれる強力な情報源といえる

Section 4.

# パネルデータの記述的分析

# はじめの一歩

- 最初に、ファイルを開こう

  PM030subset3_JAMS.sav　・・・SPSSデータファイル

  PM030subset3_JAMS.dta　・・・STATAデータファイル

  - subset1というのは、第1波のみのデータ
- STATA使用の方は、パネルデータであることの宣言を

  xtset ID WAVE

# パネルデータの記述的分析

- まずは、ロング形式のパネルデータを用いた、記述的分析について
- やるべきこと

  0）データの基本構造に関する記述

  そもそも調査へと誰がどれだけ回答したか？

  1）注目する1変数の分布に関する記述

  どのような値がどれくらい出現していたか？

  2）注目する1変数の変化に関する記述

  変化のパターンはどのようなものであったか？

# 0）データの基本構造

```
. xtdes

     ID:  13, 20, ..., 4755                                 n =       400
   WAVE:  1, 2, ..., 3                                      T =         3
          Delta(WAVE) = 1 unit
          Span(WAVE)  = 3 periods
          (ID*WAVE uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         3       3       3         3         3       3       3

     Freq.  Percent    Cum. | Pattern
------------------------------+---------
       400   100.00  100.00 | 111
------------------------------+---------
       400   100.00         | XXX
```

# 0）データの基本構造

## SPSSでのやりかた

* STATAのように簡単にはいかず、一度wide形式に直してから必要なものを出力する.

```
des ID WAVE.
casestovars /id=ID /index=WAVE /autofix=no.
recode WAVE.1 WAVE.2 WAVE.3
(1 thru 3=1) (else=0)
 into W1R W2R W3R.
compute T_i=sum(W1R, W2R, W3R).
fre T_i /percentiles=.05 .25 .50 .75 .95.
compute RP=W1R*100+W2R*10+W3R.
fre RP.
```

あえて無理矢理やるならこのような感じ…

## STATAでのやりかた

```
* kaitou no patern
xtdes
```

これのみでよい

# 1）注目する1変数の分布

- 量的変数の場合

```
. xtsum SS
```

| Variable | | Mean | Std. Dev. | Min | Max | Observations | |
|---|---|---|---|---|---|---|---|
| SS | overall | 5.22 | 1.59595 | 1 | 10 | N = | 1200 |
| | between | | 1.348469 | 1 | 8.666667 | n = | 400 |
| | within | | .8554068 | 1.553333 | 10.22 | T = | 3 |

# 1）注目する1変数の分布

## SPSSでのやりかた

```
* 量的変数の分布.
compute graSS=mean(SS).
aggregate /break=ID /groSS=mean(SS).
compute Wi_SS=SS-groSS+graSS.
ren var groSS=Bw_SS.
compute CHECK=lag(ID,2).
recode CHECK (sysmis=0).
if (ID~=CHECK) Bw_SS=$sysmis.
des SS Bw_SS Wi_SS.
```

これも無理矢理に

## STATAでのやりかた

```
* ryoteki hensu no bunpu
xtsum SS
```

# 1）注目する1変数の分布

- 質的変数の場合

```
. xttab LSAT
```

| | Overall | | Between | | Within |
|---|---|---|---|---|---|
| LSAT | Freq. | Percent | Freq. | Percent | Percent |
| -2 | 28 | 2.33 | 22 | 5.50 | 42.42 |
| -1 | 108 | 9.00 | 75 | 18.75 | 48.00 |
| 0 | 265 | 22.08 | 161 | 40.25 | 54.87 |
| 1 | 624 | 52.00 | 305 | 76.25 | 68.20 |
| 2 | 175 | 14.58 | 112 | 28.00 | 52.08 |
| Total | 1200 | 100.00 | 675 | 168.75 | 59.26 |
| | | | (n = 400) | | |

# 2）注目する1変数の変化

```
. xttrans HLTH, freq
```

| HLTH | HLTH -2 | -1 | 0 | 1 | 2 | Total |
|---|---|---|---|---|---|---|
| -2 | 1 | 1 | 0 | 0 | 0 | 2 |
| | 50.00 | 50.00 | 0.00 | 0.00 | 0.00 | 100.00 |
| -1 | 3 | 60 | 33 | 5 | 0 | 101 |
| | 2.97 | 59.41 | 32.67 | 4.95 | 0.00 | 100.00 |
| 0 | 1 | 59 | 199 | 74 | 5 | 338 |
| | 0.30 | 17.46 | 58.88 | 21.89 | 1.48 | 100.00 |
| 1 | 0 | 15 | 96 | 135 | 22 | 268 |
| | 0.00 | 5.60 | 35.82 | 50.37 | 8.21 | 100.00 |
| 2 | 0 | 0 | 10 | 34 | 47 | 91 |
| | 0.00 | 0.00 | 10.99 | 37.36 | 51.65 | 100.00 |
| Total | 5 | 135 | 338 | 248 | 74 | 800 |
| | 0.63 | 16.88 | 42.25 | 31.00 | 9.25 | 100.00 |

# 2）注目する1変数の変化

## SPSSでのやりかた

```
* 変数の変化.
compute LAGHLTH=lag(HLTH,1).
if (WAVE=1) LAGHLTH=$sysmis.
cro LAGHLTH by HLTH /cel=count row.
```

比較的無理矢理感が小さめ

## STATAでのやりかた

```
* hensu no henka
xttrans HLTH
```

# パネルデータの記述的分析

- いきなり回帰分析の推計にとびつかないで、まずはパネルデータの特徴をちゃんとみておくことは重要
  - 回答パターンから、脱落の特徴がわかる
  - 1変数の分布から、注目した変数のwithin、between、totalでのケース数や値のばらつきかたを知ることができる
  - 1変数の変化から、どのように値が移り変わっていくのか、平均的傾向が読み取れる

Section 5.

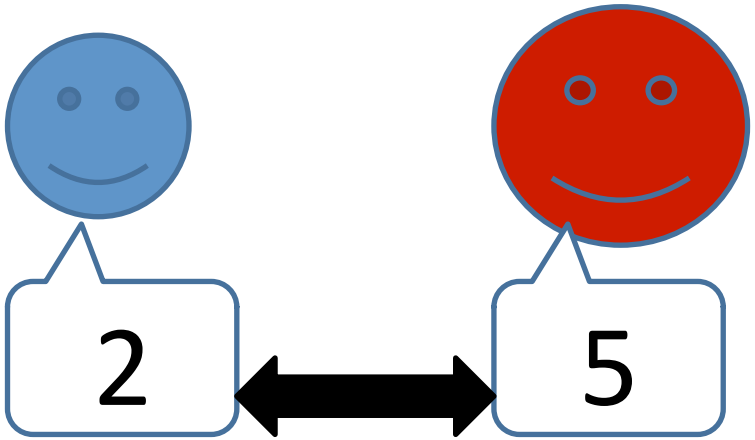
# パネルデータの回帰分析１
## ～BETWEENとWITHIN

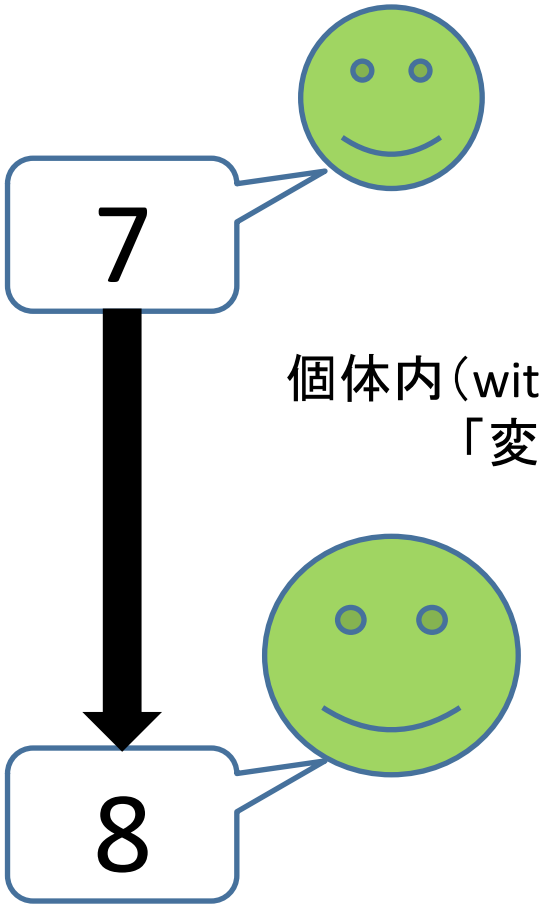
# betweenとwithin

- パネルデータの回帰分析結果を読む際には、何はともあれ、between（個人間での差）とwithin（個人内での差）を区別することが重要
- 現在用いられるパネル調査データのほとんどは、nが多く、Tが少ないデータゆえ、プールして分析したりマルチレベル分析をしても、クロスセクションの分析結果とあまり変わらない!?
- だが、withinでの推定（＝固定効果モデル）はまったく違う結果を生むかも
  - betweenの情報と独立、今まで触っていない部分

# betweenとwithin

個体間（between）の差は、
「異質性」

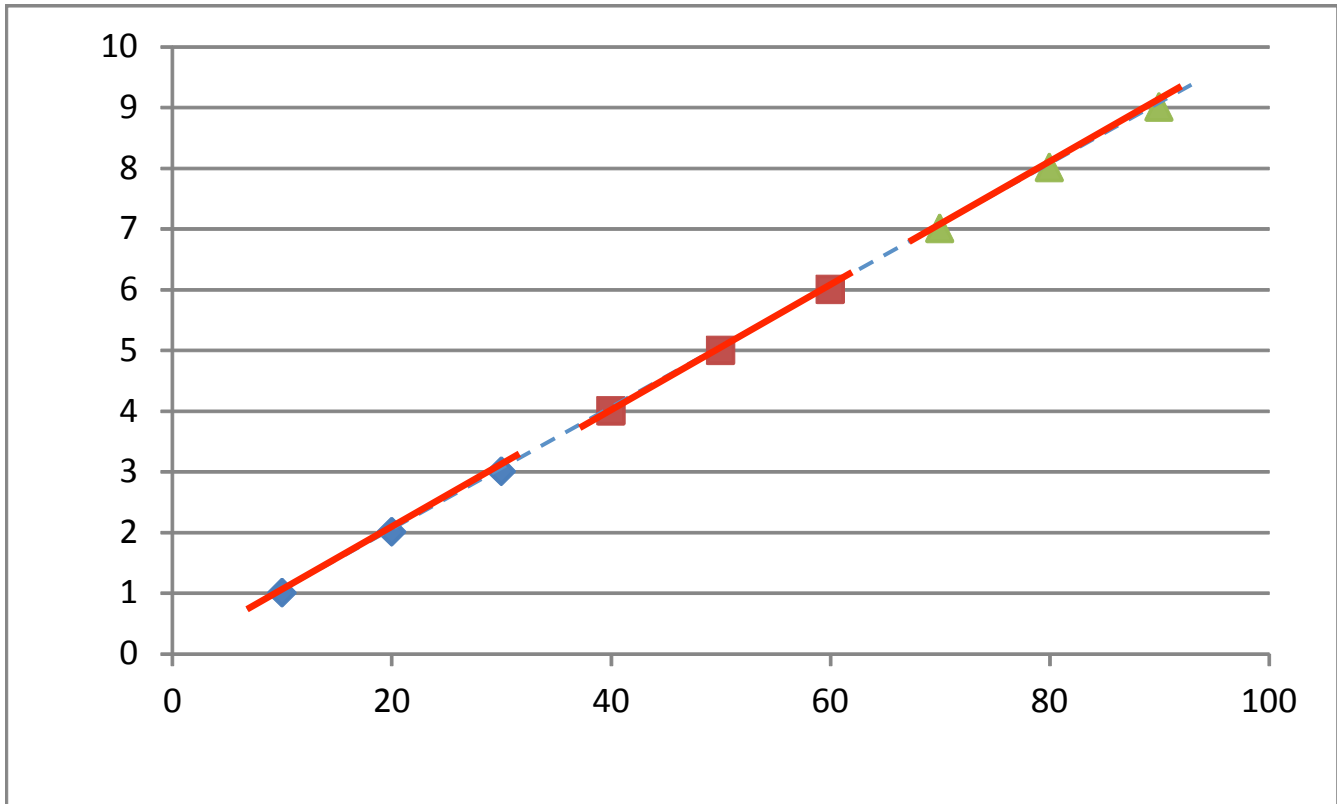



個体内（within）の差は、
「変化」

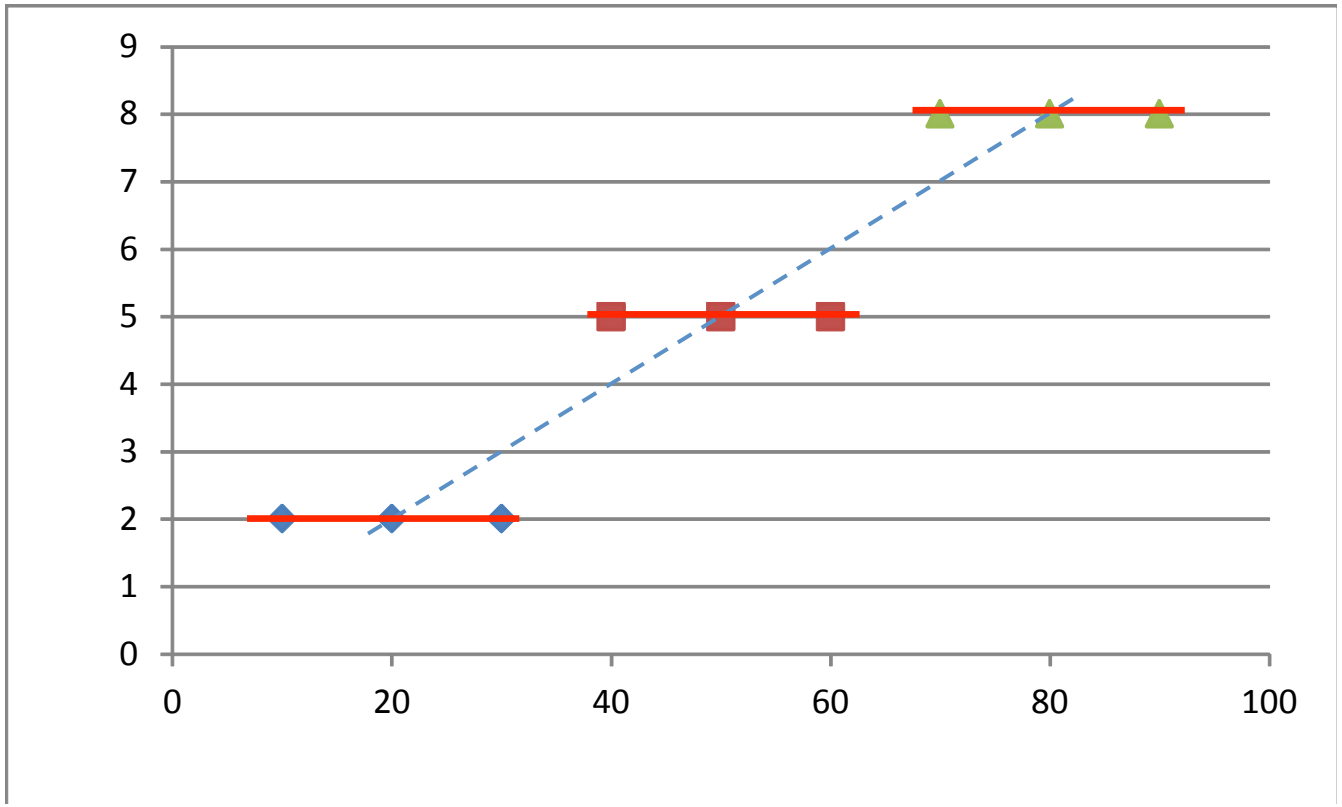
# betweenとwithin

- betweenの回帰係数とwithinの回帰係数がともに正の同じ値である場合



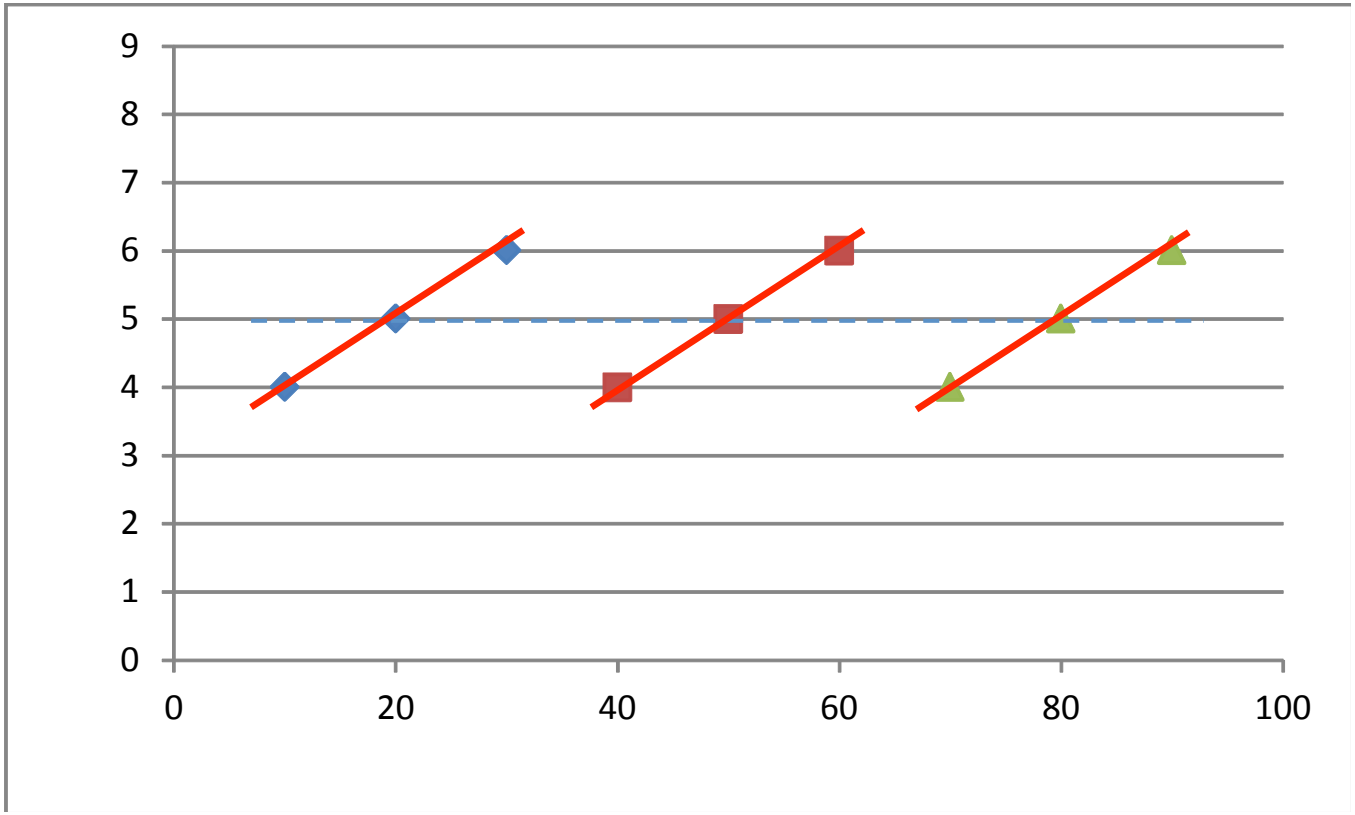
# betweenとwithin

- betweenの回帰係数だけが正で、withinの回帰係数が0の場合



# betweenとwithin

- betweenの回帰係数が0でwithinの回帰係数だけが正になる場合



# pooledの結果

```
. reg SS OPS95
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 303.7026 | 1 | 303.7026 |
| Residual | 2750.2174 | 1198 | 2.29567396 |
| Total | 3053.92 | 1199 | 2.54705588 |

| | |
|---|---|
| Number of obs | 1200 |
| F( 1, 1198) | 132.29 |
| Prob > F | 0.0000 |
| R-squared | 0.0994 |
| Adj R-squared | 0.0987 |
| Root MSE | 1.5151 |

| SS | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| OPS95 | .5813309 | .0505422 | 11.50 | 0.000 | .4821698 | .680492 |
| _cons | 5.085383 | .0452774 | 112.32 | 0.000 | 4.996551 | 5.174215 |

# betweenの結果

```
. xtreg SS OPS95, be

Between regression (regression on group means)  Number of obs      =       1200
Group variable: ID                              Number of groups   =        400

R-sq:  within  = 0.0003                         Obs per group: min =          3
       between = 0.1528                                        avg =        3.0
       overall = 0.0994                                        max =          3

                                                F(1,398)           =      71.76
sd(u_i + avg(e_i.))=  1.242766                  Prob > F           =     0.0000
```

| SS | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| OPS95 | .6314035 | .074536 | 8.47 | 0.000 | .4848701 | .7779369 |
| _cons | 5.073788 | .0644909 | 78.67 | 0.000 | 4.947003 | 5.200573 |

# withinの結果

```
. xtreg SS OPS95, fe

Fixed-effects (within) regression               Number of obs      =      1200
Group variable: ID                              Number of groups   =       400

R-sq:  within  = 0.0003                         Obs per group: min =         3
       between = 0.1528                                        avg =       3.0
       overall = 0.0994                                        max =         3

                                                F(1,799)           =      0.24
corr(u_i, Xb)  = -0.4083                        Prob > F           =    0.6211
```

| SS | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| OPS95 | -.0644307 | .130285 | -0.49 | 0.621 | -.320172 | .1913106 |
| _cons | 5.23492 | .0427196 | 122.54 | 0.000 | 5.151064 | 5.318776 |
| sigma_u | 1.3703832 | | | | | |
| sigma_e | 1.0477133 | | | | | |
| rho | .63110515 | (fraction of variance due to u_i) | | | | |

```
F test that all u_i=0:     F(399, 799) =     4.28                 Prob > F = 0.0000
```

# pooled, between, within

## SPSSでのやりかた

* pooled.
reg /dep=SS /ent=OPS95.

* between.
if (WAVE=1) MM_SS=M_SS.
if (WAVE=1) MM_OPS95=M_OPS95.
reg /dep=MM_SS /ent=MM_OPS95.

* within.
reg /origin /dep=D_SS /ent=D_OPS95.
この結果からさらに、標準誤差に

$$\sqrt{N-k/N-k-n}$$

を掛け算し、自由度調整する（Nは全ケース数、nはサンプルの人数、kは独立変数の数）

## STATAでのやりかた

* pooled
reg SS OPS95

* between
xtreg SS OPS95, be

* within
xtreg SS OPS95, fe

# 時間との交互作用

- 固定効果モデルでは、時間不変の変数は、用いることができない
- ただし、時間不変変数×時間依存変数の交互作用は入れることができる
- しばしば、時間依存変数として「時間」そのものをとりあげ、時間不変変数との交互作用をみることがある
- 交互作用があるならば、それを含めなければ、変数間の真の関連を見誤ることにつながる

# betweenとwithin

- 固定効果モデル（within）のまとめ
  - 固定効果モデルは、個人差の次元をつぶし、個人内変化のみに特化した、縦断的問いの検証のためのツール
  - 観察されない異質性、要は、モデルに含めていない個人レベルの変数による影響を、個体特有の固定効果が引き受けてくれて、丸ごと除去できる
  - だがそれはいささか効き過ぎ(?)で、観察した個人レベル変数さえ、モデルの中に入ることはできなくなっている
  - 個人内で変わりうる変数については、観察されていない重要なものがあれば、当然交絡要因となりえてしまい、推定結果に偏りをもたらすかも…
    - 時間不変の変数でも、時間依存変数との交互作用という形で交絡要因となってくるかもしれない

Section 6.

# パネルデータの回帰分析２
## ～FIXEDとRANDOM

# fixedとrandom

- 固定効果モデルの難点： 時間不変の変数の効果をまったく考慮できないところ
  - 例） 性別、学歴、出身階層　etc・・・
- 少し強めの仮定をおいたモデルなら、時間不変変数も含めることができるようになる

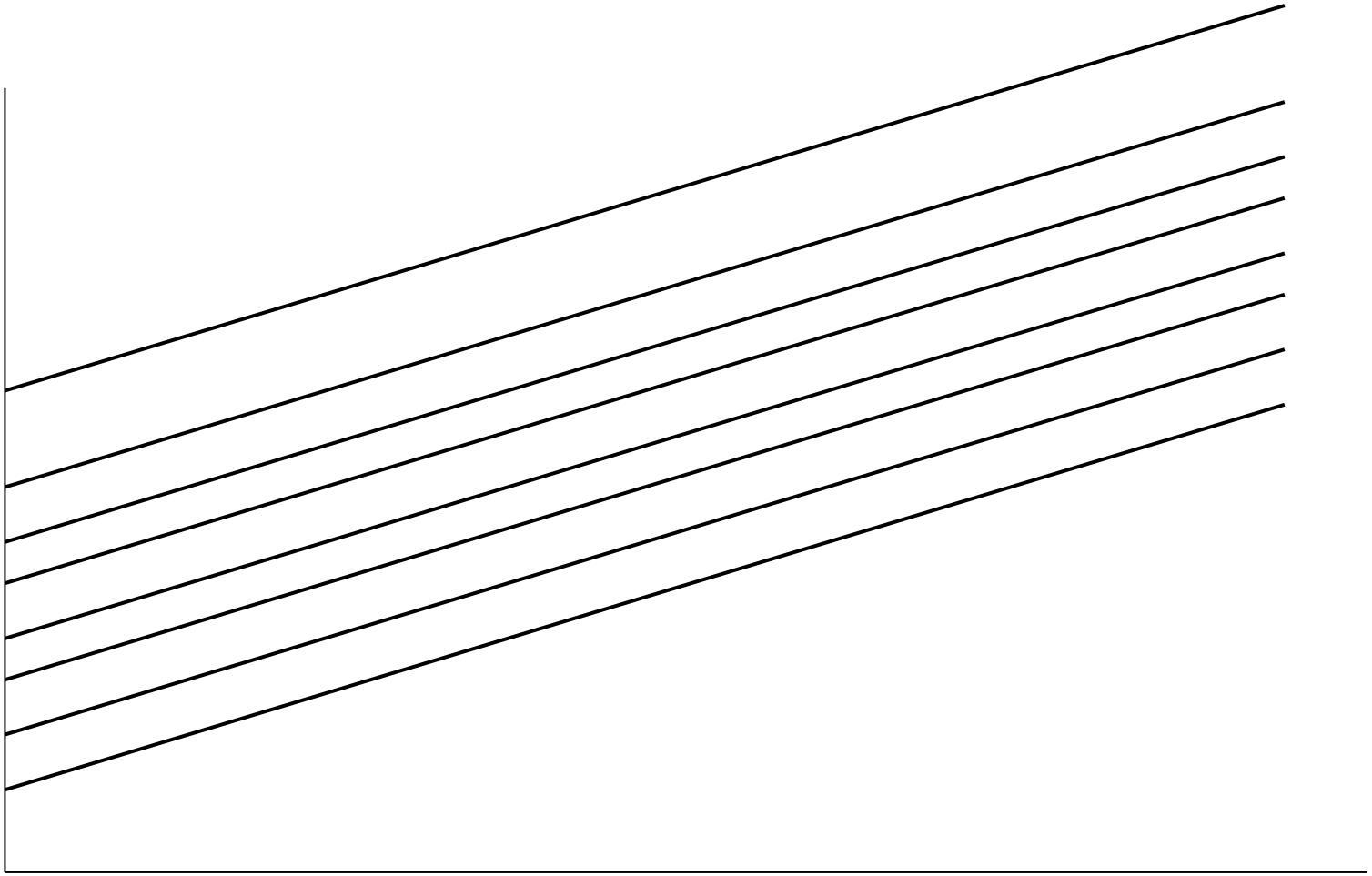
  ⇒ランダム効果モデル

$$Y\downarrow it = \alpha + \beta X\downarrow it + \gamma Z\downarrow i + u\downarrow i + \varepsilon\downarrow it$$

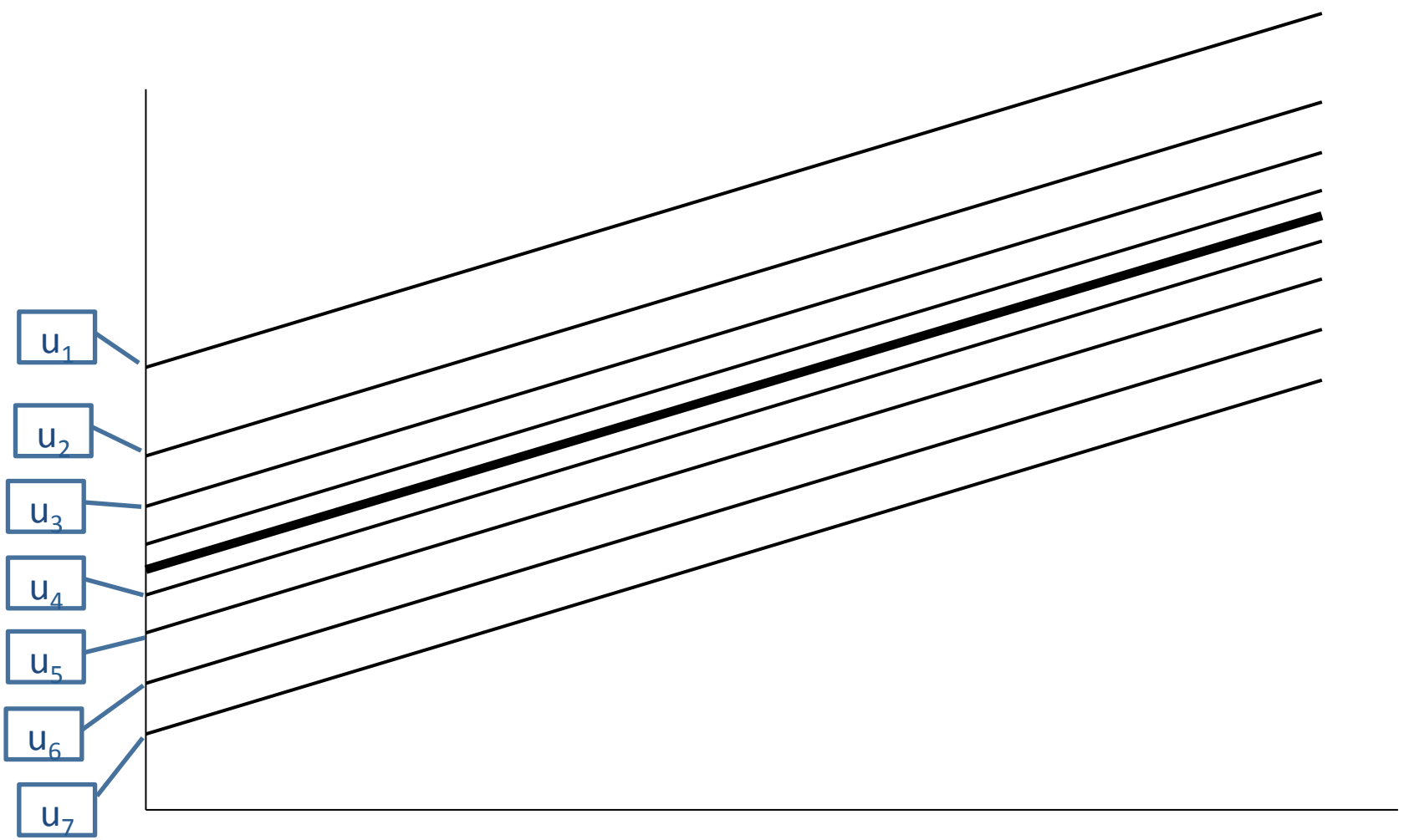
仮定：個人効果$u_i$は独立に平均0、分散$\tau^2$で同一の分布

個人効果$u_i$はXやZと無相関

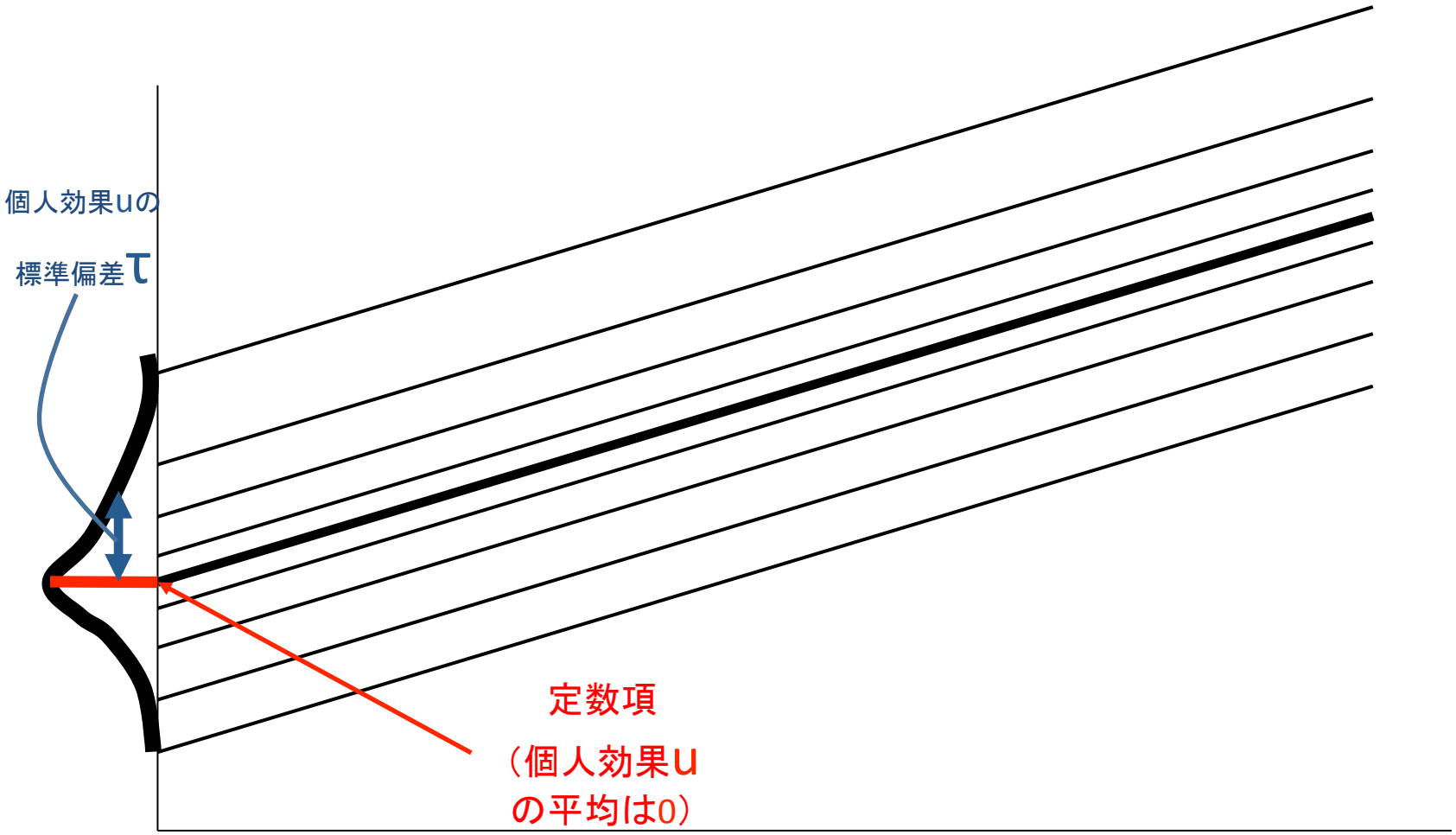
# fixedとrandom



# fixed



# random



# 一般化最小二乗法（GLS）の要点

ランダム効果モデルでは同一個人内の誤差項$u_i$と$\varepsilon_{it}$が相関するため、GLSで推定。ここでGLSの係数は右の通り。

$$\hat{\beta}_{GLS} = \frac{\sum_{i=1}^{n}[w_{xyi} + \psi_i b_{xyi}]}{\sum_{i=1}^{n}[w_{xxi} + \psi_i b_{xxi}]}$$

このとき、 $\psi_i = \sigma^2 \Big/ (\sigma^2 + T_i\tau^2)$

$$w_{xxi} = \sum_{t=1}^{T_i}(x_{it} - \bar{x}_i)^2, \quad b_{xxi} = T_i(\bar{x}_i - \bar{x})^2 \quad \text{etc.}$$

$\sigma^2$=0ならGLSはwithin推定と同じ。またTが多くなってもwithin推定に近づく。逆に、$\tau^2$=0ならばGLSはpooling推定（OLS）と同じ。

# fixedとrandom

- 固定効果モデル
  - 利点

    自由度をn-1個使って個体効果$u_i$を推定し、観察されない異質性が統制できる
  - 欠点

    個体効果$u_i$と他の変数との相関を許容、それゆえに時間不変変数が共線性のため除外される
- ランダム効果モデル
  - 利点

    $u_i$の分散の1個だけ自由度を使って個体効果をあらわし、時間不変変数もモデルに入れられる
  - 欠点

    個体効果uiが（仮定に反して）xやzと相関をもっていると、回帰係数に偏り

# fixed effects modelの結果

```
. xtreg SS OPS95 EDUY, fe
note: EDUY omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =      1200
Group variable: ID                              Number of groups   =       400

R-sq:  within  = 0.0003                         Obs per group: min =         3
       between = 0.1528                                        avg =       3.0
       overall = 0.0994                                        max =         3

                                                F(1,799)           =      0.24
corr(u_i, Xb)  = -0.4083                        Prob > F           =    0.6211
```

| SS | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| OPS95 | -.0644307 | .130285 | -0.49 | 0.621 | -.320172 | .1913106 |
| EDUY | 0 | (omitted) | | | | |
| _cons | 5.23492 | .0427196 | 122.54 | 0.000 | 5.151064 | 5.318776 |
| sigma_u | 1.3703832 | | | | | |
| sigma_e | 1.0477133 | | | | | |
| rho | .63110515 | (fraction of variance due to u_i) | | | | |

```
F test that all u_i=0:     F(399, 799) =     4.28                 Prob > F = 0.0000
```

# random effects modelの結果

```
. xtreg SS OPS95 EDUY, re

Random-effects GLS regression                   Number of obs      =      1200
Group variable: ID                              Number of groups   =       400

R-sq:  within  = 0.0003                         Obs per group: min =         3
       between = 0.1757                                        avg =       3.0
       overall = 0.1210                                        max =         3

                                                Wald chi2(2)       =     73.54
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

| SS | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| OPS95 | .3490351 | .0688903 | 5.07 | 0.000 | .2140127 | .4840576 |
| EDUY | .1634967 | .0353237 | 4.63 | 0.000 | .0942636 | .2327298 |
| _cons | 4.806868 | .0914546 | 52.56 | 0.000 | 4.62762 | 4.986116 |
| sigma_u | 1.0630202 | | | | | |
| sigma_e | 1.0477133 | | | | | |
| rho | .50725155 | (fraction of variance due to u_i) | | | | |

# fixedとrandom

## SPSSでのやりかた

* fixed effects model (within).

reg /origin /dep=D_SS /ent=D_OPS95 EDUY.

この結果からさらに、標準誤差に

$\sqrt{N-k/N-k-n}$

を掛け算し、自由度調整する（Nは全ケース数、nはサンプルの人数、kは独立変数の数）

* random effects model.

mixed SS with OPS95 EDUY

/fixed= OPS95 EDUY /method= ml

/print= solution testcov

/random intercept | subject(ID) covtype(VC).

最尤法によるランダム効果モデル

## STATAでのやりかた

* fixed effects model (within)

xtreg SS OPS95 EDUY, fe

* random effects model

xtreg SS OPS95 EDUY, re

最後のreをmleに変えると、最尤法によるランダム効果モデルに

# Hausman検定

- STATAでは、推定結果を内部的に保存し、利用できる

  estimates store *name*　下線部を好きな名に変える

  ereturn list　　保存されている結果の確認

  hausman *name*　下線部の結果と、直前の結果とのあいだで、回帰係数に系統的差異があるかを検定（ハウスマン検定）

```
. hausman feresult
```

| | Coefficients | | | |
|---|---|---|---|---|
| | (b) | (B) | (b-B) | sqrt(diag(V_b-V_B)) |
| | feresult | . | Difference | S.E. |
| OPS95 | -.0644307 | .3490351 | -.4134658 | .1105817 |

```
                b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

  Test:  Ho:  difference in coefficients not systematic

                chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                        =       13.98
              Prob>chi2 =      0.0002
```

# hybrid model

- ランダム効果モデルの枠内で、固定効果モデルと同等の回帰係数を求めることができる

  ⇒　hybrid model

- hybrid modelの走らせかたは、random effects modelと同じ
  - ただし、個人のレベルにおいて、時間依存変数の個人内平均値を投入する点のみ異なる
  - 時間依存変数を、そのまま投入するのと、個人内平均値からの偏差にしてから投入するやりかたがある

# hybrid modelの結果

```
. xtreg SS OPS95 EDUY M_OPS95, re

Random-effects GLS regression                   Number of obs      =      1200
Group variable: ID                              Number of groups   =       400

R-sq:  within  = 0.0003                         Obs per group: min =         3
       between = 0.1815                                        avg =       3.0
       overall = 0.1294                                        max =         3

                                                Wald chi2(3)       =     88.25
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

| SS | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| OPS95 | -.0644307 | .130285 | -0.49 | 0.621 | -.3197846 | .1909232 |
| EDUY | .1342905 | .0359959 | 3.73 | 0.000 | .0637399 | .2048411 |
| M_OPS95 | .5715647 | .1531818 | 3.73 | 0.000 | .2713338 | .8717956 |
| _cons | 4.829619 | .091169 | 52.97 | 0.000 | 4.650931 | 5.008307 |
| sigma_u | 1.0630202 | | | | | |
| sigma_e | 1.0477133 | | | | | |
| rho | .50725155 | (fraction of variance due to u_i) | | | | |

# hybrid modelのもう1つの結果

```
. xtreg SS D_OPS95 EDUY M_OPS95, re

Random-effects GLS regression                   Number of obs      =      1200
Group variable: ID                              Number of groups   =       400

R-sq:  within  = 0.0003                         Obs per group: min =         3
       between = 0.1815                                        avg =       3.0
       overall = 0.1294                                        max =         3

                                                Wald chi2(3)       =     88.25
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

| SS | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| D_OPS95 | -.0644307 | .130285 | -0.49 | 0.621 | -.3197846 | .1909232 |
| EDUY | .1342905 | .0359959 | 3.73 | 0.000 | .0637399 | .2048411 |
| M_OPS95 | .507134 | .0805636 | 6.29 | 0.000 | .3492322 | .6650359 |
| _cons | 4.829619 | .091169 | 52.97 | 0.000 | 4.650931 | 5.008307 |
| sigma_u | 1.0630202 | | | | | |
| sigma_e | 1.0477133 | | | | | |
| rho | .50725155 | (fraction of variance due to u_i) | | | | |

# hybrid model

- 時間依存変数の投入の仕方による違い
  - センタリングしなければ、平均値変数の回帰係数は、「between回帰係数とwithin回帰係数の差」に相当
  - センタリングすると、それは「between回帰係数」に
- hybrid modelの利点

  (1) 時間依存変数の回帰係数に関しては、固定効果モデルと同じ結果が得られる

  (2) 時間不変変数をモデルに含めることができる

  (3) withinとbetweenの係数の比較・検定ができる

  (4) ランダムスロープへの拡張ができる

  (5) 誤差の構造に関する制約が少ない

Section 7.

# 誤差の分散共分散

# 誤差の分散共分散

- パネルデータの回帰分析において、誤差の分散共分散構造に「手を入れる」ことがある
  - 何もしなければ、誤差は、等分散で、独立に同一の正規分布にしたがうとしている
- なぜパネルデータでは、誤差の分散共分散を考慮することが重要なのか？
  (1) 同じ従属変数を何度も観察するうちに、時点間で従属変数の分散が異なっている危険性
  (2) パネルデータは時系列の要素を含むゆえ、個人内において時点間で誤差に系列相関が生じる危険性

# 誤差の分散共分散構造

- パネルデータ分析においては、時点間での誤差の分散共分散について、仮定を変えることができる

（1）Identity

SPSS: id

STATA: ind

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

（2）Diagonal

SPSS: diag

STATA: ind, by(WAVE)

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

（3）Exchangeable

SPSS: cs

STATA: ex

$$\sigma^2\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

（4）AR(1)

SPSS: ar1

STATA: ar 1, t(WAVE)

$$\sigma^2\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

（5）Toeplitz

SPSS: tp

STATA: to, t(WAVE)

$$\sigma^2\begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

（6）Unstructured

SPSS: un

STATA: un, t(WAVE)

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

# 誤差の分散共分散構造

Identity？　　or　　Diagonal？

モデル次元[a]

| | | レベル数 | 共分散構造 | パラメータ数 | 被験者変数 | 被験者数 |
|---|---|---|---|---|---|---|
| 固定効果 | 切片 | 1 | | 1 | | |
| | OPS95 | 1 | | 1 | | |
| | EDUY | 1 | | 1 | | |
| 変量効果 | 切片[b] | 1 | 分散成分 | 1 | ID | |
| 反復効果 | WAVE | 3 | 一致 | 1 | ID | 400 |
| 合計 | | 7 | | 5 | | |

a. 従属変数: SS。

b. バージョン 11.5 では、RANDOM サブコマンドのシンタックスの規則が変更されています。同じコマンドのシンタックスを指定しても、前のバージョンとは違う結果が得られることもあります。バージョン 11 のシンタックスを使用している場合、詳細は現在のバージョンのシンタックスのマニュアルを参照してください。

情報量基準[a]

| | |
|---|---|
| -2 対数尤度 | 4090.438 |
| 赤池情報基準 (AIC) | 4100.438 |
| Hurvich and Tsai 基準 (AICC) | 4100.489 |
| Bozdogan 基準 (CAIC) | 4130.889 |
| Schwarz's Bayesian 基準 (BIC) | 4125.889 |

情報量基準は、smaller-is-better 形式で表示されます。

a. 従属変数: SS。

モデル次元[a]

| | | レベル数 | 共分散構造 | パラメータ数 | 被験者変数 | 被験者数 |
|---|---|---|---|---|---|---|
| 固定効果 | 切片 | 1 | | 1 | | |
| | OPS95 | 1 | | 1 | | |
| | EDUY | 1 | | 1 | | |
| 変量効果 | 切片[b] | 1 | 分散成分 | 1 | ID | |
| 反復効果 | WAVE | 3 | 対角 | 3 | ID | 400 |
| 合計 | | 7 | | 7 | | |

a. 従属変数: SS。

b. バージョン 11.5 では、RANDOM サブコマンドのシンタックスの規則が変更されています。同じコマンドのシンタックスを指定しても、前のバージョンとは違う結果が得られることもあります。バージョン 11 のシンタックスを使用している場合、詳細は現在のバージョンのシンタックスのマニュアルを参照してください。

情報量基準[a]

| | |
|---|---|
| -2 対数尤度 | 4084.927 |
| 赤池情報基準 (AIC) | 4098.927 |
| Hurvich and Tsai 基準 (AICC) | 4099.021 |
| Bozdogan 基準 (CAIC) | 4141.558 |
| Schwarz's Bayesian 基準 (BIC) | 4134.558 |

情報量基準は、smaller-is-better 形式で表示されます。

a. 従属変数: SS。

# 誤差の分散共分散構造

Identity？　　or　　Diagonal？

**共分散パラメータの推定[a]**

| パラメータ | | 推定値 | 標準誤差 | Wald の Z | 有意 | 95% 信頼区間 | |
|---|---|---|---|---|---|---|---|
| | | | | | | 下限 | 上限 |
| 反復測定 | 分散 | 1.110222 | .055706 | 19.930 | .000 | 1.006238 | 1.224952 |
| 切片 [被験者 = ID] | 分散 | 1.128828 | .108513 | 10.403 | .000 | .934982 | 1.362864 |

a. 従属変数: SS。

**残差共分散 (R) 行列[a]**

| | [WAVE = 1] | [WAVE = 2] | [WAVE = 3] |
|---|---|---|---|
| [WAVE = 1] | 1.110222 | 0 | 0 |
| [WAVE = 2] | 0 | 1.110222 | 0 |
| [WAVE = 3] | 0 | 0 | 1.110222 |

一致

a. 従属変数: SS。

**共分散パラメータの推定[a]**

| パラメータ | | 推定値 | 標準誤差 | Wald の Z | 有意 | 95% 信頼区間 | |
|---|---|---|---|---|---|---|---|
| | | | | | | 下限 | 上限 |
| 反復測定 | Var: [WAVE=1] | 1.229346 | .114113 | 10.773 | .000 | 1.024855 | 1.474640 |
| | Var: [WAVE=2] | 1.176040 | .110922 | 10.602 | .000 | .977549 | 1.414836 |
| | Var: [WAVE=3] | .902939 | .094656 | 9.539 | .000 | .735235 | 1.108896 |
| 切片 [被験者 = ID] | 分散 | 1.150277 | .110399 | 10.419 | .000 | .953031 | 1.388346 |

a. 従属変数: SS。

**残差共分散 (R) 行列[a]**

| | [WAVE = 1] | [WAVE = 2] | [WAVE = 3] |
|---|---|---|---|
| [WAVE = 1] | 1.229346 | 0 | 0 |
| [WAVE = 2] | 0 | 1.176040 | 0 |
| [WAVE = 3] | 0 | 0 | .902939 |

対角

a. 従属変数: SS。

# 誤差の分散共分散構造

## Identity？　　or　　Exchangeable？

**モデル次元[a]**

| | | レベル数 | 共分散構造 | パラメータ数 | 被験者変数 | 被験者数 |
|---|---|---|---|---|---|---|
| 固定効果 | 切片 | 1 | | 1 | | |
| | OPS95 | 1 | | 1 | | |
| | EDUY | 1 | | 1 | | |
| 変量効果 | 切片[b] | 1 | 分散成分 | 1 | ID | |
| 反復効果 | WAVE | 3 | 一致 | 1 | ID | 400 |
| 合計 | | 7 | | 5 | | |

a. 従属変数: SS。

b. バージョン 11.5 では、RANDOM サブコマンドのシンタックスの規則が変更されています。同じコマンドのシンタックスを指定しても、前のバージョンとは違う結果が得られることもあります。バージョン 11 のシンタックスを使用している場合、詳細は現在のバージョンのシンタックスのマニュアルを参照してください。

**情報量基準[a]**

| | |
|---|---|
| -2 対数尤度 | 4090.438 |
| 赤池情報基準 (AIC) | 4100.438 |
| Hurvich and Tsai 基準 (AICC) | 4100.489 |
| Bozdogan 基準 (CAIC) | 4130.889 |
| Schwarz's Bayesian 基準 (BIC) | 4125.889 |

情報量基準は、smaller-is-better 形式で表示されます。

a. 従属変数: SS。

**モデル次元[a]**

| | | レベル数 | 共分散構造 | パラメータ数 | 被験者変数 | 被験者数 |
|---|---|---|---|---|---|---|
| 固定効果 | 切片 | 1 | | 1 | | |
| | OPS95 | 1 | | 1 | | |
| | EDUY | 1 | | 1 | | |
| 変量効果 | 切片[b] | 1 | 分散成分 | 1 | ID | |
| 反復効果 | WAVE | 3 | 複合シンメトリ | 2 | ID | 400 |
| 合計 | | 7 | | 6 | | |

a. 従属変数: SS。

b. バージョン 11.5 では、RANDOM サブコマンドのシンタックスの規則が変更されています。同じコマンドのシンタックスを指定しても、前のバージョンとは違う結果が得られることもあります。バージョン 11 のシンタックスを使用している場合、詳細は現在のバージョンのシンタックスのマニュアルを参照してください。

**情報量基準[a]**

| | |
|---|---|
| -2 対数尤度 | 4090.438 |
| 赤池情報基準 (AIC) | 4102.438 |
| Hurvich and Tsai 基準 (AICC) | 4102.509 |
| Bozdogan 基準 (CAIC) | 4138.979 |
| Schwarz's Bayesian 基準 (BIC) | 4132.979 |

情報量基準は、smaller-is-better 形式で表示されます。

a. 従属変数: SS。

# 誤差の分散共分散構造

Identity？　　or　　Exchangeable？

**共分散パラメータの推定[a]**

| パラメータ | | 推定値 | 標準誤差 | Wald の Z | 有意 | 95% 信頼区間 | |
|---|---|---|---|---|---|---|---|
| | | | | | | 下限 | 上限 |
| 反復測定 | 分散 | 1.110222 | .055706 | 19.930 | .000 | 1.006238 | 1.224952 |
| 切片 [被験者 = ID] | 分散 | 1.128828 | .108513 | 10.403 | .000 | .934982 | 1.362864 |

a. 従属変数: SS。

**残差共分散 (R) 行列[a]**

| | [WAVE = 1] | [WAVE = 2] | [WAVE = 3] |
|---|---|---|---|
| [WAVE = 1] | 1.110222 | 0 | 0 |
| [WAVE = 2] | 0 | 1.110222 | 0 |
| [WAVE = 3] | 0 | 0 | 1.110222 |

一致

a. 従属変数: SS。

**共分散パラメータの推定[a]**

| パラメータ | | 推定値 | 標準誤差 | Wald の Z | 有意 | 95% 信頼区間 | |
|---|---|---|---|---|---|---|---|
| | | | | | | 下限 | 上限 |
| 反復測定 | CS 対角オフセット | 1.1102 | .0557 | 19.931 | .000 | 1.0062 | 1.2250 |
| | CS 共分散 | .0133 | .1084 | .123 | .902 | -.1991 | .2258 |
| 切片 [被験者 = ID] | 分散 | 1.1165[b] | .0000 | . | . | . | . |

a. 従属変数: SS。

b. この共分散パラメータは冗長です。検定統計量と信頼区間を計算できません。

**残差共分散 (R) 行列[a]**

| | [WAVE = 1] | [WAVE = 2] | [WAVE = 3] |
|---|---|---|---|
| [WAVE = 1] | 1.1236 | .0133 | .0133 |
| [WAVE = 2] | .0133 | 1.1236 | .0133 |
| [WAVE = 3] | .0133 | .0133 | 1.1236 |

複合シンメトリ

a. 従属変数: SS。

# 誤差の分散共分散構造

## SPSSでのやりかた

* 時点間での誤差の分散共分散構造の仮定を変える.

```
mixed SS with OPS95 EDUY
 /fixed= OPS95 EDUY
 /method= ML
 /print= r solution testcov
 /random=intercept | subject(ID) covtype(vc)
 /repeated= WAVE | subject(ID) covtype(id).
```

下線部のid（このidはidentityの意味）を、diag、cs、ar1、tp、unなどに変えることで、時点間の誤差分散共分散の仮定を変えることができる

## STATAでのやりかた

```
* gosa no bunsan-kyobunsan kouzou
xtmixed SS OPS95 EDUY, ///
 || ID:, cov(ind) residuals(ind, by(WAVE))
```

residualのかっこ内を変更する

by (WAVE)がつくと、diagnal

つかないと、identity

ex で exchangeable

ar 1, t(WAVE) で AR(1)

to, t(WAVE) で Toeplitz

un, t(WAVE) で unstructured へと、変えることができる

# 誤差の分散共分散

- 誤差の分散共分散は、より時系列データに近づいた時（Nが少なく、Tが多い）ほど注意が必要になる
- 誤差の分散共分散構造にどのような仮定をおくのが相応しいかは、尤度比の差の検定や、情報量基準の比較から、判断できる

Section 8.

# さいごに

# パネルデータ分析の強み（再掲）

（１）観察されない異質性の統制が可能

（２）個人レベルでの変化の分析が可能

（３）より精確な因果推論が可能

（４）情報量が豊富であることによる技術的強み

自由度が多い、多重共線性の緩和、推計上の利点

# さいごに

- アメリカ社会学や、日本でも経済学あるいは政治学などと比べて、日本の社会学はパネルデータ利用に関しては立ち遅れているように思える
- ぜひ皆さんが、ご自身の研究にパネルデータ分析をとりいれてくださることを願っています
  - JLPSデータは、SSJDA-Directシステムにより、ダウンロードできます！

# 文献


Allison, P. D. 2009. *Fixed Effects Regression Models*. Sage.

Bollen, K. A. & P. J. Curran. 2006. *Latent Curve Models*. Wiley.

Collins, L. M. & S. T. Lanza. 2009. *Latent Class and Latent Transition Analysis*. Wiley.

Heck, R. H., S. L. Thomas & L. N. Tabata. 2010. *Multilevel and Longitudinal Modeling with IBM SPSS*. Routledge.

Hsiao, C. 2003. *Analysis of Panel Data (2nd edition)*. Cambridge University Press.

北村行伸. 2005.『パネルデータ分析』岩波書店.

中澤渉. 2012.「なぜパネル・データを分析するのが必要なのか」『理論と方法』27(1): 23-40.

Rabe-Hesketh, S. & A. Skrondal. 2012. Multilevel and Longitudinal Modeling Using Stata (3rd edition*)*. Stata Press.

Singer, J. D. & J. B. Willett. 2003. Applied Longitudinal Data Analysis. Oxford.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge, J. M. 2009. *Introductory Econometrics(4th edition)*. South-Western.

山口一男. 2004.「パネルデータの長所とその分析方法」『季刊家計経済研究』62: 50-58.