

# SSM職歴データのパーソンイヤー データへの変換

**麦山 亮太 Ryota MUGIYAMA**

学習院大学法学部政治学科

ryota.mugiyama@gakushuin.ac.jp

資料： <https://github.com/ryotamugiyama/ssm2015personyear>

# パーソンイヤーデータの作成

# パーソンイヤーデータとは

**パーソンピリオドデータ**：同一個人（person）の複数時点（period）にわたる観察（observation）からなるデータ。時点の単位が1年の場合を、**パーソンイヤーデータ**という

パーソンピリオドデータは、個人の状態がいついかに変化したかを記録しやすい

id	year	age	work	student
1	2000	15	0	1
1	2001	16	0	1
1	2002	17	0	1
1	2003	18	1	0
1	2004	19	1	0
1	2005	20	1	0
1	2006	21	1	0
⋮				

# なぜパーソンピリオドデータが必要か

## 個人の状態の変化に関する問いに答えるため

- 雇用形態によって結婚への移行の起こりやすさはどの程度異なるか？
- 子どもを持つことによって就業率はどの程度変化するか？
- ..... etc

パーソンピリオドデータは、同一個人につき複数時点の観察を記録するため、  
時間による個人の状態の変化を分析しやすいデータ形式

縦断的な分析を行う場合には、（ほとんどの場合）パーソンピリオドデータの形式となっている必要がある

# wide形式とlong形式

wide形式（1つの行が1つの個体）

id	income1	income2	income3	income4
1	112.5	200	200	300
2	50	.	.	.
3	700	525	525	.

long形式（1つの行が1つの観察）

id	year	income
1	2007	112.5
1	2008	200
1	2009	200
1	2010	300
2	2007	50
3	2007	700
3	2008	525
3	2009	525

# SSMデータから作成できる時変の変数

今回作成するのは以下：

- 職業経歴（以下職歴とよぶ）
- 教育歴（部分的）
- 婚姻歴（部分的）、子ども歴（SSM2015は第4子まで）

以下の項目を使って時変の変数を作成することも可能だが、今回は作成しない

- 配偶者と知り合ったときの年齢
- 父母との死別
- はじめての離家
- はじめての帰家
- 資格取得年齢（SSM2005）

# 職歴の収集方法：ライフヒストリー・カレンダー（SSM2005～）\*

## あなたの職業経歴年表

支局番号		地点番号			対象番号	

この用紙は、回答者の方が経歴を思い出す手助けをするためのものです。清書の必要はありません。

I. 準備

(1) 「西暦」「和暦」欄に、回答者の誕生年から今年までの数字を、回答者の前で順に記入する（1989 年が平成元年）。

(2) 今年の左下か右下のマスが、回答者の現在の年齢になっていることを確認する。その年の誕生日が来る前であれば左下、誕生日が来た後であれば右下の年齢になる。

II. 年表づくりのための質問と年表の記入のしかた

**(質問A)まず、あなたが学校を出て初めて職業についたのは、いつでしたか。または何歳でしたか。**

⇒該当する年齢の「従業先など」の欄に「①」と記入する。暦年で回答された場合、年表で対応する年齢欄を確かめる。

**(質問B)その後、勤め先を変わられましたか。[変わった場合]その勤め先が変わったのはいつ(何歳の時)ですか。無職となった時期も含めて勤め先の変化を教えてください。[この質問を現在に至るまで繰り返す]**

⇒該当する年齢の「従業先など」の欄に何番目の従業先かを番号（②,③…）で記入する（従業先番号）。勤めていた期間が年齢をまたぐ場合は、「③→」のように、その勤め先を離れた（現在の勤め先の場合は現在の）年齢の欄まで矢印で結ぶ。

⇒無職期間は、該当する年齢の「従業先など」の欄に「無職」と記入し、波線で結ぶ。

III. 年表の最終確認

**最初の職業についてから現在まで、「従業先番号」「無職」の記入をたどり、切れ目がないことを確認する。**

**切れ目があるときは、空白期間の職業の有無を回答者に確認する。**→右下の記入例のような年表ができれば完成です。

IV. 注意事項

- 1 年に満たないものも無職期間とし（3 か月以上が目安）、該当する年齢の「従業先など」の欄に「無職」と記入する。
- 同じ年齢のときに有職期間と無職期間が混ざっている場合は、同じ欄に位置をずらして両方を記入する。
- 同じ組織の中で異動があっても、従業先は同じとみなす。
- 前と同じ従業先に戻った場合は、従業先番号も前の従業先の番号を使う（従業先 1 に戻った例：①→②→①→③）。
- 回答者が従業先の名前を答えてくれた場合には、その名前を余白に記入してもよい。
- 同じ年齢で複数の従業先に移った場合も、すべて記入する。
- 短期のパートやアルバイトも含む。ただし、学生時代のアルバイトは除く。
- 同時に複数の職業を持っていた場合は、主な職業についてのみ尋ねる。
- 必要に応じて卒業や結婚などの時期もメモし、職業の時期と正しく対応していることを確かめる。

西暦

和暦

年齢

従業先など

19

.....

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

参考：  
標準的な在学期間

※専門学校は短大と同じが多いが様々

中学 3 年間

高校 3 年間

高専 5 年間

短大 2 年間

大学 4 年間

下段に続く

【卒業後すぐの就職は、卒業と同じ年になる。翌年と間違えないように注意】

西暦

和暦

年齢

従業先など

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

記入例：1977 年生まれの現在 37 歳（まだ今年の誕生日がきていない）。大卒後すぐの 22 歳から商社に勤めるが、2 年で辞める。しばらくコンビニで働くも半年で辞め、2 年間無職。26 歳で出版社に再就職したが、32 歳で出産を機に退職。しばらく無職だったが、今年、別の出版社で働き始めた。

西暦

和暦

年齢

従業先など

77

.....

87

88

89

90

91

92

93

94

95

96

97

98

99

2000

01

02

03

04

05

06

07

08

09

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

参考：  
標準的な在学期間

※専門学校は短大と同じが多いが様々

中学 3 年間

高校 3 年間

高専 5 年間

短大 2 年間

大学 4 年間

商社

コンビニ

無職

出版社

無職

別の出版社

結婚

出産

# 職歴の収集方法：職歴頁

問 9 (a) 【回答票 8】その後、その同じ従業先の中で、「従業上の地位」「仕事の内容」あるいは「役職」が変わったことがありますか。【回答票を示し、「従業上の地位」と「役職」の意味を回答者に確認させる】

職歴頁番号  
02

1 はい

2 いいえ

【直前の職歴頁が「無職（学生含む）」の場合、問 9 (c) から聞く】

変わった内容を早いものから順に教えてください（変わらなかった欄には斜線）

問 9 (b) では、その従業先を終えられた後、3 か月以内に新しい従業先で仕事につきましたか。それとも仕事につかれていない期間が続きましたか。

1 3 か月以内に新しい従業先で仕事についた

2 無職になった（学生を含む）

①離職理由 前の従業先をやめた理由を 1 つ選んでください。【無職になった場合も聞く】

1 (ア) 定年、契約期間の終了など

2 (イ) 倒産、廃業、人員整理など

3 (ウ) よい仕事が見つかったから

4 (エ) 家庭の理由（結婚、育児など）

5 (オ) 家業を継ぐため

6 (カ) 職場に対する不満

7 (キ) 健康上の理由（病気やケガなど）

8 (ク) 年齢のため

9 (ケ) その他

99 わからない

問 9 (c) 次の従業先についてうかがいます。

【年表を参照して何個目の従業先かを記入する→】

従業先番号

②事業 新しい勤め先は、どのような事業をいとなんでいましたか。

【具体的に記入すること】【派遣社員の場合、派遣元の事業】

98 非該当（内職）

99 わからない

③従業員数 従業員（働いている人）は、会社全体で何人ぐらいでしたか。

【家族従業者、パート・アルバイトも含む】【派遣社員の場合、派遣元の企業規模】

1 (ア) 1 人

2 (イ) 2～4 人

3 (ウ) 5～9 人

4 (エ) 10～29 人

5 (オ) 30～99 人

6 (カ) 100～299 人

7 (キ) 300～499 人

8 (ク) 500～999 人

9 (ケ) 1000 人以上

10 (コ) 官公庁

98 非該当（内職）

99 わからない

そこで働き始めたときの、あなたのお仕事について教えてください

④従業上地位 そのお仕事は大きく分けてこの中のどれにあたりましたか。

1 (ア) 経営者、役員

2 (イ) 常時雇用されている一般従業者

3 (ウ) パート・アルバイト

4 (エ) 派遣社員

5 (オ) 契約社員、嘱託

6 (カ) 臨時雇用

7 (キ) 自営業主、自由業者

8 (ク) 家族従業者

9 (ケ) 内職

99 わからない

⑤仕事内容 職場ではどのような仕事をしていましたか。

【仕事の中身がわかるように具体的に記入】

999 わからない

⑥役職名 何かの役職についていましたか。

1 (ア) 役職なし

2 (イ) 監督、職長、班長、組長

3 (ウ) 係長、係長相当職

4 (エ) 課長、課長相当職

5 (オ) 部長、部長相当職

6 (カ) 社長、重役、役員、理事

7 (キ) その他

9 わからない

⑦年齢 そのお仕事には何歳の頃つかれましたか。

歳 ～ 歳

99 わからない

⑧現職確認 【この頁が現職か確認。現職ならば、問 2 と従業先、「a 従業上の地位」、「f 仕事内容」、「g 役職」との一致も確認】

1 現職にあたる→ 9 頁の問 10 へ

2 職歴が続く→ 次頁の質問を続ける

⑦年齢 無職だった期間を教えてください。

歳 ～ 歳

99 わからない

⑧現職確認 【この頁の無職が現職か確認】

1 現職にあたる→ 9 頁の問 10 へ

2 職歴が続く→ 次頁の問 9 (c) へ

従業先が変わるたびに（無職になったり、無職から有職になった場合も含む）、または同じ従業先の中で「従業上の地位」「仕事の内容」「役職」のいずれかが変わるたびに、新たな職歴頁にその内容を記入  
これを現職に至るまで繰り返し聴取

8



# 教育歴の収集方法

問18 【回答票16】あなたがこれまで通ったことのある学校等をすべてあげてください。(M.A.)

〔通った学校すべてに○をつける。在学中や中退も含む〕

- 1 (ア) 幼稚園
- 2 (イ) 保育所 (保育園)
- 3 (ウ) 小学校
- 4 (エ) 中学校
- 5 (オ) 高校
- 6 (カ) 専修学校高等課程
- 7 (キ) 専門学校 (専修学校専門課程)
- 8 (ク) 短大 (短期大学)
- 9 (ケ) 高専 (高等専門学校)
- 10 (コ) 大学
- 11 (サ) 大学院
- 12 (シ) その他 (具体的に)
- [
- ]
- 99 わからない →12 頁問 21 へ
- ③①

問19 「高校に通ったことのある人 (問18で「5 (オ) 高校」に○)」にたずねる。「高校に通ったことのない人 (問18で「5 (オ) 高校」に○がない)」は11頁問20に進む

【回答票17】あなたが通った高校についてお答えください。複数の高校に通われたことのある方は a～d について、最後の高校についてお答えください。

a その学校は次のどれにあたりますか。〔公立高校には、都道府県立高校以外に市立高校や組合立高校を含む〕	1 (ア) 国立	2 (イ) 公立	3 (ウ) 私立	9 わからない						
b 学科はどれにあたりますか。	1 (ア) 普通科・理数科・英数科・外国語科・国際科	2 (イ) 工業に関する学科	3 (ウ) 商業に関する学科	4 (エ) 農業・水産に関する学科	5 (オ) 看護・福祉に関する学科	6 (カ) 家庭・家政・保育に関する学科	7 (キ) 芸術・体育に関する学科	8 (ク) 総合学科	9 (ケ) その他 (具体的に)	99 わからない
c その高校では大学・短大進学者の割合はどのくらいでしたか。	1 (ア) ほぼ全員	2 (イ) 7～8 割	3 (ウ) 半数くらい	4 (エ) 2～3 割	5 (オ) ほとんどいない	9 わからない				
d その高校を卒業しましたか。	1 (ア) 卒業した	2 (イ) 中退した	3 (ウ) 在学中	9 わからない						

問20 あなたが通った専門学校・短大・高専・大学・大学院についてお聞きします。

(1) 【回答票18】それらの学校をすべて通った順に教えてください。

〔A 大学の後に B 大学に通ったなど、同じ学校種に複数通った経験がある場合は、別々に記録する。通った学校が 4 つ以上ある場合は、欄外に記入する〕

	1 番目の学校	2 番目の学校	3 番目の学校
〔該当する学校に○をつける。問 18 の回答と対応しているか確認する〕	7 (ア) 専門学校 8 (イ) 短大 9 (ウ) 高専 10 (エ) 大学 11 (オ) 大学院	7 (ア) 専門学校 8 (イ) 短大 9 (ウ) 高専 10 (エ) 大学 11 (オ) 大学院	7 (ア) 専門学校 8 (イ) 短大 9 (ウ) 高専 10 (エ) 大学 11 (オ) 大学院

③⑦

④⑥

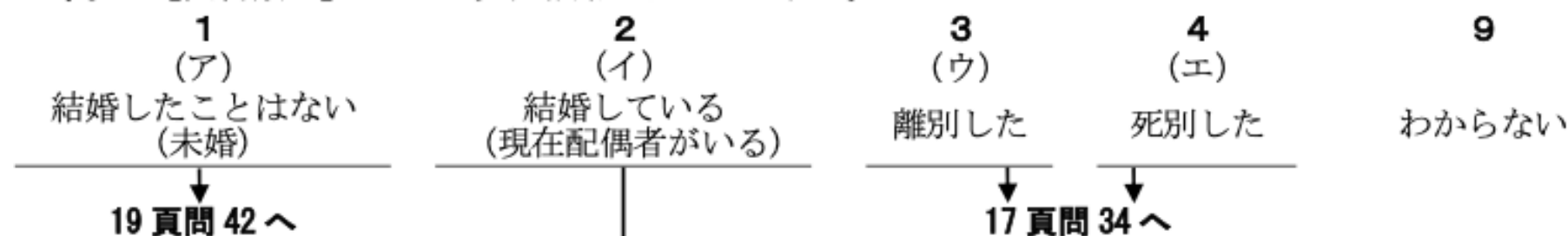
⑤⑤

(2) 【回答票19】それぞれの学校についてお聞きします。〔各学校について、a～d2を順に聞いていく〕

	1 番目の学校	2 番目の学校	3 番目の学校
a あなたは、その学校を卒業 (修了) しましたか、中退しましたか、それとも在学中ですか。	1 (ア) 卒業・修了 2 (イ) 中退 3 (ウ) 在学中 9 わからない	1 (ア) 卒業・修了 2 (イ) 中退 3 (ウ) 在学中 9 わからない	1 (ア) 卒業・修了 2 (イ) 中退 3 (ウ) 在学中 9 わからない
b1 その学校に入学したのは何歳のときですか。〔標準では高卒後すぐは 18 歳〕	<div>□□ 歳</div> <div>99 わからない</div>	<div>□□ 歳</div> <div>99 わからない</div>	<div>□□ 歳</div> <div>99 わからない</div>
b2 その学校には何年間、在籍していましたか。〔端数切り上げ (1 年半→2 年、半年→1 年)、休学期間も含む〕	<div>□□ 年間</div> <div>99 わからない</div>	<div>□□ 年間</div> <div>99 わからない</div>	<div>□□ 年間</div> <div>99 わからない</div>
c その学校は、国立ですか、公立ですか、私 (わたくし) 立ですか。	1 (ア) 国立 2 (イ) 公立 3 (ウ) 私立 9 わからない	1 (ア) 国立 2 (イ) 公立 3 (ウ) 私立 9 わからない	1 (ア) 国立 2 (イ) 公立 3 (ウ) 私立 9 わからない
d1	学部・学科など 〔大学・短大・高専〕学部・学科をお教えてください。 〔大学院〕研究科名をお教えてください。 〔専門学校〕学科または分野をお教えてください。 9 わからない	学部・学科など 〔大学・短大・高専〕学部・学科をお教えてください。 〔大学院〕研究科名をお教えてください。 〔専門学校〕学科または分野をお教えてください。 9 わからない	学部・学科など 〔大学・短大・高専〕学部・学科をお教えてください。 〔大学院〕研究科名をお教えてください。 〔専門学校〕学科または分野をお教えてください。 9 わからない
d2	学校名 〔すべての学校について〕学校名をお教えてください。 9 わからない	学校名 〔すべての学校について〕学校名をお教えてください。 9 わからない	学校名 〔すべての学校について〕学校名をお教えてください。 9 わからない

# 婚姻歴の収集方法

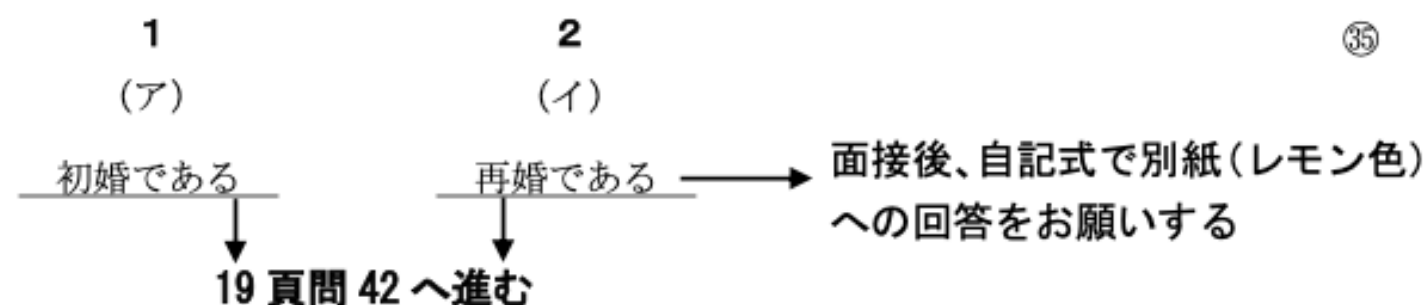
問25 【回答票26】あなたは現在結婚されていますか。



問26 現在の配偶者（奥さま、ご主人）と結婚されたのは、あなたが何歳のときですか。

満  歳      99    わからない

問33 【回答票31】現在の結婚について、あなたは次のどちらですか。



〔ここから問 41 までは、離死別の方（問 25 「3（ウ）離別した」「4（エ）死別した」と回答した方）に〕  
あなたが最初に結婚されたときのことをお聞きします。

問34 そのとき、あなたは何歳でしたか。

満  歳      99    わからない      ③⑥③⑦

問41 その方と離別（死別）されたのは、あなたが何歳のときですか。

満  歳      99    わからない

# 子ども歴の収集方法

問 12 これまでにお持ちになったお子さんは何人ですか。養子・継子（連れ子）・亡くなったお子さんを含めてお答えください。

人

0

子どもを持ったことはない

8 ページの間 15 へお進みください

⑪⑫

問 13 お子さんひとりひとりについて、以下の質問にお答えください。  
すでに成人されているお子さんや別居中のお子さんについてもお答えください。  
(お子さんが5人以上いる場合には、上から4番目のお子さんまでについてお答えください)

1 番年上のお子さん	2 番目のお子さん	3 番目のお子さん	4 番目のお子さん
(1) この方の性別をお答えください。(○は各 1 つだけ)			
1 男性    2 女性	1 男性    2 女性	1 男性    2 女性	1 男性    2 女性
(2) この方の出生年をお答えください。			
1 昭和    年	1 昭和    年	1 昭和    年	1 昭和    年
2 平成    年	2 平成    年	2 平成    年	2 平成    年
(3) この方とあなたとの続柄をお答えください。(○は各 1 つだけ)			
1 実子	1 実子	1 実子	1 実子
2 養子・継子(連れ子)	2 養子・継子(連れ子)	2 養子・継子(連れ子)	2 養子・継子(連れ子)

# SSM調査データからパーソニヤーデータを作る

## 1. 個人の複製

## 2. 時変の変数にまとめる

id	year	age	jobst1	status1	jobst2	status2	status	work
1	1985	15	18	2	26	7	.	0
1	1986	16	18	2	26	7	.	0
1	1987	17	18	2	26	7	.	0
1	1988	18	18	2	26	7	2	1
1	1989	19	18	2	26	7	2	1
⋮								
1	2012	42	18	2	26	7	7	1
1	2013	43	18	2	26	7	7	1
1	2014	44	18	2	26	7	7	1

# doファイルの整理

\_master\_ssm2015personyear.do 全体を統合するdo-file

1\_expand\_data.do

個人の複製

2\_1\_job\_history.do

職歴から時変の変数を作成

2\_2\_school\_history.do

教育歴から時変の変数を作成

2\_3\_marriage\_history.do

婚姻歴から時変の変数を作成

2\_4\_child\_history.do

子ども歴から時変の変数を作成

時変の変数  
にまとめる

2\_9\_variablelists.do

作成した変数の一覧表を出力

3\_1\_scatter.do

レキシス図の出力

3\_2\_eventhistory.do

イベントヒストリー分析の例

3\_3\_fixedeffect.do

固定効果モデルの例

# 1. 個人の複製

「観察開始時点」から「観察終了時点」に至るまでの時点の数だけデータを複製

**1970年生まれ、2014年12月31日時点で満44歳の個人の場合**

id	birthyear	age2014	year	age
1	1970	44	1985	15
1	1970	44	1986	16
1	1970	44	1987	17
⋮				
1	1970	44	2012	42
1	1970	44	2013	43
1	1970	44	2014	44

30時点分の観察

# 観察開始時点をいつとするか

今回は観察開始時点を**15歳（出生年 + 15年）**としている

- ごくわずかだが15歳よりも前に初職に就いている人はいる
- ごくわずかだが15歳よりも前に子どもを生んだことになっている人（連れ子等かもしれない）もいる

問題関心によっては、**18歳（結婚が法的に可能となる年齢／成人年齢）**などを観察開始時点としてもよい

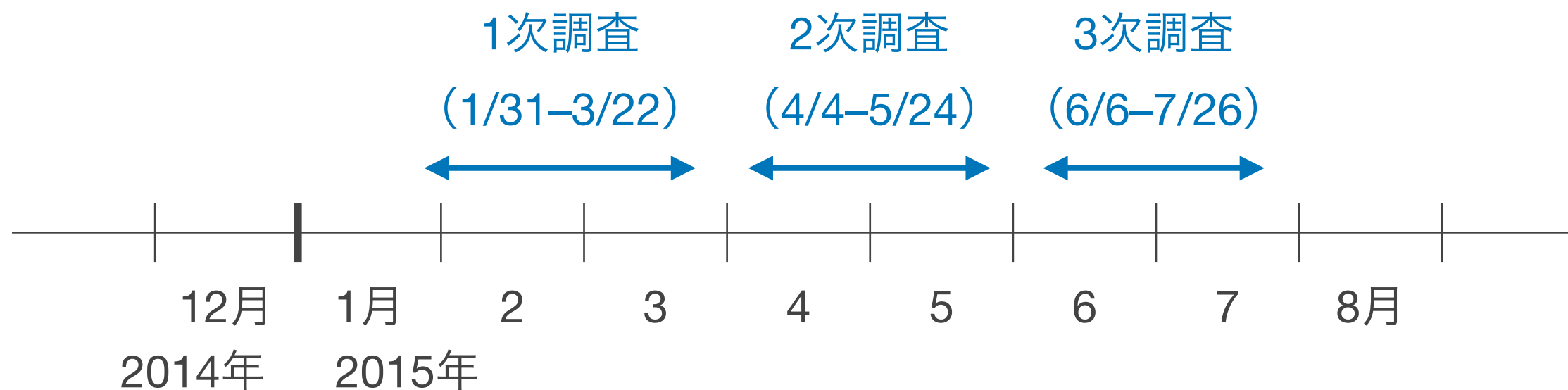
- 15歳からのパーソンイヤーデータを作成して、15～17歳の観察を削除するという手順でもよい



# 観察終了時点をいつとするか

今回は**2014年12月31日（の満年齢）**\*を観察終了時点としている

- SSM2015の調査対象：2014年12月31日現在満20歳～79歳の日本国籍を持つ男女 = 1935年1月1生まれ～1994年12月31生まれ
- 調査期間は2015年のため、調査時点の年齢 ≠ 2014年12月31日現在の満年齢
- 調査時期の関係から、2015年中の変化はすべて捕捉できない



\*正確には調査時点の年齢の状態が最後の観察となるので、たとえば1970年9月生まれの人は2015年5月時点ではまだ44歳であり、44歳のときに生じた変化はデータに反映される（保田時男先生より）



## 2. 時変の変数にまとめる

変化が起こった年齢を参照して、時変の変数を作成する

18歳で従業上の地位が2（常時雇用の一般従業者）、26歳で7（自営業）になった個人の場合

id	year	age	jobst1	status1	jobst2	status2	status	work
1	1985	15	18	2	26	7	.	0
1	1986	16	18	2	26	7	.	0
1	1987	17	18	2	26	7	.	0
1	1988	18	18	2	26	7	2	1
1	1989	19	18	2	26	7	2	1
⋮								
1	1995	25	18	2	26	7	2	1
1	1996	26	18	2	26	7	7	1
⋮								

# 基本的な加工の手順：職歴関連変数の場合

年齢（時変）の値と、変化が起こった年齢とを比較して、値を代入していく

1. 年齢（時変）の値が職歴頁1の開始年齢に満たないときには無職とみなし、従業上の地位、職業などは欠損とする
2. 年齢（時変）の値が職歴頁1の開始年齢（jobst1とする）に到達したときには、1番目の職歴の諸変数の値を代入する
3. 年齢（時変）の値が職歴頁xの開始年齢（jobstx,  $x = 2, 3, \dots$  とする）に到達したときには、x番目の職歴の諸変数の値を代入する
4. これを最後の職歴（SSM2015は職歴頁22が最長）に至るまで繰り返す

# 基本的な加工の手順：子ども人数の場合

調査項目が本人の年齢ではなく年yearを尋ねている場合は、年（時変）との比較から作成しても良い

1. 年（時変）の値が1人目の子どもの出生年に満たないときには子どもの人数は0人とする
2. 年（時変）の値が1人目の子どもの出生年に到達したときには、子どもの人数は1人とする
3. 年（時変）の値が $x$  ( $x = 2, 3, \dots$ ) 人目の子どもの出生年に到達したときには、子どもの人数は $x$ 人とする
4. これを最後の子どもの出生年（SSM2015は4人目まで）に至るまで繰り返す

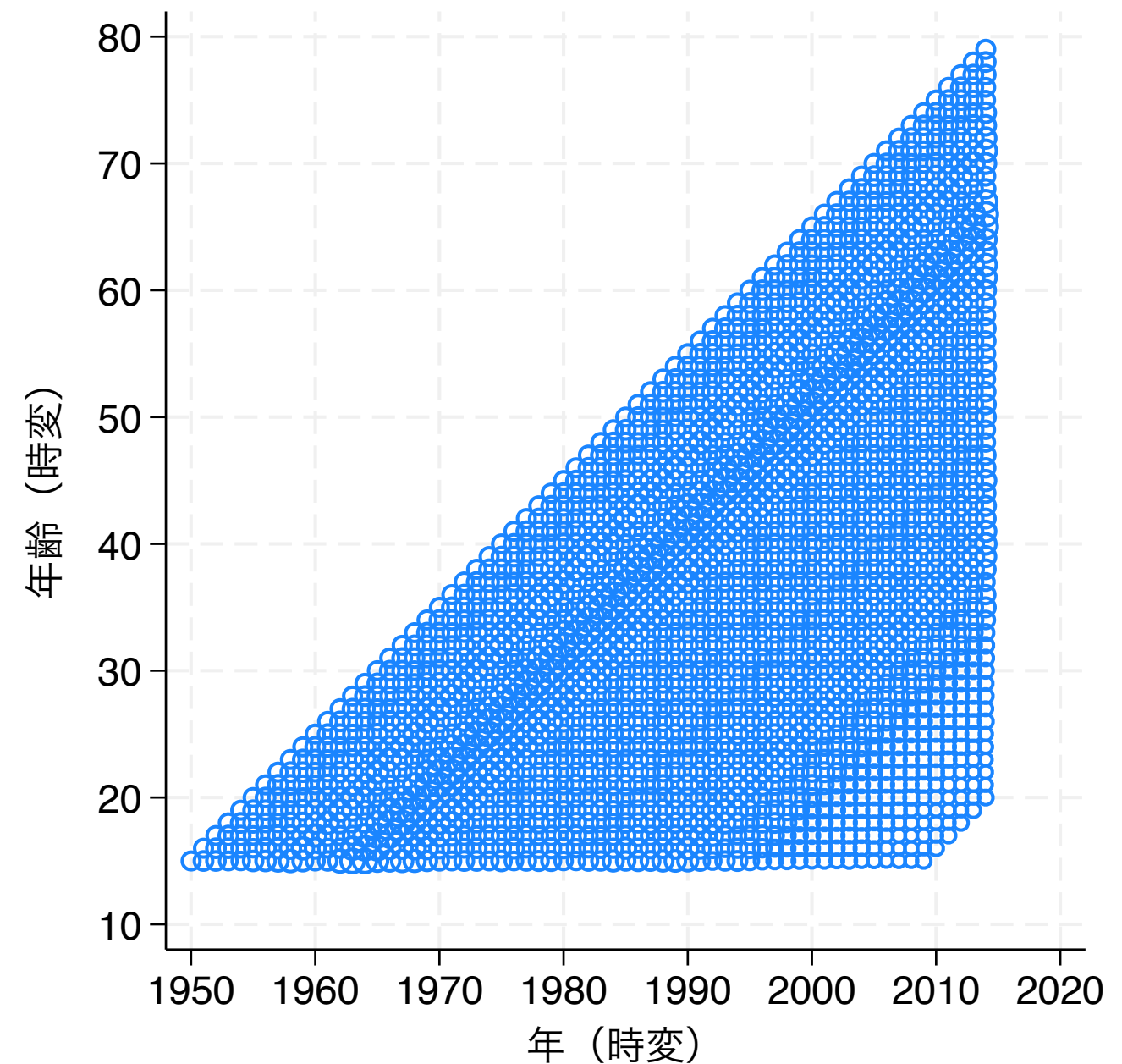
# パーソンイヤードデータの全体像：カバレッジ

新しいコーホートほど、若い年齢までの観察しか得られない（自明）

調査時点で年齢の低い対象者は回答率が低い

調査時点で年齢の高い対象者にはすでに死亡している人もいる = コーホートを代表するサンプルといえるかは検討の余地がある（c.f. 余田 2018）

調査時点からみて過去の情報になるほど、とくに職歴については、精度が低い可能性がある（後述）



# 状態変化に関するSSM職歴データの癖

**従業先の変化**：若干過小評価の可能性（離職率や転職入職率は「雇用動向調査」などと比べて若干低い）

同じ従業先のなかでの.....

- **従業上の地位の変化**：ある程度取れている（JLPSと似ている）
- **役職の変化**：役職変化を飛ばして回答していると思われるケースがあり、低めの役職（係長など）への昇進は過小報告の可能性（西澤和也さんとの雑談）
- **職業の変化**：大幅に過小に報告されている（JLPSよりもはるかに低い）

属性間の相対的な差や正負の方向に関心があるならば問題ないが、このデータから職業の変化の発生率の絶対的な水準について論じるのは相当の注意が必要

\* 正確な集計はしておらず、あくまで麦山の「感触」

\*\* ライフヒストリー・カレンダー導入以前（~SSM1995）はさらに過小評価？

# その他の職歴に関わる細かな注意点

## 古い職歴や複雑な職歴はおおざっぱに報告される（と思われる）

たとえば「近年のコーホートほど従業先・職業・従業上の地位の変化が起こる頻度が高い」（不安定性の増大）という結果が得られたとしても、古いコーホートが若い時期の変化を回答していないだけの可能性がある（cf. Manzoni et al., 2010）

## 短期の職歴の扱い

職歴は原則1年単位で収集されているが、無職期間は3ヶ月を目安に聴取されている。新しい職歴頁を常に優先してデータを作成した場合は（= 職歴頁 $x$ と職歴頁 $x+1$ の開始年齢が同じ場合に職歴頁 $x+1$ の値を優先すること。さらに、まれに有職でも同じ年齢で複数回異なる職歴を報告している回答者がいる）無職期間は見かけ上現れにくくなる（麦山 2017: 注7）

Manzoni, Anna, Jeroen K. Vermunt, Ruud Luijkx, and Ruud Muffels. 2010. “Memory Bias in Retrospectively Collected Employment Careers: A Model-Based Approach to Correct for Measurement Error.” *Sociological Methodology* 40(1):39–73.

麦山亮太, 2017, 「キャリアの中断が生み出す格差：正規雇用獲得への持続的影響に着目して」『社会学評論』68(2):248–64.

# 分析例

# イベントヒストリー分析（生存分析）

パーソンイヤードータを離散時間の観察を行ったパネルデータとみなしてイベントヒストリー分析を実施する。

時点 $t - 1$ までにイベントが起こらなかったという条件のもとで、時点 $t$ においてイベントが生起する確率（ハザード）：

$$p(t) = \Pr(Y_{it} = 1 \mid Y_{i,t-1} = \dots = Y_{i1} = 0)$$

ハザードを予測するモデルとして、たとえばロジットモデルを用いる：

$$\log \frac{p(t)}{1 - p(t)} = \alpha(t) + \beta_1 X_{1it} + \dots + \beta_k X_{kit}$$



# 初婚への移行を例としたデータ構造

age < age\_firstmarriage となる時点を0

age = age\_firstmarriage となる時点を1

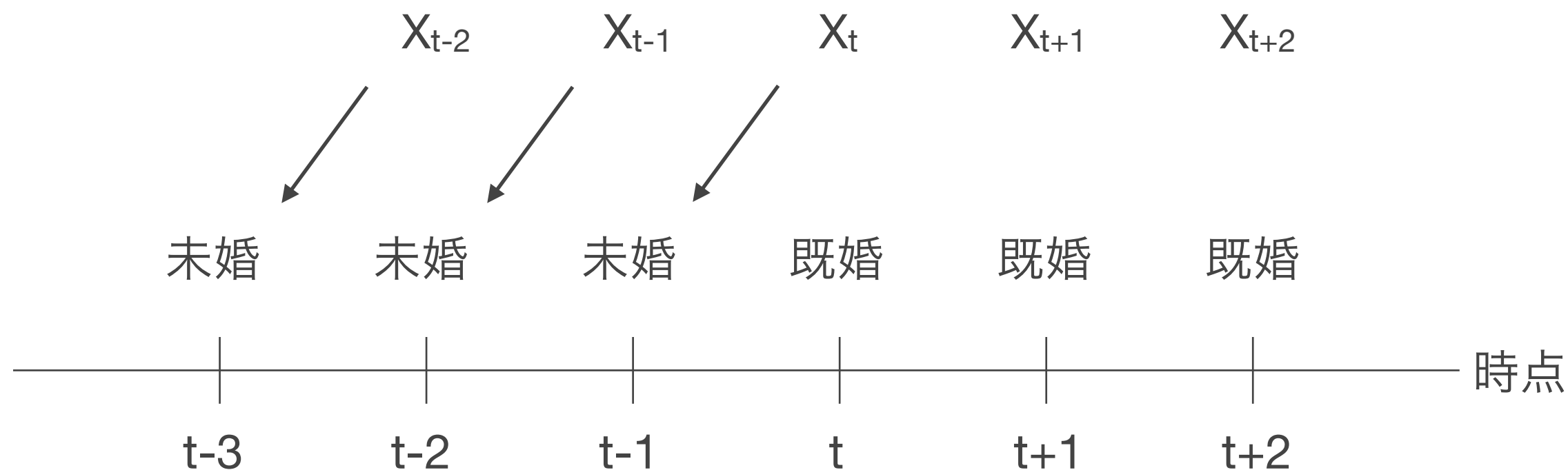
age > age\_firstmarriage は非該当（欠損）とする

id	age	age_firstmarriage	transition	status
1	18	26	0	2
1	19	26	0	2
...				
1	25	26	0	1
1	26	26	1	1
...				
2	18	.	0	3
2	19	.	0	3
...				
2	41	.	0	2
2	42	.	0	2

イベント発生時点

打ち切り

# 独立変数または従属変数の測定時点の問題



先の定義の場合、「 $t$ 時点の状態が $t-1$ から $t$ にかけての移行に影響する」と想定。  
すると、「結婚をきっかけに仕事を変える」といった行動が起こる場合は逆因果  
となってしまう

そこで、独立変数のlagを取るか、従属変数のleadを取ることで対処

どれくらいのlag/leadを取るのが適切かは、扱う問題や時間の単位に依存

# lag/leadを取る場合のデータの構造

id	age	age_firstmarriage	transition	transition_lead	status	status_lag
1	17	26	.	0	2	.
1	18	26	0	0	2	2
1	19	26	0	0	2	2
1	20	26	0	0	1	2
1	21	26	0	0	1	1
1	22	26	0	0	3	1
1	23	26	0	0	3	3
1	24	26	0	0	3	3
1	25	26	0	1	1	3
1	26	26	1	.	1	1

独立変数のlagを取る場合：  $\log \frac{p(t)}{1 - p(t)} = \alpha(t) + \beta_1 X_{i,t-1}$

従属変数のleadを取る場合：

$$\log \frac{p(t)}{1 - p(t)} = \alpha(t) + \beta_1 X_{it}, \text{ where } p(t) = \Pr(Y_{i,t+1} = 1 \mid Y_{it} = \dots = Y_{it} = 0)$$

# 固定効果モデル

パーソニイヤーデータを離散時間の観察を行ったパネルデータとみなして固定効果モデルを推定する：

$$Y_{it} = \beta_1 X_{1it} + \cdots + \beta_k X_{kit} + u_i + e_{it}$$

時点固定効果を含めたモデル（Two-way fixed effect model）：

$$Y_{it} = \beta_1 X_{1it} + \cdots + \beta_k X_{kit} + u_i + \tau_t + e_{it}$$

従属変数がカテゴリ変数の場合は線形確率モデル（Linear probability model）がよく使われる

# 固定効果モデルを推定するコマンド

## Demeaned estimation

```
xtset id year
```

```
xtreg, fe cluster(id)
```

## Least-square dummy variable estimation (LSDV)

```
ssc install reghdfe // if not installed
```

```
reghdfe, absorb(id) cluster(id)
```

\*`areg, absorb(id) cluster(id)` というコマンドもあるが、ただし、 $T$ が固定で $N \rightarrow \infty$ の場合（ほとんどのパネル調査データ）には標準誤差を過大に推計する（Cameron and Trivedi 2021）ため、LSDVをするならば`reghfre`のほうが望ましい

結語

# パーソンイヤーデータのすすめ

パーソンイヤーデータを作るのは、考え方を理解すれば実はさほど難しくない

今回のパーソンイヤーデータ作成で用いた考え方は、パネル調査データなどにも応用できる（例：JLPS）。

StataだけでなくRでも同じように適用可能（そのうちコード作ります.....）

自分が知りたい問いにとってSSM調査から構築したパーソンイヤーデータを用いる必要があるのか（知りたい問いは何か）を考えることを忘れてはいけない

- 回顧データは万能ではないという点にも注意が必要

コードに誤りなどを見つけた方はぜひお知らせください！

Stataの使い方、あるいは一般に計量分析をするときの考えかた：

- Stataを用いた計量分析入門（夏のCSRDA計量分析セミナー）
- Stataによるパネル調査データ分析の実践（春のCSRDA計量分析セミナー）

Stataコードつきの縦断データ分析の教科書：

- Cameron, Colin A., and Pravin K. Trivedi. 2022. *Microeconometrics Using Stata, Second Edition*. Stata Press.
- Rabe-Hesketh, Sophia, and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata*. Stata Press.
- Andreß, Hans-Jürgen, Katrin Golsch, and Alexander W. Schmidt. 2013. *Applied Panel Data Analysis for Economic and Social Surveys*. Springer.



# パネルデータの加工・設定でよく使うStataコマンド

**sort** データの並び替えを行う。個人→時点、の順にソートしてあるとみやすい

**by id:** idごとに何らかの変数を作ったり計算をしたりする際に用いる

**browse** データを見る（目で見てきちんとできているかを確認するのが大事）

**forvalues** 指定した値に対して繰り返し処理を実行

**variable[\_n+1]** ある変数の1行後ろの値を参照する際に用いる。

例) **by id: change = 1 if marriage == 0 & marriage[\_n+1] == 1**

**xtset id year** idを個体、yearが時点を表すパネルデータであることを宣言

**L.** 変数の前につけることで1時点前の値を参照することができる。**L2.**とすると、2時点前の値を参照できる。1時点前の値が存在しない場合は、.を返す。

**F.** 変数の前につけることで1時点後の値を参照することができる。上に同じ。