

公的統計2次利用などに関わる取り組み

独立行政法人 統計センター
樁 広計

統計センターの紹介：3つのミッション

国民の合理的な意思決定に必要な不可欠な公的統計を正確に作成し、遅滞なくかつ効率的に提供するため、統計センターに対して次の使命を果たすことが求められている。

1. 統計をつくる

人口や失業率、消費者物価指数等の我が国の基幹的な統計の作成



2. 統計を活かす

統計利用者、調査対象者、研究者が便利に安心して活用できる統計サービスの提供



3. 統計を支える

各府省、地方公共団体、国際機関、各国政府等の統計作成を支えるシステムの運用管理やプロジェクトの遂行



統計におけるオープンデータの高度化

所属組織
独立行政法人
統計センターの
最近の活動
「統計を活かす」

統計データの提供方法を高度化し、新たな付加価値を創造するサービスや革新的な事業の創出などを支援する取り組みを、総務省統計局と連携し行っており、政府が取り組んでいるオープンデータの推進を先導。

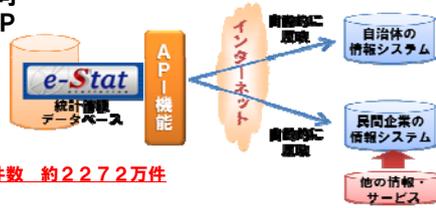
JSTAT MAPは、事業所に多様な近傍を定義し、性別年齢別人口分布をすぐに表示

ある研修：浦安市の介護施設配置のあるべき政策？

A P I 機能による統計データの提供

2014.10.31から運用開始

統計データを機械判読可能な形式で提供するAPI機能 (Application Programming Interface) を提供中



利用登録者数 3162人
統計データへのリクエスト件数 約2272万件
(平成27年3月26日現在)

活用例1：利用者の情報システムにe-Statのデータを自動的に反映

活用例2：ユーザー保有やインターネット上のデータ等と連動させた高度な統計データ分析

開発支援情報も提供中



地図による小地域分析 (jSTAT MAP)

2015.1.20から運用開始

任意に指定したエリアによる集計や利用者が保有するデータの取り込み集計する機能などを提供



利用登録者数 2334人
ログイン件数 約23000件
(平成27年3月26日現在)

活用例1：任意に指定したエリアによる集計や、利用者が保有するデータと統計データを組み合わせ、集計結果を地図上で視覚的に把握可能

活用例2：選択したエリアの年齢構成等の基本的な分析結果のレポート作成

タブレット用アプリも提供中



Needs: 人間・社会研究における 公的統計活用

2009年統計法全面改訂

- 旧統計法(1947年)法の目的
 - 第一条 この法律は、**統計の真実性**を確保し、統計調査の重複を除き、統計の体系を整備し、及び統計制度の改善発達を図ることを目的とする。
- 新統計法第一条
 - この法律は、**公的統計が国民にとって合理的な意思決定を行うための基盤となる重要な情報**であることにかんがみ、公的統計の作成及び提供に関し基本となる事項を定めることにより、公的統計の体系的かつ効率的な整備及びその有用性の確保を図り、もって**国民経済の健全な発展及び国民生活の向上に寄与**することを目的とする。
 - オーダーメイド集計や匿名データの作成・研究目的などでの提供を可能とする**統計データの二次的利用の制度**設立

統計データの二次利用促進に関する研究会

- 総務省政策統括官室:2007年スタート
 - 廣松毅(東京大学, 現情報セキュリティ大学院大学)座長
 - 内閣府統計委員会へのInput
 - 統計法改正に向けた二次利用のスタイルと展開
 - **オーダーメイド集計・匿名データ・疑似マイクロデータ**の提供
 - 利用目的としての公益性
- 統計法公布後の活動
 - 提供方法:各国制度比較と日本の特異性
 - 目的外申請:個人情報・法人情報が付随したデータをセキュアな環境を持たない研究者が管理する可能性
 - オンサイト拠点
 - 目的外申請で得た個票をセキュアな監視環境下で分析
 - 探索的なモデリングを可能とする。しかし、各拠点ごとの人員・設備整備にかなりのコスト
 - リモートアクセスによるネットワーク形成:応用統計学会からの学術会議マスタープラン提案
 - 各拠点にはデータは置かない:監視は中央で一括
 - 統計データ・アーカイブの整備

匿名データの提供

一般からの申出を受け、利用要件を満たした申出者に対し、特定の個人又は団体等が識別できないように加工して作成した調査票情報の利用を一定期間認める制度。匿名データを利用することで、行政機関が作成していない統計表の作成のみならず、多変量解析など**マイクロデータ**に基づく実証分析を行うことが可能。

【利用要件】

- 統計の作成または統計的研究にのみ利用されること
- 学術研究目的または高等教育目的の用に供することを直接の目的とすること
- 学術研究の成果または高等教育の内容が公表され、社会に還元されること
- 匿名データが適切に管理されること

オーダーメイド集計

一般からの委託を受けて、利用要件を満たした申出者に対し、調査票情報を用いて集計を行い、その結果の提供を行う制度。オーダーメイド集計を利用することで、行政機関等が作成していない統計表に基づいた分析が可能。

【利用要件】

- 統計の作成または統計的研究にのみ利用されること
- 学術研究目的または高等教育目的の用に供することを直接の目的とすること
- 学術研究の成果または高等教育の内容が公表され、社会に還元されること
- ⇒平成28年度から民間においても研究に資するものへ拡大

擬似マイクロデータの試行的提供

擬似マイクロデータは、我が国の行政機関が実施した統計調査の集計表から作成したマイクロデータ形式の擬似的なデータで、**大学等の教育機関における授業や演習**及び公的統計の二次的利用の際のテストデータなどの利用が可能であり、マイクロデータ利用者の裾野を広げ、公的統計の二次的利用の拡大を図るため無償で提供している。

【利用要件】

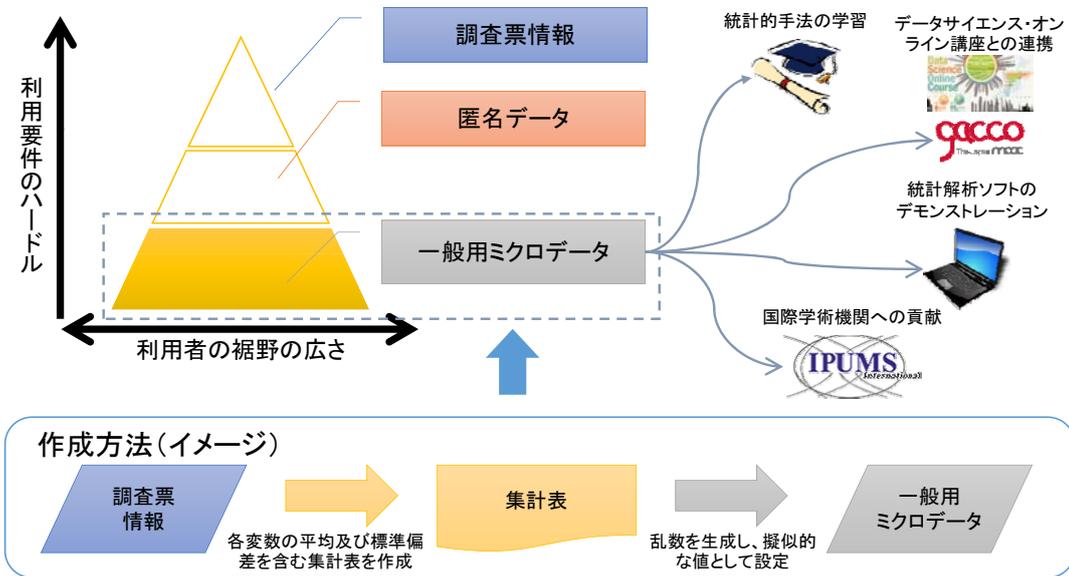
- 申請者及び利用者以外の者に利用させないこと
- 集計表から擬似的に作成したデータであるので、分析結果は実証研究の結果と見なすことができないことを理解すること
- 利用者アンケートを提出すること

試行提供している擬似マイクロデータ (平成27年5月末現在)

- 平成16年全国消費実態調査に関する試行版の擬似マイクロデータ
 - 大規模データ（CSV形式のみ）
 - 約200項目：世帯情報、収入、実支出（消費支出、非消費支出）、実支出以外、繰越金
 - 約3万2千レコード
 - 簡易データ（CSV形式 及び Excel形式）
 - 25項目：世帯情報、消費支出
 - 約8千レコード

3 今後の予定

擬似マイクロデータの提供は、平成27年度末で終了する。
平成28年度からは、一般用マイクロデータの提供に移行する。



統計センターと学術研究機関等との連携について 学術研究機関等との連携協力の取組

統計センターでは、平成21年4月施行の統計法において創設された公的統計の二次的利用制度の充実と学術研究の発展を図っていくため、学術研究機関等と連携した取組を展開。公的統計の利用拡大に係る取組に賛同する法人と連携協力協定を締結し、公的統計の二次的利用に関する研究・開発、普及・啓発を推進するほか、これらの法人に**統計データアーカイブのサテライト機関の役割を担ってもらい、研究者等に向けた二次的利用のサービスの充実を図っている。**

法人の要件

1. 国立大学法人法に基づき設置された国立大学法人及び大学共同利用機関法人
2. 私立学校法に基づく学校法人により設置された私立大学
3. 独立行政法人通則法及び個別法の定めるところにより設立された独立行政法人
4. 法人税法別表第1に掲げる公共法人
5. 公益社団法人及び公益財団法人の認定等に関する法律により公益性の認定を受けた公益法人（特例民法法人を含む。）

連携協力事項

- 公的統計に関するデータアーカイブの運営に関すること（**施設基準に適合**）
 - ・匿名データの提供
 - ・オンサイト利用環境の提供など
- 公的統計の二次的利用に関する研究・開発
- 公的統計の二次的利用に関する普及・啓発
- 人材交流
- その他協定の目的を達成するために必要な事項

連携協力協定を締結している大学等

サテライト機関名	匿名データの是非	オンライン利用環境	二次的利用に係るURL
国立大学法人 一橋大学 経済研究所附属社会科学統計情報研究センター	○	○	http://rcisss.ier.hit-u.ac.jp/Japanese/micro/index.html
国立大学法人 神戸大学大学院 経済学研究科	○	—	http://www.econ.kobe-u.ac.jp/kuma/satellite/index.html
法政大学 日本統計研究所	○	—	http://www.hosei.ac.jp/toukei/micro/index.html
大学共同利用機関法人 情報・システム研究機構 新領域融合研究センター(統計数理研究所)	○	○	http://www.rois.ac.jp/tric/tokumei/tokumei.html

進行形の活動： 公的統計マイクロデータ 活用のためのネットワーク基盤

総務省「統計の2次利用研究会」

川崎茂応用統計学会長(当時)「学術会議マスタープラン」提出

内閣府統計委員会の「公的統計の整備に関する基本計画

基本計画2014年3月25日閣議決定

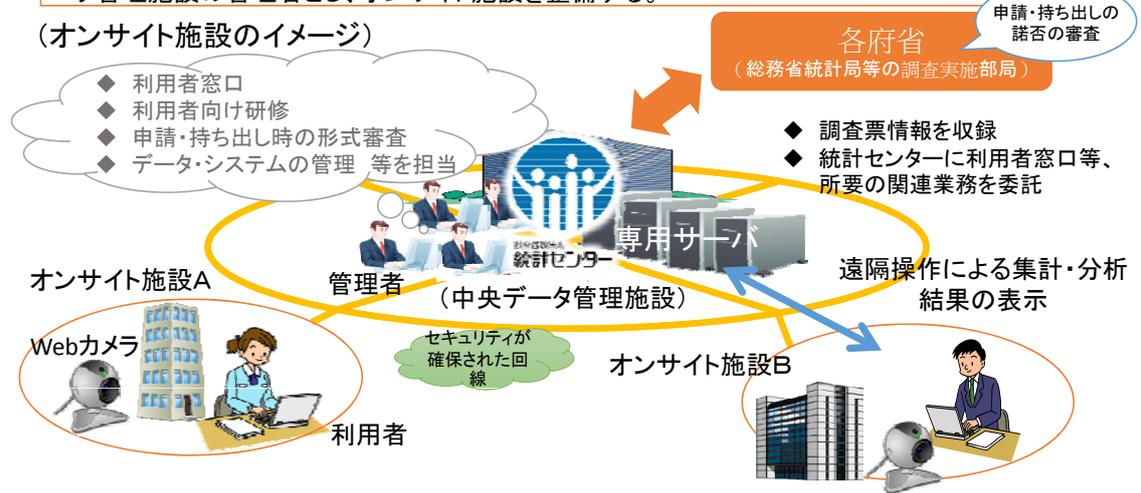
公的統計の整備に関する基本計画への組み込み

- 内閣府統計委員会
 - 2013年10月30日総務大臣より諮問, 2014年1月31日答申
- 2014年3月25日閣議決定
- 調査票情報等の提供及び活用については、
 - セキュリティレベルや調査票情報等の匿名性の程度に応じた利用形態ごとの特性
 - 諸外国における取組状況等を総合的に勘案した上
- 法制度上の整理を含め、**以下の取組**を行う
 - オーダーメイド集計における利用条件の緩和に向けた検討
 - 調査票情報の提供における**リモートアクセスを含むオンサイト利用**や**プログラム送付型集計・分析の実現に向けた整理**・検討
 - 匿名データの作成及び提供における提供対象統計調査の種類や年次の追加等によるサービスの充実
- その際、効率性及び利便性の観点から、**政府一体として一元的な取組**を推進する。
 - 主担当総務省政策統括官, 府省横断の中核機関: (独)統計センター
 - 二次利用研究会での検討⇔リモートアクセス拠点形成⇔**大学などの公的統計マイクロデータ2次利用フォーラム形成**
 - **情報・システム研究機構平成28年3月29日一橋講堂で立ち上げ**:平成28年度概算要求:統計センター人員の確保

(参考)調査票情報のオンサイト利用(現在検討中)

○ オンサイト利用については、長年にわたる課題であったが、IT技術の進展等を踏まえ、平成28年度を目的に、リモートアクセスを活用したオンサイト利用の試行を開始する。今後も、総務省政策統括官(統計基準担当)は制度面の検討・関係府省と調整するとともに、総務省統計局と(独)統計センターが連携して技術面を検討する。

○ 具体的な試行については、統計局において、学界や各府省の協力を得つつ、統計局の統計調査の調査票情報を主たる対象に、(独)統計センターを政府共通の基盤である中央データ管理施設の管理者とし、オンサイト施設を整備する。



公的統計 マイクロデータ分析の課題

政府統計部局の役割 研究の障害除去：データがない、使いにくい

- 提供可能な公的統計マイクロデータの範囲拡大
 - 匿名データ・オーダーメイド集計の範囲拡大
 - 事業所・企業統計などの匿名化困難
- 適切なメタデータの付与
 - 現状：統計局：CSVあるいはXMLでの提供
 - 研究者の負担
 - 研究に用いる複数データセットの変数結合 * 国内外の類似データとの結合
 - 適切なメタデータの付与：LOD (Linked Open Data) 形式によるデータ提供
 - LOD：関連するデータ同士が相互に結びついているデータ
 - 統計オープンデータモデル事業
 - 福井県・統計局・統計センター協働事業
 - 統計局提供：国勢調査 ¥ 社会・人口統計体系等の統計データ
 - 福井県提供各種行政データ

調査環境の悪化と 統計精度向上の社会要求

- 未回答
 - あってはならない未回答率の増大
 - Imputationの増大(回答データとして扱う問題)
- 欠測値だけのImputation(モデルベースの予測値)か？
 - 無作為抽出と不偏推定の意味
 - 抽出(観察)データの実測値 Y_t
 - 非抽出データのモデルベースの予測値 $Y(x_{t-1})$: Best Available Predicts
 - データの観測確率: $P(z_{t-1})$: 抽出確率 \times 未回答確率モデル
 - 不偏推定量の構成
 - 実測した場合: $Y(x_{t-1}) + \{1/P(z_{t-1})\}\{Y_t - Y(x_{t-1})\}$
 - 実測しない場合: $Y(x_{t-1})$

おわりに

- 社会のための科学の推進
 - 一学会、一大学、一研究所でできることの限界
 - 多様な価値を受容したネットワーク形成は必然
 - 産官学の連携
- 社会のための科学の支援基盤
 - データ基盤: 公的統計のみならず重要
 - そのセキュリティ
 - 知識基盤:
 - モデル・プロセスの記述標準・産官学の共有・相互評価
- この種の事業を進める人財の価値評価

参考資料1

- 匿名データの利用に関するFAQ:
 - <http://www.nstac.go.jp/services/faq-anonymity.html>
- 擬似マイクロデータの利用に関するFAQ
 - <http://www.nstac.go.jp/services/faq-gijimicro.html>

(参考資料2) 匿名データの利用実績例

利用目的	調査名	研究の名称
学術研究目的	社会生活基本調査	正規雇用者における平日の労働時間と休息時間 —「社会生活基本調査」マイクロデータによる分析— 生活行動からみる高齢者の行動特性について —社会生活基本調査の匿名データを用いて— 子供のいる世帯における夫と妻の2次活動時間の差異について —社会生活基本調査の匿名データを用いて— 趣味・娯楽活動の時間について 個人・世帯属性と行楽・観光旅行行動の関係
	全国消費実態調査	等価尺度の推計と比較—消費上の尺度・制度的尺度・OECD尺度— 『季刊社会保障研究』Vol.48 Spring 2013 No.4 所得格差変動の年齢階級別要因分解 :全国消費実態調査マイクロデータを用いて
	全国消費実態調査 住宅・土地統計調査	持家取得における既婚女性の就業の役割
	就業構造基本調査	転職経験および転職理由と転職希望意識との関連について —就業構造基本調査匿名データによる統計分析 若年者就業率における賃金弾力性の推定 女性事務職の賃金と就業行動 —男女雇用機会均等法施行後の三時点比較— 税負担と労働供給 —「日本労働研究雑誌」No.605 2010.12 — 若者の有業・無業状態における属性の考察 女性の働き方と少子化に関する考察
高等教育目的	就業構造基本調査 全国消費実態調査 社会生活基本調査	一橋大学大学院経済学研究科「演習」(労働経済学Ⅰ)

論文等: <http://www.nstac.go.jp/services/jisseki.html>

(参考資料3) オーダーメイド集計の利用実績例

調査名	研究の名称
国勢調査	岡山大学大学院社会文化科学研究科紀要第35号(2013.3)
	・2005年国勢調査にみる在日外国人女性の結婚と仕事・住居
	・2005年国勢調査にみる外国人の教育
	・2005年国勢調査にみる在日外国人の仕事
	Access to childcare and the employment of women with preschool-aged children in Tokyo
	看護人材の就業率の推移
	ー再検討した潜在者数推計方法による結果からー
	地域別経済指標に基づくSDモデルの開発
	在日外国人の仕事
	ー2000年国勢調査データの分析からー
全国消費実態調査	家族・ジェンダーからみる在日外国人
	ー2000年国勢調査データの分析からー
	在学率と通学率から見る在日外国人青少年の教育
	ー2000年国勢調査データの分析からー
就業構造基本調査	エリア・サンプリングにおける問題点
	新・家計消費論
	ー高齢層を支える都市部消費ー
	近年における都道府県別貧困率の推移について
	ーワーキングプアを中心に

論文等: <http://www.nstac.go.jp/services/jisseki.html>

参考資料4: 公的統計部局の課題: 提供データ範囲の拡大:

統計センターにおける匿名データ及びオーダーメイド集計の提供

提供データ一覧(厚生労働省が別途国民生活基礎調査提供)

調査名(年次等)	匿名	オーダー	調査名(年次等)	匿名	オーダー	調査名(年次等)	匿名	オーダー
国勢調査			住宅・土地統計調査			社会生活基本調査【調査票A】		
昭和55年	○		昭和53年		○	生活時間編・生活行動編		
昭和60年	○		昭和58年		○	昭和56年		○
平成2年	○		昭和63年		○	昭和61年		○
平成7年	○		平成5年	○	○	平成3年		○
平成12年	○	○	平成10年	○	○	平成8年		○
平成17年	○	○	平成15年	○	○	平成13年		○
平成22年	○		平成20年	○	○	平成18年		○
			平成25年		○	平成23年		○
就業構造基本調査					●	社会生活基本調査【調査票B】		
昭和54年	○		全国消費実態調査			生活時間編		
昭和57年	○		平成元年		○	平成13年		○
昭和62年	○		平成6年		○	平成18年		○
平成4年	○	○	平成11年		○	労働力調査【基礎調査票】		
平成9年	○	○	平成16年		○	昭和55年1月～63年12月		○
平成14年	○	○	平成21年		○	平成元年1月～23年12月		●
平成19年	○					平成24年1月～26年12月		○
平成24年	○					労働力調査【特定調査票】		
						平成14年1月～26年12月		○
						家計調査		
						昭和56年1月～平成26年12月		●
						家計消費状況調査		
						平成14年1月～26年12月		○
						消費動向調査		
						平成16年4月～27年3月		○
						企業行動に関するアンケート調査		
						平成18年度～26年度		○
						賃金構造基本統計調査		
						平成18年～26年		○
						学校基本調査【高等教育機関編】		
						平成20年度～26年度		●
						学校基本調査【初等教育機関編】		
						平成20年度～22年度		○
						建設着工統計調査		
						平成21年4月～27年3月		○

○印は提供中、●印は(複数年存在するものについては最新年(度)分を)平成27年度中に提供予定

～ 利用可能な統計調査は、今後、さらに拡大予定 ～

地方創生と人口統計

— まち・ひと・しごと再生の課題 —

国立社会保障・人口問題研究所

金子隆一

話しのアウトライン

1. 経緯

日本創成会議～創生本部～自治体

2. 地域人口の概観

地方消滅 と 東京一極集中

“少子化対策”と人口1億人維持

3. 統計データの利用

RESAS

(1) 日本創生会議 増田寛也+人口減少問題検討分科会



(2013.12)

2040年に若年女性が2010年の50%以下に減少する自治体 (= “消滅可能性都市”) が半数に昇る!



(2014.6)

(2014.7)



(2014.8)



(2) 政府の動き

報告書ポイント

- (1)「1970年代モデル」から「21世紀(2025年)日本モデル」へ
- (2)すべての世代を対象とし、すべての世代が相互に支え合う仕組み
- (3)女性、若者、高齢者、障害者などすべての人々が働き続けられる社会
- (4)すべての世代の夢や希望につながる子ども・子育て支援の充実
- (5)低所得者・不安定雇用の労働者への対応
- (6)地域づくりとしての医療・介護・福祉・子育て
- (7)国と地方が協働して支える社会保障制度改革
- (8)成熟社会の構築へのチャレンジ (世界の先頭を歩む)

永井良三 自治医科大学学長
西沢和彦 日本総合研究所調査部上席主任研究員
増田寛也 野村総合研究所顧問
宮武 剛 目白大学大学院生涯福祉研究科客員教授
宮本太郎 中央大学法学部教授
山崎泰彦 神奈川県立保健福祉大学名誉教授

- ・ 人口構造（の変化）を前提とした社会保障の重要性の認識
- ・ 社会保障の枠の拡大 → 少子化への対策等を（明確に）組み入れる
- ・ 地域への着目

選択する未来

未来への選択 — 人口急減・超高齢社会を超えて、日本発成長・発展モデルを構築 —

50年後も、人口減少が続き、加速。
現状のままでは、「人口急減・超高齢社会」の到来

人口 (2013年)12,730万人 → (2060年)8,674万人
減少幅 { 2008~2013年 ▲16万人/年
{ 2010年代後半~2020年代初頭 ▲50~60万人/年
{ 2040年代初頭 ▲100万人/年
高齢化率 (2013年)25% → (2060年)40%

- 【現状のまま何もしない場合の未来像】
- ・ プラス成長を続けることは困難になり、マイナス成長が定着
 - ・ 「人口オーナス」と「縮小スパイラル」の双方が作用し、国民生活低下のおそれ
 - ・ 女性、高齢者、若者が活躍できない労働市場の二極化、格差の固定化・再生産
 - ・ 地方で4分の1以上の自治体が消滅可能性、東京では超高齢化
 - ・ 医療・介護費の増加により財政破たんリスクの高まり



【未来への選択】 ・制度、政策や人々の意識が速やかに変われば、「人口急減・超高齢社会」への流れは変えられる
・若い世代や次の世代が豊かさを得て、結婚し、子どもを産み育てることができるよう集中して改革・変革

<p>①人口</p> <p>50年後に1億人程度(この場合、その一世代後には微増に転じる)</p> <ul style="list-style-type: none"> ・ 国民の希望どおり子どもを産み育てられる環境により、1億人程度の人口を保持 ・ 資源配分を高齢者から子どもへシフト、出産・子育て支援を倍増。費用は現世代で負担 ・ 子どものための政策推進 	<p>②成長・発展</p> <p>経済を世界に開き、「創意工夫による新たな価値の創造」により、成長し続ける</p> <ul style="list-style-type: none"> ・ イノベーションが生産性向上の切り札 ・ 産業・企業の「新陳代謝・若返り」(ダイナミズム) ・ オープンな国づくりと、外国人材の戦略的受け入れ ・ 債務残高対GDP比引下げ等の明確な目標 	<p>③人の活躍</p> <p>年齢、性別に関わらず能力発揮</p> <ul style="list-style-type: none"> ・ 男女の働き方改革により、能力や意欲に応じた活躍の機会充実 ・ 70歳まで働ける社会(新生産年齢人口) ・ 未来の技術や産業に適応したプレイヤーの育成 ・ 格差の再生産の回避 	<p>④地域の未来</p> <p>個性を活かした地域戦略、集約・活性化</p> <ul style="list-style-type: none"> ・ 新しい発想で資源を利活用し、働く場所をつくる(農業、観光等) ・ 「集約・活性化」によるコンパクトな地域・地方中枢都市圏の形成 ・ 東京への若者の人口流出を抑制 ・ 東日本大震災の復興を地域のモデルに 	<p>⑤信頼・規範</p> <p>基盤的な制度、文化、公共心など社会の土台を大切にする</p> <ul style="list-style-type: none"> ・ 日本の国土に育まれた伝統、文化、美意識、価値観の継承・発信 ・ 国際貢献やルールづくりへ参加、世界に発信し続ける ・ 社会保障制度や財政の持続可能性の確保
--	---	--	---	---

骨太の方針と改革の基本方針2014 ~デフレから好循環拡大へ~ (平成26年6月24日閣議決定) (抜粋)

(2014.6)

第1章 アベノミクスのこれまでの成果と今後の日本経済の課題

4. 日本の未来像に関わる制度・システムの改革 (「人口急減・超高齢化」の克服)

デフレ脱却・経済再生の先に、もう一つ超えなければならない高いハードルがある。現在の日本は、「人口急減・超高齢化」へ確実に向かっている。この流れを変えなければ、持続的・安定的な成長軌道に乗っていくことはできない。

人口急減・超高齢化の流れを変えることは容易でなく、流れが変わっても効果が現れるまで長期間を要する。人口急減・超高齢化の流れを変えられない場合には、経済規模が収縮し、縮小スパイラルに陥るおそれがある。そこに至っては、もはや回復は困難となろう。従来の少子化対策の枠組みにとらわれず、福祉分野以外にも、教育、社会保障、社会資本整備、地方行財政、産業振興、税制など、あらゆる分野の制度・システムを若者・子ども世代や次の世代のためになっているか、結婚しやすく子育てしやすい環境を実現する仕組みになっているかという観点から見直し、2020年を目途にトレンドを変えるために抜本的な改革・変革を推進すべき時期にきている。

希望通りに働き、結婚、出産、子育てを実現することができる環境を整え、人々の意識が大きく変わり、2020年を目途にトレンドを変えていくことで、50年後にも1億人程度の安定的な人口構造を保持することができると見込まれる。

7

まち・ひと・しごと創生本部の組織体制 (2014.9)



創生本部

まち・ひと・しごと創生本部

○設置根拠: 閣議決定

○構成:

本部長 総理大臣
副本部長 地方創生担当大臣、官房長官
本部員 他の全ての国务大臣
※その他必要に応じて本部長が出席を求める



まち・ひと・しごと創生本部事務局

○設置根拠: 総理決定

○構成:

事務局長 官房副長官(事務)
事務局長代行 ・総理大臣補佐官(地方創生等担当)
・官房副長官補(内政)
事務局長代理 (3名)

まち・ひと・しごと創生会議

○設置根拠: 本部長決定

○構成:

議長 総理大臣
副議長 地方創生担当大臣、官房長官
議員 ・経済財政担当大臣、少子化担当大臣、復興大臣、総務大臣、財務大臣、文部科学大臣、厚生労働大臣、農林水産大臣、経済産業大臣、国土交通大臣
・民間有識者

※その他必要に応じて議長が出席を求める

まち・ひと・しごと創生会議 民間有識者 (12名)
池田 弘 公益社団法人日本ニュービジネス協議会連合会会長
伊東 香織 岡山県倉敷市長
おおくま 大社 充 NPO法人グローバルキャンパス理事長
奥田 麻依子 高松環海土町 藤岐高前高松魅力化コーディネーター
坂根 正弘 コマツ相談役
清水 志摩子 NPO法人全国商店街おかみさん会理事長
田中 進 農業生産法人(株)サラダボウル代表取締役
富山 和彦 経営共創基盤代表取締役CEO
中橋 恵美子 NPO法人わははネット理事長
樋口 英雄 慶應義塾大学商学部教授
増田 寛也 東京大学公共政策大学院客員教授
山本 顕樹夫 帯広畜産大学監事、前小樽商科大学長

まち・ひと・しごと創生本部幹事会

○設置根拠: 本部長決定

○構成:

議長 地方創生担当大臣
議長代理 地方創生担当副大臣、官房副長官(事務)
副議長 地方創生担当大臣政務官、総理大臣補佐官、官房副長官補
議員 事務局長代理、全事務次官・長官



8

高齢化の進展に的確に対応し、人口の減少に歯止めをかけるとともに、東京圏への人口の過度の集中を是正し、それぞれの地域で住みよい環境を確保して、将来にわたって活力ある日本社会を維持していくために、まち・ひと・しごと創生（※）に関する施策を総合的かつ計画的に実施する。

※まち・ひと・しごと創生：以下を一体的に推進すること。

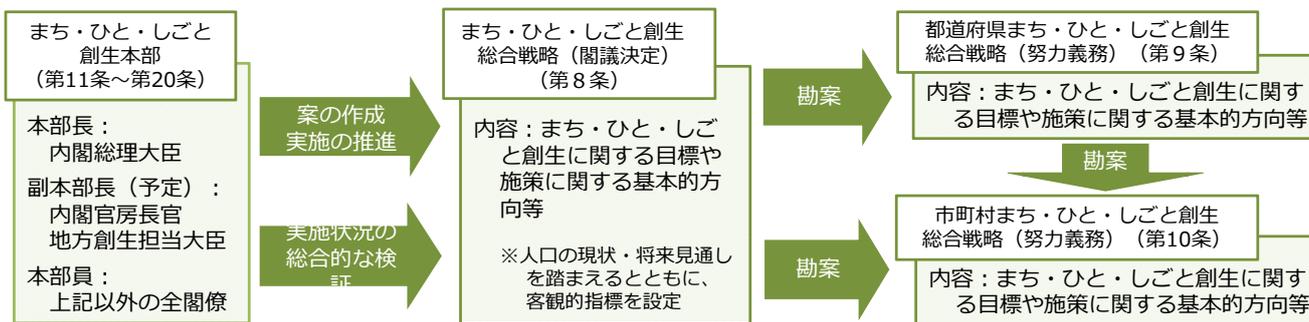
まち…国民一人一人が夢や希望を持ち、潤いのある豊かな生活を安心して営める地域社会の形成

ひと…地域社会を担う個性豊かで多様な人材の確保

しごと…地域における魅力ある多様な就業の機会の創出

基本理念（第2条）

- ①国民が個性豊かで魅力ある地域社会で潤いのある豊かな生活を営めるよう、それぞれの地域の実情に応じた環境を整備
- ②日常生活・社会生活の基盤となるサービスについて、需要・供給を長期的に見通しつつ、住民負担の程度を考慮して、事業者・住民の理解・協力を得ながら、現在・将来における提供を確保
- ③結婚・出産は個人の決定に基づくものであることを基本としつつ、結婚・出産・育児について希望を持てる社会が形成されるよう環境を整備
- ④仕事と生活の調和を図れるよう環境を整備
- ⑤地域の特性を生かした創業の促進・事業活動の活性化により、魅力ある就業の機会を創出
- ⑥地域の実情に応じ、地方公共団体相互の連携協力による効率的かつ効果的な行政運営の確保を図る
- ⑦国・地方公共団体・事業者が相互に連携を図りながら協力するよう努める



施行期日：公布日（創生本部・総合戦略に関する規定は、公布日から1か月を超えない範囲内で政令で定める日）

まち・ひと・しごと創生「長期ビジョン」

長期ビジョン

3つの視点

若い世代の
就労・結婚・子育て
の希望の実現

『東京一極集中』
の歯止め

地域の特性に
即した地域課題
の解決

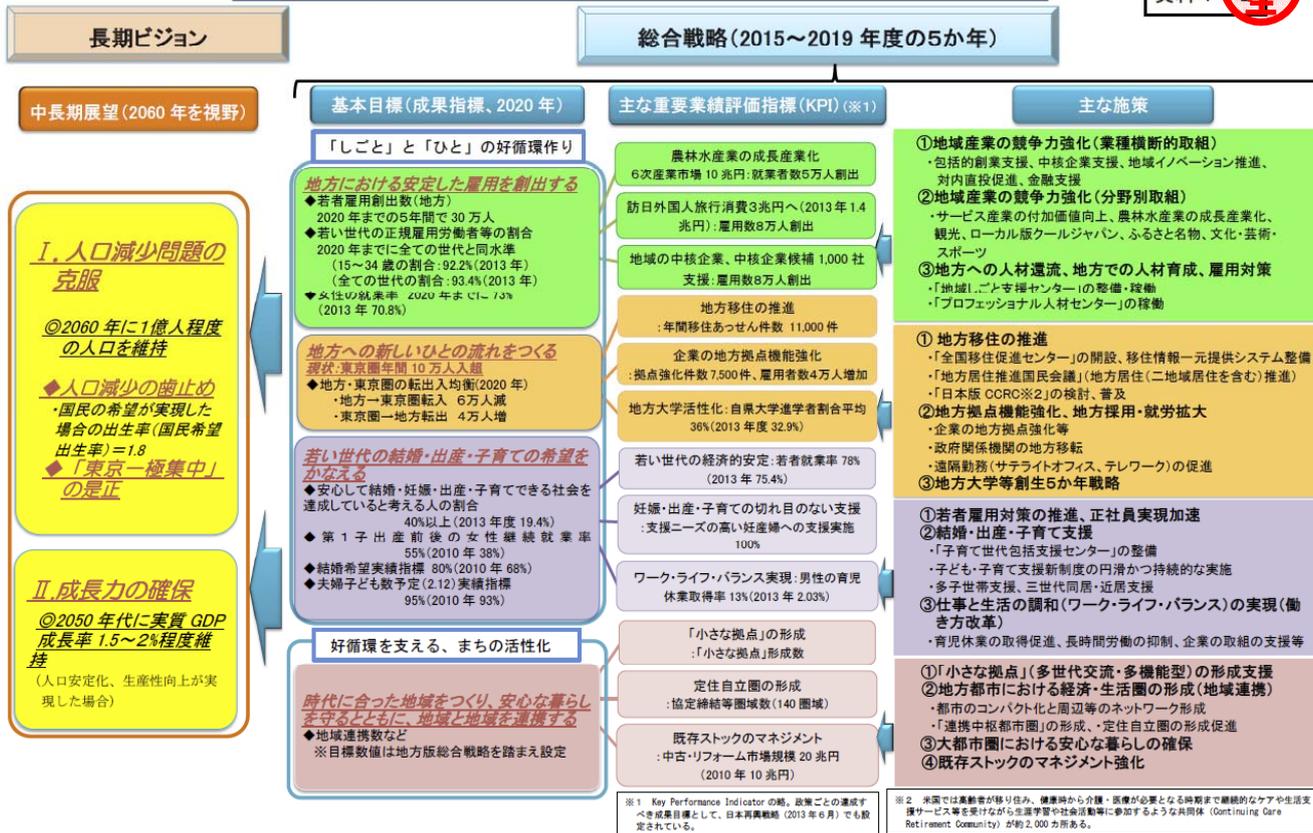
「しごと」と「ひと」の好循環を実現するための、4つの目標

- ①地方における安定的な雇用を創出
- ②地方への新しいひとの流れをつくる
- ③若い世代の結婚・出産・子育ての希望をかなえる
- ④時代に合った地域をつくり、安心な暮らしを守るとともに、地域と地域を連携する

魅力あふれる地方を創生

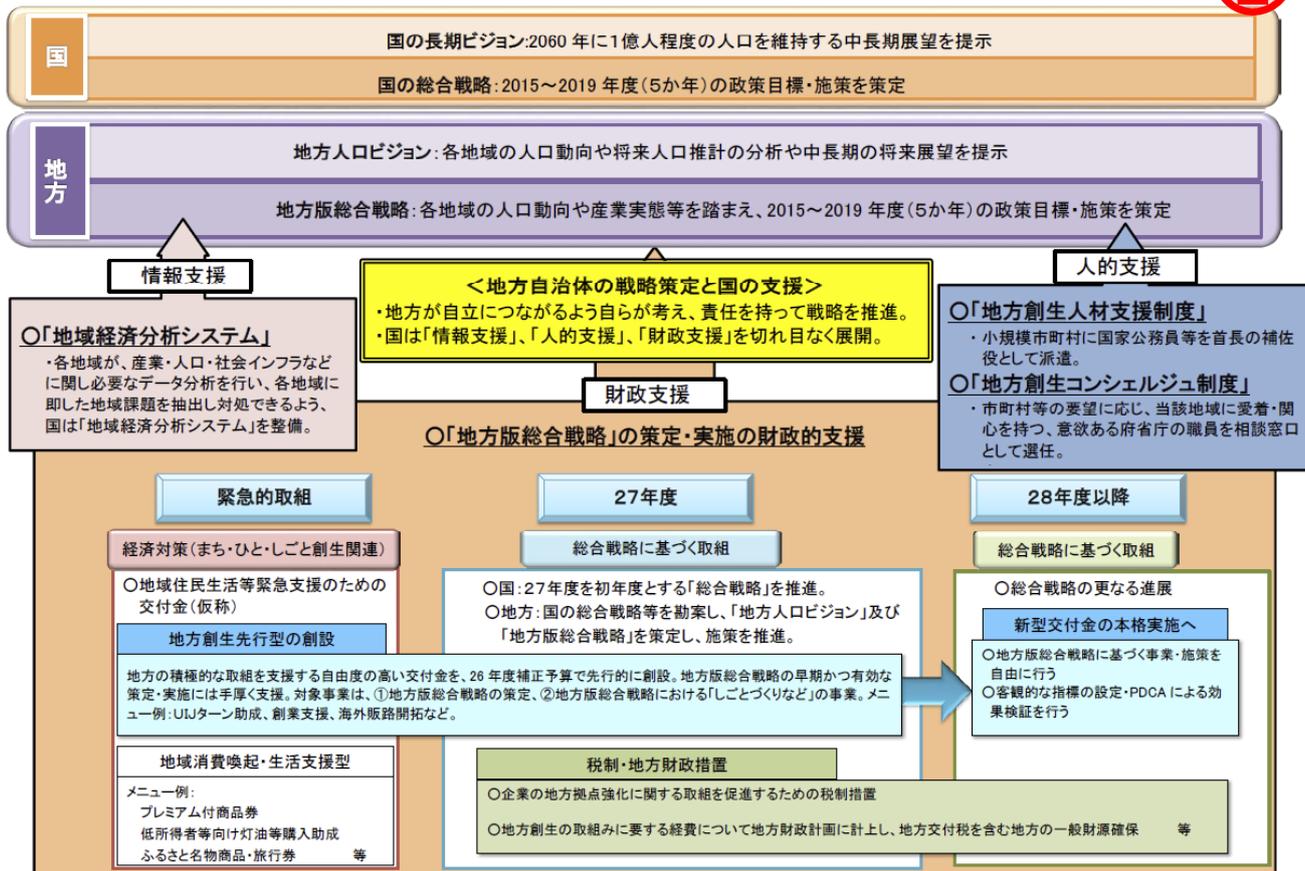
まち・ひと・しごと創生「長期ビジョン」と「総合戦略」の全体像

資料4



地方への多様な支援と「切れ目」のない施策の展開

資料4



2. 地域人口の概観

地方消滅 と 東京一極集中

“少子化対策”と人口1億人維持

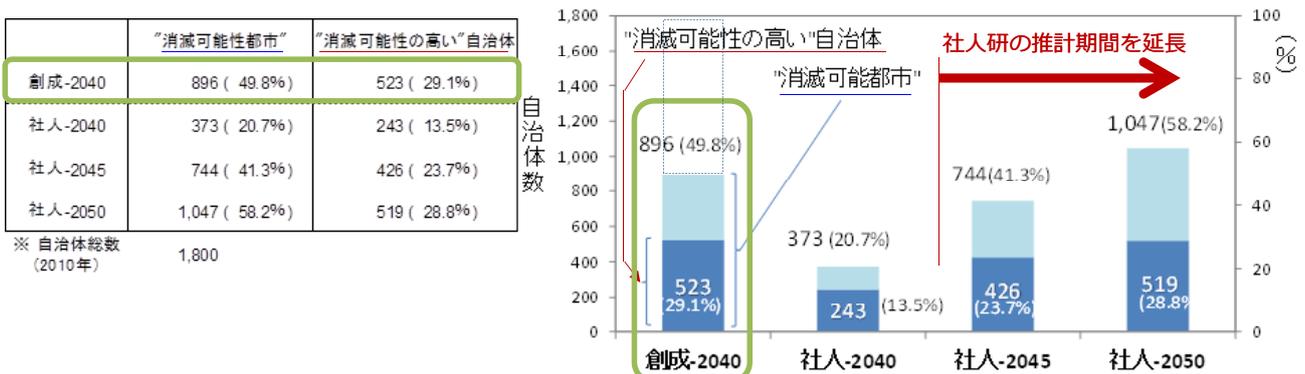
■ 自治体の「消滅可能性」について

【日本創成会議（人口減少問題検討分科会）】

今後の「人口移動が収束しない」とすると、2040年に若年女性が2010年の50%以下に減少する自治体（＝“消滅可能性都市”）は 896（全体の 49.8%）となり、さらにその中で人口規模が1万人を下回る自治体（＝“消滅可能性が高い”自治体）は 523（全体の 29.1%）となる。

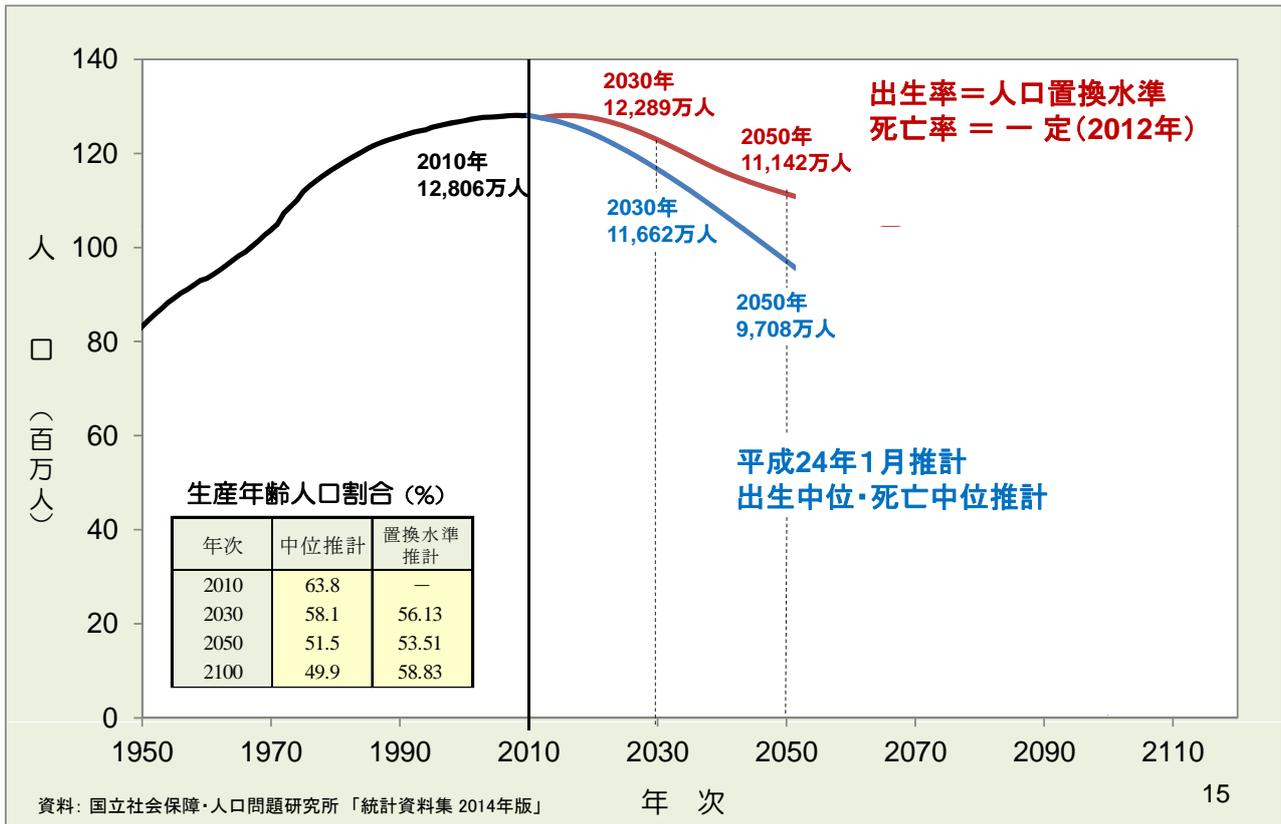
- 社人研推計（平成24年3月推計）では、2005～10年の純移動率は2015～20年に1/2となることを基本仮定としているが（自治体の類型化による差異あり）、これによると 2040年の“消滅可能性都市”は、373（全体の20.7%）、“消滅可能性が高い”自治体は、243（13.5%）である（図表）。
- しかし、社人研推計においても約10年の差で、創成会議推計とほぼ同様の状況が現出する（図表）。
- すなわち、創成会議推計は社人研推計と比較して、人口減少、年齢構造変化（高齢化）についての変化ペースが速いだけで、長期的な帰結に違いはないと見ることができる。

創成会議推計と社人研推計の比較



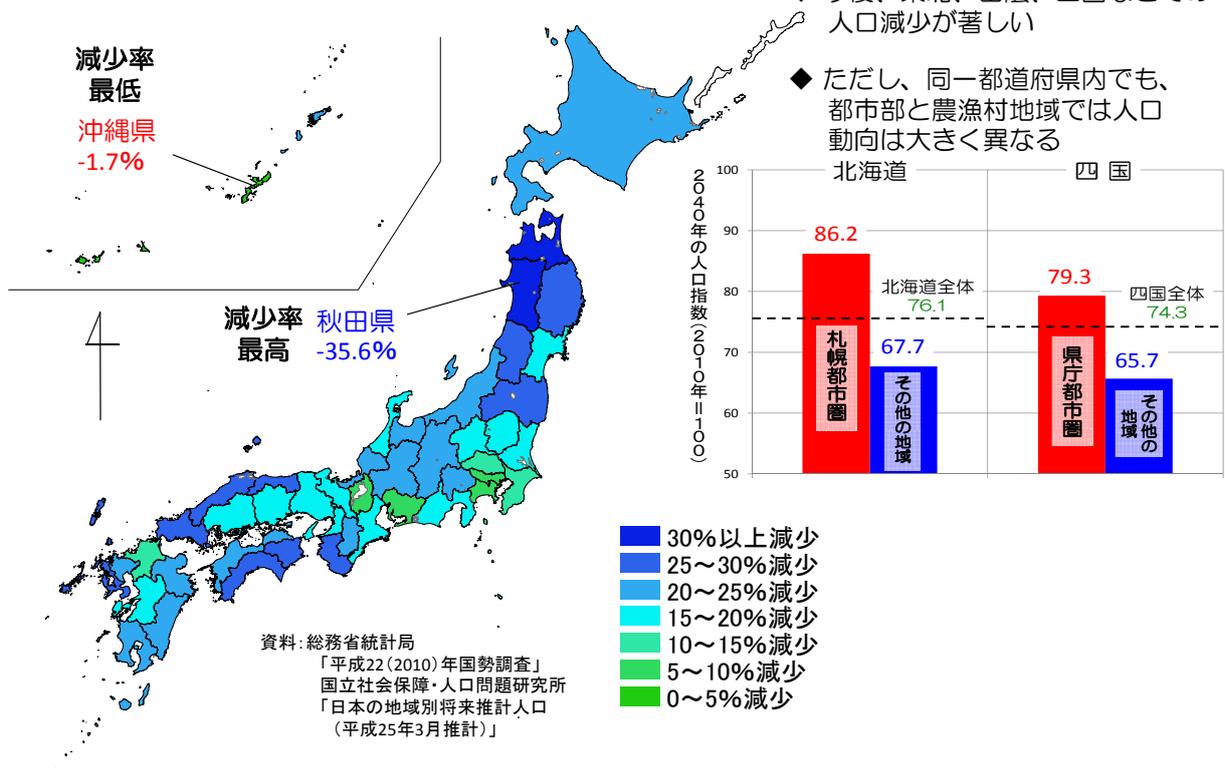
※ 社人-2045,2050は、2035～40年の移動率を含めた動態率仮定値を固定して延長推計試算したもの（公表値ではない）。

人口減少に対する人口モメンタムの効果

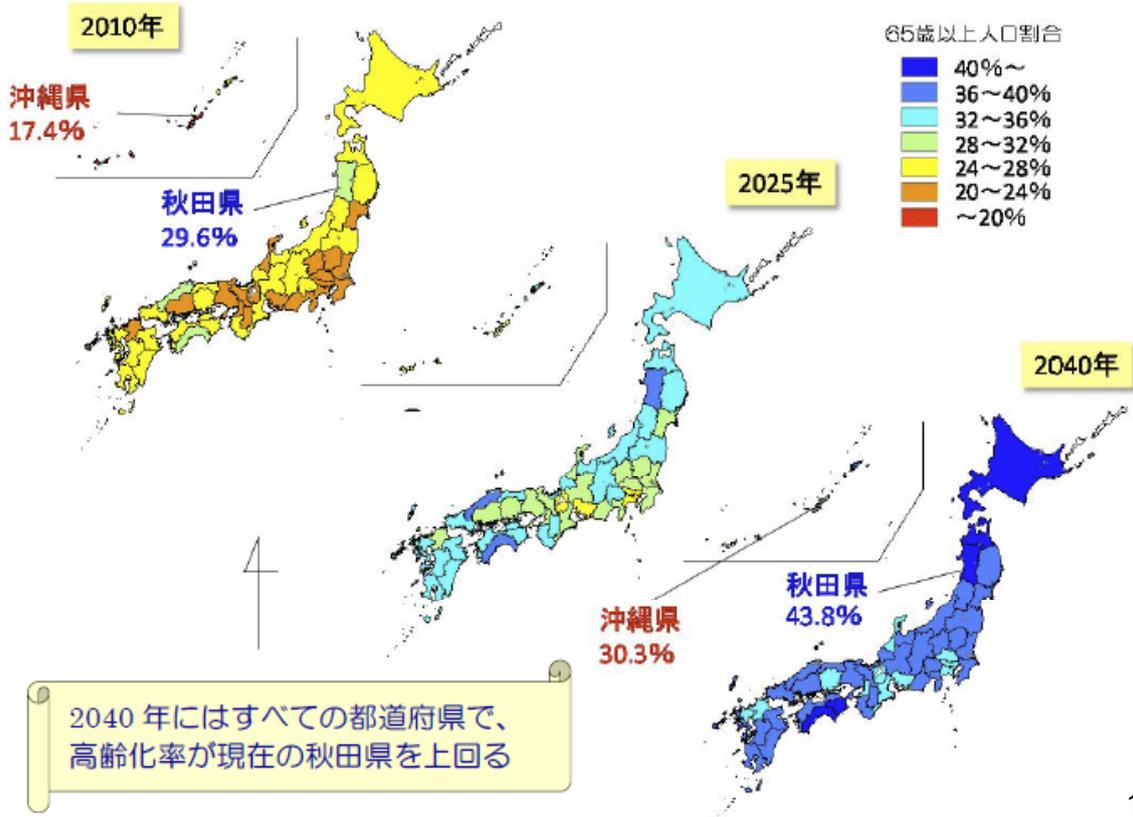


人口減少の程度は、地域差が大きい

都道府県別にみた 2010年 → 2040年の人口増減率

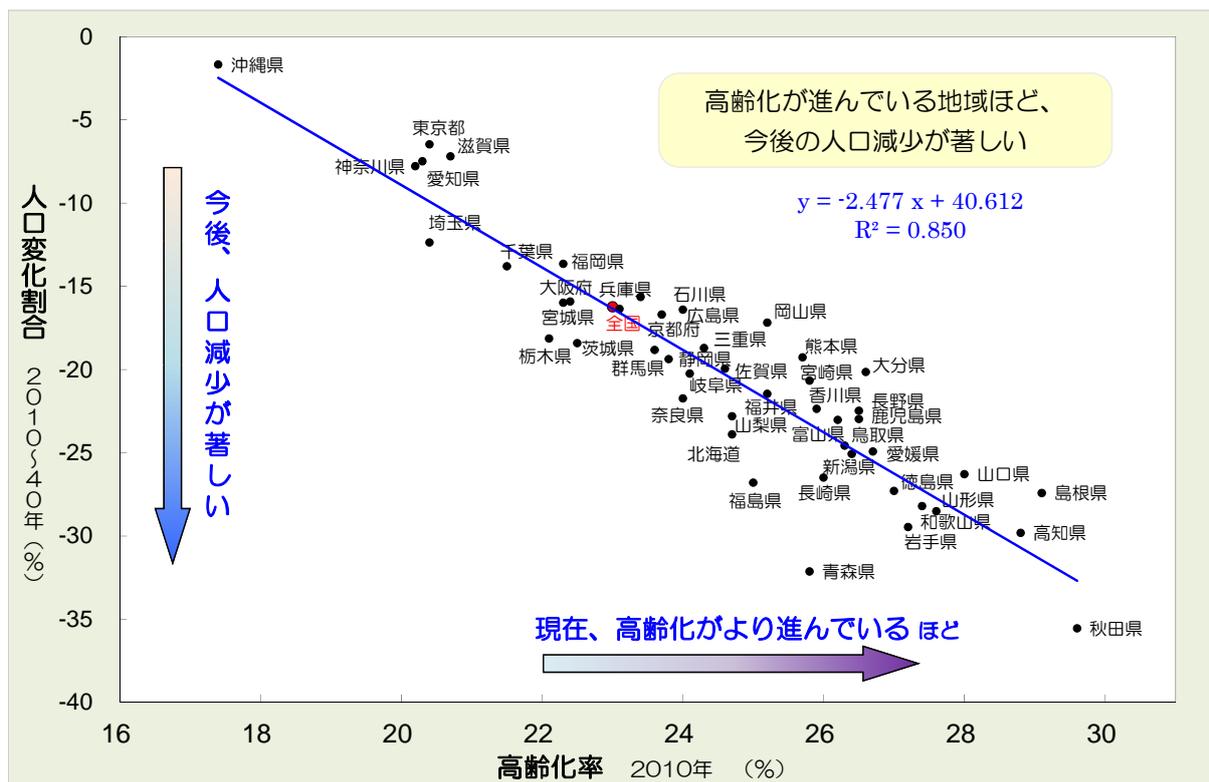


都道府県別にみた高齢化率の増加：2010, 25, 40年



17

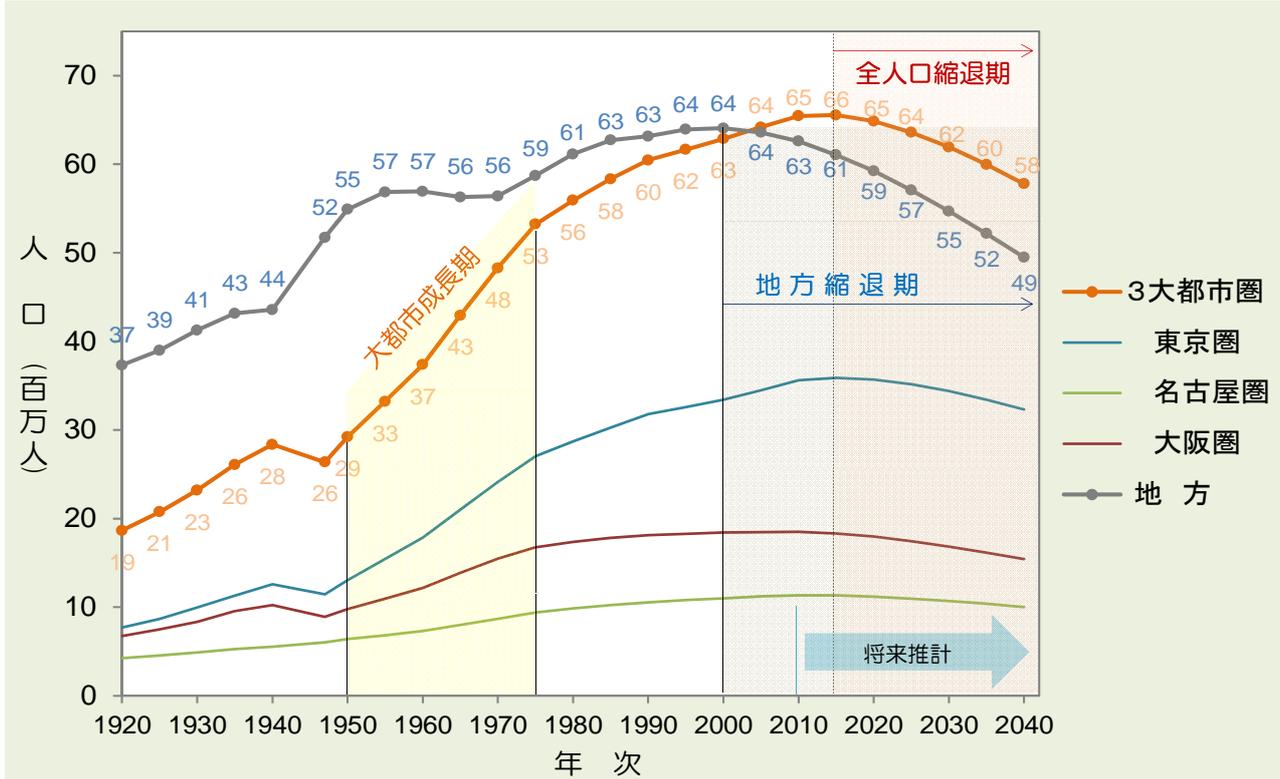
高齢化が進んでいる地域では、今後の人口の減少が著しい



資料：総務省統計局「国勢調査」、国立社会保障・人口問題研究所「日本の地域別将来推計人口（平成25年3月推計）」

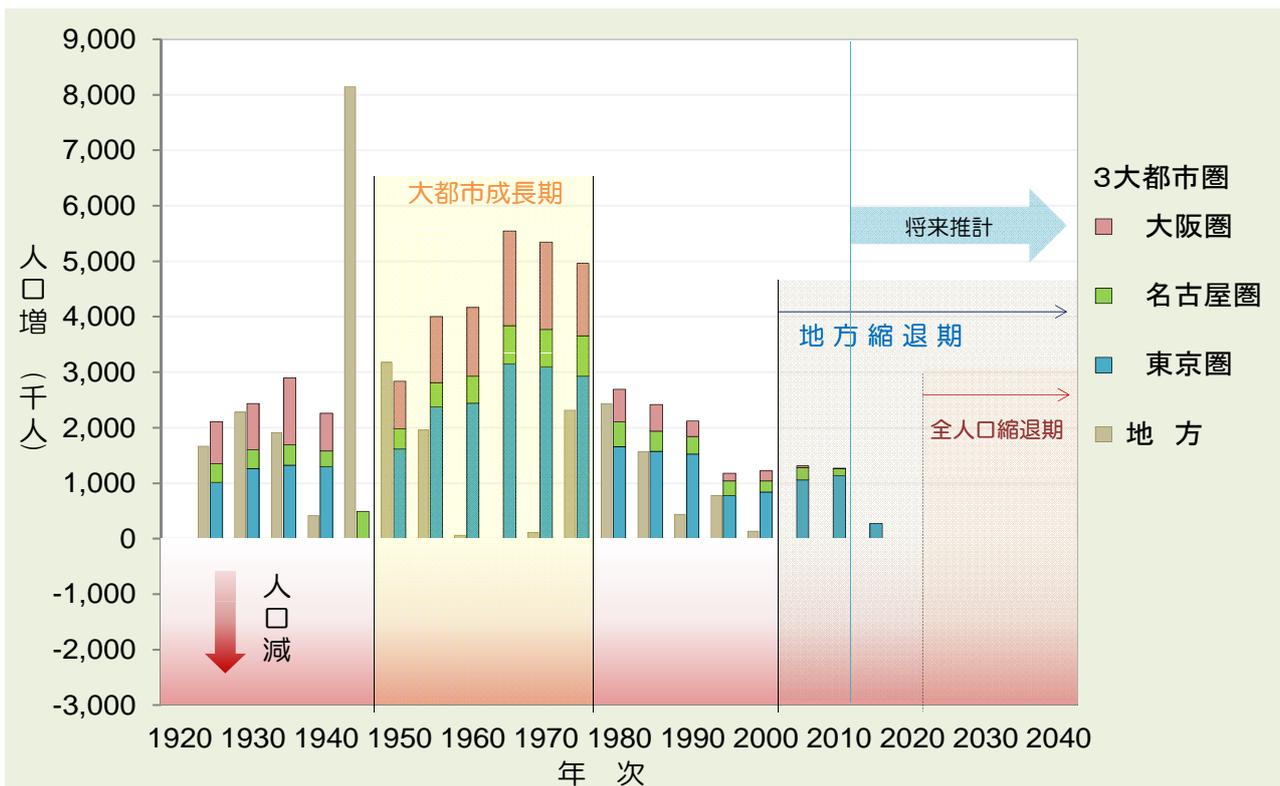
18

一極集中？ 都市圏と地方—人口の推移



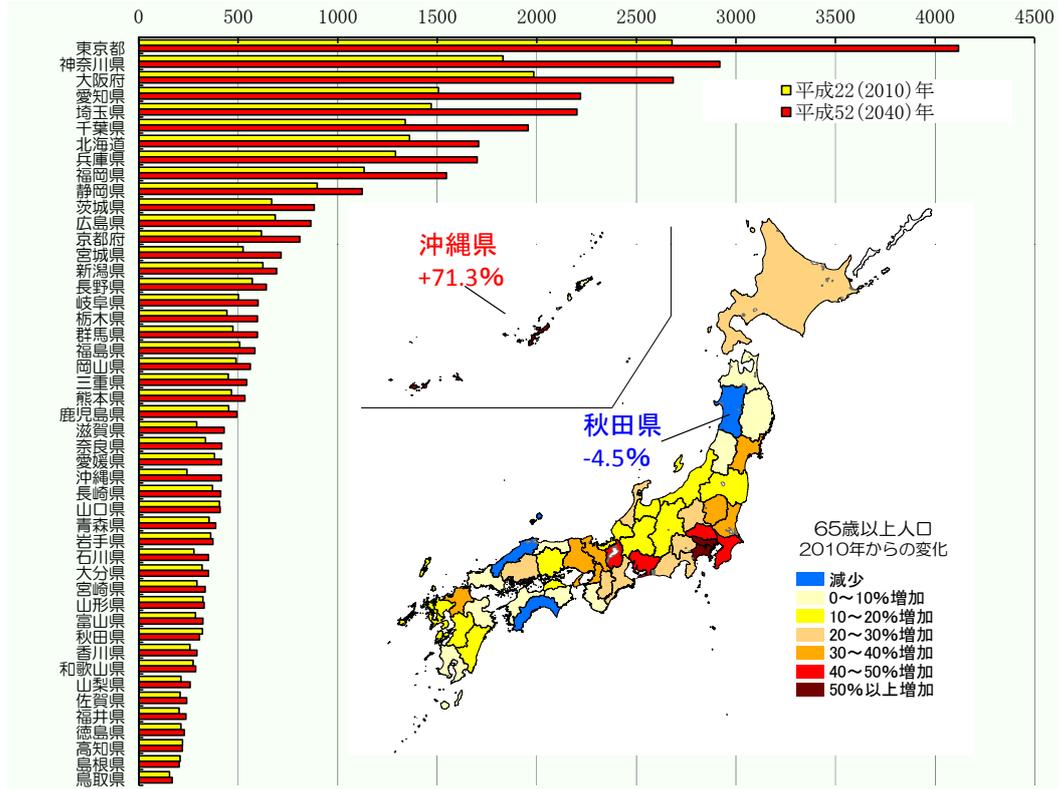
資料：2010年以前：総務省統計局「国勢調査」、2015年以降：国立社会保障・人口問題研究所「日本の地域別将来推計人口（平成25年3月推計）」

都市圏と地方—人口増減の推移（5年間隔）



資料：2010年以前：総務省統計局「国勢調査」、2015年以降：国立社会保障・人口問題研究所「日本の地域別将来推計人口（平成25年3月推計）」

都道府県別,65歳以上人口の変化：2010年、2040年比較



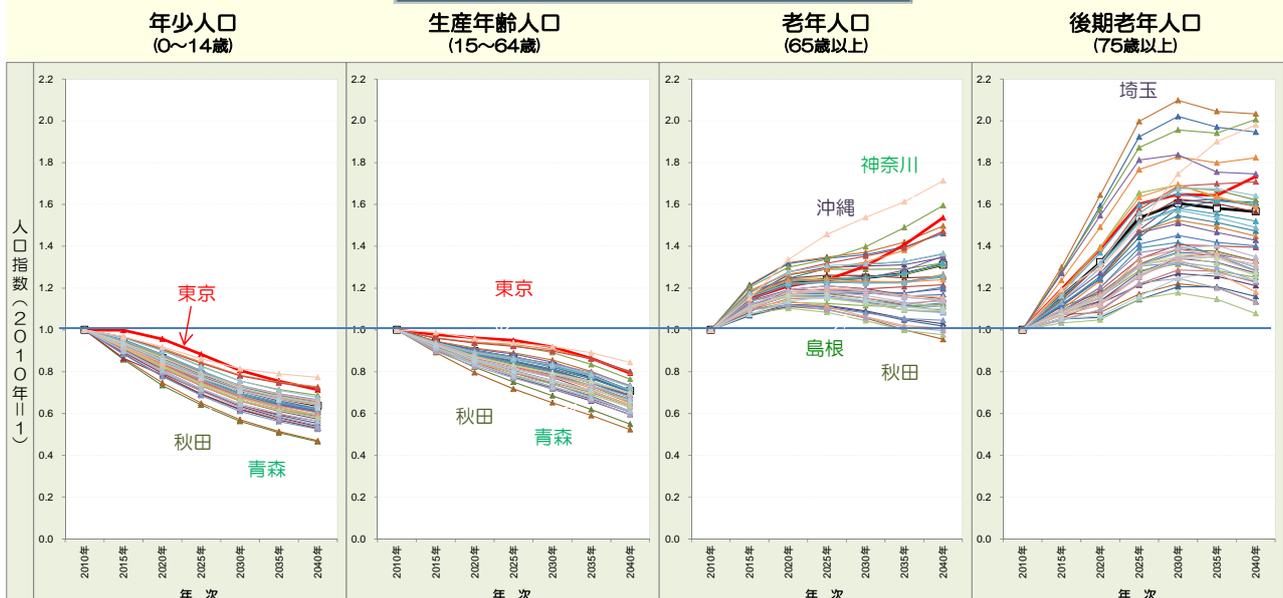
資料：総務省統計局「国勢調査」、国立社会保障・人口問題研究所「日本の地域別将来推計人口（平成25年3月推計）」

年齢階層による人口増減の違い：都道府県別

- ★ わが国の人口は、今後しばらく年齢階層（年少、生産年齢、高齢）ごとに大きく異なる推移をする。
→ 年齢が若いほど減少が著しく、高齢になるほど増加が著しい。
- ★ 相対的变化には、地域（都道府県）による違いがあるが、その違いは高齢ほど著しい。

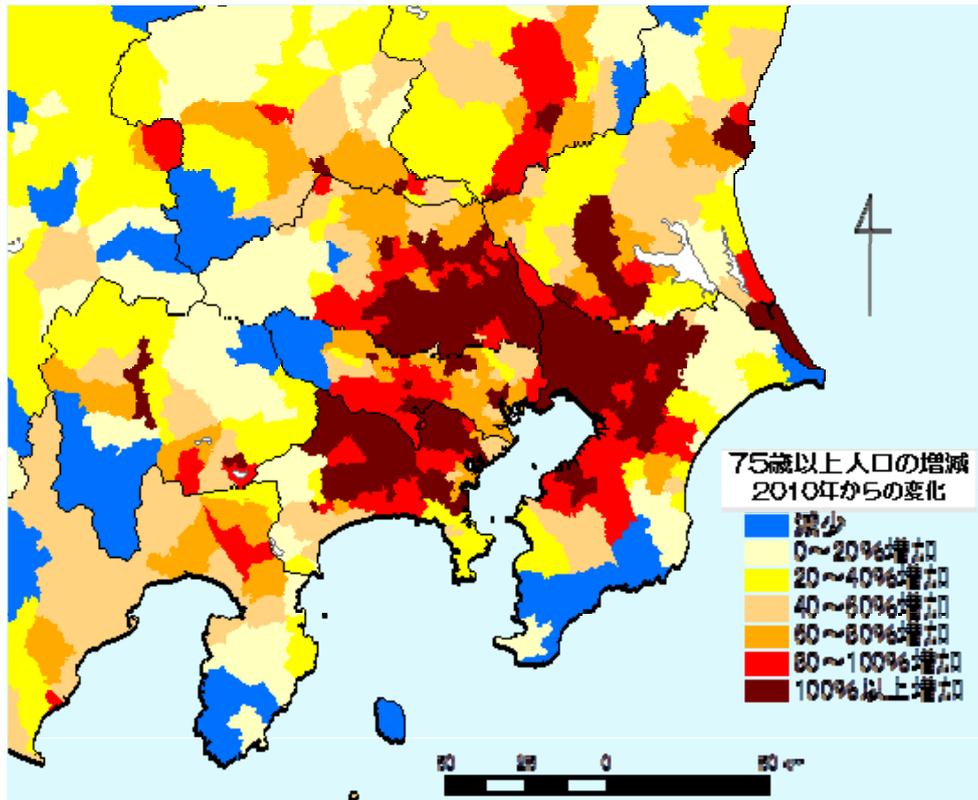
都道府県別に見た年齢階層別人口推移

(2010年人口=1.0)



資料：国立社会保障・人口問題研究所「日本の地域別将来推計人口（平成25年3月推計）」。

首都圏の高齢化：2010→2040年 75歳以上 人口増減率



資料：総務省統計局「国勢調査」、国立社会保障・人口問題研究所「日本の地域別将来推計人口（平成25年3月推計）」 25

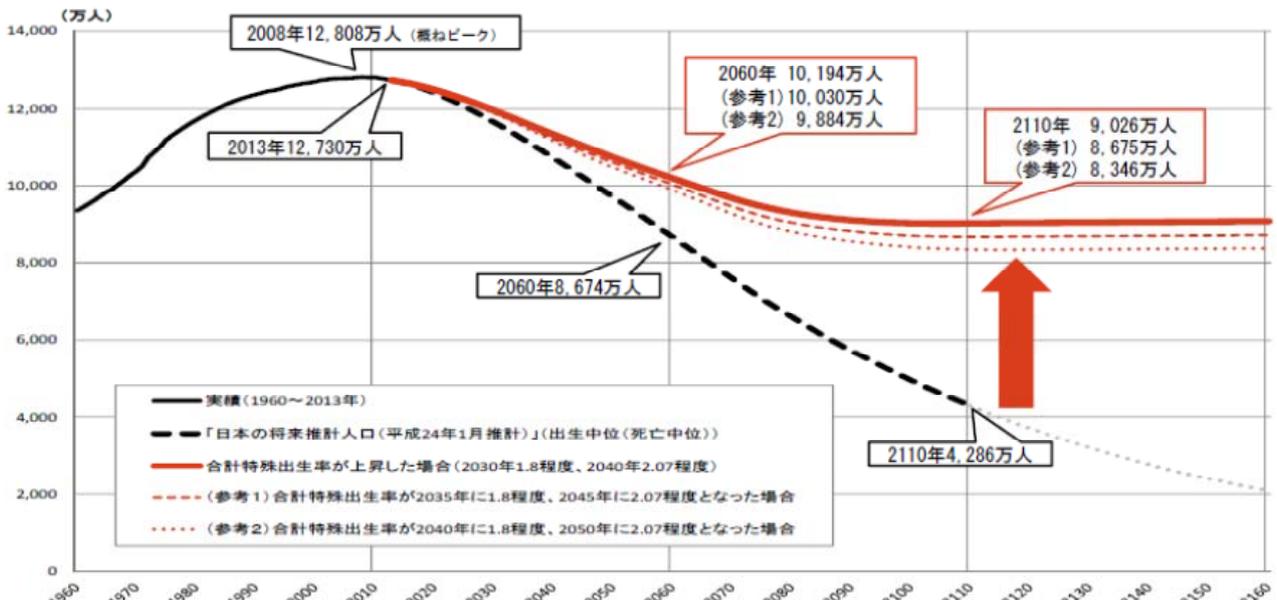
一億人？

図1. 我が国の人口の推移と長期的な見通し

創生

によると、2060年の総人口は約8,700万人まで減少すると見通されている。

- 仮に、合計特殊出生率が2030年に1.8程度、2040年に2.07程度（2020年には1.6程度）まで上昇すると、2060年の人口は約1億200万人となり、長期的には9,000万人程度で概ね安定的に推移するものと推計される。
- なお、仮に、合計特殊出生率が1.8や2.07となる年次が5年ずつ遅くなると、将来の定常人口が概ね300万人程度少なくなると推計される。

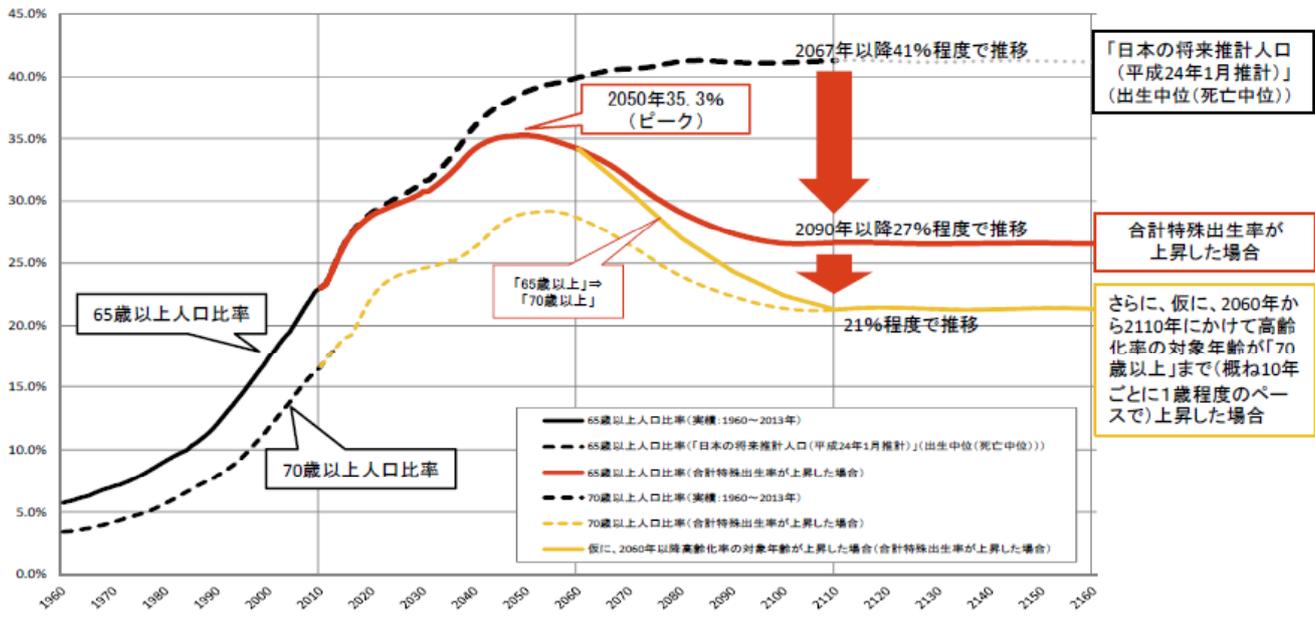


(注1)実績は、総務省統計局「国勢調査」等による(各年10月1日現在の人口)。国立社会保障・人口問題研究所「日本の将来推計人口(平成24年1月推計)」は出生中位(死亡中位)の仮定による。2110~2160年の点線は2110年までの仮定等をもとに、まち・ひと・しごと創生本部事務局において機械的に延長したものである。
 (注2)「合計特殊出生率が上昇した場合」は、経済財政諮問会議専門調査会「選択する未来」委員会における人口の将来推計を参考にしながら、合計特殊出生率が2030年に1.8程度、2040年に2.07程度(2020年には1.6程度)となった場合について、まち・ひと・しごと創生本部事務局において推計を行ったものである。

図2. 我が国の高齢化率の推移と長期的な見通し



- 「日本の将来推計人口（平成24年1月推計）」（出生中位（死亡中位））では、高齢化率（65歳以上人口比率）は、将来的に41%程度まで上昇すると見通されているが、仮に、出生率が上昇すれば、2050年の35.3%をピークに、長期的には、27%程度まで低下するものと推計される。
- さらに、将来的に健康寿命の延伸等に伴って高齢化率の対象年齢が「70歳以上」まで上昇するとすれば、高齢化率（70歳以上人口比率）は、概ね21%程度まで低下することとなる。



(注1) 実績は、総務省統計局「国勢調査結果」「人口推計」による。国立社会保障・人口問題研究所「日本の将来推計人口(平成24年1月推計)」は出生中位(死亡中位)の仮定による。2110～2160年の点線は2110年までの仮定等をもとに、まち・ひと・しごと創生本部事務局において機械的に延長したものである。
 (注2) 「合計特殊出生率が上昇した場合」は、経済財政諮問会議専門調査会「選択する未来」委員会における人口の将来推計を参考にしながら、合計特殊出生率が2030年に1.8程度、2040年に2.07程度(2020年には1.6程度)となった場合について、まち・ひと・しごと創生本部事務局において推計を行ったものである。

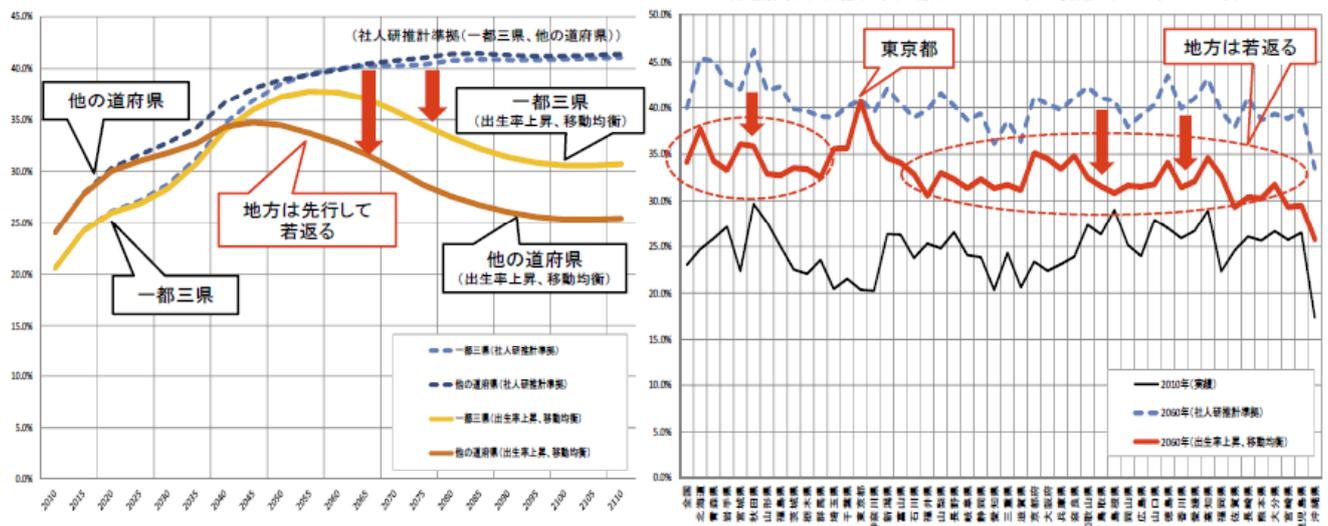
図3. 地域別の高齢化率の長期的な見通し



- 現状のまま推移したとすれば、一都三県においても、他の道府県においても、2070～80年頃以降、高齢化率は41%程度で推移するものと推計される。
- 仮に、2040年までに、現行程度の地域間の出生率格差を残しつつ全国の合計特殊出生率が2.07程度まで上昇し、移動が均衡した場合には、高齢化率は、他の道府県では、2045年頃の35%程度をピークに25～26%程度まで低下、一都三県では、2055年頃の38%程度をピークに30～31%程度まで低下すると推計される。

一都三県、他の道府県別 高齢化率(65歳以上人口比率)の見通し

都道府県別 高齢化率(65歳以上人口比率)の見通し(2010年→2060年)



(注1) 2010年(実績)は、総務省統計局「国勢調査結果」による。
 (注2) 「社人研推計準拠」は、国立社会保障・人口問題研究所「日本の地域別将来推計人口(平成25年3月推計)」の2040年までの傾向を延長して、まち・ひと・しごと創生本部事務局において推計したもの。性・年齢階級別人口が同研究所の「日本の将来推計人口(平成24年1月推計)」(出生中位(死亡中位))の値に一致するよう補正を行っている。
 (注3) 「出生率上昇、移動均衡」は、上記「日本の地域別将来推計人口」のデータを用いて、現行程度の地域間の出生率格差を残しつつ、全国の合計特殊出生率の水準が2030年に1.8程度、2040年に2.07程度と上昇し、かつ、2040年までに移動が均衡した場合(純移動率がゼロとなった場合)について、まち・ひと・しごと創生本部事務局において推計を行ったものである(全国の推計値で補正を行っている)。

「日本創成会議」（日本生産性本部）の提言（平成26年5月）



- 日本の希望出生率(1.8)を2025年に達成することを基本目標。
- 20歳代の結婚割合増加が、目標達成のための鍵。

基本目標＝国民の「希望出生率」の実現

◆現状(2013年)出生率＝1.43

↓

◆基本目標(2025年)
『希望出生率』＝1.8

●国民の「希望出生率」として出生率＝1.8を想定。

- ・夫婦の意向や独身者の結婚希望等から算出。
- ・最も出生率が高い沖縄県は出生率＝1.8～1.9
- ・OECD諸国の半数が出生率＝1.8を超えている。

↓

(参考)人口置換基準 出生率＝2.1

●将来人口が安定する「人口置換水準」は2.1

- ・日本の夫婦の理想平均子ども数は2.42人
- ・米、仏、英、スウェーデンの出生率は2前後

<出生率向上の要因>

1. 結婚割合の上昇

◎20歳代～30歳代前半に結婚・出産・子育てしやすい環境を作る

○出生率1.8

- ・20歳代後半の結婚割合(現在40%)が60%になれば実現可能

○出生率2.1

- ・20歳代前半の結婚割合(現在8%)が25%に、20歳代後半が60%になれば実現可能。

2. 夫婦の出生数増加

◎第2子、第3子以上の出産・子育てがしやすい環境を作る

具体的施策

若年世代の経済基盤確保

- 雇用・生活の安定化

結婚・妊娠・出産の支援

- 出会いと結婚の機会づくり

子育て支援

- 待機児童の解消

働き方改革

- 育児休業の拡充

多子世帯への支援

- 税・社会保障制度見直し

政策総点検

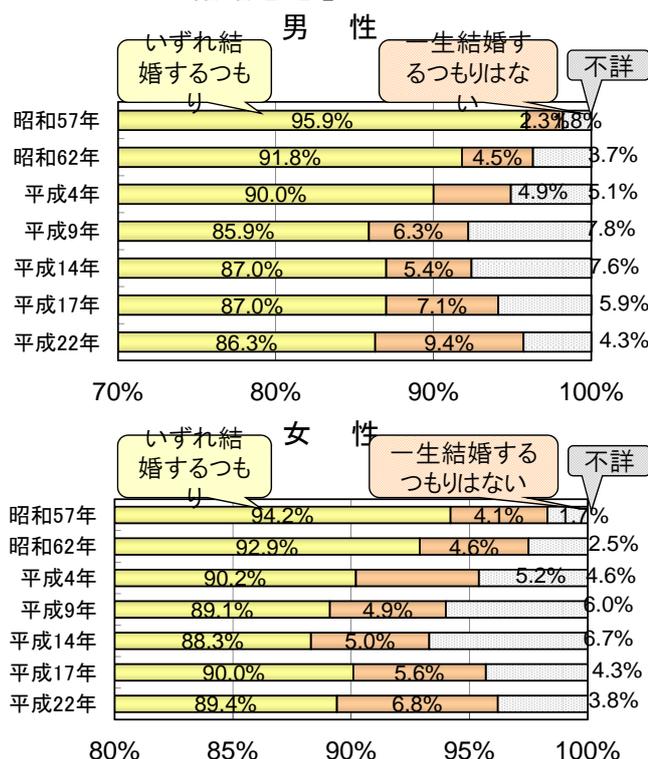
高齢者政策の見直し

国民の結婚や出産に対する希望

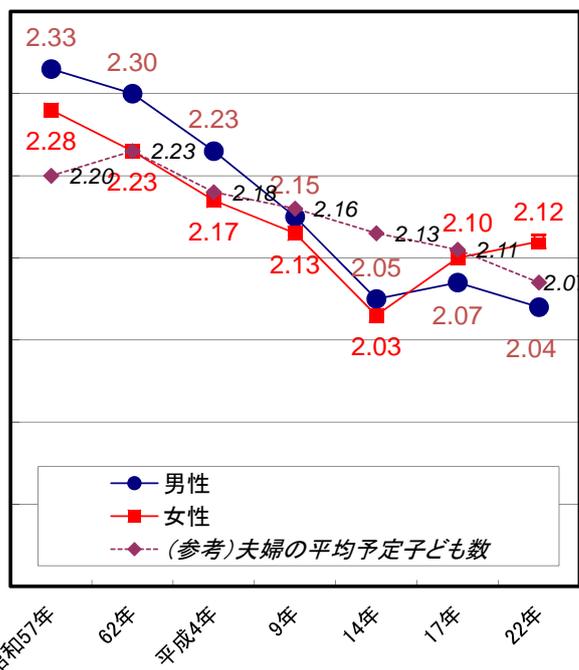


独身男女の約9割は結婚意思を持っており、希望子ども数も男女とも2人以上。

○「生涯の結婚意思」について



○「いずれ結婚するつもり」の未婚男女の希望子ども数

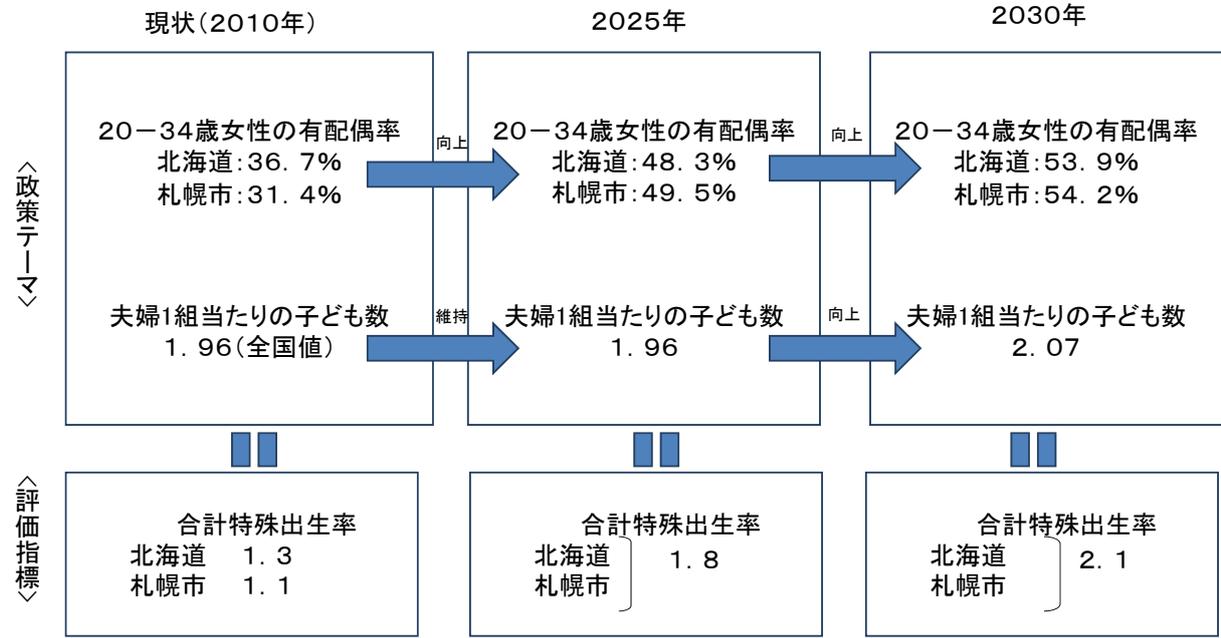


「地域人口ビジョン」の作成例（人口数値目標：北海道、札幌市）



- 20-34歳女性の有配偶率が10%ポイント~20%ポイント改善することで、出生率は希望出生率(1.8)に到達する見込み。
- さらに夫婦1組当たりの子ども数が上昇すれば、人口置換水準の出生率(2.1)まで出生率が改善することも可能。

人口数値目標(例)



少子化への対処

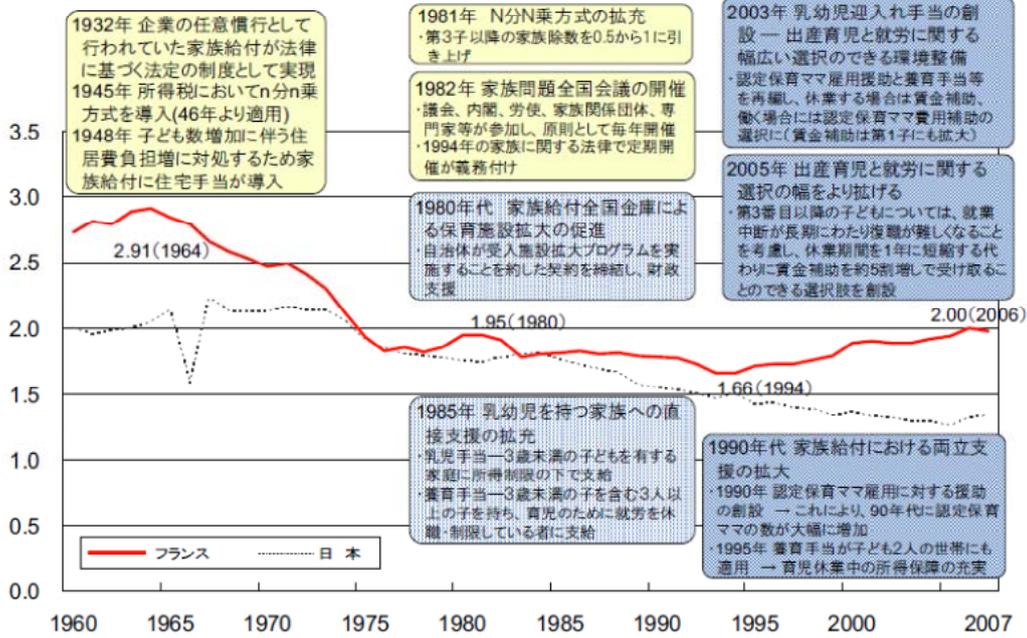
政策的対応

フランスの少子化対策



○ フランスでは、80年以上以前から少子化対策を実施しており、1980年代以降徐々に出生率が上昇。

フランスの出生率の推移と家族政策



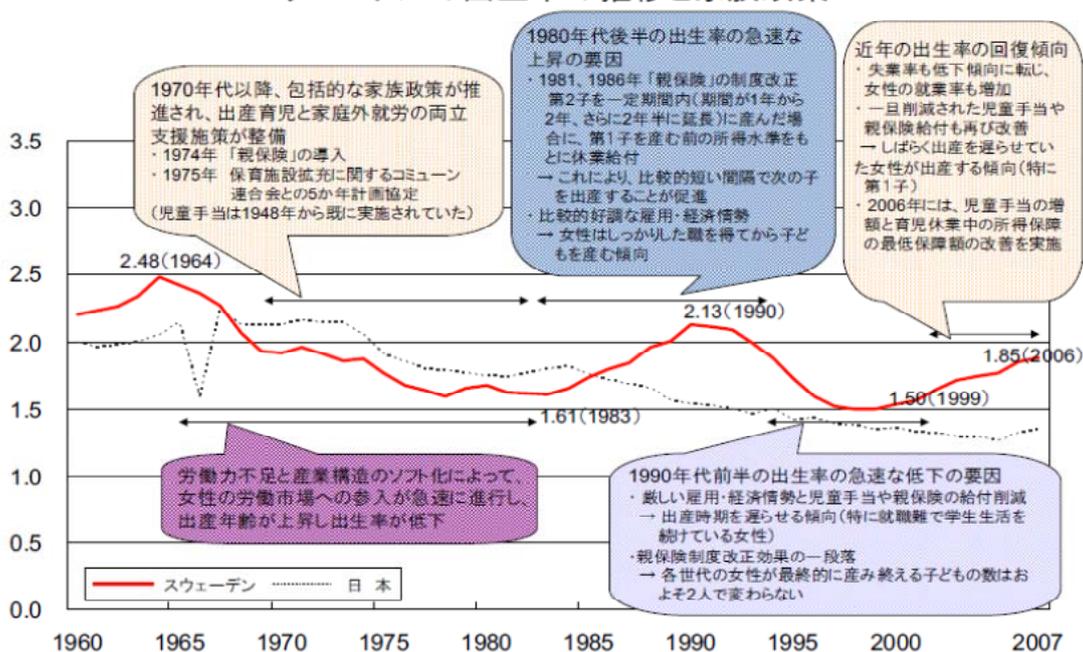
(出所) 明治大学・加藤久和教授資料

スウェーデンの少子化対策



○ スウェーデンは、直近50年で二度の出生率低下を経験するも、休業給付や児童手当をはじめとする政策対応により改善。

スウェーデンの出生率の推移と家族政策



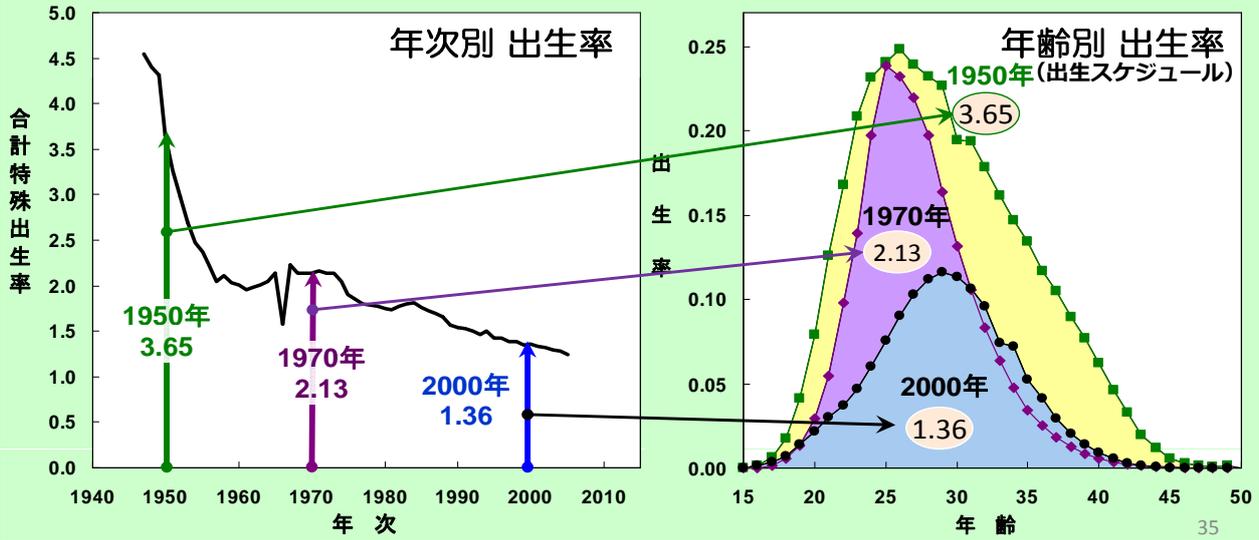
合計特殊出生率と年齢別出生率（出生スケジュール）

合計特殊出生率とは、

- ・ ある年次に観察された女性の **年齢別出生率** を、全年齢にわたって合計した数値。
- ・ それは、女性がその年齢別出生率にしたがって子どもを生んだ場合の **生涯の出生子ども数** と解釈できる。
- ・ 年齢別出生率は、出生スケジュールとも呼ばれ、(女性の)ライフコースを反映する。

下図右：1950年頃までの日本女性の50歳までの人生は、妊娠・出産・子育てに終始
 → 1970年頃までに30歳代、10代後半は妊娠・出産から解放 → その後、20歳代も妊娠・出産離れ = 少子化

年次別 合計特殊出生率 と 年齢別 出生率



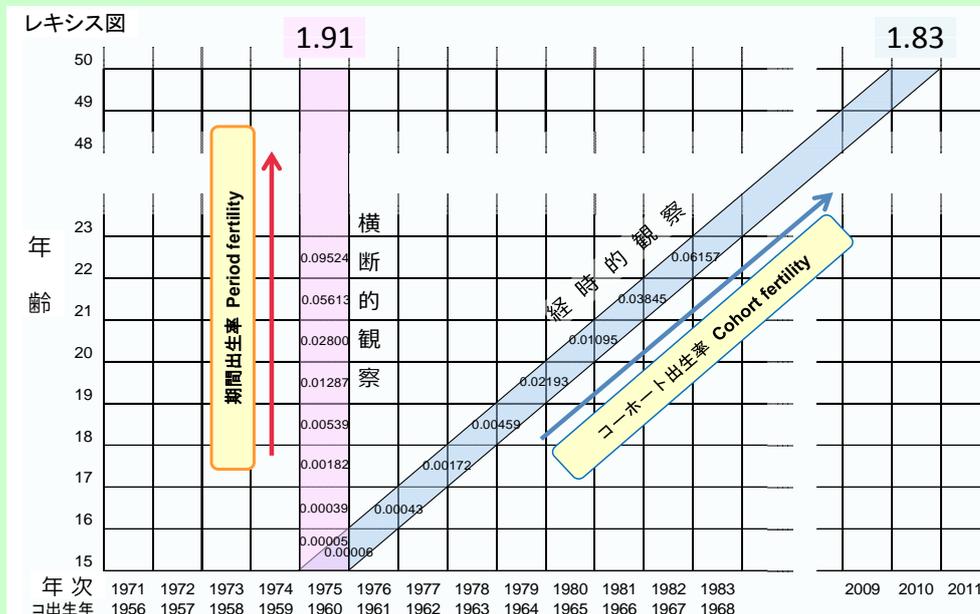
2種類の出生スケジュールと合計特殊出生率

- ・ 出生スケジュールおよび合計特殊出生率の観察軸には2種類ある。→ **年次(期間) 合計特殊出生率** VS **コーホート 合計特殊出生率**

年次別の
年齢別出生率

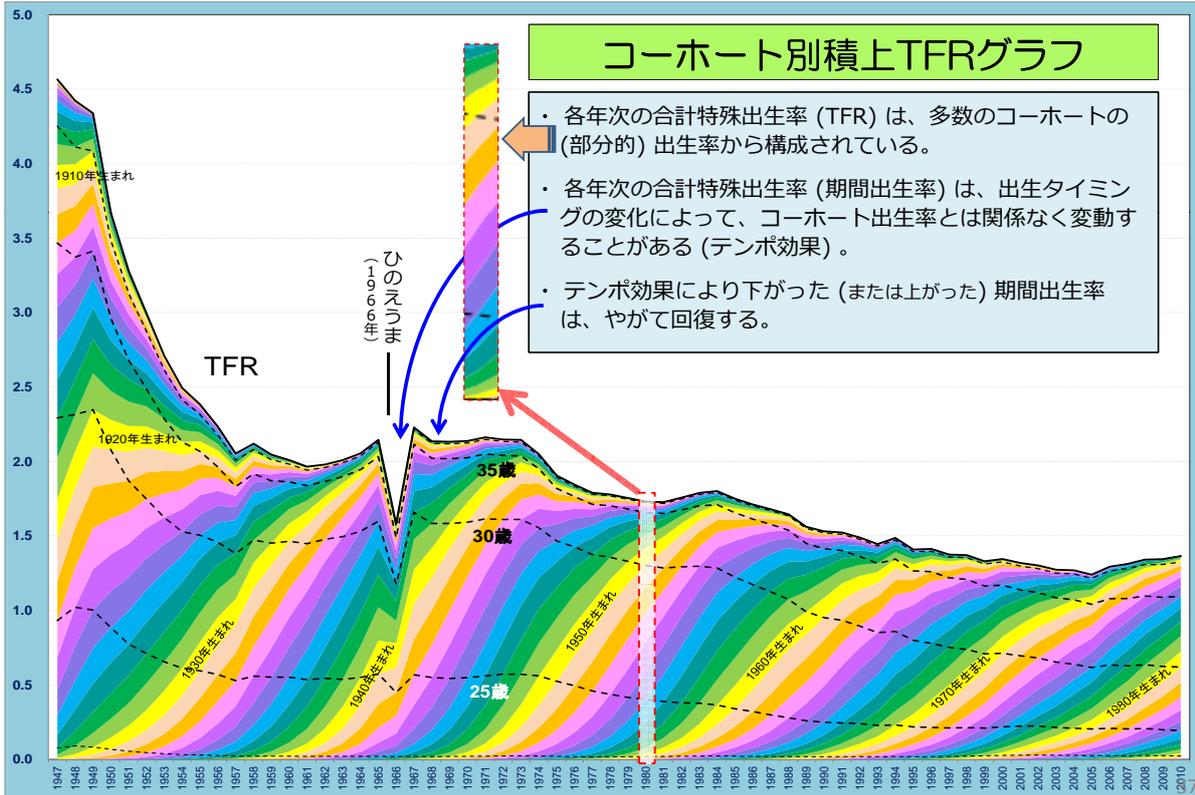
組み換え可

コーホート別の
年齢別出生率

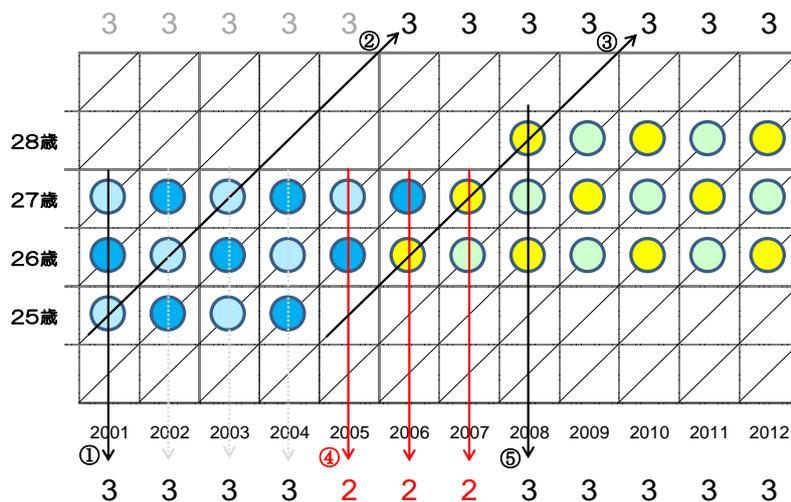


年次(期間)出生率の成り立ち

・各年次の合計特殊出生率は、必ずしもライフコース的整合性のない多数のコーホートの出生率から構成されている。



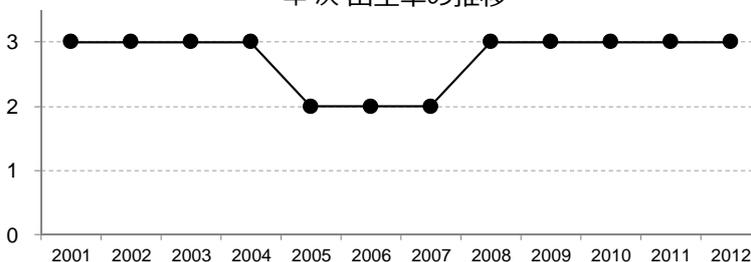
“晩産化”の年次(期間)出生率に対する効果



“晩産化”の実験

- ① 年次 2001年の出生数は 3人
- ② コーホート[2001年25歳]の生涯の出生数は 3人
- ※この例ではすべてのコーホートの生涯出生数は 3人
- ③ コーホート[2005年25歳]が 1年、晩産化 (しかし生涯の出生数は 3人)
- ④ すると 年次 2005年の出生数は 2人に減少!
- ⑤ コーホートの晩産化が止まるとピリオドの出生数が 3人に回復。

年次出生率の推移

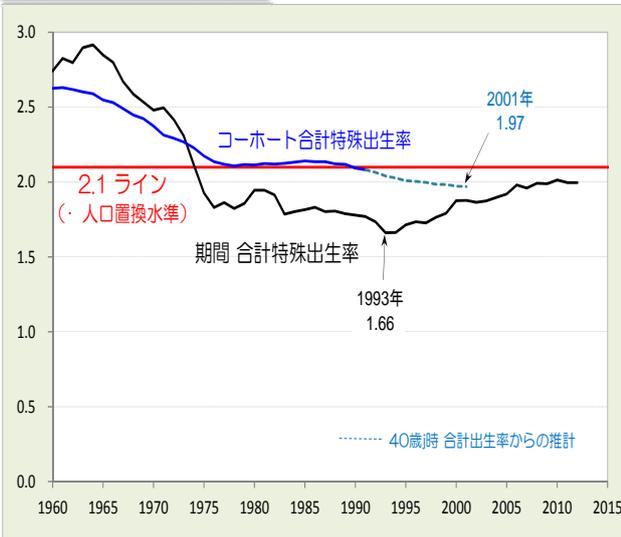


コーホート出生率が安定していても、「晩産化」が起こると年次の出生率は低下する。

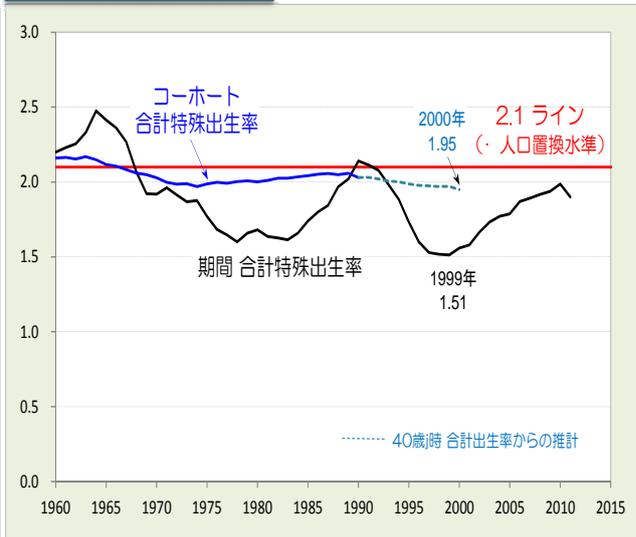
年次的に観察される出生率の注意点

フランスやスウェーデンは、施策の充実等によって少子化の状態（人口置換水準下の出生率）から脱し、出生率を回復させた国の例として取り上げられることがある。しかし、これらの国でみられた年次的な出生率（期間合計特殊出生率）の低下は、主に出生年齢の遅延にともなう効果（テンポ効果）によるもので、その背景にある実質的な出生力（コーホート合計特殊出生率 ≒ 生涯の平均出生子ども数）はもともと人口置換水準付近で推移していた。したがって、近年の出生率回復はこのテンポ効果が弱まったことが原因であり、人々の持つ実質的な子ども数が増加したわけではない。

フランス



スウェーデン



資料：The Human Fertility Database (<http://www.humanfertility.org>), 取得年月日：2014/12/17 より作成。

39

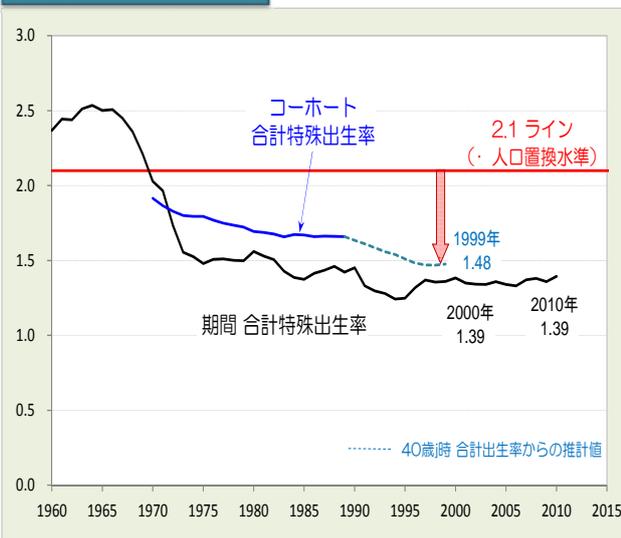
年次的に観察される出生率の注意点

これに対して、ドイツ や 日本の出生率低下（下図）は、かなりの部分が実質的な出生力（コーホート合計特殊出生率）の低下に起因しており、今後、少子化が解消するためには（出生率が人口置換水準へ復帰するためには）、フランスやスウェーデンとは異なる特段の努力が必要となる。

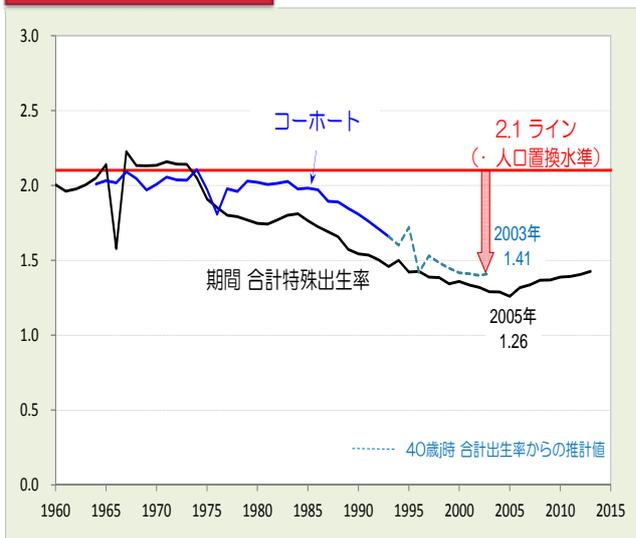
★ 少子化を正しく捉えるには、指標の背後にある“実質”を読み取る必要がある。

→ 少子化に正しく対処するには、指標の示す出生の“量”ではなく、“質”に着目する必要がある。

ドイツ



日本



資料：The Human Fertility Database (<http://www.humanfertility.org>), 取得年月日：2014/12/17 より作成。

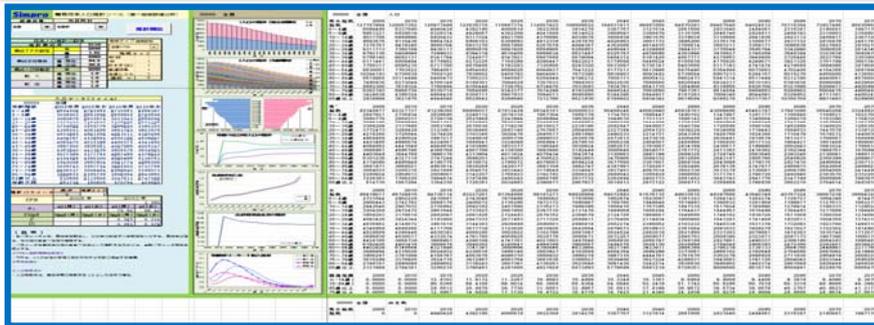
40

地方創生：統計データツールの開発・利用

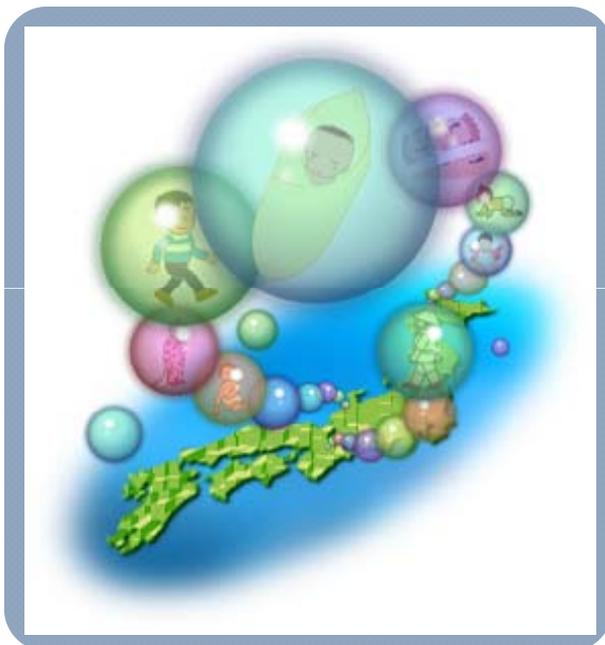
地域経済分析システム：RESAS



小地域将来人口・世帯数推計システム：Simpro (仮)



41



地方創生と人口統計

— まち・ひと・しごとと再生の課題 —

科研「経済統計・政府統計の理論と応用」

2016年1月29日(金)

東京大学経済学部小島ホール

国立社会保障・人口問題研究所

金子隆一

42

少子化指標としての出生率について

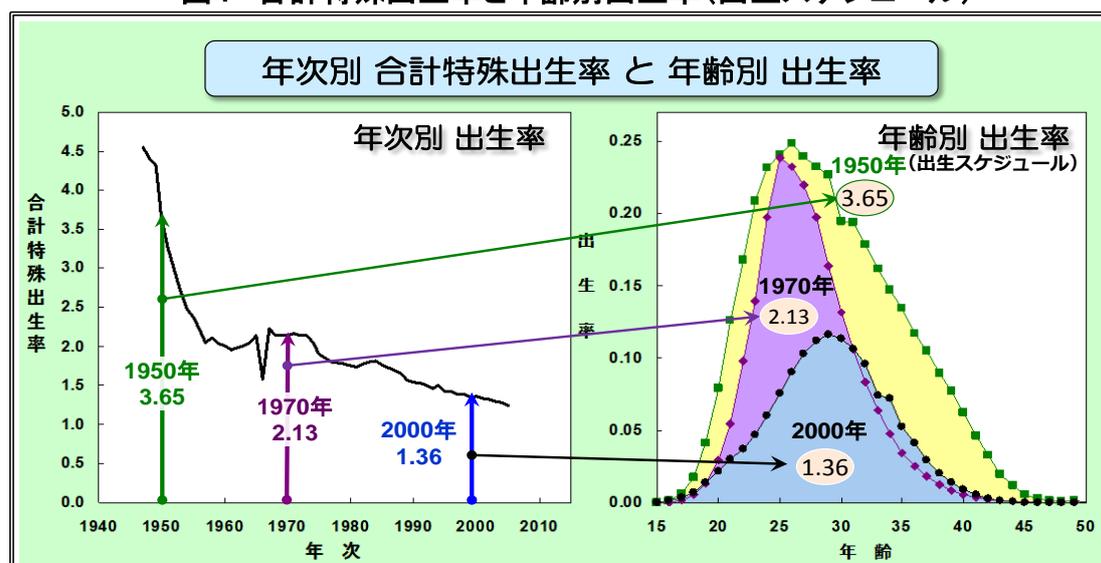
金子隆一（国立社会保障・人口問題研究所）

少子化対策を考える際には、統計指標によって少子化の現状を正しく把握することが必要となるが、その際に指標の性質に関する十分な理解を欠くと、誤った現状認識に陥ることがある。出生率（合計特殊出生率）は、少子化を論ずる上で最も基本的な指標と考えられるが、実はその見方についてはいくつか注意すべき点が存在する。すでに広く使われている指標ではあるが、ここでは一旦基礎に立ち戻って、この出生率の成り立ちや性質について解説し、その正しい見方や少子化対策における使い方について考えて見たい¹。

1. 出生率と出生スケジュール

図1の左には、一般によく見られる合計特殊出生率の年次推移を示した。この年々の出生率の元となるのは、年齢別出生率であり、これをグラフとして描いたものを同図右に示した。横軸は女性の年齢、縦軸は各年齢における出生の頻度（すなわち年齢別出生率）であり、女性の再生産年齢とされる15歳から50歳までの変化を描いている。これは「出生スケジュール」とも呼ばれる。この出生スケジュールを観察することで、女性が人生の中でどのように出産しているかがある程度見えてくる。

図1 合計特殊出生率と年齢別出生率(出生スケジュール)



資料：厚生労働省「人口動態統計」による。

¹ 本稿は、人口統計の専門家でない一般の読者を想定したものであり、一部の説明において専門的厳密性よりも、わかりやすさを優先した表現を用いている。

同図には、3つの時期のスケジュールを示しているが、一番後ろに見える黄色のグラフは1950年の女性の出生スケジュールである。それは10代の終わりに急速に立ち上がり、40歳代前半に至るまで高い水準を示しており、当時広い年齢範囲で出生があったことを示している。全体のレベルを表す合計特殊出生率は3.65と、戦前の水準(4~5)よりは低いものの、この頃までの日本女性では、成長直後の10代終わりから50歳頃までの人生はほとんど妊娠・出産・子育てと共にあったことが読み取れる。

これに対して、1970年のグラフ(紫色)では、10代ならびに30代以降の出生がすっかり消失しており、女性がこれらの年代において、妊娠・出産から解放されたことがわかる。これは当時、女性に高学歴化と社会進出が始まったことを反映しており、またその後の進展の前提ともなっている。なお、出生スケジュールの曲線が囲む面積は、合計特殊出生率に相当するため、このような年齢別出生率のグラフによって、各時期の出生水準を視覚的に比較することができる。左図に見た出生率の年次推移は、この面積の時系列変化を示したものに他ならない。

続けて2000年の出生スケジュールのグラフを見ると、20代で劇的に「出産離れ」が起きており、これによって全体の面積が大幅に削り取られ、少子化が進んだ姿を顕している。このように、出生スケジュールを見ることで、単一の数値である合計特殊出生率の背後にある女性のライフコースの姿を垣間見ることができる。

2. 2つの合計特殊出生率：年次観察とコーホート観察

人口学では、官庁統計として年次ごとに把握されるマクロの人口統計(国勢調査結果や人口動態統計報告など)から、人々のライフコースを再構築して分析することが行われるが、その際に用いられるツールの一つとして「レキシス図」と呼ばれるものがある。それは図2に示したように、横軸に年次、縦軸に年齢を配置し、格子を描いた座標平面である。

この平面上で、たとえば1960年に生まれて75年に15歳を迎えた世代のライフコースは、ブルーで示した斜めの帯として描かれる。右方向に1年進めば、個人の年齢は上方に1年上がることになるので、彼らの人生は右斜め45度上方に向けて進む。

図では、その途中で年齢別出生率を数値として示してあるが、これを斜めに足し合わせた値1.83がこのコーホートの合計特殊出生率である²。これはこのコーホートの生涯の平均出生子ども数に相当し、かれらの出生の量的水準を表す指標となる³。したがって、世代ごとに少子化がどのように進んでいるかなどを測るには適した指標となる。

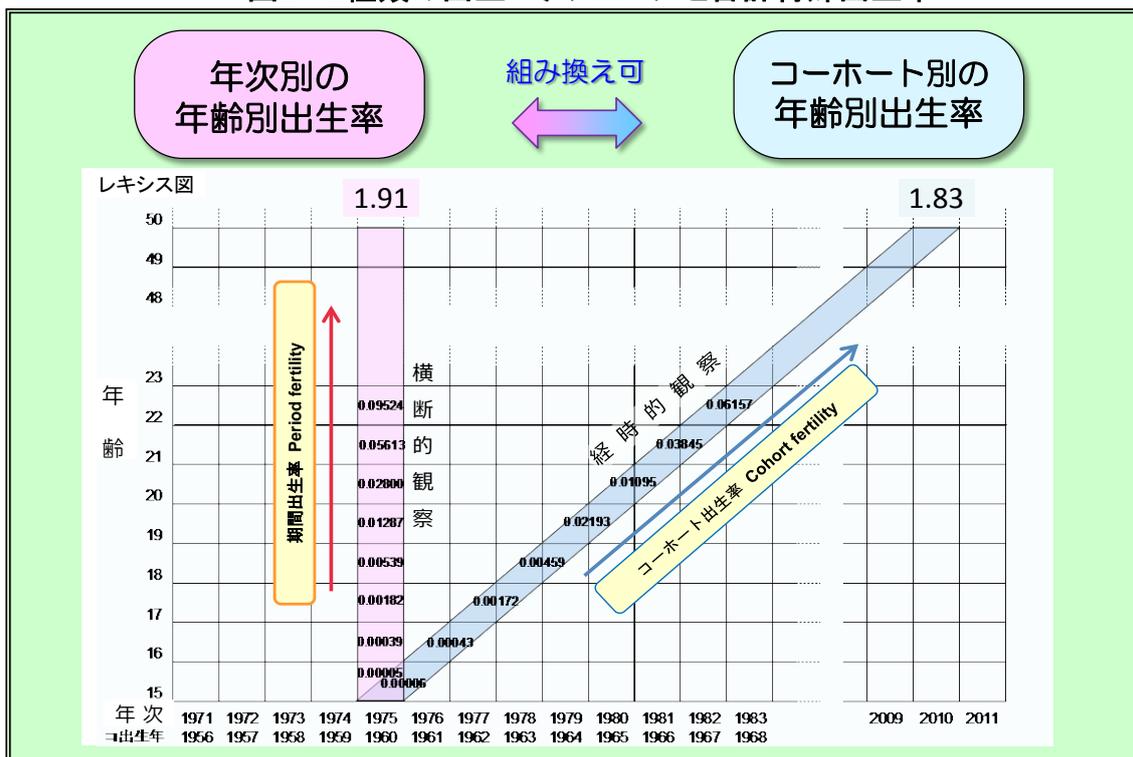
しかしながら、この指標は生涯の出生を終える年齢(便宜的に50歳と定義)に達した世代についてしか計算できないため、現在の状況を知るためには少なくとも出生の主力となっている25~35歳世代が50歳に達する25年後まで待たなければならない。

² コーホートとは、ここでは同じ年に生まれた人々の集合を指す。「世代」とほぼ同義。

³ 厳密には、合計特殊出生率は、「再生産期間を通して女性集団に死亡が発生せず、所与の年齢別出生率に従って子どもを生んだ場合の生涯の平均出生子ども数」を与える。

これに対して、たとえば去年とか今年などのように、ある1年次の出生率を知るためには、**図2**においてピンクの帯で示されるような横断的観察を行うことが考えられる。15歳から縦上方向に年齢別出生率を足して行けば、この年次の出生スケジュールにしたがって子どもを生んだ場合に期待される生涯の平均子ども数の値が算出できる。実は、これが毎年報道され、一般に用いられている「合計特殊出生率」であり、図の例では、「1975年の合計特殊出生率は、1.91である」と表現されることとなる。

図2 2種類の出生スケジュールと合計特殊出生率



資料：数値（出生率）は厚生労働省「人口動態統計」による。

このように合計特殊出生率には、世代（コホート）のライフコースに沿って算出される「コホート合計特殊出生率」と、対象年次の年齢別出生率を足し上げた「年次の合計特殊出生率」の2種類が存在する⁴。前者は実在する世代の生涯（ブルーの帯）について算出した平均出生子ども数であるのに対して、後者は多数の世代の部分的出生率を合成して算出した、いわば「仮想的なライフコース」（ピンクの帯）の平均出生子ども数となる。

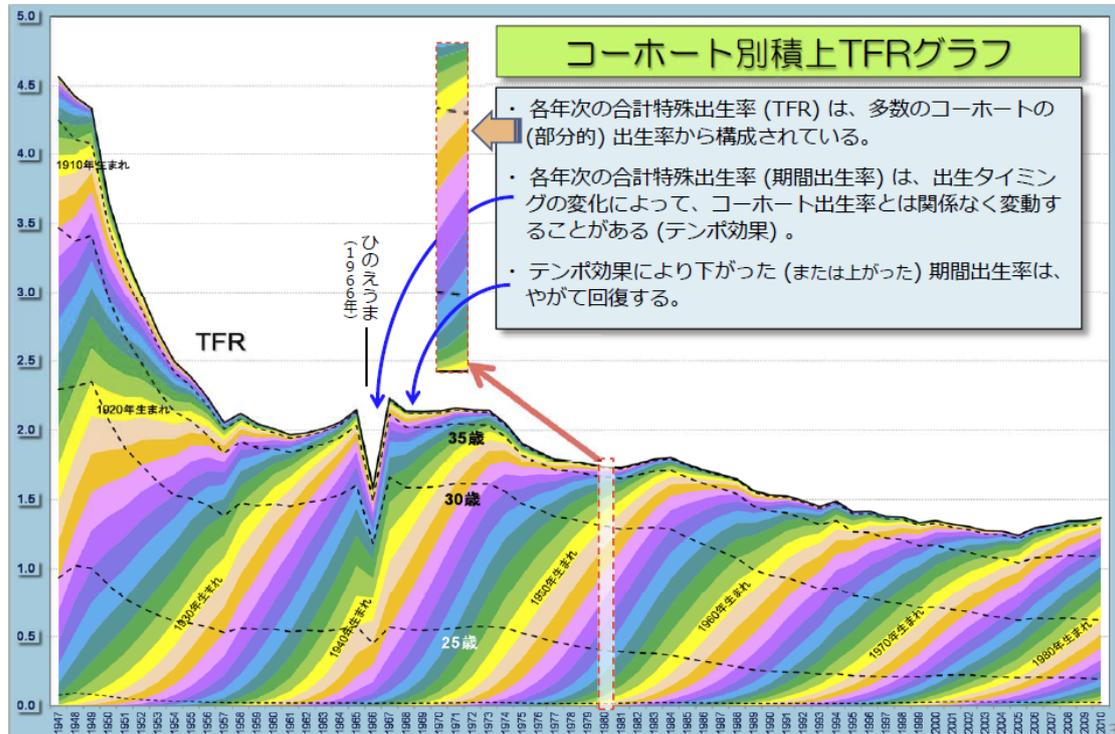
3. 合計特殊出生率の成り立ちと性質

図3には、我々がよく目にする「年次の合計特殊出生率」の推移を、その内訳となるコホートの年齢別出生率（出生スケジュール）で色分けして描いた。ある一年次を抜

⁴ ここで年次の合計特殊出生率と呼んでいる指標は、専門的には「期間合計特殊出生率」あるいは「ピリオド合計特殊出生率」と呼ばれている。

き出すと、それは多数のコーホートの部分的出生率によって構成されていることがわかる。

図3 2種類の出生スケジュールと合計特殊出生率



資料：厚生労働省「人口動態統計」による。

ここで問題となるのは、この抜き出した年次の年齢別出生率のセットは、一応すべての年齢の出生情報を含んでいるが、各年齢での出生行動は、実際のライフコースのように整合的であるとは限らないということである。たとえば、「晩産化」（本稿では、生涯に生む子ども数を変えずに出産年齢が遅くれることを意味するものとする。以下同じ。）が生じたとき、実際のライフコースでは 20 代など若年での出生率は低下し、その分を 30 代以降の高年齢で出生率が上昇することで取り戻すので、コーホート合計特殊出生率は変わらない⁵。ところが、年次の合計特殊出生率では、各年齢の出生率の間にそのように協調して動くしくみはないため、その時々的情勢に反応して、全年齢で一斉に上昇したり、低下したりすることの方が多くなる。

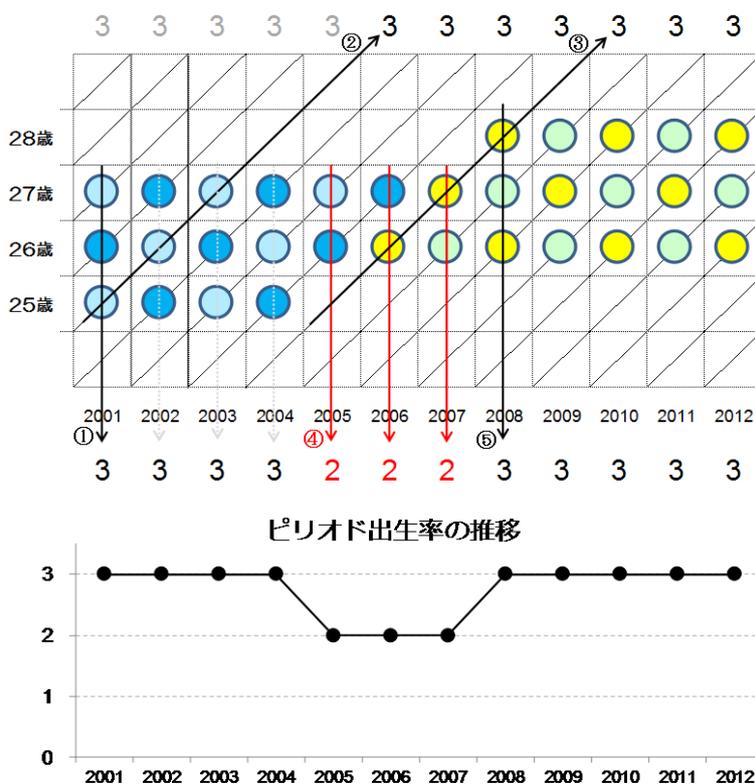
「ひのえうま」（1966 年）を例にすると分かり易い。この年に生まれた女兒は不幸になるとの迷信から、人々がこの年の出生を翌年以降に先延ばしした結果、年間の合計特殊出生率は 1.58 と例年の 3 / 4 程度にまで低下した。しかし、この現象によって日本人女性が生涯にもつ平均の子ども数が 3 / 4 に縮小したわけではない。実際は、「ひのえうま」の年に子どもを生んでいたすべての世代のコーホート合計特殊出生率（平均

⁵ 現実には出産年齢が高まると妊孕力（にんようりょく）が低下するため、高年齢での出生の取り戻しは完全ではなく、コーホート合計特殊出生率は幾分減ることが一般的であるが、ここでは簡単のため出産年齢の遅れだけのケースで説明する。

出生子ども数) は、前後の世代同様にほぼ2人となっていた。

この例は、人々が実際に持つ子ども数は変わらないのに、年次の合計特殊出生率は、見かけ上低下し得ることを示しており、この場合には「一人の女性が生涯に生む平均の子ども数」という合計特殊出生率の解釈が成り立たないことを意味する。「ひのえうま」の例では、このことはむしろ明白であろうが、日本の少子化過程で見られるような長期にわたる晩産化においても、程度の差こそ違え、同様のメカニズムが働いていたことはあまり知られていない。分かり易く表現するならば、日本の少子化過程(合計特殊出生率低下過程)においては「ミニひのえうま」が毎年生じており、人々が実際に生涯に生む子ども数と年次合計特殊出生率の値との間には齟齬が生じていたのである⁶。

図4 “晩産化”による年次の合計特殊出生率低下の実験



このことを簡単な実験により説明しよう。図4には、女性が25歳から27歳まで毎年一人ずつ出生するような仮定の集団を考え、ある世代以降、出生スケジュールが1年「晩産化」する例を示した。この例ではスケジュールが遅れるだけなので、すべての世代の最終的な子ども数は3人である(図上部の数値)。しかし、晩産化が右斜め上方への移動であることから、縦方向の合計値には一時的に「ずれ」が生ずる。この結果、晩産化過程の3年間については、合計特殊出生率が2人に低下することが示されている(図下部の数値とグラフ)。そして、この過渡期の直後、世代の出生スケジュールには

⁶ 後述するように、日本の少子化においては「ミニひのえうま」効果とともに、実際の出生子ども数の低下が同時に進行しており、年次の合計特殊出生率の低下はこれらが複合したものであった。

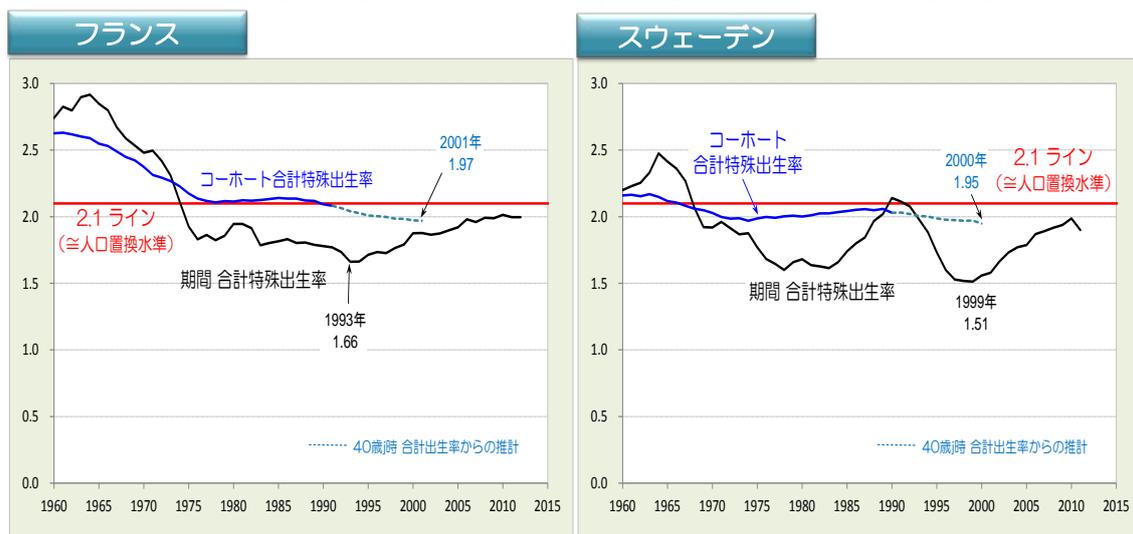
もはや変化がないにも関わらず、年次合計特殊出生率は自動的に3人に回復している⁷。

さて、以上の説明で「コーホート出生率が安定していても、晩産化が起これると、年次の出生率は低下する」という性質が確認できた。とすると、年次の合計特殊出生率は、その額面通りの数値を、少子化の動きとして理解することが適当でない場合があることになる。以下にその様な実例を取り上げて、問題点を検討してみよう。

4. 少子化指標としての合計特殊出生率の問題点

図5には、フランスとスウェーデンについて、それぞれ2種の合計特殊出生率の推移を描いた。年次の合計特殊出生率の変化は黒の曲線で描かれている。これらを見ると、いずれの国でも1970年代後半以降において年次出生率が大きく低下し、人口置換水準（赤の線）から離れた時期があることがわかる。スウェーデンにおいては、90年前後の回復を挟んで、2回の大きな低下を経験した（この著しい変動は「ジェットコースター出生率」と呼ばれたことがある）。

図5 フランス、スウェーデンの年次出生率とコーホート出生率



注：コーホート合計特殊出生率（青の曲線）は、各コーホートが29歳であった年次にプロットしている。
資料：The Human Fertility Database (<http://www.humanfertility.org>), 2014/12/17 データ取得。

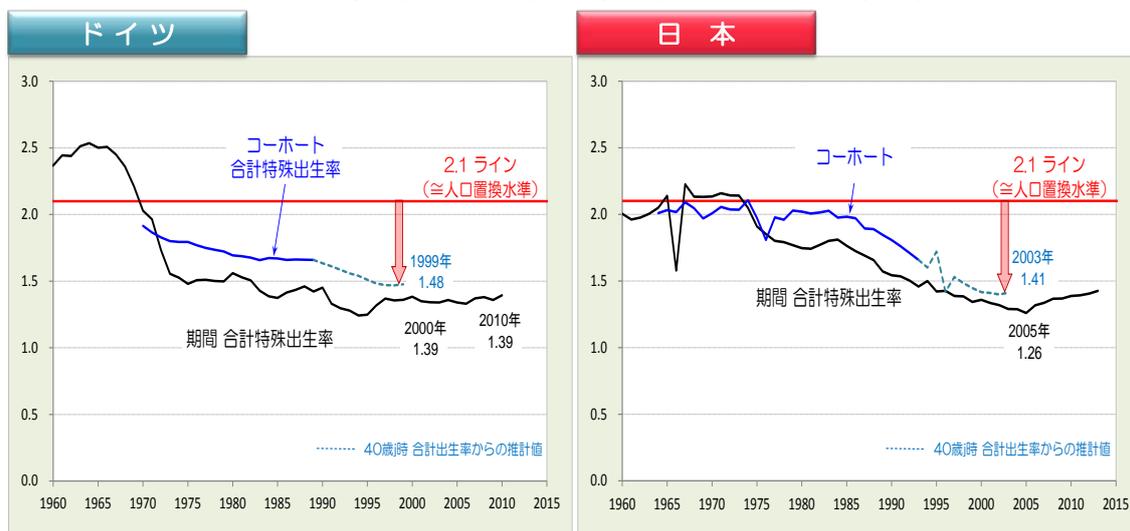
ただ、どちらの国においても90年代後半（あるいは末）以降は、年次出生率が回復基調にあり、人口置換水準付近にまで上昇していることが見られる。これらの出生率の回復は、低出生に喘ぐ他の国々からみると目覚ましいものであり、また、両国とも手厚い家族政策が実施されていることで知られているため、しばしば政策効果によって「少子化を克服」した例として紹介をされている。

しかしながら、同図に青色で示したコーホート出生率の推移を見ると、実はこの2ヶ国とも、ずっと人口置換水準付近で推移しており、これは年次出生率に大きな低下が

⁷ ちなみに年次出生率が低下した期間に失われた出生は、晩産化が終了して出生率が回復しても、もはや戻っては来ない。不思議に思うかもしれないが、それは晩産化による世代間隔の延長の代償として消費された形となる。

見られた時期においても、実質的な出生力は低下していなかったことを示している。すなわち、年次出生率の一時期の低下は、先に紹介した晩産化による見かけ上の現象（ミニひのえうま効果）であったことになる⁸。つまり現時点で振り返ってみると、当該時期の出生率の回復は、この晩産化が緩んだことによるミニひのえうま効果の解消が主な理由であって、必ずしも人々の持つ子ども数が増えたことを示すものではない。要するに、両国のこの時期の年次出生率の回復は、私たちが単純に考える「少子化を克服」したという図式、すなわち一時期人々が生涯に持つ子どもの数を減らしたが、これに即応した施策等によってその後の世代では子ども数が復活してきたという物語とは、どうも違うようなのである⁹。だから、他の少子化国でこの両国の経験から学ぼうとするなら、最近の年次出生率を回復させたという事例ばかりに注目するのではなく、コーホート出生率を長期に人口置換水準に安定させている理由にこそ注目しなくてはならない¹⁰。

図6 ドイツ、日本の年次出生率とコーホート出生率



注：コーホート合計特殊出生率（青の曲線）は、各コーホートが29歳であった年次にプロットしている。
資料：The Human Fertility Database (<http://www.humanfertility.org>), 2014/12/17 データ取得。

⁸ いわゆる「ミニひのえうま」効果は、専門的にはテンポ効果と呼ばれるものであり、厳密には晩産化などの平均出生年齢の上昇・下降だけでなく、分散変化などを含む出生年齢分布の種々の変化によって生ずるものであるが、ここでは簡単のため、両国の出生年齢分布に生じたこれらの変化をまとめて晩産化と呼んでいる。

⁹ これらの例でも、出生率回復期において晩産化が緩んだ原因として政策効果を挙げることは可能である。政策の変更等はしばしば人々の出生年齢に大きな影響を与えることがあり、スウェーデンの「ジェットコースター出生率」は実はその典型的な例にあたる。ただし、それらが実質的出生力（コーホート合計特殊出生率）の変化をとまなわれないならば、単に出生の起こる年次が変わるだけであって、「少子化の克服」にはあたらない。それどころか、年々の出生数に上下動をもたらし、その後の人口の年齢構造に長期の不均衡をもたらすという点で、負の影響を残すことにすらなる。

¹⁰ フランス、スウェーデンの出生率低下期について、後の晩産化の緩みがなければ、先送りされた出生がそのまま遺失されてしまった可能性があり、少子化国に転ずる危機にあったと考えれば、出生率回復期の政策的対応が少子化阻止に有意義であったとの見方ができるだろう。しかし、その場合であっても両国の長期にわたる手厚い家族政策の下支えが前提条件になっていたはずであり、短期的な政策のみに注目することでは、他の少子化国に应用可能な知見は得られないと考えられる。そのことは同時期の他の欧州諸国との比較などから導かれる。

これに対して、図6にドイツと日本に関する同様のグラフを描いた。この2国においても年次の合計特殊出生率（黒線）が70年代半ば頃から人口置換水準を離れ、大きく低下を示しているが、フランス、スウェーデンとの違いは、最近において低下こそ止まったものの、1.5を下回る低水準に留まっており、置換水準に向けての回復の兆しが見られないことである。そして、より大きな違いは何と云ってもコーホート合計特殊出生率（青線）が、人口置換水準を大きく割り込んで低迷している点である。すなわち、この2ヶ国の年次出生率の動向は一時期のフランス、スウェーデンにおける年次出生率の低下とは実質的に異なるものであり、晩産化の緩みのみでは置換水準付近にまで回復することはできない。実際、日本においても2006年以降、女性30代後半以降の「駆け込み出産」という形の晩産化の緩みによって年次出生率の向上が見られているが、その水準は未だ置換水準には遠く及ばない¹¹。

考察

さて、本資料では、合計特殊出生率の成り立ちについて解説するとともに、少子化の動向把握において年次別合計特殊出生率だけを見ていると、人々が生涯にもつ子ども数の動向に目が向かず、少子化の動向を正確に把握できない危険があることを紹介した。それでは、年次別の合計特殊出生率は、少子化を把握する指標として放棄すべきかといえば、決してそうではない。それは国や地域における特定の期間（年次）における人々の出生行動の強度についての的確な数値を与えるものである。問題はそれが少子化の議論に必須な視点であるライフコースに沿った出生行動についての情報をもたらさないということと、もたらさないにも関わらず、「1人の女性が生涯に生む平均子ども数」という単純化したライフコース的解釈によって理解されている場合がある点にある。

つまり、年次別合計特殊出生率は、単年に起きた年齢別の出生行動を、理解しやすいように「生涯の平均出生子ども数」に相当する数値として提示してくれる親切な指標なのだが、晩産化など出生スケジュールに変化が起きていると、「ひのえうま」年の1.58のようにライフコース視点から見ると意味のない数値になってしまい、親切が仇となってしまうということである。しかも、現実の少子化は、晩婚化・晩産化などのタイミング変化と切っても切れないので、この指標による観察は常に誤解と背中合わせになる憾みがある。

残念ながらこうした問題点をすべてクリアする単一の指標はない¹²。そもそも「指標」というものは、本来複雑な現実の複雑性を捨象し、特定の部分だけを強調して表現するためのものであるから、その特性や限界をよく知り、目的に合わせて使い分ける必要がある。欠けている部分は、他の指標や情報を駆使することになる¹³。

¹¹ 筆者の考えでは仏、典の長期にわたる手厚い家族政策の継続的、安定的、発展的持続こそが、他の少子化諸国との違いを生んでいるものである。この点には、次節「考察」においても再び言及する。

¹² 年次の合計特殊出生率の晩産化にともなう変化をテンポ効果として分離し、これを差し引いた指標（Bongaarts-Feeneyの調整合計特殊出生率など）も提案されているが、数値の解釈などに対して批判がある。

¹³ 地方における少子化対策のために、まち・ひと・しごと創生本部から提供されている未婚率や出生順位

さて、以上では、少子化という現象を扱う上での出生率という指標の見方、使い方について人口統計学分野から解説と注意喚起を行った。以下では、そのような知見が、少子化対策に対して示唆することを、私見も交えつつ、もう一步踏み込んで考えてみたい。すなわち、「出生率」という指標は、何を捨象し、何を強調しているのかということである。この指標が純化して提示しようとしているものは、ある時期の国や地域における出生の「量」である。平たくいえば、それはある年の出生の「頭数」の多寡を、「1人あたり」や「ライフコースあたり」の数値に換算して示しているものである。一方で、少子化対策が国民、とりわけ当事者となる「生まれてくる子ども達」と「その親世代」の福祉（well-being）のために行われるものであるなら、まず注視すべきものとしては、出生の「質」が中心となるべきである。それは倫理的にそうあるべきというだけでなく、少子化対策における個々の施策の実効性に対しても重要な意味を持つと考えられる。なぜならば、少子化対策を受ける側にとっては、それが「頭数合わせ」などではなく、その人々のライフコースの質を高めようとするものなのだという「メッセージ」が受信され、受容されたときにはじめて、個々の施策に反応する余地が生ずると考えられるからである。たとえば、フランス、スウェーデンの場合、長期に及ぶ家族政策の歴史の中で、人々は自らが受益者であることを経験を通して確信しているように見える。

ここで出生の質とは、たとえば全ての出生が親や社会から望まれたものであり、それゆえに出生後の心身の健全や良質な教育が保証された状態のことなどを指す。そのためには、妊娠・出産・子育てだけでなく、たとえば家庭や地域が安心、安全で親世代が将来に希望を持てるようなものでなくてはならないだろう。もちろん現在の「少子化対策」もこれらに貢献するものであるが、それらが誰のためになされているのかが不明瞭なときがあるように思う。たとえば、経済成長、社会保障制度維持や地域存続のための少子化対策（次世代の確保・健全育成）なのではなく、人々、とりわけ若年層や次世代の well-being 確保のための少子化対策であり、経済、社会保障、地域の維持発展であるということがもっと明瞭であってよい。

少子化対策を行う側は、用いる指標の特性や見方について正しく理解していることはもとより、出生を質として捉え、当事者達との継続的な対話などを通して、かれらのライフコースに寄り添って行く姿勢が求められるように思う。蛇足をいえば、フランスやスウェーデンの家族政策が実効性を有し、実質的出生率が人口置換水準付近で推移していること背景には、両国の施策が第二次大戦直後以来のきわめて長期にわたる継続的で発展的な経過の中で、人々のライフコースに自然に寄り添うものになっているということがあるように思われる。

別出生の指標、働き方に関する指標などの分析は、こうした努力の例である。

マクロデータとマイクロデータの 統計的データ融合について

慶應義塾大学 経済学部・大学院経済学研究科
星野 崇宏

bayesian@jasmine.ocn.ne.jp

問題意識

政府統計やオープンデータの利活用の重要性

⇒特に基幹調査は報告義務を有することから代表性が担保されているものと考えることができる

* 調査の誤記入や学習効果などは別途考慮すべき

可能であれば個票データ(家計・個人・企業単位)を公開し利用
一方匿名性の担保は取得情報が豊富になるほど困難

⇒一般には集計情報として公開

民間としてはマーケティングや経営意思決定のために
代表性のある集計情報をいかに活用するかは重要な課題

⇒“マイクロデータとマクロ情報の統計的融合”へのニーズ

内容

- ◆ 統計的データ融合のレビュー
- ◆ 特にマイクロデータとマクロデータの統計的データ融合
- ◆ 時系列データに対する統計的データ融合と選択バイアスの除去
- ◆ 準ベイズ推定によるマイクロデータとマクロデータの融合解析
- ◆ 適用例

統計的データ融合とは？

複数の異なる情報源から得られたデータを統合し統計的推測を行う
方法論

統計学／マーケティング分野では

Data fusion(Kamakura and Wagner,1997;Gillula and McCulloch,2013)

計量経済学では

Data Combination(Imbens and Lancaster,1994;Ridder and
Moffitt,2008; Fan et.al,2014)

特に観測単位(以降ユニット:消費者／企業／地域など)が異なる複
数のデータを活用して推論を行う方法

(c.f.)擬似パネル pseudo panel

シングルソースデータとマルチソースデータ

変数・項目

シングルソースデータ

自分の関心のある変数

すべてが、同じ対象者から

得られているデータ

⇒関連(例:広告効果)が分かる

人

購買履歴

広告接触

マルチソースデータ

自分の関心のある変数が

別々の対象者から分割して

得られているデータ

⇒普通はこれらからは関連は分からない

購買履歴

広告接触

データ融合

購買データA

調査データB

変数群A (購買履歴)	データAでの結果	データが欠測
変数群B (調査データ)	得られていない = データが欠測	データBでの結果
共変量	対象者すべてに得られている変数	

ユニットが共通しない2つのデータから

- ・ 「変数Aと変数Bの関係（相関や回帰）」の推定
- ・ 欠測値の補完によるシングルソースデータ化

を行うことが目的とされる

共変量情報を用いて データ融合が可能となる条件

y_A の条件付き分布は？ (データBでの)

$$p(y_A | y_B, z = 0, X) = \frac{p(z = 0 | y_A, y_B, X) p(y_A | y_B, X)}{p(z = 0 | y_B, X)}$$

これは推定できない ($z = 0$ では y_A 欠測)

データA $z = 1$ データB $z = 0$

変数群A y_A	データAでの結果	欠測
変数群B y_B	欠測	データBでの結果
共変量 X	調査対象者すべてに得られている変数	

共変量情報を用いて データ融合が可能となる条件

そこで $p(z = 0 | y_A, y_B, x) = p(z = 0 | y_B, x)$
=「ランダムな欠測」ならば

$$p(y_A | y_B, z = 0, x) = p(y_A | y_B, x)$$

さらに「条件付き独立」

$$p(y_A, y_B | x) = p(y_A | x) p(y_B | x)$$

ならば

$$p(y_A | y_B, z = 0, x) = p(y_A | x)$$

従って条件は

【1】 「ランダムな欠測」である

【2】 y_A と y_B が条件付き独立である

2つの条件さえ成立すれば

欠測データがある場合の完全尤度

$$\prod_{i:z_i=1}^N \int p(y_{iA}, y_{iB} | x_i) p(z_i | y_{iA}, y_{iB}, x_i) dy_{iB}$$

$$\times \prod_{i:z_i=0}^N \int p(y_{iA}, y_{iB} | x_i) p(z_i | y_{iA}, y_{iB}, x_i) dy_{iA}$$

成立なら

$$= \prod_{i:z_i=1}^N \int p(y_{iA} | x_i) p(y_{iB} | x_i) p(z_i | x_i) dy_{iB}$$

$$\times \prod_{i:z_i=0}^N \int p(y_{iA} | x_i) p(y_{iB} | x_i) p(z_i | x_i) dy_{iA}$$

$$= \prod_{i:z_i=1}^N p(y_{iA} | x_i) p(z_i | x_i) \times \prod_{i:z_i=0}^N p(y_{iB} | x_i) p(z_i | x_i)$$

観測値だけから推定が可能

擬似パネル

経済学でよく利用される方法論(Hsiao,2003)

クロスセクションデータから「本来パネルデータでしかわからない結果を得る」

例) 定年後の就業行動(岩本,2000)

定年後まで働きたいかどうか?には大きな個人差

説明変数: 生年、学歴など

⇒説明変数以外の個人差は無視して解析

* 性年齢居住地域等で集計したデータを用いる場合もある

疑似パネル(Pseudo Panel)

	2000年の調査	2005年の調査
2000年での収入	2000年の調査結果	欠測
2005年での収入	欠測	2005年の調査結果
共通項目	調査対象者すべてに得られている変数	

2つの方法

【統計的マッチング】他のデータ融合手法に同じ

【モデル仮定】例「性学歴生年コーホート内で個人差なし」として推定

Browning, Deaton and Irish, 1985; Moffitt, 1993; Hsiao, 2003

マクロデータとの融合について

ミクロデータに加えてマクロデータがある場合は？

データA $Z = 1$ データB $Z = 0$

y

データAでの結果

欠測

ケース1

+

データBでの
 y
の平均等

または

ケース2

+

全体での y の平均等モーメント

* 上記標本全体(データAとB)で無作為抽出と仮定

ケース1) データAが無作為抽出である

ケース2) データAは無作為抽出ではない

* 通常はケース2(モーメントだけ代表性ある調査から)

Imbensらの一連の方法

【先ほどのケース1を仮定する場合】

(マイクロデータからのモーメント条件)

通常の不偏推定方程式から誘導されるモーメント

(マクロデータからのモーメント条件)

マクロデータにより得られる周辺分布のモーメント条件

例) プロビットモデル サンプルサイズN

失業率を属性ベクトルxで説明 $P(y_i = 1 | x_i) = \Phi(x_i^t \beta)$

マイクロ: xの次元+1次元
$$\sum_{i=1}^N \frac{y_i - \Phi(x_i^t \beta)}{\Phi(x_i^t \beta)(1 - \Phi(x_i^t \beta))} \phi(x_i^t \beta) x_i = 0$$

マクロ: 離散化した年代(j=1~J)ごとの失業率 p_j の制約

$$\sum_{i:agec \cdot j=1} \Phi(x_i^t \beta) / \sum_{i=1}^N agec \cdot j - p_j = 0 \quad (j = 1 \cdots J)$$

* マクロデータにおける周辺分布推定時の推定誤差も考慮可能

Imbensらの一連の方法

一般化モーメント法(GMM)による推定 具体的には

$$h(y_i, x_i | \theta) = \begin{pmatrix} h_1(y_i, x_i) \\ h_2(y_i, x_i) \end{pmatrix} = \begin{pmatrix} \frac{y_i - \Phi(x_i^t \beta)}{\Phi(x_i^t \beta)(1 - \Phi(x_i^t \beta))} \phi(x_i^t \beta) x_i \\ z_{ij} \Phi(x_i^t \beta) / \sum_{i=1}^N z_{ij} - p_j \end{pmatrix} \quad g(y, x | \theta) = \frac{1}{N} \sum_{i=1}^N h(y_i, x_i | \theta)$$

として

$$\hat{\theta}_{GMM} = \arg \min_{\theta} g(y, x | \theta)^t W g(y, x | \theta)$$

- ・一般化線形モデルでの経験尤度法を用いた推定(Chaudhuriら 2008; JRSS, ser.B)

【先ほどのケース2を仮定する場合】

- ・同時分布のモーメント制約をもつ重み付き最小二乗法(Hellerstein and Imbens, 1999)
- ・経験尤度法(Qin and Zhang, 2007)
- ・Calibration estimator
- ・傾向スコア重み付け推定方程式 など

⇒方法論の仮定としてselection on observables(またはIgnorability)

マイクロデータに所属するかどうかを観測変数に依存

Calibration estimator

ケース2を仮定する方法のうち最もよく利用されている枠組み

事後層化(post-stratification)・Raking・(一般化)回帰推定を含む

1:すべての補助変数の重み付き集計値が(周辺)母集団集計値に一致するように「重み」を決める

2:それだけでは重み決定の自由度があるので、「抽出ウェイト」に”近く”なるように重みを決める

3:「近さ」を測る距離関数によって推定量が異なる

距離関数が線形関数(の2乗)の場合＝回帰推定量

ロジット関数の場合の特殊な例

＝傾向スコアによる重み付けと形式的に一致

* 但しその係数は補助変数の周辺情報によって決定される

問題点: ケース1の場合に利用できない／推定量の性質が不明確

周辺分布推定時の誤差を考慮できない

先行研究の問題点

通常マクロ情報とマイクロデータを融合させるニーズはケース2

例) マクロ統計情報は全数調査ないしは偏りのないサンプリングによる調査(政府調査)

マイクロデータは偏りあるサンプリングでユニット抽出(企業データ)

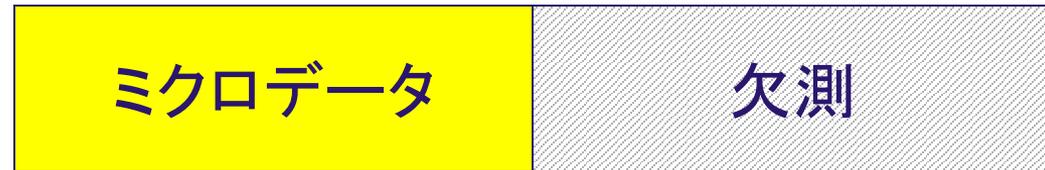
利用するデータセット間でのユニットの抽出方法の差異(ケース2)を考慮した手法はいくつかの点で問題

具体的には

- 1) selection on unobservables または missing not at random の状況では推定にバイアス
- 2) 回帰モデルで説明変数にのみ抽出が依存するという仮定(MAR)ならばそもそもweightによる調整は不要

抽出の偏りについて

無作為抽出標本からのマイクロデータの偏りを考えるために以下の
 欠測データの枠組みを
 データA $z = 1$ データB $z = 0$
 考える y



Selection on observable

/Ignorability/MARの仮定は $p(z = 1 | y, x) = p(z = 1 | x)$

ここで回帰モデル $E(y | x, \beta)$ の母数 β の推定に際して
 不偏推定方程式 $\sum_{i=1}^N m(y_i | x_i, \beta) = 0$ s.t. $E_{y|x}(m(y | x, \beta)) = 0$
 はデータAだけでも不偏

$$E(z \times m(y | x, \beta)) = E_{z|x}(z) E_{y|x}(m(y | x, \beta)) = 0$$

⇒抽出がアウトカムには依存しない場合にはマクロ情報を利用した
 重み付けはあまり意味がない

古典的な方法としてのHeckman選択モデル

アウトカム y が観測されるメカニズムを考慮

Outcome equation $y_i = x_i^t \beta + u_i$ $y_{is} > 0$ なら y_i が観測

Selection equation $y_{is} = x_i^t \beta_s + u_{is}$

同時分布を仮定

このとき y が観測される確率は

$$\begin{pmatrix} y \\ y_s \end{pmatrix} \sim N \left(\begin{pmatrix} x_i^t \beta \\ x_i^t \beta_s \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right)$$
$$p(y_s \leq 0 | y, x) = 1 - \Phi \left(\frac{1}{\sqrt{1-\rho^2}} \left\{ x_i^t \beta_s + \frac{\rho}{\sigma} (y - x_i^t \beta) \right\} \right)$$

⇒MARではない($y_s < 0$ でunobservableである y に依存)

推定の識別性問題(特に ρ) / 分布仮定からの逸脱に敏感

Little(1983), Copas and Li(1997)

⇒実際の解析ではあまり利用されない

新しい解析手法 GMMの準ベイズ推定

新手法の目的

「マイクロデータ」と「マクロ情報」の融合において
先行研究で考慮されなかった以下の点を解決

- ・マイクロデータがMNARであることを考慮することができる
- ・マクロ情報があることで識別性が担保される
- ・マクロ情報を推定する際の推定誤差も考慮する

具体的な適用の想定例としては

(マイクロデータ)企業の保有する購買履歴データ

(マクロデータ)家計調査の集計データ

⇒マイクロデータは特定の小売りチェーンで購入しているので“偏りを有する”データ

例えば

自社以外での購買と自社で得られる情報の関係の理解

MNARのモデルについて

モデルの尤度は以下のように表現される。

$$\prod_{i=1} p(y_i | x_i, \theta)^{z_i} \left[\int p(y_i | x_i, \theta) p(z_i | y_i, x_i, \phi) dy_i \right]^{1-z_i}$$

上記尤度から構成される不偏推定方程式は下記

$$g_1(y, x | \theta) = \frac{\partial}{\partial \theta} \sum_{i=1}^N z_i \log p(y_i | x_i, \theta) + (1 - z_i) \log \left[\int p(y_i | x_i, \theta) p(z_i | y_i, x_i, \theta) dy_i \right] = 0$$

または

$$g_1(y, x | \theta) = \frac{\partial}{\partial \theta} \sum_{i=1}^N \frac{z_i}{p(z_i | y_i, x_i)} \log p(y_i | x_i, \theta) = 0$$

但し一般的には上記のモデルの識別性のために

補助情報としてマクロ情報が与えられるとする。

$$g_2(y, x | \theta) = m(y, x | \theta) = 0$$

* 具体的形式はモデルと情報に依存: y の平均や分散など

合わせるとモーメント制約は

$$g(y, x | \theta) = \begin{pmatrix} g_1(y, x | \theta) \\ g_2(y, x | \theta) \end{pmatrix} = 0$$

準ベイズ推定

スコア関数の部分で欠測値についての積分が必要

本研究では擬ベイズ推定法に基づくマルコフ連鎖モンテカルロ法を用いた推定を利用する。

【準ベイズ推定】quasi-Bayesian estimation using MCMC

Chernozhukov and Hong (2003)

通常の事後分布

$$p(\theta | y) = \frac{p(\theta) \exp l(\theta | y)}{\int p(\theta) \exp l(\theta | y) d\theta}$$

但し $l(\theta | y)$ は対数尤度

ではなく擬似事後分布

$$q(\theta | y) = \frac{p(\theta) \exp l^q(\theta | y)}{\int p(\theta) \exp l^q(\theta | y) d\theta}$$

但し $-2l^q(\theta | y)$ はGMMの基準関数

とすると擬似事後平均は一致性があり漸近正規性を有する

準ベイズ推定によるミクロマクロ融合

【準ベイズ推定の利点】

擬似事後分布 $q(\theta | y)$ からの乱数列を利用して
 マルコフ連鎖モンテカルロ法(MCMC)により統計的な推測が可能
 ⇒ 欠測値もMCMCの途中で擬似予測事後分布から発生すればよい

【基準関数】(Igari and Hoshino,2015;Hoshino,2016)

今回の状況ではミクロとマクロで独立な情報のため

$$\begin{aligned}
 l^q(\theta | y) &= \begin{pmatrix} g_1^t(y, x | \theta) & g_2^t(y, x | \theta) \end{pmatrix} \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix} \begin{pmatrix} g_1(y, x | \theta) \\ g_2(y, x | \theta) \end{pmatrix} \\
 &= g_1^t(y, x | \theta)W_1g_1(y, x | \theta) + g_2^t(y, x | \theta)W_2g_2(y, x | \theta)
 \end{aligned}$$

* 注意: ベイズでは事前知識を事前分布で表現できるとされるが、
 一般に母数の関数についての情報は明示的に表現できない

具体的なモデルの例

(1) 標本調査における未回収バイアスの補正

回収有無がモデリングの対象となる変数 y や x に依存

【マクロ情報】 Y や X の周辺分布、一部変数での同時分布

(2) 競合他社での購買金額の推定

$y = y_I + y_C$: y_I は自社での購入額 y_C は他社での購入額
自社が持っている情報は Y_I と個人属性・来店履歴等説明変数 X

$$p(y | x, \theta) \text{ または } p(y_C | x, \theta)$$

の母数推定／予測分布の推定に関心がある

【マクロ情報】 Y の周辺分布、あるいは X の一部と Y の集計表

⇒観測されない潜在変数 $y_C = y - y_I$ について y の周辺分布の
モーメント情報と y_I , X のモーメント情報から推論可能

具体的なモデルの例

(3) 真の購入頻度の推定(Igari and Hoshino,2015)

他社でも当該商品
購入

⇒自社の情報だけ
では真の間隔は
不明

繰り返しのある

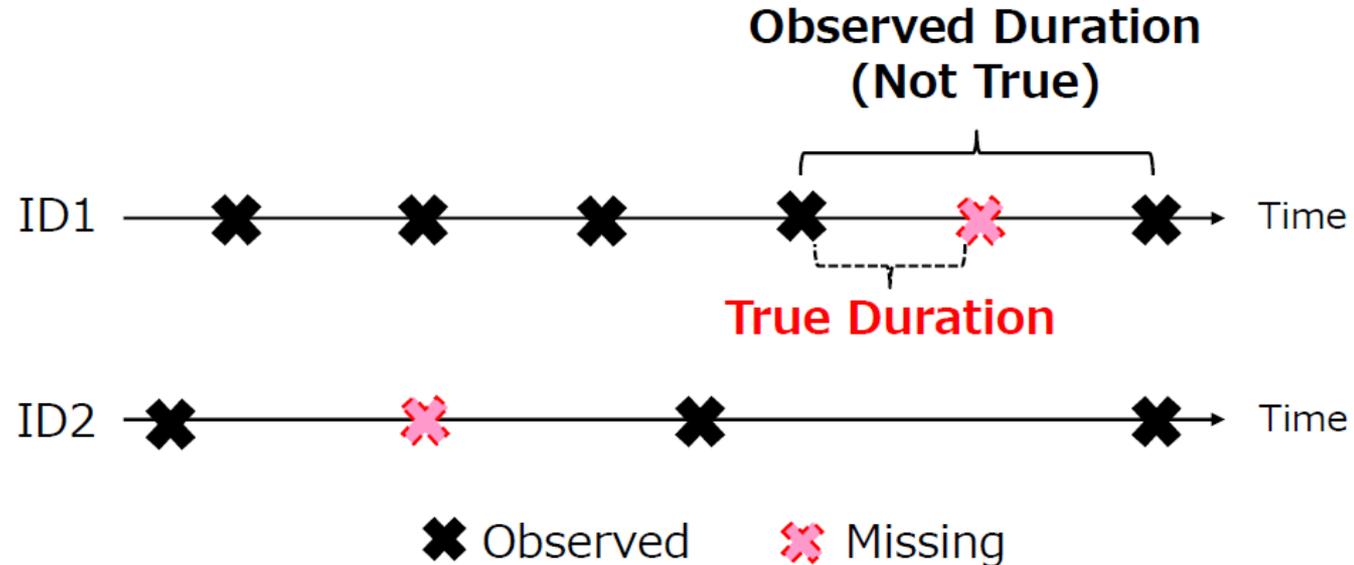
Duration analysis

属性や価格、企業のプロモーション変数で説明したい

【マクロ情報】

家計調査での当該商品の属性別年間購入数量のクロス集計表

(4) 集計レベルの時系列データと個人・家計レベルの個票データの融合: 状態空間モデルでの時変係数を仮定を置かず推定



まとめと議論

- ◆ シミュレーションの結果と具体的な解析例については当日報告
 - ◆ 今回は公的統計を民間がいかに利活用するかという観点から“マイクロデータとマクロ情報の統計的融合”の方法論の紹介と新しい解析手法の枠組みについて紹介
 - ◆ 但し誤記入や学習効果の可能性がある公的調査の“バイアス補正”にも活用可能と考えられる
- 例) 家計レベルのマイクロデータ「家計調査」と売上ベース・小売業による集計された地域別品目別の「商業動態統計調査」の融合
⇒ 周辺分布は「商業動態統計調査」に従う「家計調査」個票データ
- ◆ 公的調査の匿名データが利用可能になった場合には“マルチソースのマイクロデータの統計的データ融合”の方法論のさらなる発展と利用が望まれる