



2014.09.10

仕事術

データプレゼン入門 政策をつくる人のためのデータで伝える技術 第2回 統計データはどこにあるか

どんな統計データがあるか

今回はプレゼンの原となる統計データそのものがどこにあるか、どのように入手するか、また、どう利用するか、について触れましょう。本誌の読者層を考慮して、市町村別のデータに焦点を当てることにします。

全国一律の基準で調べられていて、他の市町村と比較可能な市町村別統計データのうち主要なものを表1に掲げました。

統計調査には、調査統計と業務統計があります。住民の身長と体重を調べるため、自治体が行っている健康診断の際に測定した結果を集計したとしましょう。これは業務統計です。本来の行政目的から副産物的に得られたデータを「業務統計」といい、そうではなくて、データを得ることを目的とした調査を「調査統計」といいます。付け加えて言及しておく、GDP統計など複数の統計を組み合わせて作成される統計は「加工統計」と呼ばれます。

犯罪統計が典型的な業務統計です。警察は犯罪人の認知、検挙という本来の業務から派生的に犯罪認知件数や検挙人数を把握することになります。この場合、警察庁が全国の警察署で業務統計をとる際の統一的な犯罪分類やカウントのタイミング、集計表のフォーマットなどの基準を徹底していますので、地域ごとに犯罪件数が比較できることになります。

市町村自治体は様々な業務統計を作成していますが、省庁が全国的な統一基準を作成し、市町村に義務付けていない場合もあり、その場合は、市町村相互には単純に比較できないので注意が必要です。例えば、自治体が自ら管理している街路樹や花壇の数の統計を作成している場合、どの自治体でもその統計を必ず作成しているとはいえ、またどのくらいの大きさのものまでカウントするかは基準もないでしょう。したがって、時系列変化などその自治体にとっては大切なデータとなっているかもしれませんが、他市町村と人口当たりの数を単純に比較したりはできないのです。

また、業務統計はバイアス（偏り）がかりやすいので注意が必要です。上で述べた健康診断の結果から住民の身長・体重を集計した結果は、会社で健康診断をする人や健康診断をそもそも忌避している人を除いたデータになります。したがって、対象者が女性に偏ったり、むしろ健康者に偏ったりしている可能性もあることから、住民全体を代表するデータにはなりません。調査統計では、調査対象集団（母集団）から無作為に抽出した者に対して調査を行うことから、こうしたバイアスがかからないので信頼が置けます。身長・体重については、厚生労働省が行っている国民健康・栄養調査という調査統計の中で調査されており、信頼が置けるデータとなっています。

また、業務統計には、データ改ざんのおそれもあります。この7月末に大阪府警で全国ワースト1の地位を避けるため、すぐに自転車が見つかった自転車盗を省いたりして、管内の街頭犯罪の件数を過少報告したことが明らかとなり、関係した幹部が処分されました。配下の大名からの石高の自己申告による業務統計が信用できないため、秀吉は、太閤検地という調査統計を強行したのでした。

表1に掲げた統計は、調査統計も業務統計もありますが、全て省庁によって基準が設けられており、それに従って作成されているので信頼が置けるデータといえます。したがって、他の類似した町村との比較も十分可能です。

ただし、観光統計だけは、同じ地域の異なった観光施設を訪れるといった場合の観光流動のダブルカウントをどう回避するか、また地元民がほとんどの盆踊り客まで対象とするかなど、基準の統一が難しいため、自治体独自の基準でデータがとられており、かつ同じ基準で毎年調査されているため時系列変化は追えても、他地域と観光客数規模を比較することはできない場合が多いので注意が必要です（観光庁が統一基準の観光統計の作成を都道府県に指示して開始していますが、市町村別データは公表されていません）。

	統計調査名	対象データ	調査頻度	主管省庁等
人口・世帯	国勢調査	人口、就業者、通勤者、世帯、住宅等	5年	総務省統計局
	住民基本台帳人口・世帯数	人口・世帯、出生・死亡、転入・転出、外国人数	毎年(自治体では毎月)	総務省
	人口動態統計	出生、死亡(死因別等)、婚姻、離婚	毎年	厚生労働省
	将来人口推計(注2)	性・年齢別将来人口	5年	社会保障・人口問題研究所
産業	農林業センサス、漁業センサス	農家数、漁業事業体数、農林漁業就業者数	5年	農林水産省
	作物調査、漁業・養殖業生産統計	耕地面積、作付面積・収穫量、漁獲量	毎年	農林水産省
	経済センサス 工業統計 商業統計	産業別事業所・企業数・従業員数、工業出荷額、商業販売額、サービス売上額	5年(工業統計は毎年)	総務省統計局、経済産業省
	観光統計(注2)	入込客数、宿泊者	毎年	自治体
行政の各分野	建築着工統計	着工建築物数・床面積	毎年	国土交通省
	医師・歯科医師・薬剤師調査	医師、歯科医師、薬剤師	2年	厚生労働省
	学校基本調査	小中学校・高校数・生徒数、卒業後の状況	毎年	文部科学省
	犯罪統計、交通事故統計(注2)	犯罪認知件数、交通事故件数・死傷者数	毎年	都道府県警察
行財政(注2)	公共施設状況調	公共施設の状況	毎年	総務省
	決算状況調、決算カード	決算の状況	毎年	総務省
統計集	社会・人口統計体系	全分野	毎年	総務省統計局

(注) 1. 市データのみ得られ、町村データの得られないものは除く。

2. 市町村データは、担当省庁のホームページか、政府の統計データを全て収録したe-Statサイトから入手可能。どちらからでも入手できる場合が一般的だが、以下のものは、括弧内に示した担当部局等のサイトからしか入手できない。

- 将来人口推計(社会保障・人口問題研究所)
- 観光統計(各都道府県の統計部局サイト)
- 犯罪統計、交通事故統計(各都道府県警察)
- 行財政(総務省統計情報)

表1：市町村データが得られる主な統計データ一覧

なお、統計データのうち、市町村別のデータが得られる統計は、全数調査の調査統計か、市町村単位までのデータが必要な業務統計に限られます。

調査統計には、国勢調査や経済センサスといった世帯や事業所などの調査対象を全て調べる「全数調査」と、全体を代表するような一部のサンプル的な調査対象に対して失業率などを調べている労働力調査、あるいは家計収支を調べている家計調査などの「標本調査」とがあり、件数的には、標本調査が圧倒的に多くなっています。

標本調査の場合は、サンプル数の制約により、せいぜい都道府県別のデータが得られるのみであり、市町村別のデータは得られません。

国勢調査のような全数調査は、標本調査のサンプリングの基礎としての役割を果たすとともに、市町村やそれより狭い調査区などの区分の集計結果や性・年齢・学歴・地域など多重クロス集計結果が得られるため、統計調査の全体体系の中でも重要な役割を有しています。

ただし、全数調査は、大変なコストや国民負担を要しますので、しょっちゅう行うというわけにはいきません。そこで、毎月の失業率データがすばやく必要な労働力調査は、国勢調査の調査区を基礎にした無作為抽出による標本調査として実施されているのです。このため、市町村別の失業率データは、5年に1度の国勢調査の結果としてだけ得られることとなります。町村部の雇用情勢のデータとしては、これをベンチマークに据えながら、最寄りのハローワーク管内の業務統計である有効求人倍率の月次データの動きで判断していくこととなります。

もうひとつ統計調査について知っておくべきことは、統計法上の位置付けの違いです。調査統計であれ、業務統計であれ、国や地方公共団体がまとめる統計を公的統計といいますが、公的統計のうちでも重要なものは、総務大臣によって「基幹統計」に指定されています(統計法の改正前は「指定統計」と呼ばれていました)。

統計法は、基幹統計について、報告(回答)を求められた者が、報告を拒んだり虚偽の報告をしたりすることを禁止しており(13条2項)、これらに違反した者に対して、50万円以下の罰金が定められています(61条)。国勢調査の人口数を水増しするなどといった実際にもあった統計担当公務員の違法行為には、さらに重たい刑が科せられています。犯罪統計は基幹統計に指定されていませんから、上に触れた大阪府警のデータ改ざんは警察内の処分にとどまっています。こうした強制措置を持つ基幹統計は、他方で、結果データについて国民への公表が政府に対して義務付けられています。

統計調査の分類の3つのポイント

1. 調査統計か業務統計か
2. 全数調査か標本調査か
3. 基幹統計かそうでないか

どこにデータそのものがあるか

統計データを入手するのは、以前はかなりの困難を伴う作業でした。一般人なら、図書館の統計書を当たる、省庁の図書館や担当部署に問い合わせでデータをもらう、省庁統計担当部局のプレス発表資料を入手する、といった手段が必要でしたし、地方議会議員であれば、自治体（市町村あるいは県）の担当課や議会事務局からデータを得る、地元国会議員の秘書を通じて省庁の担当課からデータを手に入れる、といった手段もあったでしょう。対象となる地域だけのデータであれば比較的容易であっても、これを他の地域と比較する場合は、かなりの困難を伴うのが実情でした。

ところが、近年のインターネットの発達によって事態は激変しました。ほとんどの統計データがインターネットのサイトから一般に提供されるようになりました。入手先については、表1の下の注を参照してください。データ入手に関する主な困難は、どこにどんなデータが存在するかが分からない点にあります。しかし、これもインターネットの検索サイトの発達によって、かなりの程度、解消されつつあります。先日、NHKのクローズアップ現代という番組で男性は女性より幸福度が低い点が報じられました。データ出典は男女共同参画白書とパネルで示されました。私も早速、「男女共同参画白書」と検索窓に打ち込んだところ、原データにたどり着くことができました。現代は、こんな形でデータを手に入れることができるのです。

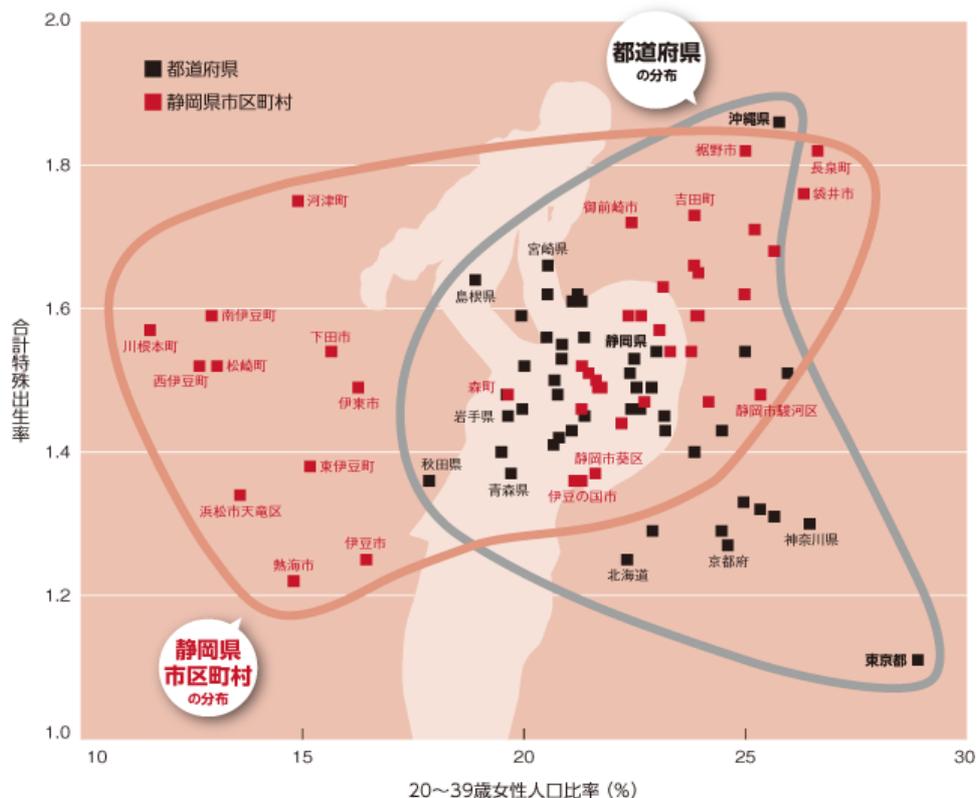
市町村統計データの活用例その1～出生率と20～30代女性人口～

こうして得られる市町村データの活用例を掲げてみましょう。

厚生労働省の人口動態統計では出生数を調べていますが、母親の年齢も同時に調べることから、女性が一生に産むであろう子どもの数を合計特殊出生率として算出しています。小さな規模の町村では毎年の出生数の変動が大きくなりますので、市町村のデータまで算出するのは、前後5年間のデータを基に5年に1回の頻度となっています。

図1では、都道府県のばらつきと市町村のばらつきを比較するため、都道府県と静岡県市区町村のデータを散布図にしています。静岡県を多くの県の代表として取り上げたのは、静岡市が新商品売り出しのサンプル地域として採用されることが多いなど、静岡が日本の中央にあって、しばしば日本の地方を代表する地域として取り上げられることが多いためであり、他意はありません。

散布図のX軸には、子どもを産む中心的な年齢である、20～39歳の女性人口比率（女性総数に占める割合）をとり、Y軸には、合計特殊出生率をとっています。右上の地域は、出生率も高く、子どもを産む可能性のある女性の数も多いので、出生数も多くなり、左下では、逆に出生数が少なくなる傾向となります。東京都のような右下の地域では、若い女性は多いけれど、子どもを産む女性の割合は高くないため、子どももあまり増えないという状況にあると判断できます。また左上の静岡県河津町は、出生率自体は高いのですが、女性そのものが少ないため出生数規模はあまり期待できない地域だと見ることができます。



(注) 合計特殊出生率については、人口規模の小さい自治体は周辺地域の値を考慮したベイズ推定値、合計特殊出生率は2008～2012年の平均、20～39歳女性人口比率は2010年国勢調査データ。

資料：厚生労働省「平成20～平成24年人口動態保健所・市区町村別統計」、総務省統計局「平成22年国勢調査」

図1：都道府県のばらつきと市町村のばらつきの比較（出生率の例）

都道府県のばらつきと市町村のばらつきを比較すると、ばらつきの程度は、ほぼ同等と見ることができます。市町村の指標を県の指標で代表させることがあります。ここで取り上げた例から分かります。静岡県と沖縄県、あるいは東京都との値の差異より、静岡県の県内市区町村の値のばらつきの方が大きいことから、やはり無理であると断ぜざるを得ません。

なお、この散布図から、ばらつきの分布パターンが都道府県と市町村で異なっている点も目につきます。県内には20～39歳の女性人口比率が都道府県のばらつきを超えて少ない市町村が存在していることが分かります。そうした地域には、県内の伊豆地方の市町が多くなっていますが、川根本町や浜松市天竜区といった山間部地域も含まれています。

市町村統計データの活用例その2～復興過程の検証～

次に、戦災や震災などからの復興過程を統計データで検証した例を2つ掲げましょう。

ひとつ目は、広島市の長期人口です（図2）。データは1990年の広島市の市域に合わせて合併した市町村を足し合わせた時系列データです。

1945（昭和20）年の原爆投下による被害で人口も大きく減少した後、人口規模から見れば、1975年くらいまで、すなわち30年くらいかけて、ほぼ同じ中国地方の県庁所在地都市である岡山市程度の人口趨勢に復帰したことがうかがわれます。

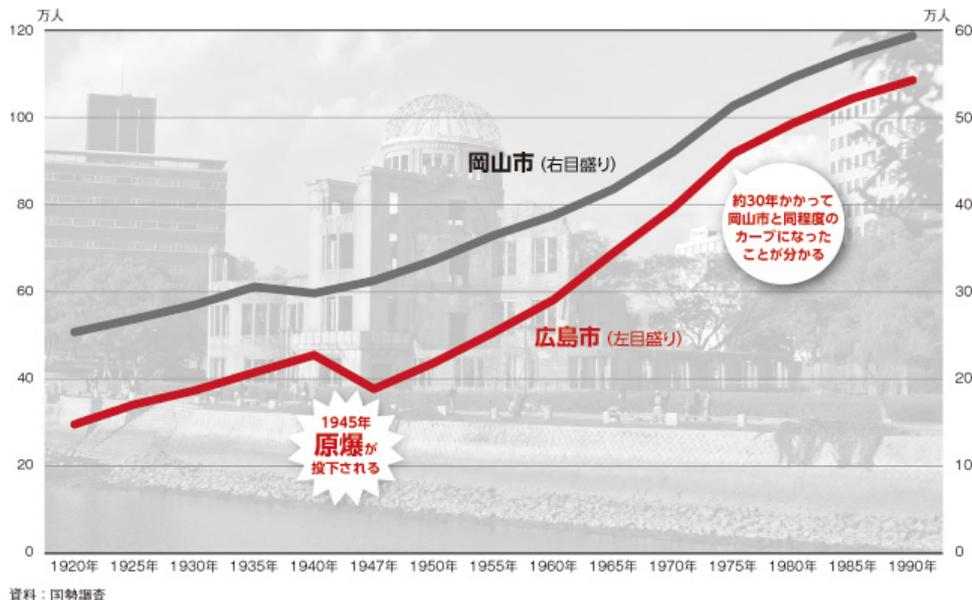


図2：広島市人口の長期推移を近隣類似市と比較する

もうひとつは、1995年の阪神・淡路大震災で被災した神戸市の地方税収の推移から復興過程の状況を調べたものです（図3）。

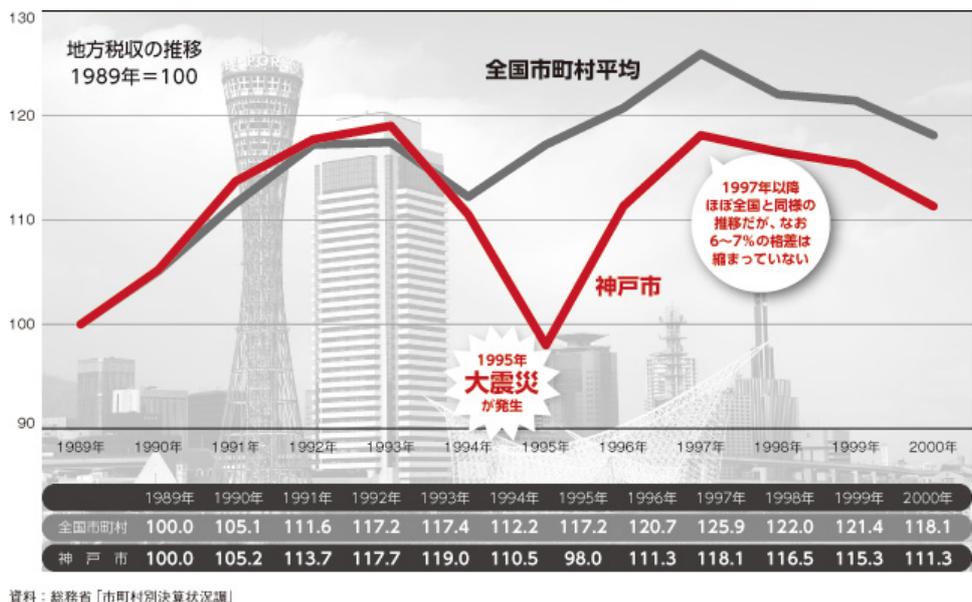


図3：阪神・淡路大震災の財政的影響を神戸市の地方税収の動きから見る

地方税収の推移を1989年を100とする指数で全国市町村平均と神戸市とで比較してみると、1989年から1994年までは神戸市も全国市町村とほぼ同様の推移をたどっていましたが、大震災の発生した1995年には全国平均より20%減の地方税収となっ

ています。その後、全国平均との格差は縮まり、1997年度以降はほぼ全国と同等の毎年の動きとなっています。ただし、震災前のレベルにまでは復帰することはできず、なお6~7%の格差は残り続けているといわざるを得ません。



この記事の著者

本川裕

アルファ社会科学株式会社 主席研究員

アルファ社会科学(株)主席研究員。あらゆるジャンルの統計データをユニークな視点でグラフィック化した人気サイト「社会実情データ図録」(<http://www2.ttcn.ne.jp/honkawa/index.html>)を主宰。2004年のスタートから、毎週2件のデータ更新を行い、現在1,200件を超える統計グラフと解説は、1日1万件以上のアクセスがある。東京大学農学部農業経済学科卒、(財)国民経済研究協会常務理事研究部長を経て現職。立教大学兼任講師。1951年神奈川県生まれ。著書『統計データはおもしろい！—相関図でわかる経済・文化・世相・社会情勢のウラ側—』(技術評論社、2010年)、『統計データはためになる！—棒グラフから世界と社会の実像に迫る—』(技術評論社、2012年)、『統計データが語る 日本人の大きな誤解 (日経プレミアシリーズ223)』(日本経済新聞出版社、2013年)。