

## 統計学習用補助教材について

## 補助教材

統計とは？ ～学習の入り口～	P 1
統計のできるまで	P 1 3
標本調査とは？ ～仕組みと設計～	P 1 6
労働力調査における標本抽出方法の概要	P 2 3
統計をグラフにあらわそう ～種類と特徴～	P 2 4
統計を使ってみよう	
家計調査の見方・使い方	P 3 9
都道府県・市区町村別データの調べ方	P 4 7
基本用語集	P 4 9
統計エピソード集 ～授業導入部で～	P 5 6

## 統計とは？～学習の入口～

### 目 次

#### 1 統計ってどんなもの？

- 辞書にはどう書いてあるの？
  - どんな実例があるの？
- <みんなで考えてみよう>

#### 2 統計はどのように役に立つの？

- テストの平均点はどのように役に立つの？
  - ボール投げの記録はどのように役に立つの？
- <みんなで考えてみよう>

#### 3 「統計」、「データ」、「情報」は同じ意味なの？

- 「データ」と「統計」はどこが違う？
  - 「情報」は「統計」や「データ」とどこが違う？
- <みんなで考えてみよう>

#### 4 中学校についての統計を見たことがありますか？

- 私たちの学校と日本全国の学校を統計で比べてみると？
  - 1クラス（学級）当たりの生徒数は？
- <みんなで考えてみよう>

#### 5 日本にはどんな統計があるの？

- どんな統計を知っていますか？
  - 何のために国の統計があるの？
  - 国の統計はどうやって作るの？
  - 国勢調査ってどんなもの？
  - 国勢調査には必ず答えなければいけないの？
  - 国勢調査では個人の秘密は守られるの？
- <みんなで考えてみよう>

#### 6 統計はだれのもの？

- みんなで作る統計、みんなで社会に活かす統計
- <みんなで考えてみよう>

## 1 統計ってどんなもの？

### ○ 辞書にはどう書いてあるの？

「統計」という言葉は日常生活でよく耳にしますが、あらためて「統計」とはどのようなものか説明しようとする、意外に難しいものかもしれません。統計とはどのようなものか、考えてみましょう。

まず、辞書で「統計」という言葉を調べてみます。

「集団における個々の要素の分布を調べ、その集団の傾向・性質などを数量的に統一的に明らかにすること。また、その結果として得られた数値。」(例 統計をとる)(広辞苑より)

「集団現象を数量的に把握すること。一定集団について、調査すべき事項を定め、その集団の性質・傾向を数量的に表すこと。」(例 統計をとる)(大辞林より)

他のどの辞書でも、言い回しは少しずつ違いますが、ほぼ同様の解説がされています。いずれにも共通なのは、「集団」の「傾向・性質」を「数量的」に明らかにすることです。

### ○ どんな実例があるの？

このような辞書の説明は抽象的で分かりにくい面もありますが、実例で考えてみると案外分かりやすいものです。

例えば、あるクラスに 20 人の生徒がいて、英語のテストの成績は次のとおりだったとしましょう。

表 1 いろは中学校 1 年 A 組の生徒の英語テストの成績

生徒番号	1	2	3	4	5	6	7	8	9	10
得点	78	89	85	82	83	85	80	85	87	83

生徒番号	11	12	13	14	15	16	17	18	19	20
得点	80	87	89	85	90	85	95	87	83	92

このクラスのテスト結果の平均点を計算してみると、

$$(\text{生徒の合計得点}) \div (\text{生徒数}) = 1710 \div 20 = 85.5 \text{ 点}$$

となります。

この平均点 (85.5 点) は、20 人の生徒という「集団」について、英語のテストによって英語の力を「数量的」に明らかにする数値ですので、辞書で言う「統計」の定義に当てはまるものです。そして、英語のテストを行うことは、教育の一環ではありますが、「統計をとる」ための活動であると言えます。

このような「集団の傾向・性質などを明らかにする」数値は、世の中には他にもたくさんあります。例えば、みなさんのボール投げも何回か記録をとれば統計の一種です。ボール投げでなくても、50m 走の記録でも、走り幅跳びでも、何でもかまいません。

「ボール投げの記録は、自分一人についての数字なのに、どうして「集団の傾向・性質」を明らかにしているのか」と疑問を持つ人がいるかもしれません。ここでいう「集団」には、学校のクラスのような人の集まりという意味での「集団」だけではなく、ボールを投げるといふ「出来事」の集まりとしての「集団」という意味も含まれています。ですから、何回かのボール投げの記録は、あなたがボール投げをしたという出来事の「集団」の性質・傾向を表わす統計ということになります。

### <みんなで考えてみよう>

このように、「統計」は私たちの生活のいろいろなところで用いられていますが、暮らしの中では、その数字を「統計」と呼ばないで使うことがしばしばあるため、「統計」に気づかないこともあります。生活の中で、ほかにどのような「統計」があるのか、みんなで考えてみるとよいでしょう。

## 2 統計はどのように役に立つの？

### ○ テストの平均点はどのように役立つ？

さて、このような「統計」は何の役に立つのでしょうか？

ここでも、平均点とボール投げの記録を例に考えてみましょう。あるクラスの英語のテストの平均点が分かると、そのクラスの英語の力が他のクラスに比べてどのくらい優れているのか、あるいは足りないのか、といった比較ができます。この数字によって、英語の授業がどれだけ生徒たちの身に付いたか、生徒たちがどれだけよく勉強したか、ということが分かります。また、他のクラスとの比較だけではなく、同じクラスのテストの平均点の時間的な変化を追いかけていくことで、そのクラスの実力が時間とともに向上しているかどうか、といった様子を観察することができます。

仮に平均点を計算せずに個人個人の点数しか見ないでいると、点数が上がった人も下がった人も混じっていたりして、クラス全体の傾向を見るのが難しくなってしまいます。そこで、「集団の傾向・性質などを数量的に統一的に明らかにする」ために、クラス全体の傾向をまとめて表わす平均点が計算されているのです。

つまり、「統計」には、異なる集団の間で比較をしたり、同じ集団での時間的な変化をとらえたりするための数字としての役割があるのです。

### ○ ボール投げの記録はどのように役に立つの？

ボール投げの記録という「統計」についても同じことが言えます。ボール投げの記録が去年と比べてどれくらい良いか（悪いか）かということは、去年からの体力・技術の向上（低下）の程度を一目で表した数字です。このような数字で示さずにボール投げが得意になった、苦手になったと言ってみても、ただカンに頼って「調子が良い」、「調子が悪い」と言うのと変わりはありません。本当にボール投げが得意になったのか、得意になったとしたら、どのくらい得意になったのかといった「程度」をはっきりさせるには、それを数字で表すのが最も分かりやすいと言えます。成長して体力が向上しているはずなのにボール投げの記録が伸びないということなら、投げる技術に問題があるということであり、投げるフォームや投げるタイミングを見直したりして記録を上向かせる努力が必要になることでしょう。クラス全体のボール投げの記録があれば、自分の記録と比較することで自分のがんばり具合もはっきりと分かるはずで

統計の数字は、きちんとした数量によって現在の状態をとらえたり、変化を見たりするために必要なのです。

一流選手であれば、仮に記録を見なかったとしても、自分を上手にコントロールできることでしょう。しかし、部活などで陸上競技や野球をしたりするみなさんは、個々の選手やチーム全体の記録あるいは打率（野球の試合でヒットを打つ確率）などの統計をとっておくと、ひとり一人の調子やチームとしての成績がよりよく分かります。それを参考にしながら、どのような練習やプレーをしたらもっと上達するか、もっと成績がよくなるのか、研究するのに役立ちます。

### <みんなで考えてみよう>

暮らしの中にある統計の数字を見ながら、それがどのような意味を持っているのか、どのような実態を表しているのか、みんなで考えてみるとよいでしょう。

### 3 「統計」、「データ」、「情報」は同じ意味なの？

#### ○ 「データ」と「統計」はどこが違う？

現代の社会は「情報化社会」と言われており、至るところにデータや情報があふれています。この「データ」や「情報」といった言葉は、「統計」ともよく似ているので、間違っって使われることもあるようです。これらの言葉はそれぞれどのような意味で使われ、どのような違いがあるのでしょうか。

「データ」というのは、辞書によると「立論・計算の基礎となる、既知あるいは認容された事実・数値。資料。与件。」（「広辞苑」より）という意味だそうです。難しい説明のように思えますが、要するに、計算をするための「基になる数字の集まり」のことと考えればよいのです。

例えば、表1「いろは中学校1年A組の英語テストの成績」に含まれている20人分の点数は「データ」です。表1の「データ」を基にして計算することにより、平均点という「統計」が得られます。つまり、「データ」は「統計」を計算するための基になるものです。

なお、「データ」は必ず統計を計算するために用いられると決まっているわけではありません。「統計」を計算するためには基になる「データ」が必要ですが、「データ」があるからといって、それが必ず「統計」として表されているとは限りません。

#### ○ 「情報」は「統計」や「データ」とどこが違う？

さて、「情報」は、「統計」や「データ」とどこが違うのでしょうか？辞書（「広辞苑」）には、「情報」という言葉について二つの意味が書かれています。1番目の意味は「あることがらについてのしらせ」で、2番目の意味は「判断を下したり行動を起こしたりするために必要な、種々の媒体を介しての知識」です。

1番目の意味の「情報」の例としては、「スパイからの情報」があります。これは、スパイが発見した事実（秘密！）についてのしらせです。2番目の意味の「情報」の例としては、「ガソリン価格の情報」があります。これは、ガソリンの価格が上がっているのか（下がっているのか）、どこの町のガソリン価格が安いのか（高いのか）といった「知識」を指します。この「情報」があれば、いつごろ、どこでガソリンを買ったら得になる（損になる）のかといった「判断」に役立ち、実際に買いに行くという「行動」の助けにもなります。

「情報」のうち「統計」や「データ」と似ていてまぎらわしいのは、2番目の意味のものです。「ガソリン価格の情報」を例として言えば、ある月のガソリン価格の平均値は、それだけであればただの数字ですが、その数字の動きを何か月も分析していくと、ガソリン価格が上がっているのか、下がっているのか、今後の見通しをどう考えたらよいか、といった「情報」が読みとれます。ガソリン価格が値上がりしている時には高くなる前に早めに買っておいたほうがよいでしょうし、逆に値下がりしている時にはもっと下がるまで待ったほうがよいでしょう。また、ガソリンの安い町が分かれば、そこに行ってガソリンを買うほうが得かもしれません。

このように、「判断を下したり行動を起こしたりするために必要な「知識」のことを「情報」と呼んでいます。「統計」の数字をひと目見ただけで判断を下すのは難しいですが、「統計」の数字をよく分析して、その背景にある事情などが分かれば判断を下すことができるでしょう。つまり、「統計」の数字があるだけでは「情報」とは言えませんが、それを分析して数字の意味を解釈して何らかの知識が得られたら、その知識が「情報」と呼ばれるものになります。「情報」には、「統計」や「データ」を分析して得られる数量的な（定量的な）「情報」ばかりではなく、ニュースや報告などを分析して得られる定性的な「情報」もあるので、「情報」という言葉のほうが「統計」や「データ」よりも広い意味を持っています。

学校の授業で「統計」を学ぶことの大きな目的の一つは、「統計」や「データ」を正しく的確に分析して、私たちそれぞれの判断や行動に役に立つ「情報」を導き出す力を養うことです。「統計」を学ぶことは、私たちの暮らしに大いに役に立つことなのです。

<みんなで考えてみよう>

身の回りにある数字について、それが「情報」、「統計」、「データ」のどれに当たるのかみんなで考えてみるとよいでしょう。同じ数字でも、見方によってどれに当てはまるかが変わることがあるかもしれません。また、実際の「統計」や「データ」を見て、それから役に立つどのような「情報」が得られるのか、といったことも考えてみましょう。

#### 4 中学校についての統計を見たことがありますか？

- 私たちの学校と日本全国の学校を統計で比べてみると？

自分のクラスに何人の人がいるかという数も「統計」の一種です。では、自分の学校全体だと、クラスの数はいくつあって、児童・生徒数は何人いるのでしょうか？これらの数字も「統計」です。

みなさんの町（市・町・村）には中学校がいくつあり、クラス（学級）の数や生徒数はどうなっているのでしょうか。日本全国ではどうでしょう。

これらの数字を総務省統計局の日本統計年鑑（平成 22 年版）を使って調べると、次のようになっています。（<http://www.stat.go.jp/data/nenkan/22.htm> の22-4表より）

表 2 全国の中学校の状況（中学校数、学級数、生徒数、平均生徒数）

年 （西暦）	中学校数	学級数	生徒数 （1000 人）	1 学級当たり 平均生徒数（人）
1985	11,028	156,516	5,990	38.3
1995	11,194	137,075	4,570	33.3
2005	10,960	118,182	3,626	30.7
2008	10,839	119,933	3,592	30.0

- 1 クラス（学級）当たりの生徒数は？

表 2 によると、2008 年には日本全体に中学校は 10,839 校あり、クラス（学級）は 11 万 9,933 あり、中学生は 359 万 2 千人いることが分かります。

359 万人というと、あまりに大きすぎて、どれくらいの大きさか想像がつかいません。でも、これを 1 クラス当たりの平均の生徒数に直してみると、身近な大きさの数字になります。2008 年の 1 クラス当たりの平均生徒数は、30.0 人となっています。今のクラスは、これよりも大きいでしょうか、それとも小さいでしょうか？このように全国平均の数字があれば、自分のクラスの大きさを日本全体と比べることができます。

また、この表 2 の数字から、1985 年から 2008 年までの 20 年余りの間に、日本の中学校の生徒数は 599 万人から 359 万 2 千人に減り、これにつれて 1 クラス当たりの生徒数も 38.3 人から 30.0 人へと大幅に減っていることが分かります。さて、今いる学校の 1 クラス当たりの生徒数は、10 年前や 20 年前に比べて増えたでしょうか、減ったでしょうか？日本全国と同じような傾向があるでしょうか？

このように、大きな数字を見る時には、例えば「1 クラス当たり」のように目に見えやすい単位に合わせて割り算をしてみると、統計の数字をより身近なものとして見ることが出来ます。

#### <みんなで考えてみよう>

表 2 の統計を使って日本全国の中学校 1 校当たりのクラス（学級）数を計算して、自分の学校が、全国平均に比べて大きいのか、小さいかなどを考えてみるとよいでしょう。また、この数字が過去からどのように変化してきたか、また、この表からほかにどのようなことが分かるかなどについて考えてみましょう。

## 5 日本にはどんな統計があるの？

### ○ どんな統計を知っていますか？

日本では、社会や経済の実態をとらえるために様々な統計が定期的に作られ、公表されています。そのような統計は、テレビや新聞のニュースなどで毎日のように報道されたり、社会科などの教科書や参考資料などで数字が引用されたりしています。

教科書などでしばしば引用される主な統計には次のようなものがあります。これらは、毎年や5年に1回の周期で数字が公表されるものです。日本全体だけではなく、都道府県別、市区町村別など詳細な地域別の統計が得られるので、自分の住んでいる地域のことを調べるのにも利用することができます。

国勢調査・・・5年ごとに日本全国と地域別の人口・世帯とその内訳を調べる。

工業統計調査・・・毎年、日本全国と地域別の工業の従業者数、出荷額などを調べる。

商業統計・・・5年で2回調査。日本全国と地域別の商業の従業者数、販売額などを調べる。

農林業センサス・・・5年ごとに日本全国と地域別の農家などの就業状態、農業生産、販売などの状況を調べる。

また、次の統計は、小中学校の教科書などに引用されることはあまりありませんが、最新の経済情勢を示す統計として毎月公表され、テレビや新聞などでも報道されます。これらは、会社を経営する人だけでなく、社会で働く多くの人々が注目しています。

消費者物価指数・・・私たちの購入する商品やサービスの物価の変化が分かります。

労働力調査・・・日本中に何人の就業者や失業者がいるかが分かります。また、その内訳（男女、年齢別など）も分かります。

家計調査・・・日本中の家庭の家計の収入・支出の状況が分かります。

鉱工業生産指数・・・日本中の工場などでの生産高の動きが分かります。

以上は、日本の統計のほんの一部で、このほかにも様々な統計が公表されています。日本にどのような統計があるか調べるには、次の資料が便利です。

・「日本統計年鑑」（編集：総務省統計研修所 発行：総務省統計局）

<http://www.stat.go.jp/data/nenkan/index.htm>

・総務省統計局サイト <http://www.stat.go.jp/>

### ○ 何のために国の統計があるの？

日本にある様々な統計は、何のために作られているのでしょうか？国の統計には、いくつかの大切な役割があります。

まず第一に、国民自身が自分の国の状態を正しく知るための役割です。私たちは、社会の一員として、日本の社会がどのような状態にあり、どのような方向に向かっているかを常に正しく知っておく必要があります。私たちの暮らし向きは社会全体の動きによって左右されますので、個人個人が社会の変化に対応してどのように行動したらよいか考えておく必要があります。

また、国の統計を見れば、人々の暮らしが豊かになっているのかどうか、お金や資源が効率的に使われているのかどうか、生活に困っている人がどれくらいいるのか、など国民の暮らしの実態や経済社会全体の動きなどが分かります。国民が自分の暮らしを工夫したり、政治に参加したりする場合には、このような統計や情報が役に立ちます。

つまり、国民それぞれが国の状態を正しく知るためには、それを客観的な数字で示した統計が必要なのです。そこで、政府は、日本の国のことがだれにでも分かるように、国の状態を表す統計を作成して公表しています。

第二に、国や地方の行政の運営を公平・公正に行う基準を与えるための役割です。国や地方公共団体は、国民、県民、市民などを対象にして様々な行政を行っています。その例としては、公共施設の整備、公共サービスの提供、補助金の給付などがよく知られていますが、それを公平に行うためには、正確で信頼できる統計が必要なのです。もしも統計がなかったとしたら、例えばどこに施設を建設するか、だれにどのようなサービスや補助金を提供するかなどといった重要なことが、それを担当する職員の個人的な勘や判断、これまでの慣習などで決められかねません。そのような決め方は公平・公正と

は言えませんし、お金（税金）の使い方として効率的でもないでしょう。

このような問題が起こることのないように、行政を行う上で多くの人々が納得できる客観的な基準として、国が統計を作成しています。

第三に、国際社会の中で日本の置かれた状況を正しく理解するための役割です。日本は、外交、貿易、人的交流など様々な形で国際社会と密接なつながりを持っています。日本の人々が今後も豊かな暮らしを続けていくためには、「世界の中の日本」がどのような状態にあるのか、理解しておくことが必要です。例えば、日本は世界の中で何番目の人口規模でしょうか？どの国からどれだけ天然資源を輸入しているでしょうか？どのような物をどれだけ輸出しているでしょうか？地球温暖化ガスをどれだけ排出しているでしょうか？日本が世界の国々とともに平和に発展していくために、自分の国のことを知ることは大切なことです。また、外国の人たちも、日本はどのような国なのか、どのように変化しているのかということを知りたいがっています。世界の国々がお互いの状況を正しく知ることができれば、無用な誤解や衝突を避けることもできます。統計は、世界の共通語として、日本の国が世界の中で置かれている状況を示してくれるものです。

#### ○ 国の統計はどうやって作るの？

国の統計の作り方には、およそ3種類の方法があります。第一は、統計調査による方法です。第二は、官庁の持っている資料を集計する方法です。第三は、ほかの統計やデータを加工計算して推計する方法です。中でも特に重要なのは統計調査による方法ですので、ここでは第二、第三の方法は省略し、第一の方法について説明します。

統計を作るためには、対象となる個人や世帯（家庭など、生計をともにして一緒に暮らしている人の集まり）に対して、同じ形にのっとしてデータを提供してもらう必要があります。このため、一般には、まず「調査票」という統一的形式の書類を作り、調査員が地域を巡回して対象となる世帯にこの調査票を配って回答を依頼します。調査票を受け取った世帯は、調査員からの説明や付属の資料など参考にして回答を調査票に記入して、それを調査員に提出します。

以上の例は、調査員が世帯を訪問して調査する方法ですが、このほかにも、郵送による調査方法や、新しいものではインターネットを利用したオンライン調査の方法もあります。

正確な統計を迅速に作るためには、調査員による調査が最も有効です。郵送による調査であれば、調査員の労力などは節約できますが、世帯の側で提出を忘れて、提出が遅れたりすることが起こりやすくなります。調査票の提出の漏れや遅れがあると、統計が不正確になったり、公表時期が遅れたりして、統計の利用に差しさわりが出てきます。

みなさんの家庭（世帯）も、国の統計調査の対象となる場合があると思われます。調査票がみなさんの家庭に届けられた時には、それをよく見てみることで、統計調査がどのようなものか、実例としてよりよく理解することができます。

なお、統計調査がどのような手順で行われるかということについて、詳しくは「統計のできるまで」を参照してください。

#### ○ 国勢調査ってどんなもの？

国の統計調査の中で、国勢調査は最も基本となる調査として広く知られています。国勢調査では、日本の人口と世帯（家族など、生計をともにして一緒に暮らしている人の集まり）について最も信頼できる詳しい統計が作成されています。日本という国や地域社会の状況を正しく理解するためには、人口を正確に知ることが何よりも大切です。

また、人口の総数だけではなく、その詳しい内訳として、どの地域にどれだけの人が住んでいるか、また、どのような仕事に就き、どのような家族や世帯を形作って暮らしているかといった実態についての情報も必要です。

そのような統計を得るために行われるのが国勢調査です。

国勢調査は世界の多くの国々で定期的に行われていますが、日本では5年ごとに行われています。調査は、西暦の末尾が0と5の年の10月1日を期して行われ、日本にふだん住んでいる人は日本人でも外国人でもすべて対象となります。今年（2010年）は国勢

調査の行われる年であり、9月最後の約1週間の間に、すべての世帯に調査票が配布されます。この機会をとらえてみなさんの家庭に届く調査票を見てみると、国勢調査がどのようなものか実感できることでしょう。

前回の国勢調査は平成17年（2005年）に行われ、その結果によると、日本の総人口は1億2776万7994人でした。調査結果は、都道府県別、市区町村別、さらに細かい地域区分についての人口も公表されています。また、総人口だけではなく、男女別、年齢別、国籍別、家族構成別、配偶関係（結婚しているか未婚かなど）の別、職業別などいろいろな切り口でみた統計も公表されています。

国勢調査の結果は、様々な法律の中で基準人口として用いられており、「法定人口」と呼ばれることがあります。例えば、地方公共団体の財政の格差を少なくすることを目的として国が地方公共団体に配分する「地方交付税交付金」の配分の基準には、国勢調査による人口が用いられています。また、衆議院議員選挙の選挙区を画定するときにも、国勢調査による人口を基準とすることとされています。

国勢調査の結果は、日本社会の実情を分析し、将来のあり方を考えるためにも広く用いられています。最近、日本の社会は人口減少に転じたと見られていますが、そのような見方も、国勢調査の結果及びそれに基づいて推計された「推計人口」によって分析されたものです。また、日本の将来の人口の見通しを示す「将来人口推計」も、国勢調査の結果に基づいて推計されています。

国勢調査の結果は、現状の確認だけではなく、将来の日本社会の在り方を考え、対策を立案するためにも用いられています。このように、国勢調査の結果は、国でも地方公共団体でも、公平・公正な行政を行うためや、国や地域社会の将来像を考えるためなど、様々な目的で利用されています。

国勢調査の詳しい情報については、総務省統計局の国勢調査のサイトをご覧ください。  
<http://www.stat.go.jp/data/kokusei/2010/index.htm>

#### ○ 国勢調査には必ず答えなければいけないの？

国勢調査は、「統計法」という法律に基づいて行われるものです。統計法には、国勢調査に代表される国の重要な統計調査（法律上、「基幹統計調査」と呼ばれます。）については、調査の対象となる人に対して答える義務（報告義務）があると定められています。

統計調査では、すべての人が漏れなく正確な回答を提出することによって、初めて正確な統計が作成できます。調査対象になった人が、仮に「回答するのが面倒だから」とか「自分一人くらい回答しなくても平気だろう」といった安易な気持ちで回答しない場合には、得られる統計は不正確なものになってしまいます。統計が不正確になれば、行政の判断に誤りや不公平が生じることとなります。

例えば、国勢調査の人口に基づいて、地方交付税交付金が国から地方公共団体に毎年支払われていますが、もしどこかの自治体で調査から漏れている人がいれば、その人数分だけ自治体が受け取る交付金が少なくなり、その地域の行政サービスが低下してしまうかもしれません。

#### ○ 国勢調査では個人の秘密は守られるの？

国勢調査など、「統計法」に基づいて行われる統計調査では、調査の対象となった個人や企業などの秘密は厳格に保護されます。統計法では、調査に携わる調査員や調査を行う総務省統計局などの統計機関の職員に秘密保護の義務を課しており、それに違反した場合の罰則も設けられています。

最近、日本では、振り込め詐欺などで個人情報が悪用される事件がしばしば発生することから、他人に個人情報を知らせることを不安に思う人が増えているようです。しかし、正確な統計を作るためには、統計調査の対象となったすべての人に漏れなく正しい回答を提出してもらう必要があります。そのためには、だれでも安心して回答できるよう、提出された個人情報を保護する万全の仕組みを設けることが必要です。このため、統計法では、個人の秘密をしっかりと保護した上で調査を行うことが規定されています。なお、個人に対する調査だけではなく、企業に対する調査の場合も、企業の秘密

は統計法によって個人情報と同様にしっかりと保護されます。

<みんなで考えてみよう>

国勢調査では、どのようなことがらを調べているのでしょうか。また、調査票とはどのようなものなのでしょうか。国勢調査の結果として、どのような統計表やグラフが公表されているのでしょうか。みんなでインターネットを使って調べてみるとよいでしょう。

## 6 統計はだれのもの？

### ○ みんなで作る統計、みんなで社会に活かす統計

国の統計は、国の統計機関（総務省統計局など）が作成していますが、これは一体だれのものだと言えばよいのでしょうか？国の統計機関は、通常、国の予算に基づいて、各省の大臣の名において統計調査を行います。ということは、統計は国のものということでしょうか？

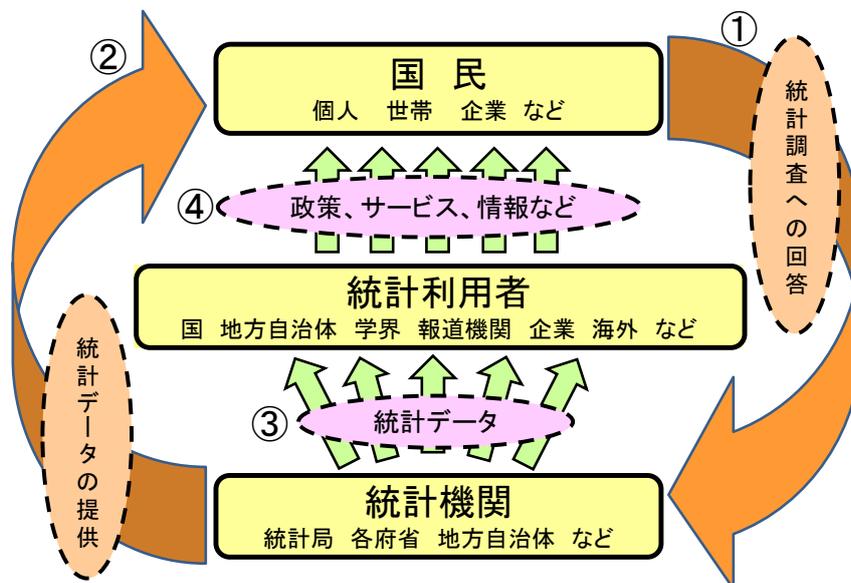
ここで少し見方を変えてみると、統計の基になっているデータは、統計調査により日本中の個人や企業などから集められたものです。ということは、一人一人の個人や企業のものなののでしょうか？

このことは、社会の中で情報がどのようにめぐっているのか、次の図を見ながら考えるとよいでしょう。この図は、統計がどのように作成され、利用されているか、大まかに示したものです。（丸印の番号は、図の中の矢印に対応しています。）

- ① 統計の基になるデータは、国民、つまり個人や世帯や企業（一番上の枠）から統計機関（一番下の枠）に提供されます。
- ② 集められたデータは、統計機関において統計として集計・加工され、世帯や企業に提供されます。つまり、データは、秘密を守った上で、最終的には国民に戻されるという循環ができています。
- ③ 統計データは、一般の世帯や個人にも利用されますが、中央の枠が示すように、統計を専門的に利用している人たちもたくさんいます。例えば、国や地方公共団体は、公平・公正で効率的な行政を行うために統計を利用しています。学界（大学・研究機関など）では、統計を使って研究を行っています。報道機関では、ニュース報道や論説などで統計を利用します。企業では、商品・サービスの企画や販売戦略を作るために統計を使います。海外の利用者は、日本の状況を理解するために統計を利用します。
- ④ 統計を専門的に利用する人たちは、統計から得られた情報を基にして、国民に対してサービスなどを提供しています。例えば、国や地方公共団体は、統計に基づいて行政サービスを国民に提供します。学界では、統計を用いた研究成果を世に発表して、社会に役立つ提言などをします。報道機関はニュースや解説記事を通じて、国民に知識や情報を提供します。企業は、統計を用いて消費者のニーズをとらえ、消費者の求める商品・サービスを提供しています。

このように、統計を利用して得られた結果は、政策、サービス、商品、情報などの形で最終的には国民に還元されるのです。

## 統計は国民の共通財産 — 統計情報は社会をめぐる



以上のように、統計情報は、ちょうど体の中を血液が循環するように、社会の中を循

環しています。国民は、統計調査への回答を国の統計機関に提出し、国の統計機関は、それを統計に取りまとめて国民に還元しています。その統計は、多くの機関や企業、個人に活用されることを通じて、国民にとって豊かで暮らしやすい社会を作るのに役立てられます。

このように見ていくと、統計は社会の共有財産であることが分かります。

人によっては、統計調査に回答することを面倒に感じたり、自分の情報の取扱いについて不安を感じたりする場合もあるのではないかと思います。しかし、この図のように、統計は社会全体の協力によって作成され、社会の共通財産として活用されるものです。統計法では厳格な秘密保護の仕組みも設けられています。統計調査に対して漏れなく正確に回答することがひいては豊かで暮らしやすい社会につながっていくのです。

また、統計調査の結果として得られる統計については、国民一人一人がその見方・使い方をよく身に付け、自分たちの暮らしている社会の実情をよく理解することが必要です。

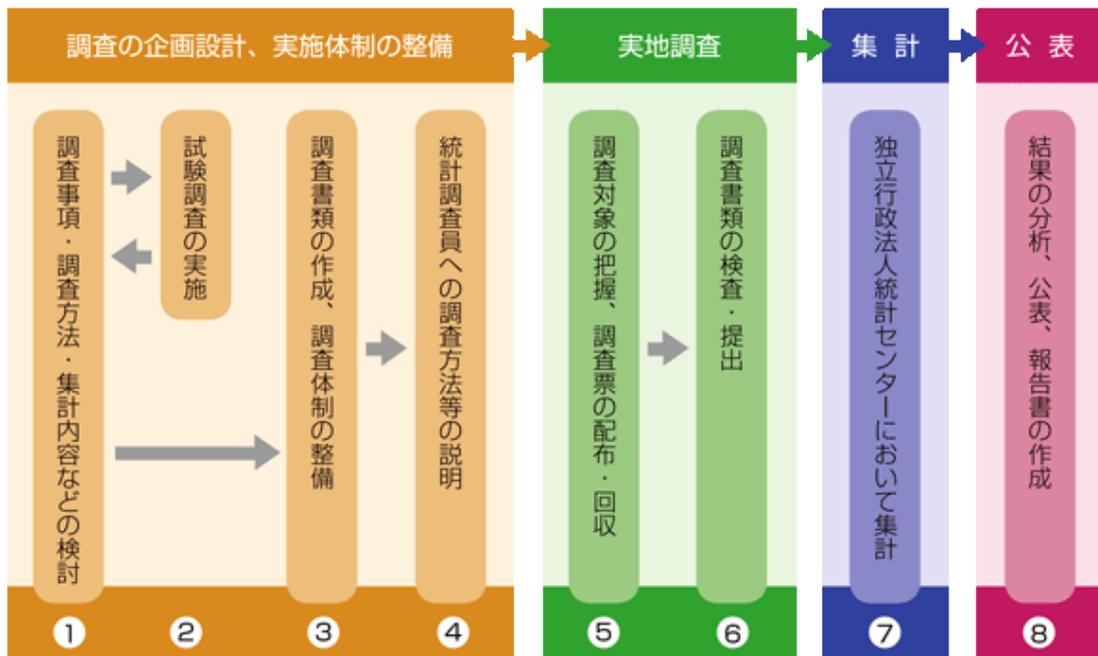
統計は、みんなで作るもの、みんなで社会に活かすものなのです。学校で「統計」について学ぶのは、私たち一人一人が様々な統計から役に立つ情報を導き出す力を身につけるためです。みんなで「統計」を学んで、様々な統計を上手に暮らしに活用しようではありませんか。

#### <みんなで考えてみよう>

日本にある様々な統計を活用するためには、統計についての用語や見方・使い方についてよく学ぶことが必要です。統計局ホームページの中にある「統計学習サイト」を入口として、自分の興味のある統計を見つけ、その見方などを自分で考えてみるとよいでしょう。

## 統計のできるまで

《統計調査の企画から結果の公表まで》流れ図（統計調査員による調査の場合）



### ◇ 調査の企画設計、実施体制の整備（統計局）

#### ① 調査事項・調査方法・集計内容などの検討

調査の性格や目的に応じて調査事項、調査方法、集計内容などについての検討を行います。

調査事項については、回答者が記入しやすいような調査票の設計などについても検討します。

調査方法については、全数調査か標本調査か（標本調査の場合は標本数をどのくらいにするか）、郵送調査か調査員調査か、などを検討します。

#### ② 試験調査の実施

大規模な調査や新しく実施する調査などの場合には、実際の調査に当たって調査票の設計や調査方法などに支障がないかなどをテストするための「試験調査」を実施します。

### ③ 調査書類の作成、調査体制の整備

調査票、記入者向けの説明書や統計調査員向けの説明書類など、調査実施に必要な書類を作成します。また、統計調査員を配置するなど、調査体制を整えます。

### ④ 統計調査員への調査方法等の説明

統計局及び地方公共団体で調査の実施事務について打合せを行った上で、統計調査員に対して調査内容や調査方法などの説明を行い、その徹底を図ります。

## ◇ 実地調査（地方公共団体、統計調査員）

### ⑤ 調査対象の把握、調査票の配布・回収

統計調査員は、所定の時期（期間）に定められた方法によって、調査対象の把握、調査票の配布・記入依頼、調査票の回収などを行い、都道府県又は市（区）町村に調査票を提出します。

### ⑥ 調査書類の検査・提出

都道府県又は市（区）町村は、回収された調査票の記入内容に不備がないかなどを確認した後、調査票を統計局に提出します。

※統計局で実施しているほとんどの基幹統計調査の実地調査については、統計法に基づく法定受託事務として地方公共団体が行っています。

なお、実地調査を民間事業者に委託している統計調査もあります。

## ◇ 集計（独立行政法人統計センター）

### ⑦ 独立行政法人統計センターにおいて集計

提出された調査票は、統計局から独立行政法人統計センターに送付され集計が行われます（詳しくは、「[統計調査の集計について](#)」を参照）。

## ◇ 公表（統計局）

## ⑧ 結果の分析、公表、報告書の作成

集計結果は独立行政法人統計センターから統計局に送付され、統計局において分析した上で、公表（発表）するとともに、結果の概要や統計表を掲載した報告書を作成します。

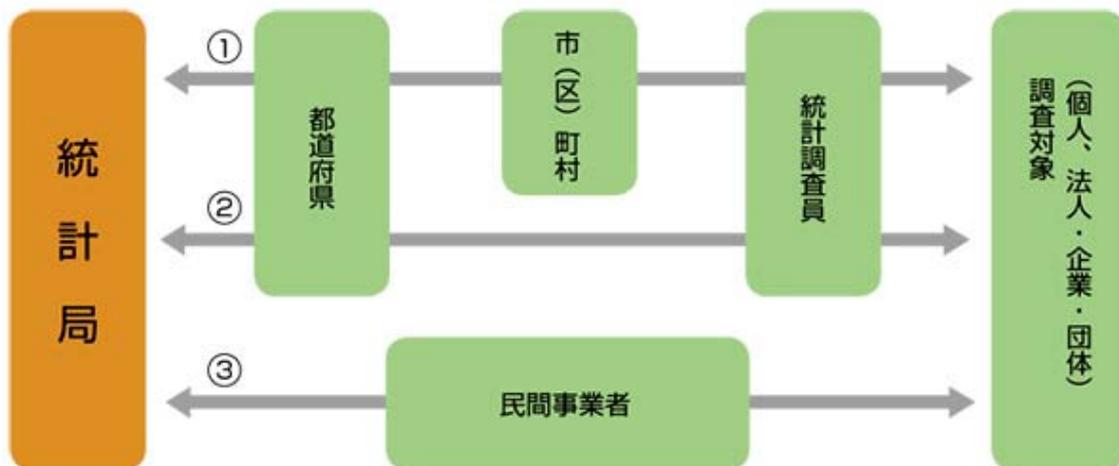
調査結果は統計局のホームページから御覧いただけます。

報告書は都道府県・市（区）町村の統計主管課や主な図書館などで御覧いただけます。

報告書や結果を収録した電磁的記録媒体（MO、CD-R など）は、政府刊行物サービス・センターなどで入手できます（詳しくは、「[統計情報の提供状況一覧](#)」を参照）。

### 調査の流れ

統計局が実施する調査では、調査票など調査関係書類の配布・回収などの調査事務は、通常、次の流れで行っています。



- ① 国勢調査など大規模な調査の場合
- ② 労働力調査、家計調査などの場合
- ③ 科学技術研究調査、家計消費状況調査、サービス産業動向調査の場合

# 標本調査とは？

## ～調査のしくみと設計～

### 全数調査と標本調査

#### ・全数調査とは

統計調査によって何かを調べたい時、例えばある中学校で全校生徒の平均身長を調べたいと思ったら、生徒全員の身長を測って平均を計算すれば正確な結果が得られます。このように、対象となるすべてを調べる調査を「**全数調査**」といいます。一つの中学校の全生徒の身長を調べることは、それほど大変な手間ではないでしょうが、日本中のすべての中学生の身長を調べるのは大変な手間と費用がかかります。このような場合には、手間や費用を省くために、一部の人だけを選んで調べる方法もあります。このような調査を「**標本調査**」といいます。

全数調査は、集団の中をすべて調査しますので、集計した結果には、標本調査では必ず生ずる「**標本誤差**」が含まれません。したがって、全体の結果はもちろんのこと、男女別の結果や詳細な地域別の結果なども統計として利用できます。

我が国で実施されている最も大規模な全数調査は**国勢調査**ですが、国勢調査では、全国結果だけでなく、都道府県別、市町村別はもちろん、町丁・字などの小地域の結果や、男女別・年齢別などの詳細な結果も公表されています。

#### ・何のために全数調査が必要なのでしょうか？

全数調査の結果には、上述のように標本誤差が含まれませんので、例えば国勢調査の結果は、衆議院議員の小選挙区の画定基準や地方議会の議員定数の決定、地方交付税の算定基準など、「法定人口」としてさまざまな場面で利用されます。

また、全数調査は、統計体系の中で、定期的に他の統計の基準となる数値を提供するという、統計の正確性を担保する根源となっています。具体的には、例えば5年ごとに行われる国勢調査から作られる統計は、日本の現在人口と将来人口の推計の基礎とされており、国勢調査の結果が得られるごとに、これらの推計の基礎が見直されています。このような基準とされる数値のことを、「ベンチマーク」と呼んでいます。また、**国民経済計算**などの加工統計（統計調査の結果だけでなく、さまざまなデータを集めて加工して作成する統計）でも、推計の基準とされる人口は国勢調査を基にしていることから、同様に国勢調査の結果が得られるごとに改定が行われます。

もう一つの重要な役割は、標本調査を設計する際の基礎としての役割です。

標本調査は、ある集団の中から一部の対象だけを抽出して調査するもので、我が国で行われている統計調査の多くはこの方法により行われています。標本調査により正確な統計を作るためには、調査対象とする集団全体の数や性質がわかっていないといけません。したがって、標本調査の企画・設計の段階で、その集団に含まれる対象すべてを含んだリストが必要になります。

例えば、経済関係の統計調査では、多くの場合、全国の事業所・企業の中から一部のものを抜き出して調べます。その場合に必要となる事業所・企業のリストは、多くの場合、全数調査である事業所・企業統計調査（平成21年から経済センサス）によって整備されたリストが用いられます。つまり、全数調査は、他の多くの標本調査を実施するために必要不可欠な情報を与えるものとしても活用されています。

このように、全数調査は、その統計自体にも重要な役割がありますが、他の統計調査を行うための基礎にもなっており、統計調査の体系の中で中心的な役割を果たしているのです。

### ・何のために標本調査が必要なのでしょう？

---

しかし、すべての統計調査で対象者全員を調査した場合、膨大な費用と手間がかかります。身近な例の中から、テレビ番組の視聴率を例に挙げてみましょう。テレビや新聞などで、「昨年末の紅白歌合戦の視聴率は〇〇%だった」とか「先週1週間で最も視聴率が高かった番組は□□□だった」などの報道を目にしますが、皆さんの中に視聴率の調査を受けたことのある方はほとんどいないと思います。視聴率は、対象となる地域（例えば関東地方）の全世帯を対象に調査すると莫大な経費を必要としますし、結果をすぐに（例えば次の日などに）出すことができませんので、このような場合、一部の世帯だけを統計的に偏りがないように選んで調査をします。これを「**標本調査**」といい、選ばれた調査対象を「**標本**」（**サンプル**）といいます。

標本調査は、選ばれた一部の標本を対象に調査を行い、すべての対象を調べるわけではありませんので、その結果には誤差（これを「**標本誤差**」といいます）が含まれますが、標本の選び方を工夫すれば、誤差をあまり大きくすることなく、調査にかかる費用と期間を大幅に縮減することができます。したがって、多くの場合、統計調査は標本調査により行われています。

総務省統計局が実施している、例えば労働力調査や家計調査、社会生活基本調査やサービス産業動向調査などのほとんどの統計調査は、標本調査により行われています。

## 標本調査

### ・標本調査を行う前に

---

標本調査は、ある集団の中から一部の調査対象を選び出して調べ、その情報を基に、元の集団全体の状態を推計するものです。調査の対象とされている「元の集団全体」のことを「**母集団**」と呼びます。

標本調査の目的は、標本を用いて母集団の状況をできるだけ正確に復元推計することです。正確な推計結果を得るためには、標本が母集団全体の特徴をよく表したものになるように、つまり、母集団のよい縮図となるように抽出することが大変重要です。

### ・有意抽出法と無作為抽出法

---

標本の選び出し方には、「**有意抽出法**」と「**無作為抽出法**」とがあります。

有意抽出法は、「代表的」あるいは「典型的」と考えられる調査対象を抽出する方法です。この方法によれば、調査の企画者が様々な情報を基に、よく考えて調査対象を抽出すれば、母集団のよい縮図となる標本を選べるのでは

ないか、と思われるかも知れません。しかし、実際には、この方法は主観的な判断に頼ることになり、この方法によって集団全体を代表する標本を選ぶことは極めて困難です。有意抽出法の大きな問題は、選ばれた標本が母集団のどのような部分を代表しているのか、統計的に評価ができないことです。

これに対して無作為抽出法は、調査の企画者の主観的判断を排除して、標本をくじ引きのような方法で、無作為に抽出する方法です。くじ引きのような確率的な出来事に関しては、確率論など統計数理の理論を当てはめて、標本から全体の数値を推計したり、推計結果の誤差を評価したりすることができます。無作為抽出によってある程度多数の標本を集めると、推計結果は安定し、精度の高い推計結果が得られます。

### ・無作為抽出の手順

---

無作為抽出を行うには、実際にどのようにやればよいのでしょうか。

「無作為」という言葉から、単に「でたらめ」に選ばばいい、と考えられがちですが、本当の意味で無作為に選ぶためには、きちんとした手順を踏む必要があります。

無作為抽出の方法として最も典型的なものは、次のとおりです。

まず、母集団を構成する全員のリストを用意します。このような母集団全員を含むリストは、「標本フレーム」または「標本枠」と呼ばれることがあります。

次に、リストに含まれる全員に一連の通し番号を付けます。そして、その中から、乱数表などによって得た乱数に従って、調査対象を選び出します。

こうして得られた調査対象を調査すれば、無作為抽出による標本調査となります。

### ・「無作為抽出」＝「でたらめ」ではない

---

では、単に「でたらめ」に調査対象を選んだ場合には、どんな問題があるのでしょうか。

ここで一つの事例を考えてみましょう。ある町で、住民を対象に標本調査を行うとします。「でたらめ」に調査対象を選ぶために、簡単な方法として、その町の一番大きな駅の前立って、通行する人に誰でもかまわず無作為に(でたらめに)声をかけて、調査してみたとしみましょう。

この方法は、無作為抽出法となっているのでしょうか。

この方法は、次のような理由から、これは無作為抽出法となりません。

駅前を歩いている人は、町の住民すべてを代表しているとは限りません。住民の中には、その駅を使わない人もいます。また、調査をしている時間帯に、住民すべてが駅前を歩いているわけではないので、ほかの時間帯にしか出歩かない人は調査されません。さらに、忙しい人は、呼びかけても立ち止まって調査に協力してくれないかも知れませんが、立ち止まって協力してくれる人は、時間に余裕のある人だけかも知れません。

つまり、この方法で調査に回答してくれる人たちは、その町の住民全体の中から選ばれたというよりは、むしろ、その町の住民のうちで、その駅を利用する人で、かつ、調査の時間帯に駅前を歩いている、さらに、立ち止まって答えてくれた人、ということになります。したがって、調査に協力してくれる人は、町の住民の中でも、上述のような特定の特徴を持った人たちと考えられます。そのような人たちが、その町の実態を反映した縮図になっているとは言えません。

したがって、このような方法で統計調査を行っても、その結果が何を意味するのか、わからないものとなってしまいます。

## ・集落抽出法

---

母集団を構成する全員のリストがあれば上述のような標本抽出が可能ですが、多くの場合、このようなリストは存在しません。すべての調査単位を含むリストが利用できない場合、どのようにして標本の抽出を行えばいいのでしょうか。

このような場合に用いられる代表的な方法として、「**集落抽出法**」と呼ばれる方法があります。これは、個々の調査単位を直接選ぶのではなく、まず調査を行う地域を選び、次に、その地域の中に含まれる調査単位をすべて調べる方法です。例えば、世帯を対象とした標本調査を行う場合、全国の全世帯のリストは存在しませんので、国勢調査の調査区の中から、調査の対象とする調査区を一定数、無作為抽出法により選び出し、その調査区の中のすべての世帯を調査するのです。

このように、調査対象を地域のまとまり(集落)ごとに抽出することから、「集落抽出法」と呼ばれています。

この集落抽出法では、国勢調査の調査区のように、調査の対象となる地域をもれなく重複なくカバーしている、区画の明確な地域の単位を用いることが必要です。国勢調査の調査区は、平均的に約50世帯が含まれるように日本全国を区画して作られ、全国に約90万の調査区が設定されています。各調査区に関しては、調査区地図が整備され、境界が明確になっていますし、国勢調査の結果によって、各調査区に関する統計が整備されていますので、国勢調査の調査区は、集落抽出法における抽出の対象として用いられることがしばしばあります。

## ・二段抽出法

---

母集団に含まれるすべての単位のリストが利用できない場合に、集落抽出法以外でよく用いられているのが「**二段抽出法**」です。

二段抽出法は、最初に調査地域を無作為に抽出するところまでは集落抽出法と同じですが、その先が異なります。調査地域内をすべて調べるのではなく、その地域の中から一部を無作為に選び出して調査するものです。この場合、すべての調査対象のリストが存在しなくても、選ばれた調査地域内の調査対象のリストさえ作れば、標本を抽出することができます。

このように、一段目では調査地域を選び出し、さらに二段目では世帯などの調査対象を選び出すという2段階の方法で標本を選ぶので、「二段抽出法」と呼ばれます。(同様に「三段抽出法」などもあり、これらを総称して「多段抽出法」と呼ぶこともあります。)

同じ地域に特徴の似ている世帯が集まっている場合には、調査区内のすべての世帯を調べる代わりに、その中の一部の世帯を抽出して調査しても、それほど調査結果の精度は低下しません。このため、二段抽出法によれば、1地域にかかる調査費用を節約することができ、そうして節約された経費を活用して、調査地域の数を増やして全国に広く散らばらせることができるので、同じ費用で調査結果の精度を向上させることが可能です。

## ・標本抽出の具体例

---

以上で説明した標本抽出の具体例を、総務省統計局が実施している労働力調査を例にとって解説します。

労働力調査では、下図のように、一段目に国勢調査の調査区を、二段目に調査区内にある住戸を抽出する二段抽出法によって標本を抽出しています。

(ここに「労働力調査における標本抽出方法の概要」の例の図を入れる)

## 標本誤差

### ・標本誤差と非標本誤差

---

統計調査の結果には、必ず何らかの誤差が生ずることは避けられません。例えば、標本調査では、調査されなかった調査対象がありますので、全数調査を行えば得られたはずの値(これを仮に「真の値」と呼ぶことにします)と調査結果には差が生じます。また、全数調査を行ったとしても、回答者が回答誤りをしたり、回答をしなかったりすることにより生ずる誤差もあります。このうち、全数調査を行わずに標本調査を行ったことにより生ずる差のことを「**標本誤差**」といい、それ以外の、例えば誤回答や未回答による誤差を「**非標本誤差**」といいます。

非標本誤差には、上述の誤回答や未回答によるもののほか、標本が正しく母集団の縮図となっていなかったことによる誤差、集計の際の誤りによる誤差など、いろいろな要因によるものがありますが、このような誤差は、どの程度の誤差が発生しているのか、数字で評価することができません。したがって、調査の設計の際には細心の注意を払ってなるべく起こらないようにすべきです。例えば回答者の回答誤りについては、誤解が生じにくいように調査票を設計するなどの工夫が必要です。

一方、標本誤差は、標本調査には必ず存在する誤差ですが、その誤差の大きさを数学的に評価することが可能です。詳しくは次項以降で解説します。

### ・標本誤差とは？

---

標本調査では、調査対象を無作為に抽出して調査をしますので、どの対象が選ばれるかは偶然によって左右されます。このため、標本調査の結果は必ずしも母集団の値、すなわち真の値とは一致せず、何らかの差があります。このように調査対象の一部を選定することによって起こる、真の値と調査結果との差を「**標本誤差**」といいます。

標本調査を行うときは、この標本誤差の存在を忘れてはなりません。標本誤差がどのくらいになるかを予測した上で標本の大きさ(標本に含まれる調査対象の数)などを決定する必要があります。

また、標本調査を行う場合だけでなく、調査結果を見る際にも標本誤差の存在を忘れてはなりません。調査結果を見る人に統計を正しく利用してもらうためにも、標本誤差を測っておく必要があります。

### ・標本誤差を測るには？

---

標本誤差を測るときには、次のことに注意する必要があります。

一つは、標本調査を行うときには、普通、真の値は分からないということです。(この値が分かっていたら、標本調査をする必要がありません！)このため、真の値が得られていることを前提にして標本誤差を測ることはできません。

したがって、標本誤差(真の値と調査結果との差)を直接的に測ることは困難であり、現実には、一定の理論式に基づいて標本誤差の大きさを推定する方法が一般的です。

また、標本誤差は抽出された標本ごとに異なります。もし、同じような標本調査を何度か繰り返して行くとすれば、調査のたびごとに異なる対象が選ばれ、そのたびに調査結果は変動します。そこで、標本誤差を表すには、仮に同じ標本調査を何回も繰り返し行ったとした場合、真の値との差が確率的にどの程度ばらつくか、そのばらつき具合を示す数値が用いられます。

したがって、標本誤差は、確率的なばらつきの幅を示す尺度で表されます。

標本誤差は、調査結果を見る際に、どの程度確かな数字であるかを判断するための情報で、例えば「この調査結果と真の値との差は95%の確からしさで〇〇円以内である」などと表されます。

なお、標本誤差とはこのような性質のもので、推定結果に何かの数値を足したり引いたりして、修正できるようなものではありません。このため、もしも誰かに、「標本誤差があるならば、その誤差を修正した値を出してもらえないか」と頼まれても、修正することはできません。

### ・どのくらいの標本を調べればいいのですか？

先に述べたように、標本調査を行うときには、どのくらいの標本誤差が見込まれるかを予測し、それを踏まえて標本の大きさを決めて調査を設計する必要があります。

まずは簡単な例で、10,000世帯が住んでいる町でパソコンの普及率を標本調査により調べる例を考えてみましょう。詳しい説明は省略しますが、この場合、標本誤差は、95%の確率でおおよそ以下の値以下になります。

$$2 \times \sqrt{\frac{p(1-p)}{n}} \quad (\text{ただし、} n \text{ は標本に含まれる調査対象数、} p \text{ はパソコンの普及率})$$

この数式は、 $p=0.5$ のとき、すなわちこの町のパソコンの普及率が50%だった場合に最大値をとり、 $\sqrt{\frac{1}{n}}$  となりますので、仮に100世帯を調査対象とした場合、「調査結果と真の値との差は95%の確率で0.1以内にある」と予測できることとなります。調査結果によるパソコンの普及率が50%だった場合、真の値は95%の確率で40%~60%の間にある、ということになります。

したがって、調査結果に求める精度がこのくらいでよければ、標本となる世帯を100世帯選んで調査すればよい、ということになります。

ここで注目していただきたいのは、上の式は母集団の大きさには関係しない、ということです。つまり、1万世帯の町であっても、100万世帯の大都市であっても、100世帯を調べた時の標本誤差は(ほとんど)同じになる、ということです。ここでも詳しい説明は省略しますが、母集団の大きさが相当大きく、それに対して調査を行う標本の大きさがそれほど大きくない場合(注)は、標本誤差の大きさは母集団の大きさとは関係なく、標本の大きさだけによってほとんど決まります。

(注)母集団の大きさに対する標本の大きさの割合が高い場合(例えば40人のクラスから20人を選んで調査をしようとした場合など)には、上に挙げた式の形がもう少し複雑になり、標本誤差は上式で計算した結果よりも小さくなります。

参考までに、上のような例の場合の標本の大きさと標本誤差の関係を以下に示します。

標本の 大きさ	調査結果が50%だった場合(注1)に 真の値が95%の確率で存在する範囲(注2)
100	50.0%±10.0、すなわち40.0%～60.0%
200	50.0%±7.1、すなわち42.9%～57.1%
500	50.0%±4.5、すなわち45.5%～54.5%
1000	50.0%±3.2、すなわち46.8%～53.2%
2000	50.0%±2.2、すなわち47.8%～52.2%
3000	50.0%±1.8、すなわち48.2%～51.8%

(注1) 調査結果が50%ではなかった場合、この幅は上例よりも小さくなります。

(注2) 母集団に対する標本の大きさの割合が高い場合には、この幅は上例よりも小さくなります。

# 労働力調査における標本抽出方法の概要

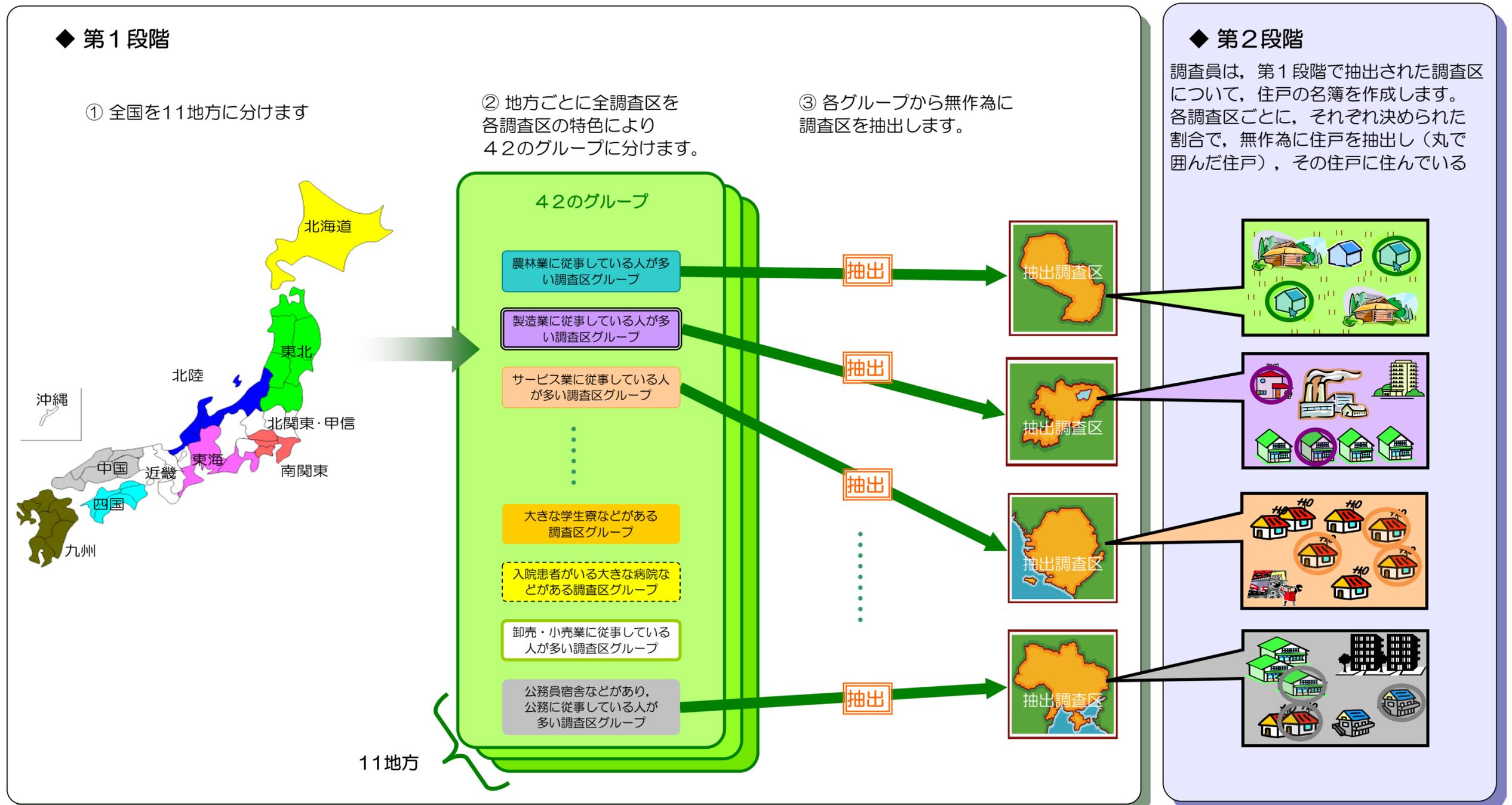
労働力調査は、毎月、全国の約4万世帯に居住する15歳以上の者約10万人を対象に実施している標本調査です。労働力調査では、二段抽出法により、次のように標本を抜き出しています。その結果、全国の正しい縮図になるように、標本世帯が選ばれます。

◆第1段階（「調査区」の抽出）

全国を網羅している国勢調査の調査区を用います。各調査区は約50世帯を含んでおり、実地調査の面からは一人の調査員が受け持つのに適切な大きさとなっています。

◆第2段階（「住戸」の抽出）

調査区内にある住宅の各戸で、一つの世帯が居住できる建物又は建物の一区画（「住戸」）を第2次抽出単位とします。



## 統計をグラフにあらわそう

統計は、データを集めて集計しただけでは、単なる数字の集まりであり、そこから何が読み取れるか必ずしも明らかではありません。

統計を作成するときは、必ず、「○○について知りたい!」という目的があるはずですから、得られた結果を、その目的に合わせて上手に使うことが重要です。グラフは、結果を視覚的に表す便利な道具であり、グラフをうまく使うことによって、自分の考えていることを相手に的確に伝えることができます。

グラフにはいくつかの種類があり、それぞれ、得手・不得手があります。自分が伝えたい目的に応じて、適切なグラフを使うことにより、説明力もぐっと高まります。ここでは、そういったグラフの種類やそれぞれの用途、注意点について説明します。

グラフの種類	主な用途
棒グラフ	棒の高さで、量の大小を比較する。
折れ線グラフ	量が増えているか減っているか、変化の方向をみる。
円グラフ	全体の中での構成比をみる。
帯グラフ	構成比を比較する。
ヒストグラム	データの散らばり具合をみる。
レーダーチャート	複数の指標をまとめてみる。
散布図	2種類のデータの相関をみる。
箱ひげ図	データの散らばり具合をみる。
三角グラフ	3つの量からなる構成比をみる。

### ○量の大小をあらわすときには、どのグラフを使えばいいですか？

→棒グラフを使えば、量の大小を、棒の高さであらわすことができます。

### ○増えている、減っている、といったことをあらわすには、どのグラフを使えばいいですか？

→折れ線グラフを使えば、線の傾きで、増減をあらわすことができます。

### ○割合をあらわすときには、どのグラフを使えばいいですか？

→円グラフや帯グラフを使います。三角グラフを使うこともあります。

### ○データの散らばり具合をあらわすときには、どのグラフを使えばいいですか？

→ヒストグラムや箱ひげ図を使います。

### ○複数の指標をまとめてあらわすときには、どのグラフを使えばいいですか？

→棒グラフや折れ線グラフでもあらわすことができますが、レーダーチャートを使えば、平均と比較したときにどの項目が大きいかな等をわかりやすくあらわすことができます。

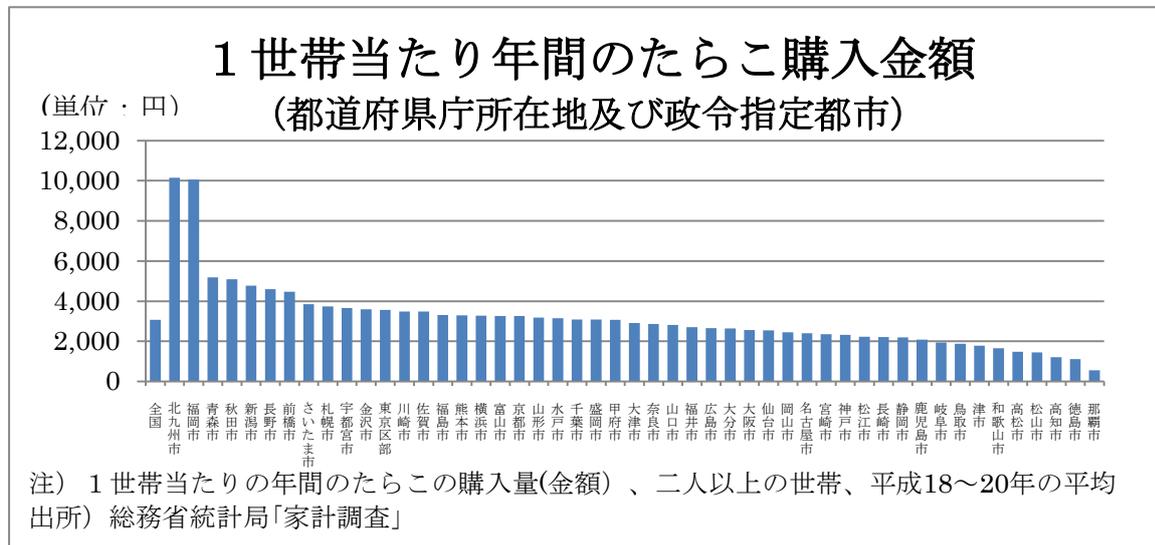
### ○2つの量に関係があるかどうかをあらわすときには、どのグラフを使えばいいですか？

→散布図を使います。

## 1 棒グラフ

### ○たとえば・・・

日本列島は東西・南北に長いため、どんなものを食べているか、地域によってかなり特色があります。ここでは、総務省統計局「家計調査」を使って、たらこの消費量にどのような地域特色があるか、みてみましょう。



これをみると、福岡県の北九州市、福岡市でのたらこ購入量が飛び抜けて大きいことが一目でわかりますね。福岡は「辛子明太子」で有名ですから、ご家庭でたらこを買う量も多い、ということでしょう。(「世帯」という言葉は耳慣れないかも知れませんが、「いっしょに住んでいる家族のこと」と思って下さい。)

### ○どんなときに使うの？

棒グラフは、縦軸にデータ量を取り、棒の高さでデータの大小を表したグラフです。(稀に縦横が逆の場合もあります。)

データの大小が、棒の高低で表されるので、データの大小を比較するのに適しています。

上記の例だと、北九州市、福岡市でのたらこ購入量が他と比べてどんなに多いか、一目でわかりますよね。

### ○気をつけることは？

棒グラフを描くときに、データをどの順に並べるかは、とくに決まりはありません。ただし、あまり意味もなく並べてもグラフが見にくいので、

- ① データの多い順(または少ない順)に並べる。・・・上記の例
- ② 都道府県データの場合、北から順に並べることもある。
- ③ 五十音順に並べる。
- ④ 横軸が年や月といった時間軸である場合、時間の順に並べる。
- ⑤ 質問に対する回答を並べる場合、質問票の順と同じ順に並べる。

などの方法をとります。

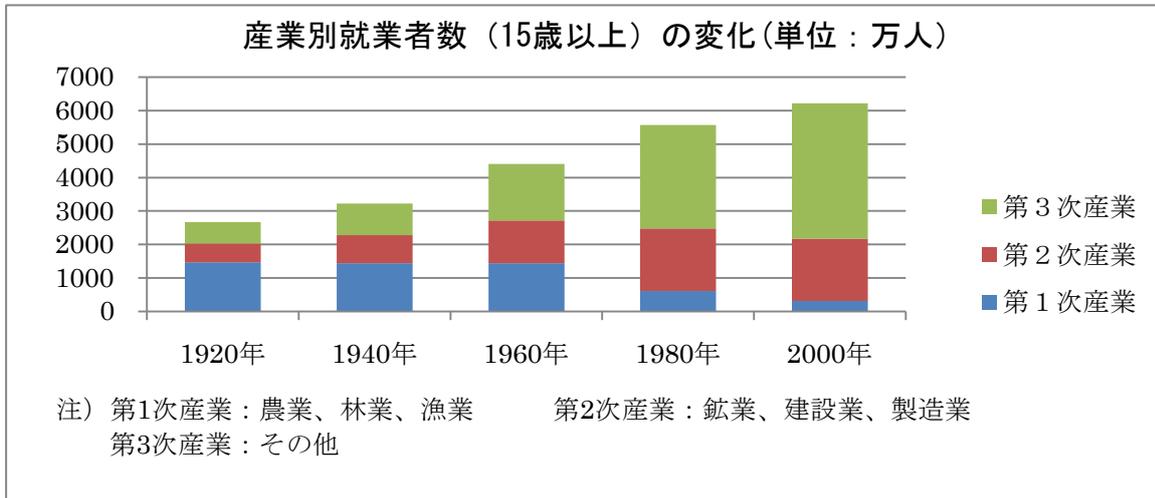
## ○その他には？

先ほど紹介したのは、棒グラフのうちでも、もっとも単純なものですが、もっと複雑な棒グラフもあります。

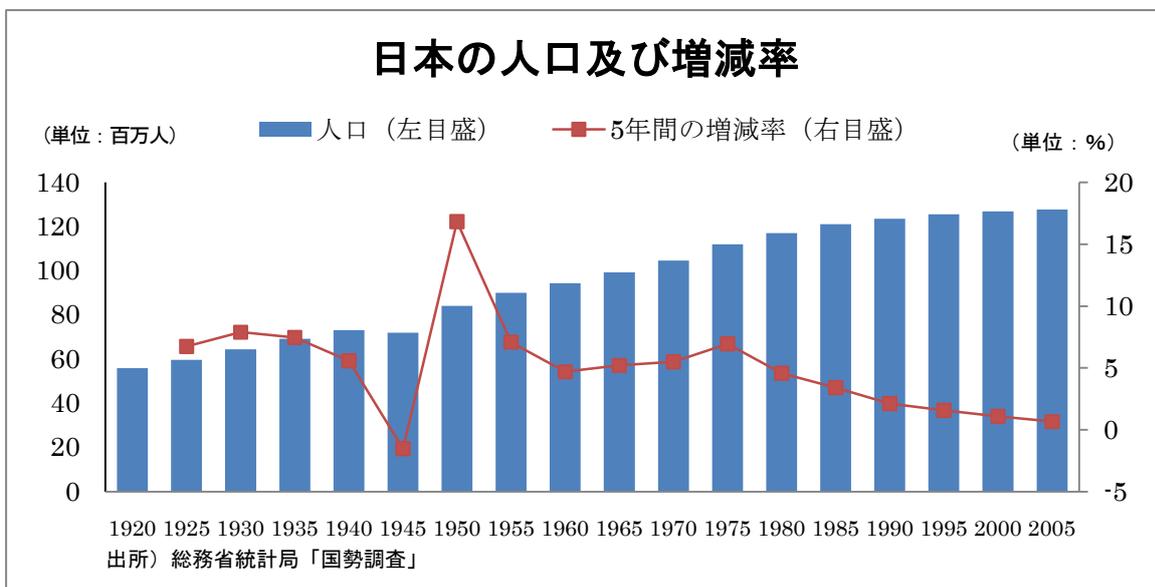
そういった複雑なグラフのいくつかを紹介しましょう。

### ①積み上げ棒グラフ

1本の棒に、複数のデータを積み上げて表示したもの。



### ②棒グラフと折れ線グラフの複合グラフ



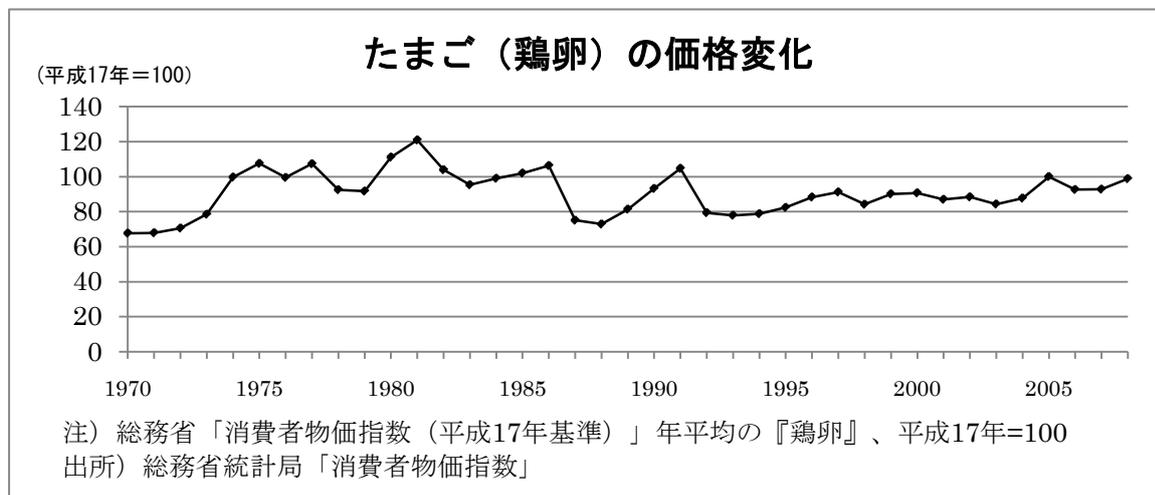
## ○やってみよう

総務省統計局「家計調査」では、たらこ以外の品目についてもデータを掲載しているので、それらについてもグラフを描いてみましょう。<http://www.stat.go.jp/data/kakei/5.htm>

## 2 折れ線グラフ

### ○たとえば・・・

たまごは、昔は「物価の優等生」と呼ばれ、様々なものの値段が変動する中で価格が安定していたのですが、最近では、「たまごの値段が上がった」というニュースも見かけるようになりました。たまごの値段の変化を、総務省「消費者物価指数」で見してみましょう。



ここでは、たまご（鶏卵）の価格の年平均をグラフにしました。ここで使った「消費者物価指数（平成17年基準）」は、平成17年の平均価格がちょうど100になるように計算されています。

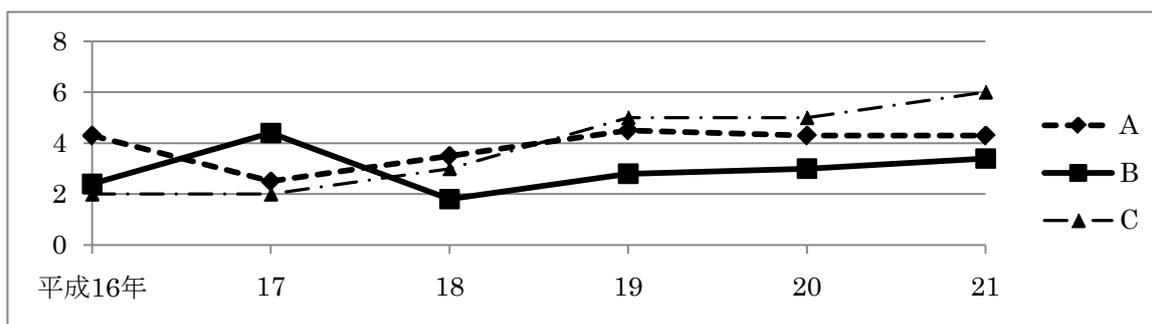
これをみると、たまごの値段もかなり上下動していることがわかります。ニワトリの飼料は輸入穀物に頼っていること、養鶏場の暖房等の費用もかかることから、たまごの価格も、国際的な穀物価格や原油価格の動向に左右されやすい、という側面もあります。

### ○どんなときに使うの？

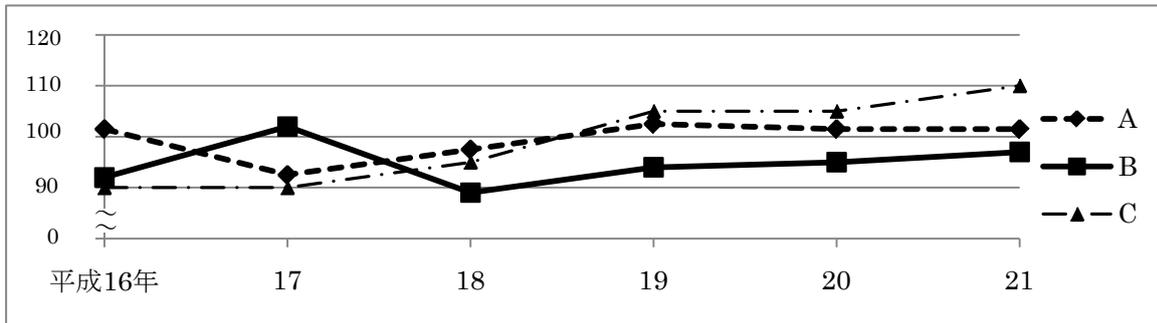
折れ線グラフは、横軸に年や月といった時間を、縦軸にデータ量を取り、それぞれのデータを折れ線で結んだグラフです。線が右上がりならその期間はデータが増加（上昇）、右下がりならデータが減少（下降）していることになるので、データの増減を見るのに適しています。

### ○気をつけることは？

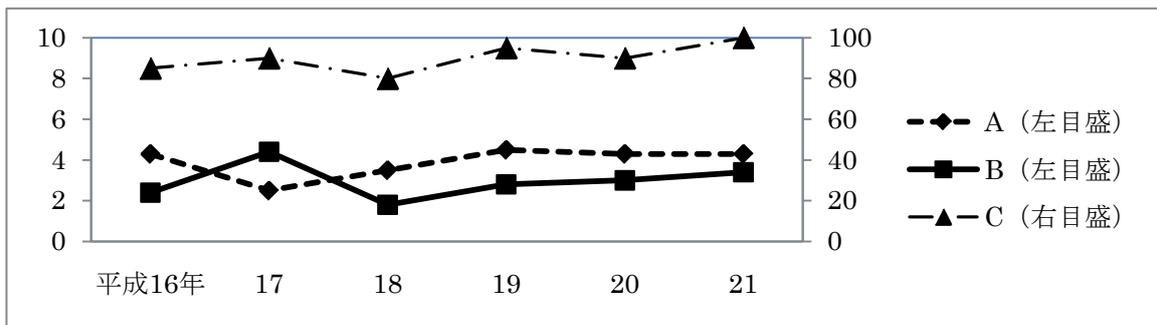
折れ線グラフでは、複数のデータを一つのグラフに重ねて描き、「こちらのデータは増加しているが、こちらは減少」という比較をすることがありますが、その場合、線の区別が付きやすいよう、線の色分けしたり、実線と破線を使い分けたりします。



また、折れ線グラフの縦軸の目盛りは、ゼロから始めるのが普通ですが、あまり数字に変化がなくて見にくいときは、縦軸に波線を入れて途中を省略することもあります。



また、Cだけ数値が大きく違って、一つのグラフに描きにくい、という場合があります。その場合は、Cだけ右目盛りで表示する、ということにして、一つのグラフに収めることもできます。



### ○やってみよう

総務省統計局「消費者物価指数」では、その他の様々な品物の価格を調べています。

たまご以外のいろいろな品物やサービスの価格について調べ、グラフにしてみましょう。

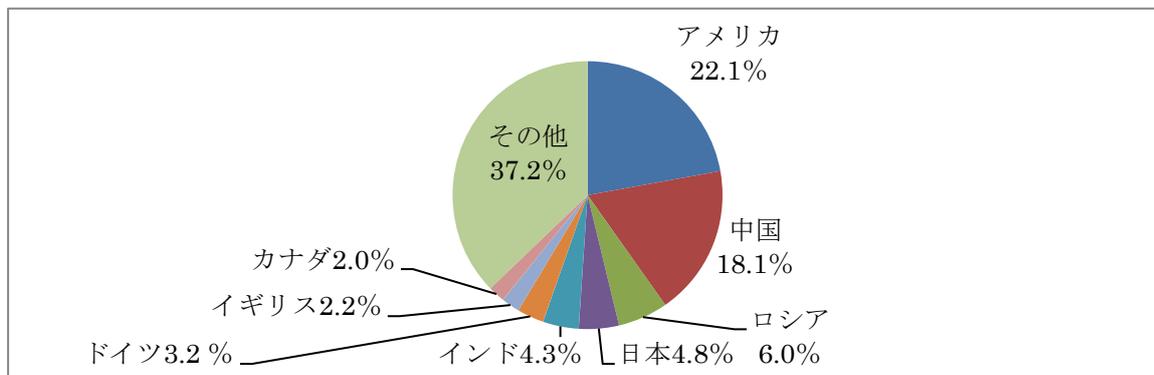
<http://www.stat.go.jp/data/cpi/historic.htm>

### 3 円グラフ

#### ○たとえば・・・

最近、「地球温暖化防止のため、二酸化炭素（CO<sub>2</sub>）排出量を減らそう」「日本が率先してCO<sub>2</sub>削減に努力すべきだ」とのニュースをよく見かけます。いったい、どの国がどのくらい、二酸化炭素を排出しているのでしょうか。

世界のCO<sub>2</sub>排出量(2004年)



出所) エネルギー・経済統計要覧

アメリカと中国のCO<sub>2</sub>排出量が飛び抜けて多いこと、上位4か国（アメリカ、中国、ロシア、日本）で全体の半分以上を占めている、等がわかります。

#### ○どんなときに使うの？

円グラフは、円を全体として、その中に占める構成比を扇形で表したグラフです。扇形の面積により構成比の大小がわかるので、構成比を示すのに使われます。

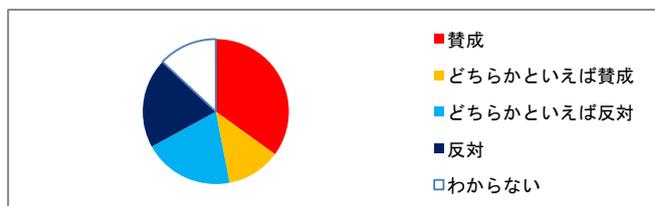
#### ○気をつけることは？

データは、時計の針の12時の位置から時計回りに、大きい順に並べます。

「その他」はいくら大きくても、一番最後に表示します。

あまり扇形が小さくなるとグラフが見にくくなるので、構成比が小さいものは、まとめて「その他」にしてしまった方がよいでしょう。

なお、「〇〇に賛成ですか」のような質問に対する回答を円グラフに表示する場合など、必ずしもデータの大きい順に並べないこともあります。「賛成」と「どちらかといえば賛成」を合わせて過半数かどうか、ということが一目でわかるようにするのが便利だからです。



円グラフを描くときは、合計が100%になるようにします。上の例で「わからない」を除いてしまうと合計が100%にならなくなりますが、それで円グラフを描いてはいけません。

また、複数回答を認める質問の場合、合計が100%を超えますが、このような場合も円グラフにはしてはいけません。

## ○やってみよう

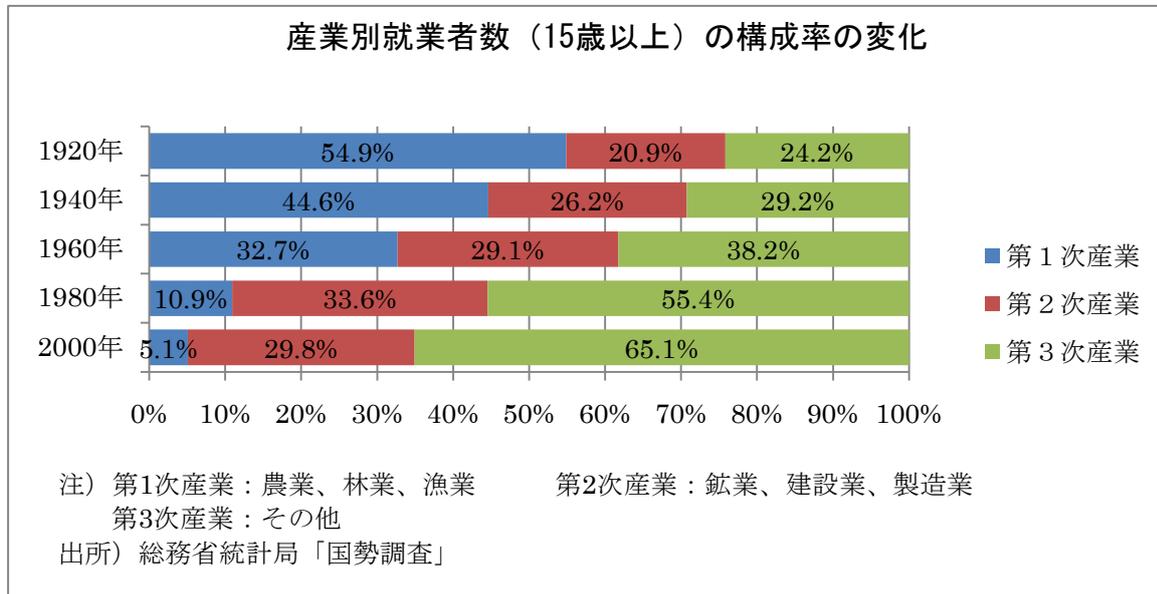
総務省統計局「家計調査」では、ご家庭で、どのようなものを買っているか、調べています。円グラフを描いて、どのようなものにお金を使っているか、見てみましょう。

<http://www.stat.go.jp/data/kakei/index.htm>

## 4 帯グラフ

○たとえば・・・

「最近、農業をする人の割合が減ってきた」との話を聞きますが、総務省「国勢調査」で、第1次産業・第2次産業・第3次産業別の就業者数の割合の変化をみてみましょう。



ここでは、仕事をしている人の数ではなく割合を知りたいので、棒の長さをそろえ、それぞれでの構成比をみてみました。1920年では第1次産業で働いている人が全体の半分以上であったのに、この割合は低下を続け、現在では5%程度であることがわかります。

○どんなときに使うの？

帯グラフは、長さをそろえた棒を並べ、それぞれの棒の中に構成比を示すことによって、構成比の比較をするためのグラフです。

○気をつけることは？

構成比をみるのが目的なので、棒の長さは全て同じにします。

また、項目を並べる順番を途中で変えると、割合の変化がグラフを見てわからなくなってしまうので、一つのグラフの中では順番は変えないでおきます。

(上の例では、途中から第3次産業の就業者数が一番多くなっていますが、多い順に並べることはせずに、すべての年次にわたって第1次産業・第2次産業・第3次産業、の順に並べています。)

○やってみよう

総務省統計局「社会生活基本調査」では、1日の時間の使い方（睡眠、食事、学業、仕事、家事、趣味、等）を調べています。

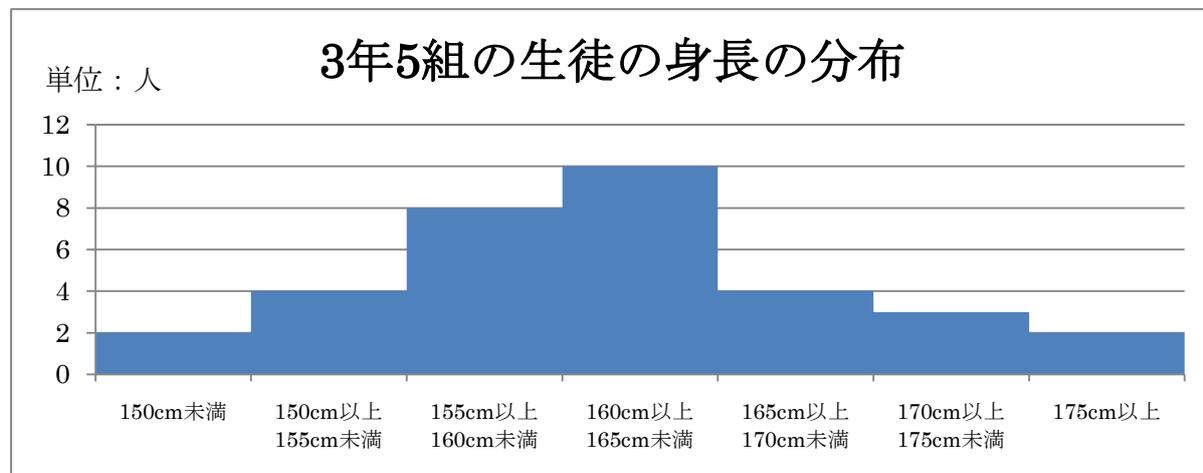
年代別に、時間の使い方によどのような違いがあるか、帯グラフを描いて比べてみましょう。

<http://www.stat.go.jp/data/shakai/2006/pdf/youyaku2.pdf>

## 5 ヒストグラム

### ○たとえば・・・

クラスの中にも、背の高い人、低い人いろいろいます。どれくらいの身長の人が何人いるか、グラフで示してみましょう。



160cm 以上 165cm 未満の人が一番多いことがわかりますね。

### ○どんなときに使うの？

ヒストグラムは、データの散らばり具合をみるのに使われます。

横軸にデータの階級を、縦軸にその階級に含まれるデータの数（人数、個数など）をとって棒グラフで表します。

### ○気をつけることは？

データの階級幅が異なるときには、棒グラフとは異なった扱いが必要です。

例えば、上記の例で、「165cm 以上 170cm 未満と 170cm 以上 175cm 未満のところは人数も少ないから、まとめてしまおう」とした場合、「165cm 以上 175cm 未満 7人」となりますが、それをそのままグラフに描くと、階級をまとめて人数が増えただけなのに、グラフの高さが高くなってしまって、誤解してしまうおそれがあります。そのため、そのような場合には、階級幅が2倍なので横幅を2倍にし、その代わりに高さを半分にして表示します（つまり、面積がデータの個数に比例するように描きます）。

また、グラフの両端（上の例では「150cm 未満」と「175cm 以上」）も、高さを隣よりも、（その階級の幅は5cmより長いので）低く表示します。

ヒストグラムを描く際には、「階級数をいくつに分ければよいか」というのが問題になります。通常は、5～10 くらいの階級数でヒストグラムを描いてみて、あまりデータの散らばりが見られなければ階級数を変えてみる、という方法をとります。

### ○やってみよう

総務省統計局では、都道府県別や市町村別のデータをまとめてデータベースにしています。これを使って、「市町村別の中学校数の分布」などのヒストグラムを描いてみましょう。

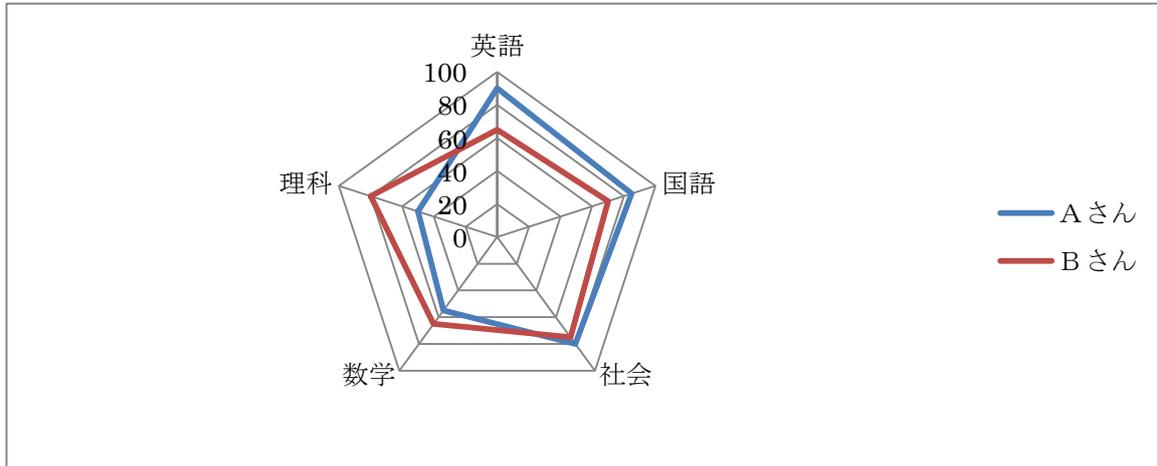
「統計でみる都道府県・市区町村(社会・人口統計体系)」

<http://www.stat.go.jp/data/ssds/index.htm>

## 6 レーダーチャート

### ○たとえば・・・

みなさんの中には、「理科系には強いが、文科系がちょっとねえ」とか、逆に「私は文科系の方が得意だ」という人もいるでしょう。そういったことを一目で示せないでしょうか。



テストの成績をレーダーチャートで示すことにより、Aさんは英語や国語といった文科系は得意だが数学や理科といった理科系は苦手、Bさんは逆に理科系が強い、といったことがみてとれます。

### ○どんなときに使うの？

レーダーチャートは、複数のデータ（指標）を一つのグラフに表示することにより、全体の傾向をつかむのに用いられるグラフです。

データの項目数に応じて中心から放射状に線を引き、それぞれの線上に、データを表示します。そしてそれらを線で結んで、その形状を見ます。

### ○気をつけることは？

レーダーチャートを描く際には、通常、「外に行くほど（データ値が大きいほど）良い」となるようにデータを選びます。

例えば、体力測定の結果をレーダーチャートに描く場合、幅跳びや反復横飛びは数字が大きいほど良いのですが、50m走などはタイムが短いほど良いので、単純に50m走のタイムでレーダーチャートを描くとおかしなグラフになってしまいます。そのような場合には、タイムではなくその逆数（つまり、「秒速〇〇m」とか）に直して、「データ値が大きいほど良い」データに変換する必要があります。

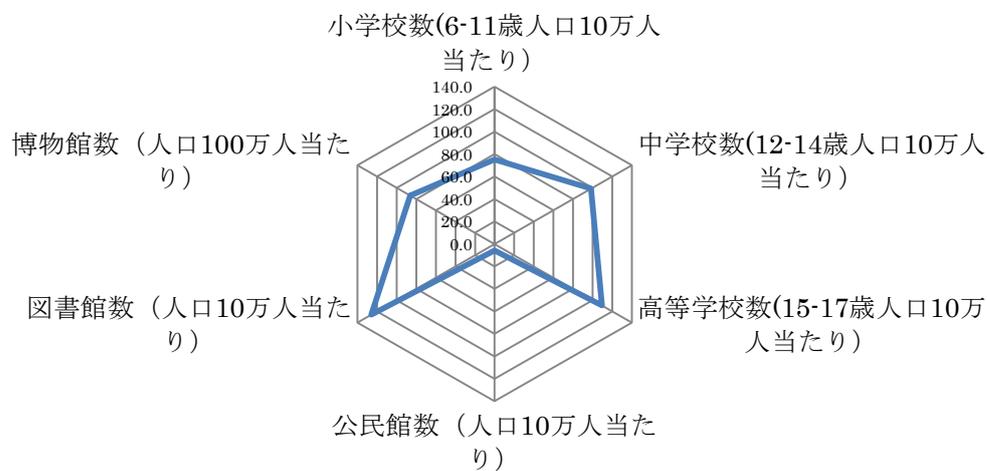
### ○やってみよう

総務省統計局では、都道府県別や市町村別のデータをまとめてデータベースにしています。これを使って、皆さんの住んでいる都道府県・市町村の指標をレーダーチャートに描き、全国平均と比べてどうか、見てみましょう。

「統計でみる都道府県・市区町村(社会・人口統計体系)」

<http://www.stat.go.jp/data/ssds/index.htm>

## 東京都の学校数、社会教育施設数 (人口当たり、全国平均=100)



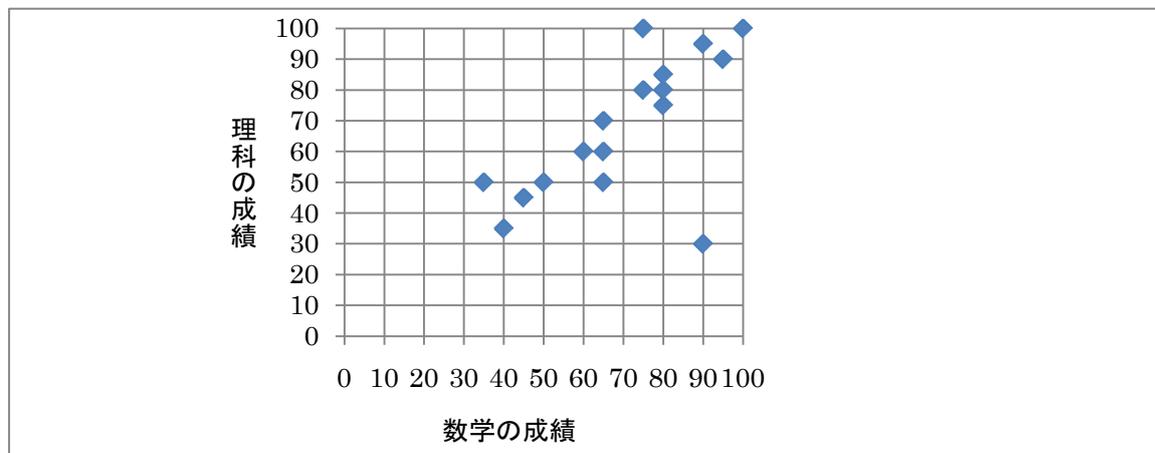
注) データは2005年値

出所) 総務省統計局「社会生活統計指標」

## 7 散布図

○たとえば・・・

「数学が得意な人は理科も強い」とよく言われますが、本当にそうなのか、クラスの人の点数をグラフにしてみました。



これをみると、確かに、数学の成績と理科の成績には関係がありそうに見えますね。

○どんなときに使うの？

散布図は、縦軸と横軸にそれぞれ別の量を取り、データがあてはまる場所に点を打って示す（「プロットする」といいます。）グラフです。2つの量に関係があるかどうかをみるのに使います。

○気をつけることは？

散布図でわかることは、2つの量の間に関係があるかということだけであり、因果関係（どちらかが原因となって、もう一方が起こる）を示すものではありません。

例えば、働いている人について、血圧と給料の関係を散布図に描いたとします。普通は、年齢が高い人ほど血圧が高く、また年齢が高いほど給料も高くなるので、「血圧と給料には関係があります」という散布図になります。

でも、これから、「血圧を高くすると、給料が上がります」ということにはなりませんよ。

○やってみよう

総務省統計局では、都道府県別や市町村別のデータをまとめてデータベースにしています。これを使って、県民所得や水道普及率、ごみの量などで、散布図を描いてみましょう。

e-stat「都道府県・市町村のすがた」

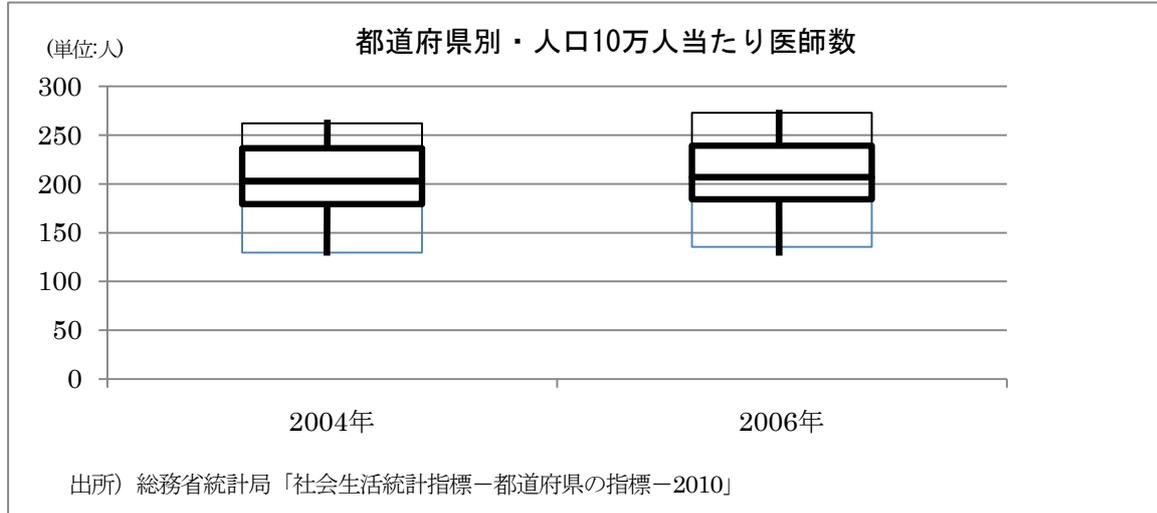
<http://www.e-stat.go.jp/SG1/chiiki/Welcome.do?lang=01>

## 8 箱ひげ図

### ○たとえば・・・

お医者さんの数は、地域的に偏りがあるとされています。

都道府県別に見て、人口10万人当たり医師数（医療施設に従事する医師数）にどの程度ばらつきがあるか、グラフに表してみましょう。



四角い箱の上下に、ひげが生えている形をしており、このようなグラフを「箱ひげ図」と呼びます。

この図では、ひげの一番下が最小値、箱の下部の辺が第1四分位数、真ん中の線が第2四分位数（中央値）、上部の辺が第3四分位数、そしてひげの一番上が最大値を表しています。

2004年と2006年とを比べると、全体として増加しているものの、最大値の増加が大きいことが見てとれます。

### ○どんなときに使うの？

データのばらつき具合を比較するのに用います。1つのデータのばらつきを見るだけでしたらヒストグラムでも見られますが、ヒストグラムでは異なる複数のデータのばらつきを比較することは困難です。そのような場合、箱ひげ図により比較することができます。

### ○気をつけることは？

ここで紹介したのは、ひげの両端がそれぞれ最小値、最大値になっている箱ひげ図ですが、中には、上下から10%点をひげの両端として描いている箱ひげ図もあります。

新聞やインターネットで箱ひげ図を見かけたときは、ひげの両端が何を表しているか、きちんと確認するとよいでしょう。

### ○やってみよう

総務省統計局では、都道府県別や市町村別のデータをまとめてデータベースにしています。いろいろなデータを使って、箱ひげ図を描いてみましょう。

e-stat「都道府県・市町村のすがた」

<http://www.e-stat.go.jp/SG1/chiiki/Welcome.do?lang=01>

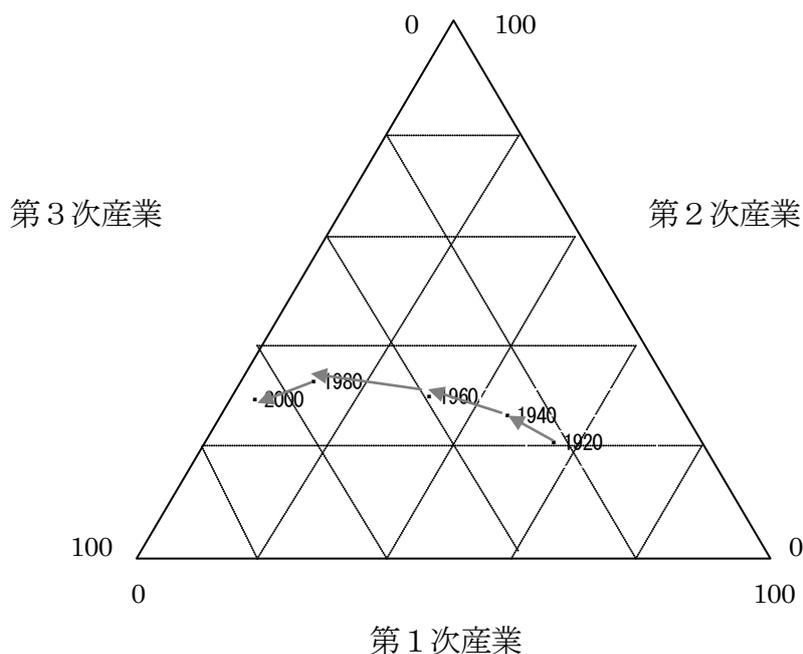
## 9 三角グラフ

○たとえば・・・

日本でも、以前は農業のような第1次産業で働いている人が多かったのですが、工業化の進展により製造業などの第2次産業で働く人が増え、その後、サービス業などの第3次産業で働く人が増えてきました。

これを、国勢調査のデータで確認してみましょう。

産業別就業者割合の推移（1920年～2000年）



	1920年	1940年	1960年	1980年	2000年
第1次産業	1467 (54.9%)	1439 (44.6%)	1439 (32.7%)	610 (10.9%)	317 ( 5.1%)
第2次産業	560 (20.9%)	844 (26.2%)	1280 (29.1%)	1874 (33.6%)	1857 (29.8%)
第3次産業	646 (24.2%)	943 (29.2%)	1684 (38.2%)	3091 (55.4%)	4048 (65.1%)
合 計	2673 (100.0%)	3226 (100.0%)	4403 (100.0%)	5575 (100.0%)	6223 (100.0%)

注) 「分類不能の産業」は合計から除外した。

出所) 総務省統計局「国勢調査」 産業別15歳以上就業者数

第1次産業の就業者の割合は、右上がりの斜線が目盛りになっています。1920年だと54.9%になっています。同様に、第2次産業の就業者の割合は底辺に平行な線が目盛り、第3次産業の就業者の割合は右下がりの斜線が目盛りです。

○どんなときに使うの？

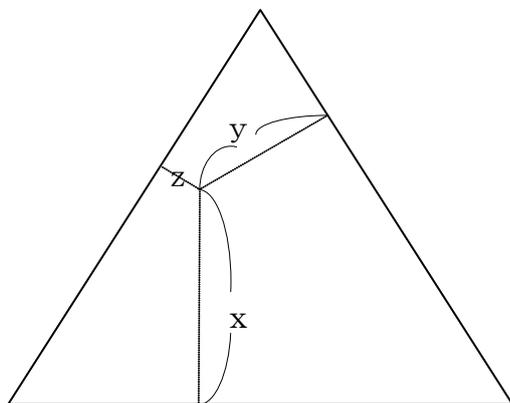
三角グラフは、3つの構成要素の比率を表すのに用いられます。社会科では上の例のような産業別の比率、理科では鉱物の組成比率を表すときなどに使われます。

なお、4以上の構成要素がある場合には三角グラフで表すことができません。(2次元

の紙の上には書いているので、やむを得ないのですが・・・) その場合は、重要でない要素は「その他」にまとめてしまって、3つの構成要素にしてグラフを描きます。

### ○気をつけることは？

ここで紹介したのは斜めの目盛線が入っている三角グラフですが、この他にも、辺からの距離でデータを表す三角グラフもあります。



例えば、底辺からの高さが、 $x$ の構成比を表す、ということになります。自分が見ている三角グラフがどちらなのか、気をつけましょう。

### ○やってみよう

国勢調査には、都道府県別の、産業別就業者数のデータもあります。あなたが住んでいる県でどうなっているか、隣の県はどうか、三角グラフに表してみよう。

<http://www.stat.go.jp/data/kokusei/2005/index.htm>

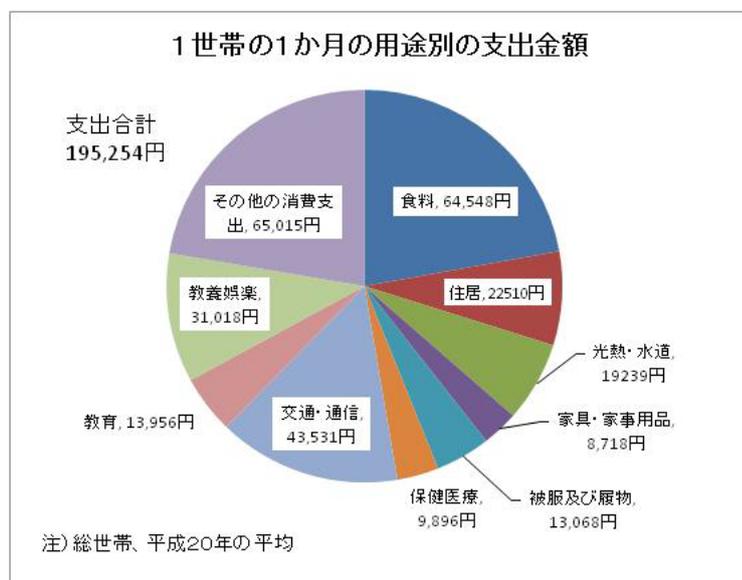
## 統計を使ってみよう ～家計調査の見方・使い方～

### ◇ 家計調査とは？

家計調査は各家庭で毎月家計簿をつけてもらい、1か月間のその家庭の収入がどのくらいあり、どんなものをどのくらい買っているのかを調べている調査です。

下のグラフは家計調査で調べた結果で、平成20年に1か月あたり1世帯がどんな支出をどのくらいしているかを示したものです。

- 世帯とは、1つの住宅と一緒に住んで一緒に生活している家族のことです。
- 支出とは、世帯の支払ういわゆる生活費（趣味・娯楽や耐久財への支出を含む）のことです。



食料、交際費などの「その他の消費支出」、自動車のガソリン代や携帯電話代などの交通・通信への支出が多くなっていることがわかります。

家計調査では、そのほかに、世帯の過去1年間の収入や世帯が貯蓄や負債をどのくらい持っているかも調べています。

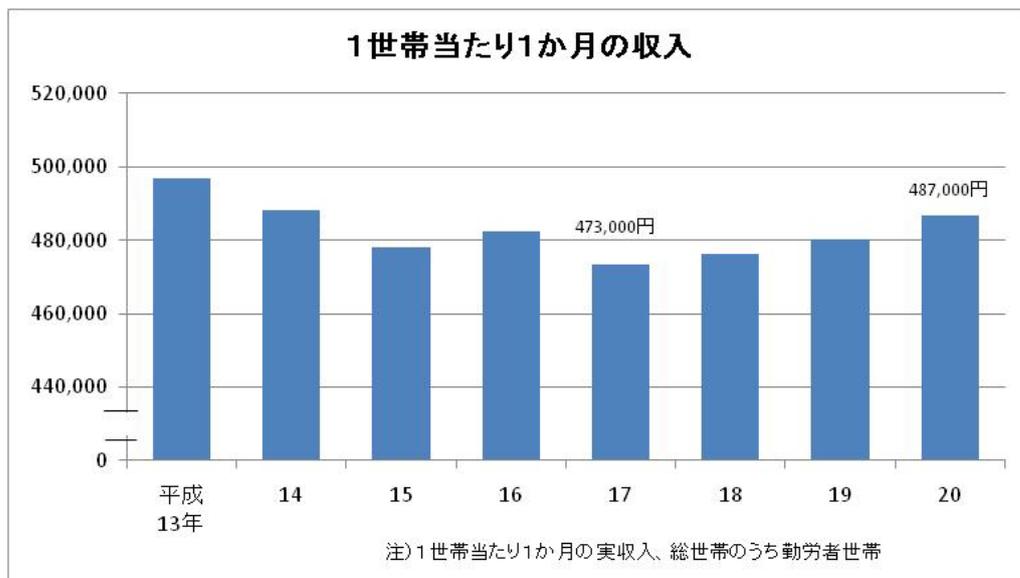
- 貯蓄とは、貯金や株式、保険などのことで、世帯がいざという時のために蓄えているお金のことです。
- 負債とは、住宅ローンのように大きな買い物をするために、銀行などから借りたお金のことです。

## ◇ 家計調査でわかること

家計調査の結果から、次のようなことがわかります。

- 世帯の1か月間の収入はいくらですか？
- 1年間の収入がいくら世帯が多いのですか？
- 世帯の1か月間の生活費（支出）はいくらですか？
- 一人暮らしの世帯の生活費はどうなっていますか？
- 高齢者の世帯の生活費はどうなっていますか？
- 品目ごと、例えば携帯電話への支出はいくらですか？
- 世帯主の年齢が違くと携帯電話への支出も違いますか？
- 地域ごと、例えばどの地方で牛肉をたくさん買っていますか？
- 1世帯あたりの貯蓄額はいくらですか？

- 世帯の1か月間の収入はいくらですか？

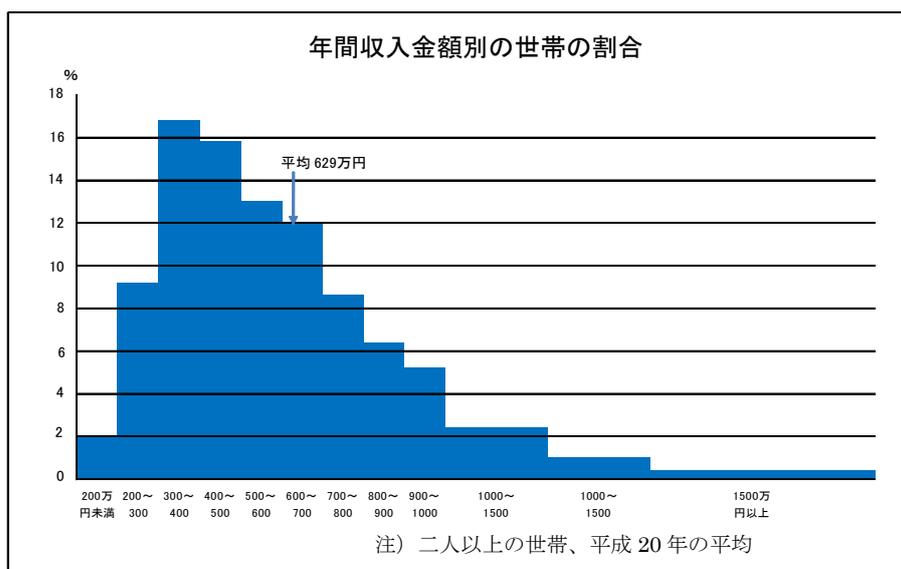


平成20年の勤労者世帯の1世帯当たり1か月の収入は約48万7千円です。景気の悪かった平成17年以前は毎年収入が減ることが多かったのですが、その後は景気が回復したことから徐々に収入が増えてきています。しかし、平成20年9月頃から景気が大きく落ち込んだので、平成21年の収入は減少すると予想されています。

- **勤労者世帯とは、**いわゆるサラリーマンの世帯のことで、家計調査では毎月の収入は勤労者世帯だけで調べています。

[データはこちら](#) **総世帯（勤労者世帯）の時系列表（年報）にリンク**

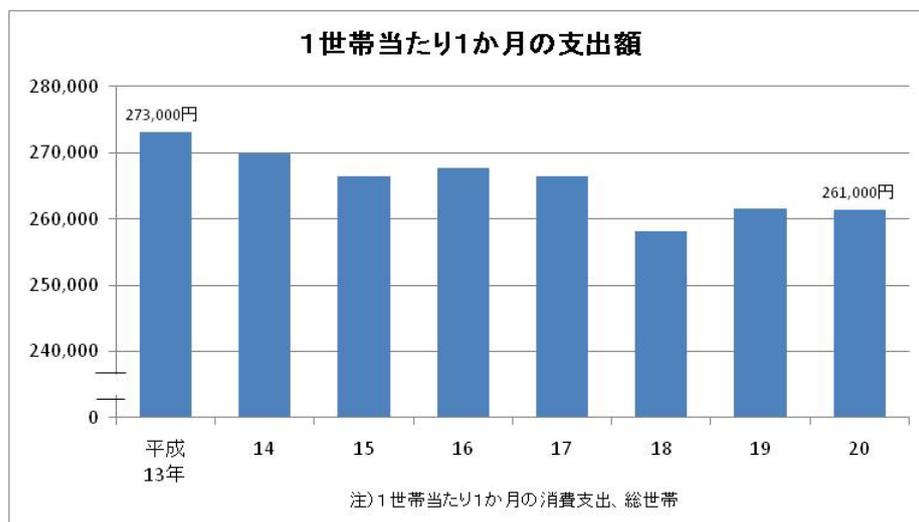
○ 1年間の収入がいくらの世帯が多いのですか？



1年間の収入が300万円以上400万円未満の世帯が約17%で、一番多くなっています。また、グラフをみると、平均629万円に比べて、それより収入の低い世帯がおおくなっていることがわかります。

[データはこちら](#) [二人以上の世帯の年平均の結果表にリンク](#)

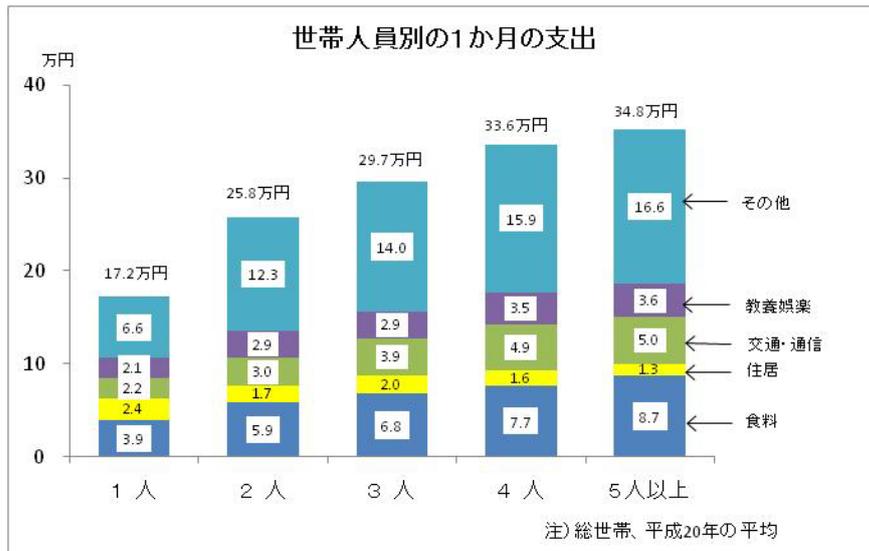
○ 世帯の1か月間の生活費（支出）はいくらですか？



平成20年の1世帯当たり1か月の支出額は約26万1千円です。7年前に比べると1万円以上少なくなっており、支出を減らしていることがわかります。

[データはこちら](#) [総世帯（勤労者世帯）の時系列表（年報）にリンク](#)

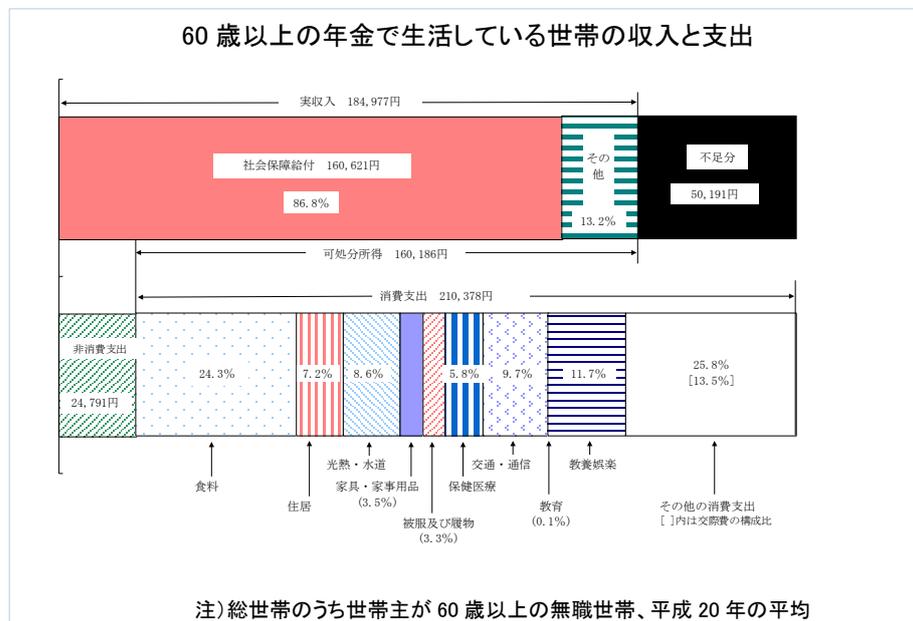
○ 一人暮らしの世帯の生活費はどうなっていますか？



一人暮らしの世帯（単身世帯）の1か月間の支出は17.2万円で、4人の世帯（33.6万円）の約半分になっています。1人あたりで考えると、逆に一人暮らしの世帯は、4人の世帯（1人あたり8.4万円）の約2倍の支出をしていることになります。

[データはこちら](#) [総世帯の年平均の結果表にリンク](#)

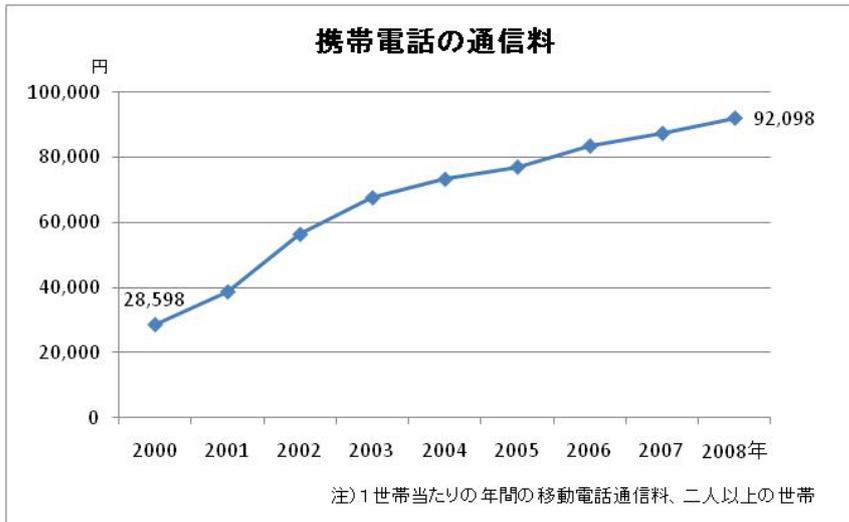
○ 高齢者の世帯の生活費はどうなっていますか？



高齢者のうち主に年金で生活していると思われる世帯では、収入が支出に比べ約5万円不足していることがわかります。この不足分は主に貯蓄の取り崩しなどでまかなわれています。

[データはこちら](#) [総世帯の年平均の結果表にリンク](#)

○ 品目ごと、例えば携帯電話への支出はいくらですか？

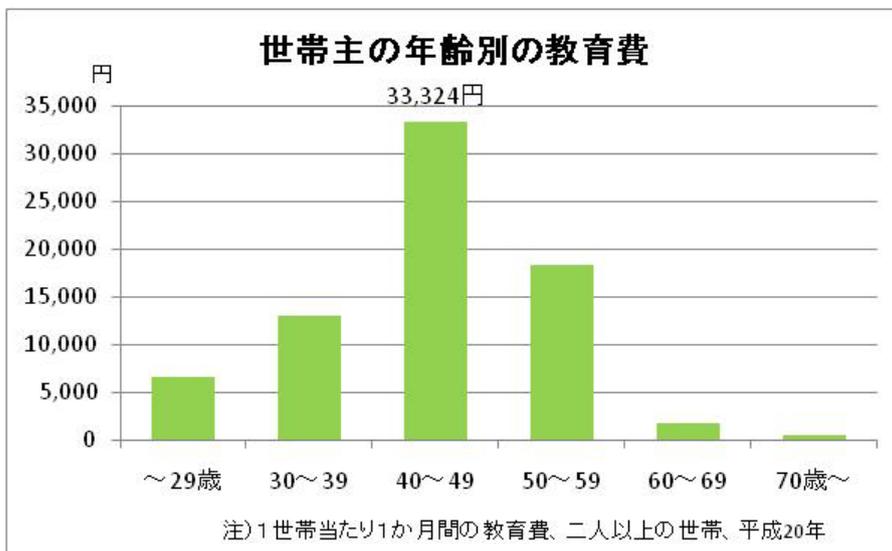


2008年には1世帯当たり1年間で92,098円、携帯電話に支出しています。8年前の2000年(28,598円)に比べると3倍以上と大きく増加していることがわかります。

●どんな品目の結果がわかるか、[収支項目分類一覧](#)でさがしてみよう。

[データはこちら](#) [二人以上の世帯の時系列表\(年報\)にリンク](#)

○ 世帯主の年齢によって子どもの教育費はどのくらい違いますか？

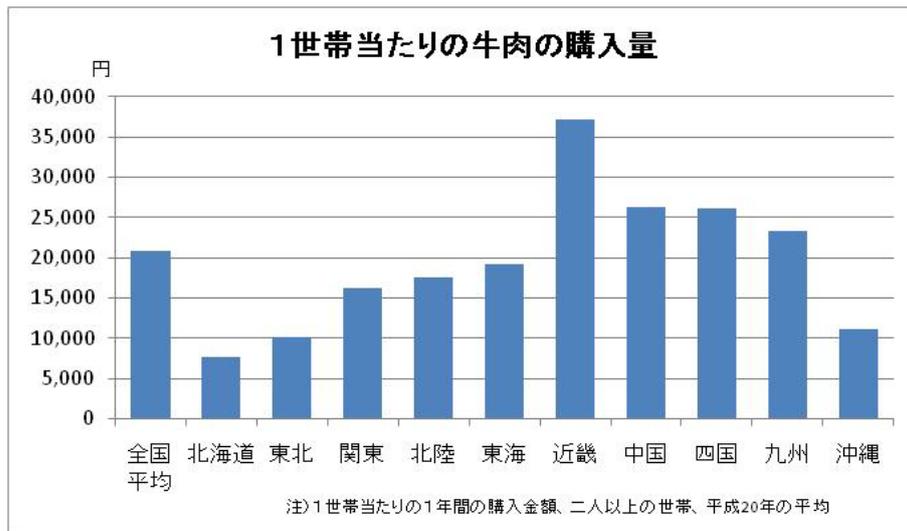


世帯主が40歳代の世帯が他に比べて飛びぬけて多くなっています。高校生や大学生の子どもがいる親がちょうどこの年代にあたっているからです。

●世帯主とは、世帯の主な収入を得ている人のことです。

[データはこちら](#) [二人以上の世帯の年平均の結果表にリンク](#)

○ 牛肉をたくさん買っているのはどの地方ですか？

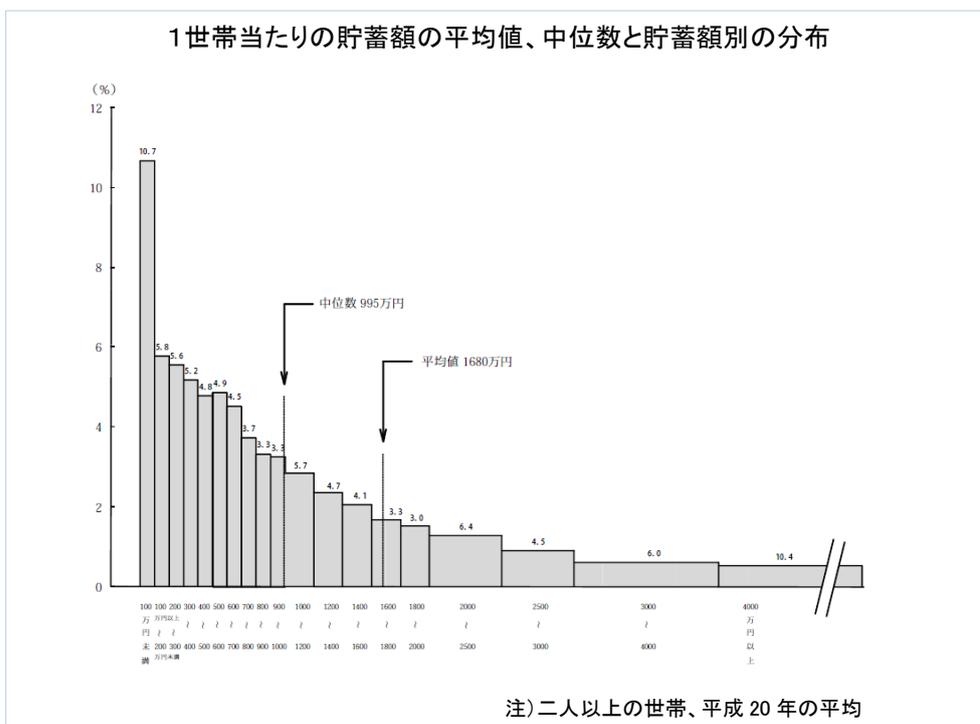


神戸牛（兵庫県）や近江牛（滋賀県）など有名ブランドの牛肉のある近畿地方が1年間に牛肉を4万円近く買っていて、他の地方より断然多くなっています。

このグラフから、沖縄を除く西日本では東日本より牛肉が多く買われていることもわかります。

[データはこちら](#) [二人以上の世帯の年平均の結果表にリンク](#)

○ 1世帯あたりの貯蓄額はいくらですか？



平成 20 年の 1 世帯当たりの貯蓄額は平均では、1680 万円になります。しかし、ヒストグラムをみると 100 万円未満の世帯の割合が 10.7%と最も高く、貯蓄の少ない世帯に分布が偏っていることがわかります。そのため中位数では 995 万円と平均より低くなっています。

●**中位数**とは、世帯を貯蓄の少ない世帯から多い世帯に順番に並べた時、ちょうど真ん中にくる世帯の貯蓄額のことです。

**データはこちら** [貯蓄・負債編の 7-1 表にリンク](#)

## ◇ 家計調査をもっと知りたい人に

**調査の概要**

家計調査のしくみをおおまかに説明しています

**家計調査のしくみと見方**

家計調査のしくみと見方を詳しく解説しています

## ◇ 家計調査のデータをもっと見たい人に

## ○ 家計調査でわかることがもっと知りたい人に

**家計調査年報（家計収支編）平成20年 家計の概況**

家計調査の平成20年平均結果からわかることを詳しく解説しています。

**統計トピックス** **家計ミニトピックス**

季節やそのときどきの話題，家計調査の代表的な結果などを図表で示しています。

**家計簿からみたファミリーライフ**

家計調査の代表的な結果を図表で示しています。

## ○ 収入と支出のデータが見たい人に

**家計調査（家計収支編）調査結果**

結果表には3つの種類があります。目的に応じてご使用ください。

- ① 二人以上の世帯 → 二人以上で住んでいる人の世帯の結果です。  
月別の結果はこの結果だけでみられます。  
世帯の種類ごとの結果など詳細な結果があります。また、購入数量（年平均）や都道府県庁所在市別の結果も二人以上の世帯だけでしかみられません。
- ② 単身世帯 → 一人暮らしをしている人の結果です。  
四半期、年ごとの結果で月別の結果はありません。
- ③ 総世帯 → 二人以上の世帯と単身世帯を合わせた結果です。  
日本全体の世帯を表すデータですが、四半期、年ごとの結果で月別の結果はありません。

## ○ 貯蓄と負債のデータが見たい人に

**家計調査（貯蓄・負債編）調査結果**

貯蓄・負債編の結果は二人以上の世帯だけで、四半期、年ごとの結果です。

## 都道府県別・市町村別データの調べ方

社会科の勉強のときなど、「わたしが住んでいる県（市）のデータが知りたい」「他の県（市）と比べてどうなのか知りたい」ということがあるでしょう。そういったときに便利なのが、総務省統計局の「統計でみる都道府県・市区町村(社会・人口統計体系)」です。

### ○ 使ってみよう

インターネットを使って、「統計でみる都道府県・市区町村(社会・人口統計体系)」の「社会生活統計指標-都道府県の指標」のページを見てみましょう。

<http://www.stat.go.jp/data/ssds/5.htm>

すると、

- I 社会生活統計指標
- II 基礎データ
- III 基礎データの説明

の3つが出てきます。そして、それぞれについて、「A 人口・世帯」などの分野に分かれています。

**基礎データ**は、都道府県ごとの人数や面積、施設数などの、統計調査等で得られたデータを掲載しています。しかし、「どの県で医療が充実しているだろう」といったことを調べようとすると、大きな県はそれだけ病院の数も多いでしょうから、単純に病院の数を比較するのでは不十分です。都道府県比較を行うために、人口や面積で割って比率にしたのが**社会生活統計指標**です。なお、社会生活統計指標には、比率だけでなく、各県庁所在市の年平均気温などのデータも含まれています。

また、データを正しく使うためには、そこで使われている用語の意味などについて、きちんと理解することが必要です。それは**基礎データの説明**にまとめられています。

### ○ どういうデータが載っているの

「都道府県の指標」には、人口や自然環境、経済、行政等、様々なデータが掲載されています。

例えば、「A 人口・世帯」では、基礎データとして、都道府県別の人口総数、年齢階級別人口、出生数、世帯数などのデータがあり、社会生活統計指標としては人口密度、年齢別構成比、出生率、**各種世帯の割合**などが掲載されています。

「B 自然環境」では、基礎データとして、都道府県別の総面積、森林面積、田畑の面積、宅地面積などが、社会生活統計指標としては森林面積割合や可住地面積割合、県庁所在市の年平均気温、最高気温、最低気温、降水量などが掲載されています。

その他にも、県内総生産額や県民所得のような経済基盤データ、県の財政支出のような行政基盤データ、学校数や生徒数、教員数のような教育データ、病院数や医療費、平均余命のような健康・医療データなどが掲載されています。

これらのデータを使って、みなさんが住んでいる県のデータを他県と比較したり、高校で学ぶ相関係数を使っていろいろなデータの相関を分析したりするとおもしろいでしょう。

## ○ 市区町村データの調べ方

市区町村別のデータについては、インターネットの「統計でみる都道府県・市区町村(社会・人口統計体系)」の「統計でみる市区町村のすがた」に載っています。

<http://www.stat.go.jp/data/ssds/5b.htm>

こちらについては、分量の関係などのため、掲載されているデータ項目が、都道府県のものに比べて少なくなっていますが、様々な分析を行うために有用なデータが数多く掲載されています。

## 基本用語集 目次

### (ア行)

移動平均 moving average

### (カ行)

階級 class

確率 probability

確率分布 probability distribution

確率変数 random variable

近似値 approximate value

グラフ graph

構成比 proportion

誤差 error

五数要約 five-number summary

### (サ行)

最頻値 mode

散布図 scattering diagram

四分位数 quartile points

四分位範囲 quartile range

四分位偏差 quartile deviation

信頼区間 confidence interval

正規分布 normal distribution

全数調査 census

相関 correlation

相関係数 coefficient of correlation

増減数・増減率 increase(decrease), rate  
of increase (decrease)

相対度数 relative frequency

### (タ行)

中央値 median

統計地図 statistical map

度数分布表 frequency distribution

ドットプロット dot plot

### (ナ行)

二項分布 binomial distribution

### (ハ行)

パーセント percent

箱ひげ図 box and whisker plot

外れ値 outlier

範囲 range

ヒストグラム histogram

非標本誤差 non-sampling error

標準偏差 standard deviation

標本、標本調査 sample, sample survey

標本誤差 sampling error

分散 variance

分布 distribution

平均値 average

ポイント (ポイント差) percent point

母集団 population, universe

### (マ行)

無作為抽出 random sampling

(収録用語 42語)

**移動平均 moving average**

データの季節変動の影響等を取り除くため、前後のいくつかのデータの平均をとること。

たとえば「3か月移動平均」とは、連続する3か月のデータを平均し、その中央の月（2か月目）のデータの代替値として使用するものであり、1月の移動平均値であれば前年12月、当年1月、2月の3か月の平均値による。

**階級 class**

度数分布表やヒストグラムを作るときの、データ範囲の刻みのことをいう。

例：身長 165cm 以上 170cm 未満の階級

**確率 probability**

ある事柄が起こる確からしさのこと。

例：コインを投げた時に表が出る確率は 50%である。

**確率分布 probability distribution**

変数がある値をとる確率がいくらか、表したもの。

例：サイコロの目の確率分布

X	1	2	3	4	5	6	計
確率	1/6	1/6	1/6	1/6	1/6	1/6	1

**確率変数 random variable**

どの値をとるか確実には分からず、確率的に決まる変数のことをいう。

例えば、サイコロを振ったときに出る目の数は、確率変数である。

**近似値 approximate value**

必ずしも真の値ではないものの、真の値に近いものを近似値という。

例えば、円周率は  $3.14159265 \dots$  であるが、3.14 はこの近似値である。

身長を測るときも、実際にはミクロン単位、あるいはさらに細かい単位で正確に測ることはできない。その意味では「A君の身長は 175cm だ」というのも、近似値である。

**グラフ graph**

数量を、見やすく図にあらわしたもの。

統計では、棒グラフや折れ線グラフ、円グラフなどを用いて、結果を分かりやすく表すことが重要である。

**構成比 proportion**

全体を 100 パーセントとしたときの、それぞれの内訳の割合。

例：3年5組の男女の構成比は、男子 46%、女子 54%である。

**誤差 error**

真の値からのずれの大きさ。

標本調査では、必ずしも真の値、あるいは全数調査をすれば得られたはずの値は分からないが、「真の値はおおよそ 1800 円±50 円の範囲にあるだろう」ということは推定できる。この場合、「誤差は 50 円以下だろう」ということになる。(⇒「標本誤差」の項参照)

**五数要約 five-number summary**

データのばらつきの様子をあらわすのに、

- ・最小値
- ・第1四分位数 (小さいほうから 1/4 のところのデータ)
- ・第2四分位数 ( " " 2/4 " " 、中央値と同じこと)
- ・第3四分位数 (小さいほうから 3/4 のところのデータ)
- ・最大値

の5つの数を用いて表すこと。

**最頻値 mode**

起こる頻度が最も高い値のこと。モードともいう。

最頻値を求めるには、度数分布表を作成し、度数の最も大きいところが最頻値である。

下記の度数分布では、野球部が最頻値である。

クラブ	野球部	サッカー部	水泳部	演劇部
部員の数(人)	20	15	10	8

連続的な変数の場合には、度数の最も大きい階級の中央値を最頻値とする。

下記の度数分布では、「155cm」が最頻値である。

身長	150cm 未満	150cm 以上 160cm 未満	160cm 以上 170cm 未満	170cm 以上
度数	10	50	42	30

**散布図 scattering diagram**

2つの量の関係を見るために、グラフの縦軸・横軸にそれぞれの変数を取り、該当するデータを点であらわしたもの。

**四分位数 quartile points**

データを小さい順に並べて、下から 1/4 のところのデータを第1四分位数、2/4 のところのデータを第2四分位数 (これは中央値と同じ)、3/4 のところのデータを第3四分位数という。そして、第1四分位数、第2四分位数、第3四分位数をまとめて、四分位数という。

**四分位範囲 quartile range**

(第3四分位数－第1四分位数)の値のことを四分位範囲といい、データがどのくらい散らばっているかの目安として用いる。

**四分位偏差 quartile deviation**

(第3四分位数－第1四分位数) ÷ 2の値のことで、四分位範囲の半分。四分位範囲と同じく、データがどのくらい散らばっているかの目安として用いる。

**信頼区間 confidence interval**

真の値がどの範囲にあるのかを表す方法であり、「真の値が存在する 95%信頼区間は 150cm ± 5cm である」と言った場合、真の値は 95%の確率で 145cm から 155cm の間にある、ということの意味している。

**正規分布 normal distribution**

確率密度関数が  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$  で表される確率分布。これを、平均  $\mu$ 、分散  $\sigma^2$  の正規分布といい、記号  $N(\mu, \sigma^2)$  で表す。

同じ確率分布に従う確率変数  $X_1, X_2, \dots, X_n$  を独立に取り出したとき、もともとの確率分布がどのようなものであろうとも、その平均値 ( $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ ) は近似的に正規分布になる (中心極限定理) など、正規分布は統計学では特に重要な意味を持つ。

**全数調査 census**

対象となるものの全てを調査するものを全数調査といい、対象のうち一部だけを取り出して調査するものを抽出調査 (あるいは標本調査) という。総務省統計局が5年ごとに実施している国勢調査は代表的な全数調査である。

**相関 correlation**

2つの変数の間で、一方が増えれば他方も増える (または減少する) という関係がある場合、2つの変数の間に相関があるという。

**相関係数 coefficient of correlation**

相関の強さを表す指標で、-1 から 1 の間の値をとる。2変数の一方が増えれば他方も増えるという関係にある時は正の値をとり、一方が増えれば他方が減るという関係にある時は負の値をとる。いずれの場合も相関が強いほど高い絶対値をとる。例えば、散布図を描いたときにデータが完全に一直線上に乗っており、その直線の傾きが正であるときは相関係数が 1、負であるときは -1 となる。

**増減数・増減率 increase(decrease), increase (decrease) rate**

ある量の以前からの増加、減少を増減数、それを元の値で割って率にしたものを増減率という。

例えば、2000年の日本の人口は1億2692万6千人、2005年の人口は1億2776万8千人なので、

- ・この期間の人口の増減数は、1億2776万8千-1億2692万6千=842千人
- ・ " 増減率は、842千÷1億2692万6千×100=0.7%

### 相対度数 relative frequency

データがある値をとる頻度を度数というが、その度数を全体で割って割合で表したものを相対度数という。

	A	B	C	D	合計
度数	10	16	8	6	40
相対度数	25%	40%	20%	15%	100%

### 中央値 median

データを小さいほうから順にならべたときに、ちょうど中央にくるデータの値。中位数ともいう。第2四分位数と同じになる。

### 統計地図 statistical map

統計データを地図上にあらわしたもの。  
一般には、データの大きさを、色の濃淡で表す。

### 度数分布表 frequency distribution

データがどのように散らばっているかを示す表であり、値の階級に対して、その範囲にいくつデータがあるかの頻度（データの個数）を表したものをいう。

### ドットプロット dot plot

度数分布表を視覚的に表したグラフで、横軸にデータ階級を、縦軸に度数を点で表したもの。度数が10であれば点を10個、縦に並べて表示する。

### 二項分布 binomial distribution

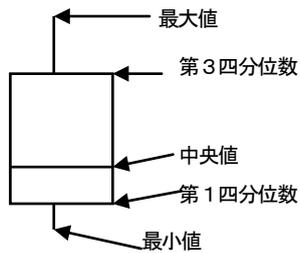
表が出る確率がpであるようなコインをn回投げたときに、表が全部で何回出るか、という確率分布。 $P(X=r) = {}_n C_r p^r (1-p)^{n-r}$ で表わされ、記号  $B(n,r)$  で表す。nが十分大きい場合、この分布は平均 np、分散  $np(1-p)$  の正規分布にしたがう。

### パーセント percent

全体に対する割合を、全体を100として表したもの。百分率。

### 箱ひげ図 box and whisker plot

データの散らばり具合を表す五数要約を図で表したものであり、以下のような形をしている。



### 外れ値 outlier

一つの集団の中に存在する、全体の傾向とは異なって、離れた所に分布するデータのことを外れ値という。ヒストグラムや散布図を描くことによって外れ値を容易に検出することができる。

### 範囲 range

データの最小値と最大値の間のことを、データの範囲という。

### ヒストグラム histogram

度数分布表を棒グラフで表したもので、データの散らばり具合をみるのに用いる。棒の一番高いところが最頻値である。

### 非標本誤差 non-sampling error

標本誤差以外の誤差を言う。例えば統計調査では、回答者の誤回答や未回答などが主な非標本誤差の例であるが、その大きさを評価することは困難である。

### 標準偏差 standard deviation

データの散らばり度合いを表す数で、大きいほど、データが散らばっていることを表す。分散の平方根に等しく、記号 $\sigma$ で表す。

$X_1, X_2, \dots, X_n$ の標準偏差は、以下のように計算される。

$$\sigma = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}} \quad \text{ただし } \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

### 標本、標本調査 sample, sample survey

統計調査を行う際に、対象のすべてを調べるのではなく、一部だけを取り出して調査するものを標本調査といい、取り出されたものを標本という。

### 標本誤差 sampling error

全数調査でなく標本調査を行ったことにより生ずる誤差を言う。通常、調査結果から「真の値（あるいは、全数調査をすれば得られたはずの値）は〇%の確率で1800円±50円の範囲にあるだろう」などということは推定できるため、誤差の大きさを事後的に評価する

ことは可能である。

### 分散 variance

データの散らばり度合いを表す数で、大きいほど、データが散らばっていることを表す。記号  $V$  で表す。標準偏差の 2 乗に等しい。

$X_1, X_2, \dots, X_n$  の分散は、以下のように計算される。

$$V = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n} \quad \text{ただし } \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

### 分布 distribution

「確率分布」と同じ意味で用いられる。

### 平均値 average

データの合計をデータの個数で割ったもので、「算術平均」ともいう。

### ポイント（ポイント差） percent point

パーセントで表された数字同士の差を表す単位。

### 母集団 population, universe

標本調査を行うときの、調べる対象となる全体のこと。

### 無作為抽出 random sampling

標本調査を行うときの標本の選び方の一つで、選ぶ際の恣意性をなくし、全く確率的に母集団から選ぶ方法。

## 補助教材イメージ 「授業導入の素材～統計をめぐるエピソード」

### 統計の歴史を振り返る ～統計の3つの源流～

我々が今日「統計」と呼んでいるものの歴史を振り返ると、その源流は以下のように大きく3つに分けることができます。

- ① 国の実態をとらえるための「統計」
- ② 大量の事象をとらえるための「統計」
- ③ 確率的事象をとらえるための「統計」

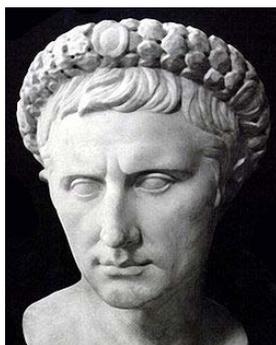
これらは別々のルートをたどって、19世紀半ば、ケトレー（Adolphe Quetelet 1796-1874）が社会統計を科学的に作成・分析するために確率論を導入したことで、社会現象・自然現象いずれも数量的にとらえる「統計」として形を整えました。ケトレーは、母国ベルギーの統計制度の整備や公的統計の改善に努めただけでなく、国際的な統計の比較可能性を高めるべく国際統計会議の設立にも力を尽くし、その功績から「近代統計学の祖」とされています。



Adolphe Quetelet

以下では、ケトレーによって統合される三つの源流について振り返ります。

#### 1 国の実態をとらえるための「統計」



Augustus

①の「統計」は、もっとも古くまで歴史を遡ります。

古来、為政者は、徴税、兵役などのために、その支配する領域内の実情をできるだけ正確に把握する必要がありました。明治初期に「統計」と訳された statistics（英）やその基になった statistik（独）はラテン語の「status」（国家・状態）に由来していますし、19世紀のフランスの統計学者モーリス・ブロックは「国家の存するところ統計あり」という言葉を残しています。こうしたことから、統計が国家経営に欠かせないものとして発展してきたことは容易に理解できます。

古代エジプトでは紀元前三千年にピラミッドを建設するための調査が行われたことが知られていますし、ローマ帝国では初代皇帝アウグストゥスの治世の頃に、人口や土地を調べる調査（Census）が行われました。今日、国勢調査のことを「人口センサス」と呼ぶのはその名残です。

16世紀以降、ヨーロッパでは各国が互いの勢力拡大を目指してしのぎを削るようになり、国家の繁栄は人口や貿易に反映されるという考え方から、17世紀になると産業や人口に関して数量的なデータを把握するための調査・研究が盛んに行われました。ドイツを中心に発展した「国勢学」がその代表です。この学派は、人口や土地面積、歳入歳出といった国家の基礎をなす事項（国家顕著事実）を記録し、比較を容易にするために表式で表すことを試みました。

イギリスでは、ウィリアム・ペティ（William Petty 1632-87）がその著書「政治算術」の中で、度重なる戦争で苦しい状況に追い込まれていた当時のイギリスの人口や経済の実態をオランダ、フランスと定量的に比較し、国政に役立てるよう国王に献上しました。その手法はドイツの「国勢学」とは異なるもので、後述する②の流れに属しませんが、国家の実情を把握し、国家運営の指標として用いようという意図において、互いに通じるものがあります。



William Petty

このように、18世紀から19世紀にかけて、各国で国家運営の基礎として統計を用いることの重要性が強く認識されるようになり、そのための体制整備や統計調査が積極的に行われるようになりました。フランスでは、統計の重要性に着目したナポレオン（1769-1821）によって1801年に統計局が設置され、政府によって統計が整備されるようになりました。各国で最初の近代的なセンサス（人口調査）が行われたのも、この時期です（デンマーク 1769年、アメリカ 1790年、オランダ 1795年、イギリス 1801年など）。

## 2 大量の事象をとらえるための「統計」

②の統計は、イギリスのジョン・グラント（1620-74）によってその道が切り開かれました。

グラントは、当時たびたびペスト禍に見舞われていたロンドンで、教会の資料を基にした死亡統計表を分析し、一見偶然とみえる人口現象に規律性のあることを明らかにしました。彼はまた、当時200万人と考えられていたロンドンの人口について、様々なデータや観察を通じて38万4千人と見積もり、限られた量のサンプルデータを注意深く観察することで全体の人口に関する推測が可能になることを示したのです。

グラントのこうした手法は、研究の基礎を数字に置きながら、単に物事の状況を描写することにとどまらず、一見不秩序に見える複雑な物事の間横たわる規律の発見に努めた点で、前述した「国勢学」（国家の基礎を成す重要事実を対象としてそれをありのままに描写・記述することに終始し、因果関係を探求

することはなかった)とは明らかに一線を画していました。前述のペティはグラントの友人でした。彼が著した「政治算術」は国の実態を明らかにするものでしたが、社会的な事象を数量的に観察し、その背後にある規則性を指摘しており、統計学上の位置づけとしてはグラントと同じ流れに属します。

グラントやペティに続いて、その手法を科学的に一層発展させたのが、ハレー彗星を発見したことで知られるエドモンド・ハレー (Edmond Halley 1656-1742) です。ハレーは、それまで偶然が支配するところと考えられていた人間の死亡に一定の規律性があること、すなわち集団的な人口に現れる死亡には、これを予測し得る一定の秩序があることを明らかにしました。当時のイギリスには、いくつかの生命保険会社がありましたが、合理的な保険料を計算する基礎を持たず、その経営はいわばギャンブルの一種であるかのように考えられていました。ハレーはグラントが手がけた生命表を更に発展させ、これによって初めて、生命保険会社が合理的な保険料金を算出できるようになったのです。その意味で、ハレーは今日の生命保険事業の基礎を築いたと言えます。



Edmond Halley

### 3 確率的事象をとらえるための「統計」



Blaise Pascal

①②の統計の流れとは別に、確率的な事象をとらえる必要から、統計に関する重要な概念や手法が発展してきました。③の統計は、サイコロ賭博のように偶然に左右されるギャンブルとの関わりの中から産み出されました。

今日、統計には欠かせない「標本空間」の基本的な考え方は、16世紀にサイコロ賭博やトランプゲームにおける偶然の仕組みを数学的に研究したイタリア人カルダーノ (Geloramo Cardano 1501-76) によっていますし、地動説を唱えたことで有名なガリレオ (Galileo Galilei 1564-1642) はトスカーナ大公から命じられて「サイコロゲームについての考察」という小論を書いています。

数学者パスカル (Blaise Pascal 1623-62) とフェルマー (Pierre de Fermat 1600年代初頭-1665) はサイコロ賭博の問題をテーマに書簡をやりとりし、その中から確率論の基礎が芽生えました。期待値、推定、検定、標本理論などは、そこから発展していったものです。

パスカル、フェルマーが基礎を作った確率論は、その後、数学における大きなジャンルとなり、18世紀に入り、ベイズ (Thomas Bayes 1702-61)、ラグランジュ (Joseph-Louis Lagrange 1736-1813)、ラプラス (Pierre-Simon Laplace



Pierre de Fermat



Pierre-Simon Laplace

計に応用されました。

1749-1827) といった一流の数学者たちの研究を経て大成します。確率論の統計への応用としては、ドゥ・モアブル (Abraham de Moivre 1667-1754) の年金論、D. ベルヌーイ (Daniel Bernoulli 1700-82) による天然痘の罹病率、死亡率の計算などがあります。オイラー (Leonhard Euler 1707-83) とラプラスは抽出調査を基にした全体の推計方法を考案し、それはフランスの人口の推

<了>

## 「統計」という言葉の起源

「統計」という言葉は、「統」（トウ・す（べる））という漢字と、「計」（ケイ・はか（る））という漢字とで構成され、「すべてを集めて計算する」という意味になっています。個々の事柄を問題にするのではなく、全体を集めてその姿を観察しようという統計の本質をよく表しています。

「統計」という言葉は英語の「statistics」の訳語として、明治年間を通じて次第に定着するようになったものです。「統計」を冠した最初の政府組織は明治4年に大蔵省に置かれた「統計司」と「統計寮」です。明治7年にはフランス語の「Elements de Statisique」の訳書として、蓑作麟祥（みつくり・りんしょう）により文部省から「統計学」という本も出されています。

「statistics」の訳語としては、当時、他にも「政表」「表記」「表紀」「形勢」などが提案されました。我が国の「統計学の開祖」とも言われる杉亨二（すぎ・こうじ）などは、無理に訳語を当てずに「寸多知寸知久（スタチスチク）」を用いるべきと主張しました。そのための国字（日本で作られた漢字：「峠・とうげ」

「榭・さかき」など）さえ考案しました。これが<sup>寸</sup>多<sup>知</sup>寸<sup>知</sup>久<sup>久</sup>です。しかし、「統計」以外の訳語は、この国字も含めていずれも定着することなく、やがて消えていきました。

それでは statistics を最初に「統計」と訳したのは誰だったのでしょうか。残念ながら確定的な証拠は残っていないのですが、当時の史料に基づいたある研究によれば、それは「柳河春三」（やながわ・しゅんさん）ではないかということです。柳河春三は、現在の愛知県（名古屋市）出身で、幼少の頃から神童の誉れ高く、1864年に開成所という江戸幕府の洋学研究教育機関の教授に登用された人物です。我が国で初めて「雑誌」と名のついた出版物を刊行したり、日本人の編集による最初の新聞を創刊したことでも有名です。

石橋重朝（明治18年から26年まで内閣統計局長を務めた人物）の証言によれば、彼が開成所の学生だった明治2年、柳河が編さんした「統計入門」又は「統計便覧」と題した小冊子の中で「統計」という言葉が使われていた、「この訳語は不完全と考えるが、とりあえず仮にこうしておく」という趣旨のことが記されていた、ということです。当時、「統計」という言葉は、もっぱら「合計」

という意味で使われていたようなので、柳河もこの訳語に自信を持ちきれなかったのかもしれませんが。

柳河は明治 3 年（1870 年）に没しますが、その生前には、「統計学」（明治 7 年）を翻訳した蓑作や「統計寮」の創設（明治 4 年）を発案した伊藤博文の米国視察に随行した福地源一郎と深い親交があったことが知られています。そのことから、柳河から二人に何らかの形で「統計」という言葉が伝わったのではないかと考えられます。

仮の訳語として生まれたらしい「統計」という言葉ではありますが、1900 年代初頭に日本の統計学関係の書籍を通じて中国に伝わってそのまま根付き、今日では中国語としても使われています。

<了>

## ナイチンゲールと統計

「近代看護教育の生みの親」とも呼ばれるイギリスの看護師フロレンス＝ナイチンゲール（1820-1910）ですが、統計とも深い関わりがあることは日本ではあまり知られていません。



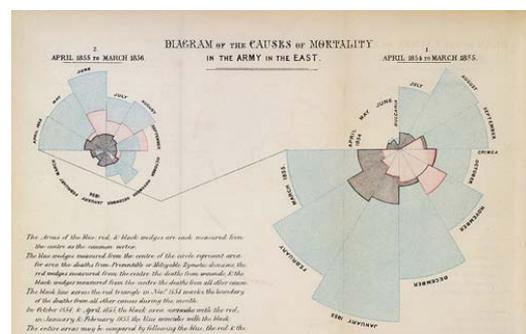
*Florence Nightingale*

彼女は、上流階級の家産に生まれ、歴史・語学・音楽など高いレベルの教育を受けました。また、若い頃から「近代統計学の父」ベルギー人アドルフ＝ケトラー（1796-1874）を信奉し、数学や統計に強い興味を持ち、優秀な家庭教師について勉強したとされています。

ナイチンゲールは、イギリス政府によって看護師団のリーダーとしてクリミア戦争（ロシアとトルコとの戦争で、イギリスはフランスとともにトルコに味方してロシアと戦った）に派遣されると野戦病院で骨身を削って看護活動に励み、病院内の衛生状況を改善することで傷病兵の死亡率を劇的に引き下げました。

彼女は統計に関する知識を存分に使ってイギリス軍の戦死者・傷病者に関する膨大なデータを分析し、彼らの多くが戦闘で受けた傷そのものではなく、傷を負った後の治療や病院の衛生状態が十分でないことが原因で死亡したことを明らかにしたのです。

彼女が取りまとめた報告は、統計になじみのうすい国会議員や役人にも分かりやすいように、当時としては珍しかったグラフを用いて、視覚に訴えるプレゼンテーションを工夫しました。今も「鶏のとさか」と呼ばれる円グラフの一種はこの過程で彼女によって考え出されたものです。



クリミア戦争における死因分析を表したグラフ

1860年には、ケトラーが立ち上げた国際統計会議のロンドン大会に出席し、統一的な病院統計のためのモデル形式を提案しました。統計のとり方がバ

ラバラであっては、有効な比較分析に支障を来し、医療技術の向上にもつながらないと考えたのです。提案は会議の分科会で討議され、各国政府に送付する決議が採択されました。

国をまたいで統計調査の形式や集計方法を標準化することは、今日でも簡単なことではありません。ナイチンゲールには現場の経験と統計の知識に裏付けられた揺るぎない信念があったのでしょう。

このような活躍が認められ、ナイチンゲールは1859年に女性として初めて王立統計協会（the Royal Statistical Society）の女性会員に選ばれ、その16年後には米国統計学会の名誉会員にもなっています。

「白衣の天使」ナイチンゲールー祖国イギリスでは統計学の先駆者として今も人々の記憶に刻まれています。

<了>

## 特定の日に消費が増える品物

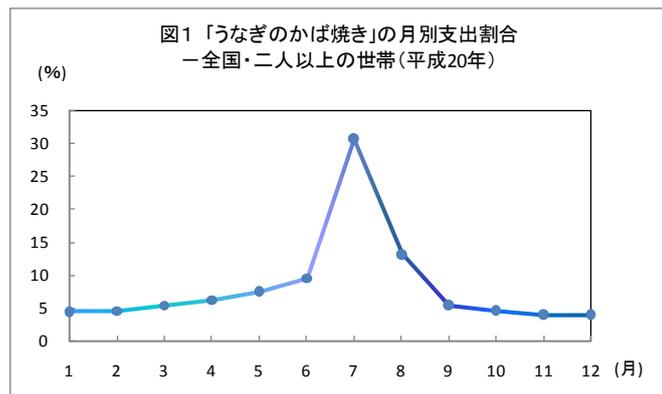
毎年7月になると、「土用の丑の日」ののぼりが店頭で目立ってきます。

「土用の丑の日」にうなぎを食べる由来については諸説ありますが、讃岐国（現在の香川県）出身の平賀源内が発案したという説がよく知られています。

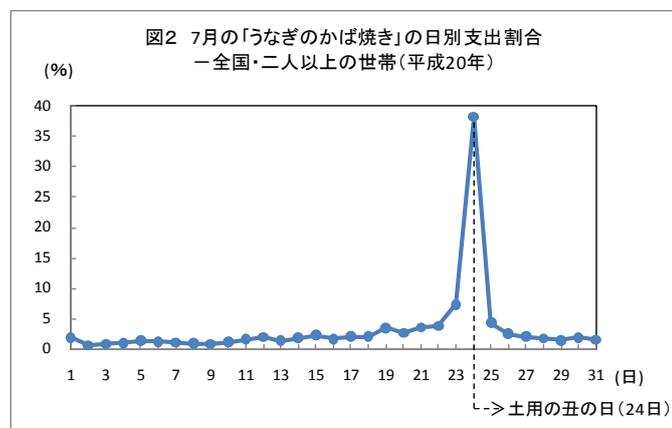
文政5年（1822年）の、当時の話題を集めた『明和誌』（青山白峰著）によると、夏に鰻が売れないことを悩んでいた鰻屋が、何とか売るため源内の所に相談に行ったところ、源内は、「丑の日に『う』の字が付く物を食べると夏負けしない」という民間伝承からヒントを得て、「本日丑の日」と書いて店先に貼ることを勧めたようです。すると、物知りとして有名な源内の言うことなら、ということで、その鰻屋は大変繁盛したようです。その後、他の鰻屋もそれを真似るようになり、土用の丑の日に鰻を食べる風習が定着したといわれています。

実際に土用の丑の日に鰻が食べられているのか、家計調査（二人以上の世帯）の結果から見てみましょう。

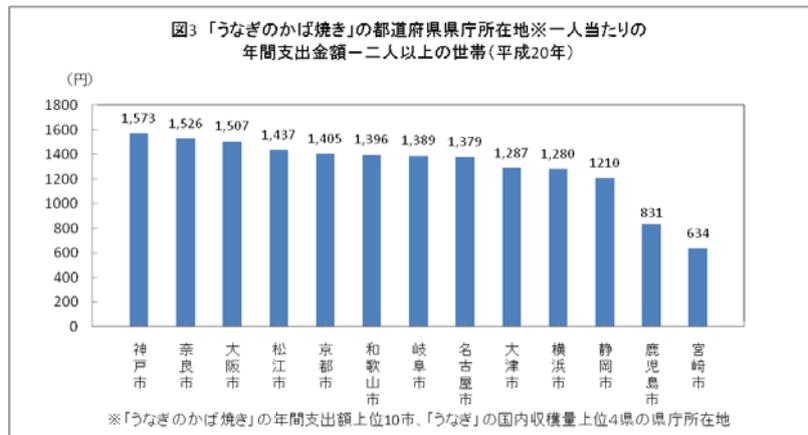
「うなぎのかば焼き」について、平成20年の1世帯当たり年間支出金額に対する各月の支出割合をみると、1年の中で土用の丑の日を含む7月が最も高く、30.7%を占めています。（図1）



さらに、7月の「うなぎのかば焼き」の月間支出金額に対する日別割合をみると、「土用の丑の日」に当たる7月24日の支出が38.3%と最も高く、また、その前後の23日及び25日もほかの日に比べると高く、この3日間の合計だけで約50%となっており、「土用の丑の日」に「うなぎのかば焼き」を食べるといふ昔ながらの慣習が顕著に表れています。（図2）



「うなぎのかば焼き」について、平成20年の1人当たり年間支出金額を都道府県庁所在市別にみると、神戸市が1,573円で最も多く、次いで奈良市(1,526円)、大阪市(1,507円)の順となっています。(図3)



昔から、「うなぎのか

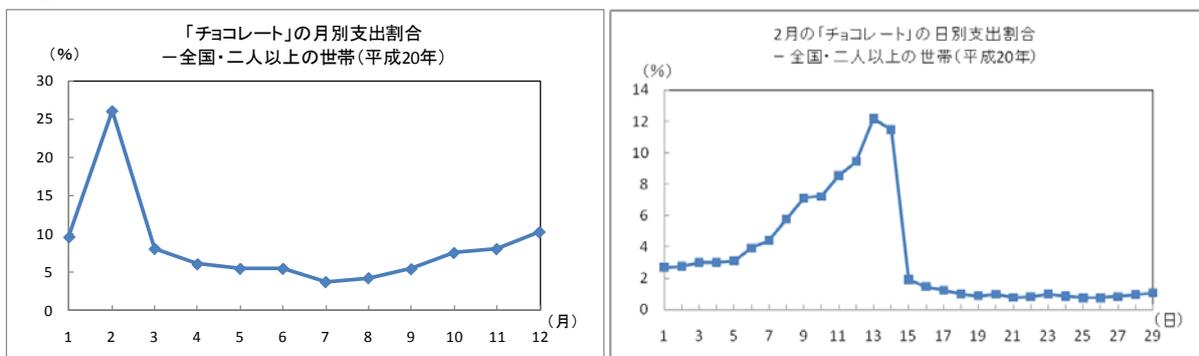
ば焼き」は、関西では「まむし」と呼ばれて有名なことなどが、ランキングに表れているのでしょうか。

一方で、平成20年の「うなぎ」の国内収穫量(農林水産省漁業・養殖業生産統計)の全国に占める割合をみると、鹿児島県(35.5%)、愛知県(29.9%)、宮崎県(16.7%)、静岡県(7.8%)などが高くなっており、必ずしも生産地でたくさん食べられているわけではないようです。

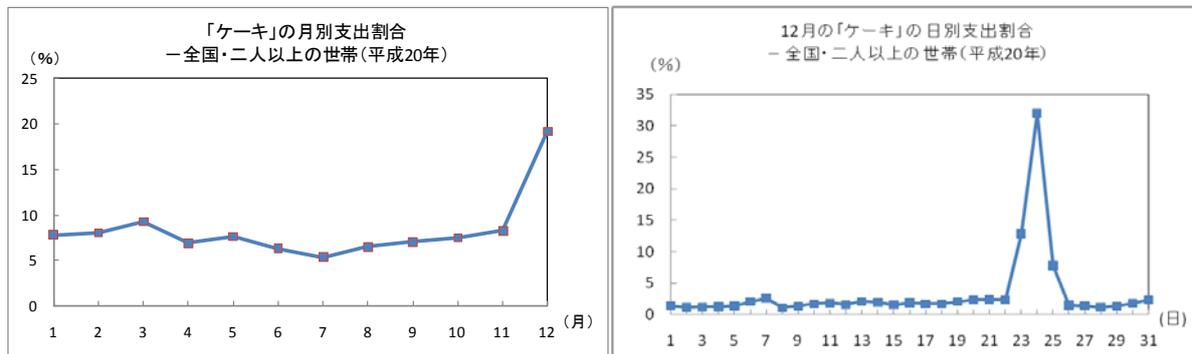
その他、家計調査に関するトピックスが知りたい方は、家計ミニトピックスをご参照ください。家計ミニトピックスでは、特定の日に消費が増える品物はもちろん、地域性や世帯主の属性(年齢別等)による支出金額の違い、時代の移り変わりによる支出金額の変化など季節やそのときどきの話題を紹介しています。

なお、特定の日に消費が増える品物として、チョコレート(2月)、ケーキ(12月)、テレビゲーム(12月)などがあります。

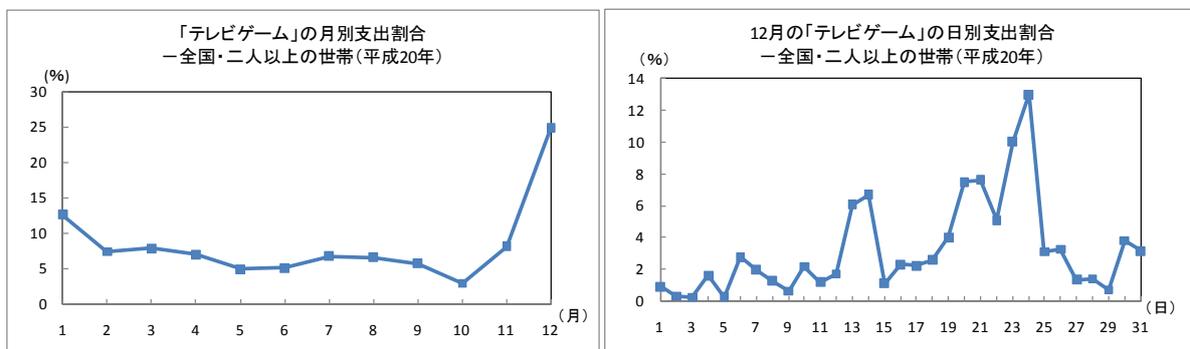
### ○チョコレート



## ○ケーキ



## ○テレビゲーム



このほか、品物による日々の消費支出のパターンは、以下の統計表をご参照下さい。

○統計局ホームページ 家計調査 統計表一覧

<http://www.stat.go.jp/data/kakei/index.htm>

家計収支編 > 4. 詳細結果表 二人以上の世帯(平成12年から掲載) \*  
月次 > 該当年月 > 第6-16表 品目分類による日別支出

### <参考>

#### 土用

暦法で、立夏の前18日を春の土用、立秋の前18日を夏の土用、立冬の前18日を秋の土用、立春の前18日を冬の土用といい、その初めの日を土用の入りという。普通には夏の土用を指している。

#### 丑の日

十二支の丑にあたる日。夏の土用の丑の日には夏やせの薬に鰻の蒲焼を食べ、また、灸をすえる。

【出典：広辞苑】