

統計学習用補助教材について

コンテンツ予定

統計基本用語集（案）・・・・・・・・・・P 1

統計グラフのいろいろ（案）・・・・・・・・P 7

家計調査の見方・使い方（案）・・・・・・・・P 1 7

物事を正確に捉えるための統計　・・・・・・・・P 2 5

授業導入の素材（案）・・・・・・・・・・P 3 7

統計基本用語集（案）

（ア行）

移動平均 moving average

（ナ行）

二項分布 binomial distribution

（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 ratio

誤差 error

五数要約 five-number summary

（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

標準偏差 standard deviation

標本、標本調査 sample, sample survey

分散 variance

分布 distribution

平均値 average

母集団 population

（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

（マ行）

無作為抽出 random sampling

（収録用語 39語）

（タ行）

中位値 median

統計地図 statistical map

度数分布表 frequency distribution

ドットプロット dot plot

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 カ月移動平均」とは、1 月データを前年 1 2 月、1 月、2 月の 3 つの平均で置き換え、2 月データを 1 月、2 月、3 月の 3 つの平均で置き換え、・・・としたものをいう。

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：今日の降水確率は 80% です。

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。

サイコロを振ったときにどの目が出るかは、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 $3.14159265 \cdots$ 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 ratio

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 組の男女別の構成比は、男子 46%、女子 54% だ。

誤差 error

真の値からのずれの大きさ。

標本調査では、必ずしも真の値は分からないが、「真の値はおおよそ 1800±50 円の範囲にあるだろう」ということが分かる。この場合、「誤差は 50 円以下」ということになる。

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 " 、中位値と同じこと）
- ・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、「171cm」が最頻値である。

X	170cm	171cm	172cm	173cm	174cm	175cm
度数	10	50	42	30	10	6

散布図 scattering diagram

2 つの量の関係を見るために、グラフの縦軸・横軸にそれぞれの量を取り、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（これは中位値と同じ）、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、「真の値の 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、ということを意味している。

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を、平均 μ 、分散 σ^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 $X_1, X_2, \dots, X_n$ を独立に取りだしたとき、もともとの確率分布がどのようなものであろうとも、 $X_1, X_2, \dots, X_n$ の平均は近似的に正規分布になることが知られており（中心極限定理）、統計学では正規分布はとくに重要である。

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つの量の間で、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つの量の間に関連があるという。

相関係数 coefficient of correlation

相関の強さを表す数で、-1から1の間の値をとる。散布図を描いたときにデータが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), rate of increase(decrease)

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、

- ・ この期間の増減は、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

データがある値を取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。

中位値 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。
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、データの大小を、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

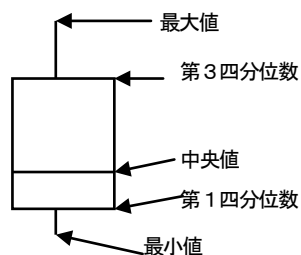
表が出る確率が p であるようなコインを n 回投げたときに、表が全部で何回出るか、という確率分布。 $P(X=r) = {}_n C_r p^r (1-p)^{n-r}$ となる。記号 $B(n, p)$ で表される

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

ヒストグラムや散布図を描いたとき、全体の傾向とは異なって、離れた所に位置しているデータのことを外れ値という。

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見るのに用いる。棒の一番高いところが、最頻値である。

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。記号 σ で表す。分散の平方根に等しい。

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

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

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。

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 σ^2 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。

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

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、「算出平均」ともいう。

母集団 population

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。

統計グラフのいろいろ

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、「〇〇について知りたい!」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。

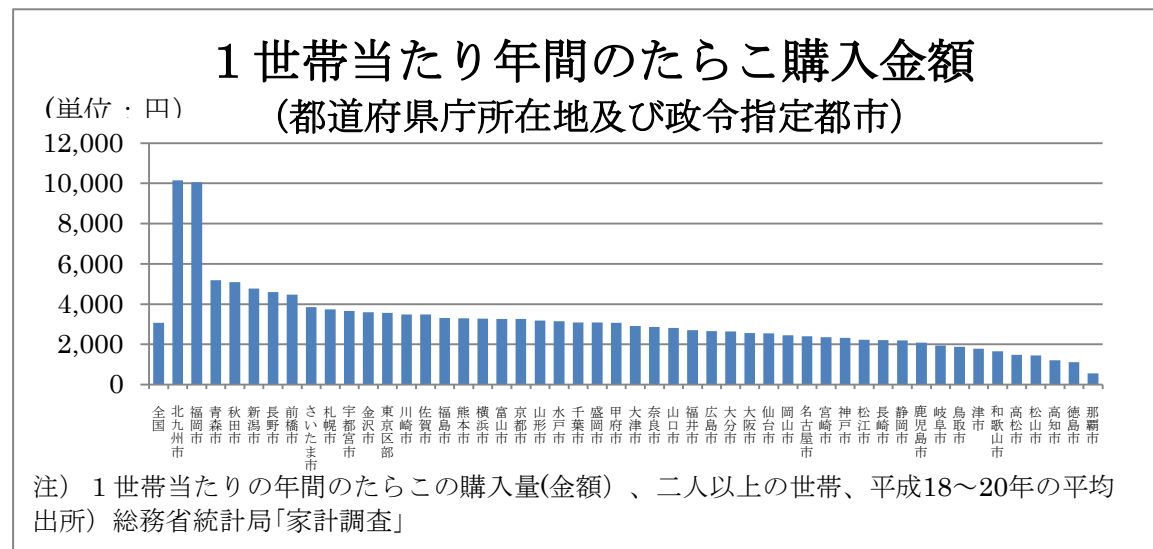
グラフにはいくつかの種類があり、それぞれ、得手・不得手があります。自分が伝えたい目的に応じて、適切なグラフを使うことにより、説明力もぐっと高まります。ここでは、そういったグラフの種類やそれぞれの用途、注意点について説明します。

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、変化の方向を見る。
円グラフ	全体の中での構成比（百分率）を見る。
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見る。
レーダーチャート	複数の指標をまとめて見る。
散布図	2種類のデータの相関を見る。

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分かりますね。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。

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。

さっきの例だと、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、一目で分かりますよね。

○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。

のような方法をとります。

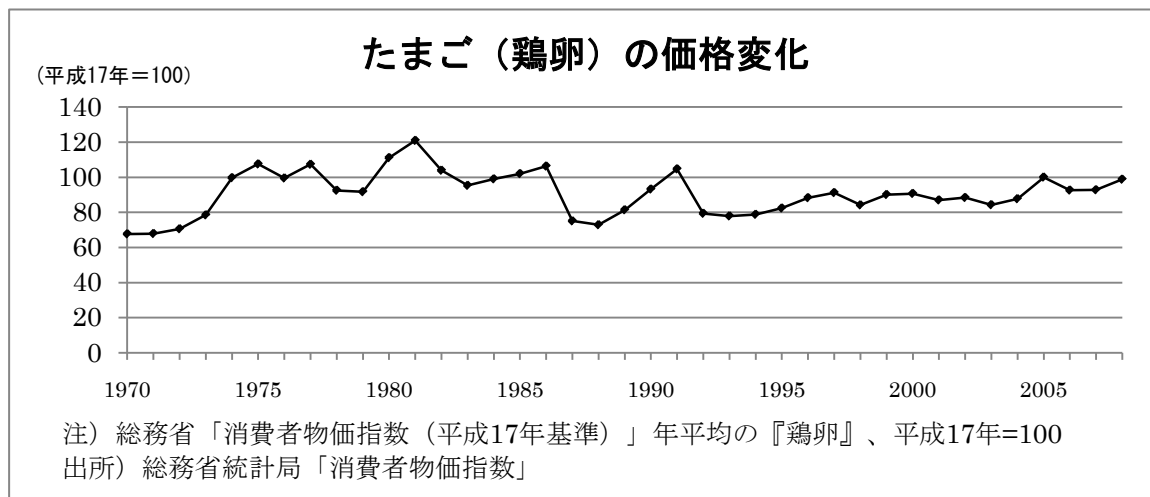
○やってみよう

総務省統計局「家計調査」では、たらこ以外の品物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あ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。

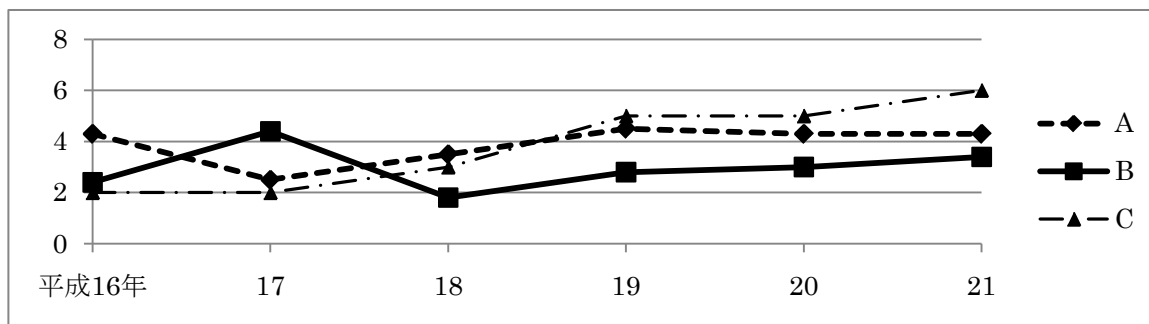
これをみると、たまごの値段もかなり上下動していることが分かります。ニワトリの飼料は輸入穀物に頼っていること、養鶏場の暖房等の費用もかかることから、たまごの価格も、国際的な穀物価格や原油価格の動向に左右されやすい、という側面もあります。

○どんなときに使うの？

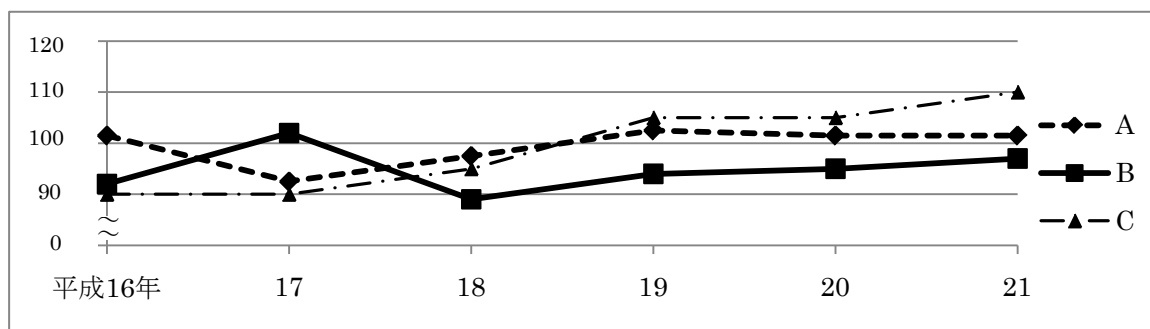
折れ線グラフは、横軸に年や月といった時間を、縦軸にデータ量を取り、それぞれのデータを折れ線で結んだグラフです。線が右上がりならその期間はデータが増加、右下がりならデータが減少していることになるので、データの増減を見るのに適しています。

○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、こっちは減少」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、線を色分けしたり、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。



○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。

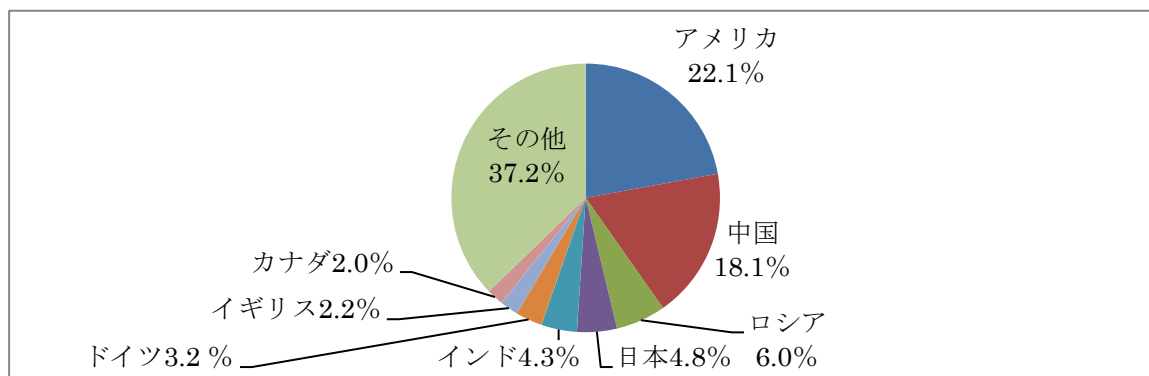
<http://www.stat.go.jp/data/cpi/historic.htm>

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（CO₂）排出量を減らそう」「日本が率先して CO₂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 CO₂ 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 CO₂ 排出量が飛び抜けて多いこと、上位 4 カ国（アメリカ、中国、ロシア、日本）で全体の過半数を占めている、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分かるので、構成比を示すのに使われます。

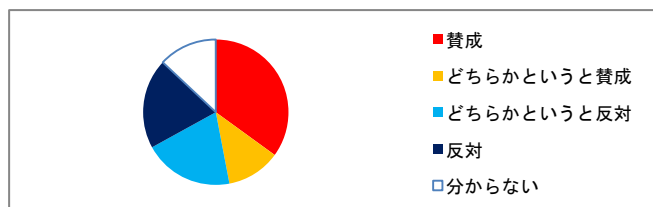
○注意することは？

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、一番最後に持ってきます。

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「どちらかという賛成」を合わせて過半数かどうか、ということが一目で分かるようにする方が便利だからです。



○やってみよう

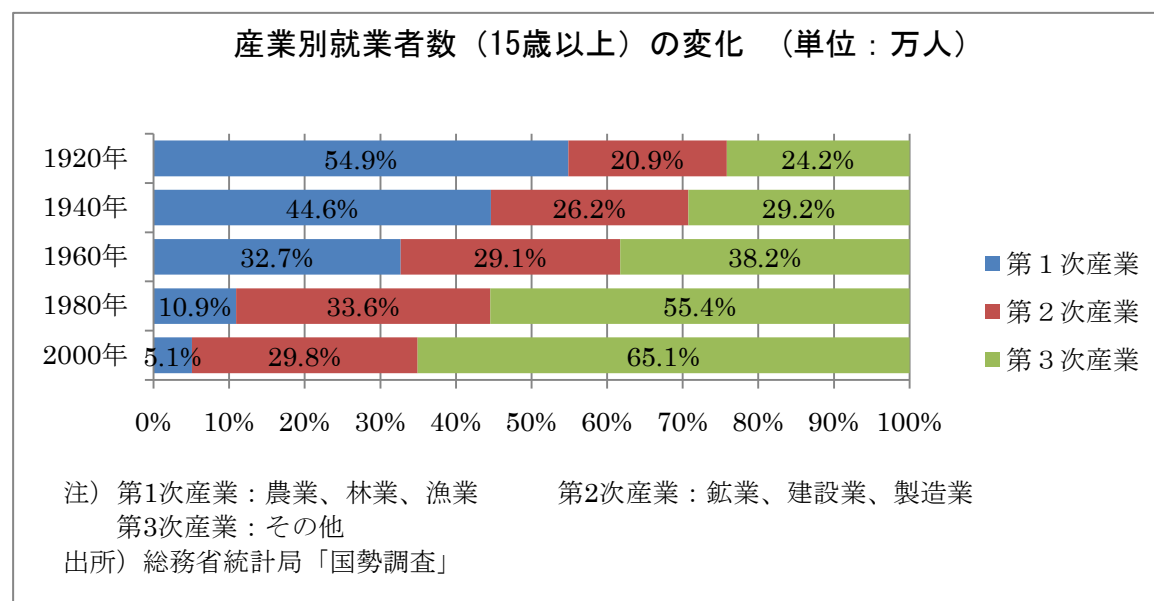
総務省統計局「家計調査」では、ご家庭で、どのようなものを買っているか、調べています。円グラフを描いて、どのようなものにお金を使っているか、見てみましょう。

<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%程度であることが分かります。

○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。

○注意することは？

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分からなくなってしまうので、順番は変えないでおきます。

（上の例では、2000年では、第3次産業の就業者数が一番多くなっていますが、順番を入れ替えずに、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。

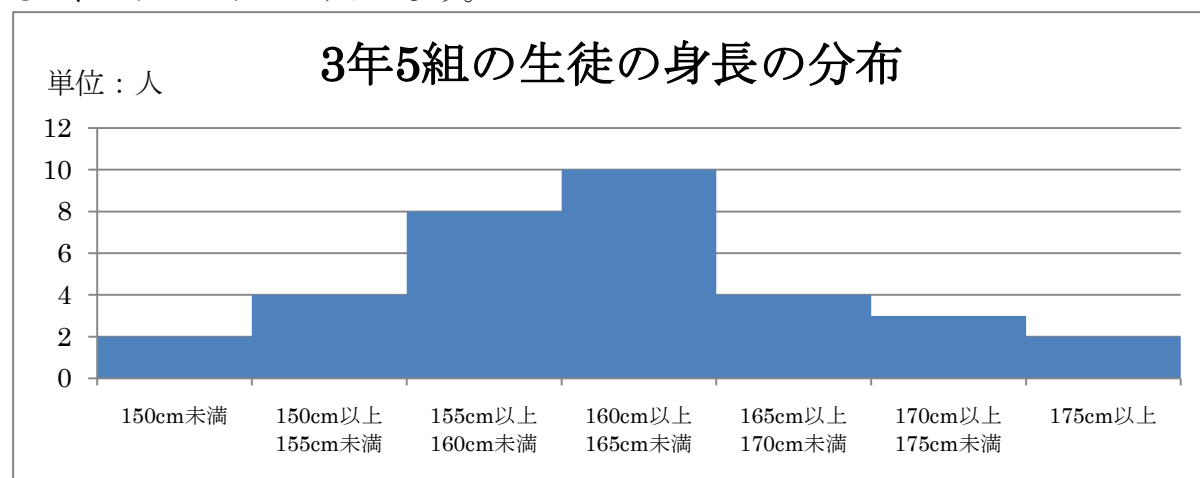
年代別に、時間の使い方にどのような違いがあるか、帯グラフを描いて比べてみましょう。

<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分かりますね。

○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。

○注意することは？

データの階級幅が異なるときには、通常の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、棒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）。

また、グラフの両端（上の例では「150cm 未満」と「175cm 以上」）も、高さを隣よりも低く表示します。

ヒストグラムを描く際には、「階級幅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。

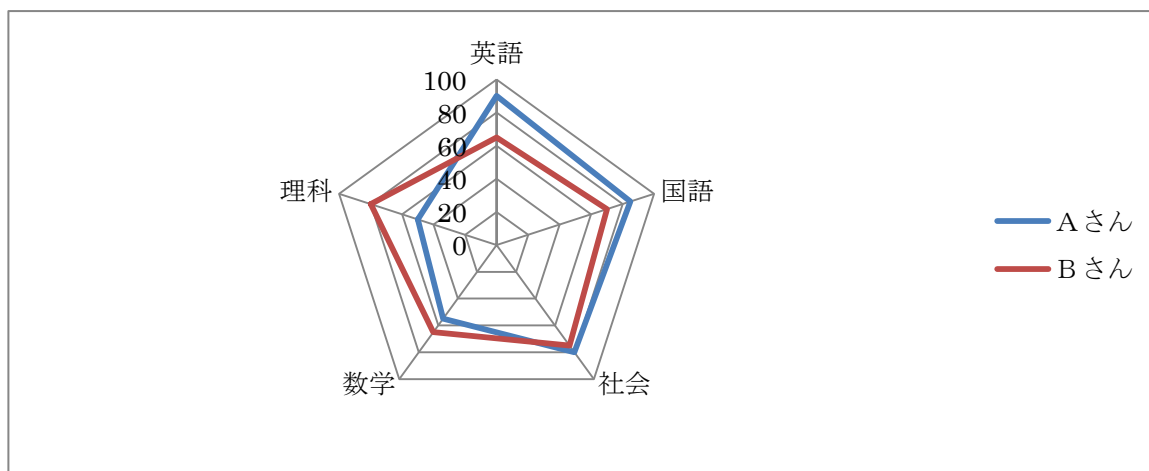
e-stat「都道府県・市町村のすがた」

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

6 レーダーチャート

○たとえば・・・

みなさんの中には、「理数系は強いが、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。



例えば、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。

円を、データの数に応じて扇形に分け、それぞれの半径に、データを表示します。そしてそれらを線で結んで、それが円に近いかどうかを見ます。

○注意することは？

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）よい」というふ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ければよいので、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い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よい」データに変換する必要があります。

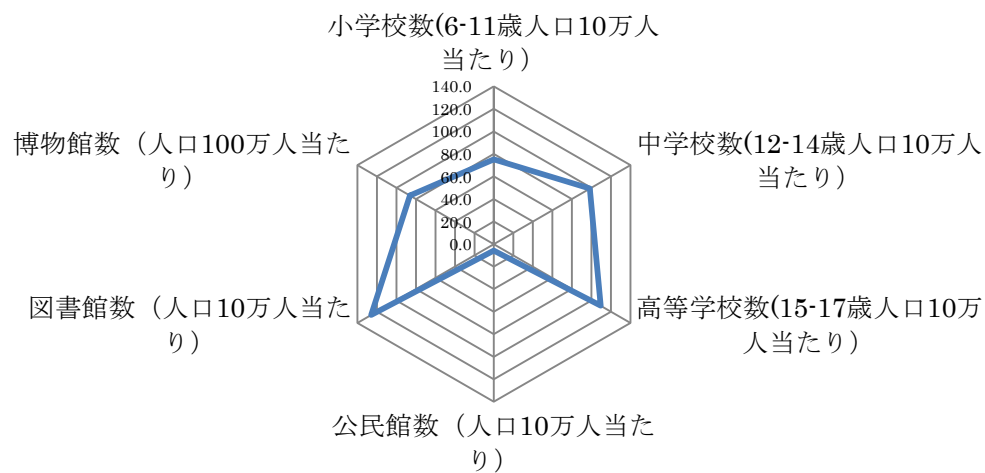
○やってみよう

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。

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>

東京都の学校数、社会教育施設数 (人口当たり、全国平均=100)



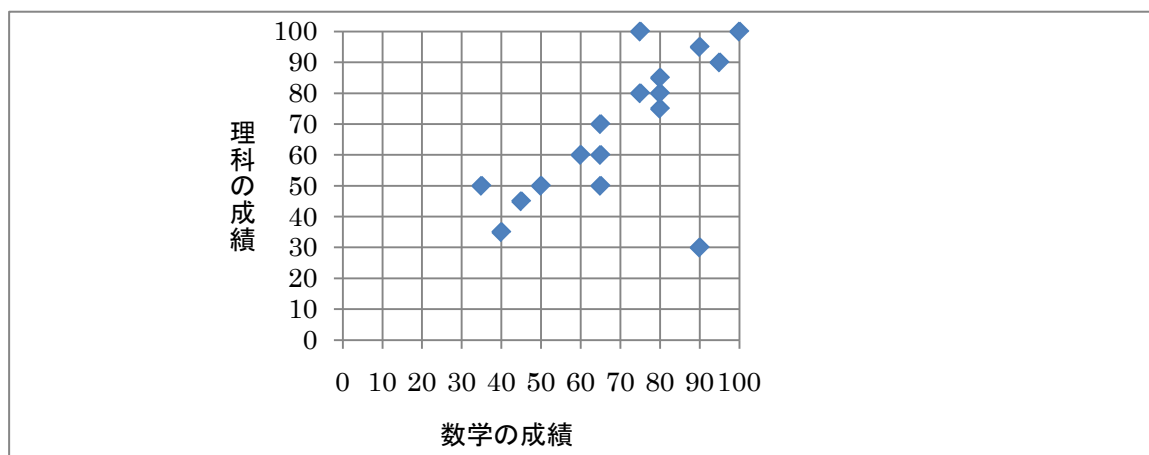
注) データは2005年値

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ところに点を打って示す（「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。

○注意することは？

散布図で分かることは、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ね。

○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>

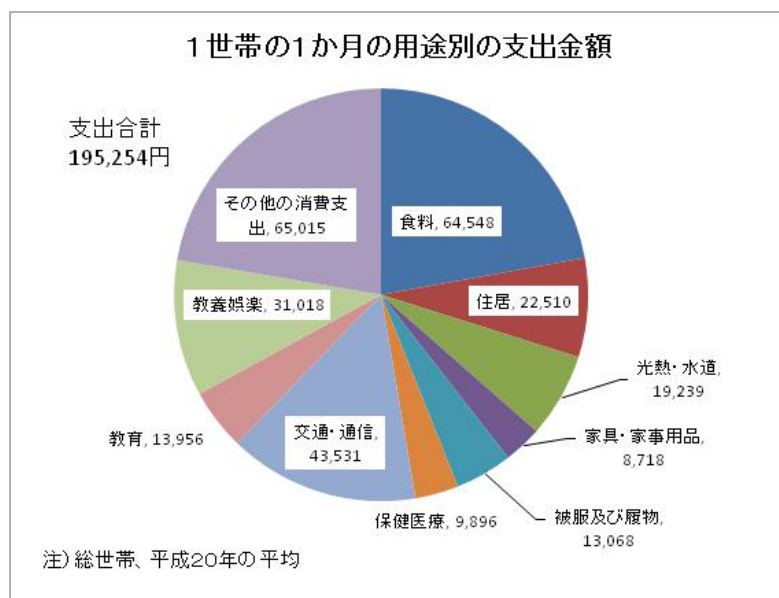
家計調査の見方・使い方

◇ 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 つの住宅に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調べています。

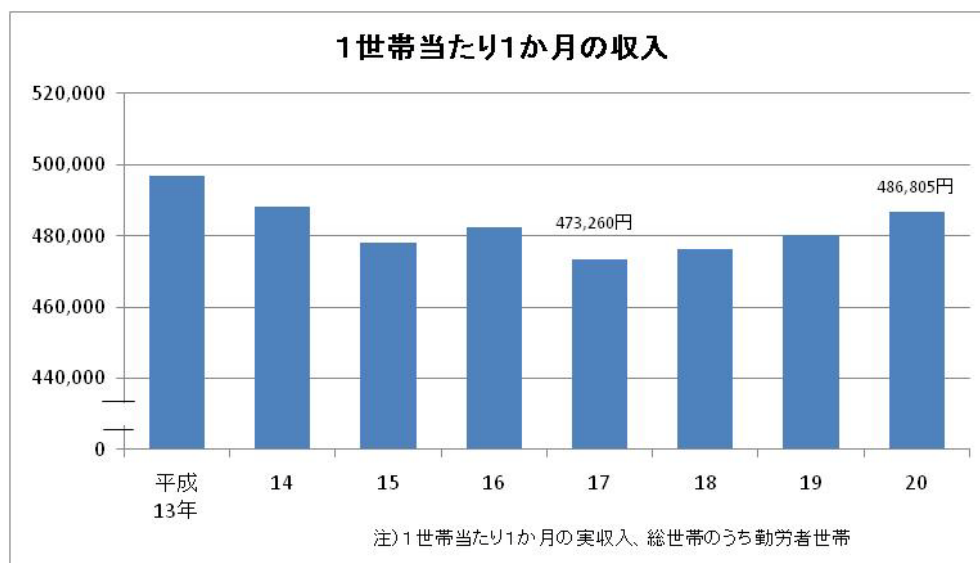
- 貯蓄とは、貯金や株式のことで、世帯がいざという時のために蓄えているお金のことです。
- 負債とは、住宅ローンのように大きな買い物をするために、銀行などから借りたお金のことです。

◇ 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、例えば牛肉をたくさん買っているのはどの地方ですか？
- 都道府県ごと、例えばみかんは何県でたくさん買われているのですか？
- 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？

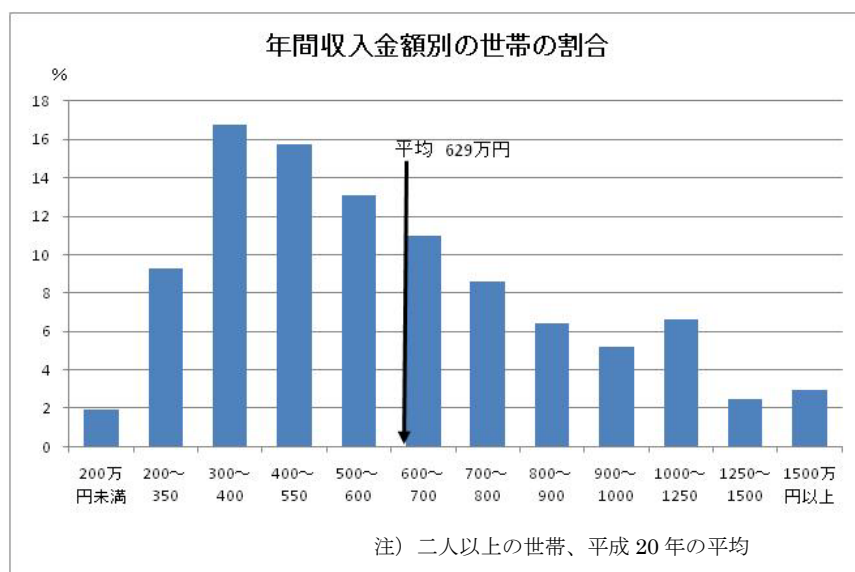


平成20年の勤労者世帯の1世帯当たり1か月の収入は486,805円です。景気の悪かった平成17年以前は毎年収入が減ることが多かったのですが、その後は景気が回復したことから徐々に収入が増えてきています。しかし、平成20年9月頃から景気が大きく落ち込んだので、平成21年の収入は大きく減るとわれています。

●**勤労者世帯とは**、いわゆるサラリーマンの世帯のこと、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](#)

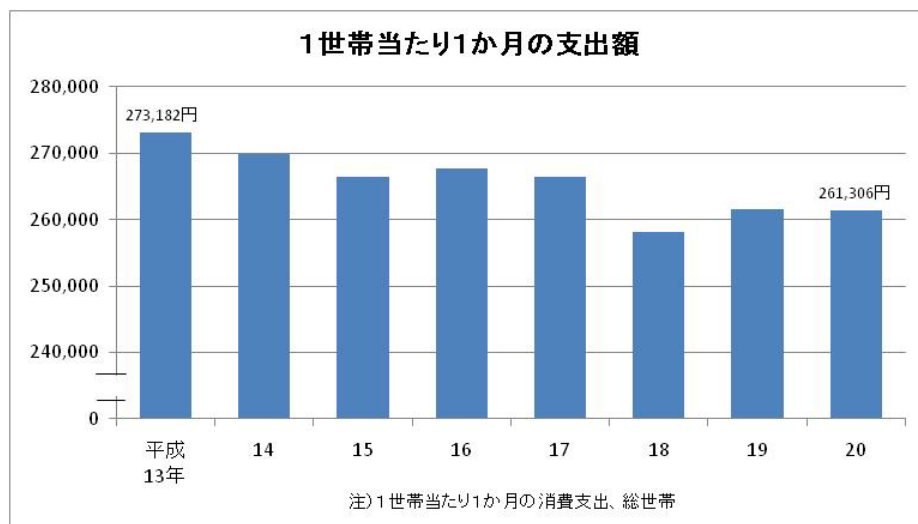
○ 1 年間の収入がいくらの方が多いのですか？



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多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](#)

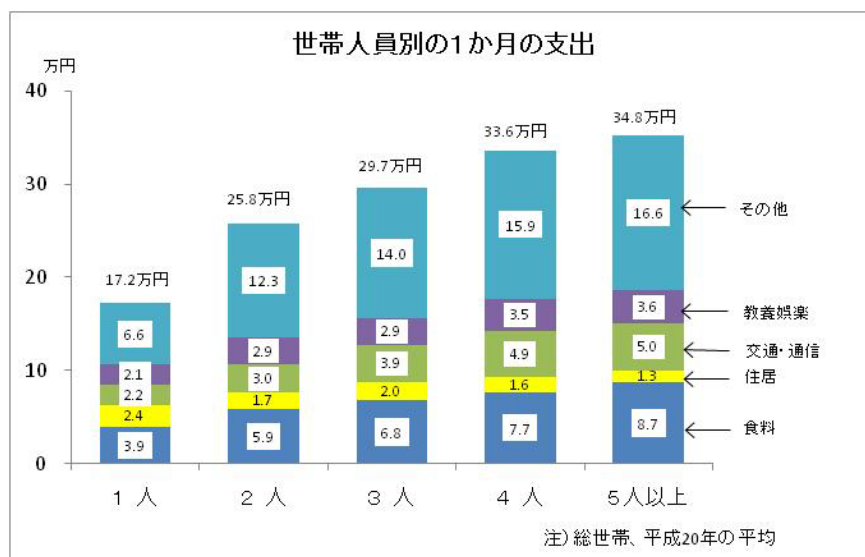
○ 世帯の 1 か月間の生活費（支出）はいくらですか？



平成 20 年の 1 世帯当たり 1 か月の支出額は 261,306 円です。7 年前に比べると 1 万円以上少なくなっており、支出を節約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](#)

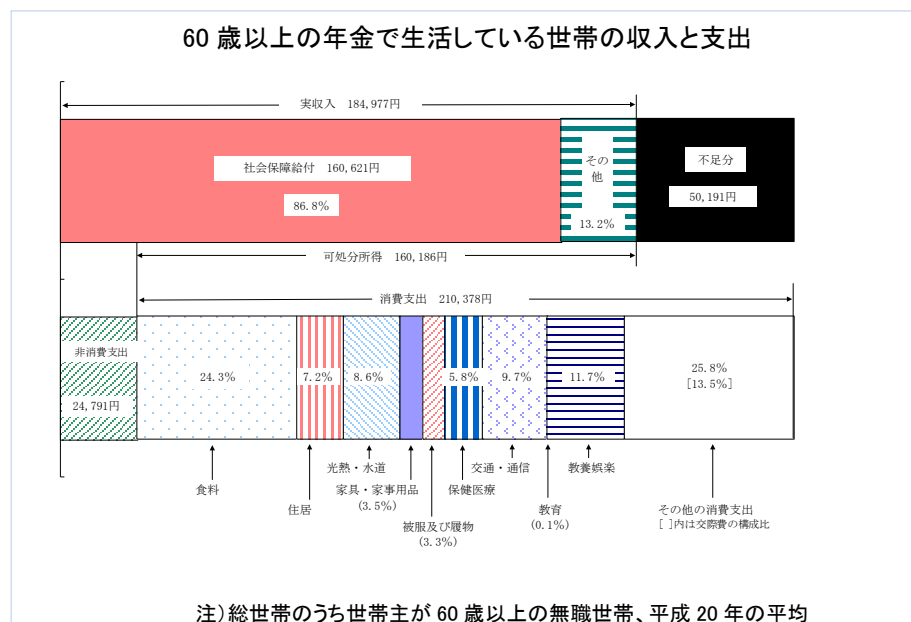
○ 一人暮らしの世帯の生活費はどうなっていますか？



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](#)

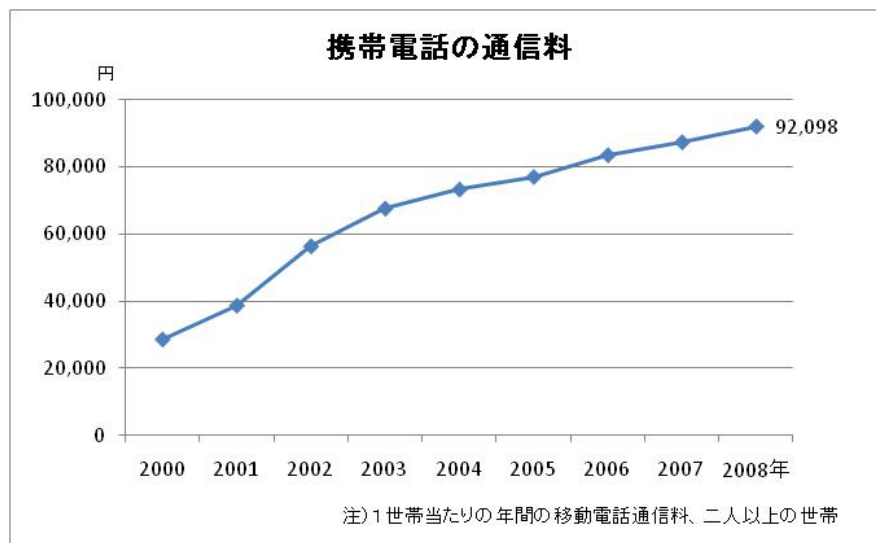
○ 高齢者の世帯の生活費はどうなっていますか？



高齢者で主に年金で生活している世帯では、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？

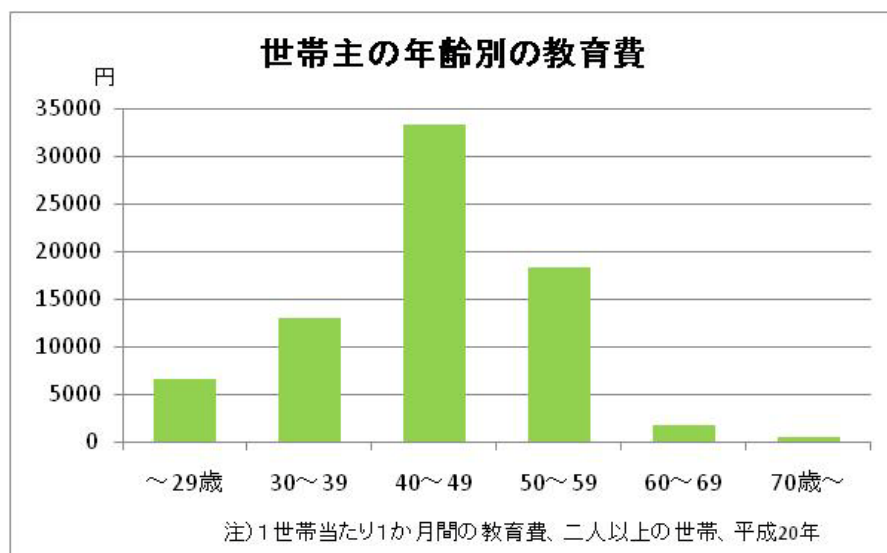


2008 年は 1 世帯当たり 1 年間で 92,098 円、携帯電話に支出しています。8 年前の 2000 年に比べると 3 倍以上と大きく増加していることがわかります。

●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？

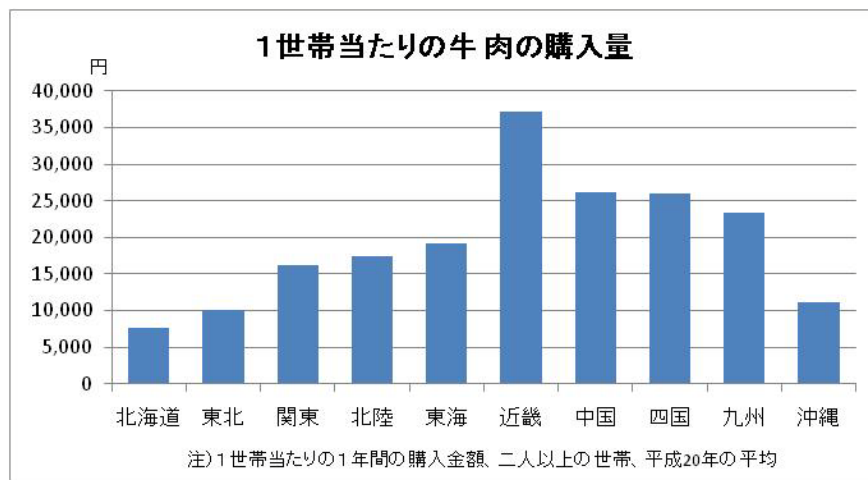


世帯主が 40 歳から 49 歳の世帯が他に比べて飛びぬけて多くなっています。子どもが高校や大学に入る子どもがいる親がちょうどこの年代にあたっているからです。

●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？



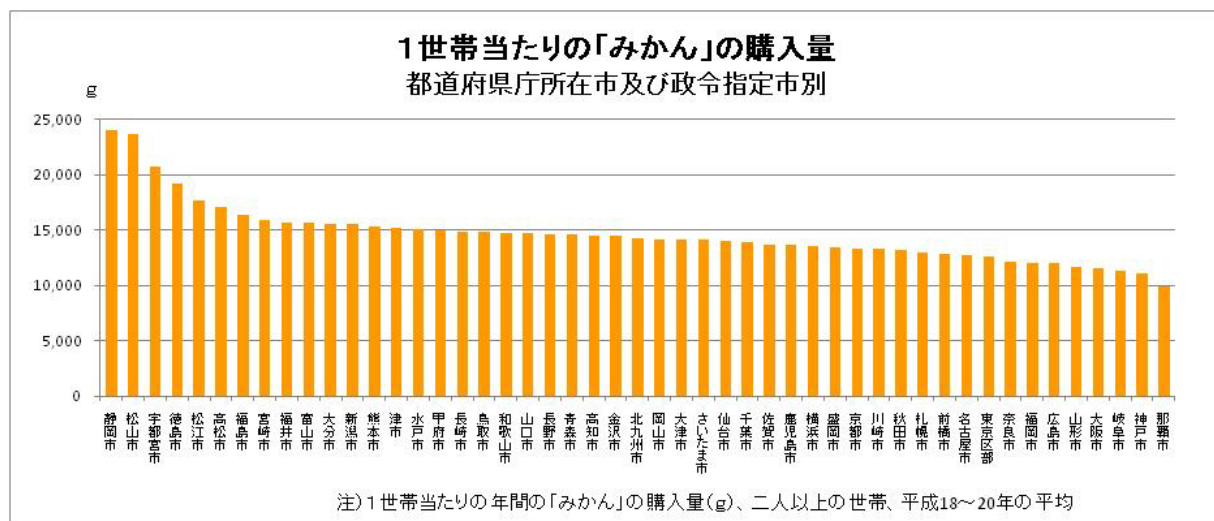
神戸牛（兵庫県）や近江牛（滋賀県）など有名ブランドの牛肉のある近畿地方が1年間に牛肉を5万円近く買っていて、他の地方よりダントツに多くなっています。

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](#)

○ みかんは何県でたくさん買われているのですか？

家計調査には、都道府県別の結果はありませんが、代わりに都道府県庁所在市別の結果があるので、その結果をみてみましょう。

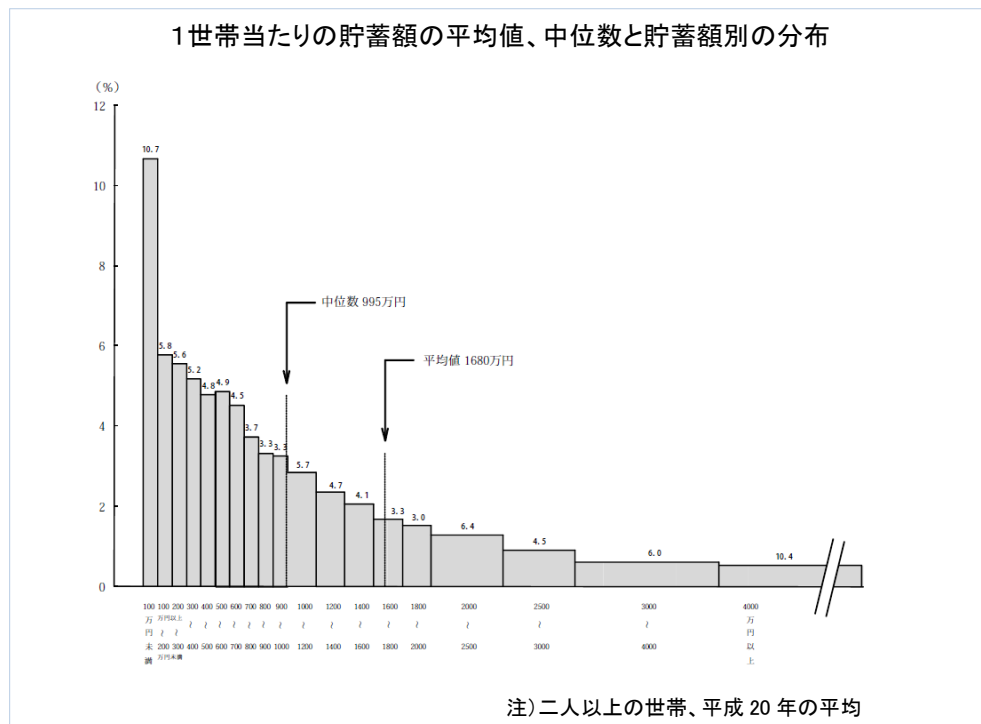


みかんの大産地の静岡県（静岡市）と愛媛県（松山市）の購入量がやはり多くなっています。産地では世帯もたくさん買っていることがわかります。

● どの都道府県庁所在市でどんなものがたくさん買われているか、**都道府県庁所在市及び**

政令指定都市別ランキングで見てみよう。

○ 1世帯あたりの貯蓄額はいくらですか？



平成20年の1世帯あ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。

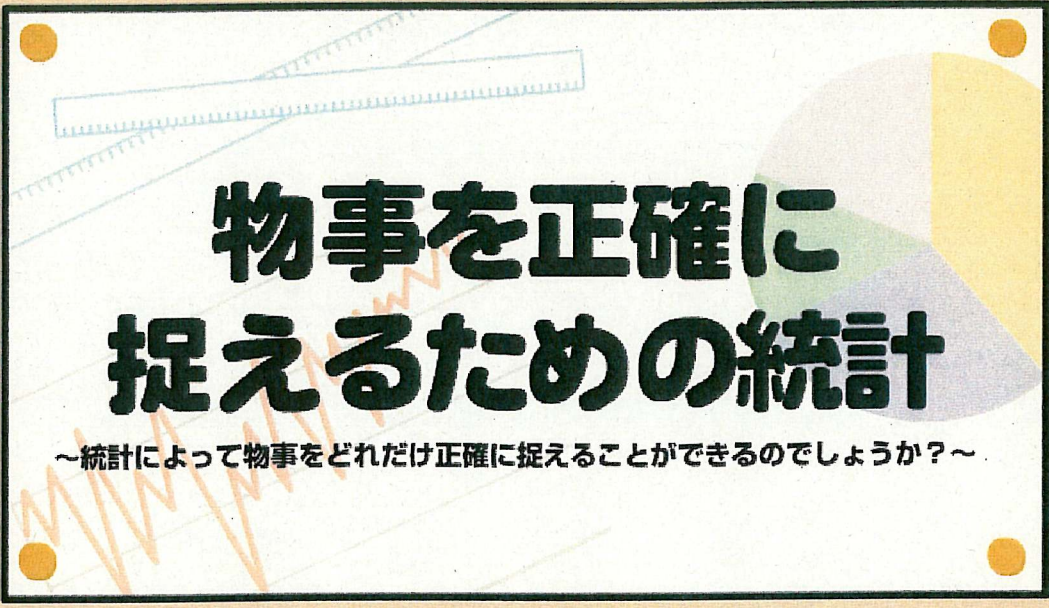
- ③ 総世帯 → 二人以上の世帯と単身世帯を合わせた結果です。

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。



物事を正確に 捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか？～

● 物事を正確に捉えるための統計 ●

～統計によって物事をどれだけ正確に捉えることができるのでしょうか？～

統計の意義と利用

標本調査

有意抽出法と無造作抽出法

無作為抽出の手順

「無作為抽出」＝「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

もどる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか？～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」＝「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

統計の意義と利用

・統計にはどのような意義があるのでしょうか？

統計によって集団の特徴を表すことには、どのような意義があるのでしょうか。

一つは、客観的、定量的な情報として表現できることです。例えば、人数が「多い」とか「少ない」といった定性的な表現では、正確な数を伝えることができませんが、統計として「何人」と言えば、正確に伝わります。

二つ目の意義は、統計により、集団の時間的な変化を捉えたり、地域間での比較をしたりすることができることです。例えば、日本の各地域間で、人口や産業の規模や構成を比較したり、日本と外国とで人口や経済の状況を比較したりすることが可能となります。

三つ目の意義は、統計を分析することによって、集団の特徴や物事の相互の関連性などを明らかにできることです。例えば、家計収入が増えるほど食料費の割合が下がるという「エンゲルの法則」がありますが、これは統計を分析することによって家計消費の関係や法則を把握することのできた事例の一つです。

・統計はどのような分野で用いられているのでしょうか？

統計は実際にどのような分野で用いられているか、いくつかの例を見てみましょう。

行政の分野では、景気、雇用情勢、国民の暮らし向きなどに関する統計が作成され、政策を立案し、実行するための基礎資料として活用されています。

工業の分野では、製品の不良を減らし、高い品質の物を生産するために、統計手法を用いた品質管理が行われています。

医学の分野では、薬や治療法の有効性を判断するために統計が用いられています。

企業経営の分野では、市場の動向や商品の売れ行きを分析したり経営戦略を立てたりするのに用いられています。

・様々な分野で統計が用いられるのはなぜでしょうか？

統計は、なぜこのように様々な分野で用いられるのでしょうか？

どのような分野のどのような問題でも、問題に対して適切に対応するためには、事実を客観的に把握し、定量的に分析し、きちんとした根拠に基づいて判断することが必要です。統計は、「事実を客観的に把握し、定量的に分析する」ための強力かつ有効な手段として、様々な分野で用いられているのです。



上にもどる



次のページ

× とじる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか?～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」＝「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

標本調査

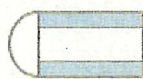
政府の統計調査は、最初は国勢調査のような全数調査や特定の対象に関して調べる典型調査として行われていましたが、戦後は標本調査の理論や手法が普及し、今日では、数多くの統計調査が標本調査として行われるようになっていきます。

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、元の集団全体の状況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。

標本調査の重要な目的は、標本を用いて母集団の状況をできるだけ正確に推計することです。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。



前のページ



上にもどる



次のページ

× とじる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか?～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」＝「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

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。この方法によると、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ぶのではない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な判断を排除して、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。一般に、無作為抽出によってある程度多数の標本を集めると、統計学でいう「大数の法則」によって、推計結果は安定し、精度の高い推計結果が得られることが知られています。



前のページ



上にもどる



次のページ

× とじる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか?～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」＝「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

無作為抽出の手順

無作為抽出を行うとすれば、実際にどのようにやればいいのでしょうか。

無作為抽出というのは、単に、「でたらめ」に選ぶこと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。

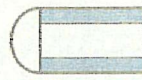
まず、母集団を構成する全員のリストを用意します。このような母集団全員を含むリストは、標本フレームまたは標本枠と呼ばれることがあります。

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いいのです。



前のページ



上にもどる



次のページ

× とじる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか?～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」=「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

「無作為抽出」=「でたらめ」ではない

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、パソコンの普及率を調べるために、標本調査を行うとします。標本調査を行うために、簡略な方法として、その町の一番大きな駅の前に立って、通行する人に誰でもかまわず無作為に声をかけて、パソコンを持っているかどうか聞いてみましょう。

この方法は、標本調査で言う無作為抽出法となっているでしょうか。

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、町の住民全てを代表しているかと言えば、そうとは限りません。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、町の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。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、駅を利用する人で、かつ、調査の時間帯に駅前を歩いていて、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、何か一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。



前のページ



上にもどる



次のページ

× とじる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか?～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」＝「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

集落抽出法

母集団の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。

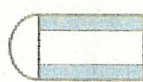
このような場合に用いられる代表的な方法として、集落抽出法と呼ばれる方法があります。これは、個別の調査単位を直接選ぶのではなく、まず調査を行う地域を選び、次に、その地域の中に含まれる調査単位を全て調べる方法です。具体的には、国勢調査の調査区の中から、調査の対象とする調査区を一定数、無作為抽出法により選び出し、その調査区の中のすべての世帯を調査するのです。

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、日本全国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、毎回の国勢調査によって、各調査区に関する統計が整備されています。このため、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。



前のページ



上にもどる



次のページ

× とじる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか?～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」=「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

■ 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。

二段抽出法では、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、その地域の中ですべての調査対象の名簿を作り、その中からもう一度、標本を選び出して調査するのです。

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、調査結果の精度を向上させるためには得策です。



前のページ



上にもどる



次のページ

× とじる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか？～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」＝「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

標本誤差

・切っても切り離せない標本誤差の存在

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。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を無作為に選定することによって起こる差を「標本誤差」といいます。

標本誤差を測る上では、次の二つの問題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、真の値が分からないということです。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、標本誤差を直接的に測ることは困難であり、現実には、標本誤差を得るには、一定の理論式に基づいて推定する方法が一般的に用いられています。

もう一つの問題は、調査結果と真の値との差は、抽出された標本ごとに異なるということです。もし、標本調査を何度か繰り返して行くとすれば、抽出のたびごとに異なる対象が選ばれ、その結果、得られる数値は変動します。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。このように標本誤差とは確率的なばらつきの幅を示す尺度です。

標本誤差とはそのような性質のもので、単なる計算間違いや、測り間違いのような意味での誤差とは性質が異なり、推定結果に、例えば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。

標本誤差は、むしろ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なのです。

・標本誤差を調べてみましょう！

標本誤差とはどのようなものかを理解するために、架空の数値実験を例として、考えてみましょう。

ここに10,000世帯が住んでいる町があり、この町でのパソコンの普及率を標本調査により調べることにしましょう。

この町でのパソコンの普及率は50%、すなわちパソコンを持っている世帯は5,000世帯あるものと想定します。ただし、現実の世界では、そのことについてはこれから調査をするのですから、この真の値は誰にも分かりません。しかし、ここでは、実験の前提としてこのような想定を置いて事例を検討してみましょう。

ここで標本調査として、単純任意抽出法によって100世帯を選んで調査し、その調査結果から町全体のパソコンの普及率を推定することにします。

ただし、標本調査の結果は偶然によって左右されますので、この実験では、このパソコンの普及率の推定値がどのくらい偶然によってばらつくものなのか試してみるために、100世帯を無作為に抽出するという調査を1000回繰り返して行い、そのばらつき具合を調べてみます。

1000回の実験の結果、パソコンの普及率は次のよう分布を示します。

パソコンの普及率	40%未満	40%以上 45%未満	45%以上 50%未満	50%以上 55%未満	55%以上 60%未満	65%以上	合計
発生回数	20	140	340	340	140	20	1000

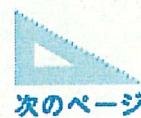
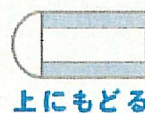
この実験結果からどのようなことが読み取れるでしょうか。

1000個の推定結果の中には、真の値である50%から大きく外れることもあります。全体として真の値を中心にした左右対称なばらつきを示しています。このように、偶然によって調査結果はばらつくのです。

現実の社会で標本調査を行う場合には、1回しか調査できませんので、実際に得た調査結果を見ると、このような形でばらつく可能性のある推定結果のどれか一つを見ているということになります。

この実験の場合、パソコン普及率の推定結果は、1000回のうち680回は、真の値である50%を中心として $\pm 5\%$ の幅、つまり45%以上55%未満の範囲内に入っています。このように標本誤差とは、推定値がどの程度の確率で真の値の近くに分布するかということを表したものです。

繰り返しになりますが、標本誤差は、標本調査の結果として得られる推定値が確率的にばらつく幅を示したものですから、いくつ多い、又は少ないと確定できる性質のものではないので、標本誤差を修正した推定値を示すことはできません。むしろ、標本誤差は、調査結果を使う時に、どの程度の幅を持って見るかを示す情報です。



× とじる

物事を正確に捉えるための統計

～統計によって物事をどれだけ正確に捉えることができるのでしょうか?～

統計の意義と利用

標本調査

有意抽出法と無作為抽出法

無作為抽出の手順

「無作為抽出」＝「でたらめ」ではない

集落抽出法

二段抽出法

標本誤差

全数調査の意義

全数調査の意義

全数調査は、集団の中を全て調査しますので、集計した結果には標本誤差が含まれません。したがって、全国など全体の結果はもちろんのこと、都道府県別、市町村別など詳細な区分に分けて集計した結果でも統計として利用できます。実際、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も公表されています。

また、全数調査は、統計体系の中で、次の2つの観点から統計の正確性を担保する根源となっています。

一つは、定期的に他の統計の基準となる数値を与えるということです。例えば、5年ごとに行われる国勢調査の統計は、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。また、国民経済計算などの加工統計でも、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と改定が行われます。

最新の国勢調査結果が得られてから、次回の国勢調査結果が得られるまでの間は、推計によって補間が行われますが、推計には誤差がつきものですので、その誤差を定期的に補正するために、全数調査としての国勢調査結果が使用されるのです。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。

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を抽出して調査するものですので、調査の企画・設計の段階では、その集団に含まれる対象すべてを含んだリストが必要です。そのようなリストがないと、標本調査を設計することができません。

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のリストは、事業所・企業統計調査によって整備されるリストがしばしば用いられます。つまり、事業所・企業統計調査は、他の多くの標本調査の最も基礎となる情報を与えるものとして活用されています。

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います。

全数調査は、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。



前のページ



上にもどる

× とじる

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」

【例1】

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」

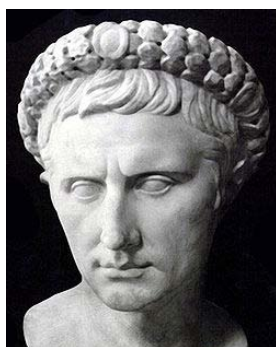
これらは別々のルートをたどって、19世紀半ば、ケトレー（Adolphe Quetelet 1796-1874）が社会統計を科学的に作成・分析するために確率論を導入したことで、社会現象・自然現象いずれも数量的にとらえる「統計」として形を整えました。ケトレーは、母国ベルギーの統計制度の整備や公的統計の改善に努めただけでなく、国際的な統計の比較可能性を高めるべく国際統計会議の設立にも力を尽くし、その功績から「近代統計学の祖」とされています。

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。



Adolphe Quetelet

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「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の流れに属しますが、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、単に物事の状態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中に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・記述することに終止し、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。彼の著した「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、それまで偶然の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。

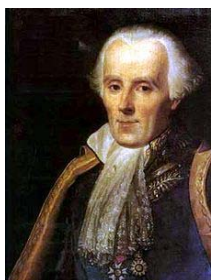
今日、統計には欠かせない「標本空間」の基本的な考え方は、16世紀にサイコロ賭博やトランプゲームにおける偶然の仕組みを数学的に研究したイタリア人カルダーノ（Geloramo Cardano 1501-76）によっていますし、地動説を唱えたことで有名なガリレオ（Galileo Galilei 1564-1642）はトスカーナ大公から命じられて「サイコロゲームについての考察」という小論を書いています。

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 1749-1827）といった一流の数学者たちの研究を経て大成しま



Pierre de Fermat



Pierre-Simon Laplace

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 1707-89）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計に応用されました。

<了>

【例2】

「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 Statistique」の訳書として、簗作麟祥（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」

「榊・さかき」など）さえ考案しました。これが^{寸多知}多智智^{寸知久}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。

石橋重朝という当時の人物（明治18年から26年まで内閣統計局長）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。

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）を翻訳した簗作や「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、今日でも中国語として使われています。

<了>

【例3】

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。



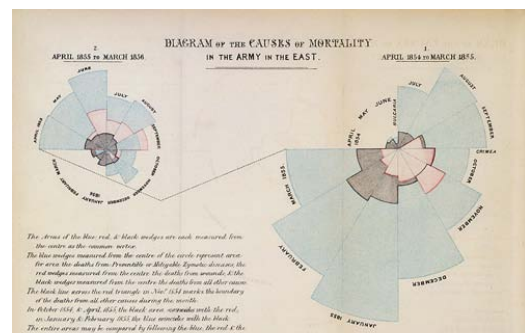
Florence Nightingale

彼女は、上流階級の家庭に生まれ、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＝ケトレー（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（ロシアとトルコの間の戦争で、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、統計になり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、ケトレ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>

【例 4】

特定の日に消費が増える品物

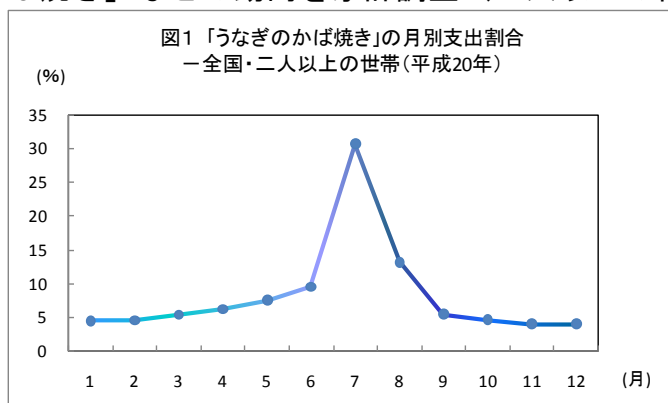
毎年 7 月になると、「土用の丑の日」ののぼりが店頭で目立ってきます。

「土曜の丑の日」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。

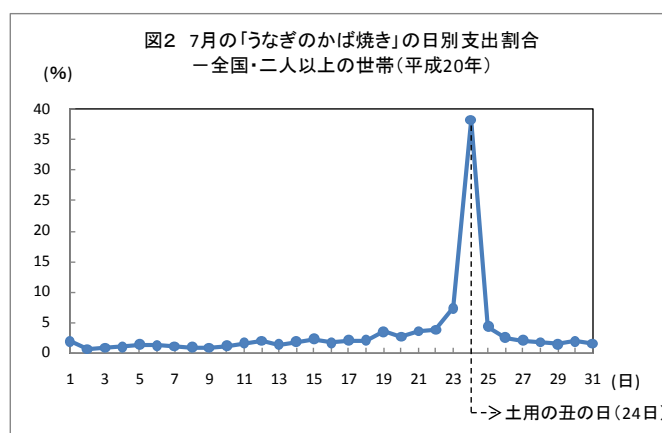
文政 5 年（1822 年）の、当時の話題を集めた『明和誌』（青山白峰著）によると、夏に鰻が売れないことを悩んでいた鰻屋が、何とか売るため源内の所に相談に行ったところ、源内は、「丑の日に『う』の字が附く物を食べると夏負けしない」という民間伝承からヒントを得て、「本日丑の日」と書いて店先に貼ることを勧めたようです。すると、物知りとして有名な源内の言うことなら、ということで、その鰻屋は大変繁盛したようです。その後、他の鰻屋もそれを真似るようになり、土用の丑の日に鰻を食べる風習が定着したといわれています。

これにちなんで「うなぎのかば焼き」などの動向を家計調査（二人以上の世帯）の結果から見てみましょう。

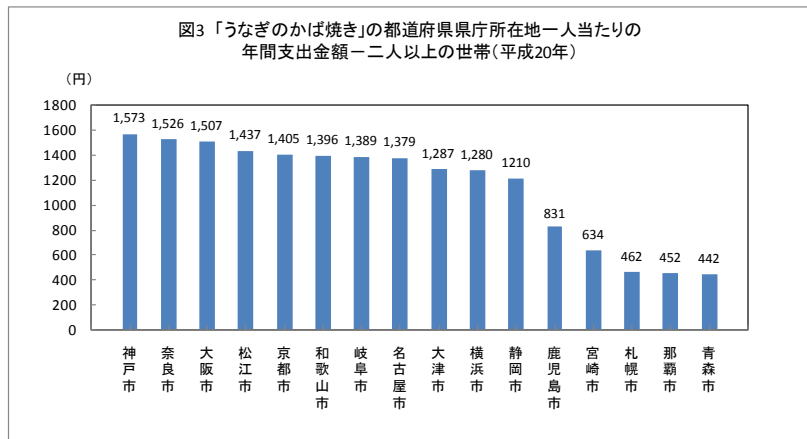
「うなぎのかば焼き」について、平成 20 年の 1 世帯当たり年間支出金額に対する各月の支出割合をみると、1 年の中で土曜の丑の日を含む 7 月が最も高く、30.7%を占めています。（図 1）



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」を食べるという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地別にみると、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)

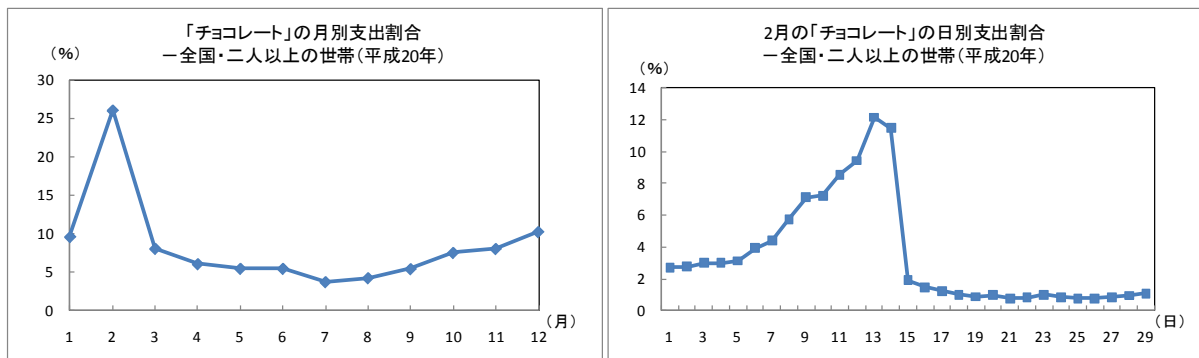


昔から、「うなぎのかば焼き」は、関西では「まむし」と呼ばれて有名なことなどが、ランキングに表れているのでしょうか。

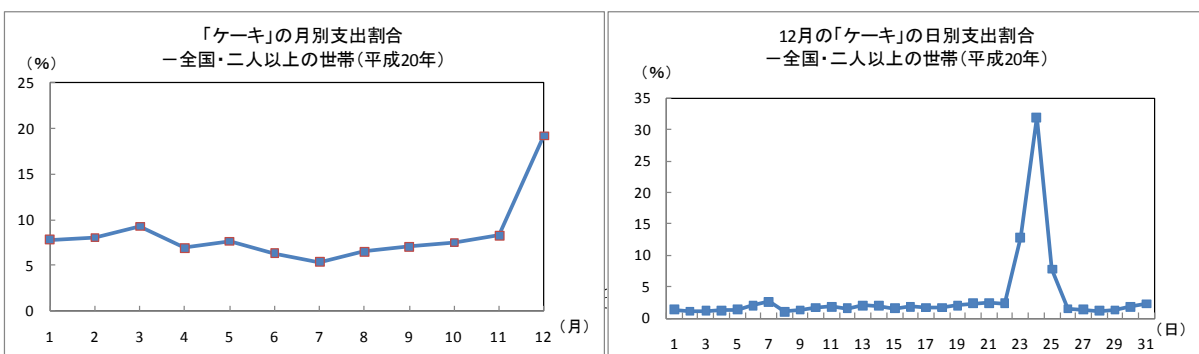
一方で、平成20年の「うなぎ」の国内収穫量(農林水産省漁業・養殖業生産統計)の全国に占める割合をみると、鹿児島県(35.5%)、愛知県(29.9%)、宮崎県(16.7%)、静岡県(7.8%)などが高くなっており、必ずしも生産地でたくさん食べられているわけではないようです。

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。

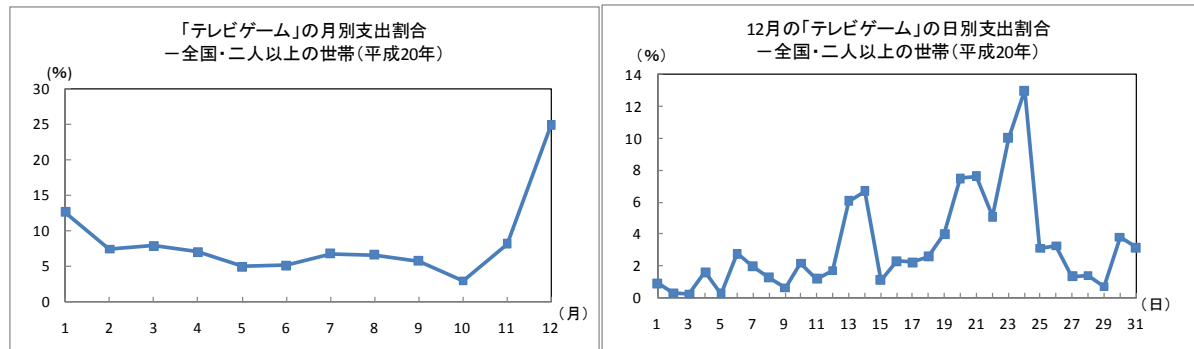
○チョコレート



○ケーキ



○テレビゲーム



<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】