

【データの代表】（その1）

データを代表するような値といえば…

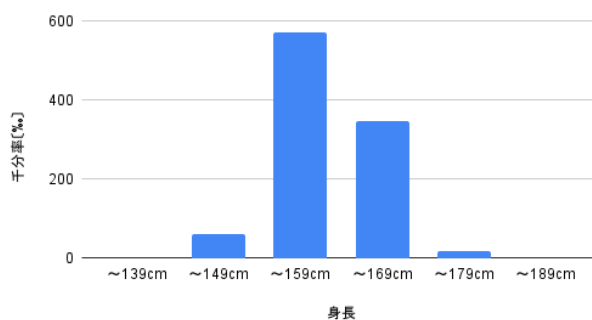
集めたデータを簡単に表したいという場合、どうすればいいでしょうか。「こんなデータが集まりました」ということを縮約して知らせる場合によく使われる統計量としては（ ）が代表といえるでしょう。実際に私たちが最も一般的に使っているのは（ ）とも呼ばれるものになります。

例1. 身長分布

総務省統計局の統計データ（2021年度 学校保健統計調査）から、全国の高校2年生（16歳）女子の身長のデータを見てみましょう。

グラフにしたのが右図です。平均値はおおよそ（ ）cmです。

身長分布の年齢別分布(16歳女)

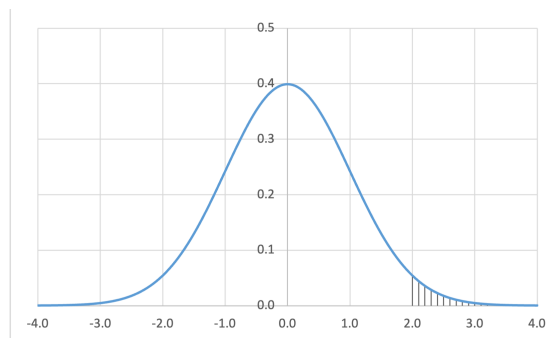


【練習6】 2022年の学校保健統計調査の結果から

の全国の16歳のデータの千分率集計表

(e-Statから得たデータを「練習6」に入力してClassroomで配信)を使って、もっと階級の細かい棒グラフを描くとどうなるか確認してみましょう。ただし、グラフタイトルや軸タイトルを適切なものに設定すること。

→練習6では身長のデータを利用したが、体重のデータなどでも階級を細かくしていったらグラフの先端をつなぐと、右図のような曲線が得られる。このような分布を（ ）といい、コインの裏表などの確率がこの分布になり、いろいろなものが似た分布を示す。

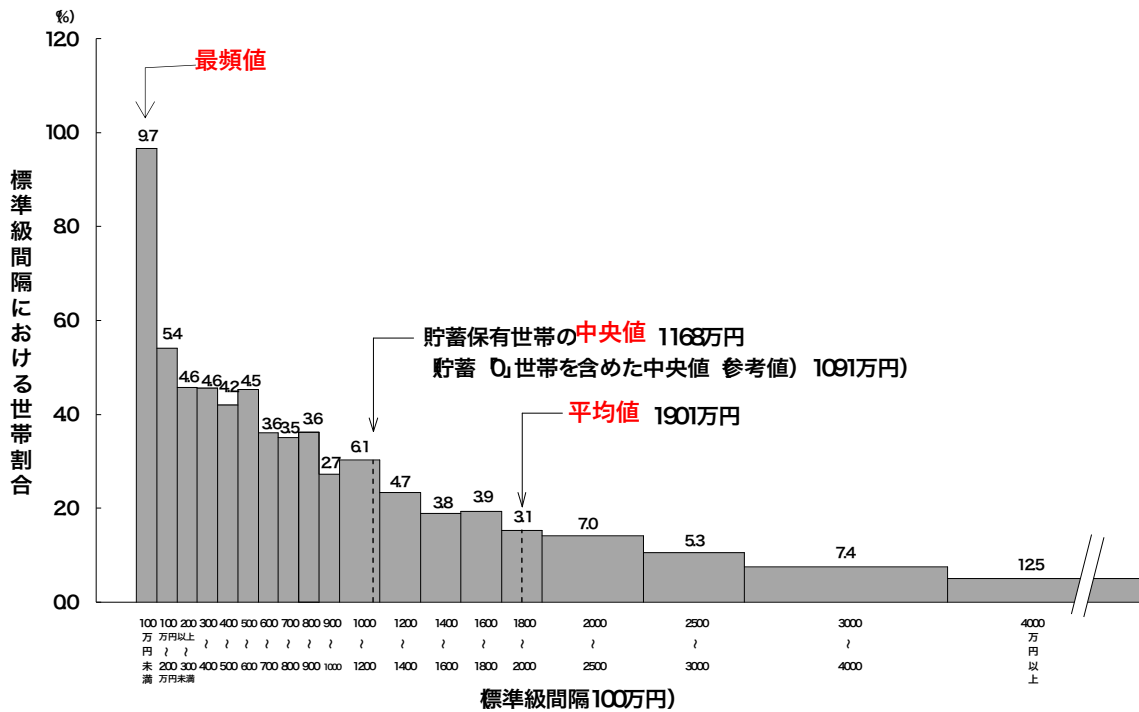


この分布だと、最も度数の大きい値＝**最頻値**あたりが（ ）になり、グラフもほぼ左右対照になるのが特徴である。

例2. 貯蓄現在高の分布

総務省統計局のデータには家計に関するものもある。2022年の家計調査の報告書では、次のように述べられています。これを見て、みなさんはどう思うでしょうか。

- 二人以上の世帯における2022年平均の1世帯当たり貯蓄現在高（貯蓄現在高が「0」の世帯を含めた平均値）は、**1901万円**で、前年に比べ21万円、1.1%の増加となり、4年連続の増加となるとともに、比較可能な2002年以降で最多となっている。
- 二人以上の世帯における2022年平均の1世帯当たり負債現在高（平均値）は、**576万円**で、前年に比べ9万円、1.6%の増加となっている。



- 注1) 貯蓄保有世帯の中央値とは、貯蓄「0」世帯を除いた世帯を貯蓄現在高の少ない方から順番に並べたときに、ちょうど中央に位置する世帯の貯蓄現在高をいう。
- 注2) 標準級間隔100万円 (貯蓄現在高1000万円未満) の各階級の度数は縦軸目盛りと一致するが、貯蓄現在高1000万円以上の各階級の度数は階級の間隔が標準級間隔よりも広いので、縦軸目盛りとは一致しない。

貯蓄現在高階級別世帯分布 (2人以上世帯) (2022年家計調査)

→明らかに偏った分布であり、平均値以下が3分の2以上を占めている。この場合、平均値よりも、データを順番に並べたときの真ん中の値を代表にした方がしっくりくる。このような統計量を () という。これは、他のデータと比べて極端に大きく異なるデータ、つまり () がある場合には平均値より代表として適していることがある。

例3. 往復での速度の違い

[考えてみよう6] 休みの日に家族で出掛けて、倉敷インターチェンジから神戸北インターチェンジまでを車で往復したとき、行きはスムーズに時速100kmで、帰りは混んでいたもので時速60kmで帰ってきました。行き帰り合わせて平均値は時速何kmになるでしょうか。

- () …対象が x と y の2つであれば、() で計算できる。往復の速度の平均値を求める計算に用いることができる。
- () …伸び率 (成長率) の平均を求めることができる。
(例) 1年の利益が 80%→20%→10%アップ

【データの代表】（その1）

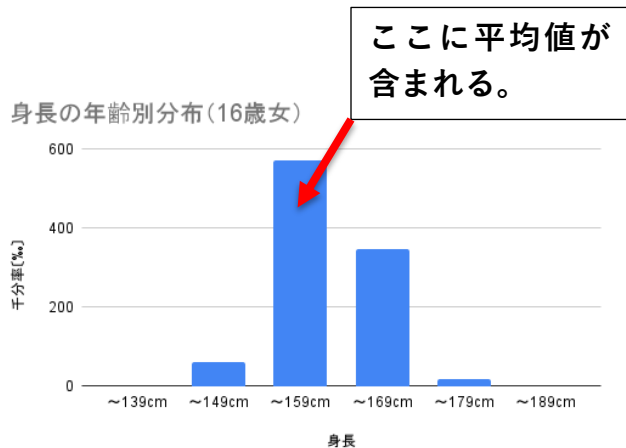
データを代表するような値といえば…

集めたデータを簡単に表したいという場合、どうすればいいでしょうか。「こんなデータが集まりました」ということを縮約して知らせる場合によく使われる統計量としては（ **平均値** ）が代表といえるでしょう。実際に私たちが最も一般的に使っているのは（ **算術平均** ）とも呼ばれるものになります。

例1. 身長分布

総務省統計局の統計データ（2021年度 学校保健統計調査）から、全国の高校2年生（16歳）女子の身長のデータを見てみましょう。

グラフにしたのが右図です。平均値はおよそ（ **158** ）cmです。

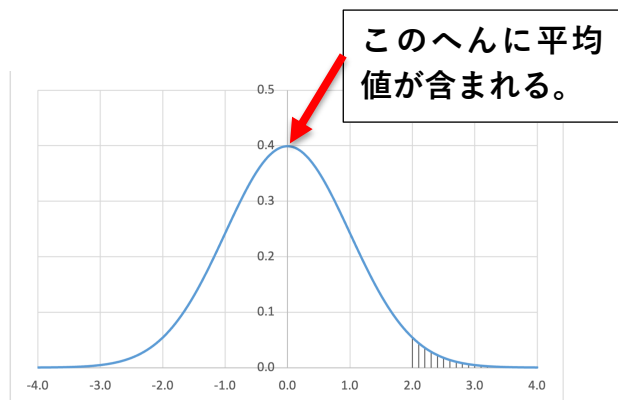


【練習6】2022年の学校保健統計調査の結果から

の全国の16歳のデータの千分率集計表

（e-Statから得たデータを「練習6」に入力してClassroomで配信）を使って、もっと階級の細かい棒グラフを描くとどうなるか確認してみましょう。ただし、グラフタイトルや軸タイトルを適切なものに設定すること。

→練習6では身長のデータを利用したが、体重のデータなどでも階級を細かくしていったらグラフの先端をつなぐと、右図の曲線と似た曲線になる。このような分布を（ **正規分布** ）といい、コインの裏表などの確率がこの分布になり、いろいろなものが似た分布を示す。

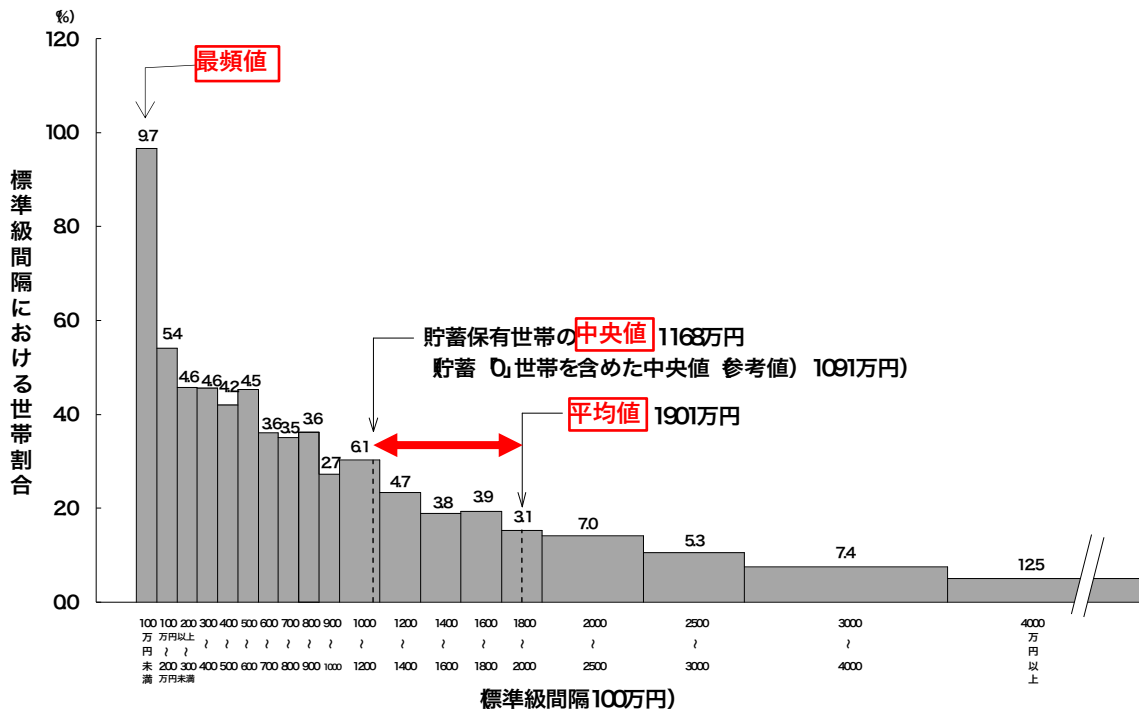


この分布では、最も度数の大きい値＝最頻値あたりが（ **平均値** ）になり、グラフもほぼ左右対照になるのが特徴である。

例2. 貯蓄現在高の分布

総務省統計局のデータには家計に関するものもある。2022年の家計調査の報告書では、次のように述べられています。これを見て、みなさんはどう思うでしょうか。

- 二人以上の世帯における2022年平均の1世帯当たり貯蓄現在高（貯蓄現在高が「0」の世帯を含めた平均値）は、**1901万円**で、前年に比べ21万円、1.1%の増加となり、4年連続の増加となるとともに、比較可能な2002年以降で最多となっている。
- 二人以上の世帯における2022年平均の1世帯当たり負債現在高（平均値）は、**576万円**で、前年に比べ9万円、1.6%の増加となっている。



- 注1) 貯蓄保有世帯の中央値とは、貯蓄「0」世帯を除いた世帯を貯蓄現在高の少ない方から順番に並べたときに、ちょうど中央に位置する世帯の貯蓄現在高をいう。
- 注2) 標準級間隔100万円（貯蓄現在高1000万円未満）の各階級の度数は縦軸目盛りと一致するが、貯蓄現在高1000万円以上の各階級の度数は階級の間隔が標準級間隔よりも広いので、縦軸目盛りとは一致しない。

貯蓄現在高階級別世帯分布（2人以上世帯）（2022年家計調査）

→明らかに偏った分布であり、平均値以下が3分の2以上を占めている。この場合、平均値よりも、データを順番に並べたときの真ん中の値を代表にした方がしっくりくる。このような統計量を（ **中央値** ）という。これは、他のデータと比べて極端に大きく異なるデータ、つまり（ **外れ値** ）がある場合には平均値より代表として適していることがある。

例3. 往復での速度の違い

【考えてみよう6】休みの日に家族で出掛けて、倉敷インターチェンジから神戸北インターチェンジまでを車で往復したとき、行きはスムーズに時速100kmで、帰りは混んでいたもので時速60kmで帰ってきました。行き帰り合わせて平均値は時速何kmになるでしょうか。（ちなみに倉敷-神戸北間は170kmですが不要です。）

実は距離を x とおくと、 $\frac{2x}{\frac{x}{100} + \frac{x}{60}} = \frac{2}{\frac{1}{100} + \frac{1}{60}} = 75$ と x は消えるのでわからなくても問題なく75km/hとわかる。 $(100+60) \div 2 = 80$ の算術平均だと違ってしまう。

- （ **調和平均** ）…対象が x と y の2つであれば、 $\left(\frac{2}{\frac{1}{x} + \frac{1}{y}} \right)$ で計算できる。往復の速度の平均値を求める計算に用いることができる。
- （ **幾何平均** ）…倍率の平均を求めることができる。 $\sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$
 (例) 1年の利益が80%up→20%up→10%up $\sqrt[3]{1.8 \times 1.2 \times 1.1} \div 1.33 \therefore 1$ 年平均33%up!