

## 資料の活用についての Q&amp;A

## 相対度数の表し方

**Q:** 相対度数は、分数で表してはいけないのでしょうか。また、各階級の相対度数を小数で表すと、それらの合計が1にならない場合もありますが、例えば0.99になった場合、度数分布表の相対度数の合計の欄には、0.99と示すのでしょうか。

**A:** 相対度数は、それをを用いる目的を考えると、小数で表す方が好都合です。

例えば、相対度数が分母の異なる分数で表されていた場合には、2つの階級でどちらの度数が大きいのかを瞬時に判断しにくかったり、その階級の度数の全体に対する割合もわかりにくかったりします。相対度数を求めること自体が目的ではなく、その資料を活用するために行うものですから、**活用しやすい形を考えて、相対度数も小数で求めるように指導するのがよい**と思います。

また、四捨五入によって各階級の相対度数を求めた場合に、その合計が1にならない場合もあります。そのときにも、全体を1として考えていることを示す意味で、相対度数の合計の欄には1.00のように書きます。

このとき、合計がちょうど1.00となるように各階級の相対度数を調整する必要はありませんが、円グラフに表すなど、特にそれが必要な場合には、相対度数のもっとも大きい階級で調整することが多いようです。

## 度数分布表から求める平均値の実際とのずれ

**Q:** 度数分布表から平均値を求めるとき、資料の個々の値が離散的な場合には実際の平均値とのずれが大きくなってしまおうに思うのですが、それでよいのでしょうか。

**A:** テストの点など、整数値しかとらない資料を、階級の幅を10として、0以上10未満、10以上20未満、…のように整理したとします。このとき、例えば、70以上80未満の階級に入るのは、70, 71, 72, …, 79の10個で、これらの中央の値は74.5となり、階級値の75を用いて求めた平均値は、実際の平均値とずれが大きくなってしまおうことを懸念されているのだと思います。実は、このようなことは資料の個々の値が離散的だから起こるわけではありません。

例えば、連続的な測定値を小数第1位までの値にして整理した場合でも、やはりずれがあります。つまり、資料の値が離散的か連続的かではなく、このようなずれが資料の傾向を読み取るのに影響するものであるかどうかで、場合によっては整理のしかたを見直さなければならない問題です。

上の、整数値しかとらない資料の例でいえば、階級のとり方を0から9まで、10から19まで、…のようにして、階級値を4.5, 14.5, …とする整理のしかたもあります。そうしなければならないかどうかは、資料の性質や調査の目的などに応じて考えることになります。

## 資料の活用についての Q&amp;A

## 度数分布表と代表値

**Q**：代表値は、資料の個々の値がわかっている場合と、度数分布表に整理された場合とで、扱いが異なります。その理由を教えてください。

**A**：代表値を求める際に、資料の個々の値が全てわかっている状況があれば、度数分布表に整理された資料しか手に入らない状況もあります。それぞれの状況で、代表値の求め方が変わります。

まず、**平均値**は、資料の個々の全ての値がわかっている場合には、それらすべての合計を資料の個数でわって計算します。一方で、度数分布表に整理された資料しか手に入らない場合には、「**各階級に入っている資料の値は、どの値もすべて、その階級の階級値である**」と考えて計算します。このようにして求めても、各階級で、階級値より大きい値、小さい値の両方があればそれらがならされ、資料の個々の値から求めた平均値と大きく変わらない値となります。

次に、**最頻値**ですが、資料の個々の値が全てわかっている場合には、その中でもっとも多く現れる値を最頻値とします。ただし、ほとんど同じ値がない資料などでは、たまたま複数現れたある値が、資料の最頻傾向を示していることにならない場合があります。そのため、度数分布表に整理し、どの階級のあたりが最頻傾向を示しているかを調べ、その階級値を最頻値として用いる方が適切な場合があります。つまり、最頻値の場合には、平均値と違い、手に入った資料のすべての値がわかっているかどうかではなく、その資料がよくとる値の傾向を知るのに**度数分布表に整理すべきかどうかを判断することも必要**になります。

最後に、**中央値**は、資料の個々の値が全てわかっている場合には、それらすべてを大きさの順に並べた際の中央の値です。資料の個数が偶数の場合には、中央に並ぶ2つの値の平均をとります。度数分布表に整理された資料しか手に入らない場合には、中央値がどの階級に入っているかまではわかりませんが、**その階級の階級値を中央値とすることは、ふつうはしません**。これは、中央値のもつ意味が、資料においてその値より大きい値をとることも小さい値をとることも、50%の確率で起こる境の値であるのに、階級値がその値である保証がなく、判断に狂いが生じるおそれがあるためだと考えられます。

