

データの平均性と中心性

九州大学 鈴木讓

1 目的

この報告では、現在流布している代表値の説明においては、平均性と中心性という2つの性質が混同されていることをまず指摘し、より正確な説明を提唱する。周知の通り、相加平均は平均値として、標準偏差は散布度として、記述統計においては最もよく使われる指標である。一般に標準偏差は、相加平均の周りの散布度を表す指標として説明される。しかしながら、現在の説明の論理では平均性と中心性という2つの性質が混同されている。平均値は平均性に関する指標、散布度は中心性に関する指標であり、この2つの指標は本来別個に論じられるべきである。

2 方法

現在の説明では、平均性と中心性の2つの性質が混同されているために、次の3つの根本的な疑問に答えることができない。第一は、標準偏差が平均偏差よりも散布度の指標としてなぜすぐれているのか。第二の疑問は、仮に平均偏差を使わないとしても、なぜ3次以上の積率（モーメント）を散布度の指標として用いないのか。第三の疑問は、なぜ相乗平均（幾何平均）の周りの散布度を用いないのか。本報告では、現在の説明の論理では、この3つの疑問に答えることができないことを、初めに示す。次に、平均性と中心性を明確に区別して定義し、その上でこの3つの疑問に対する回答を示す。

3 結果

平均値を説明するために不均整指数、散布度を説明するために散乱指数という指数を定義し、これらの指数が最小になる場合に、平均性と中心性が最大になると考える。この定義にもとづけば、減算を基礎演算としてデータ間の変位を測る場合、平均性が最大になる点は相加平均である。また、除算を基礎演算とする場合、平均性が最大になる点は相乗平均である。

次に、偏差平方値の平均により散乱の程度を測る場合、中心性が最大になる点は相加平均に一致し、この時の散乱指数の値は分散となる。また、偏差絶対値の平均により散乱の程度を測る場合の中心点は、一般には一意に定まらないが、一意に定まる場合には相加平均ではなく中央値となる。平均偏差とは、相加平均に対するこの散乱指数の値であるが、相加平均がこの散乱指数の中心点とはならない以上、散布度の指標としては適切ではないことが分かる。

また、3次以上の積率を散乱指数とした場合も、その中心点は相加平均にはならない。言い換えれば、この方式では中心としての相加平均の周りのデータのばらつきを測ることにはならない。

そして、除算を基礎演算としてデータ間の変位を測る場合、積率と同様の方法で散乱指数を定義すると、最小値をとる値は相乗平均には一致しない。従って、この方式では相乗平均は中心点にはならず、中心としての相乗平均の周りのデータのばらつきを測ることにはならない。

4 結論

平均性と中心性をともに備えている代表値は相加平均のみであり、これが相加平均と分散（標準偏差）が指標として最もよく使われる論理的な理由だと分かる。

本報告では離散分布のみを扱う。積率は連続分布においても定義されるが、平均偏差と相乗平均は連続分布においては定義されない。また、連続分布においては平均値、分散が存在するとは限らない。従って、平均性と中心性に関しての包括的な議論のためには、連続分布よりもむしろ離散分布の方が適していると考えられる。