

標準偏差 (Standard Deviation)

1999年12月4日

2005.3.20 改定

作成：橋 永久

経済データの観測値のバラツキぐあいを計る物差しとして、標準偏差という指標がよく用いられます。GDP や、その中の消費・投資系列のバラツキぐあいという、観測期間中に各データ系列がどれだけ変動したかを見ることになります。通常は、GDP そのものの変動に比べると、消費の変動は小さく、投資の変動は大きくなっています。

標準偏差と聞いて戸惑われる方がいるかもしれませんが、分散の平方根を取ったものにすぎません。まず分散の復習から始めましょう。ある経済データを8年間観測して、以下の観測値系列(系列(1))を得たとします。

$$(2, -1, 0, -1, 1, -2, 1, 2) \quad (1)$$

系列(1)の平均値は、0.25になります。分散は、観測値が全体としてどれだけ平均値から離れているかを示します。具体的には、各観測値と平均値の差を求め、その差の2乗の平均値が分散です。まず各観測値と平均値の差を取ると、

$$(1.75, -1.25, -0.25, -1.25, 0.75, -2.25, 0.75, 1.75)$$

となります。2乗して全て足し合わせると、15.5です。さらにこれを8で割って平均を求めると、1.9375となります。これが系列(1)の分散です。

平均からの乖離でバラツキをはかるというのは分かりやすいですが、ではなぜその2乗をとるのでしょうか？ 2乗を取る事により、平均から遠く離れた値がより強調されることとなります。例えば系列(1)の4番目と5番目の観測値をよりバラツカせた系列(2)と、

$$(2, -1, 0, -2, 2, -2, 1, 2) \quad (2)$$

系列(1)の1番目と6番目の観測値をよりバラツカせた系列(3)

$$(3, -1, 0, -1, 1, -3, 1, 2) \quad (3)$$

の分散を比べてみましょう。系列(2),(3)共に、系列(1)の観測値を2個、1ずつ変化させたにすぎませんが、系列(3)の方が平均から遠い値を動かしています。皆さん、表計算ソフト等で系列(2)、(3)の分散を計算してみてください。系列(3)の分

散が、系列(2)のそれより大きくなっているはずですが、同じように値が変化しても、分散は平均から遠い値の変化を強調していることが分かります。

つぎに、なぜ分散の平方根を取って標準偏差を計算するのか考えてみます。系列(1)の観測値が、私の体重の対前年増減値だとします。すると単位は kg になります。分散の単位は、2乗を取っているので kg の2乗となり、値の意味が良く分かりません。そこで平方根を取って標準偏差にしてやると、バラツキの指標の単位がデータの単位と同じ kg に戻り、意味が分かりやすくなります。これが、分散より標準偏差を使う事が多い理由です。ちなみに系列(1)の標準偏差は、1.392 (小数第3位で四捨五入)になります。

次に、以下の系列(4)を考えてみます。

$$(12, 9, 10, 9, 11, 8, 11, 12) \quad (4)$$

系列(1)をゼロからの変動と見ていただくと、系列(4)はそれを10からの変動に置き換えたものです。よって系列(4)の分散は系列(1)の分散と同じ1.9375となります。ここから、系列(1)と系列(4)のバラツキ具合は同じ、と判断して良いのでしょうか？ いささか問題がありそうです。例えば10から11への変動は10%の増加に過ぎませんが、1から2への変動は100%の増加になります。同じように1つつ動いていても、系列(1)の変動の方が率としては大きいのではないのでしょうか？ こうした問題に対応するには、平均に対する相対的なバラツキ度を示す変動係数を用います。変動係数は、標準偏差を平均値で割ったものです。系列(1)の変動係数は5.568、一方系列(4)のそれは0.136となります。

統計の入門書は統計学の講義で数多く紹介されますが、私からは

加納悟、浅子和美 「経済学のための統計学：第2版」東京：有斐閣 (1998年)。

東京大学教養学部統計学教室編 「人文・社会科学の統計学」 東京：東京大学出版会 (1994年)。

を奨めておきます。