

「分散」と「標準偏差」について

寺林 暁良

今回は、統計値のバラつきを示す指標である「分散」と「標準偏差」を紹介する。これらは、統計値の個々のデータが分からない場合でも、データの大まかな分布状況を把握できる重要な指標である。

「分散」と「標準偏差」

統計値の散らばり具合を示す「分散」(シグマ²)は、各データから平均値を引いて個々の統計値と平均値の距離(偏差)を求め、それらの値の2乗の総和をデータ数で割ることで、データと平均値の距離を求めたものである。計算式では、

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$
 とあらわされる。

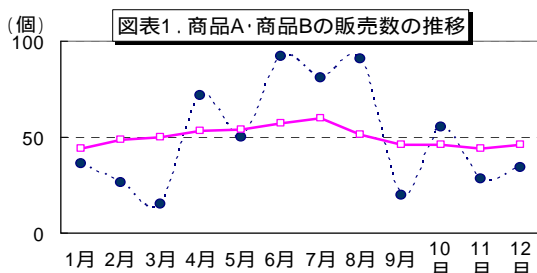
分散の値が大きいほど、平均値からの各データの散らばりが大きいということになる。

また、「標準偏差」()は、分散に平方根()を付けたもので、分散と同様にデータの散らばりを示す。標準偏差は、分散の2乗分を取り除くため、データと同じ水準の数値となる。そのため平均値等との対比がしやすいという利点がある。

商品Aと商品Bの販売数を事例に

「分散」や「標準偏差」が重要になる事例を、一つ挙げよう。

商品AとBが1年間販売されたとする(図表1)。両者は、ともに1年間で600



個を売り上げたため、12ヵ月間の月ごとの平均販売数は、どちらも50個(600個÷12ヵ月)である。しかし、商品AとBは、各月の販売数の増減(=散らばり具合)が異なっている。商品Aの販売数は、月によって販売数の差が大きいが、商品Bの販売数は、販売開始から終了まで平均の50個近くに集中している。

こうした特徴を統計的に示すのが、分散と標準偏差である。商品AとBの分散(標準偏差)を計算すると、商品Aでは分散²=779.5(標準偏差=27.9)と、商品Bの²=26.9(=5.2)と比べて値が大きい(図表2)。これにより、「商品AのほうがBよりも月ごとの販売数のバラつきが大きい」ことが統計的に示される。

図表2. 商品A・商品Bの分散と標準偏差

	月数	平均	分散	標準偏差
商品A	12	50	779.5	27.9
商品B	12	50	26.9	5.2

標準偏差の性質

データが図表3のように正規分布する場合、標準偏差は各データが含まれる割合(確率)を示す指標であり、標準正規分布表により平均±標準偏差の範囲に68.26%、平均±2の範囲に95.44%のデータが含まれることが分かる。

先述の商品Bの例では、平均=50、=5.2なので、44.8~55.2(50±5.2)の間に68.26%、39.6~60.4(50±2×5.2)の間に95.44%のデータが含まれる。

図表3. 標準偏差と分布確率

