

Lesson3 > データのバラツキ

このレッスンではデータの「バラツキ」について学習をします。バラツキを表す指標として、「分散」や「標準偏差」といったものがあり、これらの指標もエクセルを使えば簡単に求めることができます。

まずは「分散」と「標準偏差」の考え方を理解しましょう。

■ 分散の求め方

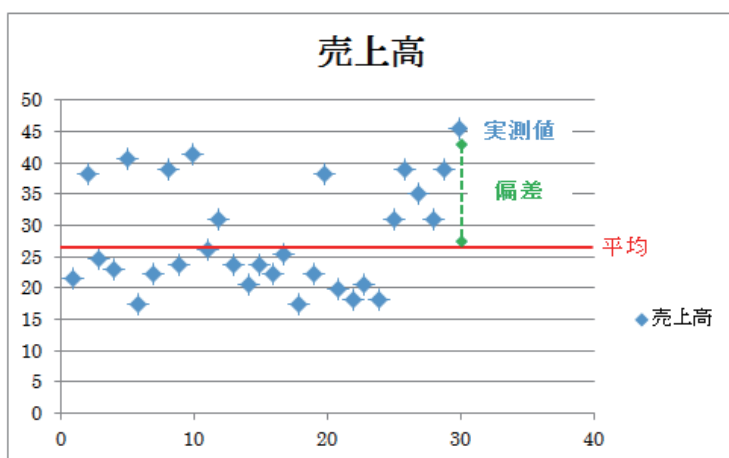
まずは分散を求める公式を確認しましょう。

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

V : 分散、n : データの数、 x_i = 各データの値、 \bar{x} = データの平均

数学のレッスンではありませんので、公式を覚える必要はありません。計算はエクセルに任せてしまいましたが、公式の意味は理解しておきましょう。

$(x_i - \bar{x})^2$ は各データの値から平均値を差し引いた値の2乗を表しています。 Σ は合計を表しますから、この式は各データの値から平均値を引いた値の2乗を合計して、n分の1かけたもの、つまりデータの個数で割るという計算式です。



この散布図は、左の表にある各商品の売上高をプロットしたものです。赤い線はデータの平均値を示しています。実測値と平均の差を「偏差」といいます。数式で $(x_i - \bar{x})^2$ の部分です。分散は偏差を2乗した数の合計をデータの個数で割り算しているので、偏差の2乗の平均値と言えます。偏差というのは、平均から見てプラスの値とマイナスの値で構成されています。これを合計していくと、ゼロになってしまい、バラツキを示すことができなくなります。そこで偏差を2乗することで、全ての偏差をプラスの値にして、バラツキを表現しているのです。これが分散という指標の考え方です。

Point

▷プロット

データをグラフなどの図表上に表すことです。

$$S = \sqrt{V}$$

分散を平方根 ($\sqrt{\quad}$) にしたのが標準偏差です。 $\sqrt{\quad}$ は 0.5 乗と同じです。分散を求める過程で、各データの値と平均値の差、つまり偏差を 2 乗して合計をしました。それにより、分散はバラツキを示す指標としては、元のデータより大きくなってしまいます。2 乗したデータに 0.5 乗すれば、元の単位に戻すことができます。0.5 乗することでデータの平均値と単位を揃えたものが、標準偏差という指標です。

●実習 分散と標準偏差の算出(1)

(ブック名 : Chapter02 / 使用シート : 分散と標準偏差 02)

分散と標準偏差を求めましょう

1. セル H4 にデータの平均値を求めます。
 - ① [オート SUM] の ▼ をクリックし、「平均」を選択します。
 - ② セル B5 から B34 を選択し、Enter キーを押します。
2. C 列に偏差を求めます。
 - ① セル C5 を選択して、データと平均値の差を求めます。「=B5-H4」と入力します。数式はコピーしますので、平均値のセル H4 が固定されるように絶対参照を設定しておきましょう。

商品ID	売上高	偏差 (平均との差)	偏差の2乗
A01	21.5	-6.19	
A02	37.6		

3. セル D5 に偏差を 2 乗する数式を入力します。「=C5^2」と入力します。

商品ID	売上高	偏差 (平均との差)	偏差の2乗
A01	21.5	-6.19	38.3161
A02	37.6		

4. セル C5 と D5 の数式を 34 行目までコピーします。