

いろいろな平均値

2019年(作成)
2025年1月(改定)
倉谷 隆博

1-1. 平均と期待値

平均(値)と期待値の違い、および平均と期待値の関係を整理する。

母集団(別途資料「いろいろな確率分布」参照)から、確率変数 $X = x_i$ ($i = 1, \dots, n$) という値が観測される回数を N_i とする(標本数 $N = \sum_{i=1}^n N_i$)。また、 $X = x_i$ という値が観測される確率を p_i ($i = 1, \dots, n$)、 $\sum_{i=1}^n p_i = 1$ とする。

(1) 平均(Average)

$$\bar{x} = \sum_{i=1}^n x_i N_i / N \quad \dots \text{(式1-1)}$$

平均は、既に発生した事象を対象とする。つまり、観測データは存在する。

(2) 期待値(Expected Value)

$$\mu = \sum_{i=1}^n x_i p_i \quad \dots \text{(式1-2)}$$

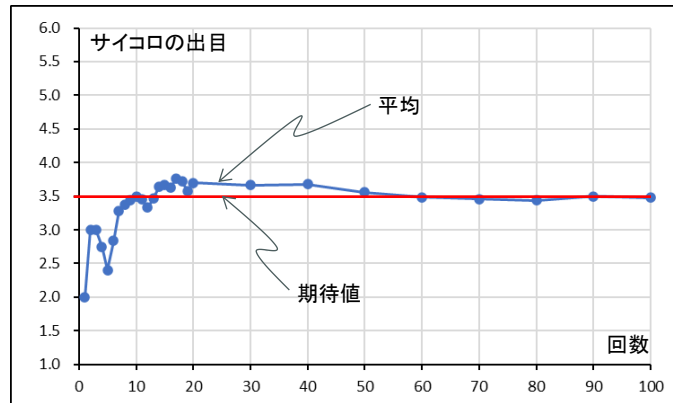
期待値は、今後発生するであろう事象を対象とする。つまり、観測データは存在しない。

(3) 平均と期待値との関係

標本数 N を無限大にすれば、 $p_i = N_i / N$ が成立するが、「母集団からの無作為抽出の試行を無限に繰り返せば、母集団がどのような確率分布であっても標本平均は母平均に近づく」という大数の法則(別途資料「いろいろな確率分布」を参照)から、標本数 N が無限大のとき、つまり、標本が母集団と一致するとき、 $\mu = \bar{x} \quad \dots \text{(式1-3)}$ が成り立つ(期待値と平均が一致する)ことが分かる。

(例)

(図1-1)は、サイコロの出目の平均と期待値を表したものである。
試行回数を増やしていくと、平均が期待値に近づいていくことが分かる。



(図1-1)サイコロの出目の平均と期待値

・サイコロの出目の期待値

サイコロの目(1, 2, 3, 4, 5, 6)の出目の確率は等しく、 $1/6$ であるから、サイコロの出目の期待値

$$= (1 \times 1/6) + (2 \times 1/6) + (3 \times 1/6) + (4 \times 1/6) + (5 \times 1/6) + (6 \times 1/6) = 3.5$$

・サイコロの出目の平均

ここでは、Excel 関数 `RANDBETWEEN(1, 6)` を使用して、サイコロを疑似的に100回振ったデータを作り、振った回数に対応するサイコロの出目の平均を求めた。

2. 平均を求めるときの注意点

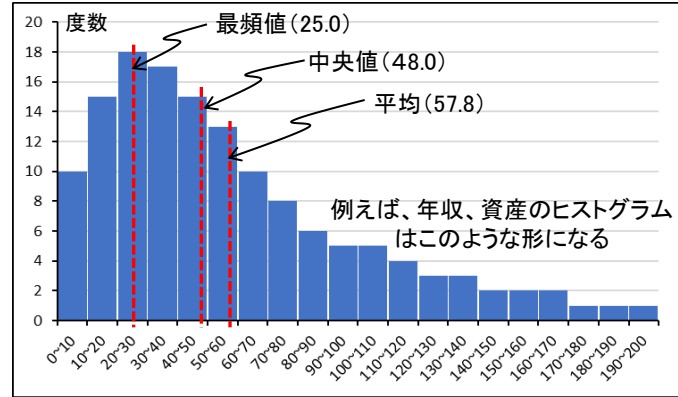
平均(Average, Mean)は、データ分布の中心(重心)を表す重要な基本統計量であるからこそ、平均を計算するにあたっては様々な視点でデータを見ておく必要がある。

- 1) データの数が少ないか。
データの数が少な過ぎてデータ分布のばらつき具合を把握できなければ、平均の計算をすべきではない。
- 2) データの分布に偏りがいないか。
データのばらつき具合の偏りを考慮しなくてはならない。つまり、(図2-1)のようなデータ分布の平均であれば、平均が分布の特徴(分布の中心)を表していると言えるが、(図2-2)のようなデータ分布の平均は分布の特徴(分布の中心)を表しているとは言い難い。
- 3) 外れ値を含んだデータでないか。
データ分布の塊から外れた外れ値(異常値)がデータに混在していると、平均の計算に悪影響を与えてしまう。ただ、外れ値が重要な情報を持っている場合があるので、平均の計算にあたって外れ値を取り除く場合には、十分な確認が必要である。

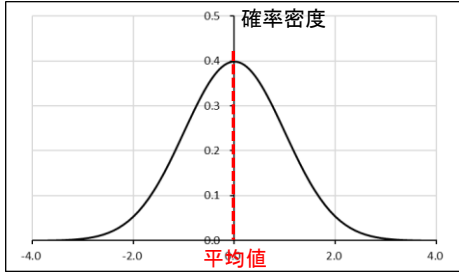
データ分布の中心を表す統計量としては、平均のほかに、中央値と最頻値がある。

- 1) 中央値(Median)
片側によっているようなデータ分布のときは、平均に加えて中央値を確認しておく必要がある。
離散型のデータ: 中央値はデータを値の大きさの順に並べたときの真ん中の値になる。
連続型のデータ: 中央値はデータ分、あるいは確率分布の左右の面積を2等分する値である。
- 2) 最頻値(Mode)
最頻値は外れ値の影響を受けない統計量である。
離散型のデータ: 最頻のデータ(ヒストグラムのピーク)になるところの値である。
連続型のデータ: データ分布、あるいは確率分布の最大になるところの値である。

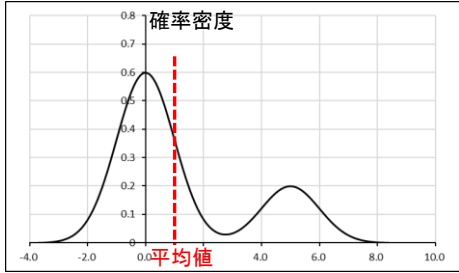
離散型のデータ(図2-3)と連続型のデータ(図2-4)の平均(値)、中央値、最頻値を示す。分布の右裾の広がり方が平均と期待値を右側に寄せている。また、(図2-1)は正規分布であるが、この場合は平均(値)、中央値、最頻値の全てが等しくなる(なお、正規分布とカイニ乗分布は、別途資料「いろいろな確率分布」参照)。



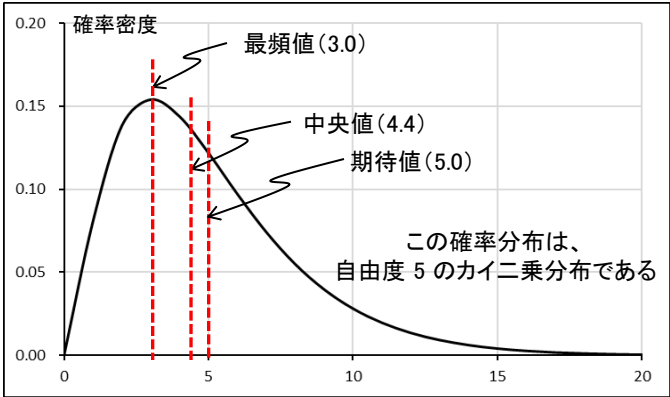
(図2-3) 離散型データ(ヒストグラム)



(図2-1) 正規分布



(図2-2) 平均の意味がない分布
(図中の平均値の場所はイメージ)



(図2-4) 連続型データ(確率分布)

3. 算術平均(相加平均)

(1) 用途

算術平均(Arithmetic Mean)は、測定データが正規分布(別途資料「正規分布の概要」参照)のような分布に従うと判断できるときに使用される平均である。最もよく使用される平均であり、単に平均というときには、算術平均を指している。

製品の成分量などの測定値、身長・体重などの測定値、日次売上高などは正規分布のような分布に従う場合が多く、算術平均が使用される。

ただし、容易に予想できることであるが、例えば身長・体重は性別、年齢別、国別で異なるはずである。また、日次売上高は曜日、天候、地域などで大きく異なることがある。平均の算出にあたっては対象データの分布に注意を払う必要がある。

(2) 計算式

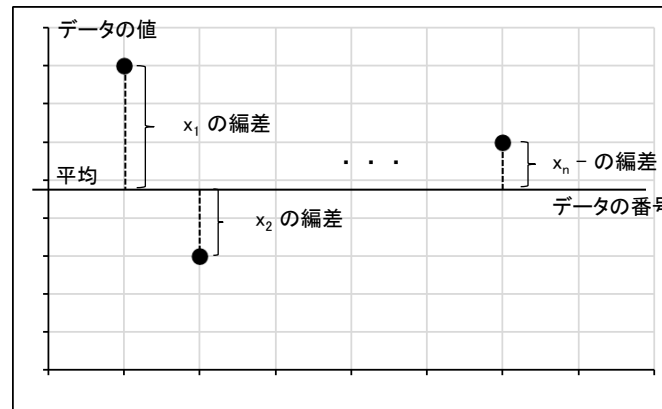
算術平均は、 n 個の対象データ x_i ($i = 1, 2, \dots, n$) の総和をとって、その総和をデータ総数 n で割って求める。

算術平均 $= \sum_{i=1}^n x_i / n \quad \dots$ (式3-1) Excel 関数では、算術平均 $= \text{AVERAGE}(x_1, x_2, \dots, x_n)$

なお、データごとの偏差 = データのごとの値 - 全てのデータの平均 で表されるが、

(式3-2)で示す様に、この偏差の和は 0 になる。(式3-2)の計算を(図3-1)で示す。

$$\text{偏差の和} = \sum_{i=1}^n (x_i - \text{算術平均}) = \sum_{i=1}^n (x_i - \sum_{i=1}^n x_i / n) = \sum_{i=1}^n x_i - \sum_{i=1}^n \sum_{i=1}^n x_i / n = \sum_{i=1}^n x_i - n \sum_{i=1}^n x_i / n = \sum_{i=1}^n x_i - \sum_{i=1}^n x_i = 0 \quad \dots$$
 (式3-2)



(図3-1) 算術平均の偏差

3) 計算例

$n = 10$, $x_1 = 8$, $x_2 = 4$, $x_3 = 5$, $x_4 = 3$, $x_5 = 5$, $x_6 = 6$, $x_7 = 9$, $x_8 = 4$, $x_9 = 6$, $x_{10} = 8$ のとき、

$$\text{算術平均} = \sum_{i=1}^{10} x_i / 10 = (8 + 4 + 5 + 3 + 5 + 6 + 9 + 4 + 6 + 8) / 10 = 58 / 10 = 5.8$$

4. 幾何平均(相乗平均)

(1) 用途

幾何平均(Geometric Mean)は経時(経年)の伸び率の平均値を求めるときに使用する。

例えば、

- ・5年間の年平均成長率 CAGR (Compound Average Growth rate)
- ・7年間の年平均物価上昇率

などを求めるときに使用する。

(2) 計算式

A : 初期値、 x_i : i 年 ($i = 1, \dots, n$) ごとの年間変化率、 x : n 年にわたる平均年間変化率(幾何平均)とすると、(式4-1)が成り立つ。

$$A x_1 x_2 \cdots x_n = A \underbrace{x x \cdots x}_{n \text{ 個の積}} \rightarrow \text{総積を表す記号 } \Pi \text{ を使用して、} A \prod_{i=1}^n x_i = A x^n \quad \cdots \text{ (式4-1)}$$

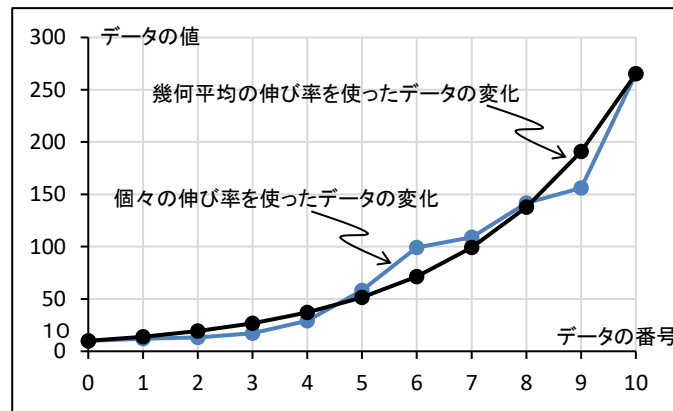
(式4-1)より、幾何平均 $x = (\prod_{i=1}^n x_i)^{1/n} \quad \cdots \text{ (式4-2)}$ Excel 関数では、幾何平均 = GEOMEAN($x_1, x_2, \dots, x_n$)

(3) 計算例

$n = 10$, $x_1 = 1.2$, $x_2 = 1.1$, $x_3 = 1.3$, $x_4 = 1.7$, $x_5 = 2.0$, $x_6 = 1.7$, $x_7 = 1.1$, $x_8 = 1.3$, $x_9 = 1.1$, $x_{10} = 1.7$ のとき、

幾何平均 = $(1.2 \times 1.1 \times 1.3 \times 1.7 \times 2.0 \times 1.7 \times 1.1 \times 1.3 \times 1.1 \times 1.7)^{1/10} = 1.39$

結果を、(図4-1)に示す。



(図4-1) 幾何平均

なお、算術平均と幾何平均の図形を使った比較を(付録1)に示す。

5. 調和平均

(1) 用途

調和平均(Harmonic Mean)は、同じ仕事を繰り返すときの労働生産性の平均を求めたり、同じ距離を走行するときの速度の平均を求めたりするときに使用できる。
また、電気回路で並列配置された複数個の抵抗の全体抵抗値を求める場合にも、この調和平均の考え方が適用できる。

(2) 計算式

生産性の逆数は所要時間にあたるので、個々の生産性の逆数の総和は所要時間の総和を表す。これは平均の生産性を用いて算出した所要時間の総和に等しい。同様に速度の逆数は走行時間にあたるので、個々の速度の逆数の総和は走行時間の総和を表す。これは平均速度を用いて算出した走行時間の総和に等しい。これらの平均を調和平均という。
調和平均は、 n 個の対象データ x_i ($i = 1, 2, \dots, n$) の逆数の総和をデータ総数 n で割り、その逆数になる。

$$\sum_{i=1}^n 1 / \text{調和平均} = \sum_{i=1}^n 1 / x_i \quad \rightarrow \quad n / \text{調和平均} = \sum_{i=1}^n 1 / x_i \quad \rightarrow \quad 1 / \text{調和平均} = (\sum_{i=1}^n 1 / x_i) / n$$

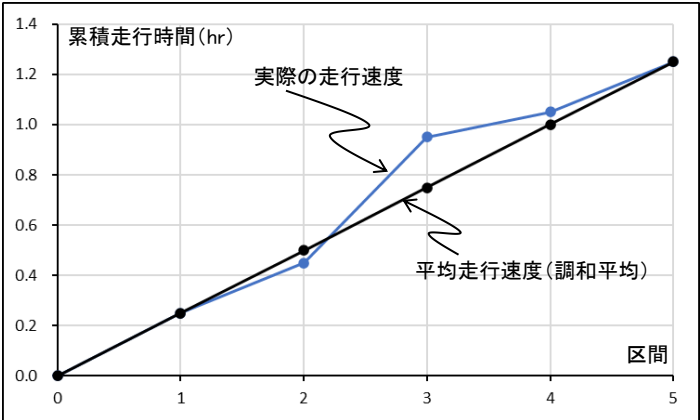
従って、 $\text{調和平均} = n / (\sum_{i=1}^n 1 / x_i)$ $\dots$ (式5-1) EXCELの関数では、 $\text{調和平均} = \text{HARMEAN}(x_1, x_2, \dots, x_n)$

(3) 計算例

(表5-1)は、総距離 100km を 20 km ずつ5 区間に分け、第 1 区間を速度 80 km/hr で、第 2 区間を速度 100 km/hr $\dots$ で走行したことを示す。
総区間の平均速度を求めるには、まず各区間ごとの走行時間(第 1 区間 0.25 hr, 第 2 区間 0.20 hr, $\dots$)を計算し、総区間での総走行時間 1.25 hr を求める。ついで、総距離 100 km を総走行時間 1.25 hr で割って、平均走行速度 = $100 / 1.25 = 80$ km/hr を求める。そして、この平均走行速度は、 $\text{調和平均} = 5 / (1/80 + 1/100 + 1/40 + 1/200 + 1/100) = 80$ で求めることができる。
なお、言うまでもなく、各区間の速度の算術平均値 = $(80 + 100 + 40 + 200 + 100) / 5 = 104$ km/hr と求めるのは間違いである。
(表5-1)をグラフで表したのが(図5-1)である。

(表5-1) 調和平均の計算例

	第1 区間	第2 区間	第3 区間	第4 区間	第5 区間	総区間	
距離(km)	20	20	20	20	20	総距離(km)	100
速度 (km/hr)	80	100	40	200	100	調和平均 平均走行速度(km/hr)	80
走行時間 (hr)	0.25	0.20	0.5	0.10	0.20	総走行時間(hr)	1.25



(図5-1) 調和平均

6. 相加平均、相乗平均、調和平均の大小関係

(1) 相加平均(算術平均) $\geq$ 相乗平均(幾何平均) $\geq$ 調和平均 の関係があることを確認する。

$$\text{相加平均: } A = \sum_{i=1}^n x_i / n \quad \cdots \text{ (式3-1)}$$

$$\text{相乗平均: } G = \left(\prod_{i=1}^n x_i \right)^{1/n} \quad \cdots \text{ (式4-2)}$$

$$\text{調和平均: } H = n / \left(\sum_{i=1}^n 1 / x_i \right) \quad \cdots \text{ (式5-1)}$$

1) 相加平均と相乗平均の大小関係

$\exp(x)$ をマクローリン展開すると(別途資料「いろいろな確率分布」参照)、

$$\exp(x) = 1 + x + x^2 / 2 + x^3 / 6 + \cdots \geq 1 + x \quad \cdots \text{ (式6-1)}$$

一方、 $x = x_i / A - 1$ ($i = 1, 2, \cdots, n$) とおけば、

(式6-1)の不等号の左辺は $\exp(x) = \exp(x_i / A - 1)$ であり、右辺は $1 + x = 1 + x_i / A - 1 = x_i / A$ である。

$$\text{ここで、} \exp(x_1 / A - 1) \exp(x_2 / A - 1) \cdots \exp(x_n / A - 1) = \exp\left(\sum_{i=1}^n x_i / A - n\right) \quad \cdots \text{ (式6-2)}$$

$$(x_1 / A) (x_2 / A) \cdots (x_n / A) = \left(\prod_{i=1}^n x_i \right) / A^n \quad \cdots \text{ (式6-3)}$$

であるから、 $\exp\left(\sum_{i=1}^n x_i / A - n\right) \geq \left(\prod_{i=1}^n x_i \right) / A^n \quad \cdots \text{ (式6-4)}$ が成立する。

また、(式3-1)より、 $\sum_{i=1}^n x_i / A - n = \sum_{i=1}^n x_i / \left(\sum_{i=1}^n x_i / n \right) - n = n - n = 0$ であるから、

(式6-2)は、 $\exp\left(\sum_{i=1}^n x_i / A - n\right) = \exp(0) = 1$ となる。

よって、(式6-4)は、 $1 \geq \left(\prod_{i=1}^n x_i \right) / A^n$ となり、 $A^n \geq \prod_{i=1}^n x_i$ が成立する。 $A \geq \left(\prod_{i=1}^n x_i \right)^{1/n} = G$ が成立する。従って、

$$A \geq G \quad \cdots \text{ (式6-5)}$$

2) 相乗平均と調和平均の大小関係

$1 / x_i$ ($i = 1, 2, \cdots, n$) の相加平均 $\sum_{i=1}^n (1 / x_i) / n$ と相乗平均 $\left(\prod_{i=1}^n 1 / x_i \right)^{1/n}$ を考える。これらを(式6-5)に代入すると、

$$\sum_{i=1}^n (1 / x_i) / n \geq \left(\prod_{i=1}^n 1 / x_i \right)^{1/n} \quad \cdots \text{ (式6-6)}$$

(式6-6)の逆数をとると、 $n / \sum_{i=1}^n (1 / x_i) \leq \left(\prod_{i=1}^n x_i \right)^{1/n}$

この式は、(式5-1)、(式4-2)より、 $H \leq G \quad \cdots \text{ (式6-6)}$ の関係を表している。

従って、(式6-5)と(式6-6)より、 $A \geq G \geq H$ の関係が成り立つ。

(2) 2変数の場合には、(相乗平均) 2 = 相加平均 $\times$ 調和平均 の関係がある。

2変数では、相加平均 $A = (x_1 + x_2) / 2$ 、相乗平均 $G = (x_1 x_2)^{1/2}$ 、調和平均 $H = 2 / (1/x_1 + 1/x_2)$

従って、 $G^2 = x_1 x_2$ 、 $AH = ((x_1 + x_2) / 2) (2 / (1/x_1 + 1/x_2)) = x_1 x_2$ 従って、 $G^2 = AH \quad \cdots \text{ (式6-7)}$

7. 加重平均

(1) 用途

加重平均 (Weighted Average) は、個々のデータに重みを付けた算出平均である。
例えば、販売単価の異なる商品の1日あたりの売上高から平均販売単価を求める場合などに使用される。この場合、総売上高を総販売個数で割って平均販売単価とするのが合理的である。これは個々の商品の販売数量を販売における貢献度と見なして、その貢献度 (重み) を反映した平均販売単価に等しい。

(2) 計算式

加重平均は、 n 個の対象データを x_i ($i = 1, 2, \dots, n$)、それぞれのデータの重みを w_i としたとき、(式7-1) で表される。

$$\text{加重平均} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad \dots \text{ (式7-1)}$$

なお、全ての w_i が等しい場合には、 $w_i = w$ として、(式7-1) に代入すれば、

$$\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} = \frac{\sum_{i=1}^n w x_i}{\sum_{i=1}^n w} = w \frac{\sum_{i=1}^n x_i}{(n w)} = \frac{\sum_{i=1}^n x_i}{n} \text{ となり、算術平均になる。}$$

(3) 計算例

製品歩留り率の加重平均を求める。

(表7-1) は、製品の 3 日間の生産量と良品数を示す。

3 日間の製品良品率 (歩留り率) は、3 日間合計の生産量 $2,000 + 25,000 + 12,000 = 57,000$ (個) と

3 日間合計の製品良品数 $19,500 + 21,020 + 11,670 = 52,290$ を使って

3 日間合計の製品歩留り率 $= 53,390 / 57,000 = 0.917 = 91.7$ (%) と求めることができる。

一方、加重平均の考え方を使用すると、

1 日目の重み $= 20,000 / 57,000 = 0.35$, 2 日目の重み $= 25,000 / 57,000 = 0.44$, 3 日目の重み $= 12,000 / 57,000 = 0.21$

を使って、3 日間の歩留り率の加重平均 $= (0.35 \times 98.0 + 0.44 \times 84.1 + 0.21 \times 92.3) / (0.35 + 0.44 + 0.21) = 91.7$ (%)

と求めることができる (なお、ここでは、重みの合計は、 $0.35 + 0.44 + 0.21 = 1.0$ である)。

(表7-1) 加重平均

月日	生産量(個)	良品数(個)	歩留り率(%)	重み	歩留率×重み
7月1日	20,000	19,600	98.0	0.35	34.4
7月2日	25,000	21,020	84.1	0.44	36.9
7月3日	12,000	11,670	92.3	0.21	20.8
	3 日間の生産量(個)	3 日間の良品数(個)	3日間の歩留り率(%)	重みの合計	3 日間の歩留り率の加重平均(%)
	57,000	52,290	91.7	1.0	91.7

8. 移動平均

(1) 用途

移動平均 (moving Average) は、種々の要因の影響を受けて変動する時系列データ (音声データ、画像データ、金融データ、気象データ、株価^(注) など) を対象に、全体の大まかな変動に目が向くように平滑化する方法である。
時系列データを一定の区間ずらしながら、その平均を求める。

(注) 株価変動への移動平均の活用方法
株価の移動平均線は市場の動向を把握するために使用されている。
短期の移動平均線が長期の移動平均線を下から上に交差して抜けると (ゴールデンクロスと呼ぶ)、相場の上昇期入りのシグナル、逆に上から下に交差して抜けると (デッドクロスと呼ぶ)、相場の下降期入りのシグナルである、とする見方がある。

(2) 計算式

- ① 中央移動平均: ある時点前後を区間とした区間データの平均
- ② 後方移動平均: ある時点以前を区間とした区間データの平均
- ③ 前方移動平均: ある時点以降を区間とした区間データの平均

これらのなかで、現在までの過去のデータを用いた後方移動平均がよく使われる。後方移動平均の算出式を下記に示す。

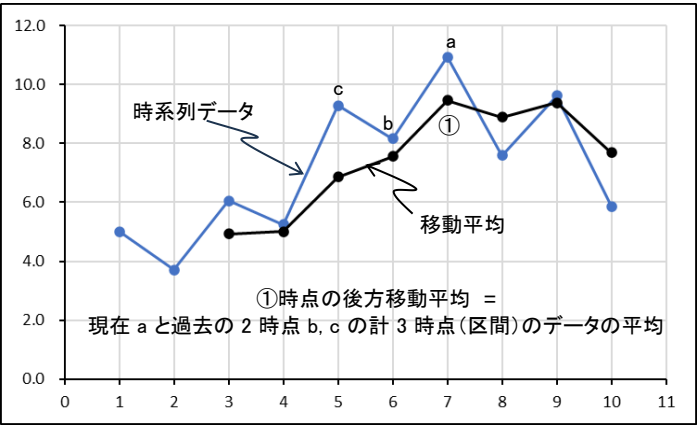
i 時点の n 区間の後方移動平均 = $(\sum_{j=i-(n-1)}^i x_j) / n \quad \dots \text{ (式8-1)}$

(3) 計算例

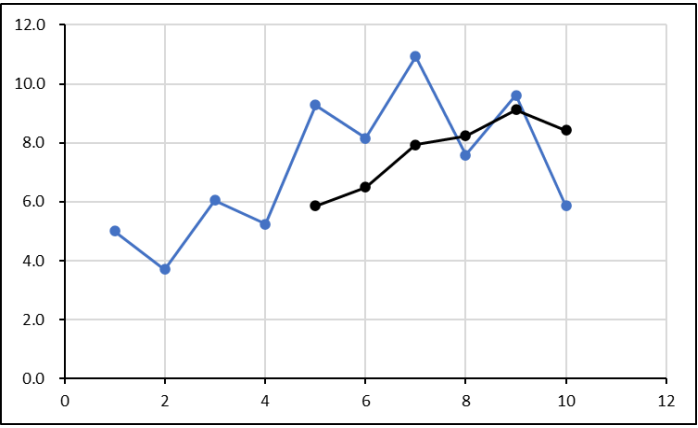
(表8-1)の時系列データを使用する。
後方移動平均 (区間 = 3) の計算結果を、(表8-1)と(図8-1)に示す。
例えば、時点 7 での後方移動平均 (区間 = 3) は、
 $(9.29 + 8.16 + 10.93) / 3 = 9.46$

(図8-1)から、移動平均によってデータ推移の平滑化が行われていることが分かる。

また、(図8-2)は、(表8-1)の時系列データの移動平均 (区間 = 5) を示す。(図8-1)の区間 = 3 に対して、(図8-2)の区間は長い。
区間を長くしたことにより、より平滑化された移動平均になっていることが分かる。また、移動平均の計算に使用するデータがより過去に遡ったものになるため、より過去のデータの影響を受けていることが分かる。



(図8-1) 移動平均の方法 (後方移動平均、区間 = 3)



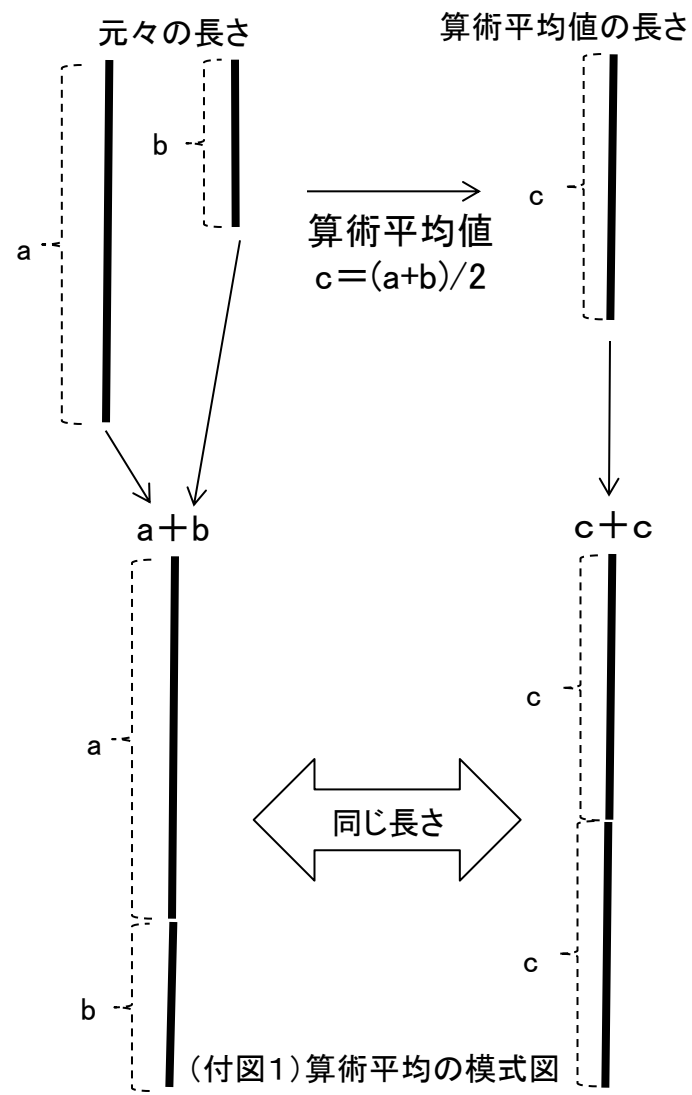
(図8-2) 移動平均の方法 (後方移動平均、区間 = 5)

(表8-1) 移動平均

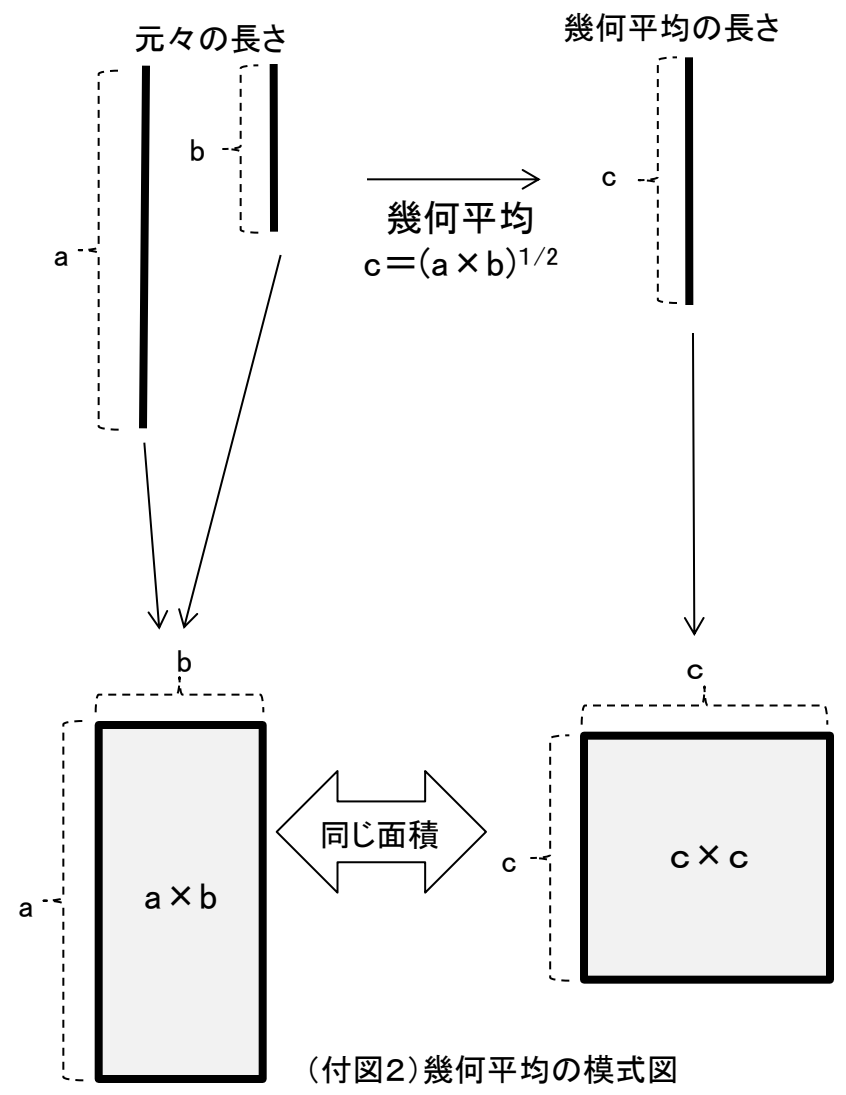
時点	1	2	3	4	5	6	7	8	9	10
時系列データ	5.01	3.70	6.06	5.24	9.29	8.16	10.93	7.58	9.62	5.86
移動平均			4.92	5.00	6.86	7.56	9.46	8.89	9.38	7.69

付録1. 算術平均と幾何平均の比較値

(1) 算術平均
2つの値の算術平均を図式化すると、(付図1)のようになる。
算術平均値を2倍した値(ここでは長さとする)は、元々の値(長さ)を2つを足し合わせた値(長さ)と等しくなる。



(2) 幾何平均
2つの値の幾何平均を図式化すると、(付図2)のようになる。
幾何平均の積(面積)は、元々の値の積(面積)と等しくなる。



参考にした書籍

西内啓：「統計学が最強の学問である(実践編)」、ダイヤモンド社

AVILE：[期待値とは？定義や性質を解説 | AVILEN](#)

数学テラス：[【いろいろな平均】算術平均・幾何平均・調和平均・加重平均の使い方を解説 | 数学テラス](#)

数学好きのすすめ：[相加平均・相乗平均・調和平均の関係 | 数学好きのすすめ](#)

遊ぶ数学：[平均値・中央値・最頻値はどう使い分ける？【3つの代表値を詳しく解説】 | 遊ぶ数学](#)

大和ネクスト銀行：[平均値だけ見ると騙される！？経済指標は「中央値」も確認しよう | マネー | おすすめコラム | 大和ネクスト銀行](#)

GMOリサーチ&AI：[外れ値の意味と求め方を解説 | 必ずしも除外することが正解とは限らない？ - GMOリサーチ&AI](#)