

データ分析で常用される4種類の平均の使い分け —算術平均・幾何平均・調和平均・平方平均—

張 興 和

目次

はじめに

- 1 算術平均以外の3種類の平均の必要性和例解
 1. 1 幾何平均
 1. 2 調和平均
 1. 3 平方平均
- 2 4種類の平均値の幾何学的解釈
 2. 1 一つの半円による4種類の平均値の視覚化
 2. 2 長方形から正方形への等価変形と平均値の対応関係
 2. 3 長方体から立方体への等価変形と平均値の対応関係
- 3 4種類の平均値の共通性と特性
 3. 1 4種類の平均値の対応関数
 3. 2 平均値の一般的定義と一般化平均
 3. 3 4種類の平均値の特性

おわりに

付録

参考文献 (五十音順)

はじめに

平均値は、小学校から親しんでいる数学の基本概念であり、日常的にも頻繁に用いられている。このため、平均値を知っているかと聞くと、誰でも「知っているよ」と自信満々に答える。しかし、平均値はどんな意味を持つのか、何のために求めるのかとさらに質問すると、答えられる人は意外と少ない。これは著者が大学での授業「数学」・「統計学」及び市民向けの講座「数学の面白さの再発見」の教育現場で経験したことである。

一方、2011年に全国で約6,000人の大学生を対象に行われた「大学生数学基本調査」では、大学生の4人に1人は「平均」の意味を正しく理解していないという結果が得られた¹。この結果は多くの人に強い危機感を抱かせたものであり、著者の経験が決して例外ではないことを証明している。

このように、平均値を十分に理解していない大学生や社会人が多数存在しているのが実状である。にも拘らず、数学や統計学の教科書には、たとえ平均値の説明が記載されていたとしても説明

1 日本数学会が2011年4月から7月にかけて国公立大48校の大学一年生約6千人を対象に実施した調査である。実施の詳細や結果は日本数学会教育委員会「第一回 大学生数学基本調査報告書」(2013.4.14)に掲載されているが、日本経済新聞(2012.2.24付)をはじめ、読売新聞、朝日新聞、毎日新聞等各全国紙やNHKで大きく取り上げられたため、社会的関心が高まった。

文が短すぎ、示された計算例だけでは平均値の本質を理解することが困難である²。一方、平均値に関する論文は、理論に偏っていたり特定部分の議論に限定されたりしており、とりわけ計算例に関してはほとんど無いに等しい³。このような論文は、平均値の更なる研究には有益ではあるものの、平均値の正確な理解に役立つとは言えない。

こうした状況のもとで、著者は一見すると簡単そうに見えるが実際は奥が深い平均値について研究し、平均値の本質を正確に理解できるような教材研究を始めた。一口に平均値と言っても、データの性質により多様な求め方があり、目的や用途に応じて相応しいものを使わなければならない。上記の「大学生数学基本調査」は算術平均に限定された問題ではあるが、算術平均以外の平均を知ることが、平均の本質の理解にとって必要不可欠であると考えられる。

平均には様々な種類があるが、本稿では、データ分析においてよく利用される算術平均、調和平均、幾何平均、平方平均という4種類の平均を取り上げる。調和平均と幾何平均は、既にそれぞれ単独で詳細に報告した⁴ので、ここでは重複する説明を省き、他の平均との関係や使い分けを中心に考察する。

抽象的な学問的理解のためには、優れた説明よりも適切な事例を示すほうが効果的である。これは、筆者が教わるときに強く実感したことであり、教える立場になってからは常に説得力のある事例作りに力を入れている。

そのため本稿では、教科書や論文の事例不足を補い、これまでの教材研究で考案した、平均値の理解に役立ちかつ面白さを感じる事例を呈示する。これらのイメージしやすい事例を通じて、ストレスを感じることなく、平均値の本質に対する理解が一段階上がることを期待する。

1 算術平均以外の3種類の平均の必要性と例解

平均値はデータ全体の中心的傾向を示す代表値⁵として、データ分析を行う際に頻繁に利用される。それぞれ個性のある数多くのデータを一つの平均値に縮約すれば、データ全体の中心的傾向を

2 例えば、岡部恒治『数学はこんなに面白い』、日本経済新聞社 (1999年)；小島寛之『完全独習統計学入門』、ダイヤモンド社 (2006年)。

3 例えば、大林昇「離散数値データの平均値について」、『日本体育大学紀要』、26巻1号 (1996年)、pp. 113-115；柳井浩「平均値の意味と構造Ⅰ、Ⅱ」、『オペレーションズ・リサーチ』1997年3-4月号、pp. 155-160、pp. 222-226；一松信、柳井浩「幾何平均と調和平均とのかくれた関係」、『メモランダム』、2000年9月号、p. 465；岐阜県立岐山高等学校「相加平均・相乗平均からの探究」、『平成13年度 岐阜県高等学校教育課程研究集会 (数学会) 発表要綱』；範春来「平均値の定義及び幾何的意義」、『中学数学教学』、1999年第6期、p. 13。

4 張興和「平均の意味と正確な計算方法に関する浅見 —調和平均の例解を中心に—」、『旭川大学経済学部紀要』、第73号 (2014年)、pp. 21-34；張興和「経済分析における幾何平均の活用 —事例を重視した教材研究を中心に—」、『旭川大学経済学部紀要』、第76号 (2017年)、pp. 19-31。

5 データ全体の中心的傾向を示す代表値は「たくさんあるデータを1つの数値で表す」というものである。たくさんあるデータを1つの数値で表すのは便利であるが、限界がある。主な代表値に平均値、中央値、最頻値があり、データが正規分布であれば、三者が一致する。しかし、非正規分布の場合は、三者が等しくなるとは限らず、データの特徴に基づき、適切な代表値を選ぶ必要がある。

データ分析で常用される4種類の平均の使い分け

知ることが可能になり、データ数が異なるデータ集合同士を比較するのに役立つ。

算術平均（相加平均）は、すべてのデータの値を足して、その和をデータ数で割った値であり、「平均値にデータ数を掛けるとデータの合計が求まる」、「偏差（個々のデータと平均値との差）の合計が常に0になる」、「偏差の2乗和は、他のいかなる一定値からの差の2乗和より必ず小さい」など重要な性質を持ち、体重や寿命、収入などの平均に利用される。

しかし、データに外れ値が含まれた場合は算術平均の代表性が弱まる⁶。特に、算術平均を用いてはいけないうえも多く存在する。例えば以下に示す3種類の平均値の問題には、算術平均を適用してはならず、以下に示す幾何平均、調和平均、平方平均に頼らざるを得ない。

【例題1】 某社の売上高は、前半期で40%減少したが後半期で50%増加したとする。この一年間で売上高は果たして平均5%増加したのか。

【例題2】 燃費が10と30km/Lの自動車を各1台所有しているとする。この2台の自動車の平均燃費は本当に20km/Lになるだろうか。

【例題3】 正方形の広場が2面あり、小さい方は辺長10m、大きい方は辺長30mであるとする。この2面の広場の平均辺長は20mで良いのか。

1.1 幾何平均

まず、1、3、9、27の4つの数の平均を考えよう。算術平均だと10であるが、これは明らかに最大値27に偏り、その4つの数の代表値としては相応しくない。

なぜ相応しくないのだろうか。この4つの数の増え方に注目すると、3倍ずつ増えていることに気づく。実は、このような幾何級数的増加（または減少）のデータの平均値を求めるには、算術平均は不適切であり、必ず幾何平均（相乗平均）を使わなければならない。

幾何平均は全データの相乗積の同次乗根で求められる。その4つの数の幾何平均は、 $\sqrt[4]{1 \times 3 \times 9 \times 27} \approx 5.2$ となる。これは、その4つの数に相応しい代表値であると考えられる。一方、 $\sqrt[4]{1 \times 3 \times 9 \times 27} = \sqrt[4]{3^0 \times 3^1 \times 3^2 \times 3^3} = 3^{\frac{0+1+2+3}{4}} = 3^{1.5} \approx 5.2$ により、幾何平均は実は指数の算術平均であることが分かる。

次に、某社の過去三年の売上高（億円）が、10、20、160であり、つまり、二年目の売上高が前年の2倍、三年目の売上高が前年の8倍だったとする。二、三年目の売上高は平均何倍であろうか。

2倍と8倍の算術平均を求めると5倍となる。しかし、一年目の売上高が10であるので、その後毎年5倍ずつ増加したとすると、三年目の売上高は $10 \times 5 \times 5 = 250$ となり、160にはならず、算出平均で求められた平均倍率は過大に見積もってしまう。

6 データの中に他の値から大きく離れた「外れ値」が含まれると、平均値は外れ値に引っ張られ、実際のデータの中心部分を示さないことがあり、その代表性が弱くなる。その場合、中央値、最頻値（データ数が多く、かつ同じ値が多い場合）、または、外れ値を除外するトリム平均（刈り込み平均、調整平均）を使う。

他方、幾何平均で計算すると、 $\sqrt{2 \times 8} = 4$ となる。平均して毎年4倍増加すると、三年目の売上高は、 $10 \times 4 \times 4 = 160$ となる。この平均倍率は与えられた問題に合致する。

直接、売上高のデータを用いて計算すると、 $\sqrt{\frac{20}{10} \times \frac{160}{20}} = \sqrt{\frac{160}{10}} = \sqrt{16} = 4$ となる。同じ結果が得られるのは当然であるが、計算式からわかるように、平均倍率の算出は、出発点と終着点のデータ(10と160)だけが利用され、中間のデータは利用されていない。

以上のように、幾何級数的変化のデータ、またはそのデータの倍率の平均値を求めるには、直接幾何平均を取れば良い。しかし、増加率(成長率、伸び率)の平均を求める場合、まずそれらの増加率を倍率に直してから幾何平均を用いて平均倍率を算出した上で、それを平均増加率に換算しなければならない。

さて、前述した【例題1】では、「前半期40%減、後半期50%増」という条件が与えられている。算術平均的な考え方だと、この一年間で平均5%増になるが、実際は以下に示すように、逆に平均約5%減になるのである。

$$G = \sqrt{0.6 \times 1.5} - 1 = 0.9487 - 1 = -0.0513 = -5.13\%$$

初期値が100だと仮定すると、【例題1】によれば、後期末(年末)の売上高が、 $100 \times (1 - 40\%) \times (1 + 50\%) = 100 \times 0.6 \times 1.5 = 90$ となり、増加はおろか初期値にも戻っていない。一方、幾何平均で算出した平均増加率-5.13%を用いて計算すると、後期末(年末)の売上高が、 $100 \times (1 - 5.13\%) \times (1 - 5.13\%) = 100 \times 0.9487 \times 0.9487 = 90$ となり、両者が一致する。

1. 2 調和平均

甲乙両地間を同じ道で往復した。往路の平均時速は60km、復路の平均時速は20kmであったとする。この場合、往復の平均時速はいくらか。

算術平均だと40kmとなるが、これは実状に合うのかを考えてみよう。両地間の距離を60kmだとすると、往路に1時間、復路に3時間かかる。往復距離が120kmに対して、往復時間が4時間なので、往復の平均時速は30kmであり、算術平均で算出した40kmは明らかに誤りである。

この場合の平均は、調和平均で算出すべきである。調和平均は、データの逆数の平均値の逆数で求めることができ、上記の例の往復の平均時速は、 $\frac{2 \times 60 \times 20}{60 + 20} = 30$ である。これは、往復距離を往復時間で割って算出した平均時速と一致する。また、平均速度は両地間の距離とは無関係である。

類似例が多くの教科書にも挙げられているが、もしこれだけの理解にとどまるならば、たとえ往復速度の平均ならば調和平均を使うことを覚えたとしても、本当に調和平均の本質的な意味を理解したと言えるのだろうか。実は、問題にある「往路60km/時間、復路20km/時間」を、速度ではな

データ分析で常用される4種類の平均の使い分け

く遅度⁷に直すと、「往路1分間/km、復路3分間/km」となる。遅度の算術平均は2分間/kmであり、時速では30kmである。調和平均速度は、算術平均遅度そのものである。

このように、走行速度（走行距離／走行時間）、購入単価（支払金額／購入数量）、人口密度（人数／面積）のような2種の量の比で構成された指標において、その指標の分子の量が同じである（分子の量を基準とする）場合、その指標の平均値は調和平均で求めなければならない。

さて、前述した【例題2】では、燃費がそれぞれ10と30km/Lの2台の自動車の平均燃費を求めている。燃費は走行距離と燃料消費量との比であり、平均燃費を求めるなら、走行距離が同じであることが前提であり、調和平均で算出すべきである⁸。

$$H = \frac{2 \times 10 \times 30}{10 + 30} = 15 \text{ (km/L)}$$

燃費が10と30km/Lの2台の自動車の平均燃費が15km/Lであり、算術平均の20km/Lより遥かに小さい。

「並列接続の平均電気抵抗」や「ドル・コスト平均法」なども、調和平均を理解するのに良い例である（これらの例に関しては前報⁹を参照されたい）。ここでは、自転車王国である中国で話題となっていた自転車タイヤ問題を紹介したい。

自転車のタイヤは、前輪よりも後輪の減りが速いのが一般的である。このため前後のタイヤを入れ替えることでタイヤの寿命を延ばすことができると考えられる。しかし、一回入れ替えるだけで、2本のタイヤがちょうど同時に寿命を迎えるようにするには、どの時点で入れ替えるべきか。ここでは実用性を無視し、単に平均値問題として考えてみる。

タイヤの寿命が、前輪500km、後輪300kmだとすると、両者の調和平均は $\frac{2 \times 500 \times 300}{500 + 300} = 375$ である。2本のタイヤの平均寿命が375kmであるため、走行距離がその半分である187.5kmに達した時点で、前後のタイヤを入れ替えれば良い¹⁰。その場合、一本のタイヤの消耗率が前輪では $187.5/500 = 37.5\%$ 、後輪では $187.5/300 = 62.5\%$ である。

1. 3 平方平均

幾何平均と調和平均は通常、正数のみしか扱えない。算術平均は0や負数を含めても扱えるが、

7 速度が移動の速さを表すに対して、遅度は速度の逆数である「時間／距離」で表し、移動の遅さを計る指標である。坂口有人<http://www.arito.jp/LecEQ16.shtml>、

ザ・グラスベクロス<https://ameblo.jp/tsuneki714/entry-11344907995.html>

8 燃費について、日本及び米国では、単位燃料量あたりの走行距離（km/L）を用いるが、欧州や中国では一定距離を走行するのに必要な燃料量（L/100km）を用いる。後者の欧州式燃費では、明確に数値が小さいほど低燃費であり、平均燃費は算術平均で算出する。日本式燃費では、逆に数値が大きいほど「低燃費」となり誤解を招きやすい。そのため「低燃費」の代わりに「燃費が良い」などの表現を使う人がいる。

9 張興和「平均の意味と正確な計算方法に関する浅見－調和平均の例解を中心に－」、『旭川大学経済学部紀要』、第73号（2014年）、pp. 21-34。

10 知乎、<https://www.zhihu.com/question/23096098/answer/116667404>

「2と-2」の場合と、「3と-3」の場合は、まったく異なるデータであるにも拘らず、算術平均は共に0であり、何の違いも生じない。

この正と負の相殺を解消するために、データを二乗してから、算術平均を求めるが、その二乗の影響を取り消すために、平方根を取る。これが平方平均¹¹である。

$$\text{「2と-2」の平方平均} \quad \sqrt{\frac{2^2 + (-2)^2}{2}} = 2$$

$$\text{「3と-3」の平方平均} \quad \sqrt{\frac{3^2 + (-3)^2}{2}} = 3$$

両方のデータの算術平均は同じく0であるが、平方平均はそれぞれ2と3である。それでは、この値の違いは一体何を意味しているのか。実は、データの算術平均が0である時に、その平方平均はそのデータの標準偏差¹²であり、後者のデータは前者よりばらつきが大きいことを表しているのである。

勿論、正と負のデータが混ざっていても、算術平均は必ずしも0ではない。しかし、データの算術平均が0ではなくても、データの偏差の算術平均は必ず0になる。「1、2、3」を例に見てみると、算術平均は2、それぞれの偏差は「-1、0、1」であり、偏差の算術平均は0となる。偏差の平方平均は $\sqrt{\frac{(-1)^2 + 0^2 + 1^2}{3}} = 0.82$ となり、これは元のデータ「1、2、3」の標準偏差であり、そのデータのばらつきを表している。

この問題はさておき、前述した【例題3】では、辺長10mと辺長30mの2面の正方形の平均辺長を求めようとしているが、その平均辺長は本当に20mになるのだろうか。

辺長の算術平均は確かに20である。しかし、問題は、決して辺長の平均を求めようとしているのではなく、面積の平均になるような辺長を求めようとしているのである。そのため、平均辺長は平方平均で算出すべきである。

$$Q = \sqrt{\frac{10^2 + 30^2}{2}} = 22.36$$

辺長10mと辺長30mの2面の正方形の平均辺長は22.36mであり、算術平均の20mより長いのである。

11 武藤真介『統計解析ハンドブック』朝倉書店(2004年) p. 30。計算式から厳密には「平方平均平方根」、「二乗平均平方根」と呼ぶべきであるが、略して「平方平均」、「二乗平均」と呼ばれている。

12 標準偏差はデータのばらつきを表す指標である。平方根を取る前の分散もデータのばらつきが表せるが、データの二乗による影響(例えば単位)が残っている。

2 4種類の平均値の幾何学的解釈

2.1 一つの半円による4種類の平均値の視覚化

平均値の大小関係を視覚的に表示するために、台形や円形等様々な表示法が考案されてきた¹³が、ここでは最もシンプルだと考えられている半円表示法を紹介する。図1は直径が $a+b$ (a と b が共に正数であり $a < b$)の半円を示す。この半円内の4本の線分の長さを用いて、4種類の平均値を表示できる¹⁴。

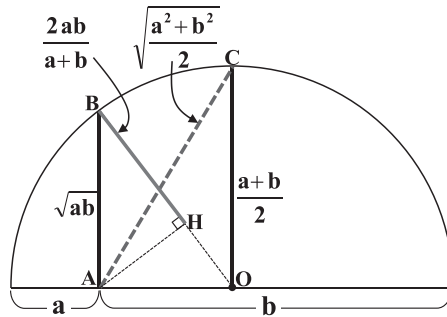


図1 半円内の線分の長さによる4種類の平均値の大きさの表示

言うまでもなく、その半円の半径 OC の長さが a と b の算術平均 $\frac{a+b}{2}$ を表している。

また、線分 a と線分 b の分界点 A から垂線を引き、その垂線と円の交点を B と置くと、垂線 AB の長さが a と b の調和平均 \sqrt{ab} を表している。

続いて、点 B と円の中心 O を結んだ線分 OB に向けて、点 A から垂線を引き、垂線と線分 OB との交点を H と置くと、線分 BH の長さが a と b の調和平均 $\frac{2ab}{a+b}$ を表している。

更に、点 A と C を結んだ線分 AC の長さが a と b の平方平均 $\sqrt{\frac{a^2+b^2}{2}}$ を示している。

線分 AB 、 BH 、 AC の長さによる幾何平均、調和平均、平方平均の大きさの表示は、三平方の定理及び相似三角形の性質より証明できる¹⁵。

上述した4種類の平均値を表す4本の線分の長さを比較すれば、4種類の平均値の大小関係が分かる。明らかに、「平方平均 \geq 算術平均 \geq 幾何平均 \geq 調和平均」が成り立ち、 $a=b$ の時のみ、4本の線分は重なり、4種類の平均が一致する。2個のデータ (a , b) の平均値の不等式の数学的証明は付録を参照されたい。

13 ネットワーク型教材データベース、<http://izumi-math.jp/sanae/MathTopic/ave/ave.htm>

14 ロボット・IT雑食日記、<https://www.yukisako.xyz/entry/average>を参考に円形を半円に改変した。

15 三平方の定理より、 $AB = \sqrt{OB^2 - OA^2} = \sqrt{\left(\frac{a+b}{2}\right)^2 - \left(\frac{b-a}{2}\right)^2} = \sqrt{ab}$ 、 $AC = \sqrt{OC^2 + OA^2} = \sqrt{\left(\frac{a+b}{2}\right)^2 + \left(\frac{b-a}{2}\right)^2} = \sqrt{\frac{a^2+b^2}{2}}$ が、また三角形 BHA と

BAOの相似性より、 $\frac{BH}{AB} = \frac{AB}{OB}$ $BH = \frac{AB^2}{OB} = \frac{ab}{(a+b)/2} = \frac{2ab}{a+b}$ が証明できる。

2. 2 長方形から正方形への等価変形と平均値の対応関係

図2に示すように、辺長がそれぞれaとbの長方形から、与えられた等価変形条件を満たす正方形に変形することは、ある種類の平均を求めることと同じである。その正方形の辺長xは、元の長方形の辺長aとbの当該平均値に等しい。

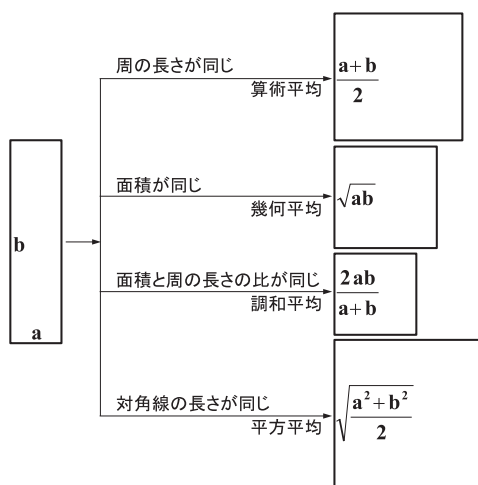


図2 長方形から正方形への等価変形条件と4種類の平均との対応関係

長方形と正方形との「周の長さが同じである」ように変形すると、 $2(a+b) = 4x$ 、 $x = \frac{a+b}{2}$ が成り立ち、正方形の辺長xは、長方形の辺長aとbの算術平均に等しい。

長方形と正方形との「面積が同じである」ように変形すると、 $ab = x^2$ 、 $x = \sqrt{ab}$ が成り立ち、正方形の辺長xは、長方形の辺長aとbの幾何平均に等しい。

長方形と正方形との「面積と周の長さの比が同じである」ように変形すると、 $\frac{ab}{2(a+b)} = \frac{x^2}{4x}$ 、 $x = \frac{2ab}{a+b}$ が成り立ち、正方形の辺長xは、長方形の辺長aとbの調和平均に等しい。

長方形と正方形との「対角線の長さが同じである」ように変形すると、 $\sqrt{a^2+b^2} = \sqrt{2}x$ 、 $x = \sqrt{\frac{a^2+b^2}{2}}$ が成り立ち、正方形の辺長xは、長方形の辺長aとbの平方平均に等しい¹⁶。

2. 3 長方体から立方体への等価変形と平均値の対応関係

同様に、辺長がそれぞれa、b、cの長方体から、与えられた等価変形条件を満たす立方体へ変形すると、その立方体の辺長xは、元の長方体の辺長a、b、cのある種類の平均値である。

長方体と立方体との「すべての辺の長さの合計が同じである」ようにする場合、 $4(a+b+c) = 12x$ 、

16 高妍麗「平均値不等式に関する検討」『山西師範大学学报』第27巻2013年12月、pp. 15-16。

データ分析で常用される4種類の平均の使い分け

$x = \frac{a+b+c}{3}$ が成り立ち、立方体の辺長 x は、長方形の辺長 a 、 b 、 c の算術平均である。

長方形と立方体との「体積が同じである」ようにする場合、 $abc = x^3$ 、 $x = \sqrt[3]{abc}$ が成り立ち、立方体の辺長 x は、長方形の辺長 a 、 b 、 c の幾何平均である。

長方形と立方体との「体積と表面積の比¹⁷が同じである」ようにする場合、 $\frac{abc}{2(ab+bc+ca)} = \frac{x^3}{6x^2}$ 、 $x = \frac{3abc}{ab+bc+ca}$ が成り立ち、立方体の辺長 x は、長方形の辺長 a 、 b 、 c の調和平均である。

長方形と立方体との「対角線の長さが同じである」ようにする場合、 $\sqrt{a^2+b^2+c^2} = \sqrt{3}x$ 、 $x = \sqrt{\frac{a^2+b^2+c^2}{3}}$ が成り立ち、立方体の辺長 x は、長方形の辺長 a 、 b 、 c の平方平均である¹⁸。

3 4種類の平均値の共通性と特性

3.1 4種類の平均値の対応関数

図3は、4種類の平均値に対応する関数を示すと共に、矢印を用いて、 $a=2$ 、 $b=8$ における a と b の平均値の計算過程を示している。

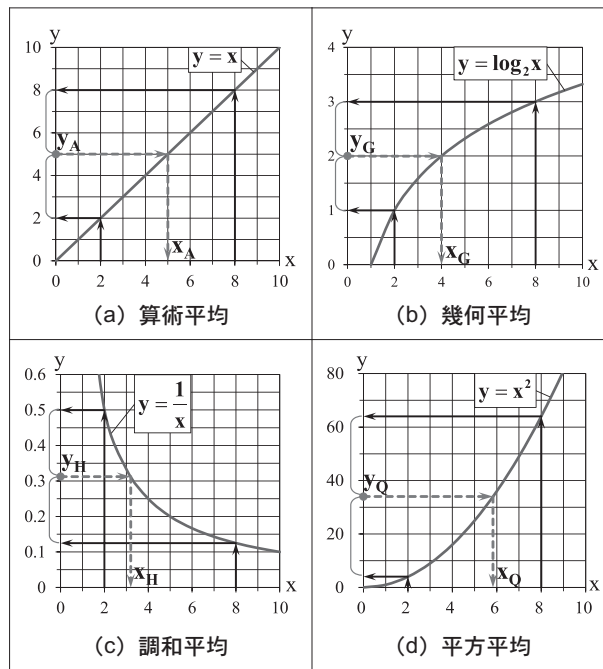


図3 4種類の平均値に関する対応関数及び2と8の平均値の計算過程

17 「体積と表面積の比」の逆数である「表面積と体積の比」、つまり単位体積当たりの表面積は、比表面積と呼ばれ、界面化学反応速度に係わる指標である。『旭川大学経済学部紀要』第73号（2014年）pp. 27-28に示したように、「粒子充填層の平均粒子径」は粒子径の調和平均で求められるが、比表面積の算術平均によるものである。

18 範春来「平均値の定義及び幾何的意義」『中学数学教学』1999年第6期、p. 13。

- (a) 算術平均には、恒等関数 $y = x$ が対応し、言うまでもなく、 a と b の算術平均 x_A は、恒等関数の算術平均 y_A と同じである。

$$x_A = y_A = \frac{a+b}{2} = \frac{2+8}{2} = 5$$

- (b) 幾何平均には、対数関数 $y = \log_2 x$ が対応し、 a と b の幾何平均 x_G は、両者の対数の算術平均 y_G の指数関数で求められる¹⁹。

$$y_G = \frac{\log_2 a + \log_2 b}{2} = \frac{\log_2 2 + \log_2 8}{2} = \frac{1+3}{2} = 2 \quad x_G = 2^{y_G} = 2^2 = 4$$

- (c) 調和平均には、逆数関数 $y = \frac{1}{x}$ が対応し、 a と b の調和平均 x_H は、両者の逆数の算術平均 y_H の逆数で求められる。

$$y_H = \frac{\frac{1}{a} + \frac{1}{b}}{2} = \frac{\frac{1}{2} + \frac{1}{8}}{2} = \frac{5}{16} \quad x_H = \frac{1}{y_H} = 3.2$$

- (d) 平方平均には、指数が2であるべき関数 $y = x^2$ が対応し、 a と b の平方平均 x_Q は、両者の二乗の算術平均 y_Q の平方根で求められる。

$$y_Q = \frac{a^2 + b^2}{2} = \frac{2^2 + 8^2}{2} = \frac{4 + 64}{2} = 34 \quad x_Q = \sqrt{y_Q} = \sqrt{34} = 5.83$$

以上示したように、4種類の平均値に対応する関数は、それぞれ恒等関数、対数関数、逆数関数、べき関数と異なるが、その関数の算術平均値の逆関数を求めることにおいては、4種類の平均値は共通である。

3. 2 平均値の一般的定義と一般化平均

4種類の平均値はその計算方法でそれぞれ定義されるのが一般的である。しかし、上述した通り、4種類の平均に対応する関数の違いはあるが、関数の算術平均の逆関数を求めることには共通性があり、4種類の平均は本質的に同じであると考えられる。ここでは、4種類の平均値を一般的に定義することを試みる。

集合データがすべてある値だと仮定したとき、そのある値を用いて決まった計算方法（以下に示す）に従って算出した値が実際のデータで算出した場合と同じになるとき、そのある値のことを当該データの平均値と呼ぶ²⁰。この定義は、各種の平均値の共通性を反映し、各種の平均値に適用できる。

$$\text{直接加法} \quad a+b = [A+A] = 2A \quad A = \frac{a+b}{2}$$

19 ここでは対数関数、指数関数の底を共に2としている。他の底でも構わない。

20 環境計画研究所、https://www.ries.co.jp/project/topic_point_of_view.htmlを参考に改変した。

データ分析で常用される4種類の平均の使い分け

対数加法²¹ $\lg(\mathbf{a}) + \lg(\mathbf{b}) = [\lg(\mathbf{G}) + \lg(\mathbf{G})] = 2\lg(\mathbf{G}) \quad \mathbf{G} = \sqrt{\mathbf{a} \cdot \mathbf{b}}$

逆数加法 $\frac{1}{\mathbf{a}} + \frac{1}{\mathbf{b}} = \left[\frac{1}{\mathbf{H}} + \frac{1}{\mathbf{H}} \right] = \frac{2}{\mathbf{H}} \quad \mathbf{H} = \frac{2\mathbf{ab}}{\mathbf{a} + \mathbf{b}}$

平方加法 $\mathbf{a}^2 + \mathbf{b}^2 = [\mathbf{Q}^2 + \mathbf{Q}^2] = 2\mathbf{Q}^2 \quad \mathbf{Q} = \sqrt{\frac{\mathbf{a}^2 + \mathbf{b}^2}{2}}$

4種類の平均をべき乗で一般化した一般化平均（p乗平均）は、次の通りである²²。

$$\mathbf{x}_p = \left(\frac{\mathbf{a}^p + \mathbf{b}^p}{2} \right)^{\frac{1}{p}}$$

この式の中で、p=1のとき算術平均が、p=2のとき平方平均が、p=-1のとき調和平均が表され、更にp→0のとき、 $\left(\frac{\mathbf{a}^p + \mathbf{b}^p}{2} \right)^{\frac{1}{p}} \rightarrow (\mathbf{ab})^{\frac{1}{2}} = \sqrt{\mathbf{ab}}$ により、幾何平均が表される²³。

4種類の平均を関数で一般化した一般化平均（f関数平均）は、次の通りである。

$$\mathbf{x}_f = \mathbf{f}^{-1}(\mathbf{y}_f) = \mathbf{f}^{-1}\left(\frac{\mathbf{f}(\mathbf{a}) + \mathbf{f}(\mathbf{b})}{2} \right)$$

恒等関数 $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ により算術平均が、対数関数 $\mathbf{f}(\mathbf{x}) = \lg(\mathbf{x})$ により幾何平均が、逆数関数 $\mathbf{f}(\mathbf{x}) = \frac{1}{\mathbf{x}}$ により調和平均が、べき関数 $\mathbf{f}(\mathbf{x}) = \mathbf{x}^2$ により平方平均がそれぞれ表される。

3. 3 4種類の平均値の特性

図4はXとYが共に0～10の値を取り、かつX+Y=10という条件を満たす場合、4種類の平均値の変化の比較を示している。

XとYが共に5（均一なデータ）である時、4種類の平均値が一致し、元のデータ5に等しい。しかし、その他のXにおいては、4種類の平均値が異なり、特に両端ではその差異が極めて大きい。

いかなるXの値においても、算術平均は常に5である。それに対して、データが均一な場合、幾何平均と調和平均は最大値、平方平均は最小値を取る。X（またはY）が5より小さくなるにつれ、幾何平均と調和平均は小さくなる。一方、X（またはY）が5より大きくなるにつれ、平方平均は大きくなる。つまり、算術平均は全ての値に、幾何平均と調和平均は小さい値に、平方平均は大きい値に、左右される。

21 対数の底は任意である。また、対数加法の代わりに、乗法を用いても良い。

22 藤井淳一、高橋真映「平均考」『数理解析研究所講究録』1452巻2005年pp. 78-85。

23 テイラー展開： $x \rightarrow 0$ 、 $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \approx 1 + x$ を利用すれば証明できる。

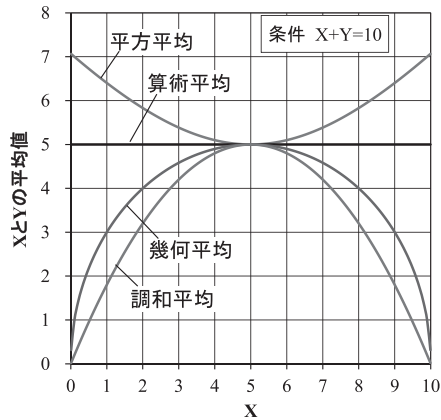


図4 XとYの4種類の平均値の変化の比較 (X+Y=10)

このことから、算術平均は「長所を取り短所を補充する」、幾何平均と調和平均は「バランスを重視し、短所を厳罰する」、平方平均は「短所を軽視し、長所を優賞する」特性を持っていると言える。ここで、この特性の利用方法を考えてみる。

表1に示したように、甲社と乙社が共に製品AとBを製造しているとする。甲社の製品Aが6点、製品Bが4点と評価されている。これに対して、乙社の製品Aが9点、製品Bが1点と評価されている。単純に比較する場合、製品Aは乙社が、製品Bは甲社が優れている。しかし、製品AとBを総合的に比較する場合は、どちらの会社が優位になるだろうか。

表1 二種類の製品の得点と総合的評価方法

	甲社	乙社	製品特徴と両社製品の優劣の総合的評価
製品A	6	9	(製品Aは乙社製が優)
製品B	4	1	(製品Bは甲社製が優)
算術平均	5	5	独立財なら算術平均で評価、両社の製品は同等
調和平均	4.8	1.8	補完財なら調和平均で評価、甲社製が優
平方平均	5.1	6.4	代替材なら平方平均で評価、乙社製が優

製品Aと製品Bを総合的に比較するには、両者の関係を考える必要がある。両者が独立財である場合は算術平均で、補完財である場合は調和平均（あるいは幾何平均）で、代替材である場合は平方平均で評価することを提案したい。なおその評価結果は、表1に示した通りである。

データ分析で常用される4種類の平均の使い分け

おわりに

平均値はデータ全体の中心的傾向を示す代表値として、データ分析を行う際に最もよく利用されている。平均値には様々な種類があるが、本稿ではデータ分析でよく利用される算術平均、幾何平均、調和平均と平方平均の4種類を取り上げた。

事例を通じて、幾何平均、調和平均、平方平均を使う必要性を明らかにした上で、それぞれの計算方法を示した。4種類の平均値の幾何学的解釈により、4種類の平均値間の大小関係を明確にし、平均値に対する理解を深めることができた。

また、平均値の一般的定義や一般化平均の数式により、平均値の共通性や本質が示された。さらに、それぞれの平均値にはそれぞれの特性があることにも言及した。4種類の平均値の使い分けを明確にし、目的に応じて適切に利用することが重要である。

計算しやすくするため、本稿では主に2つの値の平均の計算例を示した。しかし、3つ以上の値を扱う場合においても考え方は全く同じであり、Excel関数を利用すれば計算も容易である。一方、本稿では4種類の平均の単純平均のみを示したが、単純平均さえ押さえれば、重みが異なる場合の加重平均であっても対応できるだろう。付録には2つの値及びn個の値の単純平均及び加重平均の計算式、4種類の平均のExcel関数を掲載したので参照されたい。

本稿の作成にあたっては、紀要編集委員長大野成樹教授より多くの有益な助言と共に詳細な日本語校正を頂いた。この場を借りて深く感謝申し上げたい。なお、本稿での誤りは全て執筆者の責任に帰するものである。

付 録

1. 単純平均

	2個のデータ (a, b)	n個のデータ (x ₁ , x ₂ , ..., x _n)
算術平均	$A = \frac{a+b}{2}$	$A = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$
幾何平均	$G = \sqrt{ab}$	$G = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$
調和平均	$H = \frac{2ab}{a+b}$	$H = \left(\frac{1}{n} \sum_{i=1}^n x_i^{-1} \right)^{-1} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$
平方平均	$Q = \sqrt{\frac{a^2+b^2}{2}}$	$Q = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$

2. 加重平均 (重み付き平均)

	2個のデータ (a, b) 対応する重み (w _a , w _b)	n個のデータ (x ₁ , x ₂ , ..., x _n) 対応する重み (w ₁ , w ₂ , ..., w _n)
加重算術平均	$A = \frac{w_a a + w_b b}{w_a + w_b}$	$A = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i x_i = \sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} x_i$
加重幾何平均	$G = \sqrt[w_a + w_b]{a^{w_a} b^{w_b}}$	$G = \left(\prod_{i=1}^n x_i^{w_i} \right)^{\frac{1}{\sum_{i=1}^n w_i}} = \prod_{i=1}^n x_i^{\frac{w_i}{\sum_{i=1}^n w_i}}$
加重調和平均	$H = \frac{w_a + w_b}{\frac{w_a}{a} + \frac{w_b}{b}}$	$H = \left(\frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n \frac{w_i}{x_i} \right)^{-1} = \left(\sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} \frac{1}{x_i} \right)^{-1}$
加重平方平均	$Q = \sqrt{\frac{w_a a^2 + w_b b^2}{w_a + w_b}}$	$Q = \left(\frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i x_i^2 \right)^{\frac{1}{2}} = \left(\sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} x_i^2 \right)^{\frac{1}{2}}$

データ分析で常用される4種類の平均の使い分け

加重平均（重み付き平均）は、各々のデータ（観測値）に重み（ウェイト）を加味した平均である。重みのばらつきが大きければ、加重平均と単純平均の間に乖離が大きくなり、加重平均は大きい重みに引っ張られる。重みのばらつきが小さければ、加重平均と単純平均との乖離が小さくなる。特に、重みが全て同じ場合（ $w_1=w_2=\dots=w_n$ ）は、加重平均は単純平均と一致し、単純平均は加重平均の特例である。

3. 4種類の平均を求めるExcel関数

算術平均=Average ()

幾何平均=Geomean ()

調和平均=Harmean ()

平方平均=Sqrt (Sumsq ()/n)

4. 平均値不等式の証明

$$(x-y)^2 \geq 0 \text{ より、 } x^2 + y^2 \geq 2xy$$

$$x = a, \quad y = b \text{ より、 } 2(a^2 + b^2) \geq 2ab + a^2 + b^2 = (a+b)^2, \quad \sqrt{\frac{a^2 + b^2}{2}} \geq \frac{a+b}{2}$$

$$x = \sqrt{a}, \quad y = \sqrt{b} \text{ より、 } \frac{a+b}{2} \geq \sqrt{ab}$$

$$x = \frac{1}{\sqrt{a}}, \quad y = \frac{1}{\sqrt{b}} \text{ より、 } \sqrt{ab} \geq \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2ab}{a+b}$$

$$\text{よって、 } \sqrt{\frac{a^2 + b^2}{2}} \geq \frac{a+b}{2} \geq \sqrt{ab} \geq \frac{2ab}{a+b}$$

即ち、平方平均 \geq 算術平均 \geq 幾何平均 \geq 調和平均

参考文献（五十音順）

- 1) 一松信、柳井浩「幾何平均と調和平均とのかくれた関係」『メモランダム』2000年9月号p. 465。
- 2) 岡部恒治『数学はこんなに面白い』日本経済新聞社（1999年）。
- 3) 大林昇「離散数値データの平均値について」『日本体育大学紀要』26巻1号（1996年）pp. 113-115。
- 4) 岐阜県立岐山高等学校「相加平均・相乗平均からの探究」『平成13年度 岐阜県高等学校教育課程研究集会（数学部会）発表要綱』。

- 5) 小島寛之『完全独習統計学入門』ダイヤモンド社 (2006年)。
- 6) 高妍麗「平均値不等式に関する検討」『山西師範大学学报』第27巻2013年12月pp. 15-16。
- 7) 張興和「経済分析における幾何平均の活用 —事例を重視した教材研究を中心に—」『旭川大学経済学部紀要』第76号 (2017年) pp. 19-31。
- 8) 張興和「平均の意味と正確な計算方法に関する浅見 —調和平均の例解を中心に—」『旭川大学経済学部紀要』第73号 (2014年) pp. 21-34。
- 9) 範春来「平均値の定義及び幾何的意義」『中学数学教学』1999年第6期p. 13。
- 10) 藤井淳一、高橋真映「平均考」『数理解析研究所講究録』1452巻2005年pp. 78-85。
- 11) 武藤真介『統計解析ハンドブック』朝倉書店 (2004年)。
- 12) 柳井浩「平均値の意味と構造 I、II」『オペレーションズ・リサーチ』1997年3-4月号pp. 155-160、pp. 222-226。

ホームページ (五十音順)

- 1) 環境計画研究所 https://www.ries.co.jp/project/topic_point_of_view.html
- 2) ザ・グラスペコロス <https://ameblo.jp/tsuneki714/entry-11344907995.html>
- 3) 坂口有人 <http://www.arito.jp/LecEQ16.shtml>
- 4) 知乎 <https://www.zhihu.com/question/23096098/answer/116667404>
- 5) 日本経済新聞 http://www.pikkei.com/article/DGXPASDG24024_U2A220C1000000/
- 6) ネットワーク型教材データベース <http://izumi-math.jp/sanae/MathTopic/ave/ave.htm>
- 7) ロボット・IT雑食日記 <https://www.yukisako.xyz/entry/average>

要 旨

データ分析で常用される4種類の平均の使い分け
—算術平均・幾何平均・調和平均・平方平均—

張 興 和

著者の数学教育の経験と2011年に約6,000人の大学生を対象に行われた「大学生数学基本調査」の結果から、平均値を十分に理解していない大学生や社会人が多数存在していることが明らかになった。

本稿では、データ分析においてよく利用される算術平均、幾何平均、調和平均、平方平均という4種類の平均値を取り上げ、事例を通じて、各種の平均値の必要性、使い分け、計算方法を示した。

4種類の平均値の幾何学的解釈により、4種類の平均値間の大小関係が明確になった。また、平均値の一般的定義や一般化平均の数式により、共通性と特性が明らかになった。目的に応じて適切に利用されることを望んでいる。

キーワード：算術平均、相加平均、幾何平均、相乗平均、調和平均、平方平均、二乗平均、二乗平均平方根、平均

Abstract

Correct Use of Four Kinds of Mean Commonly Used in Data Analysis:
Arithmetic Mean, Geometric Mean, Harmonic Mean and Quadratic Mean

Xinghe Zhang

According to the author's teaching experience in mathematics, and the results of the "Investigation into University Students' Basics of Mathematics" involving about 6,000 people in 2011, it is clear that many of the university students and graduates do not have sufficient understanding of mean.

This paper discusses four kinds of mean commonly used in data analysis: the arithmetic mean, the geometric mean, the harmonic mean and the quadratic mean. The necessity of each kind and the distinctions between how to use them, followed by the calculation methods, are shown through specific examples.

Via geometric interpretations of the four types of mean, the inequality relations between them are clarified. Moreover, by the general definition of mean and the expression of the generalized mean, common features of every type and unique properties of each type are made explicit. The author hopes that these means can be correctly used according to the users' purposes.

Key words : Arithmetic mean; Geometric mean; Harmonic mean; Quadratic mean; Root mean square; Mean; Average