

統計計算の基礎

Y-teraya

2025 年 11 月 3 日

概要

本資料は、統計の基礎的な計算方法をまとめたものである。実験計画法を主として扱い、実務で直結する内容（Excel 関数など）を備忘録として記しておく。ほか、線形代数や微分積分も取り扱う。

また、総和 \sum の記号に苦手意識を持っている人も多いので、理解して欲しい重要な式のみ \sum 無しで表現した式も併記している。

目次

第Ⅰ部	初歩的な計算	1
1	指数 Exponent	1
1.1	累乗 Repeat multiplication	1
1.2	0 乗とマイナス乗	1
1.3	累乗根 Power root	2
第Ⅱ部	数学の記法と用語	3
2	国際単位系 International System of units	3
2.1	基本単位系 Basic units	3
2.2	接頭辞 Prefix	4
2.3	単位換算 Unit conversion	4
第Ⅲ部	統計のデータ	5
3	代表値 Average	5
3.1	平均値 Mean	5
3.2	中央値 Median	6

目次	ii
3.3 最頻値 Mode	6
3.4 最大値 Maximum・最小値 Minimum	7
4 散布度 Dispersion	7
4.1 分散 Variance・標準偏差 Standard deviation	7
4.2 範囲 Range	8
 第 IV 部 データのイメージ	 9
5 スカラー Scalar とベクトル Vector	9
5.1 ベクトルの加減法	10
5.2 単位ベクトル	10
5.3 ベクトルの成分表示	10

[青文字](#)をクリックすると、対応したページに遷移します。

留意事項

1. 色付き文字やハイライトは重要事項または強調箇所である。
2. 自身の好み（独断と偏見）で作成しているため、旧字体や座標を行列で記載している箇所がある。
3. 本資料の著作権は、[CC BY-NC-SA 4.0](#) を適応する。

第 I 部

初歩的な計算

1 指数 Exponent

1.1 累乗 Repeat multiplication

任意の^{*1} a を n 回掛けたものを a の n 乗といい a^n と書く. a を底, n を指数という. 特に, 指数が正の整数 ($n > 0$ ^{*2}) のとき累乗といい, それ以外の場合は, べき乗 Exponent という.

指数法則 Exponential laws は, 定義より自明^{*3}である. 底を a , 指数を m, n としたとき, 次の 2 法則が成立する.

$$a^m \times a^n = \underbrace{a \times a \times a \times \cdots \times a}_{m\text{個}} \cdot \underbrace{a \times a \times \cdots \times a}_{n\text{個}} = a^{m+n} \quad (1.1)$$

これは単純. a を m 回掛けたものと a を n 回掛けたものを全て乗算したら, a を $m+n$ 回掛けたものと同じになる.

$$(a^m)^n = \underbrace{a^m \times a^m \times \cdots \times a^m}_{n\text{個}} = a^{m \times n} \quad (1.2)$$

少し捉えにくいかもしれない. a を m 回掛けたものを n 回掛けたら, それは a を $m \times n$ 回乗算したものと同じになる. ただ, それだけの話.

式 (1.1), 式 (1.2) は特に暗記しなくても, 累乗がどういうものだったか? を考えれば簡単に分かる.

1.2 0 乗とマイナス乗

式 (1.1) が正しいと仮定すると, 0 乗とマイナス乗を定義できる. まず, 式 (1.1) を $n = 0$ とすると次である.

$$a^m \times a^0 = a^{m+0} = a^m \quad (1.3a)$$

a^m に a^0 を乗算しても変化しない. その数は, 1 なので次と定義される.

$$a^0 := 1 \quad (1.3b)$$

*1 「すべての」, 「どんな」, 「どれを選んでも」と同じ意味である.

*2 第 II 部 [数学の記法と用語](#) で取り扱う.

*3 一般に法則として紹介されていることが多いので, 本資料では法則として扱う. また法則とは一般に, 経験則から導かれた証明できない事象のことを指す.

式 (1.4) が正しいと仮定し、式 (1.1) を $n = -m$ とすると次である。

$$a^m \times a^{-m} = a^{m-m} = a^0 = 1 \quad (1.4a)$$

a^m に a^0 を乗算すると 1 になる。その数は、[逆数](#)^{*4}なので次と定義される。

$$a^{-m} := \frac{1}{a^m} \quad (1.4b)$$

1.3 累乗根 Power root

n 乗すると任意の数 a になる数のことを累乗根といい $\sqrt[n]{a}$ と書く。 n 乗を強調して、 **n 乗根**ともいう。また、 $n = 2$ のとき、**平方根**といい \sqrt{a} と書き、ルートの左肩の 2 は省略できる。ただし、ルートで表記するのは見ずらくなるので指数で表現したい。指数方程式を解くことで累乗根を指数で定義できる。 a を**底**、 n を**指数**と次である。

$$\sqrt[n]{a} = a^x \quad (1.5a)$$

辺々^{*5} n 乗すると、次である。

$$\left(\sqrt[n]{a}\right)^n = \left(a^x\right)^n \quad (1.5b)$$

左辺はルートが外れ、指数で表現される。また、分かりやすいように 1 乗として書くと次である。

$$a^1 = a^{nx} \quad (1.5c)$$

底が同じ方程式^{*6}は、指数も同じである^{*7}。指数だけ取り出して等式にすると、次である。

$$1 = nx \quad (1.5d)$$

よって、

$$x = \frac{1}{n} \quad (1.5e)$$

である。そのため、 n 乗根を指数で表すと次で定義される。

$$\sqrt[n]{a} := a^{\frac{1}{n}} \quad (1.5f)$$

^{*4}任意の数に**乗算すると 1 になる数**のこと。

^{*5}両辺と同じ意味。左辺（左側の式）と右辺（右側の式）の両方の式という意味である。

^{*6}等号で結ばれた式のこと。

^{*7}これが指数方程式の基本的な考え方である。

第 II 部

数学の記法と用語

2 国際単位系 International System of units

物理学 Physics に関わらず科学 Science の世界では，この単位系を用いて**数値と単位の組**を物理量という．7つの**基本単位系**を組み合わせ**組立単位**にすることで，さまざまな単位を作ることができる．組立単位を知ることによって，求めたい単位に変える**単位換算**を行うことができるようになる．

2.1 基本単位系 Basic units

基本単位は次の7つである．

表 1: 基本単位系

データの種類	英語	単位
時間	<i>time</i>	[s]
距離	<i>route</i>	[m]
電流	<i>Intensity of current</i>	[A]
質量	<i>mass</i>	[g]
絶対温度	<i>Temperture</i>	[K]
物質質量	<i>number</i>	[mol]
光度	<i>Luminous Intensity</i>	[cd]

【余談】〈基本単位系の定義〉

基本単位系には，**一義性**^aを持った基準が必要であり世界共通のものである．その**厳密な定義**を紹介する．

2.1.1 時間の定義

^aただ1つに定まること．

2.2 接頭辞 Prefix

接頭辞とは、基本単位系よりも**大きい・小さいことを表す指標**のことである。Si 接尾辞では $\times 10^{30}$ まで定まっているが、よく使われる $\times 10^{12}$ までを紹介する。

表 2: Si 接頭辞

(+) 接頭辞	英語	指数乗	(-) 接頭辞	英語	指数乗
T	<i>Tera</i>	$\times 10^{12}$	p	<i>pico</i>	$\times 10^{-12}$
G	<i>Giga</i>	$\times 10^9$	n	<i>nano</i>	$\times 10^{-9}$
M	<i>Mega</i>	$\times 10^6$	μ	<i>micro</i>	$\times 10^{-6}$
k	<i>kilo</i>	$\times 10^3$	m	<i>milli</i>	$\times 10^{-3}$
h	<i>hecto</i>	$\times 10^2$	c	<i>centi</i>	$\times 10^{-2}$
da	<i>deca</i>	$\times 10^1$	d	<i>deci</i>	$\times 10^{-1}$

2.3 単位換算 Unit conversion

単位は、**定義に基づいて組み立てることで新しい単位ができる**。基本的には、**同じ意味を持つものの同士を分母と分子に置き、約分**することで求めたい単位へと変える。次式は、「**1 日は何秒か?**」を求めたものである。恐らく、一度は計算したことがあるだろう。

$$1 [\text{day}] = 1 \cancel{[\text{day}]} \times \frac{24 \cancel{[\text{hr}]} }{1 \cancel{[\text{day}]} } \times \frac{60 \cancel{[\text{min}]} }{1 \cancel{[\text{hr}]} } \times \frac{60 [\text{sec}]}{1 \cancel{[\text{min}]} } = 86,400 [\text{sec}] \quad (2.1)$$

単位換算を行う上で押さえるポイントは、最初の単位と**求めたい単位に注目し定義としてどの物理量同士が等しいか**を考え約分する必要がある。

第 III 部

統計のデータ

3 代表値 Average

データ全体を分布中心のデータ 1 つで表したものを代表値という。主に 3 つの値のことを指し、平均値 Mean, 中央値 Median, 最頻値 Mode である。ただし、これらの値がデータの代表ではない可能性もあるため、扱うときには必ずデータの代表として機能しているのか確認する必要がある [4, 5]。

3.1 平均値 Mean

主に算術平均のことを指す。全データを合計し、データの数で割ることで求められる。平均値を \bar{x} ^{*8}, データ数を n , 各データを x_k とすると、次である^{*9}。

$$\bar{x} := \frac{1}{n} \sum_{k=1}^n x_k = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad (3.1)$$

【平均値 概要】

メリット [2] 全てのデータを考慮できる。

デメリット 外れ値に弱い。

エクセル =AVERAGE(Cell₁: Cell₂)

【余談】〈平均にもいろんな種類がある！ [1]〉

上記では、算術平均（相加平均）を取り扱った。一般に平均と言われればこの算術平均だと思えばよい。ただ、他にも場合によっては使える平均があるので紹介する。

3.1.1 幾何平均（相乗平均）

算術平均は足し算においての平均であったが、これは掛け算においての平均である。平均値を G , データ数を n , 各データを x_k とすると、次である^a。

$$G := \left(\prod_{k=1}^n x_k \right)^{\frac{1}{n}} = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} \quad (3.2)$$

^{*8} μ と置くこともある。

^{*9} \sum の計算は、基本的に $1 \leq k \leq n$ ($k \in \mathbf{Z}$) の範囲内での総和を示す記号である。 \sum の下を書いてある数字から、上に書いてある数字までをカウントアップして足したものである。

主に、成長率や増加率など**比率や割合で変化する数の平均**を算出したいときに用いる。

3.1.2 調和平均

調和平均は、一度逆数にしたデータの算術平均のことである。平均値を H 、データ数を n 、各データを x_k とすると、次である。

$$H^{-1} := \frac{1}{n} \left(\sum_{k=1}^n \frac{1}{x_k} \right) = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}{n} \quad (3.3)$$

主に、平均速度や並列接続時の抵抗、直列接続時のコンデンサなど**比率で表わされる数の平均**を算出したいときに用いる。

^a \prod の計算は \sum の足し算を**掛け算に変えたものである**、基本的に $1 \leq k \leq n$ ($k \in \mathbf{Z}$) の範囲内での**総積**を示す記号である。 \prod の下に書いてある数字から、上に書いてある数字までを**カウントアップして掛けたものである**。

3.2 中央値 Median

データを**大きさの順に並べた際の**、真ん中の値を指す。例えば、データの集合 A を次と設定する。

$$A = \{x_k | 1 \leq k \leq 5\} \quad (3.4)$$

数列 x_k が昇べきの順に並んでいるとすると、集合 A の中央値は x_3 である。

【中央値 概要】

メリット ^[2] 外れ値に強い。

デメリット 全てのデータを十分に考慮に入れることができない。

エクセル =MEDIAN(Cell₁: Cell₂)

3.3 最頻値 Mode

データの集合内で、同じデータが存在する場合がある。その同じデータが最も出現する値を指す。例えば、集合 A 、式 (1.2) に次と設定する。

$$x_2 = x_3 \quad (3.5)$$

このとき、集合 A の最頻値は x_2, x_3 である。

【最頻値 概要】

メリット ^[2] 外れ値に強い.

デメリット 1つに決まらないことがある. また, サンプルサイズが少ないと使えない.

エクセル =MODE.SNGL(Cell₁: Cell₂)

3.4 最大値 Maximum・最小値 Minimum

データの集合内で, 一番大きい値を最大値, 小さい値を最小値という. 例えば, 集合 A , 式 (1.2) を基にして考える. このとき, 最大値は x_5 , 最小値は x_1 である.

【最大値・最小値 概要】

エクセル (最大値) =MAX(Cell₁: Cell₂)

エクセル (最小値) =MIN(Cell₁: Cell₂)

4 散布度 Dispersion

4.1 分散 Variance・標準偏差 Standard deviation

各データと平均との差の平均である. 一般にデータのバラつきを表す. 分散を s^2 ^{*10}, データ数を n , 各データを x_k とすると, 次である.

$$s^2 := \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} \quad (4.1)$$

標準偏差はこの2乗を外したもののことである. すなわち, 分散 s^2 をルート $\sqrt{\quad}$ に入れたものである. そのため, 標準偏差を s ^{*11} とすると次である.

$$s := \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2} \quad (4.2)$$

【余談】〈なぜ2乗した誤差を考えるのか? ^[3]〉

平均との差であれば, 絶対値を用いれば良いと考えられる. 絶対値で計算する平均偏差と呼ばれるものを紹介する. 平均偏差は次で求められる.

^{*10} σ^2 と置くこともある.

^{*11} σ と置くこともある.

$$MAE = \frac{1}{n} \sum_{k=1}^n |x_k - \bar{x}| \quad (4.3)$$

ただし、これは数学的に扱いにくい。次に 3 点挙げる。

1. 絶対値は函数として微分可能でなく扱いづらい。
2. 場合分けが必要になることがある。
3. $\bar{x} \neq \arg \min_{\bar{x}} \sum_{k=1}^n |x_k - \bar{x}|$ である。

特に 3 つ目が問題であるが、説明は参考文献の記事に譲る。

4.2 範囲 Range

単に最大値と最小値の差である。例えば、集合 A ，式 (1.2) を基にして考える。このとき、範囲は $x_5 - x_1$ である。

第Ⅳ部

データのイメージ

5 スカラー Scalar とベクトル Vector

スカラーは**1つの数**であり、ベクトルは**2つ以上の数を束ねたもの**である。スカラーはそのまま、ベクトルは**太字 x** または **2重文字 x** (文字に余計な線を1つ入れるだけ) で表現する^{*12}。よく、スカラーは**大きさだけ**持つ量で、ベクトルは**大きさと向きも**持つ量と理解している人も多い。**それは何故か？**

$$A = 2$$

図 1: スカラー量

$$B = (2 \quad 1 \quad 3)$$

図 2: ベクトル量

スカラーのイメージは、**1次元**である。すなわち、 x 軸だけの数直線を考えると単なる**大きさ**にすぎない。次にベクトルのイメージは、**2次元**や**3次元**である。 x 軸だけでは、大きさしか表せなかったのに対し、数を束ねることで**2つ目以降の数によって方向が決まる**。

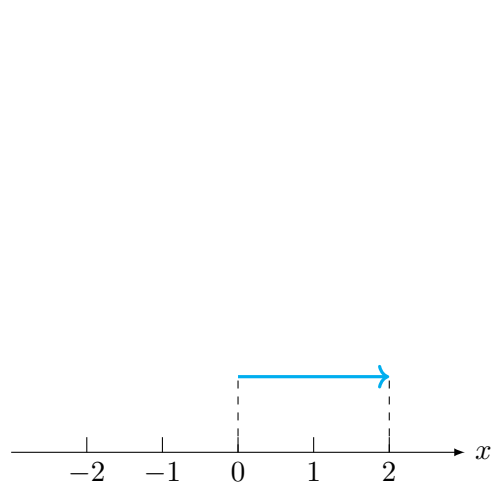


図 3: スカラーのイメージ

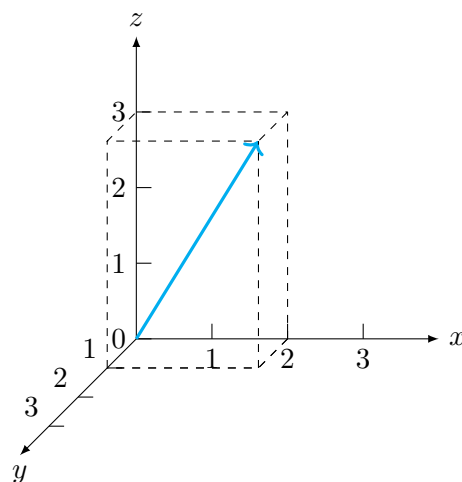


図 4: ベクトルのイメージ

簡単に説明すると、スカラーは**1つの数**であり、ベクトルは**数と数の組**^{*13}である。ベクトルは方向を表すことから、基本的に**矢印**で表現することが多い。そのとき大きさは、矢印の**長さ**で表す。

^{*12}高校までは文字の上に矢印 \vec{x} を描いてベクトルを表現した。ただし、**見づらくなりやすい**という欠点があるため、大学以降ではあまり使われない。

^{*13}厳密には、**線形空間の元**である。線形空間の元として考えると、多項式もベクトル、関数もベクトル、微分方程式の解もベクトルとして捉えられる。これらのように、この世界にはベクトルでありふれている！

5.1 ベクトルの加減法

スカラーはそのまま加減乗除できるが、ベクトルはそう上手くいかない。ベクトルの乗法には**内積**と**外積**の2種類あり、除法はできない。分かりやすい加減法から説明する。

言葉で説明すると、矢印の終点ともう一方の矢印の始点を合わせ、1つの折れ線矢印と見なして始点と終点を線で結ぶ。減法の場合は、矢印を逆にしてから足す。矢印が逆のベクトルを**逆ベクトル**という。

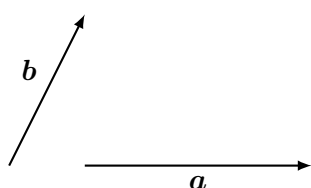
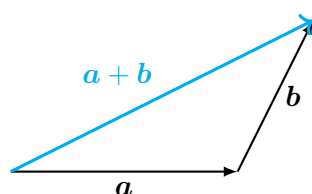
図 5: a と b 

図 6: 三角形を作る方法

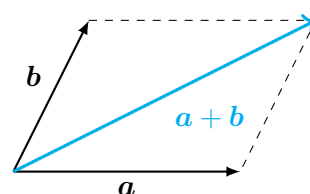


図 7: 平行四辺形を作る方法

5.2 単位ベクトル

ユークリッド空間（実数を n 個並べた全体の集合）において、3つの直交座標をそれぞれ x 軸、 y 軸、 z 軸とする。そのなかで**大きさを「1」に仕立てた**ベクトルを単位ベクトル^{*14}という。

また、 x 軸と平行な単位ベクトルを i 、 y 軸と平行な単位ベクトルを j 、 z 軸と平行な単位ベクトルを k とする。

5.3 ベクトルの成分表示

単位ベクトルと係数倍を用いて、一般にベクトルを次のような式で表せる。

$$a = Ai + Bj + Ck \quad (5.1)$$

また、係数を座標のように表して、

$$a = (A \ B \ C) \quad (5.2)$$

とも表せる。

^{*14}単位**は基本的に、**の大きさを「1」に仕立てたもののことである。

参考文献

- [1] 技術でブランドを支える ケムファク. 知っておきたい様々な平均値. <https://chem-fac.com/average>, 2024. アクセス日: 2025-11-02.
- [2] 高校数学の美しい物語. 平均値, 中央値, 最頻値の求め方といくつかの例. <https://manabitimes.jp/math/985>, 2023. アクセス日: 2025-10-28.
- [3] 数学の景色. データの分散・標準偏差の定義・具体例・性質まとめ. <https://mathlandscape.com/variance>, 2023. アクセス日: 2025-10-28.
- [4] 内田誠一. 2-1-2. データの分布と代表値. <https://mi.u-tokyo.ac.jp/consortium/pdf/2-1-2.pdf>, 2020. アクセス日: 2025-10-28.
- [5] 内田誠一. 2-1-3. 代表値の性質の違い. <https://mi.u-tokyo.ac.jp/consortium/pdf/2-1-3.pdf>, 2020. アクセス日: 2025-10-28.

索引

Average, [5](#)

Dispersion, [7](#)

Maximum, [7](#)

Mean, [5](#)

Median, [5](#), [6](#)

Minimum, [7](#)

Mode, [5](#), [6](#)

Range, [8](#)

Standard deviation, [7](#)

Variance, [7](#)

中央値, [5](#), [6](#)

代表値, [5](#)

分散, [7](#)

外れ値, [5](#)–[7](#)

平均値, [5](#), [6](#)

散布度, [7](#)

最大値, [7](#), [8](#)

最小値, [7](#), [8](#)

最頻値, [5](#), [6](#)

標準偏差, [7](#)

算術平均, [5](#)

範囲, [8](#)