統計計算の基礎

Y-teraya

2025 年 11 月 4 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 函数など）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

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.

第 I 部
数学の記法と用語

1 国際単位系 International System of units

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** Physical quantity 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** Unit conversion を行うことができるようになる．

1.1 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	$t$ime	[s]
距離	$r$oute	[m]
電流	Intensity of current	[A]
質量	$m$ass	[g]
絶対温度	Temperture	[K]
物質量	$n$umber	[mol]
光度	Luminous Intensity	[cd]

【余談】〈基本単位系の定義 [2]〉

基本単位系には，**一義性**[a] と時が経っても**べき等性**[b] 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．

1.1.1 時間の定義 [3]

原子や分子には，固有の振動数の光や電波を吸収し放射する性質がある．セシウム原子時計のマイクロ波の振動 $\Delta\nu_{\mathrm{Cs}}$ を，9,192,631,770 回数えたときを 1 秒と定義した．[Hz][c]は**単位時間あたりの振動数**のことであり，$[\mathrm{s}^{-1}]$ と同じ意味である．

1.1.2 距離の定義

真空中の光の速さ c を $299,792,458$ [m/s] と定めることによって定義した．

1.1.3 電流の定義

電気素量 e[d]を $1.602,176,634 \times 10^{-19}$ [C] と定めることによって定義した．[C] は**単位時間あたりの電気量**のことであり，$[\mathrm{A} \cdot \mathrm{s}]$ と同じ意味である．

1.1.4 質量の定義 [7]

物体の静止質量を m，光子の周波数を ν，プランク定数を $\hbar$ とすると，アインシュタインが導いた質量とエネルギーの等価性の式は次である．

$$E = mc^2 = \hbar\nu \tag{1.1}$$

ここでプランク定数 $\hbar$ を $6.626,070,15 \times 10^{-34}$ $[\mathrm{J} \cdot \mathrm{s}]$ と定義することで，周波数 ν [Hz] の光子のエネルギー E と等価な質量が m [kg] である．

1.1.5 絶対温度の定義 [5]

理想気体の場合，個々の分子は他の分子と衝突するとき以外は自由に動く．1 個の単原子の平均の運動エネルギーは，質量を m，2 乗平均速度を $\bar{v^2}$，ボルツマン定数 k，絶対温度 T とすると次である．

$$\frac{1}{2}m\bar{v^2} = \frac{3}{2}kT \tag{1.2}$$

一方で，理想気体の分子 n 個を体積 V の容器に入れた場合，その圧力 p は，質量 m と 2 乗平均速度 $\bar{v^2}$ とすると次である．

$$p = \frac{1}{3}\frac{nm}{V}\bar{v^2} \tag{1.3}$$

式 (1.2) と (1.3) より，次の理想気体の状態方程式が得られる．

$$p(T)V = nkT \tag{1.4}$$

エネルギーは $[\mathrm{J}] = [\mathrm{kg} \cdot \mathrm{m}^2\mathrm{s}^{-2}]$ で表される．ここでボルツマン定数 k を $1.380,64910^{-23}$ [J/K] と定義することで，ケルビンを定めることと同じになる．

---

[a] ただ 1 つに定まる性質のこと．

[b] 同じ操作を何度繰り返しても，同じ結果が得られる性質のこと．

[c] ヘルツは，**回数**を示す単位である．

[d] ファラデー定数（電子 1 [mol] が持つ電気量の値のこと）を F，アボガドロ定数を N_A とすると，$e = \frac{F}{N_A}$ でも求められる．

1.2 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	Tera	$\times 10^{12}$	p	pico	$\times 10^{-12}$
G	Giga	$\times 10^{9}$	n	nano	$\times 10^{-9}$
M	Mega	$\times 10^{6}$	μ	micro	$\times 10^{-6}$
k	kilo	$\times 10^{3}$	m	milli	$\times 10^{-3}$
h	hecto	$\times 10^{2}$	c	centi	$\times 10^{-2}$
da	deca	$\times 10^{1}$	d	deci	$\times 10^{-1}$

1.3 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．

$$1\ [\mathrm{day}] = 1\ \cancel{[\mathrm{day}]} \times \frac{24\ \cancel{[\mathrm{hr}]}}{1\ \cancel{[\mathrm{day}]}} \times \frac{60\ \cancel{[\mathrm{min}]}}{1\ \cancel{[\mathrm{hr}]}} \times \frac{60\ [\mathrm{sec}]}{1\ \cancel{[\mathrm{min}]}} = 86,400\ [\mathrm{sec}] \tag{1.5}$$

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**考え約分する必要がある．

1.4 ギリシャ文字 Greek alphabet

よく数式の記号でギリシャ文字が用いられるので，知っていると役に立つだろう．

表 3: ギリシャ文字

読み方		大文字	小文字	読み方		大文字	小文字
alpha	ｱﾙﾌｧ	A	α	beta	ﾍﾞｰﾀ	B	β
gamma	ｶﾞﾝﾏ	Γ	γ	delta	ﾃﾞﾙﾀ	Δ	δ
epsilon	ｲﾌﾟｼﾛﾝ	E	ε, ϵ	zeta	ｾﾞｰﾀ	Z	ζ
eta	ｲｰﾀ	H	η	theta	ｼｰﾀ	Θ	θ, ϑ
iota	ｲｵﾀ	I	ι	kappa	ｶｯﾊﾟ	K	κ
lambda	ﾗﾑﾀﾞ	Λ	λ	mu	ﾐｭｰ	M	μ
nu	ﾆｭｰ	N	ν	omicron	ｵﾐｸﾛﾝ	O	o
xi	ｸｼｰ, ｸﾞｻﾞｲ	Ξ	ξ	pi	ﾊﾟｲ	Π	π, ϖ
pho	ﾛｰ	P	ρ, ϱ	sigma	ｼｸﾞﾏ	Σ	σ, ς
tau	ﾀｳ	T	τ	upsilon	ｳﾌﾟｼﾛﾝ	Υ	υ
phi	ﾌｧｲ	Φ	ϕ, φ	chi	ｶｲ	X	χ
psi	ﾌﾟｻｲ	Ψ	ψ	omega	ｵﾒｶﾞ	Ω	ω

第 II 部
初歩的な計算

2　指数 Exponent

2.1　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[*2] と書く．a を**底** Base，n を**指数** Exponent という．特に，指数が正の整数 $(n > 0)$ のとき**累乗** Repeat multiplication といい，それ以外の場合は，**べき乗** Power 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．

$$a^m \times a^n = \underbrace{a \times a \times a \times \cdots \times a}_{\boldsymbol{m}\,\text{個}} \cdot \underbrace{a \times a \times \cdots \times a}_{\boldsymbol{n}\,\text{個}} = a^{m+n} \tag{2.1}$$

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．

$$\left(a^m\right)^n = \boldsymbol{n}\text{個}\left\{\underbrace{\begin{matrix} a & \times & a & \times & \cdots & \times & a \\ a & \times & a & \times & \cdots & \times & a \\ \vdots & & \vdots & & \ddots & & \vdots \\ a & \times & a & \times & \cdots & \times & a \end{matrix}}_{\boldsymbol{m}\,\text{個}}\right. = a^{m \times n} \tag{2.2}$$

少し捉えにくいかもしれない．式 (2.2) のように，a を m 回掛けたものを縦に n 個並べる．それは a を $m \times n$ に並べた長方形となり，$m \times n$ 回乗算したものと同じだと分かる．このように，視覚的に理解できるだろう．

式 (2.1) と (2.2) 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．

2.2　0 乗とマイナス乗

式 (2.1) 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 (2.1) を $n = 0$ とすると次である．

[*1] 「すべての」，「どんな」，「どれを選んでも」と同じ意味である．

[*2] 指数が複雑な場合，$a \exp(n)$ と表記される場合もある．

[*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．また一般に法則とは，経験則から導かれた証明できない事象のことを指す．

$$a^m \times a^0 = a^{m+0} = a^m \tag{2.3a}$$

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．

$$a^0 := 1 \tag{2.3b}$$

式 (2.3b) が正しいと仮定し，式 (2.1) を $n = -m$ とすると次である．

$$a^m \times a^{-m} = a^{m-m} = a^0 = 1 \tag{2.4a}$$

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．

$$a^{-m} := \frac{1}{a^m} \tag{2.4b}$$

2.3　累乗根 Radical root

n 乗すると任意の数 a になる数のことを累乗根 Radical root 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** Square root 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．

$$\sqrt[n]{a} = a^x \tag{2.5a}$$

辺々[*5]n 乗すると，次である．

$$\left(\sqrt[n]{a}\right)^n = \left(a^x\right)^n \tag{2.5b}$$

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．

$$a^1 = a^{nx} \tag{2.5c}$$

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．

$$1 = nx \tag{2.5d}$$

x について解き，n 乗根を指数で表すと次で定義される．

$$\sqrt[n]{a} := a^{\frac{1}{n}} \tag{2.5e}$$

[*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 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 [8, 9]．

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，次と定義*9する*10．

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

【平均値 概要】

メリット [4] 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(Cell$_1$: Cell$_2$)

【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]．

*8 μ と置くこともある．

*9 := は**左辺を右辺と定義する**と示す記号である．

*10 $\sum$ の計算は，基本的に $1 \leq k \leq n$ $(k \in \boldsymbol{Z})$ の範囲内での**総和**を示す記号である．$\sum$ 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足したもの**である．

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

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．

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

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 $1 \leq k \leq n$ $(k \in \boldsymbol{Z})$ 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．

$$A = \{x_k | 1 \leq k \leq 5\} \tag{3.4}$$

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 [4] 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(Cell$_1$: Cell$_2$)

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 (3.4) に次と設定する．

$$x_2 = x_3 \tag{3.5}$$

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 [4] 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(Cell$_1$: Cell$_2$)

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 (3.4) 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(Cell$_1$: Cell$_2$)

エクセル（最小値） =MIN(Cell$_1$: Cell$_2$)

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*11，データ数を n，各データを x_k とすると，次である．

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

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 $\sqrt{}$ に入れたものである．そのため，標準偏差を s*12とすると次である．

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

*11 σ^2 と置くこともある．

*12 σ と置くこともある．

【余談】〈なぜ 2 乗した誤差を考えるのか？ [6]〉

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．

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

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 (3.4) 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** $\boldsymbol{x}$ または **2 重文字** $\mathbb{x}$（文字に余計な線を 1 つ入れるだけ）で表現する[*13]．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？**

$$A = 2 \tag{5.1}$$

図 1: スカラー量

$$\boldsymbol{B} = \begin{pmatrix} 2 & 1 & 3 \end{pmatrix} \tag{5.2}$$

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，数を束にすることで **2 つ目以降の数によって方向が決まる**．

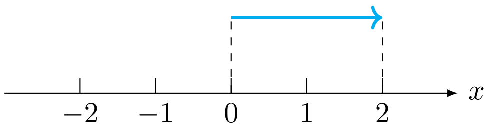

図 3: スカラーのイメージ

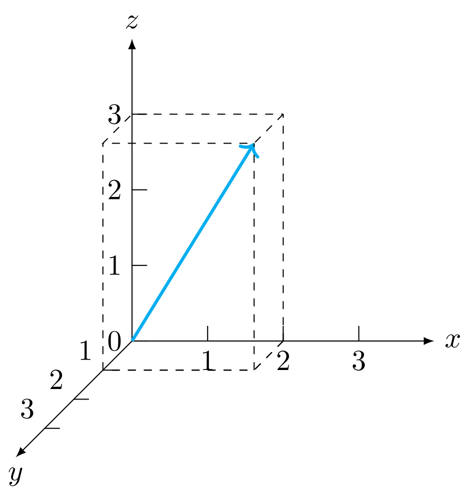

図 4: ベクトルのイメージ

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**[*14]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．

---

*13 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．

*14 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，函数もベクトル，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．

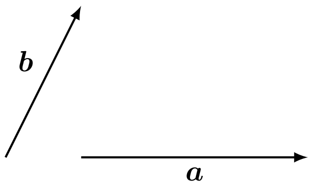

図 5: $\boldsymbol{a}$ と $\boldsymbol{b}$

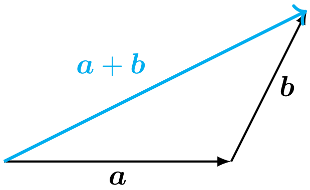

図 6: 三角形を作る方法

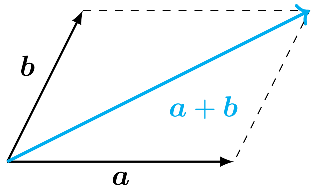

図 7: 平行四辺形を作る方法

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[*15]という．

また，x 軸と平行な単位ベクトルを $\boldsymbol{i}$，y 軸と平行な単位ベクトルを $\boldsymbol{j}$，z 軸と平行な単位ベクトルを $\boldsymbol{k}$ とする．

5.3　ベクトルの成分表示 Component form

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．

$$\boldsymbol{a} = A\boldsymbol{i} + B\boldsymbol{j} + C\boldsymbol{k} \tag{5.3}$$

また，係数を座標のように表して，

$$\boldsymbol{a} = (A \quad B \quad C) \tag{5.4}$$

とも表せる．

[*15] 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.

[2] 産総研 計量標準総合センター. 国際単位系（si））. https://unit.aist.go.jp/nmij/library/si-units.

[3] 科学技術振興機構 香取創造時空間プロジェクト. 世界を変える 1 秒の誕生. https://www.jst.go.jp/erato/katori/feature.

[4] 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.

[5] 中野享 山田善郎. 熱力学温度の単位「ケルビン」の定義改定. https://unit.aist.go.jp/nmij/public/report/si-brochure/pdf/5_SI_%E3%82%B1%E3%83%AB%E3%83%93%E3%83%B3.pdf.

[6] 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.

[7] 藤井賢一. プランク定数にもとづくキログラムの新しい定義. https://www.jstage.jst.go.jp/article/jaesjb/61/10/61_716/_pdf/-char/en.

[8] 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.

[9] 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.

索引